

# Basketball Plots

Samantha Rodriguez

1/26/2022

## 1 Introduction

Basketball players have a lot of statistics about their career. Over the years the game of basketball has changed as well as the playing style. Looking deeper into players position and performance over the years we can see that change.

## 2 Data & Organization

To analyze basketball player data, I collected the information from a Kaggle data set<sup>1</sup>. The data set was organized in a SQLite format. SQLite holds different tables of information separately. There is no manual way of extracting the data into a usable R format. In order to change the SQLite file into separate CSV files of the tables it has, I needed to download an IDE of SQLite called SQLiteStudio<sup>2</sup>. The file had sixteen separate tables. I extracted nine CSV files.

After extracting, I can now import the files to clean and organize the different data sets. The data sets included information on the players, the teams, the draft, the draft combine, and the team and player salary. Looking into each data set closer, I decided to analyze three of the tables. These will be looking at the players, draft, and draft combine. After factoring some columns and changing the names of the columns to match, they were merged together. Beginning with the `draft` table, I dropped all the NAs from the rows. I will be using this table as the main data set to combine the other two tables. The data set being 7890 columns and 16 columns. Moving to the next table `draft_combine` I dropped columns that weren't necessary or had over a 90% NA percentage and then again dropping the NAs from all the columns. Then merging the two tables. The last table used, `player_attributes`, I only selected four of the twenty columns and merged it with the final data set.

Leaving the final table is a 217 rows and 30 columns. Some of the columns basis are to be connected with the other tables, such as the `idTeam` and `idPlayer`. These columns will not be used in the further analysis but it being kept in case further changes needed to be made and the tables need a similar column. Other columns such as the name of the player and the team the player played for are not included in the analysis but are kept for reference. **Table 1** includes all the columns that were kept. The next section, will go further into the analysis being made of the data that has remained.

## 3 Models & Analysis

Height is a large factor in basketball players, even in player positions that don't need to be tall. **Figure 1** shows the height of the basketball players that went through the combine from 2005-2020. The players are separated by what position they play. There are some players who play up to two positions and they are put into their own category.

---

<sup>1</sup><https://www.kaggle.com/wyattowalsh/basketball>

<sup>2</sup><https://sqlitestudio.pl/>

```
levels(player_info$yearDraft) <- c("1987", "2004", "2005", "2006", "2007", "2008",  
                                   "2009", "2010", "2011", "2012", "2013", "2015",  
                                   "2017", "2018", "2019", "2020")  
player_info$yearDraft <- as.Date(player_info$yearDraft, "%Y")
```

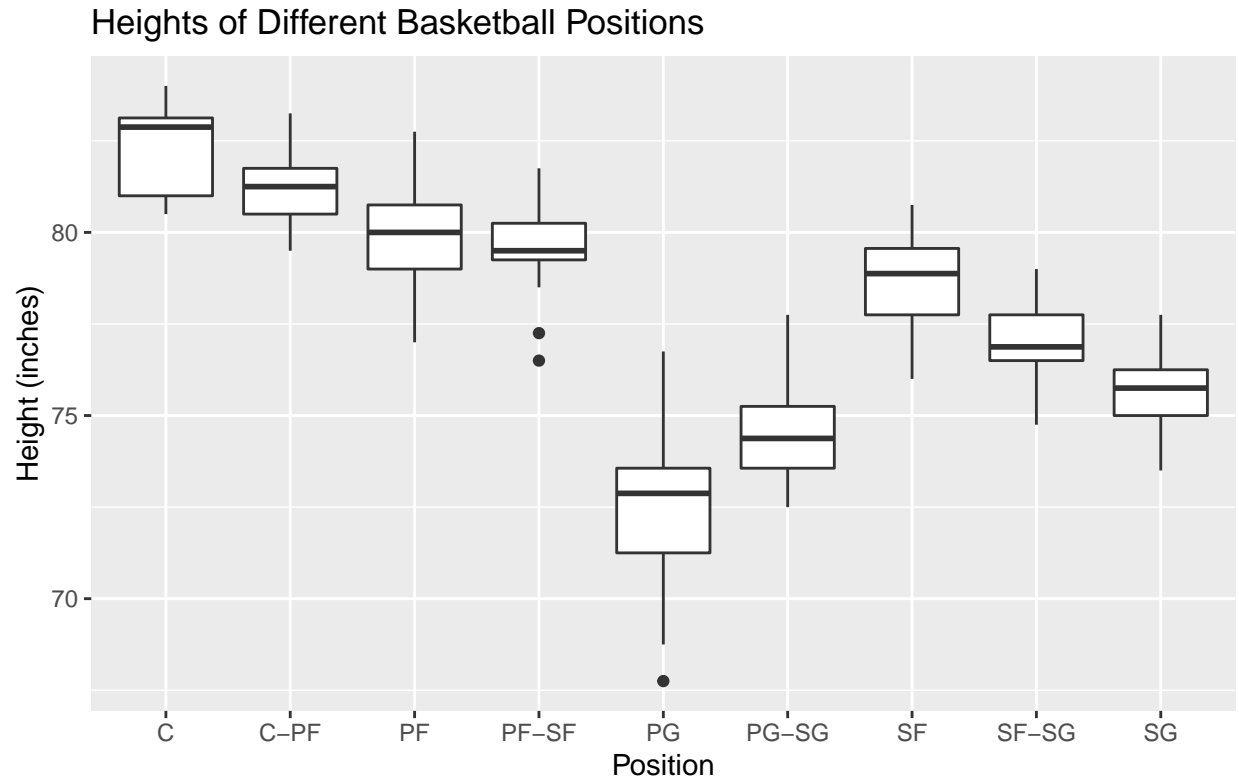


Figure 1. Height of basketball players of different positions over the years of 1947 and 2020. If a player played positions they are in a group together and not put in each separate group.

The box plots show that there is clear pattern based on the player height and what position they play. With center having the tallest height because they need to be the biggest players on the court since they're the closest to the basket. Also, point guards, being the shortest player, because they are doing most of the dribbling during the plays.

Now, that we have looked at the height of the players, let's look at the weight. **Figure 2** displays box plots of the weight of the players based on their position. Similar to **Figure 1**, players who play multiple positions are put in their own category. Based off these two graphs, we can see that height and weight are important for basketball players. The farther the player is positioned from the basket the shorter and less they weigh.

The strategy of basketball is each player bring their own set of skills based on their athletic ability and combine them as a team to beat the other team. The three main skills needed for basketball is rebounds, assist, and point scoring. The different positions favor each skill differently.

## Rebounds

Centers and power forwards need to be the players rebounding the most on the court. Whether that be offensively or defensively. As mentioned before, height is important to these two positions as well. **Figure 3** compares the height with their career average rebounds over the years the players were drafted. We're also looking at the height of the players. They are separated by whether they are above or below the median height of the positions selected. The graph also includes the players that play both center and power forward positions.

Ignoring when the height of the player, there is a decreasing trend on the average rebounds of center and power forwards. There was a steady trend on the average number of rebounds in the early 2000's but moving into the 2010's the trend is decreasing. It seems center and power forwards don't rebound as much as before. Now looking at the height, with red circle being above median height and blue triangle being below. Visually, there is no immediate trend. There are outliers for both sides, but they seem to be more red circles averaging a higher rebounds per game. This does make sense as taller players have a better chance of rebounding.

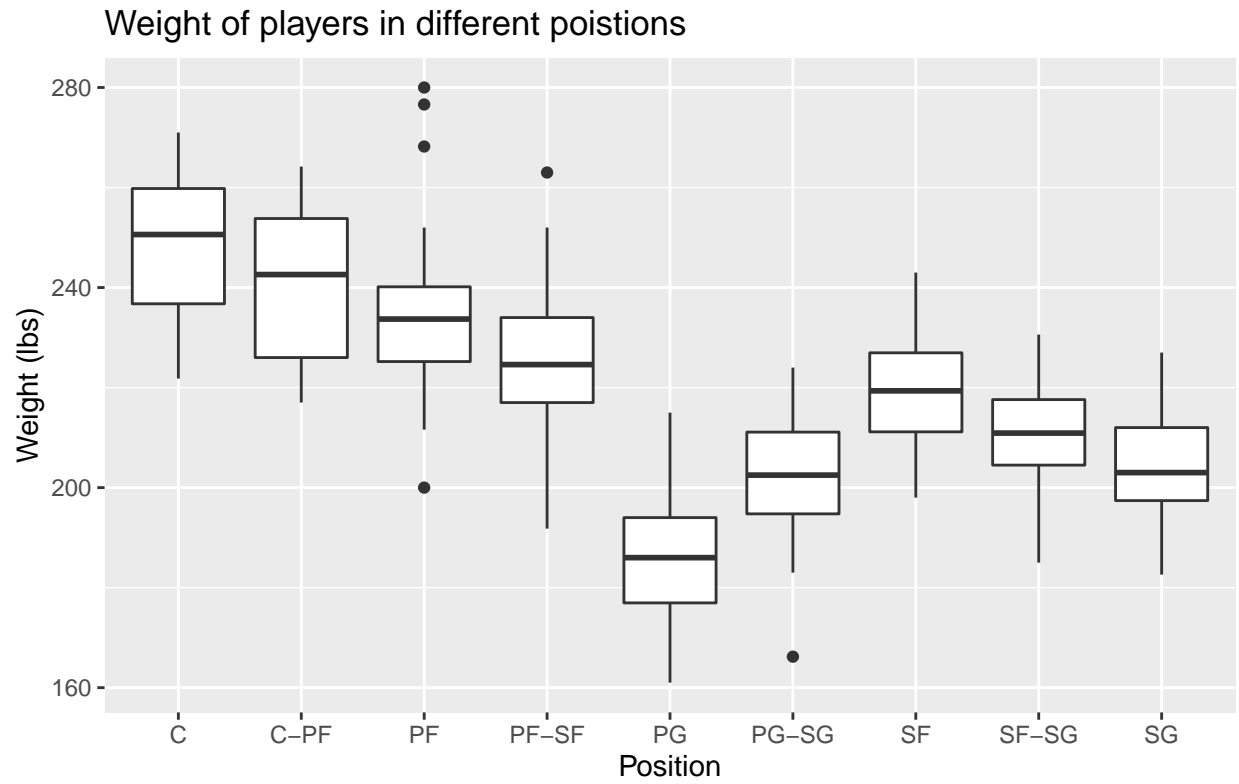
## Points

The next position we'll be looking at is shooting guards. These players are supposed to score the most of the team. This is the largest group of players since they are in the middle of the field. **Figure 4** includes shooting guards, point guards, and shooting forwards. Since, these are the scorers of the team we will be comparing the average points per game in the players career and what year they were drafted, as well as which round they were drafted in. There are two rounds in the NBA draft, the better players are drafted from the first round. Because of this we would expect them to perform better.

Similar to **Figure 3**, **Figure 4** shows a negative trend overall with a couple bumps in 2008 specifically. In the 2000's shooting guards had a high average of points per game. Now in the 2010's, the average has decreased. Looking at which round the players were drafted there is a clear pattern of the first round picks averaging more than the second round picks, as expected. All the shooting guards in the 2010's average what the second round picks did in the early 2000's. In the 2010's there is also less first round picks that are shooting guards than second round picks.

## Assists

The last skill is assists, which is given to the point guard position. For **Figure 5** we will be including the players who played point guard and shooting guard. The graph compares the players assists average of their career and the year they were drafted as well as what round they were drafted in.



*Figure 2. Weight of players in different poistions in the NBA from the year of 1947 to 2020. If a player played different poisitons they are in a seperate group and not put in two seperate groups at the same time.*

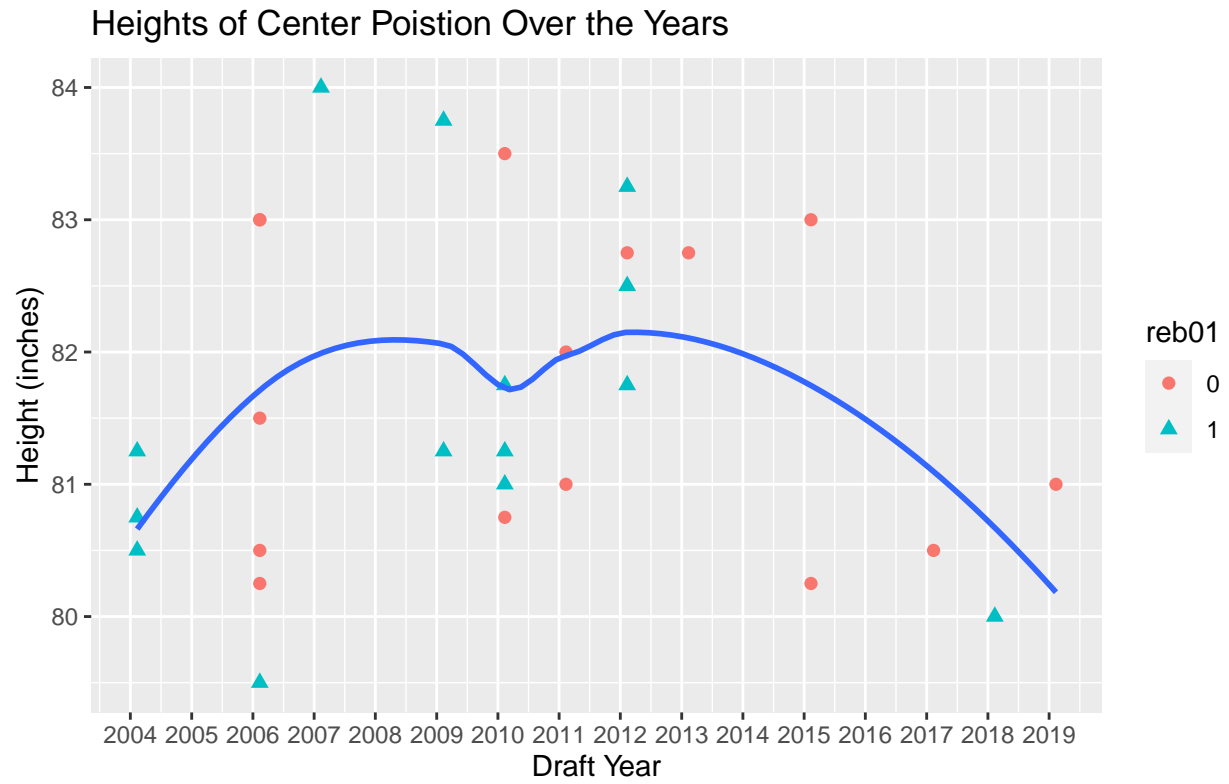


Figure 3. Comparing the center positions height as well as if they are above or below the median rebounds.

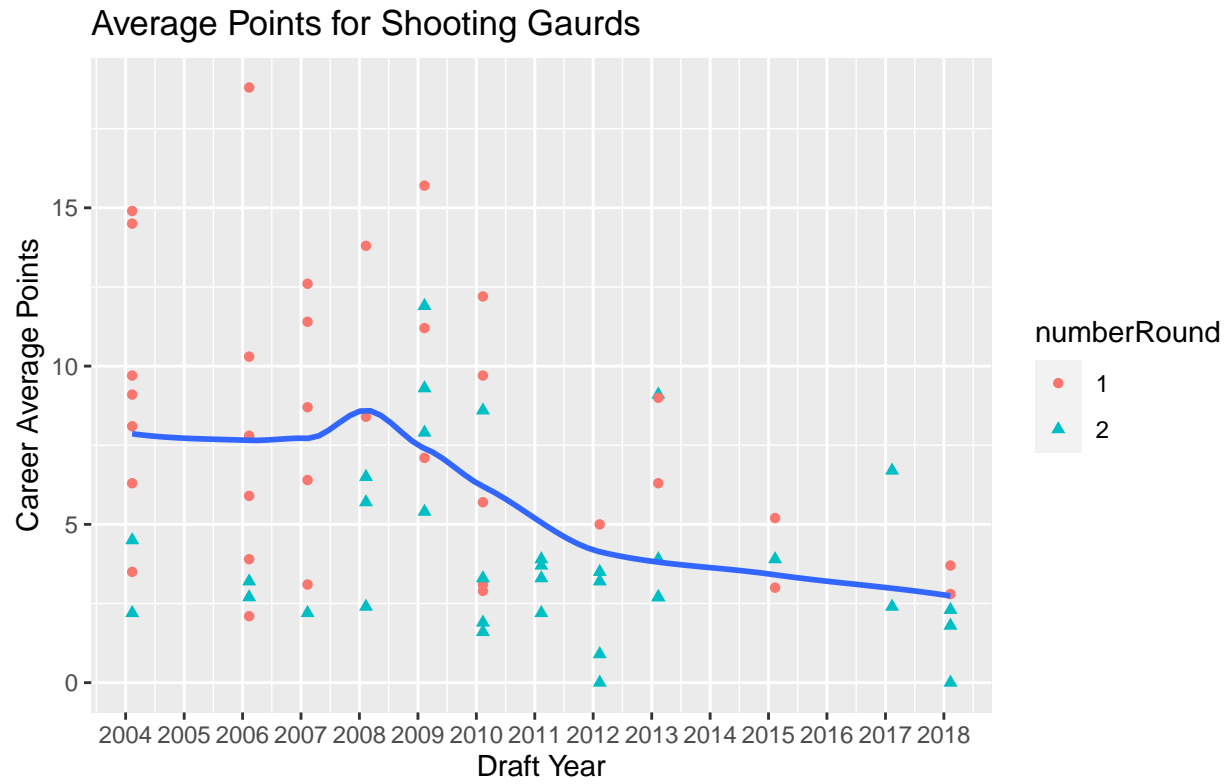


Figure 4. Comparing the average points of shooting guards also seeing in what round they player was drafted.

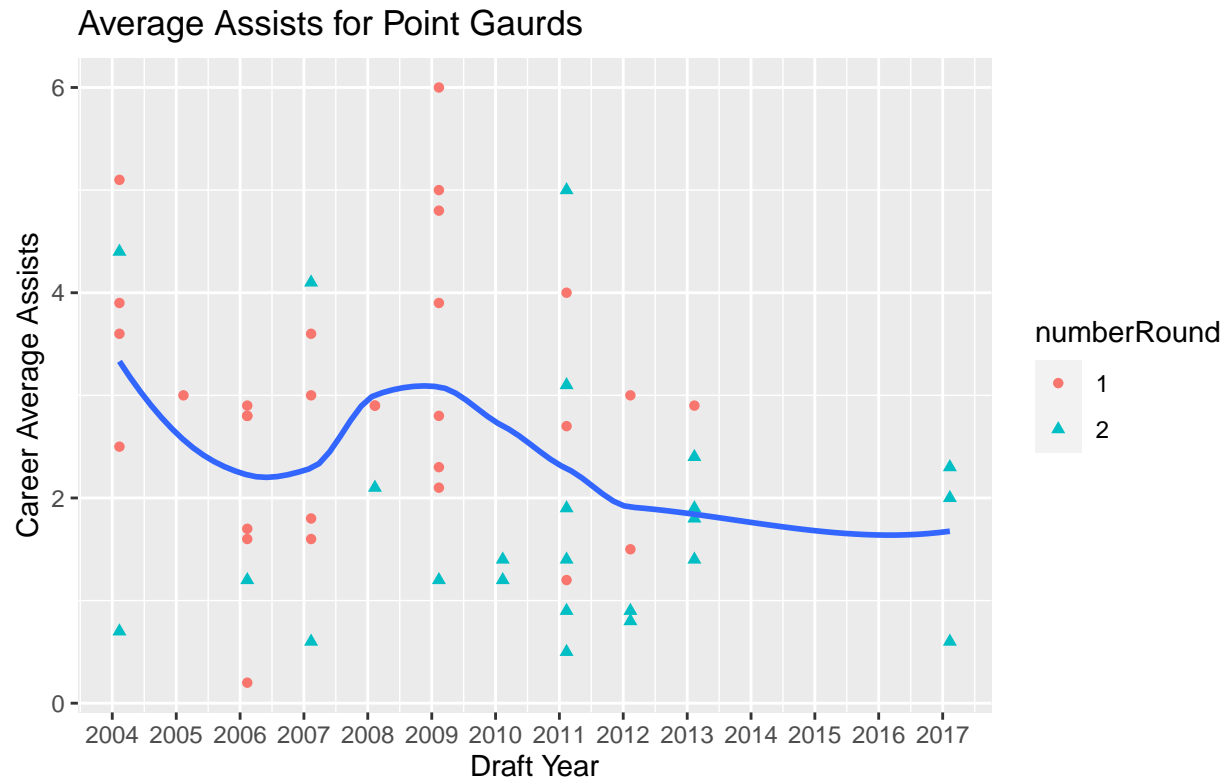


Figure 5. Comparing the average assists of point guards also seeing in what round they player was drafted.

Similar to the previous two graphs, there is a negative trend overall but a bump on 2008-2009. As before, round 1 players averaged more assists than players drafted in round 2. Another similarity, there is not many point guards and shooting guards drafted from round 1 in the 2010's as compared to the 2000's.

The last figure, **Figure 6** is the conclusion of these models. The graph is a stacked bar chart of the different positions and the draft year. Specifically, the number of players drafted for each year. This data set does not include all years which is why there are some blanks in the x-axis. For the years, that are included, the positions that require more height are drafted more than the other positions. In the 2010's the number of positions that have shorter players are drafted in a very small amount or almost none. As compared to the taller positions who stay relevant for the full timeline. We will discuss this further in the next section.

## Discussion of Major Results

The game of basketball changes over the years. From the models listed above it is clear that is true. There wasn't as much athletic ability in tall player, thus the need for shorter players who were faster and more agile. Looking at the figures, point guards and shooting guards are doing less and less for the game. There was a peak of both players in 2008 and 2009. There seems to be less of a demand for those positions. 2008 and 2009 includes the best players of those positions ever since. A couple of those players are Derrick Rose and Russell Westbrook, thought to be the best point guards of that time<sup>3</sup>. The game is changing and taller players can be faster and more agile. This correlates to the change in athletic training that has been made. Different training regimines has created players that have changed how the game is played and how teams carry out their strategy.

There has always been a difference in the performance of a round 1 and round 2 player. As one is supposed to be better than the other. In the 2000's the graph shows that the statement is true. Round 1 players performed better in their respective positions. Shooting guards scored more points and point guards had more assists. Moving into the 2010's that trend is not the same. Those positions were drafted more in round 2 than in round 1. Backing up the previous statement, that those positions aren't as important to the strategy anymore. Players have become more well rounded in skills and a specified skill set is not seen as helpful.

This can be seen in **Figure 6**, the positions that are mixed are drafted more than specialized positions. A player that can play a shooting forward and a shooting guard is better for the team than a shooting guard. Teams are looking for players that can do the three main skills needed for the game not just one.

While, not every year was included in the analysis, there was enough of a trend that the years wouldn't have affected the analysis of the graphs in the end.

---

<sup>3</sup><http://www.espn.com/nba/draft2008/index>



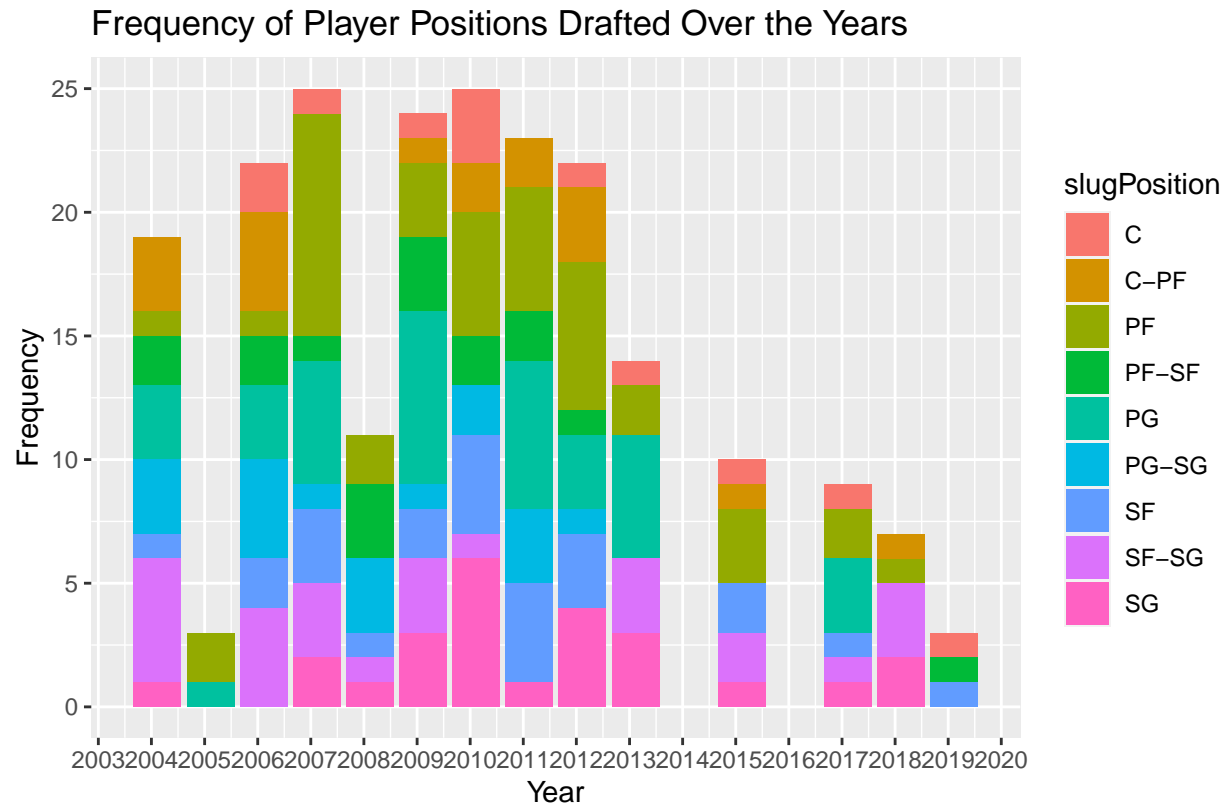


Figure 6. The frequency of how each poistion changes importance over the years.

## Conclusion

The game of basketball has a strategy to win just like any game. The strategy may change and the players and teams have to adjust accordingly. The results showed this change. Well rounded players have a higher chance of being drafted than single specialized players. With the advancement in sports medicine, players can become more well rounded.

## References

1. <https://www.kaggle.com/wyattowalsh/basketball>
2. <https://sqlitestudio.pl/>
3. <http://www.espn.com/nba/draft2008/index>