Insights into NBA: Looking at Player Statistics, Evolution of NBA, and the G.O.A.T Debate

1. Data

I gathered data from Kaggle and Github which included .csv and .xlsx files respectively based on information from Basketball-reference.com. A total of 6 files were used as preliminary data. These contained data on players, their statistics, seasons played, positions, country, performance statistics, salaries, and such variables. One file contained the weekly awards given to players over the years and corresponding statistics. The files contained data from various decades but the general years seen in most files were from the 1980s - 2020s, although some files had observations from the 1960s as well as the most recent 2022-23 season.

2. Research Questions

To gain insights into the NBA, I had several primary research questions.

i. USA vs the World: Do American and international (non-American) NBA players differ significantly in terms of their average points per game (PPG) during the regular season?
ii. East vs West: Is one of the conferences in the league more dominant than the other?
iii. The G.O.A.T debate: Who is the better NBA player – LeBron James or Michael Jordan?

Further although not direct research questions, I wanted to gain insights into the NBA. For this I looked at:

- How the number of points scored has evolved over time,
- The team and player with most weekly awards,
- The best (/ideal) player based on key primary statistics.

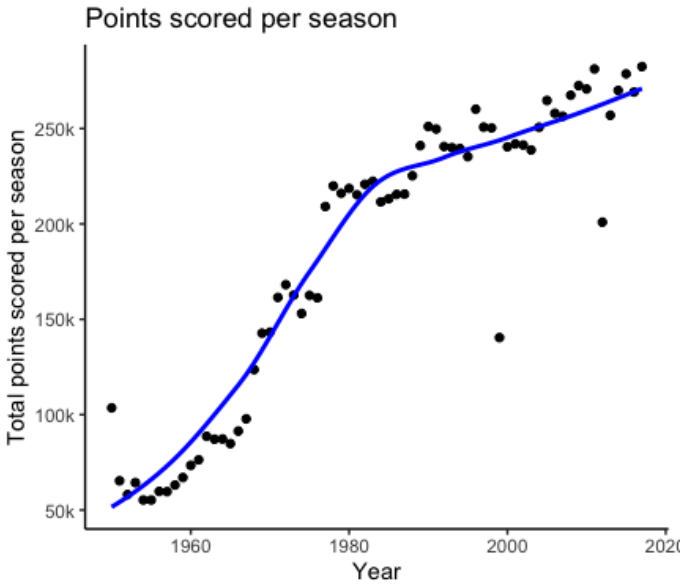
3. Analysis

To go about the analysis, I first merged all necessary datasets into groups based on the player's name. In such a fashion I could utilise several datasets and combine it into one larger dataset which increases the number of observations and the variables accounted for, increasing the reliability of the datasets. After that based on the question at hand, I went about grouping the data into required variables with filters I considered key for the visualization. For some relations, I used a scatterplot, while other times a bar graph, or histogram as I felt that they allowed better visualizations for the relation. For East vs West, the dataset was filtered grouped by the name of the player to see if they played in the eastern or western conference which was then plotted based on the number of weekly awards obtained by players. For USA vs the World, the dataset was grouped based on if they players were from USA or US territories, or from other regions of the world (!="USA"). Players who have been in the league since 2018 were considered since this allows them to be in the league for a few years to check for longevity. The top 10 players from USA and from other regions were considered based on the highest number of points per game (ppg) scored. The idea behind using ppg as a filter was more points a player scores, the better the player is, assumably. Some data like countries, coordinates needed to be manually added to the dataset with the top 10 players to map the values. The ppg scored by Americans and non-Americans was geospatially plotted on a map to see if there is one region more dominant than the other. A separate interactive plot was also created using the package leaflet which allowed to see the number of points scored and the name of the player. This visualization was also plotted as a histogram which showed a much vivid relation. LeBron vs Jordan – for this, values such as average points per game, assists, rebounds, and salary over the duration of their careers was filtered for to see if one player is better "theoretically" than the other. Ppg, assists, rebounds, and salary are key factors when assessing a player and their performance, thus these statistics were chosen. Finally, to check for the best player, a similar filter process was performed however with some additional filters to make sure that the players in the resulting dataset have a minimum number of points, assists, rebounds, and games played.

4. Findings & Conclusions

From Figure1, we can satisfactorily conclude the number of points scored every season in the NBA has increased over time with an almost linear relationship.

Figure 1: Points scored over the seasons in the NBA



From Figure2 we can evidently say that with an increase in the overall number of points scored over the seasons in the NBA, the number of three pointers scored have increased again almost co-linearly as well.

Figure 2 : The number of three pointers scored in the NBA over time

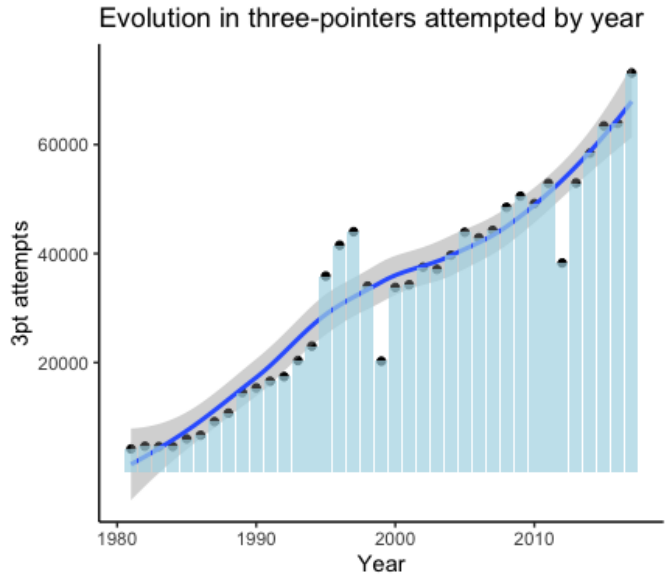


Figure 3 looks at the number of weekly awards given to each team in the NBA. Simply based on the number of awards, we can say that Los Angeles Lakers is the best team in the league. However, there are many more factors which should be considered when determining the best team in the league. Championships, wins, playoff appearances, and total spending are just some of the factors which should also be considered.

Figure 4 looks at the same weekly awards given to the East and the West conference. There is a difference in the number of awards given between the years 1985 – 2003, however after that the number of weekly awards is more or less uniform. Again, although the number of weekly awards can be used as a primary reference, it cannot be the sole determinant of assessing the more dominant conference.

Figure 3 : The number of weekly awards per team in the NBA

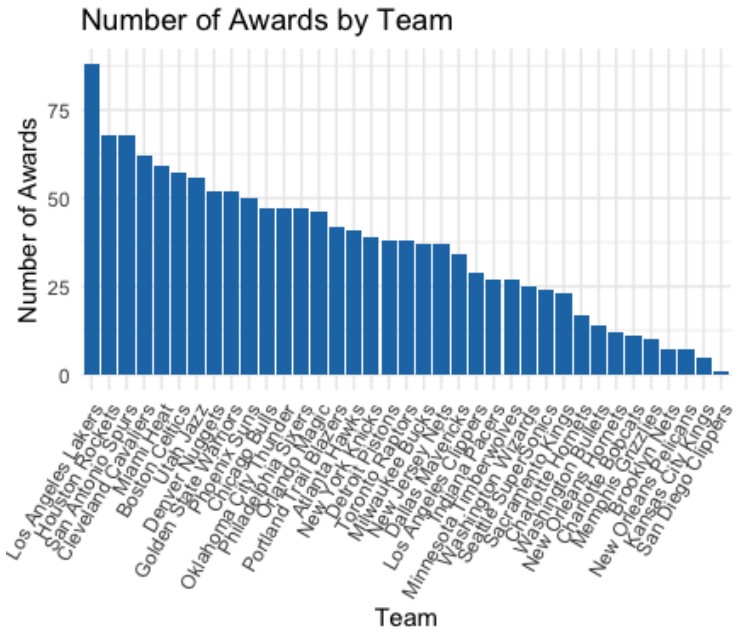


Figure 4 : Distribution of weekly awards based on the conference in the NBA

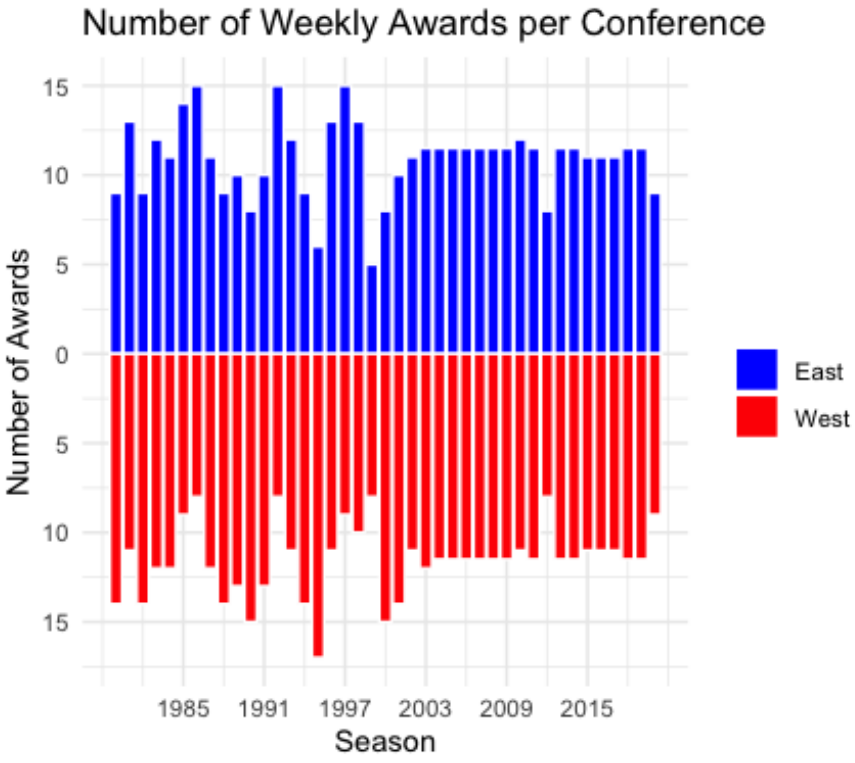
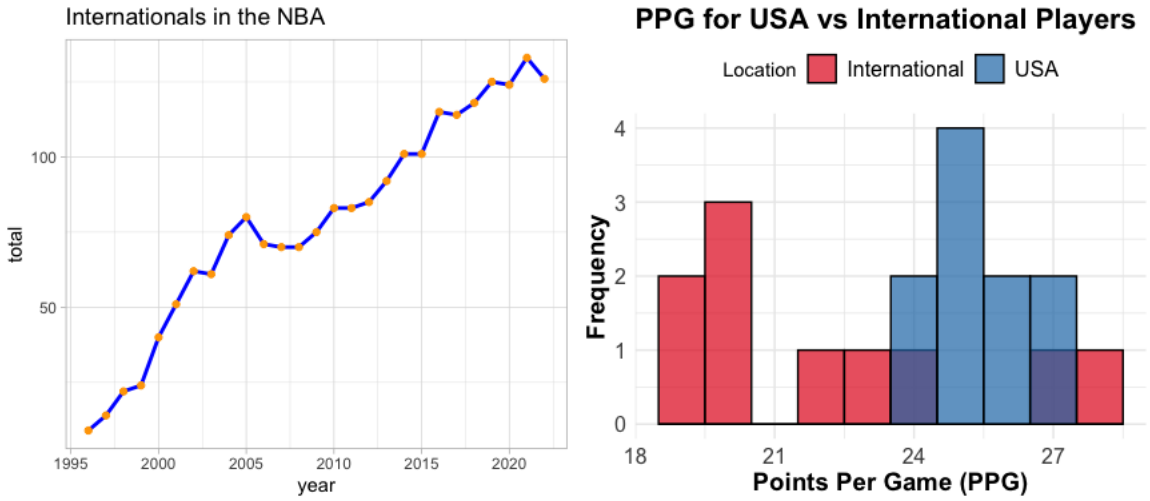
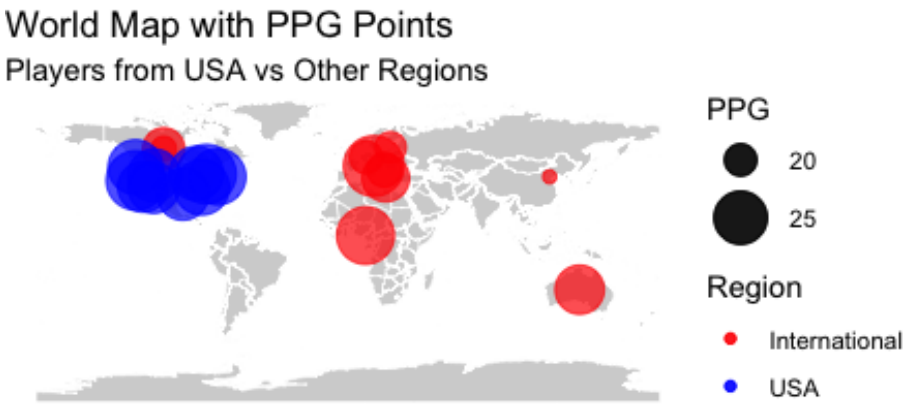


Figure 5 visualizes the trend between American and international players. We can successfully say that over time, the number of international players has increased in the NBA. The geospatial plot is good to visualize the countries whose players have scored the most points; however, it does not necessarily convey any evident relation. A bar graph, in this case, although a more primitive method of visualization lets us conclude that the number of points scored per game is significantly higher for American players than international players. However, again to see if American players are significantly better than internationals, more statistics and factors need to be taken into consideration. Often, non-numeric values need to be assessed as well like player impact and role.

Figure 5: USA vs the World statistics based on points per game





From Figure 6 we can look at the statistics like average points per game, assists, rebounds and salary for LeBron James and Michael Jordan. Solely based on this, we can conclude that LeBron James has scored more points, assists, and rebounds over the duration of his career compared to Jordan. Although his average salary per season is not that higher on this scale, it should be noted that salary here is on a logarithmic scale, so actual values represent a much greater difference. This however could also be accounted for difference in eras played and inflation and more resources being directed towards athletes and sports. Here again, other factors like championship wins, other

personal accolades, team performances, other player statistics, era played, player impact and more need to be considered when assessing a player. But for now, we can say that LeBron James is the better player and the greatest of all time.

Figure 6 : Comparing LeBron James and Michael Jordan

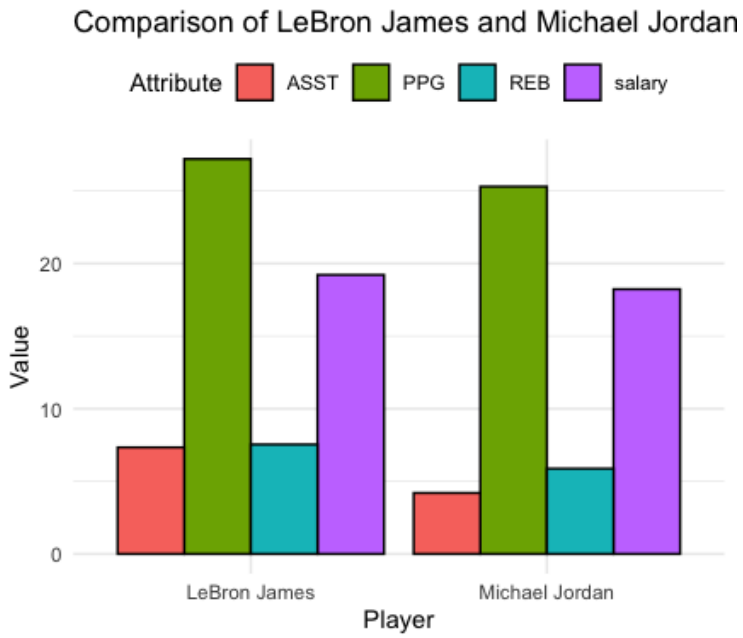


Figure 7 here shows the number of weekly awards given to individual players over the duration of their careers along with the number of points scored, assists, and rebounds averaged for their entire career. Their average salary on a logarithmic scale is also included. Solely based on these factors, we can arguably infer that LeBron James is the best player in the league. However similar to the conclusion above for Figure 6, there are other factors which need to be considered.

Figure 7 : Comparing the best player in the NBA based on weekly awards, ppg, assists, rebounds, and salary