NBA Player Statistics & Data Insights Herb Lau

In the past decade, analytics has made itself visible in professional sports in a way that is significantly moreso than I remember in my childhood of the 1980s. This is no less true in professional basketball’s the National Basketball Association (NBA) than it is in any other professional league. The primary objective for this project is to show interesting insights into this data on the professional basketball players, including such ideas as which players are not only the highest scoring players, but the most efficiently scoring players in the NBA. The secondary objective of this project is to show those insights visually with data visualization tools and packages available within the R statistical language and the RStudio development environment. The particular dataset used for this project was downloaded from the website <https://www.basketball-reference.com/leagues/NBA_2017_per_game.html> and loaded into a .csv file before being made a little easier to read including adjustments to the column names to make the data easier to read to the casual viewer.

The dataset consists of approximately 600 rows of data tracking all NBA players who played during the 2016-2017 season and approximately 30 columns representing the players’ measures in different statistical categories during the 2016-2017 regular season. The primary objective for this project is to show interesting insights into this data on the professional basketball players, including such ideas as which players are not only the highest scoring players, but the most efficiently scoring players in the NBA. The secondary objective of this project is to show those insights visually with data visualization tools and packages available within the R statistical language and the RStudio development environment.

One of the more interesting statistics for analysts looking at efficiency of the guard position (especially the point guard) is the assist to turnover ratio. Generally speaking, for a point guard (who would be considered most responsible for distributing and passing the basketball), a 3 to 1 assist to turnover ratio is considered excellent, while a 2 to 1 ratio is considered average at best. Within the data set, the Assist column is titled “AST” while the Turnover column is titled “TOV”. With R, I used a function titled “asst2turnover” to calculate the assist to turnover ratio for each player within the data set. There is a weird quirk with this calculation, as the code does run and produce a result, but there is an error displayed after the output.

As the project is further improved over time, adding the data created by this function and adding it to the data set would be in the works. As it is now, a cursory glance at both the results and data set tells me that Kyle Anderson, Jimmy Butler and Jerian Grant have assist to turnover ratios of more than 2.5 to 1, but that is still far below the highest measures of that statistic.

NBA Player Statistics & Data Insights Herb Lau

Next, there are boxplots of certain player statistics by player position, to better show the relative frequency and ability for certain positions to compile certain statistics over other statistics. For example, players at the positions of power forward and center (represented by “PF” and “C”) are near the basket more often than small forwards, point guards, and shooting guards (represneted by “SF”, “PG”, and “SG”) and therefore often get more rebounds, both defensive rebounds, offensive rebounds, and therefore total rebounds as well (represented by “DRB”, “ORB”, and TRB”.) (Rebounds are when a player gets the ball after a missed shot, either by a teammate (therefore, an offensive rebound because the team was in possession of the ball and trying to score) or by the opposing team (a defensive rebound.) On the other hand, point guards and shooting guards would normally get more assists (when an offensive player passes the ball to a player who proceeds to gain a field goal scored) than other players because they are expected to be in control of the ball more for their teams while they are on the offensive side. R did separate the entire dataset’s statistics in each specific statistic by the player’s position, creating hundreds of entries in each grouping.

One thing that is noticable from the data is that players at the center position are noticably better at collecting rebounds and blocks than those at other positions. This does match the expectation mentioned previously about centers and rebounds because of proximity to the basket. Another noticable idea is that players at the point guard position are much better at collecting assists than those at other positions. This also matches prevailing widom as mentioned before about the point guard and handling the ball, and is especially important when distinguishing between the shooting guard and the point guard. While the former does handle the ball often, he or she is not nearly expected to be the same level of facilitator of scoring opportunities for teammmates as is the point guard.

While the data on statistics by position of player is interesting visually as presented, it is not useful quite yet. It is promising, perhaps only to me though, as I find some of the ways the statistics are skewed and displayed curious and worthy of further development and manipulation.

I also wanted to better understand the top 20 players in field goals scored to understand their scoring efficiency. To do so, I compared their field goals scored to their minutes played. I then plotted them based on those numbers. Based on that, Lebron James, Karl-Anthony Towns, and Anthony Davis would probably be considered the players with the highest scoring efficiency. On the other hand, Klay Thompson, Bradley Beal, and DeMarcus Cousins would be the among the players with the lowest scoring efficiency, at least out of the 20 players with the most field goals made per game that is.

NBA Player Statistics & Data Insights Herb Lau

It should be noted that this does not include their free throws scored, which is a notable exception at this time. It should be noted that the field goals scored totals both the player’s 2 point and 3 point field goals, also a notable distinction. Devising a formula to allow better weighting for the different field goals and incorporate free throws would better represent a player’s scoring efficiency, and is a future goal for this project. Additionally, the plot doesn’t label the names of the players well, causing their names to be cut off by the edges of the plot. It would be much more useful to the casual viewer of the plot to have those names fully visible. Another issue with the player’s name on the label is that names can overlap often on the plot, causing the names to be illegible to the viewer.

Team scoring efficiency per game, compiled by averaging the field goals per game of all a team’s players and dividing by the average minutes played per game of all the team’s players, was calculated for all 30 NBA teams as well.

Overall, this project is a good first step into using R for data visualization. It shows a certain amount of value for data visualization within RStudio and the R statistical language, while suggesting areas where improvements could be made either to offer better, more useful insights to readers, or data visualization improvements such that the data could be more readily understood to viewers.