Optimization of Strategic NBA Team Drafting

Jean Park

Introduction

Our dataset contains over 20 years of data on every player that has been on an NBA team's roster. The data ranges from the 1996-1997 NBA season to the most recent 2020-2021 NBA season. The dataset contains information on a variety of features for each player. One feature is personal characteristics such as height, weight, and place of birth. Another feature is biographical characteristics such as team played for, draft year, and draft round. The final feature is player statistics including basic statistics shown in box scores, as well as more advanced statistics such as usage percentage and true shooting percentage. This dataset contained a good mix of quantitative and categorical variables which allowed us to explore some very interesting questions relating to the strategy of NBA team building.

Research Questions

Our group used the dataset to answer four main questions. Our first question was: What is the value of players being drafted in the three main groups (1st round, 2nd round, and Undrafted)? We used important statistics such as games played, usage percentage, points, and net rating to compare the value of players. These important statistics help to determine the value that a player brings to their team because they display a player's longevity and talent. Our second question was: How has the value of 1st round picks compared to the other groups changed over time? We were interested in investigating whether the other groups were becoming more or less valuable over time compared to 1st round picks. Our third question was: What is the impact of being a

2nd round pick or Undrafted player on playing time? We were interested to see if players drafted later on, or not at all, are given less chances to play compared to players who were drafted higher. Our fourth and final question was: How have player profile tendencies changed over time due to philosophy and player development strategy changes? We wanted to find if the types of players drafted in each group, based on personal characteristics such as height and weight, has changed over time as the NBA has shifted styles.

Data Cleaning

Our first step in our analysis was cleaning the data. For our analysis, we wanted to focus on the three main groups of drafted players discussed previously: 1st round picks, 2nd round picks, and Undrafted players. This was a point of concern because we wanted to be able to make assessments about the NBA today, which only has 2 rounds in the draft therefore including players who were not in this category could negatively impact our analysis. We found that the NBA draft switched to two rounds in 1989. Thus, the first step of data cleaning was only included NBA players who were drafted in 1989 or later. The second step of data cleaning was correcting an error with the Undrafted group. There were some players who had a draft round of 0, so we manually corrected this to fall into the Undrafted group. The final step of data cleaning was removing players who played less than 10 games in a season. These players are often players that did not contribute to their team's success in a meaningful way and are not the type of players we wanted to analyze in this project.

Graphics Discussion
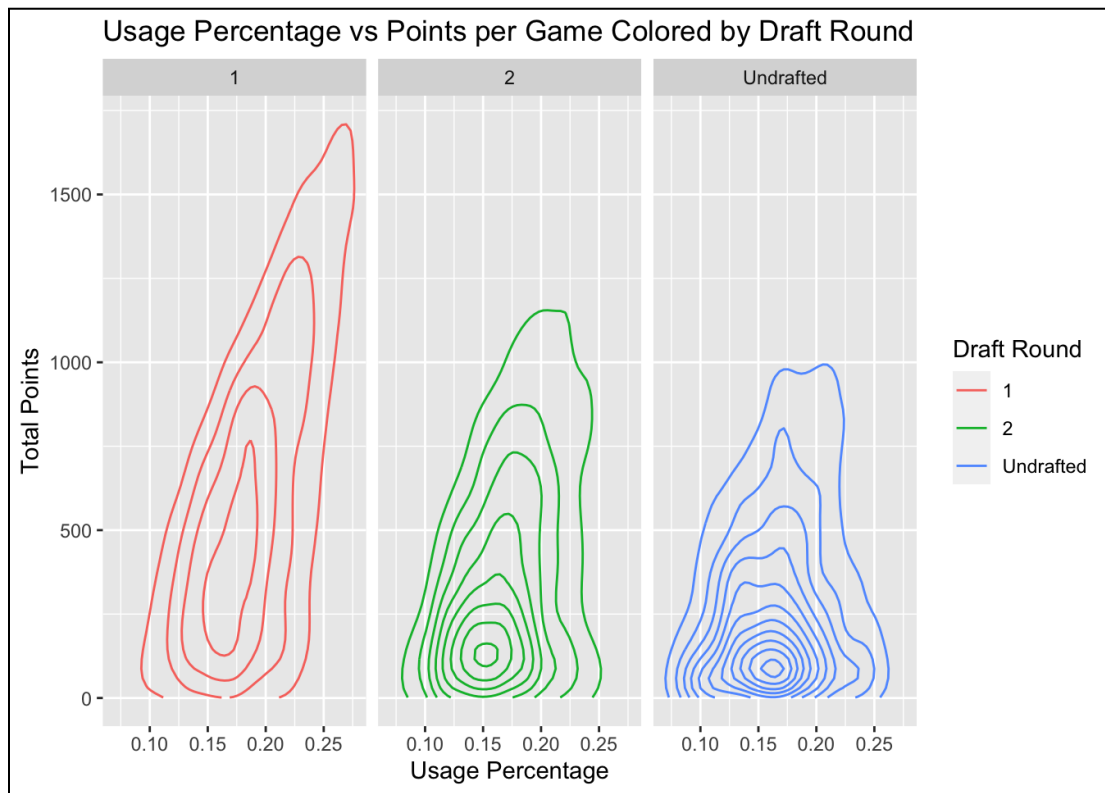
We wanted to learn about the value of players in each of the draft groups, which means we needed to examine the draft round of the players, their usage percentage, and their total points scored in a season. We needed to manually create a variable of the total points for each player by multiplying the average points per game times the number of games played. We found the total points each player scored in a season to weight the number of games played more evenly to avoid any bias. This graph (Figure 1) shows that 1st round picks have higher usage rates and points per game than the other two groups. While there are some 2nd round picks who have both high usage rate and points per game, the majority of 2nd round picks and Undrafted players are concentrated at the bottom center of the plot. In comparison, the performance of 1st round picks

are concentrated at the center of the x-axis and widely distributed across the y-axis. This suggests that 2nd round picks and Undrafted players generally have lower total points and less or similar usage rates to the less valuable 1st round picks. Furthermore, the spread in the 1st round picks to the upper right of the plot, shows that they score more points and these players tend to also have higher usage rates. The conclusion from this graph is that 1st round picks appear to be more valuable offensively than the other two groups of players by contributing more points and having a larger share of their offenses.
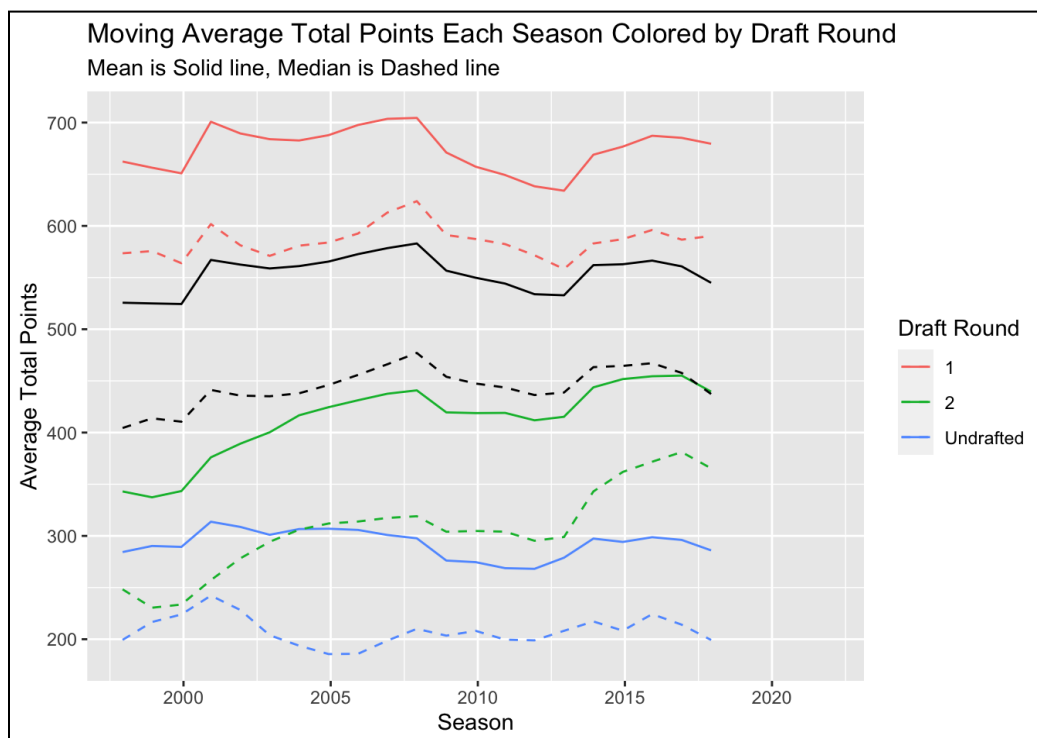
Figure 2.
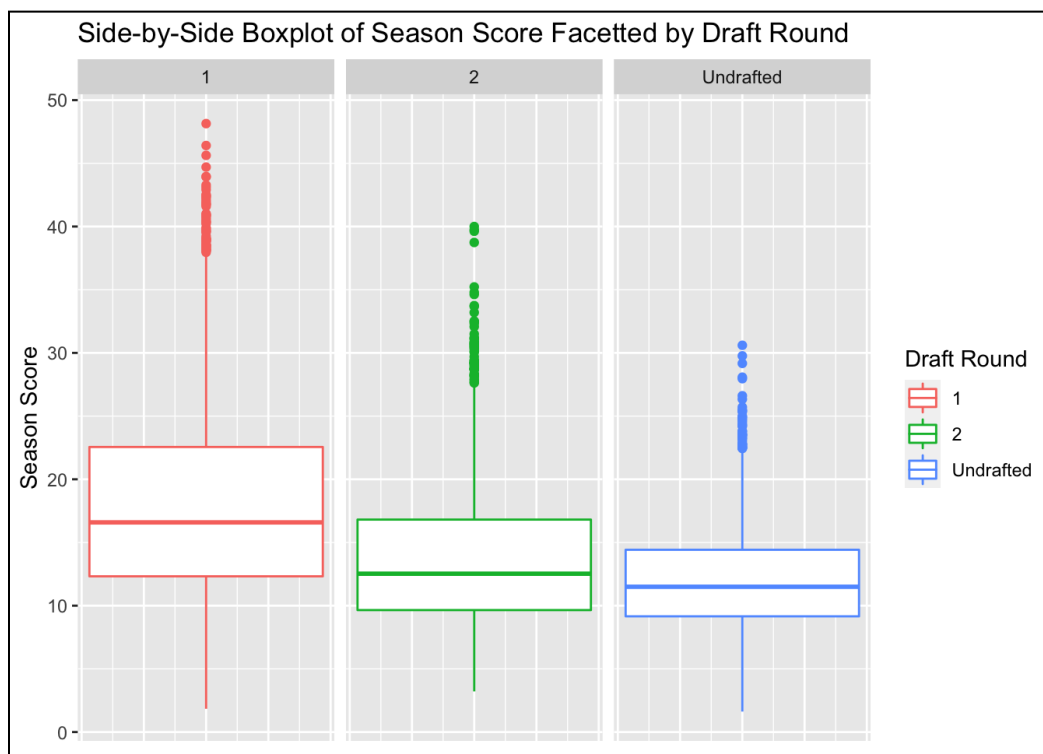
We also wanted to find a way to measure and compare the overall impact of players in each draft round. We first tried to use net rating, but found that net rating is heavily influenced by the team that a player is on since it is the overall point differential of a team while the player is playing.
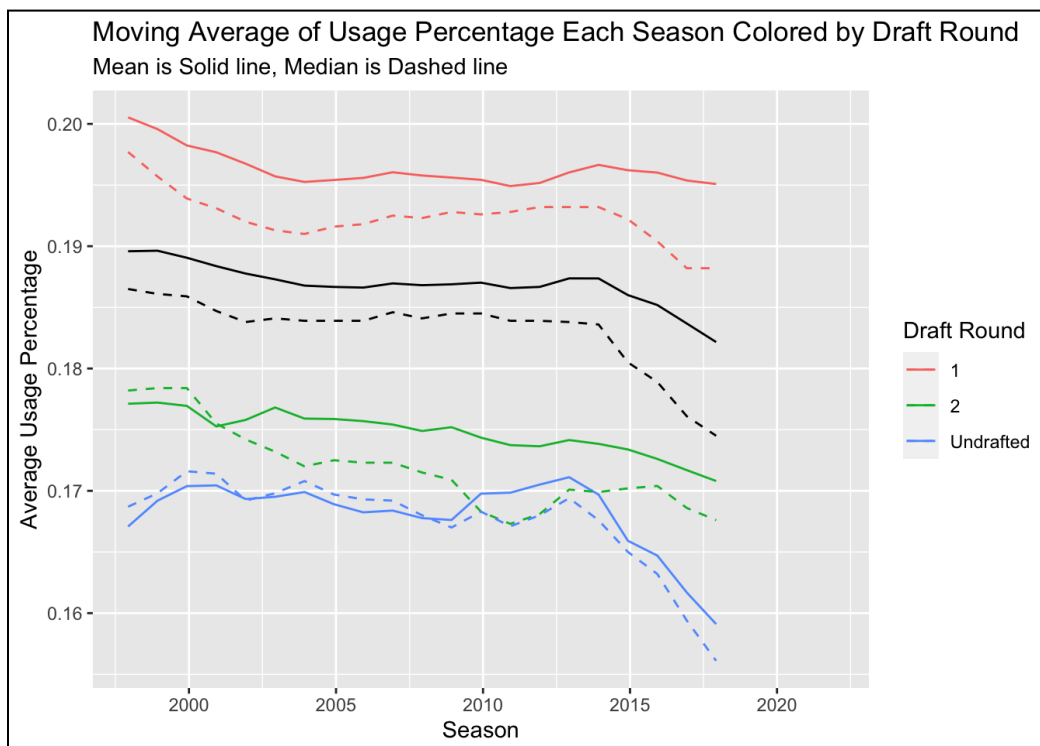
However, this means that a good player on a bad team could have a low net rating. Likewise, a bad player on a good team could have a high net rating. Then, we found a statistic called game score that looks to aggregate the overall impact of a player on a single game. In order to measure the impact of a player throughout the entire season, we created a column called season score which followed the game score formula, but used season averages rather than the single game statistics that game score uses. The season score formula that we used (based on the game score formula) is as follows: points per game + 7(offensive rebounding percentage) + 3(defensive rebounding percentage) + .7(assists per game) + 10(true shooting percentage). From the above graph (Figure 2), we can see that 1st round picks tended to have higher season scores than 2nd round picks and Undrafted players. We also see that 2nd round picks and Undrafted players tended to have similar median season scores, but the median season score of 1st round picks was much higher.

Figure 3.

We then wanted to investigate the changing value of players in each of the draft groups compared to each other, so we decided to examine the draft round of the players, and the average total points for each draft round over time. This graph (Figure 3) shows the average total points from each of the three draft groups over time along with the league average shown in black. It appears that 1st round picks score much more than 2nd round picks and Undrafted players. This separation grew around roughly 2013, where scoring increased across the NBA. The conclusion from this graph is that first round picks continue to be more valuable scoring assets than the other two groups of picks, contributing more total points on average. Also of note, the gap between the mean and median for these groups shows that the "superstars" that are found in each group score much more than the typical player drafted in that group. This difference is larger for 2nd round picks and Undrafted players than 1st round picks, indicating that these players are more "boom or bust".

Figure 4.

Moving Average of Usage Percentage Each Season Colored by Draft Round
Mean is Solid line, Median is Dashed line

Continuing with the previous analysis we conducted a similar analysis to examine new variables, keeping the draft round of the players, but shifting to the average usage percentage for each draft round over time. This graph (Figure 4) shows the average usage percentage from each of the three draft groups over time along with the league average shown in black. It appears that 1st round picks have higher usage percentages than 2nd round picks and Undrafted players. This separation has persisted over time and appears to be widening slightly. The conclusion from this graph is that 1st round picks continue to be involved in their team's scoring more than the other two groups of picks. Unlike the graph above, the gap between the mean and median for these groups is not as large. This means that "superstars" and average players drafted in these groups do not have dramatically different usage percentages, meaning teams are giving similar opportunities to both types of players.

Figure 5.
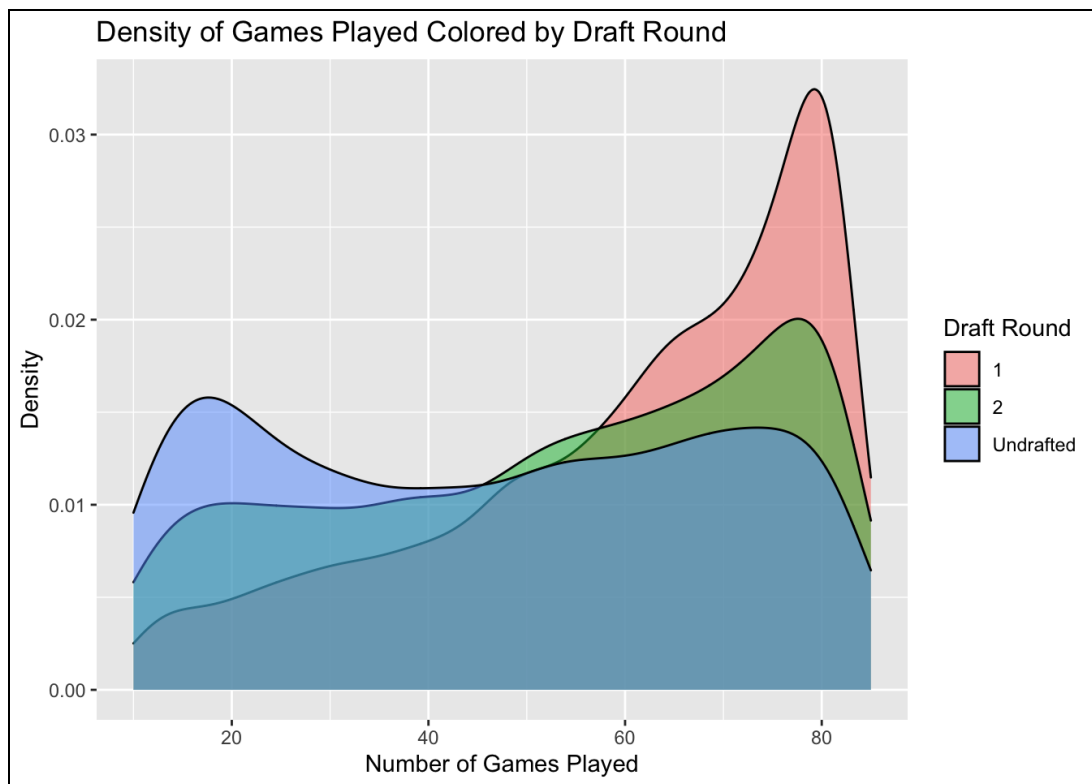
We wanted to investigate whether being a late draft pick affects a player's play time. Since we do not have a variable of the number of minutes to show play time, we instead use the number of games played as this measure. So, we decided to look at the number of games played and the draft round of the players. The graph (Figure 5) above shows a density plot of the number of games played by players in each draft round. The biggest takeaway is that 1st round picks appear to have a higher density of games played around a large number of games, meaning that they play more games than the other groups of picks. Another takeaway is that players in the other two groups have roughly even densities across all values of games played, meaning that players are just as likely to play a large number of games, as they are to play a very small number of games.

Figure 6.
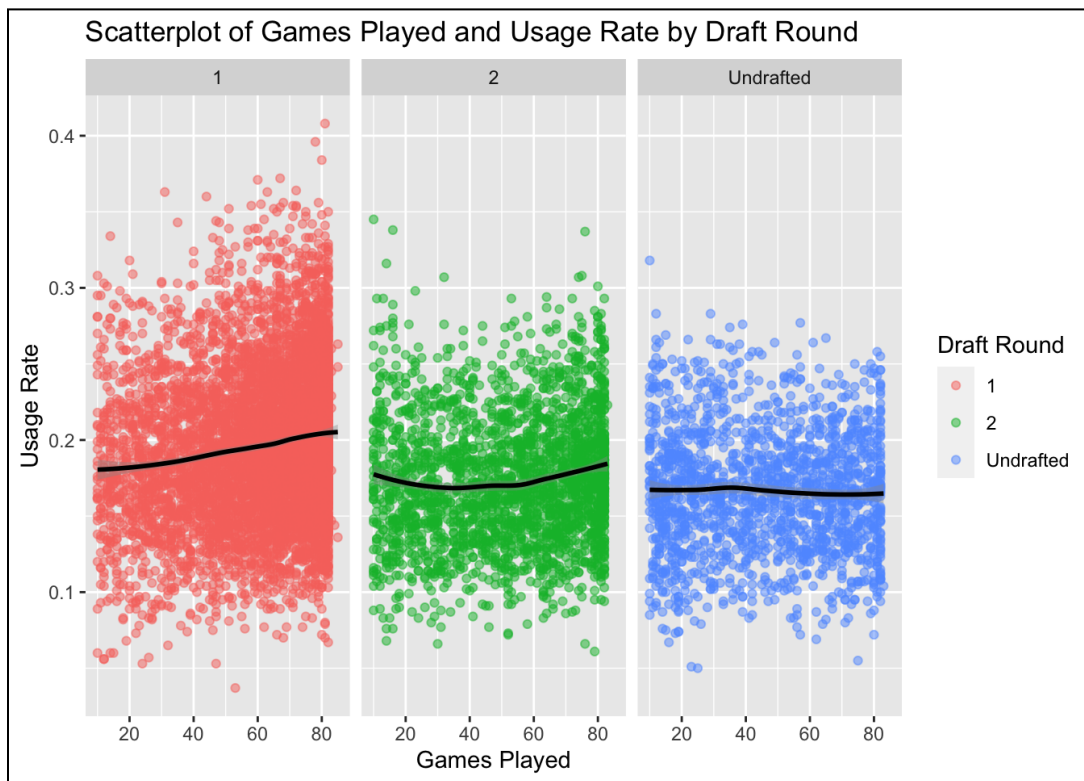
We wanted to continue investigating whether being a late draft pick affects a player's play time. We Turned to plotting Usage Rate against Games Played, to get a sense of how often players are used in the games that they play. So, we decided to look at the number of games played and the draft round of the players. This scatterplot (Figure 6) above shows that 1st round players have higher usage rates and games played compared to 2nd round and Undrafted players. While there are some 2nd round and Undrafted players who have played similar amounts of games as the 1st round players, a majority of the 1st round players have played more games than the rest. Additionally, we can see from the loess-smoothed regression that 1st round players' usage rate increases as they number of games played increases. This is interesting since both 2nd round and Undrafted players do not have an increasing trend with their usage rates. From this graph, we can conclude that 1st round players play more games and contribute more to the team than both 2nd round and Undrafted players.

Figure 7.
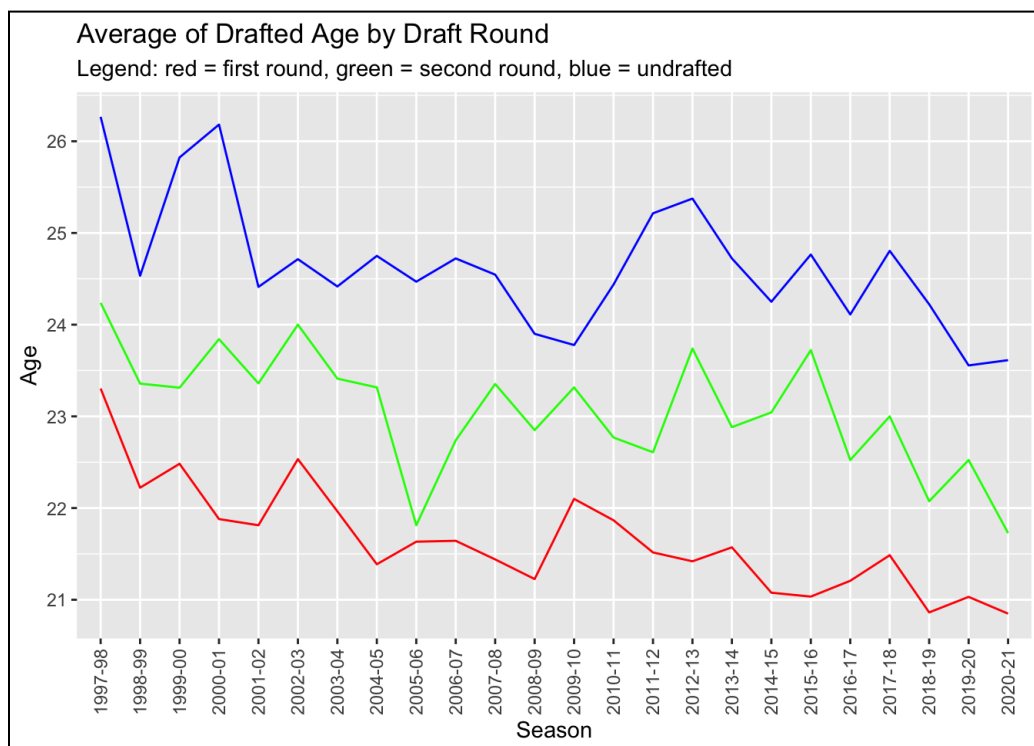
We wanted to examine how the measurements (age, weight, and height) of drafted players has changed over time and if there is any difference between 1st round picks, 2nd round picks, and Undrafted players. Since our dataset does not include age, height, or weight of a player when they were drafted, we created a subset of the dataset that includes the 1st time a player was seen. We did not include the 1st season (1996-97) in the dataset since our logic of a player's 1st appearance being their draft year would not hold. From the graphs above (Figure 7), we see that 1st round draft picks tend to be younger than 2nd round draft picks which tend to be younger than Undrafted players. This seems to follow the idea that younger players have more potential to improve and thus warrant being drafted higher. Another noticeable trend is that the average age for players across all draft categories have slowly decreased, with 1st round draft picks having the most consistent, gradual decrease.

Figure 8.

Welch Two Sample t-test

data: x_draft1 and y_draft1

t = 3.6033, df = 55.429, p-value = 0.0006729

alternative hypothesis: true difference in means is not equal to 0

95 percent confidence interval:
0.6098448 2.1376299

sample estimates:

mean of x mean of y
22.22222 20.84848

Welch Two Sample t-test

data: x_draft2 and y_draft2

t = 1.902, df = 15.251, p-value = 0.07623

alternative hypothesis: true difference in means is not equal to 0

95 percent confidence interval:
**-0.1940301 3.4537704**

sample estimates:

mean of x mean of y
23.35714 21.72727

To check if the decrease in average age we saw above was statistically significant for each draft round, we conducted 2 t-tests (Figure 8 and Figure 9). For the first t-test, we compared the average age for players who were 1st round draft picks in season 1998-99 against players who were 1st round draft picks in season 2020-21. We replicated this t-test for players who were 2nd round picks. Since the p-value for our first t-test was less than 0.05, we reject the null hypothesis and conclude that the difference in average age for 1st round picks drafted in 1998 vs. those drafted in 2020 are statistically significant. Since the p-values for our other two t-tests were greater than 0.05, we fail to reject the null hypothesis and conclude that the difference between average ages for 2nd round picks and Undrafted players who played in 1998-99 vs. those who played in 2020-21 are not statistically significant. However, we would like to note that the p-values (0.076 and 0.076) were only slightly greater than 0.05.

Figure 10.
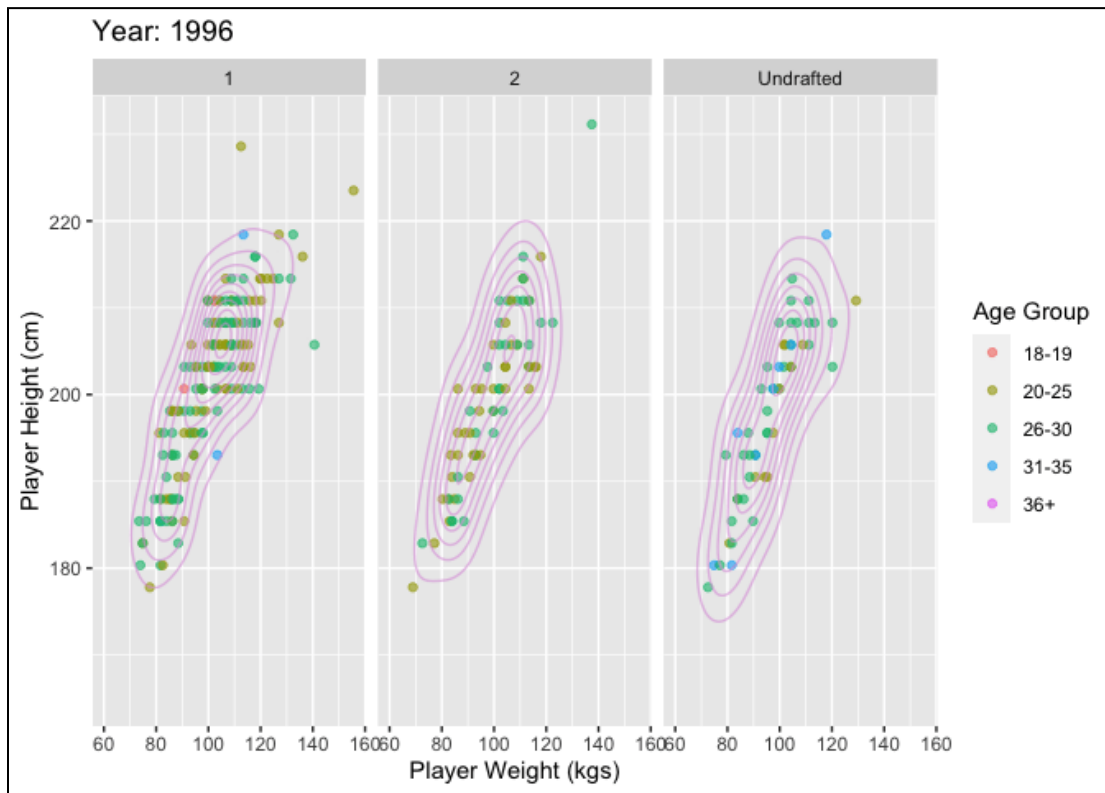
Since there was no information about the players' weights and heights when they were drafted, we decided to look at the weight and height of players for each NBA season. We also wanted to look at players' age each season which could give us additional information unavailable in the last graph. In order to visualize three quantitative variables, we made a scatterplot with weight on the x-axis, height on the y-axis, and age as the data point colors. Since we wanted to make sure age was interpretable, we decided to turn the quantitative variable into a categorical variable by separating the ages into 5 age groups. On top of the scatterplot (Figure 10), we overlaid contour lines in order to make shifts in density more obvious. Then, to add in a time element, we created an animation that transitioned through distinct states of the scatterplot in time. Finally, we faceted by draft round to stay consistent with our previous analysis and show differences in preferences for each draft round. There are many takeaways we can gather from the

time-dependent graph above: 1) During early years, the spread of 1st round draft picks was larger than the spread of 2nd draft picks. However, as time passed, the difference in the two spreads seemed to decrease 2) Despite the increasing spread of 1st round draft picks, the density consistently stays the highest around the area where height equals 210 cm and weight equals 100 kg 2) 2nd round draft picks are more diverse in later years 3) 1st round draft picks seem to have more points in the far top right than 2nd round draft picks 4) The Undrafted data had the most points in the lower bottom left portion of the graph 5) The number of players over the age of 36 appears more frequently in later years. All in all, we conclude that the spread of height and weight preferences for 1st round draft picks slightly increased while preferences for 2nd round draft picks and Undrafted players more drastically increased. Player profiles across all three profile dimensions become more diverse over time.

Conclusion

We have shown that 1st round picks are the most valuable of any draft group. In particular, their offensive ability makes them very valuable players for NBA teams. This value has persisted over time, and the difference in value has even increased in recent years. Another form of value is the availability of a player. To this point, 1st round picks appear to play more games meaning that they tend to be more valuable in the fact that they play more games than the other groups of picks. Finally, over time it appears that in general, players that are drafted are weighing less and shorter on average. What stands out is that 2nd round picks appear to be much more volatile in the types of players that are drafted, meaning a more "standard" NBA player can be found in the 1st round, while the type of player available in the 2nd round could be quite random.