

# **2019-20 NBA Season Players Comparison Project**

Group Member: Yiqi Ren, Xiaopeng Lan, Zhongrui Wang, Yi Zhu

## **I. Introduction**

People also regard our present time as a time of "Big Data". Indeed, with the increasing progress of the data collecting techniques and the computational power, nowadays there are abundant amounts of data that exist in nearly every business and are waiting for people to use. So, it is for the sports industry. Sports fans now can easily get a much larger body of data than those professional scouts once could have back to the 50s. Then, however, the overwhelmed data also raise the question of efficiency. Then, all in all, what constitutes a good application for sport Data? This project is an experimental effort to solve the question; by incorporating both the graphic characteristics and the choice to compare the performances between different NBA players, this project presents a Shiny app based on the newest data from NBA official statistics for sports fans to better deal with the overwhelmed statistics.

As mentioned previously, the major feature for this app is the combination of visualized data and the comparison of the statistics between players. This app visualized six essential stats for 263 players through both bar chart and radar chart and allows the user to compare these two kinds of plots between any of two current NBA players. This innovation thus brought two advantages that made the app itself more user-friendly: first of all, through visualized data, users now can acquire a more direct impression of a certain player's performance than merely reading the data. Secondly, taking this convenience brought by visualization, the comparison based on the graphical approach further emphasizes the differences between players. In addition to that, our selection of the plot type is also suitable for comparison. While the histogram illustrates a certain stat among the seven stats better, the radar plot can help the user better explore the comprehensiveness and stability of players. Therefore, this project is an app that indeed helps the user to explore the benefits of Big Data while not being overwhelmed by it.

## **II. Data Summary:**

The data is acquired from the official statistics from the NBA website, consisting of the six statistical summaries that measure the average performance of each of the 236 players in 2020 seasons so far. The six statistics are the following:

PTS: The average points a player scored in each game so far

REB: The average rebounds of a player in each game so far

AST: The average assistances of a player in each game so far

STL: The average steals of a player in each game so far

BLK: The average blocks of a player in each game so far

EFF: The average efficiency of a player calculated through the efficiency function, a popular measurement of a player's performance that also takes time into consideration.

These six data are selected because they are both the most widely used statistics to measure a player's contribution. The PTS and AST are good measurements for a player's performance during the attack, while STL and BLK are for during defense. The REB is important in both attack and defense. The efficiency is a more general measurement of the performance.

### **III. Application Development**

The development of the application consists of two parts: the data scraping and the data visualization, which includes both the bar plot and the radar chart.

As mentioned previously, the data is acquired through dynamic web scraping techniques. As there are only limited numbers of data shown on each page, therefore, it is necessary to set up a remote drive to go through each page and scrape the data by emulating the action of clicking the button to the next page. This remote drive is achieved through the R package RSelenium. In terms of the web scraping process, this app takes advantage of HTML structure of the NBA website. By setting up the node according to each kind of statistics, we are thus able to scrape the data through R package rvest.

The data visualization consists of two different tasks: creating the bar chart for each player's statistics and the radar plot for them. In addition to that both the objects demand to be able to overlap with each other for comparative purposes. As we use the reactive tool provided by the shiny package, we are now able to create the column containing the information that the user selected, namely, the player's information. The plotting of the bar chart is conducted based on the reactive response we acquired through letting the users input the information they need through selecting the options provided by the app. The plotting itself is done through R package Plot-ly, as it is more suitable for the web presentation.

The radar plot required a different approach. Given each individual statistics varied in their ranges by the nature of basketball, to have a balanced shape of radar plot, the data need slightly more treatment. We then standardize the data to make the radar plot output more regularly. And we plot the radar plot through R package fmsb and colorized it through package RcolorBrewer.

#### **IV. User Guide**

NBA Stats Lite is a user-friendly app that is convenient for any basketball fan or anyone without much basketball knowledge to compare essential basketball statistics of any two active NBA players in the 2019-20 NBA regular season. We selected several basketball statistics which are essential to evaluate and compare the performances of basketball players. NBA Stats Lite visualizes the related essential attributes through a bar chart and radar chart.

*Procedure:*

1. Choose any two NBA players that you want to compare.
2. Select the tabs of essential attributes that you are interested in comparing.
3. Related comparison histogram and radar chart would appear afterwards.