# Hawks Intern Project

Ben Thorpe

## Part 1

**A)**

a. The three teams who drafted the most players from Duke between 1989 and 2000 (inclusive) were the Dallas Mavericks, Minnesota Timberwolves, and Phoenix Suns.

b. The team who drafted the most players whose first name starts with the letter D and were drafted in an even year were the OKC Thunder. (Note: if OKC and Seattle were not combined then the correct answer would have been a tie between the Boston Celtics, Milwaukee Bucks, and Seattle Supersonics).

**B)**
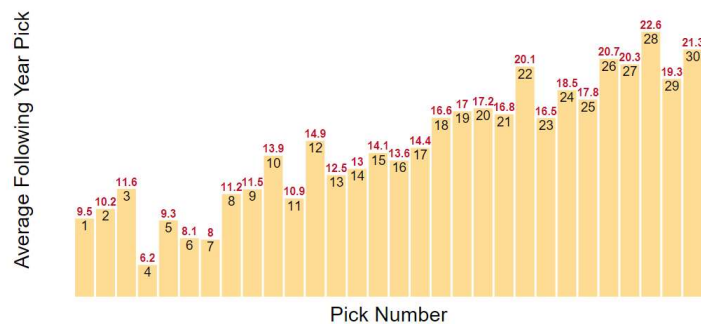


As we can see from the plot above, there is a relatively strong positive linear relationship between the current year and following year first round pick number for a certain team. However, teams drafting in the top ten are expected to improve the following year, on average, and end up with a lower pick in the subsequent draft, and the opposite looks to be true for most teams that are outside of the lottery, also on average. One other interesting finding is that teams picking 4-7 end up with next year draft picks that are higher, on average, than the top three picks obtain. It is possible that the bottom seven teams are often similar in how good they are, and the difference between the value of the top three picks and the next four are substantial enough to affect the average next year draft position.

## Part 2
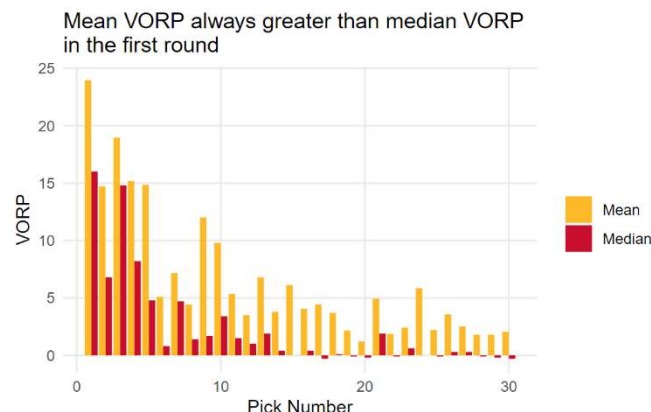
The metric I will be using to evaluate each draft slot will be with the median VORP value of that specific slot among all values in the dataset. I made all NAs in the dataset (for players who never played in the league) -2, since this is the replacement level VORP value, and these drafted players who were unable to get minutes likely would have been replacement level, on average. I chose VORP because it was one of the few metrics provided which attempt to model a player's overall value from box score stats alone and has advantages over the other two similar metrics. I chose it over win shares since it is based on BPM, which performed better than win shares in a player metric comparison from dunksandthrees.com, and over BPM itself to account for career longevity since BPM is an efficiency stat while VORP is a counting stat. Once I settled on VORP, I wanted to choose between using the median or mean value across drafts for my analysis. From the above graph, we can see that the mean VORP for a given draft slot in the first round (only plotted first round for graph clarity) will always be greater than the corresponding median VORP, which makes sense as most players who are net negatives will not be played nearly as much as those who are positives to their team, and thus their VORP will not build up as much over time. This is also shown by the average career VORP of players in this dataset being 3.56, which is greater than zero.
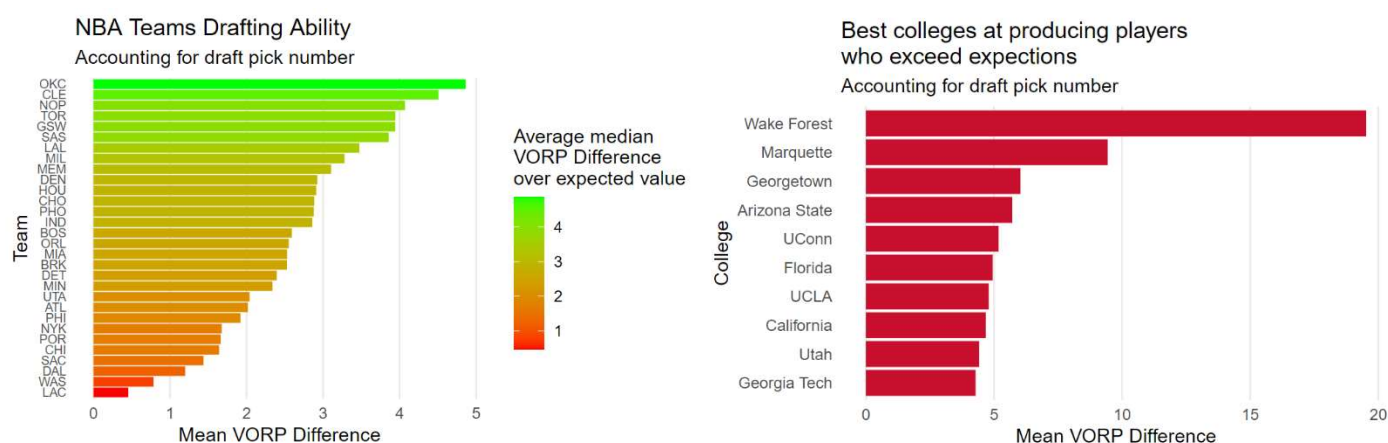


The above graph uses the average difference in median VORP to evaluate the relative draft success of each NBA team. This value was calculated by evaluating the difference between the VORP of a draft pick and then taking the average of these values for each team, respectively. From the graph, we can see that the three best teams at drafting according to this metric are OKC, Cleveland, and New Orleans, and the three worst teams are the Clipper, Wizards, and Mavericks.

I used a very similar process evaluating the NBA teams as I did in seeing which college teams had the most over performance in their draft picks. Here Wake Forest stands out as the team having the most over performance, mainly due to producing two legends in Chris Paul and Tim Duncan.

One area of my project that could be improved with future work is to do two different sets of analysis for lottery and non-lottery picks. I think that being able to draft well at the top end of the draft may be fairly different than the ability to find and develop gems later on, and it would be interesting to see if this is the case based on the data available in this study. Also, if as a part of further research I was allowed to use additional data sources, I would replicate the study using a metric such as Estimated Plus Minus since it incorporates player tracking data (since it has been available) and can be a more accurate measure of overall player success. Lastly, I would try to dive more into how I could include measures to reduce outlier influence if continuing with median VORP as the evaluation variable, by potentially limiting all VORP to an arbitrary maximum number to try and get a better sense of which teams overperform the most in the draft by limiting the effect generational players have on the results.