Advances in basketball statistics

Marica Manisera

Marco Sandri

Paola Zuccolotto

Big&Open Data Innovation Laboratory (BODaI-Lab)
University of Brescia, Italy

December 2, 2019

Abstract

In recent years, the interest in basketball statistics has greatly increased and new methods and models have been proposed to analyze basketball data with several interesting aims. In this contribution, after briefly introducing the state of the art of basketball analytics, we offer an overview of possible basic and advanced investigations in performance analysis in basketball, with focus on some opportunities offered by the R package `BasketballAnalyzeR`.

1 Introduction

In recent years, basketball statistics have become more widespread and the interest in the applications of statistical methodology to basketball data has increased both among statisticians, more focused on the methods, and professionals (coaches, players, scouts, managers, but also fans and sportscasters). A huge number of statistical books and papers have been published on all the aspects of basketball, including performance analysis, sports markets, marketing strategies, psychological attributes of players and their impact on match results, medical issues related for example to injuries, physical and physiological characteristics of players, fitness and training strategies, scheduling, and many other issues (among others, see 4, 73, 85, 3, 57, 64, 90, 31, 32).

This contribution is focused on performance analysis, which includes both basic tools and more advanced statistical methods. Starting from Dean Oliver's book [60], the quantitative approach to performance analysis has been applied with a great variety of aims: predicting the result of a match (or a tournament) [84, 44, 15, 34, 45, 69, 87, 50, 83], investigating the determinants of a team's success [80, 71, 36, 19, 18, 35, 40, 41, 27, 10], analysing a player's performance and its impact on the team's play [63, 17, 70, 67, 23, 61, 62, 22, 20, 64, 25, 21], analysing the discussed "hot hand" [28, 82, 39, 81, 7, 8, 9], examining performance in high-pressure game situations [48, 47, 29, 91], monitoring playing behaviours, also in order to define new roles [2, 11], studying tactics and identifying optimal game strategies [5, 88, 77, 51].

With the help of technology, sensor data are now available. This kind of data allows statisticians to extend the range of topics to analyze, for example the kinetics of body movements, with the aim of identifying which are the movements leading to the highest shooting efficiency and timing [56, 58, 59, 1, 92]. Other interesting topics are the study of players' movements and trajectories and the network of play actions[26, 66, 76, 78, 13, 14, 65, 43, 68, 24, 79, 74, 6, 16, 33, 52, 53, 86, 12, 55, 54].

The possibilities in basketball analytics are really wide and the results that can be obtained from a statistical analysis also depend on the available data and software. The aim of this contribution is to give an overview of basic and advanced statistical analyses that can be performed, focusing on the use of the R package `BasketballAnalyzeR` [72, 89, 49]. Section 2 briefly presents the different data sets existing for statistical analysis in basketball. Section 3 offers an overview of possible basic (Subsection 3.1) and advanced (Subsection 3.2) investigations that can be addressed in basketball analytics, with focus on some of the opportunities offered by `BasketballAnalyzeR`. Section 4 is devoted to answer some of the questions that could be posed by basketball fans, in a sort of Q&A section with Questions from fans and Answers by basketball analysts.

2 Basketball statistics: state of the art and data

The application of statistical methods and models to basketball data gives birth to the so-called basketball analytics. In a schematic way, three main worlds exist at the heart of basketball analytics: institutional analyses, sport analytics services and scientific research.

Institutional analyses are often provided for teams and players in every league or championship, also in the lowest level tournaments. They include number of shots made and attempted, success percentages, number of fouls, rebounds, assists, turnovers, etc. These statistics are usually available for free on the web and are released in real time or at the end of a game. Basketball fans know them very well and sport journalists base their articles on these results. However, although useful, institutional analyses are not able to give a relevant contribution in defining winning game strategies or predicting success probabilities.

Sport analytics services are provided by companies specialized in the creation of analysis platforms against payment. These services are often customized on the basis of the customer's needs and give the possibility to record all the events of a match. They return the results of simple statistical analyses, often in the form of graphs and nice visualization tools easy to interpret. Their aim is to offer analytical tools helping the team's staff to make strategic decisions about the next match or identify the best training programmes.

Scientific research deals with basketball analytics with sophisticated analyses, using state-of-the-art statistical techniques and models and developing new methods to address research questions in basketball appropriately. Studies that deeply investigate methodological aspects of basketball statistics cannot provide usable results in real time, but require a quite long and complex evaluation, especially in the interpretation of results, also with the help of basketball experts.

In addition, we can also consider analysts working inside basketball teams. Usually, their statistical analyses and results are kept out of the spotlight, for fear that they might reveal some secret strategies or weaknesses.

All the three worlds described above give an important contribution to basketball analytics, addressing several different issues. The range of questions about basketball that may be answered by statistical analyses is growing thanks to the online availability of data sets and increasing computational power. Data can be obtained by multiple sources and according to Zuccolotto and Manisera [89] can be classified in the following macrocategories: (*i*) data recorded manually or quasi-manually; (*ii*) data detected by technological tools; (*iii*) data from questionnaires; (*iv*) other data, including, among others, information retrievable from the web.

Data manually (or quasi-manually) recorded include data from box scores and play-by-play data. Box scores are tables commonly used in basketball that provide a structured summary of the results of a match or a championship. They list the game score as well as individual and team achievements in the game. Typically, they show the number of matches played, the minutes played, the field-goals made and attempted, together with the field goal percentage, three-point and two-point shots attempted and made, with the success percentage, free throws made and attempted and the success percentage, together with some other game variables, like the number of (offensive, defensive and total) rebounds, assists, turnovers, steals, fouls, points scored, etc.

These data are sometimes considered as the basketball final statistics to look at. However, from our point of view, they are only the starting point for other analyses. This is a crucial point, because we believe that the understanding of basketball cannot be limited to the computation of these numbers. Instead, they must be collected and appropriately structured in order to allow further statistical analyses.

Play-by-play data record all the events occurred during a match; in other words, they provide a transcript of the game as a list of individual events. Each event is described together with the indication of the time of the possession, the player involved in that event, often including the player location in that instant.

Examples of data detected by technological tools (category (*ii*)) include positions of players, referees and the ball on the court, recorded by sensors or other tracking systems. This kind of data record defensive and offensive alignment relative to ball location, shot trajectories, player position and defender proximity to a player.

Data from questionnaires (category (*iii*)) are usually obtained by means of psychological scales administered to players. The aim is to measure their subjective perceptions and attitudes (for example, leadership, coping strategies, mental toughness, etc.) that can be somehow related to performance.

Category (*iv*) is a residual category, including for example twits of the fans, posts on social media like Facebook or Instagram, trends of online searches as provided by Google Trends.

While obtaining data of category (*iii*) and (*iv*) requires a customized data processing, data belonging to categories (*i*) and (*ii*) are often available, with different degrees of quality, in the informative systems of the National and International Federations, sporting organizations, basketball societies. Basketball data can be available for free or on payment. Usually, when data are free, sophisticated computer tools are needed to get the data, like web scraping procedures.

Data quality is a very important issue and the potential of a data set should be assessed with reference to the specific objective of the analysis [38]. For example, if the focus is on the game evolution in terms of single events, play-by-play data are required. However, a good play-by-play data set must track the game events multiple times at each minute. For example, the play-by-play data currently made available by the Italian Serie A website are aggregated per minute (web.legabasket.it). Without a description second-by-second, no in-depth analyses are viable. In addition, a special attention must be devoted to the appropriate contextualization of the obtained results. One must keep in mind that generalizability is not always guaranteed, so for example, results from NBA data cannot be directly extended to Europe.

3 Statistical analyses using `BasketballAnalyzeR`

Advances in basketball statistics gain traction from the wide range of questions that arise in the field of basketball analysis. The forthcoming R package `BasketballAnalyzeR` [72, 49] allows to perform basic and advanced statistical analyses in order to give a reply to several questions. It accompanies the book entitled "Basketball Data science" [89], developed with a substantially empirical approach within the activities of the international network BDsports (Big Data analytics in sports, bdsports.unibs.it), whose main aims include scientific research, education, dissemination and practical implementation of sports analytics.

3.1 Basic statistical analyses

In basketball analytics, there exists a set of well-known basic indexes and graphs that are commonly used by experts, proposed on several specialized websites and well understood by fans. Examples of **basic indexes** include the Total Basketball Proficiency Score [37], the Individual Efficiency at Games [30] and those proposed by Dean Oliver, based on the importance of pace and possessions in the definition of player and team performance and the influence of teamwork on individual statistics. Oliver proposed the definition of offensive and defensive efficiency ratings and the so-called Four Factors (field-goal shooting, offensive rebounds, turnovers and getting to the free-throw line, 42). They can easily be computed by `BasketballAnalyzeR`.

Several **graphical representations** complete the set of basic analyses. Bar-line plots, radial plots, scatter plots and bubble plots can give interesting insights on the characteristics of games, teams and players. For example, a bar-line plot helps in comparing the offensive statistics of several teams or players; radial plots can be useful to represent a profile, based on several game features, for each team or player considered. Scatterplots suggest relationships between two variables (or even three, if points are color- or symbol-coded) measured on the teams or players and allow to identify possible anomalous situations. Bubble plots represent even four characteristics of teams or players in one single plot, since size and color of bubbles vary according to two variables that are added to the two that define x and y location. For example, a bubble plot of a team's players can be defined using 2-point and 3-point shot percentages on the x and y axes, respectively, and bubbles can be colored according to the scoring percentage of free throws and dimensioned according to the total number of attempted shots. Shot charts are widely employed in basketball analytics. They give clear indications on the players' positions on the court during the match and can be enriched with useful statistics that help understanding what happens on the court.

Other simple tools of descriptive statistics can be used to analyze **performance variability**. Variability indexes and plots inform us about the extent to which players or teams perform differently from each other. Special caution should be used in interpreting variability results. For example, if we are interested in evaluating performance variability of the players in one team, high variability indicates the presence of some players very good and some other very bad with reference to the game variable analyzed. If the analyzed game variable relates to a very specific task (for example, number of assists), high variability suggests that some of the players specialize in assisting their teammates and the team balance is good. On the contrary, if the analyzed variable refers to a generic task (for example, field goal percentages), a high variability is a sign of high dependency of the team on a few players with performance higher than average.

`BasketballAnalyzeR` also allows us to run an **inequality analysis** within a basketball team. Borrowed from the economics field, inequality analysis in basketball evaluates the extent to which the distribution of some performance measure (for example, the number of points made by one team) deviates from a perfectly equal distribution (all the players score the same number of points) and from a maximally unequal distribution (one single player scores the total number of points). The value of the Gini's inequality coefficient[1] and the graphical representation of the Lorenz curve [2] allow comparisons among the inequality of several teams.

[1]The Gini's inequality index is a normalized index measuring the degree of inequality in a given distribution. It ranges from 0 (perfect equality) to 1 or 100% (maximal inequality). In the case of income distribution of a nation's N residents, a value of 0 indicates that all residents have the same income, while a value of 1 is obtained when all the nation's income is owned by only one person.

[2]The Lorenz curve graphically represents, on y-axis, the fraction y of the total variable (for example, income) that is cumulatively referred to the bottom fraction x of the population (on x-axis). The two extreme situations of perfect equality and maximal inequality are represented by the straight line $y = x$ and the line having $y = 0$ for all $x \leq (N-1)/N$, and $y = 1$ when $x = 1$, respectively. The observed Lorenz curve lies in between these two extremes. The closer the curve to the perfect equality line, the smaller the inequality level. The Gini coefficient is computed as the ratio of the inequality area (that is the area between the perfect equality line and the observed Lorenz curve) to the maximum inequality area (given by the area between the perfect perfect equality and the perfect inequality line).

3.2 Advanced statistical analyses

There exist numerous advanced statistical analyses that can be performed in basketball studies with a variety of aims. The long, although non-exhaustive, list of papers cited in Section 1 gives an idea of the recent publications on this topic. Some of them are directly available in `BasketballAnalyzeR`, some others can use the results from the package as a starting point to develop further methodologies and applications.

A first set of analyses can be carried out to study the **association among variables**. In a very wide sense, one can investigate statistical dependence (for example, to measure how much the number of rebounds by one team depends on the opponent team), mean dependence (to evaluate, for example, if the average number of points scored by all the NBA teams differs between the East and West conferences) and correlation. In particular, pairwise linear correlation among variables can be studied creating a correlation matrix and its graphical representation in order to examine the degree, direction and significance of linear relationships between game variables measured on the single players.

The **similarity among teams or players** with respect to selected game variables can also be assessed and graphically represented by a specific function of `BasketballAnalyzeR`, by resorting to the multivariate data analysis technique of Multidimensional Scaling. It reduces a high-dimensional data set (several teams or players on which a high number of variables have been measured) into a low-dimensional map (usually two dimensions are retained) displaying teams or players as points: points close to each other have similar characteristics, while distant points indicate peculiar teams or players. The dimensionality reduction implies a loss of information, which must be evaluated in order to assess the goodness of the resulting representation (the stress index is usually employed).

A very important issue in basketball statistics is the study of the relationships among players and the impact of their interaction on the team's achievements. Several statistical methods can be used to this aim, for example **network analysis**, in its meaning of the statistical analysis of network data. In basketball analytics, the system conceptualized as the network is the team and the focus is on modelling the complex statistical dependencies among players, often using high-dimensional data, with the final aim of predicting the team behavior. For example, `BasketballAnalyzeR` allows to investigate the network of assists in a team using play-by-play data. The resulting graph is a net displaying the players who mostly interact as those who make and receive most assists. This issue can then be further investigated by considering the play in the absence of some key players, in order to examine how the team reorganizes its game strategy.

The frequency of occurrence of some events with respect to some variable of interest can give interesting insights on the way of playing by a certain team or player. For example, it can be interesting to examine the **frequency of shots in time** (with respect to the seconds played in a quarter, for example) **or in space** (looking at the shot distance, for example) in order to investigate the players' performance in specific moments or areas. Statistics gives a response to these questions by density estimation methods, including several tools like histograms, naive, kernel, nearest neighbor, maximized penalized likelihood, etc. [75]. It becomes then clearer, for example, if the considered players tend to concentrate their shots in a particular moment of the match or in a given area of the court, how their scoring probability varies in time and in space, giving useful suggestions to define winning game strategies and improve training programmes.

Among the data mining algorithms allowing advanced statistical analyses in basketball we can mention **cluster analysis**, a broad methodology including several techniques which differ remarkably in their functioning. Cluster analysis is an unsupervised classification method that aims at grouping players, teams, matches, game moments, etc. into classes not a priori defined. Clusters gather together observation units that are similar to each other while different from units belonging to the other clusters. The analysis of the goodness of the solution together with the clusters' profiles and characteristics can give interesting insights on the structure of the data set and finally on the phenomenon under study. In basketball, clusters of players can be used to identify similar players and re-define their roles in playing that can differ from the traditional positions (among others, 11). Indeed, the historical five positions (point guard, shooting guard, center, small forward and power forward) were defined a long time ago, when even the basketball rules were different (for example, the introduction of the 3-point line dates back to the early 80's). New rules have led to changes in the players' physical preparation, their playing style and the way players interpret their role. In the end, new positions can reflect updated points of view about the game and can be identified by means of statistical analyses based on game variables. Also, matches can be clustered based on their ease, then further analyses can be carried out on the obtained clusters with the final aim of identifying losing and winning factors [46]. Game moments can also be classified, for example according to the game schemes adopted by players, in order to analyze the play style of a team during a match [54].

Naturally, advances in basketball statistics also concern the very broad field of **statistical models**, in their wide meaning of both models with mathematical formalization and algorithmic models. Every objective in basketball statistics can be achieved by using appropriate models, from the simplest ones, as linear regressions and non-parametric regressions, to the most complicated ones, able to model multivariate nonlinear relationships, time dependencies, interactions among variables and observation units, different nature of the involved variables, etc. A non-exhaustive literature review can be found, for example, in Zuccolotto and Manisera [89], where also some recent scientific papers are discussed in detail, namely a study on the scoring probability in the presence of high-pressure game situations [91], the definition of new roles in basketball [11], the analysis of players' movements and their effect on the team performance [54].

4 Statistics answers fans' questions

This section is devoted to answer some of the possible basketball fans' questions, with the aim of showing how basketball statistics "are more than numbers". The R package `BasketballAnalyzeR` [72, 49] is really helpful in fostering the meeting between statistics and basketball, because it can be fruitfully used by interested users without a strong statistical background.

This Q&A section, with examples of Questions from fans with the corresponding Answers by basketball analysts, is developed using the data from the regular season (82 games) of the NBA championships 2018/2019 and 2017/2018. In much detail, we used one data set including additional information (for example, Conference, Division, etc.) and three box scores data (2018/2019): the Teams' box scores, containing the achievements of all the analyzed teams, the Opponents' box scores, for the achievements of the opponents of each of the analyzed teams, the Players' box scores, for the individual achievements of each single player in the considered games. In addition, we have one play-by-play data, recording the events of the 82 games played by the Cleveland Cavaliers during the NBA regular season 2017/2018.

Q1: What were offensive and defensive performance of the four NBA Conference finalists?

Although a universal definition of performance does not exist, we can refer to the famous Four Factors by Dean Oliver [42] as a good starting point to measure offensive and defensive performance of one team or one player. The Four Factors are (1) the Effective Field Goal Percentage (eFG%); (2) the Turnovers per possession (TO Ratio); (3) the Rebounding Percentages (REB%) and (4) the Free Throw Rate (FT Rate). They can be computed for both offense and defense so giving a measure of offensive and defensive performance from four points of view. Applying the function `fourfactors` of `BasketballAnalyzeR` to the data of the four 2018/2019 Conference finalists (Milwaukee Bucks, Toronto Raptors, Golden State Warriors and Portland Trail Blazers), Figure 1 is obtained. It shows the four factors for every team, besides information about pace, possessions and offensive and defensive ratings.

In detail, Figure 1 shows that the pace of the games increases moving from Portland Trail Blazers to Toronto Raptors, Golden State Warriors and Milwaukee Bucks. The Golden State Warriors have the best offensive performance (the highest offensive rating), while the Milwaukee Bucks are the best team in defense (they have the lowest defensive rating, that is, the offensive ratings of the opponents). The bars in the bottom plots represent the Four Factors: each bar measures, for each team and each Factor, the difference between the team value and the average of the four analyzed teams. A positive (negative) bar indicates a value above (below) the mean and suggests a strength or a weakness of the team (depending on which of the Four Factors is referred to, as explained later). The Portland Trail Blazers show a very good performance in offensive rebounds and free throw rate, with a performance lower than the average on the effective field goal percentage. This picture is completed, on the defensive

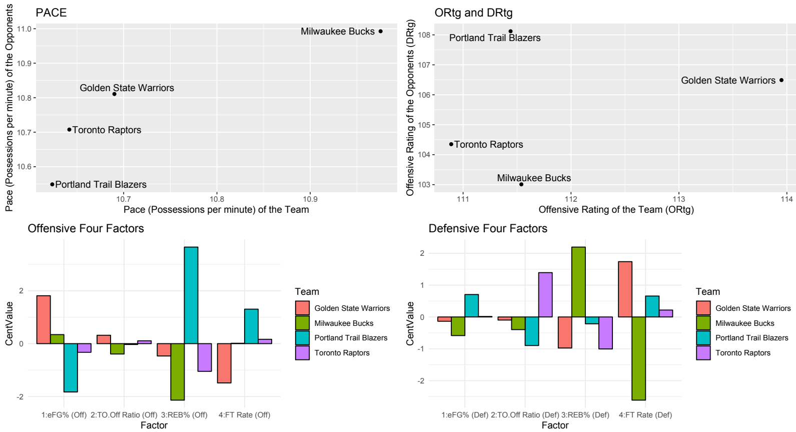

Figure 1: Pace, Offensive/Defensive Ratings and Four Factors (differences between the team and the average of the considered teams) - Conference finalists 2018/2019.

side, by a bad performance in the effective field goal percentage and free throw rate. This result is consistent with the fact that the Portland Trail Blazers lost the Western Conference final versus the Golden State Warriors (obviously, there are many other determinants of the outcome of a match).

Q2: Which NBA Eastern conference team had the best performance according to defensive statistics?

It is possible to plot in one single graph the main defensive statistics of the Eastern Conference teams. With the function `barline` we obtain Figure 2, where steals, blocks and defensive rebounds are represented on the bars, which are ordered (in decreasing order) according to the points scored by the opponents; the grey line measures the opponents' turnovers (TOV.O), whose scale is on the right vertical axis.

The Milwaukee Bucks are the team with the best performance on these defensive statistics (especially on defensive rebounds) while the Cleveland Cavaliers had the worst performance in defending (it has the lowest levels of steals, blocks and rebounds). However, there is no evidence of a relationship between the defensive statistics and

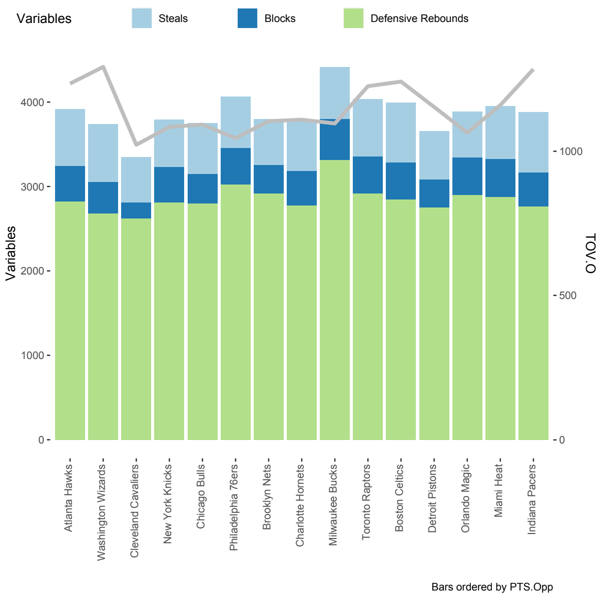

Figure 2: Defensive statistics of the NBA Eastern Conference teams (TOV: turnovers of the opponents, PTS.Opp: points scored by the opponents). Grey line: TOV.O (opponents' turnovers; right vertical axis)

the points scored by the team's opponents or their turnovers. For example, the Atlanta Hawks (the leftmost team) suffered the highest number of scored points, although they have the highest number of opponents' turnovers (grey line) and do not show so bad defensive performance in terms of steals, blocks and rebounds (it is not the team with lowest bars).

This response should be integrated with other statistics related to the defensive performance, which is indirectly measured by the offensive performance of the opponents (for example, the points scored by the opponent teams). A possible solution is the examination of the teams' Four Factors, as done in the previous Question Q1.

Q3: Can we compare the game profiles of the three best centers in NBA 2018/19?

According to some basketball experts, the three best centers in the NBA season 2018/19 have been Anthony Davis (New Orleans Pelicans), Joel Embiid (Philadelphia 76ers) and Karl-Anthony Towns (Minnesota Timberwolves). We can compare their performance according to 2- and 3-point shots made (P2M and P3M), free throws made (FTM), total (offensive and defensive) rebounds (REB), assists (AST), steals (STL) and blocks (BLK) (all per minute played) by constructing three profile plots with the function `radialprofile`. Figure 3 shows that the profiles of the three considered players are very similar. Special attention must be paid in interpreting radial plots, because the axes have all the same scale and sometimes this prevents us from seeing differences among players.

Indeed, the comparison should be complemented by analyzing Figure 4, where variables have been standardized [3]. Here, the points in each profile show whether that player is positioned above or below the average (computed on the three analyzed players) for the considered variables. Variable standardization makes it possible to highlight

[3] A standardized variable is a variable that has been rescaled to have a mean of 0 and a variance (or a standard deviation) of 1. In order to standardize a variable, it is necessary to subtract the mean from each of its observed values and divide by the standard deviation. Standardizing allows to compare variables, even when measured on different scales.

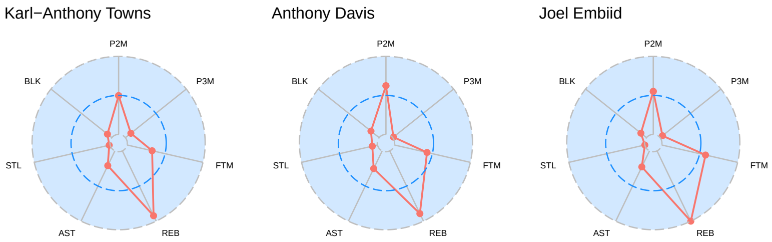

Figure 3: Radial plots of three selected centers, non-standardized variables. Dashed blue line: midpoint between minimum and maximum.

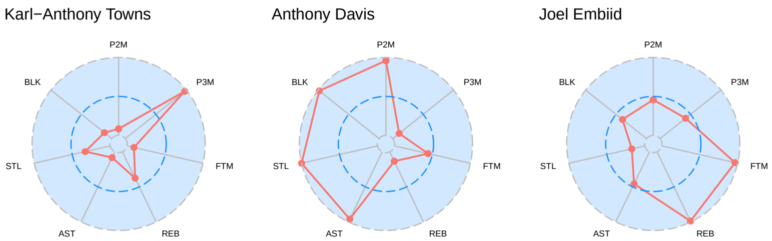

Figure 4: Radial plots of three selected centers, standardized variables. Dashed blue line: zero (average of each variable).

differences among the three players. For example, focusing on the 3-point field goals, it is evident that Karl-Anthony Towns had the best perfomance, far above the average, followed by Joel Embiid (close to the average) and then Anthony Davis (below the average). Anthony Davis overperformed the two other centers in several other variables (P2M, BLK, STL, AST), while Joel Embiid is the best player (among the three) in free throws and rebounding.

Q4: Which team had the best shooting performance in NBA 2018/19?

Shooting performance can be measured by the success percentage of 2-point shots, 3-point shots and free throws. The number of attempted

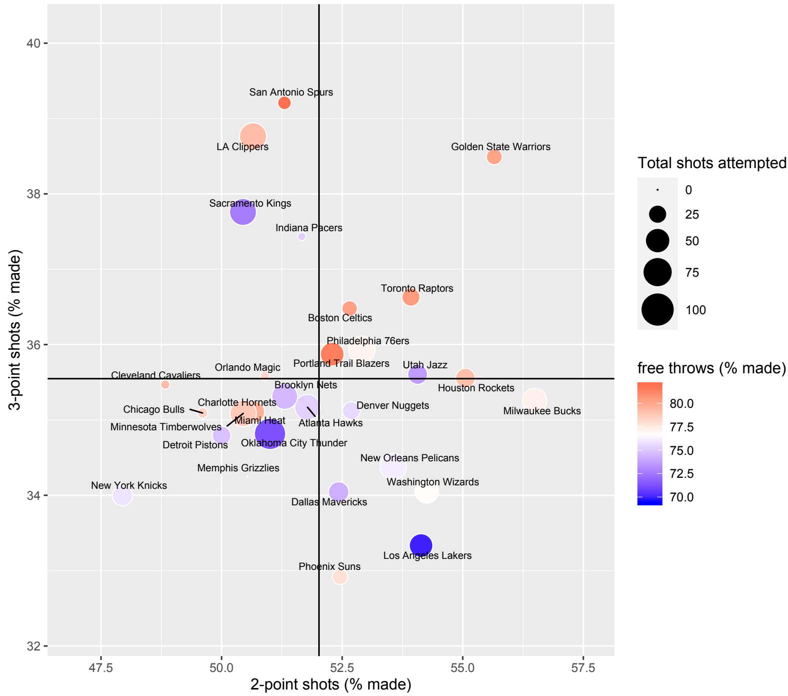

Figure 5: Bubble plot of teams according to the scoring percentages of 2-point shots (x-axis), 3-point shots (y-axis), free throws (red-blue color scale); bubble size: number of attempted shots.

shots is also important, because it is different if a team's players have made all the shots (success percentage equal to 100) having attempted 10 times or 100 times. All these four variables can be represented for the NBA teams in one single plot using the function `bubbleplot`. Figure 5 shows every team as a bubble, whose location reflects its scoring percentages in 2-point shots (x-axis) and 3-point shots (y-axis), colored according to the scoring percentage of free throws (on the red-blue color scale) and sized according to the number of attempted shots (of any kind), rescaled between 0 and 100. Vertical and horizontal black lines indicate the average scoring percentages in 2-point shots and 3-point shots, respectively, computed over the teams considered in the plot.

Figure 5 highlights some outstanding teams: the Golden State Warriors show very high shooting percentages with a medium-low number of attempted shots. The San Antonio Spurs and the LA Clippers have similar performance on 2- and 3-point shots (exceptionally good in 3-point shots and just below the average in 2-point shots), but the Spurs exhibit a higher percentage of free throws made and a lower number of attempted shots. The Toronto Raptors, the NBA 2018/2019 champions, have above-average performance on all the three types of shot but do not differ sharply from other teams (for example, Boston Celtics). The Portland Trail Blazers, finalists of the Western Conference together with Golden State Warriors, excel on free throws, but are close to the average for 2- and 3-point shots performance, with a high number of attempted shots.

Q5: Which players of the two NBA finalists had the best shooting performance and the best defense in NBA 2018/19?

A bubble plot analogous to the one in Figure 5 can be created to represent players instead of teams. Figures 6 and 7 show, respectively, the shooting and the defense performance of the players of the two NBA finalists (Golden State Warriors and Toronto Raptors) who have played at least 1,000 minutes in the regular season. Names are colored to distinguish the team they belong to (red for Golden State Warriors and blue for Toronto Raptors). Vertical and horizontal black lines indicate the average of x-axis and y-axis variables, respectively, computed over the selection of players considered in the plot.

It is interesting to observe that in both plots there is a mix of players from the two teams in all the four quadrants.

If we must choose the very best player from the shooting point of view, Figure 6 suggests Stephen Curry for the Golden State Warriors and Danny Green for the Toronto Raptors. The Raptors have relied heavily on Kawhi Leonard (one of the best defenders in NBA), whose position however is not as good as that of Danny Green, at least as regards the analyzed variables.

Kawhi Leonard appears in his great defensive ability in Figure 7, together with Draymond Green of Golden State Warriors, who shows, respect to Leonard, a better performance on blocks but a lower num-

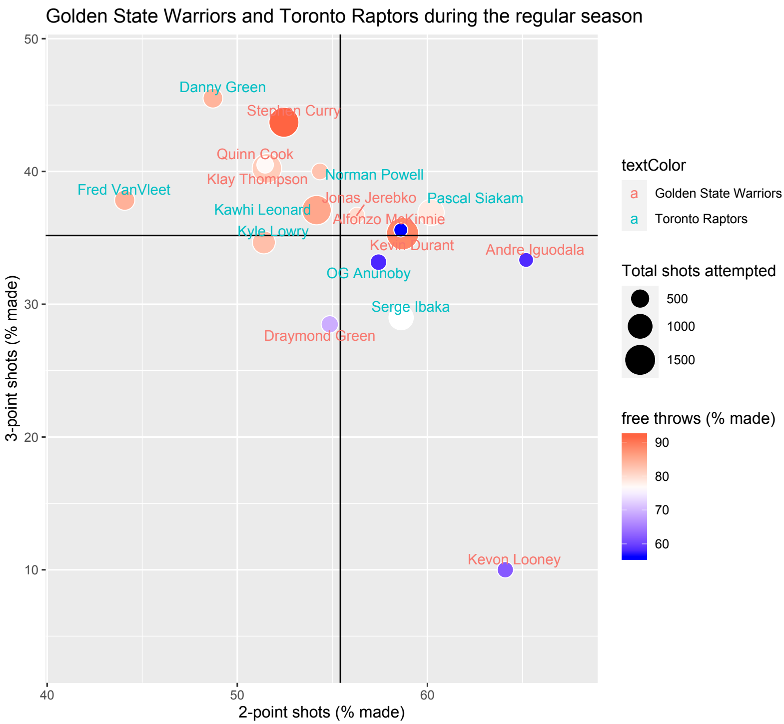

Figure 6: Bubble plot of selected players according to the scoring percentages of 2-point shots (x-axis), 3-point shots (y-axis), free throws (red-blue color scale); bubble size: number of attempted shots.

ber of steals (per minute played). Serge Ibaka shows a very good performance in blocks: this allowed him to help the Raptors to defeat the Golden State Warriors during the NBA Finals. It is nice to compare this plot with one of the top-10 list of the best defenders released by NBA, which includes Klay Thompson, Draymond Green and Kawhi Leonard. Some fans were surprised by the exclusion of some players from this list, for example Pascal Siakam and Danny Green. According to Figure 7, while Draymond Green and Kawhi Leonard have outstanding performance in defense, Klay Thompson has a position comparable to that of Danny Green while Pascal Siakam even performed slightly better than Klay Thomson, especially with respect to blocks and defensive rebounds.

A bubble plot using other interesting game variables can give another perspective for evaluating players.

Q6: How different is the performance on the 3-point shots among the players of Los Angeles Lakers?

Selecting only the players of Los Angeles Lakers who played at least 500 minutes and have attempted more than one 3-point shot, we have 11 players with different performances, as shown in Table 1.

Player	P3p	P3A
Alex Caruso	48.00	50
Lance Stephenson	37.06	197
Rajon Rondo	35.92	142
Kentavious Caldwell-Pope	34.71	435
Reggie Bullock	34.34	99
LeBron James	33.94	327
Josh Hart	33.58	274
Brandon Ingram	32.98	94
Lonzo Ball	32.89	228
Kyle Kuzma	30.33	422
JaVale McGee	8.33	12

Table 1: 3-point shots percentage (P3p) and 3-point shots attempted (P3A), Los Angeles Lakers

The average percentage is 32.92, but there is a great difference be-

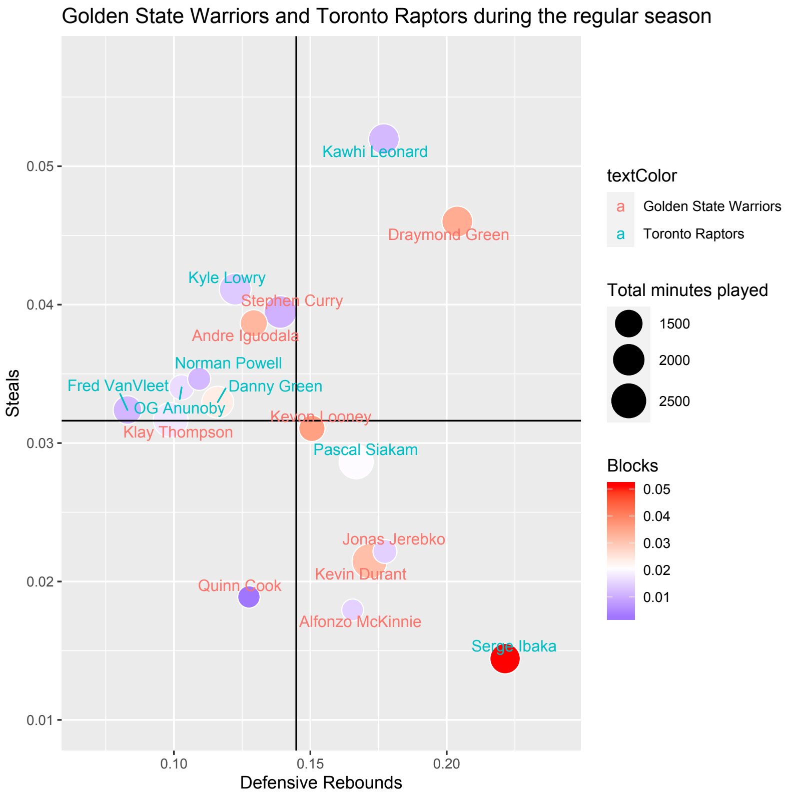

Figure 7: Bubble plot of selected players according to the defensive rebounds per minute played (x-axis), steals per minute played (y-axis), blocks per minute played (red-blue color scale); bubble size: total number of minutes played.

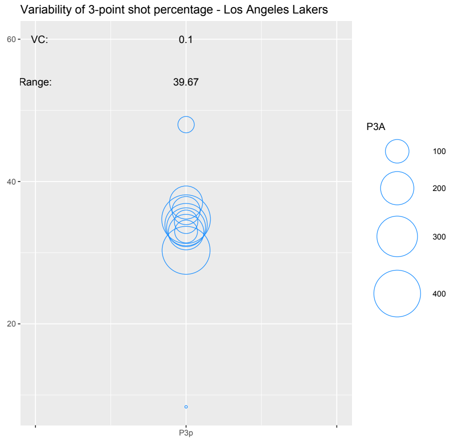

Figure 8: Variability diagram of the 3-point shot percentages weighted by the attempted shots, Los Angeles Lakers, VC=Variation Coefficient.

tween the top-player Alex Caruso (48%) and JaVale McGee (8.33%). To measure the variability of the 3-point shot scoring percentages, it is possible to compute some indexes with the function `variability`. Standard deviation, variation coefficient and range result in 8.90, 0.27 and 39.67, respectively.

An important issue is to consider the number of attempted shots when computing the variability indexes. Figure 8 shows a bubble for every player, located in correspondence of his 3-point percentage (vertical axis) with size proportional to the number of 3-point shots attempted. It is evident that there is a player with an outstanding position (Alex Caruso), with a relatively low number of attempted shots; considering only the remaining players, the bubbles are not very scattered around the average, denoting a fairly low variability.

Q7: How balanced is the San Antonio Spurs 3-point shot performance?

This question could be addressed analyzing the team variability, as in the previous question Q6. However, if we want to understand if the

San Antonio Spurs are a well-balanced team (that is, all the players contribute to its 3-point shot performance) or, on the contrary, depend too much on a few players, we can also consider the methods of inequality analysis, usually performed to evaluate the income or wealth distribution in a Country. In our context, we can resort to inequality analysis to measure the distribution of game achievements within the players of a team. With the function `inequality`, we obtain the value for the Gini coefficient, which measures the degree of inequality ranging from 0% (null inequality) to 100% (maximum inequality). In the example, focusing on the performance of 3-point shots, we can study whether only one or a few players are able to score all the 3-point shots (high level of inequality) or, conversely, all the team players give an equal contribution (null inequality).

Considering the 10 players of San Antonio Spurs who have scored the highest number of 3-point shots (Table 2), the value of the Gini coefficient equals 39.38%. This denotes a quite high level of inequality, given that the Gini coefficient computed on the 2018-19 NBA data ranges from 17.71% (Boston Celtics) to 48.13% (Golden State Warriors), with average equal to 29.65%. The San Antonio Spurs can count on a few players that score much of the 3-point shots, so it is not well-balanced from this perspective. Indeed, the top three players (Forbes, Mills and Belinelli) scored 61% of the total number of 3-point shots scored by all the 8 players considered.

Player	P3M	P3A
Bryn Forbes	176	413
Patty Mills	159	404
Marco Belinelli	147	395
Davis Bertans	145	338
Rudy Gay	74	184
Derrick White	48	142
Dante Cunningham	30	65
LaMarcus Aldridge	10	42

Table 2: 3-point shots made (P3M) and 3-point shots attempted (P3A), San Antonio Spurs

Figure 9 displays all the 30 NBA teams, according to their Gini coefficient's value and number of 3-point shots made. The San Antonio Spurs have roughly the same number of 3-point shots made as the LA

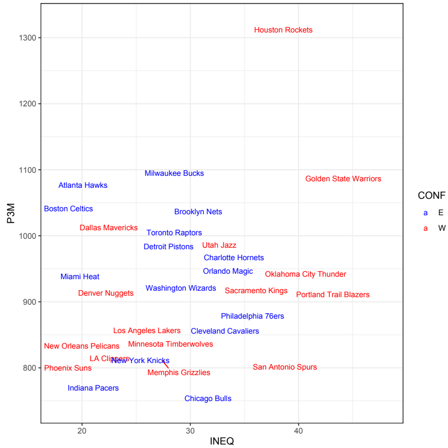

Figure 9: Gini coefficient (INEQ, x-axis) and number of 3-point shots made (P3M, y-axis) in the Western (red) and Eastern (blue) Conference teams.

Clippers or the Phoenix Suns, but with a highest level of inequality. The Houston Rockets appear as an outlier, with a huge number of 3-point shots made and a fairly high value of Gini coefficient: James Harden alone scored nearly 1/3 of the total number of 3-point shots made in the whole season.

Q8: What are LeBron James's favorite and disliked spots on the court?

LeBron James is one of the top NBA players; in 2017/2018 "King James" played with Cleveland Cavaliers, before moving to Los Angeles Lakers in 2018. To identify his favorite and disliked spots on the court, the function `shotchart`, applied to play-by-play data (with space coordinates), allows to obtain interesting shot charts, like those in Figure 10. In the left plot, points represent the shots attempted by LeBron James, colored according to whether he missed the shot (blue) or scored the basket (red). In the right plot, the court is split into sectors colored according to the average play length, that is the average time elapsed since the immediately preceding event when the shot is attempted. We note that in the first seconds of the play, Le-

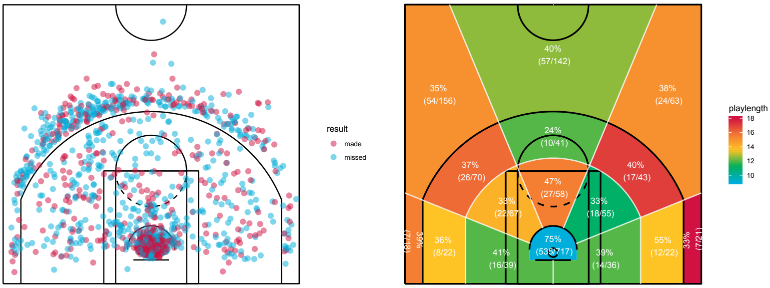

Figure 10: Shot chart (LeBron James). Left: missed (blue) and made (red) shots; Right: areas colored according to the average play length and annotated with shooting statistics.

Bron James prefers shooting from short distances (especially from his left-hand side), or attempting long-range shots from the center (where he has 40% of successful shots). Late shots are mainly attempted from the left and right sides (both mid-long and close range, with successful percentages ranging from 33% to 40%) and the center mid-range (where the proportion of shots that scored the basket is 47%). LeBron James shoots from short range, on average, very early in the play and these shots are the most successful (75%).

Q9: How does the network of assists work in the team of Cleveland Cavaliers?

The analysis of passing sequence with reference to the assists (the last pass before shot) can be analyzed using network analysis tools. In `BasketballAnalyzeR` this is implemented in the function `assistnet`, which requires play-by-play data. Investigation of interactions among teammates is very important in basketball analytics, due to its nature of team sport.

The left-hand plot in Figure 11 displays the network of assists made and received by the Cleveland Cavaliers players. Each node represents a player and the oriented edge goes from the player who

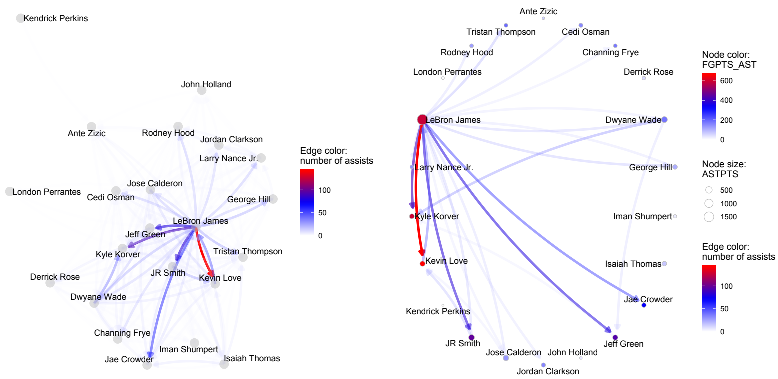

Figure 11: Network of assists of Cleveland Cavaliers. Edges are colored according to the number of assists (see the white-blue-red scale in the legend). Right plot: nodes are colored according to the points scored thanks to a teammate's assist (FGPTS_AST); node size is according to points scored by assisted teammates (ASTPTS).

made the assist to the player who received it. The edges are colored according to the number of assists. In the right plot, nodes are colored according to the points scored thanks to a teammate's assist (FGPTS_AST) and sized according to the points scored by assisted teammates (ASTPTS).

The crucial role played by LeBron James clearly emerges. He appears as the center of the game strategy, offering lots of assists to his teammates, firstly to Kevin Love and, secondly, to Jeff Green, Kyle Korver (who also receives assists from Dwyane Wade), JR Smith and Jae Crowder.

Node size and color in the right plot in Figure 11 give interesting insights on the network of assists of Cleveland Cavaliers. For example, LeBron James scored a high number of points thanks to a teammate's assist (his node is red) but he was also able to create scoring opportunities for the other players (his node is big). Kevin Love scored points thanks to the assists he received, while not offering many successful assists to his teammates. JR Smith, Jeff Green and Jae Cowder have small blue circles, indicating that they capitalized a medium-low number of assists and, at the same time, offered few fruitful assists to their

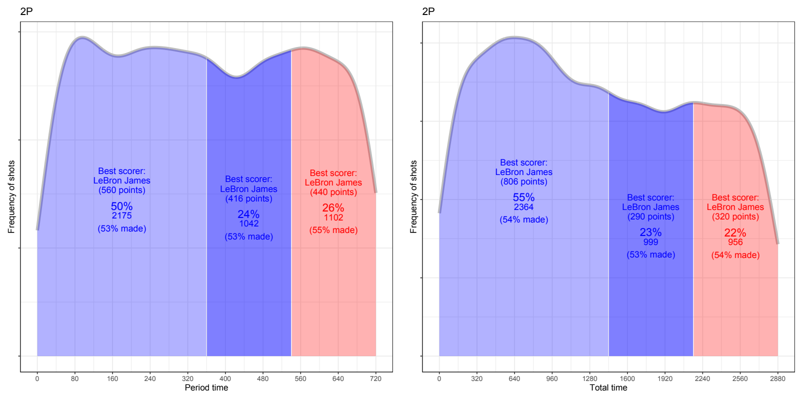

Figure 12: Density estimation of the 2-point shots by Cleveland Cavaliers with respect to period time (left) and total time (right).

teammates.

Q10: How does the frequency of 2-point shots by Cleveland Cavaliers vary during the match?

The analysis of the shooting frequency in time (or even in space) is a very relevant topic in basketball analytics and requires play-by-play data. The function `densityplot` originates nice plots, like those in Figure 12, displaying the density estimation of the 2-point shots attempted by the players of Cleveland Cavaliers, with respect to two concurrent variables: the period time (left plot), i.e. the time played in the quarter (in seconds) and the total time (right plot), i.e. the time played in the match (in seconds). This analysis is made possible thanks to the availability, in the play-by-play data, of total time and play time for every single shot.

The Cleveland Cavaliers tend to concentrate their shots in the first half of the match (55%). Looking at the period time, they tend to equally divide their 2-point shots in the two halves of each quarter, with a slightly higher concentration in the last part (26%) than in the third one (24%). The team's success percentages are fairly stable (53%-55%) during the different phases of the match or of each quarter. In every subperiod of time, LeBron James is always the best scorer (in brackets, the total number of points he scored). Several other interesting notes can be drawn comparing these results with those of the opponents of Cleveland Cavaliers or by focusing on some specific players.

Q11: Which is the best distance to shoot from for LeBron James? And for his teammates?

With the function `expectedpts` we can estimate the expected points of one team or single players with respect to some variables, like the shot distance or the total time played in the match. Focusing on the expected points rather than on the scoring probability allows to determine, for each player, his situations of maximum efficiency (for example, the best distance to shoot from), considering both the points scored and their scoring probabilities.

Figure 13 shows the expected points by LeBron James (left plot) and all the players of Cleveland Cavaliers who scored more than 500 points (right plot) in function of the shot distance. The maximum efficiency of LeBron James is for the distances where the red line is above the grey line (team average): when he shoots from a distance higher than 22 feet (i.e., in 3-point shots) he clearly overperforms in terms of expected points. Actually, he performs better than the team average from every distance, except when shooting from a distance between 10 to 22 feet. From that distance, Kyle Korver, Kevin Love, and JR Smith perform better than LeBron James and the team average. This result can really help to identify the best players the team can count on in each spot of the court and finally to define a winning game strategy.

5 Conclusions

In this contribution, the focus was on performance analysis in basketball, in an era in which most coaches and their backroom staff rely on formulas and figures to predict the most effective methods for winning. This is especially true in the US, where NBA is leading the major transformation related to the use of analytics: experts rely on data to measure a team's probability of winning and to assess a player's or a team's performance. Assessments based on the "eye-test", that is the

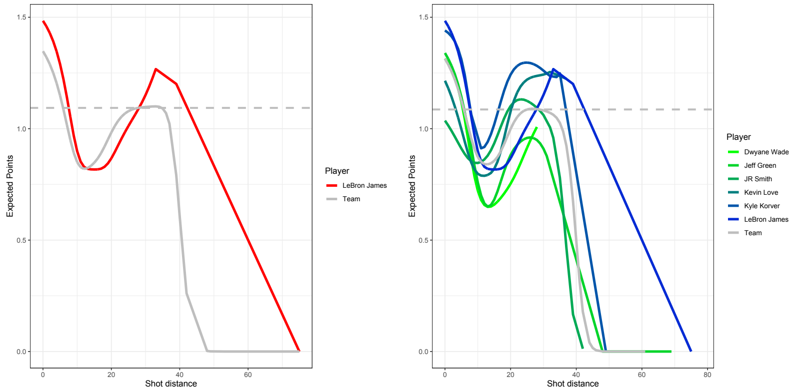

Figure 13: Expected points from a given distance (dashed line: team average independent from shot distance) of shots by LeBron James (left) and all the Cleveland Cavaliers players who scored more than 500 points (right).

impression that came from watching a game, are out of fashion. The attention is in particular on basketball analytics used to analyze performance issues related to players and teams, playing patterns, game strategies, performance drivers, interactions among players.

After briefly introducing the state of the art of basketball analytics and the different basketball data sets existing to perform statistical analyses, we offered an overview of possible basic and advanced investigations in performance analysis in basketball, with focus on some of the opportunities offered by the R package `BasketballAnalyzeR`. In order to show that statistics are more than numbers, we replied to some of the possible questions that could be posed by basketball fans, in a sort of Q&A section with Questions from fans and Answers by basketball analysts.

Basketball analytics can give the appropriate answer to many other questions. We believe that `BasketballAnalyzeR` can be a valid help for both basketball fans without a strong statistical background and expert analysts, who can subsequently apply their sophisticated methods to the results obtained from the proposed analyses.

The quality of the available data plays a crucial role in analytics. We focused on NBA data, which are complete and of good quality.

Steps are being taken also in other countries, leagues and championships, in order to improve the data collection and finally obtain high-quality data, which is the ingredient needed to perform accurate analyses with reliable results that can be appreciated by sport professionals and fans. This can foster the spread of the statistical culture and finally refine the understanding of basketball analytics in a virtuous circle that is good for both basketball and statistics.

Acknowledgments

The Authors wish to thank BigDataBall (www.bigdataball.com) for making data available.

References

[1] Aglioti, S.M., Cesari, P., Romani, M., Urgesi, C., 2008. Action anticipation and motor resonance in elite basketball players. Nature Neuroscience 11, 1109.

[2] Alagappan, M., 2012. From 5 to 13: Redefining the positions in basketball. MIT Sloan Sports Analytics Conference .

[3] Alamar, B.C., 2013. Sports analytics: A guide for coaches, managers, and other decision makers. Columbia University Press, New York.

[4] Albert, J., Glickman, M.E., Swartz, T.B., Koning, R.H., 2017. Handbook of Statistical Methods and Analyses in Sports. CRC Press, Boca Raton, FL.

[5] Annis, D.H., 2006. Optimal end-game strategy in basketball. Journal of Quantitative Analysis in Sports 2, –.

[6] Ante, P., Slavko, T., Igor, J., 2014. Interdependencies between defence and offence in basketball. Sport Science 7, 62–66.

[7] Arkes, J., 2010. Revisiting the hot hand theory with free throw data in a multivariate framework. Journal of Quantitative Analysis in Sports 6, 1–12.

[8] Avugos, S., Köppen, J., Czienskowski, U., Raab, M., Bar-Eli, M., 2013. The "hot hand" reconsidered: A meta-analytic approach. Psychology of Sport and Exercise 14, 21–27.

[9] Bar-Eli, M., Avugos, S., Raab, M., 2006. Twenty years of “hot hand” research: Review and critique. Psychology of Sport and Exercise 7, 525–553.

[10] Barber, J.M., Rollins, E.S., 2019. Factors influencing scoring in the NBA slam dunk contest. Journal of Sports Analytics 5, 223–245.

[11] Bianchi, F., Facchinetti, T., Zuccolotto, P., 2017. Role revolution: towards a new meaning of positions in basketball. Electronic Journal of Applied Statistical Analysis 10, 712–734.

[12] Bornn, L., Cervone, D., Franks, A., Miller, A., 2017. Studying basketball through the lens of player tracking data, in: Handbook of Statistical Methods and Analyses in Sports. Chapman and Hall/CRC, Boca Raton, FL, pp. 245–269.

[13] Bourbousson, J., Sève, C., McGarry, T., 2010a. Space–time coordination dynamics in basketball: Part 1. intra-and inter-couplings among player dyads. Journal of Sports Sciences 28, 339–347.

[14] Bourbousson, J., Sève, C., McGarry, T., 2010b. Space–time coordination dynamics in basketball: Part 2. the interaction between the two teams. Journal of Sports Sciences 28, 349–358.

[15] Brown, M., Sokol, J., 2010. An improved LRMC method for NCAA basketball prediction. Journal of Quantitative Analysis in Sports 6, 1–23.

[16] Clemente, F.M., Martins, F.M.L., Kalamaras, D., Mendes, R.S., 2015. Network analysis in basketball: inspecting the prominent players using centrality metrics. Journal of Physical Education and Sport 15, 212.

[17] Cooper, W.W., Ruiz, J.L., Sirvent, I., 2009. Selecting non-zero weights to evaluate effectiveness of basketball players with DEA. European Journal of Operational Research 195, 563–574.

[18] Csataljay, G., O’Donoghue, P., Hughes, M., Dancs, H., 2009. Performance indicators that distinguish winning and losing teams in basketball. International Journal of Performance Analysis in Sport 9, 60–66.

[19] De Rose, D.J., 2004. Statistical analysis of basketball performance indicators according to home/away games and winning and losing teams. Journal of Human Movement Studies 47, 327–336.

[20] Deshpande, S.K., Jensen, S.T., 2016. Estimating an NBA player’s impact on his team’s chances of winning. Journal of Quantitative Analysis in Sports 12, 51–72.

[21] Engelmann, J., 2017. Possession-based player performance analysis in basketball (adjusted +/– and related concepts), in: Handbook of Statistical Methods and Analyses in Sports. Chapman and Hall/CRC, pp. 215–227.

[22] Erčulj, F., Štrumbelj, E., 2015. Basketball shot types and shot success in different levels of competitive basketball. PloS one 10, e0128885.

[23] Fearnhead, P., Taylor, B.M., 2011. On estimating the ability of NBA players. Journal of Quantitative Analysis in Sports 7.

[24] Fewell, J.H., Armbruster, D., Ingraham, J., Petersen, A., Waters, J.S., 2012. Basketball teams as strategic networks. PloS one 7, e47445.

[25] Franks, A.M., D'Amour, A., Cervone, D., Bornn, L., 2016. Meta-analytics: tools for understanding the statistical properties of sports metrics. Journal of Quantitative Analysis in Sports 12, 151–165.

[26] Fujimura, A., Sugihara, K., 2005. Geometric analysis and quantitative evaluation of sport teamwork. Systems and Computers in Japan 36, 49–58.

[27] García, J., Ibáñez, S.J., De Santos, R.M., Leite, N., Sampaio, J., 2013. Identifying basketball performance indicators in regular season and playoff games. Journal of Human Kinetics 36, 161–168.

[28] Gilovich, T., Vallone, R., Tversky, A., 1985. The hot hand in basketball: On the misperception of random sequences. Cognitive Psychology 17, 295–314.

[29] Goldman, M., Rao, J.M., 2012. Effort vs. concentration: the asymmetric impact of pressure on NBA performance. MIT Sloan Sports Analytics Conference .

[30] Gómez Sánchez, J., Moll Sotomayor, J.A., Pila Teleña, A., 1980. Baloncesto: Técnica de entrenamiento y dirección de equipo. Madrid, Pila Teleña .

[31] Groll, A., Manisera, M., Schauberger, G., Zuccolotto, P., 2018. Guest Editorial 'Statistical modelling for sports analytics'. Statistical Modelling 18, 385–387. URL: https://doi.org/10.1177/1471082x18810264, doi:10.1177/1471082x18810264.

[32] Groll, A., Manisera, M., Schauberger, G., Zuccolotto, P., 2019. Guest Editorial 'Statistical modelling for sports analytics'. Statistical Modelling 19, 3–4. URL: https://doi.org/10.1177/1471082x18810965, doi:10.1177/1471082x18810965.

[33] Gudmundsson, J., Horton, M., 2016. Spatio-temporal analysis of team sports. A survey. arXiv preprint arXiv:1602.06994 .

[34] Gupta, A.A., 2015. A new approach to bracket prediction in the NCAA men's basketball tournament based on a dual-proportion likelihood. Journal of Quantitative Analysis in Sports 11, 53–67.

[35] Ibáñez, S.J., García, J., Feu, S., Lorenzo, A., Sampaio, J., 2009. Effects of consecutive basketball games on the game-related statistics that discriminate winner and losing teams. Journal of Sports Science and Medicine 8, 458–462.

[36] Ibáñez, S.J., Sampaio, J., Sáenz-López Buñuel, P., Giménez Fuentes-Guerra, J., Janeira, M., 2003. Game statistics discriminating the final outcome of junior world basketball championship matches (Portugal 1999). Journal of Human Movement Studies 45, 1–20.

[37] Kay, H.K., 1966. A statistical analysis of the profile technique for the evaluation of competitive basketball performance. Ph.D. thesis. University of Alberta.

[38] Kenett, R.S., Redman, T.C., 2019. The Real Work of Data Science: Turning data into information, better decisions, and stronger organizations. John Wiley & Sons, Hoboken, US.

[39] Koehler, J.J., Conley, C.A., 2003. The "hot hand" myth in professional basketball. Journal of Sport and Exercise Psychology 25, 253–259.

[40] Koh, K.T., Wang, C.K.J., Mallett, C., 2011. Discriminating factors between successful and unsuccessful teams: A case study in elite youth Olympic basketball games. Journal of Quantitative Analysis in Sports 7.

[41] Koh, K.T., Wang, C.K.J., Mallett, C., 2012. Discriminating factors between successful and unsuccessful elite youth Olympic female basketball teams. International Journal of Performance Analysis in Sport 12, 119–131.

[42] Kubatko, J., Oliver, D., Pelton, K., Rosenbaum, D.T., 2007. A starting point for analyzing basketball statistics. Journal of Quantitative Analysis in Sports 3, 1–22.

[43] Lamas, L., De Rose Jr., D., Santana, F.L., Rostaiser, E., Negretti, L., Ugrinowitsch, C., 2011. Space creation dynamics in basketball offence: validation and evaluation of elite teams. International Journal of Performance Analysis in Sport 11, 71–84.

[44] Loeffelholz, B., Bednar, E., Bauer, K.W., 2009. Predicting NBA games using neural networks. Journal of Quantitative Analysis in Sports 5, 1–15.

[45] Lopez, M.J., Matthews, G.J., 2015. Building an NCAA men's basketball predictive model and quantifying its success. Journal of Quantitative Analysis in Sports 11, 5–12.

[46] Lorenzo, A., Gómez, M.Á., Ortega, E., Ibáñez, S.J., Sampaio, J., 2010. Game related statistics which discriminate between winning and losing under-16 male basketball games. Journal of Sports Science & Medicine 9, 664.

[47] Madden, C.C., Kirkby, R.J., McDonald, D., Summers, J.J., Brown, D.F., King, N.J., 1995. Stressful situations in competitive basketball. Australian Psychologist 30, 119–124.

[48] Madden, C.C., Summers, J.J., Brown, D.F., 1990. The influence of perceived stress on coping with competitive basketball. International Journal of Sport Psychology 21, 21–35.

[49] Manisera, M., Sandri, M., Zuccolotto, P., 2019. `BasketballAnalyzeR`: the R package for basketball analytics., in: Conference Smart Statistics for Smart Applications, Pearson, Milan, Italy. pp. 395–402.

[50] Manner, H., 2016. Modeling and forecasting the outcomes of NBA basketball games. Journal of Quantitative Analysis in Sports 12, 31–41.

[51] McFarlane, P., 2019. Evaluating nba end-of-game decision-making. Journal of Sports Analytics 5, 17–22.

[52] Metulini, R., Manisera, M., Zuccolotto, P., 2017a. Sensor analytics in basketball, in: Francesco, C.D., Giovanni, L.D., Ferrante, M., Fonseca, G., Lisi, F., Pontarollo, S. (Eds.), Proceedings of the 6th International Conference on Mathematics in Sport, Padova University Press, Padova, Italy. pp. 265–274.

[53] Metulini, R., Manisera, M., Zuccolotto, P., 2017b. Space-time analysis of movements in basketball using sensor data, in: Petrucci, A., Verde, R. (Eds.), Statistics and Data Science: new challenges, new generations - Proceedings of the SIS 2017 Conference of the Italian Statistical Society, Firenze University Press, Firenze, Italy. pp. 701–706.

[54] Metulini, R., Manisera, M., Zuccolotto, P., 2018. Modelling the dynamic pattern of surface area in basketball and its effects on team performance. Journal of Quantitative Analysis in Sports 14, 117–130.

[55] Miller, A.C., Bornn, L., 2017. Possession sketches: Mapping NBA strategies. MIT Sloan Sports Analytics Conference 2017 .

[56] Miller, S., Bartlett, R., 1996. The relationship between basketball shooting kinematics, distance and playing. Journal of Sports Sciences 14, 243–253.

[57] Miller, T.W., 2015. Sports analytics and data science: winning the game with methods and models. FT Press.

[58] Okubo, H., Hubbard, M., 2006. Dynamics of the basketball shot with application to the free throw. Journal of Sports Sciences 24, 1303–1314.

[59] de Oliveira, R.F., Oudejans, R.R., Beek, P.J., 2006. Late information pick-up is preferred in basketball jump shooting. Journal of Sport Sciences 24, 933–940.

[60] Oliver, D., 2004. Basketball on paper: rules and tools for performance analysis. Potomac Books, Inc.

[61] Özmen, U.M., 2012. Foreign player quota, experience and efficiency of basketball players. Journal of Quantitative Analysis in Sports 8, 1–18.

[62] Page, G.L., Barney, B.J., McGuire, A.T., 2013. Effect of position, usage rate, and per game minutes played on NBA player production curves. Journal of Quantitative Analysis in Sports 9, 337–345.

[63] Page, G.L., Fellingham, G.W., Reese, S.C., 2007. Using box-scores to determine a position's contribution to winning basketball games. Journal of Quantitative Analysis in Sports 3, 1.

[64] Passos, P., Araújo, D., Volossovitch, A., 2016. Performance analysis in team sports. Taylor & Francis.

[65] Passos, P., Davids, K., Araújo, D., Paz, N., Minguéns, J., Mendes, J., 2011. Networks as a novel tool for studying team ball sports as complex social systems. Journal of Science and Medicine in Sport 14, 170–176.

[66] Perše, M., Kristan, M., Kovačič, S., Vučkovič, G., Perš, J., 2009. A trajectory-based analysis of coordinated team activity in a basketball game. Computer Vision and Image Understanding 113, 612–621.

[67] Piette, J., Anand, S., Zhang, K., 2010. Scoring and shooting abilities of NBA players. Journal of Quantitative Analysis in Sports 6.

[68] Piette, J., Pham, L., Anand, S., 2011. Evaluating basketball player performance via statistical network modeling. MIT Sloan Sports Analytics Conference .

[69] Ruiz, F.J., Perez-Cruz, F., 2015. A generative model for predicting outcomes in college basketball. Journal of Quantitative Analysis in Sports 11, 39–52.

[70] Sampaio, J., Drinkwater, E.J., Leite, N.M., 2010. Effects of season period, team quality, and playing time on basketball players' game-related statistics. European Journal of Sport Science 10, 141–149.

[71] Sampaio, J., Janeira, M., 2003. Statistical analyses of basketball team performance: understanding teams' wins and losses according to a different index of ball possessions. International Journal of Performance Analysis in Sport 3, 40–49.

[72] Sandri, M., Forthcoming. The r package basketballanalyzer, in: Basketball data science (Chapter 6). Chapman and Hall/CRC.

[73] Severini, T.A., 2014. Analytic methods in sports: Using mathematics and statistics to understand data from baseball, football, basketball, and other sports. Chapman and Hall/CRC.

[74] Shortridge, A., Goldsberry, K., Adams, M., 2014. Creating space to shoot: quantifying spatial relative field goal efficiency in basketball. Journal of Quantitative Analysis in Sports 10, 303–313.

[75] Silverman, B.W., 2018. Density estimation for statistics and data analysis. Routledge.

[76] Skinner, B., 2010. The price of anarchy in basketball. Journal of Quantitative Analysis in Sports 6.

[77] Skinner, B., Goldman, M., 2017. Optimal strategy in basketball, in: Handbook of Statistical Methods and Analyses in Sports. Chapman and Hall/CRC, pp. 229–244.

[78] Therón, R., Casares, L., 2010. Visual analysis of time-motion in basketball games, in: Smart Graphics, Springer. pp. 196–207.

[79] Travassos, B., Araújo, D., Davids, K., Esteves, P.T., Fernandes, O., 2012. Improving passing actions in team sports by developing interpersonal interactions between players. International Journal of Sports Science & Coaching 7, 677–688.

[80] Trninić, S., Dizdar, D., Lukšić, E., 2002. Differences between winning and defeated top quality basketball teams in final tournaments of European club championship. Collegium Antropologicum 26, 521–531.

[81] Tversky, A., Gilovich, T., 2005. The cold facts about the "hot hand" in basketball. Anthology of Statistics in Sports 16, 169.

[82] Vergin, R., 2000. Winning streaks in sports and the misperception of momentum. Journal of Sport Behavior 23, 181.

[83] Vračar, P., Štrumbelj, E., Kononenko, I., 2016. Modeling basketball play-by-play data. Expert Systems with Applications 44, 58–66.

[84] West, B.T., 2008. A simple and flexible rating method for predicting success in the NCAA basketball tournament: Updated results from 2007. Journal of Quantitative Analysis in Sports 4, 8.

[85] Winston, W.L., 2012. Mathletics: How gamblers, managers, and sports enthusiasts use mathematics in baseball, basketball, and football. Princeton University Press.

[86] Wu, S., Bornn, L., 2018. Modeling offensive player movement in professional basketball. The American Statistician 72, 72–79.

[87] Yuan, L.H., Liu, A., Yeh, A., Kaufman, A., Reece, A., Bull, P., Franks, A., Wang, S., Illushin, D., Bornn, L., 2015. A mixture-of-modelers approach to forecasting NCAA tournament outcomes. Journal of Quantitative Analysis in Sports 11, 13–27.

[88] Zhang, T., Hu, G., Liao, Q., 2013. Analysis of offense tactics of basketball games using link prediction, in: Computer and Information Science (ICIS), IEEE/ACIS 12th International Conference, IEEE. pp. 207–212.

[89] Zuccolotto, P., Manisera, M., Forthcoming. Basketball data science. Chapman and Hall/CRC.

[90] Zuccolotto, P., Manisera, M., Kenett, R., 2017. Guest Editorial 'Statistics in sports'. Electronic Journal of Applied Statistical Analysis 10, 1–2.

[91] Zuccolotto, P., Manisera, M., Sandri, M., 2018. Big data analytics for modeling scoring probability in basketball: The effect of shooting under high-pressure conditions. International Journal of Sports Science & Coaching 13, 569–589.

[92] Zuccolotto, P., Sandri, M., Manisera, M., 2019. Spatial performance indicators and graphs in basketball. Social Indicators Research URL: https://doi.org/10.1007/s11205-019-02237-2, doi:10.1007/s11205-019-02237-2.