

# Does Defense Really Win Championships? An In-Depth Analysis of the NBA Playoffs



*Jed Jacobsohn/Getty Images*

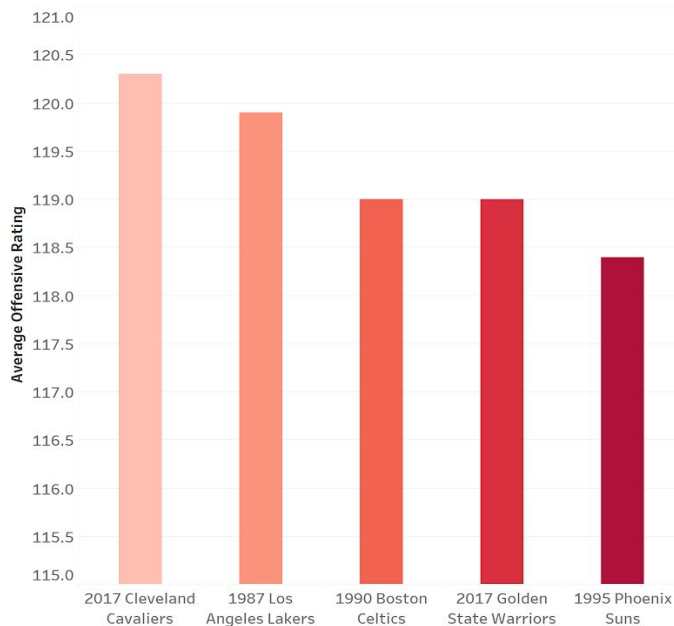
By: Donald Vu | 20 May 2018

## Introduction

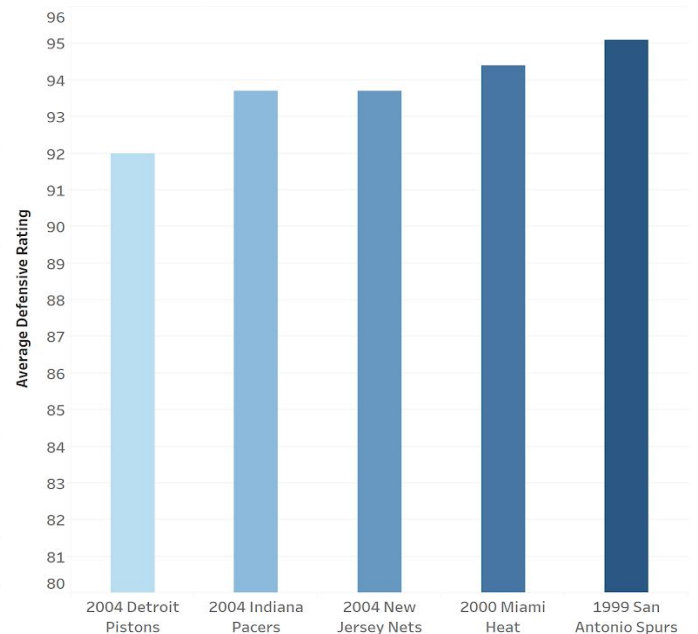
With 2 minutes left on the clock, Richard Hamilton stepped up to the free throw line. He dribbled the ball a few times and calmly knocked down both shots to seal one of the most stunning upsets in playoff history. The Detroit Pistons had defeated the heavily favored Los Angeles Lakers to become the 2004 NBA Champions. Hardly anyone could have predicted that this team would overcome the combined talents of Kobe Bryant and Shaquille O'Neal along with the brilliance of Phil Jackson. Most analysts give credit to the outstanding defense the Pistons displayed. Though

their offense just placed 7th, this Pistons team had the best ranked defense throughout the playoffs and actually had the highest average defensive rating in postseason history.

Top 5 Offensive Rated Playoff Teams



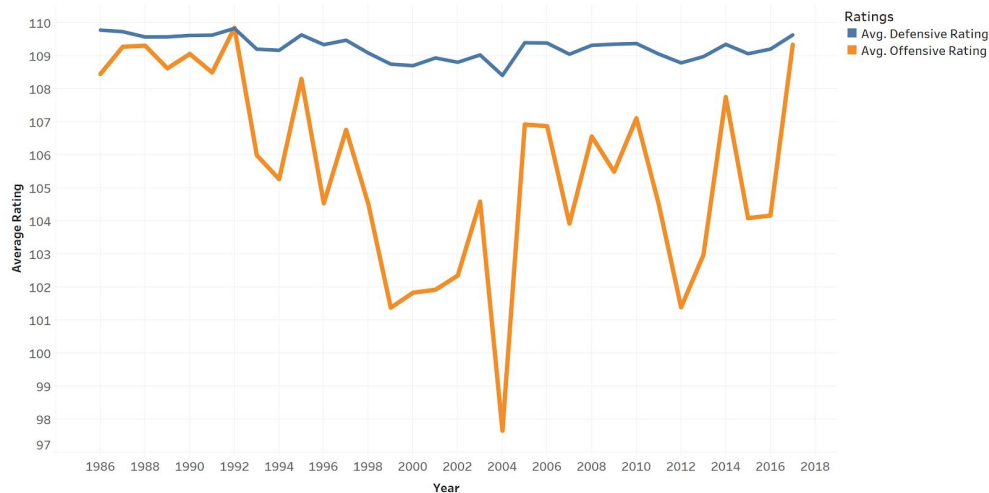
Top 5 Defensive Rated Playoff Teams



*\*Adjusted for pace  
\*Defensive rating is reverse scored*

This was largely due in part because of their defensive stalwart Ben Wallace, who averaged 19.8 rebounds, 2.7 steals, and 3.4 blocks per 100 possessions. Clearly, defense was a significant aspect of the playoffs that year, but since then, the game has started to trend in another direction. Nowadays, there has been a resurgence in robust offenses as shown by both the 2017 Warriors and Cavaliers who contributed substantially to the recent spike in average offensive rating.

Average Playoff Offensive/Defensive Rating by Year



However, even with this new style of fast-paced basketball taking over, can one say with certainty that a relentless offense is more valuable than a resilient defense? Or even, is there a way to quantify the impact of defensive prowess versus offensive tenacity on a team's postseason success? Using comprehensive data from previous playoffs, we can explore all the possible answers to these questions.

## Dataset

For this analysis, I scraped 32 seasons of playoff statistics (1986-2017) from [basketball-reference.com](https://www.basketball-reference.com) giving me 512 instances of teams to work with. The data used were the playoff averages of each statistic. The reason I excluded the regular season is because playoff basketball creates a highly competitive atmosphere. Each possession is much more valuable and players exert an exceptional level of effort when the season is on the line. Also, the goal is to measure the components of a win within the scope of the playoffs so it would be erroneous to use regular season data which includes tanking teams and high variances in player performances due to injuries and suspensions.

After extracting the information, I created a database with all the necessary offensive and defensive variables that would be viable predictors of a team's success. I determined these based on the correlation coefficients between each variable and a team's number of wins during their playoff run. All statistically insignificant predictors were removed using t-tests for individual factors and likelihood-ratio tests for joint significance tests. This step is done to avoid removing relevant variables because of multicollinearity.

The database is located in my github: <https://github.com/Donaldvu18/Playoffproject>

The predictors are described in the table below:

| Predictors                    | Description  |
|-------------------------------|--|
| Offensive Rating (offRTG)     | An estimate of points scored per 100 possessions   |
| Offensive FG % (offFG)        | Field goal percentage adjusted for 3-pointers being worth more than 2-point field goal   |
| Offensive Turnover % (offTOV) | An estimate of turnovers committed per 100 plays   |
| Offensive Rebound % (offORB)  | An estimate of available offensive rebounds grabbed which is calculated by $ORB / (ORB + Opp\ DRB)$                                  |
| Offensive FT/FGA (offFT)      | The Free Throws / Field Goals Attempted factor is a measure of both how often the team gets to the line and how often they make them |

|                              |   |
|------------------------------|---|
| Defensive Rating (defRTG)    | An estimate of points allowed per 100 possessions.  |
| Defensive FG % (defFG)       | Opponent Field Goal Percentage so lower is better   |
| Defensive TO % (defTOV)      | An estimate of opponent turnovers committed per 100 plays   |
| Defensive Rebound % (defORB) | Defensive rebound percentage calculated by DRB / (Opp ORB + DRB)  |
| Defensive FT/FGA (defFT)     | The Free Throws / Field Goals Attempted factor is a measure of both how often the opponent gets to the line and how often they make them. |
| 3-Point Attempt Rate (3PAr)  | Percentage of field goal attempts from 3-point range  |

*\*Note that defensive rating, defensive FG %, and defensive FT/FGA are correlated inversely with predicted number of wins since they measure the opponent's performance. This means the lower the value, the better a team's defensive ability.*

## Method

The objective is to create a model that can predict the total amount of playoff wins a team will achieve based on their averages for the parameters mentioned above. Note that the purpose of this model is not to predict whether a certain team will win the championship, but rather to see the trends associated with teams given their amount of wins and other variables. From these estimates, we are able to see how deep a playoff team will go, but the intention is to see how much each offensive or defensive measure contributes to an overall win count. This is better than just simply making an algorithm purely focused on win outcome because that would depend on matchups and countless other elements.

Before creating the models, I standardized and centered each predictor due to the varying magnitudes from percentages and ratings. Using forward and backwards stepwise regression, I generated the best model for each subset size based on the criteria of highest R-squared. The R-squared value for each model represents how accurately the predictors can estimate the outcome, which in this case is total playoff wins. Since we have eleven predictors, this produced eleven unique models.

| Subset Size | Model                    |
|-------------|--------------------------|
| 1           | Wins~defRTG              |
| 2           | Wins~offRTG+defRTG       |
| 3           | Wins~offRTG+defRTG+offFG |

|    |   |
|----|---|
| 4  | Wins~offRTG+defRTG+offFG+defFT  |
| 5  | Wins~offRTG+defRTG+offFG+defFG+defFT  |
| 6  | Wins~offRTG+defRTG+offFG+defFG+defFT+defTOV                                     |
| 7  | Wins~offRTG+defRTG+offFG+defFG+defFT+defTOV+defORB                              |
| 8  | Wins~offRTG+defRTG+offFG+defFG+offFT+defFT+defTOV+defORB                        |
| 9  | Wins~offRTG+defRTG+offFG+defFG+offFT+defFT+defTOV+defORB+<br>offTOV             |
| 10 | Wins~offRTG+defRTG+offFG+defFG+offFT+defFT+offTOV+defTOV+<br>offORB+defORB      |
| 11 | Wins~offRTG+defRTG+3PAr+offFG+defFG+offFT+defFT+offTOV+def<br>TOV+offORB+defORB |

The next step was to choose the optimal model out of all these options which means finding the one with the lowest mean squared error. The MSE will provide us with an idea of a model's precision since it measures how much its predictions deviate from the true observations. To acquire these MSE values, I chose to employ cross-validation methods instead of traditional criteria such as AIC, BIC, or adjusted R-squared. Cross validation is a more suitable approach since our sample size is relatively small ( $n=512$ ) compared to the amount of predictors ( $k=11$ ) which makes the error variance much harder to estimate.

I utilized a 10-folds cross validation which consists of dividing the dataset into ten groups, or folds, of equal sizes. The first fold is chosen as the validation set while the rest of the data is used to fit the proposed model. The fitted model is then tested on the first fold and we are able to find our first MSE value by comparing our estimates of each team's total wins with the actual numbers. We continue to do 10 iterations of this step with a different fold being designated as the validation set each time while all the other groups are used to train the model. Then we average the ten MSE values acquired which gives us the overall MSE of that specific model. This entire process is repeated for each subset size listed above which consequently allows us to narrow down our pool to a single quintessential model.

## Final Model

### Model:

Wins ~ offRTG + defRTG + offFG + defFG + defFT + defTOV + defORB

### Coefficients:

|             |           |            |           |            |
|-------------|-----------|------------|-----------|------------|
| (Intercept) | offRTG    | defRTG     | offFG     | defFT      |
| 4.8320312   | 2.4010377 | -2.1930858 | 0.6824176 | -0.4028614 |
| defTOV      | defORB    | defFG      |           |            |
| 0.3451101   | 0.3298418 | -0.6061492 |           |            |

**Prediction Error:**  $\pm 2.887163$  wins

## Results

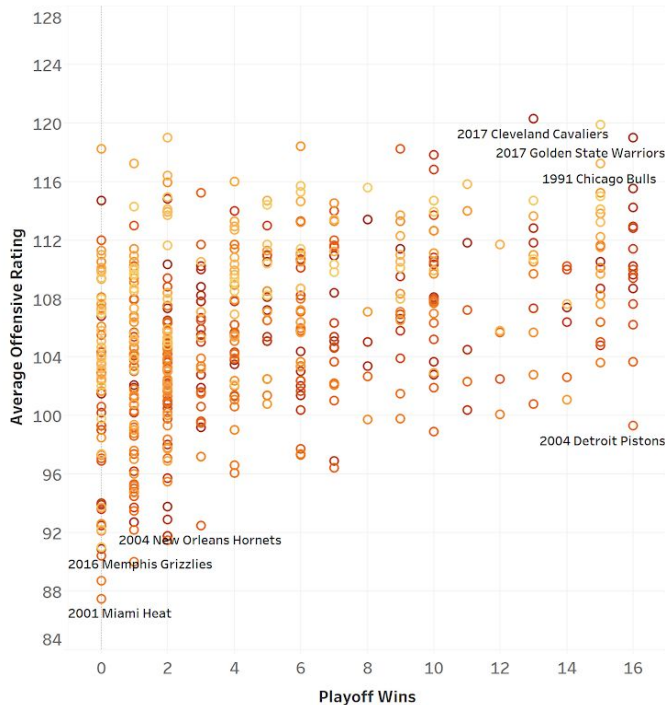
Since we have standardized and centered our predictor values, there are certain ways to interpret these numbers. The intercept value represents the base number of wins that all teams start off with and then it will increase or decrease depending on the values of the parameters. For example, keeping all else constant (fancy way of saying keep all other variables at their average value), for every increase in one standard deviation of a team's offensive rating, the amount of predicted wins increases by 2.4010377. Conversely, for every increase in one standard deviation of a team's defensive rating, the amount of predicted wins decreases by -2.1930858 (remember that defensive rating is inversely correlated).

Since I am just looking to compare the magnitude of offensive measures versus defensive ones, I will interpret the results in terms of standard deviation. However, if you want to work with concrete numbers, you can find the mean and standard deviation of each predictor by referencing back to the excel database in my github. The number of total wins remains unscaled so it can be construed normally.

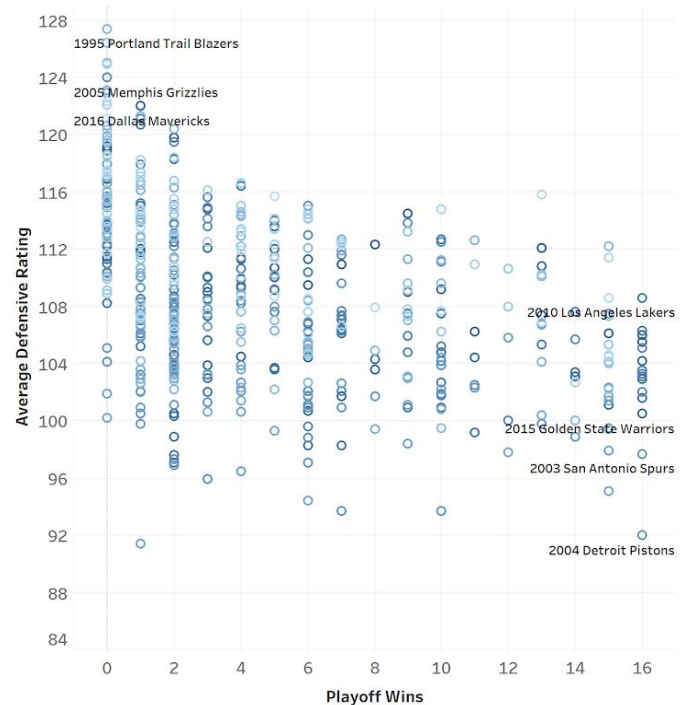


Interestingly, our model gives the edge to the defensive side of the ball by incorporating five defensive components in its algorithm versus two offensive ones. However, the strongest indicators are offensive and defensive rating with the former having the edge. The correlation is illustrated in the graphic below where one can visualize the apparent trends.

Average Playoff Offensive Rating vs Playoff Wins



Average Playoff Defensive Rating vs Playoff Wins



Overall, the results of the model dictate that all the various mechanics of a team's defensive scheme play a pivotal role in achieving wins, while on the offensive side, the main factor is if your shots are falling through the net or not.

There were four variables left out of the model as follows:

**Offensive rebound percentage-** excluded from the model probably because there are not enough instances of it in a game to impact a team's total wins.

**Offensive FG/FTA-** usually most teams do not have to rely on getting to the line often to achieve a win which could explain why offensive FG/FTA was left out of the equation. However, if a team's opponents are getting an abundance of calls from the referees, it can shift the game outcome which explains why defensive FG/FTA is used in the model still.

An example of this would be teams facing against James Harden who notoriously goes to the free throw line with high-frequency.

**Offensive Turnover Percentage-** a possible explanation for this predictor being omitted from the model is that most playoff teams are usually skilled at taking care of the ball so the offensive turnover percentages may be similar across the board.

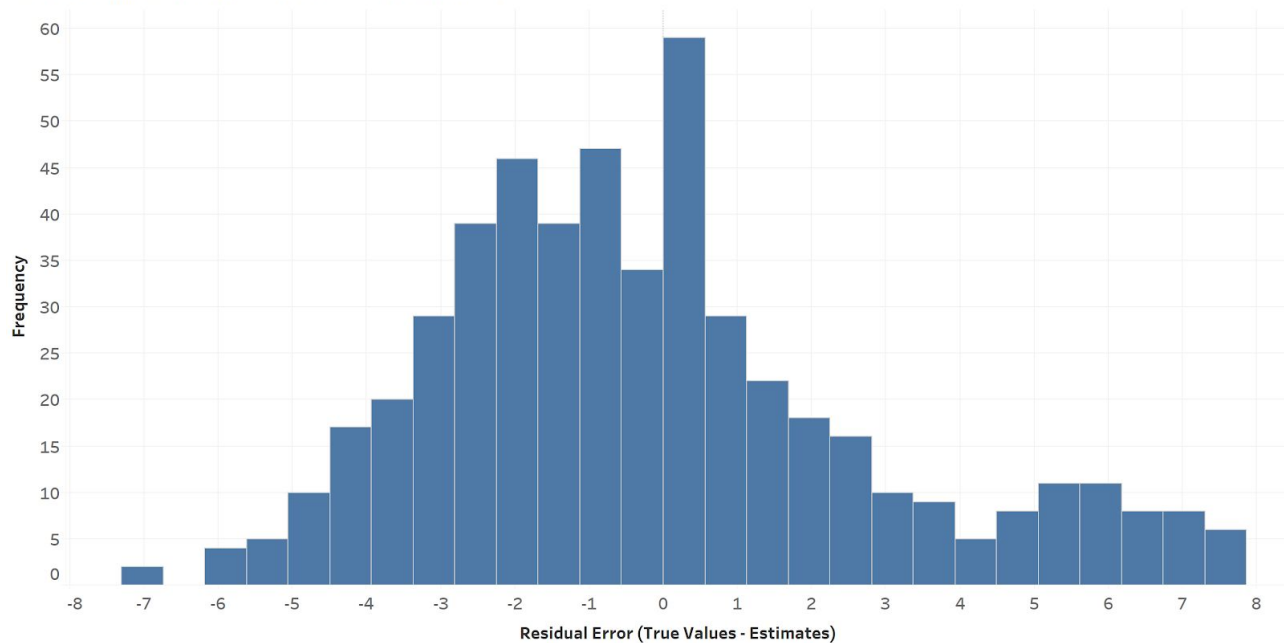
**3-Point Attempt Rate-** when considering all the playoff teams, the amount of 3-pointers attempted do not differ greatly among them. You might have a few exceptions like the recent Golden State Warriors and Houston Rockets but those are not enough to warrant the inclusion of the predictor in the model.

Let's take a glance at the predictive aspect of our model. At first, it may seem like a prediction error of 2.887163 wins is fairly high. But one has to consider that the model consist of handpicked predictors that would best showcase the influence of offensive and defensive traits. Since the primary objective was to analyze these effects, other predictors that could have refined the algorithm's prediction ability like a team's seed or number of all-stars were excluded to reduce noise. Having said that, our estimates are still in close vicinity to the actual number of wins. With the standard error being 2.887163 wins, it gives our estimates some breathing room while not forfeiting too much precision. It also helps that the playoffs are set up in a best-of-seven format because this way, the better team will usually move on and have the opportunity to add to their win total thus lowering the variance. This is in contrast to the NFL where each playoff round consists of a single game and can lead to highly unexpected results.

Though the model heavily relies on the accuracy of its predictors, it may also use a bit of luck. For example, it predicted that the 2015 Clippers would reach a total win count of seven in the playoffs. Though the model was correct with that estimate, it is far-fetched to believe the algorithm could have anticipated that the Clippers would infamously collapse and blow a 3-1 lead against the Rockets.



Distribution of Residual Errors for Predictions



The histogram above displays the distribution of our model's residual error rounded to the nearest .5. If the estimate was a negative number, I converted it to a zero which explains the spike around the 0 to 1 interval. As depicted by the graph, a majority of the residual errors fall within the range of -3 to 3.

Our model does have its weaknesses. For example, consider the 2016 San Antonio Spurs who finished their season with a 67–15 record, their best winning percentage in franchise history. With a roster comprised of decorated all-stars such as Tim Duncan, Tony Parker, Kawhi Leonard, and LaMarcus Aldridge, this Spurs team was expected to compete head-to-head with the defending champion Warriors. When we plug their average statistics into the model, the Spurs are predicted to have 11.78863 wins which is equivalent to reaching the championship finals. However, they were defeated in the conference semifinals by the OKC Thunder led by spectacular performances from future MVP Russell Westbrook and scoring champion Kevin Durant. The Spurs ended up with a total of 6 wins which deviates substantially from our estimate when you consider the model's error range.

As expected, our algorithm runs into problems when players perform well above their expectations or below them, but it is difficult to foresee these anomalies since each game is treated as an independent event.

Though these results do provide some insight into our research question, we can make things a bit more interesting. Instead of generalizing our models to the entire postseason, let's customize our model on a round-to-round basis. Since each playoff round is based on seedings, this will produce higher skilled matchups in the later stages which may lead to unique insights.

## Further Methods

I used the same methods from before, but instead of sampling the entire pool of playoff teams, I only used the teams that advanced into the next round to compute the respective model. I decided to skip the initial round just because of the high variation of statistics due to one-sided matchups (1st vs 8th seed, 2nd vs 7th, etc...)

## Further Findings

### Conference Semifinals



YONG KIM / File Photograph

## Model:

Wins  $\sim$  offRTG + defRTG + defFT

## Coefficients:

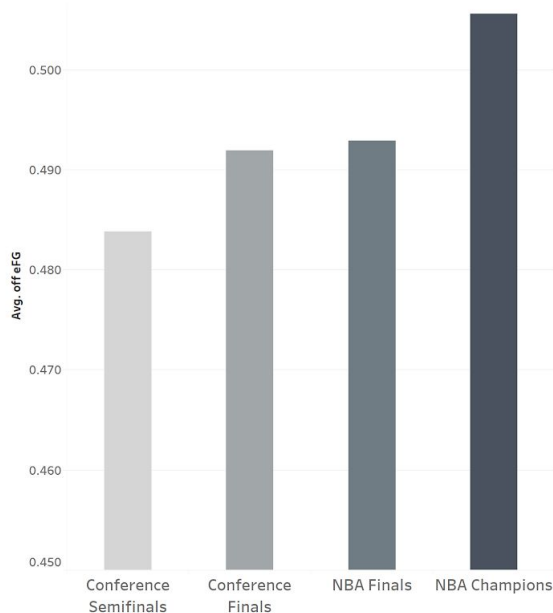
| (Intercept) | offRTG    | defRTG     | offFG      |
|-------------|-----------|------------|------------|
| 5.4645024   | 3.7962900 | -3.8757600 | -0.3886062 |

**Prediction Error:**  $\pm 2.779325$  wins

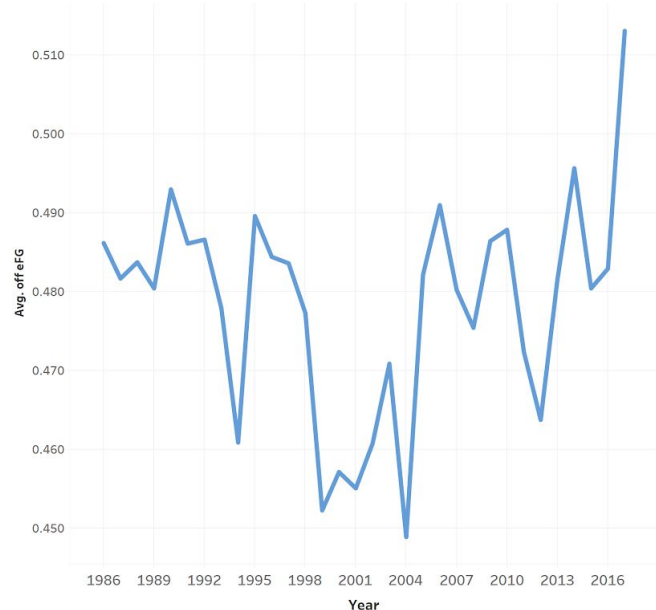
For this model, I sampled all the playoff teams that made it into their conference semifinals.

Yikes. Only three predictors remain out of the eleven that were available. Offensive and defensive ratings continue to have a profound impact on wins as expected, but offensive FG was the only other variable to be included.

Average FG Percentage by Playoff Finish



Average Playoff FG Percentage by Year



As illustrated above, there is a strong correlation between field goal percentage and a team's playoff run. Overall, this model confirms the importance of a team's shot selection

but does not give us too much to interpret other than that which may be due to the numerous one-sided matchups that are still skewing the results.

## Conference Finals



Image via Air Alamo

### Model:

Wins  $\sim$  offRTG + defRTG + defFT + defTOV + defFG

### Coefficients

| (Intercept) | offRTG    | defRTG     | defFT      | defTOV    | defFG      |
|-------------|-----------|------------|------------|-----------|------------|
| 8.8069458   | 2.3421628 | -2.5706374 | -0.4991053 | 0.1828533 | -0.1739399 |

**Prediction Error:**  $\pm 2.465866$  wins

This model will help choose what factors help teams in the conference finals advance to the championship finals. The sample size only includes teams that made it into their conference finals.



Almost all the defensive measures had a significant role in a team's wins while only one offensive factor was deemed notable. Conference final matchups involve the top two teams within each conference battling it out and usually these lineups consist of superstars who can score at will. As a result, the different offensive factors are not as remarkable since teams will usually find a way to attain points. However, defensive effort can vary for each team depending on the roster assembled and coach's personal style. Defensive plays such as forcing a turnover or a clutch block can swing the outcome of a tightly contested game.

Also, remember that in our overall model, offensive rating was the strongest predictors of wins. However, defensive rating has now surpassed it by about .20 which is a sizable difference. This further supports the notion that defensive grit is critical in these middle playoff rounds.

## NBA Finals



Associated Press

## Model:

Wins  $\sim$  offRTG + defRTG + 3PAr + defFG

## Coefficients

| (Intercept) | offRTG    | defRTG     | 3PAr      | defFG      |
|-------------|-----------|------------|-----------|------------|
| 12.1405979  | 1.3797300 | -1.3670177 | 0.3059395 | -0.4518510 |

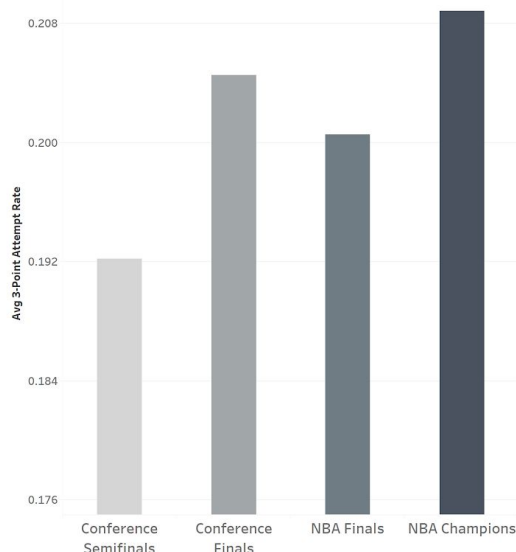
**Prediction Error:**  $\pm 1.349233$  wins

This will help us choose what factors help teams move on from the finals to become the NBA champions. Only teams that have gone to the finals were sampled.

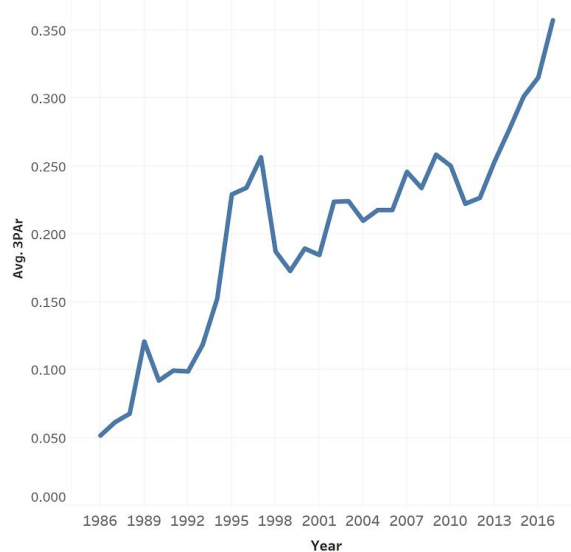
For this model, we have the usual suspects but also a few surprises. It makes sense that when the best two teams go head to head, whoever can defensively lock down their opponents will most likely get the win.

However, it is interesting to see that the 3-point attempt rate made the cut as a predictor in this model while other variables did not. I believe this is because when a team is hustling and outplaying their opponents, they are producing more open looks which correlates with the amount of 3 pointers being taken. In this case, the 3-point attempt rate would be a compelling indicator of which team is controlling the flow of the game.

Average 3 Point Attempt Rate by Playoff Finish



Average Playoff 3-Point Attempt Rate by Year



Of course there are other plausible explanations such as a team adopting an offensive strategy based on 3 pointers or individual sharpshooters having hot hands. Also, note that offensive rating has reclaimed the top coefficient over defensive rating.

## Conclusion

Depending on which playoff round it is, there are certain team attributes that are emphasized more than others. However, for the most part, defensive abilities have the potential to make more of an overall impact on an outcome even though the amount of points scored is still usually the highest weighted determinant. In the conference semifinals and conference finals, defensive potency is valued more than offensive power which demonstrates the strong leverage of defensive plays in the middle rounds. For the finals, it does shift back to an offensive based model that takes the amount of 3 pointers attempted into consideration.

Though all these statistics and equations can tell you about the various aspects that go into winning a playoff game, it is not an absolute indicator of what a team should focus on. The data used in this research goes back 32 seasons when a distinct style of basketball was being played and since then, the game has evolved immensely. Also, there are factors that one cannot quantify or account for such as the greatness of LeBron James or Draymond Green's ability to earn himself technicals. Nevertheless, these models can still provide us insight on the offensive and defensive patterns associated with successful playoff teams. In the end, these are just analytics and nothing will beat the good old-fashioned way of playing your heart out and leaving it all on the floor.