Does Having More Players Scoring Greater than 15 Points Per Game Benefit NBA Teams' Success?

Devin Nordlund

2024-04-26

0.1 Abstract

This study examines the relationship between the number of NBA players scoring over 15 points per game on each team and team success, measured using playoff appearances and wins. Using a dataset of NBA team and player statistics from 1980 to 2023, we analyzed the relationships present using ARIMA modeling, logistic regression, and correlation analysis. Contrary to what was expected, there is not a significant association between the count of players scoring over 15 points per game and playoff appearances or wins. These findings suggest that players scoring over 15 points per game is not impactful on team success, with many other factors driving team success. Scoring more points typically leads to successful outcomes, but there are many factors other than individual scoring that contribute to winning in the NBA. Further research is required to understand the variables that have the greatest contribution to winning, and if a higher point threshold would have a greater association with successful outcomes.

0.1.1 Keywords

NBA, Basketball, Team Success, Scoring, Player Performance

0.2 Introduction

0.2.0.1 Overview In sports, data analytics has been a growing part of the way teams operate. Teams have been leveraging a variety of analytics to help make informed decisions. The market for sports analytics has increased significantly since 2010. In the NBA, the game of basketball has evolved over time with a shift toward three point shooting and higher scoring outputs. Scoring points is one of the most important keys to winning basketball games, and understanding the relationship between higher scoring output from multiple players could be essential to making in-game decisions and team construction. Individual player performance is a key component of increasing overall team points (Choo et al. 2020), which subsequently can often lead to winning and successful outcomes. For a professional sports team, the main goal is to win a championship. This can be achieved by first having enough regular season success to make a playoff appearance. The NBA regular season format consists of 82 games. The playoffs currently include 16 of the 30 teams in the NBA, with the top eight teams from each of the Eastern and Western Conferences. The best regular season teams are not always the ultimate winners of the championship, so many teams are vying for a playoff spot for the opportunity. Team executives have used many strategies in team-building, with varying results. Some examples of this include "superteams", in which a team has at least three perennial superstars on the team. In a league that is typically offensively oriented, we would expect a positive correlation between the number of players scoring above 15 points per game and the corresponding team success metrics. A higher output from more players would appear to be signs of a more cohesive team, with multiple options to score points and ultimately win games in the face of different strategies.

0.2.0.2 Existing Works The article "Data-Driven Approaches to NBA Team Evaluation and Building" (Choo et al. 2020) states that teams at "different stages of the success cycle and with different resources, will use these strategies differently". The more successful teams at a point in time may allocate their resources towards maintaining current success, which may involve trading away draft picks or younger, unproven players, for proven players in the league who can benefit the team's current performance. This article also discussed an evaluation of player performance based on a tree model, with the possible outcomes of a possession being taken into consideration. This was used to gather an estimated impact on expected team points. In this article, the cohesion of a roster is also mentioned to have a positive impact on winning percentage.

A 2022 study (Cabarkapa et al. 2022) focused on finding the variables that made the largest differences in winning and losing games. There was a statistically significant difference found between scoring in wins versus losses. This was true both in the regular season and the postseason, but the difference was larger in the postseason, which could be attributed to the more competitive nature of the games. Shooting efficiently was also found to lead to more winning outcomes. In this study, the winning teams were found to have made more shots with higher shooting percentages on all field goals, three-pointers, and free throws. Shooting the ball more efficiently has been shown to lead to more shots made, points scored, and more wins. In the case of three-pointers, teams that shoot them efficiently can score more efficiently, with more points scored on less possessions and attempts. This aligns with the notion that shooting efficiency is key to determining game outcomes.

A 2020 study (Terner and Franks 2021) claims that scoring ability is the most important part of player performance. This not only includes skill, but the shot selection of the player. Shot selection takes many things into account, such as the defensive pressure, situational factors, teammate ability, and the specific skillset of the player. Players will typically choose shots that align with their skillset, with an approximately linear relationship between the log relative frequency of two-point and three-point shot attempts. Coaching and team strategy also comes into play here, with good coaching being able to give players better opportunities to take good shots. A 2010 study (Lorenzo et al. 2010) corroborates the importance of scoring to winning and team success. This study was focused on the under-16 male basketball games. These players were not professionals, but similarities were found between their play and that of NBA players. A discriminant analysis was used to find the most important variables in final outcomes. For games that were decided by a margin between 10 and 29 points and those decided by more than 30 points, the main discriminants were successful two-point field goals. At a younger age, the game is played differently, but the same principles can be applied to all levels of basketball. Making more shots and subsequently scoring more points is shown to be related with winning games.

0.2.0.3 Gap in Research These findings demonstrate the importance of individual and team scoring to winning and overall team success. The gap in research lies in the exact team scoring structure and how this relates to winning. Existing literature has explored the relationships between scoring efficiency, individual player performance, and team success in basketball. There remains a gap in understanding how having more players scoring greater than 15 points per game impacts team success. The distribution of scoring output across players on teams in the NBA has not been studied in depth. This is necessary to address because basketball is a team sport, and individual starpower is not enough to make a team successful. As three-point shooting has become more prevalent in the NBA, the total scoring outputs have increased, and the need for more high-scoring players has risen. The teams with more players consistently scoring 20 points per game may have an advantage over more one-dimensional teams, with only one or two players scoring most of the points. With the increasing usage of data analytics in basketball and sports as a whole, teams are looking to find an advantage in any way possible. Understanding the impacts of having multiple high-scoring players on team success can benefit teams by informing their strategies on the court and in team construction in order to maximize their chances at winning games, making playoff appearances, and winning championships.

0.2.0.4 Contributions This study aims to fill existing gaps in literature by examining the relationships between the number of players scoring greater than 15 points per game and how successful NBA teams are. The information gained from this study will have practical applications in the sport of basketball, by informing team decisions and coaching strategies, leading to a greater competitive balance in the NBA.

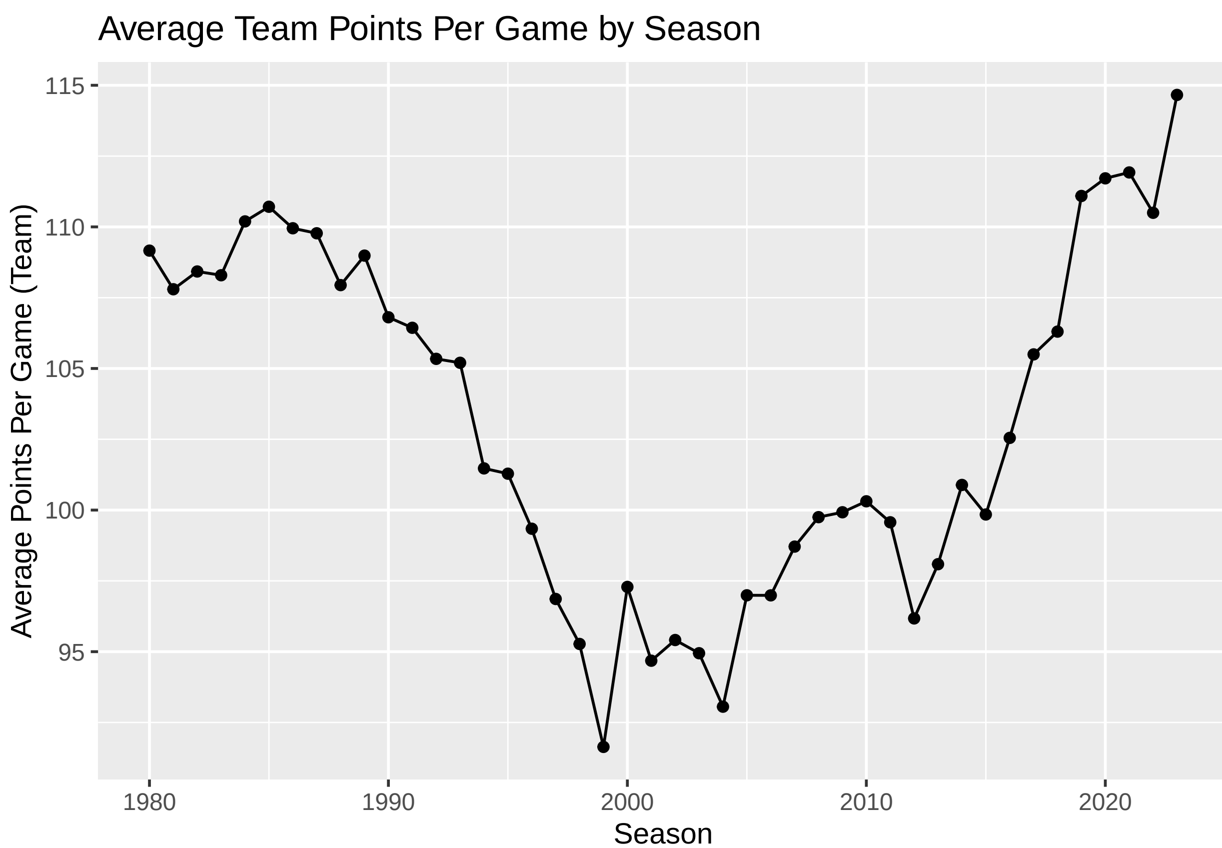
0.2.0.5 Summary Following this section, the details of the dataset will be explained. This will discuss the sourcing of the data, data preprocessing, sampling techniques used, and the usefulness of the data to the field of basketball. Exploratory data analysis is done using visualizations of the overall scoring trends in the NBA. The average points scored per game over time is shown, and a comparative box plot is used to show the relationship between average team points scored and making the playoffs. Then, the regression analysis techniques will be discussed in detail, using multiple regression to model the relationship between the number of high-scoring players and winning outcomes.

0.3 Data

The dataset used in this study is sourced from Kaggle, compiled by Sumitro Datta. The data contains observations spanning from 1947 to 2024 across the NBA and ABA, but for ease of use and validity of data, we will only be including observations from 1980 to 2023. The ABA was a competing league with the NBA from 1967-1976, but then merged with the NBA in 1976. In 1980, the three-point shot was introduced, and the game style corresponds much more with the game played today. This dataset contains exhaustive information about player and team statistics per game in professional basketball, as well as end of season summaries which provides insights into player performance and team success. There are 35 total variables included in the table of player statistics, however the variables of interest from this include: points per game, team, and year, as the goal is to calculate the count of players on each team scoring 15 points per game or more each year. The player statistics table includes observations from 5,153 players. The team statistics table contains 1681 team observations, which is due to the fluctuation in league format and team count over time. The variables of interest from this table are team points per game, and year. The team summaries table contains information about end of season success, and the variables of interest from this table include playoffs (boolean), wins, and year. In order to prepare this data for the study, the outlined variables of interest will be combined into one table, with the player and corresponding team statistics organized by year. Using the comprehensive statistics from a wide span of years will allow us to view the impact of having multiple players scoring 15 points across multiple eras of basketball.

0.4 Exploratory Analysis

The exploratory analysis of this data has shown an increase in team scoring since the early 2000s. The San Antonio Spurs, Los Angeles Lakers, and Portland Trail Blazers have the three most total playoff appearances since 1980. It has been demonstrated that teams that make the playoffs tend to score more points per game on average than those that do not. The average number of players scoring greater than 15 points per game has historically followed similar patterns to those of the overall team scoring output, increasing in recent years.



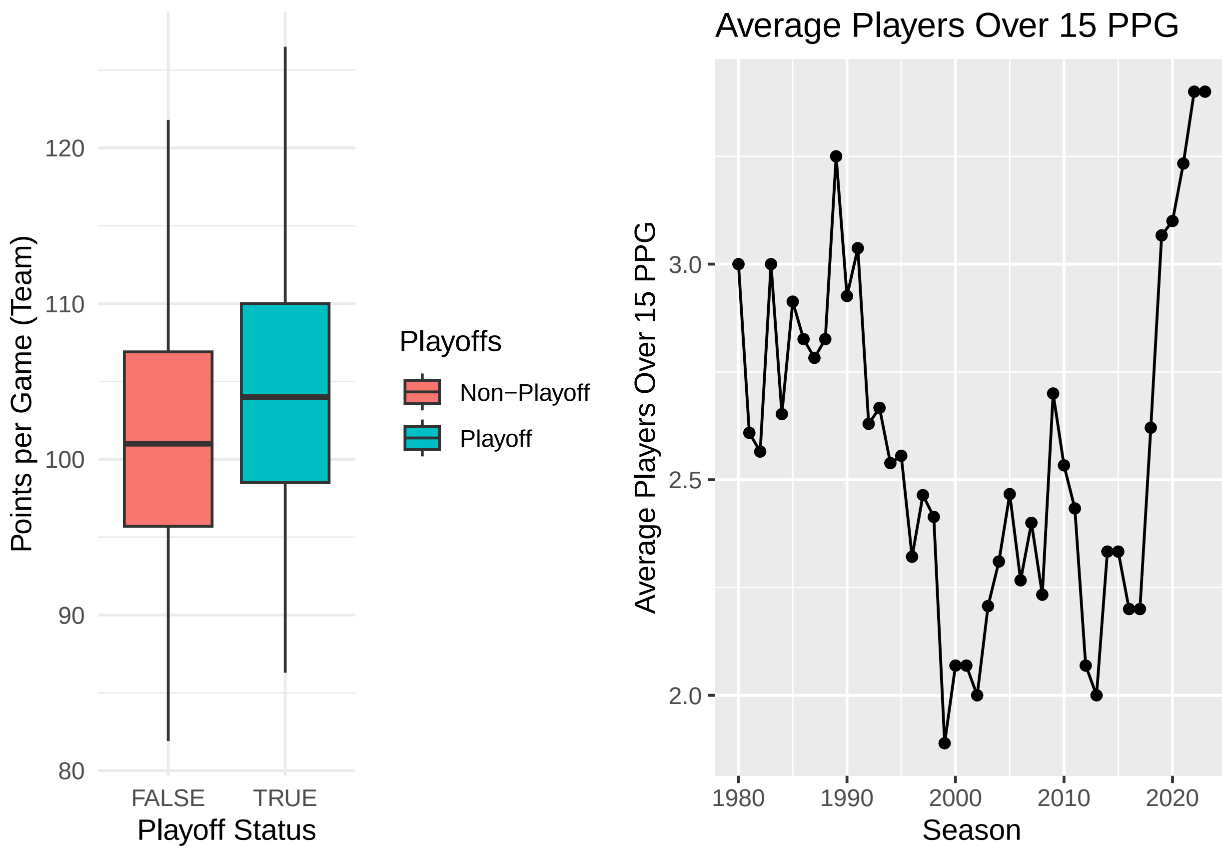
- In the plot above, the average team scoring per year is displayed. There have been noticeable increases in scoring during recent decades.
- Below on the left, the boxplot shows the scoring patterns of teams that made the playoffs against those that did not.
- The plot on the right below demonstrates the changes in players scoring greater than 15 points per game over time, following a roughly similar trend to total points scored.

##

```
## Attaching package: 'gridExtra'

## The following object is masked from 'package:dplyr':
##
##     combine

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




```
## # A tibble: 10 x 2
##    tm    total_playoffs
##    <chr>          <int>
##  1 SAS               36
##  2 LAL               34
##  3 POR               34
##  4 BOS               33
```

```
## 5 UTA 31
## 6 HOU 30
## 7 ATL 28
## 8 OKC 28
## 9 CHI 27
## 10 IND 27
```

0.5 Methods

For the purposes of this study, we will be using autoregressive integrated moving average (ARIMA) and logistic regression to model this time series data. In the first model, let X denote the number of players scoring greater than 15 points per game on an NBA team for a given season. Y will denote the chosen team success metric, the number of wins in this case. The ARIMA is shown below, where ϕ_p represents the effects of the lagged team success metric and ϵ is the white noise error term.

$$Y = c + \beta X_t + \phi_p Y_{(t-1)} + \epsilon$$

We will be estimating the parameters c (constant term), the independent variable β_1, and ϕ_p to evaluate and understand the relationship present between the number of players scoring 15 points per game and wins. For the second model, let Y denote the number of players scoring greater than 15 points per game on an NBA team, and let X denote the corresponding playoff status for that year. The Ordinary Least Squares Method will be used to estimate c and β_1 based on the observed data. For each of these coefficients, the standard errors of the point estimators will be calculated as a means to measure the level of uncertainty. The OLS method will estimate the variances of the point estimators, which are important for hypothesis testing. The method of maximum likelihood will be used to estimate ϕ_1. This model will work to examine the relationships between the number of players scoring greater than 15 points per game and wins, in addition to the relationships between an observation and its corresponding lagged value.

The hypothesis testing will be used to establish the null distribution of the testing statistics, using the permutation method to test the correlation of variables. For the alternative hypothesis, we claim that there is a significant relationship between the number of players scoring greater than 15 points per game and team success metrics. We predict that there will be a nonzero association between these. We will model these relationships to make inferences about the impacts on team success, and use correlation analysis to confirm the findings observed in the model. For the hypothesis testing, we will be using these hypotheses:

$$H_0 : \rho = 0$$

$$H_A : \rho \neq 0$$

$$\alpha = 0.05$$

The variable ρ represents the correlation between variables, and for any value greater than the chosen significance level, there will not be sufficient evidence to reject the null hypotheses.

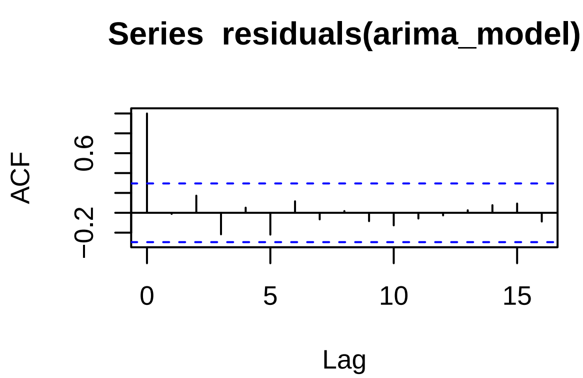
0.6 Analysis

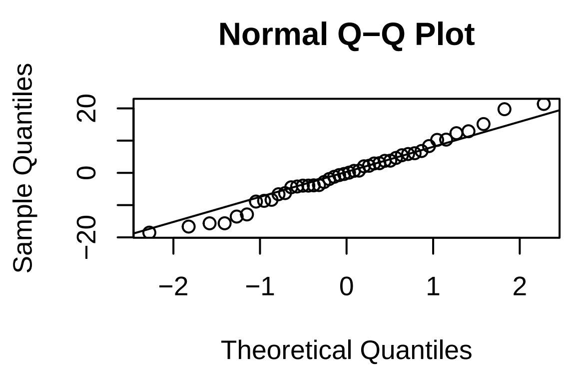
The aim of this analysis is to investigate the relationship between the number of players on a team scoring greater than 15 points per game and team success in the NBA. Specifically, the goal is to determine if having more players scoring greater than 15 points per game is associated with more wins and playoff appearances. The data to be used in this study is real, observational data, therefore resampling and simulation are not necessary. The main factor of interest is the number of players scoring greater than 15 points per game. Other factors considered are the team success metrics, which include wins and playoff appearances. Using ARIMA and logistic regression for analysis, we will explore the relationship between the number of players scoring greater than 15 points per game and team success. This will be done using ARIMA to create a model for the relationship between players scoring greater than 15 points per game and wins over time. Logistic regression will be used to model the relationship between players scoring greater than 15 points per game and playoff appearances. Additionally, diagnostic tests will be used to check for key assumptions and the validity of the models. Evaluating the performance of the ARIMA and logistic regression models will be done using the AIC, which we seek to minimize for goodness of fit.

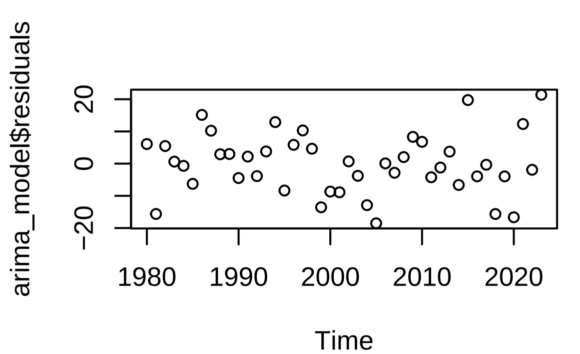
0.7 Results

- The summary statistics for the ARIMA model for wins are below. The low AIC value suggests a good fit for the data.
- The plots below show that the key assumptions for ARIMA modeling are true, including normality on the q-q plot, no significant autocorrelation on the acf plot, and heteroscedascity with no trends on the fitted values vs. residuals plot.

```
## Series: ts_w
## Regression with ARIMA(1,0,0) errors
##
## Coefficients:
##          ar1  intercept    xreg
##       0.5314    36.1411  2.2262
## s.e.  0.1332     5.5003  2.0076
##
## sigma^2 = 92.13:  log likelihood = -160.56
## AIC=329.11   AICc=330.14   BIC=336.25
##
## Training set error measures:
##                      ME     RMSE      MAE       MPE     MAPE      MASE
## Training set -0.1116636 9.265491 7.299011 -7.502489 21.44175 0.8237729
##                    ACF1
## Training set -0.0116547
```







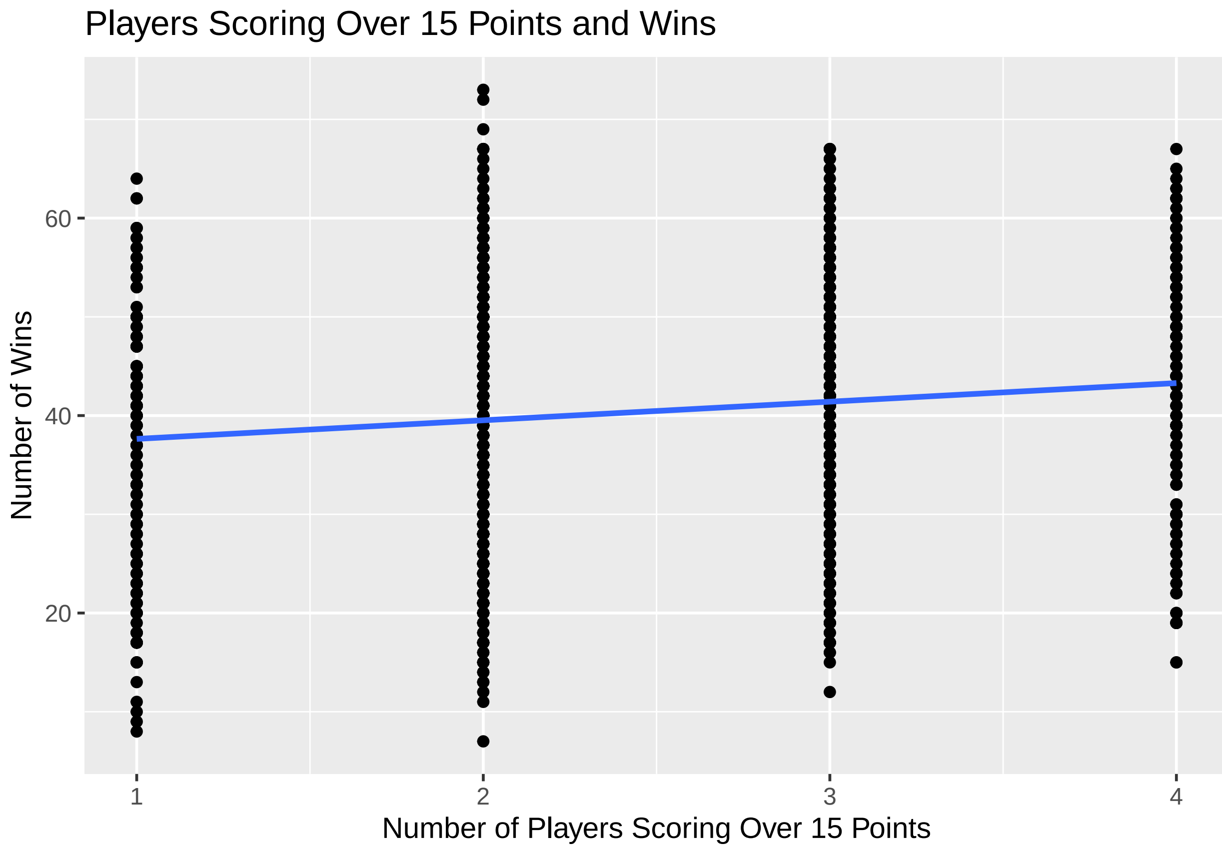
0.7.1 Wins

An ARIMA model (1,0,0) was fitted to the time series data relating the count of players scoring greater than 15 points per game to the team's win totals. This model revealed that there is insufficient evidence to support the claim that the number of players scoring greater than 15 points per game is a significant predictor of wins. To further analyze this, cross-correlation was used to calculate the correlation between the two variables.

A very low correlation value of 0.035 suggests that there is not a significant relationship between the number of players scoring greater than 15 points per game and team wins. Using the permutation test to obtain the p-value for this correlation, the p-value of 0.457 greatly exceeds the threshold, furthermore suggesting that there is not sufficient evidence to reject the null hypothesis. There is not a significant relationship or correlation observed between the players scoring greater than 15 points per game and wins.

- The plot below shows the distribution of team wins based on the count of players scoring over 15 points.
- The line indicates a slightly positive association between the variables, however this does not mean the association is significant.

```
## `geom_smooth()` using formula = 'y ~ x'
```



0.7.2 Playoff Appearances

To model playoff appearances using the number of players scoring greater than 15 points, the logistic regression model is used. The assumption of linearity between the number of players scoring over 15 points and the log odds of the playoff appearances was met.

The summary statistics for this logistic regression model are displayed below.

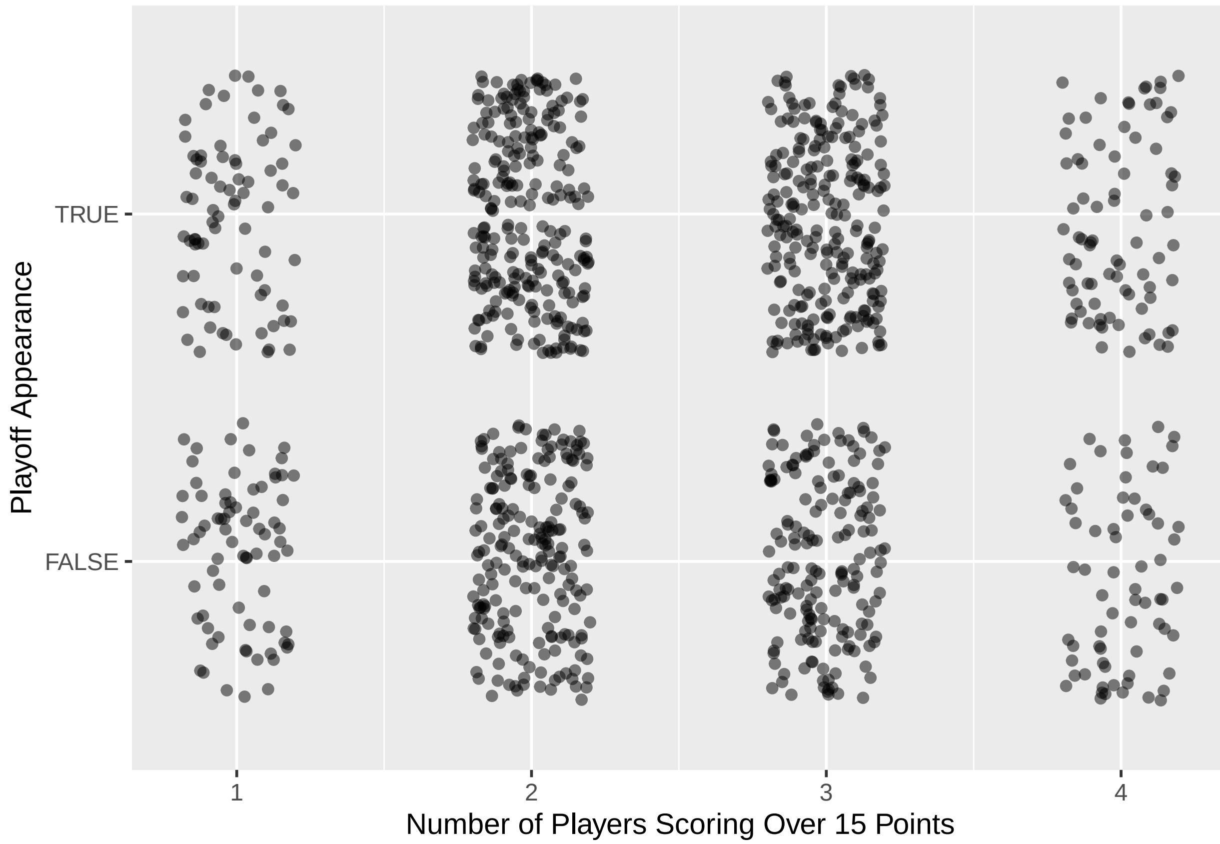
- The number of players scoring over 15 is estimated to have a positive effect, but this relationship is not significant.
- The AIC value of 1598 demonstrates that the fit of this model could be improved.
- The correlation between players scoring over 15 and playoffs is estimated to be very low (0.03), and this is not significantly correlated.

```
##
## Call:
```

```
## glm(formula = playoffs ~ num_over_15, family = "binomial", data = team_data_clean)
##
## Coefficients:
##             Estimate Std. Error z value Pr(>|z|)
## (Intercept)  0.04567    0.17750   0.257    0.797
## num_over_15  0.07171    0.06771   1.059    0.290
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 1595.1  on 1160  degrees of freedom
## Residual deviance: 1594.0  on 1159  degrees of freedom
## AIC: 1598
##
## Number of Fisher Scoring iterations: 3
```

```
##
## "Pearson’s correlation test estimate and p-value:"
##                                                  cor
##                                "0.0310918922043776"
##
##                                 "0.289817179221549"
```

- The jittered scatter plot below shows the distribution of playoff appearances versus the number of players scoring over 15 points.
- There is clustering at 2 and 3 players scoring over 15, but overall playoff appearances appear to be evenly distributed between true and false at all levels of scorers.



The results of the logistic regression model demonstrate no significant association between players scoring over 15 and playoff appearances. There is no statistically significant correlation observed between the number of players scoring over 15 points per game and playoff appearances, with a very low correlation coefficient of 0.03. The number of players scoring over 15 points on a team does not appear to be a significant predictor or have an association with playoff appearances. Logistic regression may not be the best fit for the data.

0.8 Application

Using a variety of modeling techniques and correlation analysis, there is not sufficient evidence to support the claim that having more players scoring 15 points per game or greater has any meaningful impact on team success in the NBA. For both dependent variables, wins and playoff appearances, the models revealed that the number of players scoring over 15 points per was not a statistically significant predictor. Additionally, the number of players scoring over 15 points does not have a significant correlation with either of the team success metrics.

The analysis for team wins, done by fitting an ARIMA model to the data, showed that the number of players scoring over 15 points per game was not a significant predictor of wins. A low correlation coefficient of 0.035

suggests that there is a very weak correlation between the two. The logistic regression model used to model playoff appearances using the number of players scoring over 15 points per game yielded similar results. The correlation coefficient of 0.031 indicates a very weak positive association between the number of players per team and playoff appearances. This was not a statistically significant predictor in the model.

0.9 Discussion

The findings of this study indicate that while there may be a very small association between the number of players scoring greater than 15 points per game on a team and the team's playoff appearance status and win count, there is not a substantial impact on team success and is not a direct predictor for these team success metrics. The results of these analyses provide insufficient evidence to reject the null hypothesis that there is not a significant association between the number of players scoring greater than 15 points per game and team success. This suggests that other factors play a dominant role in predicting an NBA team's success. This study is limited by the specific research question, which only takes one independent variable into question. Establishing the threshold for players scoring at 15 points may be a limiting factor as well. Multiple players scoring 15 points in an NBA game is commonplace, so 15 points may not be high enough to see an impact. If a similar research question with a higher standard for point scoring was investigated, there may be a stronger association than observed in this study. The dynamics of the NBA and the factors of team success include far more than scoring. Future research could include a modeling approach with many more offensive and defensive methods to better determine the contributing factors to the success of an NBA team.

In conclusion, the analysis did not find a strong relationship between the number of players scoring over 15 points per game and team success. This analysis provides information about scoring patterns and how they can be related to success, and how future research would benefit the understanding of the determinants of team success. Continued research into the complexities that define NBA success can greatly impact the strategies used in formulating gameplans and team construction. Relating multiple facets of the game to team success requires more research in order to gain a better understanding of their impacts on team outcomes.

References

Cabarkapa, Dimitrije, Michael A. Deane, Andrew C. Fry, Grant T. Jones, Damjana V. Cabarkapa, Nicolas M. Philipp, and Daniel Yu. 2022. "Game Statistics That Discriminate Winning and Losing at the NBA Level of Basketball Competition." *PLOS ONE*. Public Library of Science. https://journals.plos.org/

plosone/article?id=10.1371%2Fjournal.pone.0273427.

Choo, Hyunsoo, David Creegan, Olivia Majedi, Daniel Smolyak, and Brian Valcarcel. 2020. "DATA-DRIVEN APPROACHES TO NBA TEAM EVALUATION AND BUILDING." *University of Maryland DRUM.* https://api.drum.lib.umd.edu/server/api/core/bitstreams/bed7e07f-2360-4e0d-a30e-4f0b32950a7d/content.

Lorenzo, Alberto, Miguel Angel Gomez, Enrique Ortega, Sergio José Ibáñez, and Jaime Sampaio. 2010. "Game Related Statistics Which Discriminate Between Winning and Losing Under-16 Male Basketball Games." *Journal of Sports Science & Medicine.* U.S. National Library of Medicine. https://pubmed.ncbi.nlm.nih.gov/24149794/.

Terner, Zachary, and Alexander Franks. 2021. "Modeling Player and Team Performance in Basketball." *Annual Review of Statistics and Its Application.* Annual Reviews. https://www.annualreviews.org/content/journals/10.1146/annurev-statistics-040720-015536#right-ref-B38.