An Interactive Dashboard for Comparing NBA Players' Proficiency

Arnold LEE Wai Tong Lee, Alson TAN Yong Kiong an, Lai Ho LIM

***Abstract* —** With the rise in Big Data Analytics, more organizations are interested in collecting data for analysis. Through the collection of data, organizations hope to discover fresh insights, allowing them to improve their existing processes and gain an advantage over their competitors. NBA is one of such organizations to employ data analytics. In 2013, NBA installed player-tracking systems in its arenas [1]. The new data gathered offered more knowledge to the NBA community than ever before. NBA coaches can evaluate his players' performance more accurately, such as from which distance does the player have the highest Field Goal Percentage and how he stands relative to other players. This helps coaches to identify each player's strength and weakness more accurately, allowing them to improve their gaming strategies. Similarly, basketball scouters can also turn to data analytics to evaluate players. To facilitate comparison between players, we built an interactive dashboard consisting of parallel coordinates, radar chart and line graphs to allow easy comparison of players' stats.

***Keywords* —** *NBA, NBA players, Visualization*

1. INTRODUCTION

The National Basketball Association (NBA) is the pre-eminent men's professional basketball league in North America, and is widely considered to be the premier men's professional basketball league in the world. NBA is a billion dollar business [2]. According to a report by ESPN, the New York Yankies are worth $3.4 billion dollars [3]. Not only do the winning teams get a large cash payout, they are often flooded with endorsement deals. Exceptional NBA players like LeBron James, may receive even more from commercial endorsements. LeBron has signed a lifetime deal with Nike which is estimated to worth more than $1 billion dollars [4]. With so much at stake, it is no wonder that team managers and coaches are doing everything they can to improve their team's winning odds and recruiting the best players.

Big data and sports analytics is increasingly gaining attention in NBA. The Houston Rockets is one team that employs data analytics to improve their game strategy. According to an article by The Atlantic, the team rarely attempts long-range two-point jumpshots because it has the worst scoring probability to point ratio [1]. Such insights help NBA teams and players to find out what strategies are most rewarding and most suitable to carry out during their games. Recognizing the strong benefits that data analytics provide, most NBA teams now have a staff [5] dedicated to carry out analytics on the team to maximize their chances of winning. He is able to advise each player on their strengths and weaknesses and offer recommendations to improve their techniques and strategies. To support the data analytics phenomenon, NBA installed player-tracking systems in its arenas to facilitate accessible data collection [1].

This paper aims to report on our efforts in researching and developing this visualization. Section 1 provides a general background on the NBA industry. In Section 2, we will cover our motivation and objectives. In Section 3, we will discuss our data sourcing and transformation efforts. In Section 4 and 5, we will cover on our user interface design and application. Lastly, we will conclude our paper in Section 6.

2. MOTIVATION AND OBJECTIVES

Even with the increasing availability of NBA sports data, we realize that there is a lack of visualizations that evaluates and compares players against one another. Having such a visualization has several benefits. First, the visualization serves as a performance benchmark for players. Poorer performing players can identify the areas in which they are weaker in and work on those areas. Similarly, coaches can also use the tool to monitor their players and create more targeted training programmes for them. Second, scouters can use the visualization to identify top players to join their ranks. They can view and understand decisive stats quickly and compare how the player fares against other players. In all, we believe

that this visualization can offer greater analysis and awareness of each player's strengths and weaknesses. Players may also then drive themselves further to improve their skills and raise the sport of basketball to a higher level. For our visualization, we aim to:

1. Provide an overview of NBA players' statistics
2. Compare individual player's strengths and weaknesses
3. Analyze each player's shooting proficiency

3. Data Transformation

Data Sourcing

Our main dataset is obtained from Kaggle [6] and contains shots data made during the 2014-15 NBA Season. Examples of some of its contents are: if the shots were made, shot distance from the goal, 2 or 3-point shot, player name etc. As we did some initial exploration with the dataset, we realized that the players have different positions and roles. Their roles may determine their playing style and behaviour on the court. We felt that players' positions were a vital information that was missing and had to source for it. Eventually, we crawled the players' position data from ESPN website [7].

Data Transformation

In order to evaluate the players', it was necessary to do an overall analysis of their performance. The raw data recorded each shot in a single row and thus it was necessary to combine and aggregate the data.

Limitations

During our exploration, we found that some players had only one record. This made data visualization on these players not possible as there is no data to work with. As such, we removed these players from the dataset.

4. User Interface Design

For our application, we have created two dashboards. The first dashboard focuses on providing an overview and comparing between different players. The second dashboard focuses more in-depth on each player's stats and shooting performance.

Dashboard 1: Overview

For this dashboard, the main visualization we used is parallel coordinates. Parallel coordinates is an effective visualization technique to analyze multivariable data. In our case, we use the visualization to provide an overview of all the players' performance. Each vertical column corresponds to a dimension we want to display while each horizontal line corresponds to a single NBA player. By plotting all the players' stats onto the same graph, we can quickly identify general patterns and recognize better players that rise above the rest.

To further improve users' experience, the parallel coordinates visualization is interactive. By hovering over different lines, the graph will highlight the selected players. This aids the user visually, allowing him to trace his eyes easily to see the player's stats and how he compares with other players. Additionally, the user can brush a region along each vertical column to filter players within that range. Only players that are brushed within the filtered region will be displayed. This can be done concurrently with other vertical columns.

Due to development constraints, we were unable to include the player's name onto the parallel coordinates graph itself. This meant that the user would not be able to identify which player belonged to the highlighted line. To overcome this problem, we included a summary table below which displays the highlighted players' name and stats. Also on the right, we created a dropdown list in which the user can select which players to highlight for comparison. A search function is also available if the user wishes to find a specific player.

Figure 1: Dashboard 1 - Overview

Dashboard 2: Player Analysis

In the second dashboard, we focus on the player's shooting performance. As NBA teams are small, each player must be proficient in shooting goals. In this dashboard, we use a radar chart, line and bar graphs.

On the radar chart, we identify the most crucial shooting statistics and plot them. There are four plots against four axes, each corresponding to a stat that we are interested in. As each plot goes further from the center, it suggests that the player is more proficient in that area. As such, a

good player who scores highly on all four stats would have a large shaded polygon area. This offers the user a sensing of how "proficient" the player is.

On the line graph, we plot the player's shots based on distance. This supplements the radar chart to offer more insight about the player's shooting pattern. The red dotted line in the middle is the 3-point line.

In the top line graph, we plot the player's shot frequency. This is done by taking all the shots made by the player and arranging them into the distance they were shot at. From this, the user can have a visual appreciation from where does the player prefer to shoot from and his playing style. From the example below, we can see that player A tends to make his shots from 22ft onwards whereas player B tends to make shots closer to the goal.

Player A

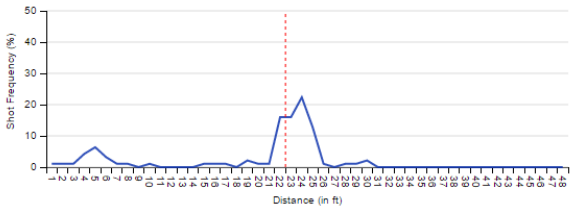


Figure 2: Example - Player A

Player B

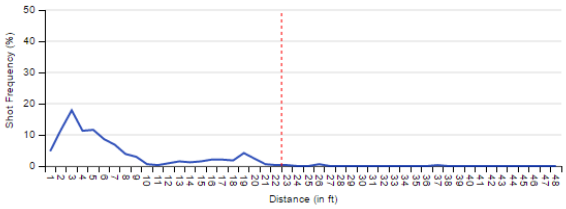
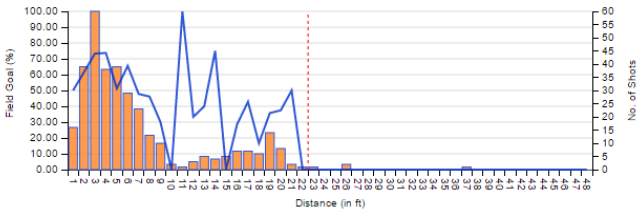


Figure 3: Example - Player B

In the lower line graph, we plot the player's Field Goal Percentage and Number of Shots. The Field Goal Percentage (FG%) is represented by the line graph whereas the Number of Shots is represented by the bar graph. The FG% is the ratio of shots made at that distance divided by the total shots at that distance; the higher the percentage the better. However, in some cases due to low shot count, this statistic may not be reliable. From the example below, we observe that the player has a 100 FG% at 11ft. On closer examination, we realized that this is because the player only made one shot from that distance at made the goal. It is premature to conclude that the player is good at shooting from 11ft because of the high FG%. Noting this flaw, we added the Number of Shots bar graph to complement the FG% graph. This allows the user to check the FG% reliability by comparing it against the number of shots made. The more shots the player made at that distance, the more reliable the FG%.



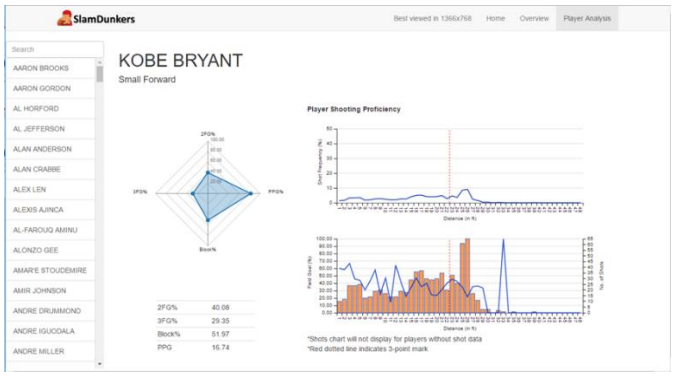


Figure 4: Dashboard 2 - Player Analysis

5. The Application

Data Visualization

To create our visualization, we decided to employ D3.js. D3.js is a JavaScript library for creating graphics from data [8]. The library is open sourced and is widely used by many developers around the world to generate charts dynamically. D3.js is also well supported with many good demo examples and forum support from its community. Furthermore, it is well integrated with HTML making it easy to develop alongside web development. However, we realized that most examples on D3.js were basic and did not provide sufficient interaction which we hoped. As such, we explored further into using other JavaScript libraries which are built on top of D3.js like Dimple.js and DC.js. These extended libraries offered more interactivity which we found more fitting for our visualization.

Search and filter

To improve user experience, it was necessary to find a way to allow user to have a search and mouse over feature. To achieve this, we used AngularJS to power the search and display function of our web application. AngularJS is a front-end web framework developed by Google. It improves user experience by eliminating the need to refresh the page each time we need to render a new page. In our case, we used AngularJS to manage and maintain the list of players on the side. Users can simply enter a player they wish to find and the list will filter accordingly.

6. CONCLUSION

In conclusion, we believe that there are much more insights that can be gained from NBA data. As NBA opens up its data sources and API, developers can leverage on the data to make new connections and draw fresh insights. This will make NBA games even more exciting as it may no longer be just having the best players but rather devising the best strategies from analytics. For this project, data limitations were a major stumbling block as it proved difficult to find a suitable and comprehensive dataset for analysis. We had hoped to go further and create a court map to plot the different spots where the shots were made. Such a visualization may offer more interesting patterns about players' shooting habits. This can be a future improvement.

Our team has also gained good exposure building web data visualizations and using data-driven libraries such as D3.js. These libraries allow us more capability to build and design our own custom visualizations in the future.

7. REFERENCE

[1] T. F. Ross, "Welcome to Smarter Basketball," 15 June 2015. [Online]. Available: http://www.theatlantic.com/entertainment/archive/2015/06/nba-data-analytics/396776/.

[2] K. Badenhausen, "New York Knicks Head The NBA's Most Valuable Teams At $3 Billion," Forbes, 20 January 2016. [Online]. Available: http://www.forbes.com/sites/kurtbadenhausen/2016/01/20/new-york-knicks-head-the-nbas-most-valuable-teams-at-3-billion/#7020ac392d3a.

[3] D. Rovell, "Forbes estimates Yankees worth MLB-high $3.4 billion," ESPN, 24 March 2016. [Online]. Available: http://www.espn.com/mlb/story/_/id/15051301/new-york-yankees-worth-mlb-high-34-billion.

[4] "Sports Illustrated," [Online]. Available: http://www.si.com/nba/2016/05/17/lebron-james-nike-deal-contract-one-billion.

[5] "NBAStuffer," [Online]. Available: http://www.nbastuffer.com/component/option,com_glossary/Itemid,0/catid,44/func,view/term,NBA%20Teams%20That%20Have%20Analytics%20Department/

[6] "Kaggle NBA Shots," [Online]. Available: https://www.kaggle.com/dansbecker/nba-shot-logs/.

[7] "ESPN," [Online]. Available: http://www.espn.com/nba/players.

[8] "D3.js," [Online]. Available: https://d3js.org/.