Analysis of NBA Draft Candidates

CS 170A: Final Project

Andrej Ivanovic

604016344

Table of Contents

Introduction

Background:

Basketball is one of the most popular sports worldwide. The following table summarizes the distribution of basketball players across various age groups in the United States.

Table 1. Number of Students Playing Basketball[1]

Level ↓ Players →	Number of Players
High School Basketball Players	545,844
NCAA Basketball Players	17,500
NCAA Athletes Drafted into NBA	48

We narrow our sample size to just high school basketball players. Referring to table 1, about half a million students play basketball on that level. Out of these players, only 17,500 go on to play college basketball in the NCAA. This leads to the following analysis: only 3.2% of high school basketball players have a chance to play at the college level. Lastly, only 48 NCAA players get drafted each year. If we only take the high school seniors into account, it leaves us with 155,955 players. Only 48 of these students will make the NBA, meaning there is a 0.031% chance of a high school senior eventually ending up playing professional basketball at the highest level in the world. However, basketball is such a popular sport that students are extremely interested in making themselves one of these lucky 48 players each year that get a chance to compete in the NBA. It leads us to a very important question: which physical characteristics lead to getting drafted into the NBA?

Dataset:

To analyze the physical characteristics of players hoping to make the NBA, it was crucial to use information gathered from a collection of NBA draft camp measurements. The website *draftexpress.com* had all the necessary information. There was information for basic categories of 1348 players. However, for analysis involving body fat and agility/sprint tests, the data set was trimmed by 300 players due to missing data. The measurements recorded by the camp were recorded using the following techniques:

<u>Height and Weight</u>

Heights of the players were measured in two ways. They were measured barefoot, with just socks on, as well as with shoes on. It was not necessary to include both measurements, as the only information that we gather from the two heights is the various thicknesses of shoes players wear. As a result I only used the height without shoes in my analysis. I used the unaltered weight reports.

Wingspan and Standing Reach

To measure wingspan, the player stood facing a wall, spread out their arms horizontally as far as possible, and reach was measured from the left fingertip to the right fingertip. To measure standing reach, the player was still facing the wall, and simply put up their hand. The highest point their fingers reached was recorded as the standing reach.

Hand Size Measurements

While it would have been interesting to observe any relationships between hand size and entering the NBA, the data size for this measurement was too small (under 300 samples).

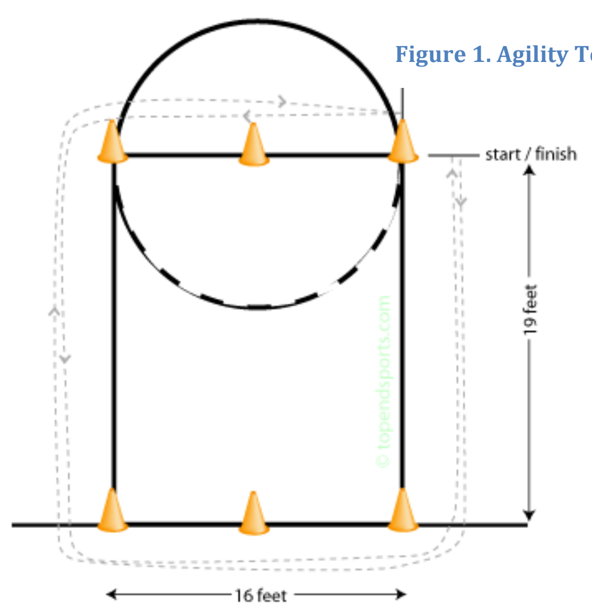
Vertical Leap Measurements

Seeing as how basketball is largely a jumping sport, vertical leap seems to be a key statistic that needs to be analyzed. Thus, I used the vertical standing leap, which involved how high the player could touch without a run up, as well as running leap, which allowed the player several steps of momentum. The standing and running vertical leap *reaches* were also measured, which was how high the player touched with either leap. The vertical leaps in both cases were measured simply by subtracting the difference between the highest points touched, and the player's previously measured standing reach.

Athletic Measurements

The player was also measured in other athletic tests. Players had their body fat percentages recorded, as well as agility and sprint speed. To test sprint speed, the player was timed running ¾ths of a basketball court. In order to test agility, a player was timed completing an agility course. Referring to figure 1, the player started at the start cone, sprinted forward to the baseline 19 feet, side shuffled 16 feet along the baseline, backpedaled 19 feet, and side shuffled once more back to the starting cone. The course was ran one more time but in reverse. Table 2 below lists the expected times for

Figure 1. Agility Test Set-Up [2]



professional basketball players in the agility test.

Table 2. Expected Agility Times [2]

position	Males (sec)	Females (sec)
guards	10.2 - 10.9	13.0 - 14.5
forwards	11.0 - 11.4	14.6 - 15.5
centers	11.5 - 12.3	14.6 - 15.5

The last category included in the data set was whether the player was drafted or not, and if they were drafted, what position they got.

Obtaining the Dataset

Obtaining the dataset involved writing a script to extract each row with its information. The data, however, was poorly formatted for analysis because all measurements were done in feet-inches in the following format, X ’ Y ”, which is difficult to analyze. So I wrote a parsing script that took as input a single column of length data, extracted the second element, Y, of each measurement, divided it by 12 to format the inches into feet, and added it to the first element X, to obtain a single number, a measurement in just feet.
*C++ script is in Appendix on page 42

Summary Log of Analysis

*Full Log of Code Included in the Appendix

It would be most obvious since basketball is a game of height and jumping that those two factors would be the most dominant in helping a player at a professional level. Before examining the data further, it is important to check that the dataset is valid, and that the distributions of physical abilities are normal, and not skewed in some way. To accomplish this, I wanted to check that the heights, wingspans, and both forms of vertical leap were normally distributed. Figure 2 exhibits the shape of a normally distributed data set. Figures 3, 4, 5, and 6 show the distributions of height, wingspans, and both forms of vertical leap, respectively.

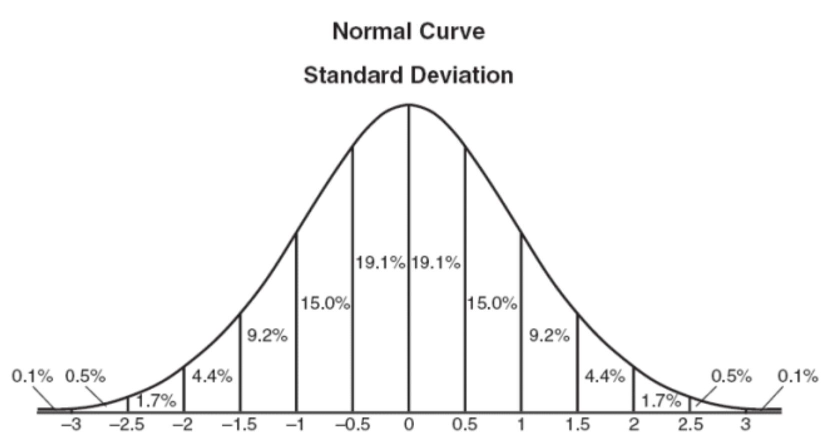


Figure 2. Normal Distribution

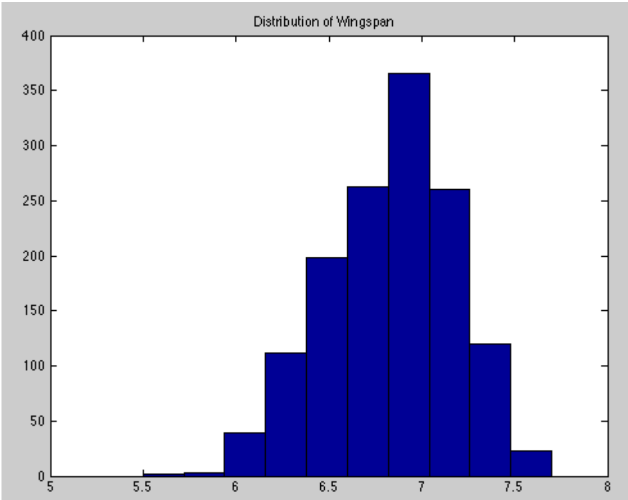


Figure 3. Distribution of Wingspan

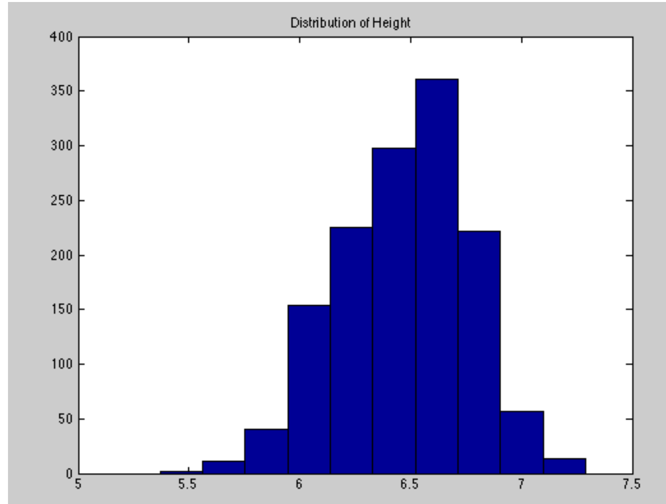


Figure 4. Distribution of Heights

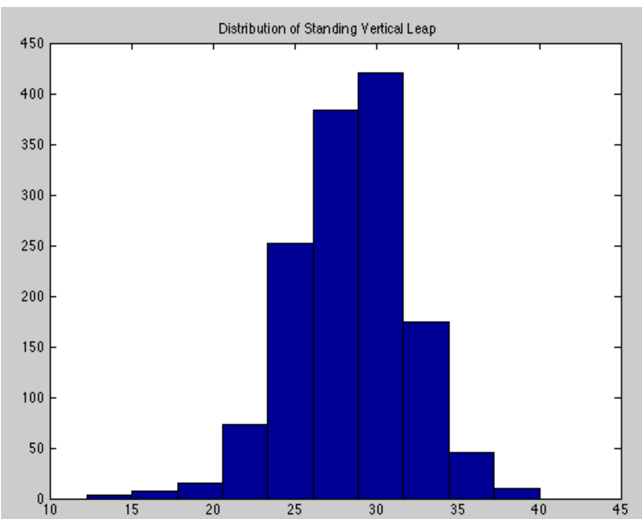


Figure 5. Distribution of Standing Leap

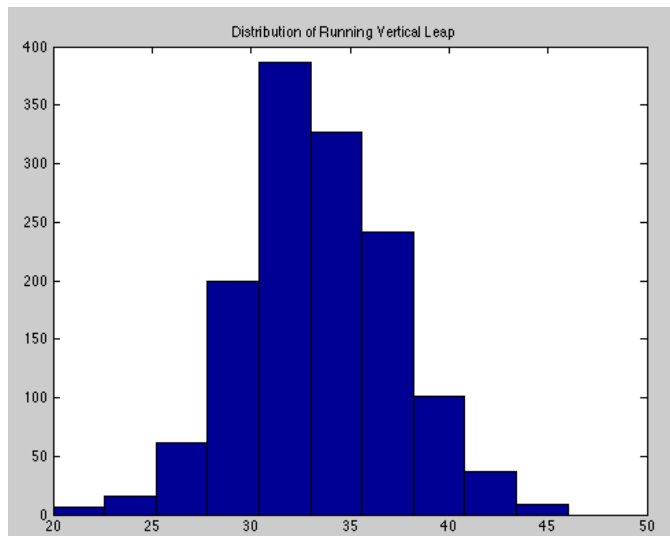


Figure 6. Distribution of Running Leap

* refer to Appendix page 25 for code

By observing the 4 charts, it can be concluded that the data is valid, as it seems to have the skeleton of a normal distribution. Therefore, analysis made from this data set can be considered accurate.

NBA scouts spend the entire NCAA season observing the performance statistics of players across NCAA Division 1 and 2. It is not as common for a Division 2 player to be observed, but it does occur. Unfortunately our dataset did not include information on the players' collegiate level, so this distinction, and its effects, could not be observed. Thus, out of all the NCAA players, only a few even get invited to the NBA draft combine. This tests their physical abilities, putting them in the running to get drafted. As a result, the first important observation was to see what chances a player that got invited to the draft combine has to get into the NBA. Keeping in mind that those under consideration are mostly NCAA players, with a small percentage of international players, the following pie chart shows the percentage that actually got drafted.

*refer to Appendix page 26 for code

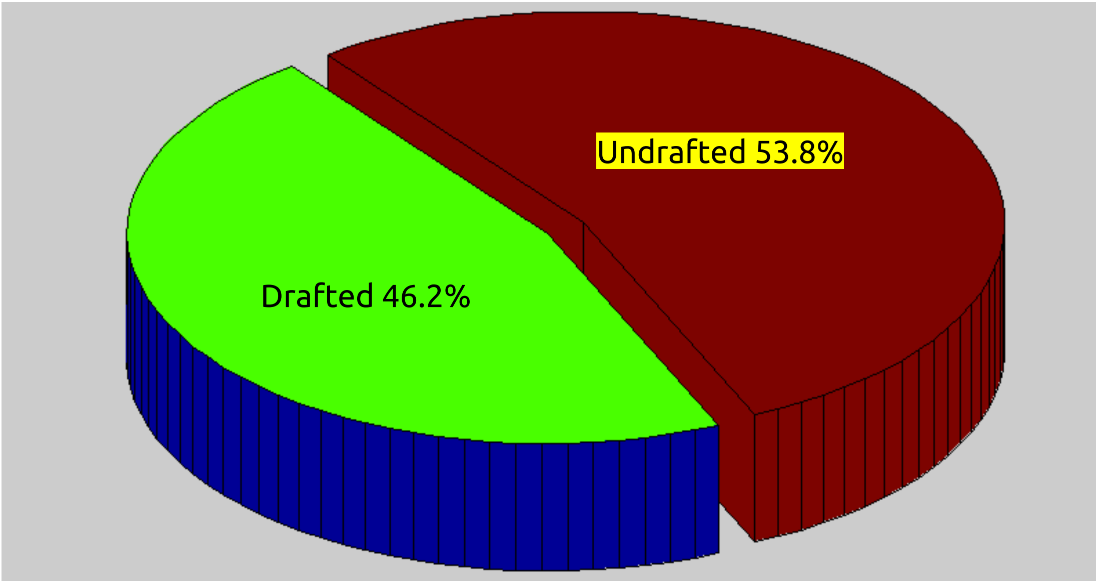


Figure 7. Chart Showing Chances of Getting Drafted At Combine

Value of the Correlation Co-Efficient	Strength of the Correlation
1	Perfect
0.8 - 0.9	Very Strong
0.5 - 0.8	Strong
0.3 - 0.5	Moderate
0.1 - 0.3	Modest
> 0.1	Weak
0	Zero

Table 3. Standard Correlation Results[3]

Thus, even if a player gets invited to the combine, he still has only a 46.2% chance of making it to NBA. The next step was to examine the correlation between getting drafted, and the player's height. The players' draft values were changed to a 1 if they player got drafted, and a 0 if he didn't make it. The correlation between height and the player's draft status was measured at just 0.0225. Table 3 is the standardized result interpretation of correlation values. From this definition we can conclude that height and a player getting drafted have a weak correlation.

It would be of value to observe, however, the characteristics of those who were lucky enough to get drafted and those that were not. Table 4, below, compares key statistics of the two sets of players.

*refer to Appendix page 27 for code

Variable ↓ Status →	Mean of Drafted	Mean of Undrafted
Height (feet)	6.48	6.47
Standing Reach (feet)	8.63	8.60
Wingspan (feet)	6.87	6.81
Running Vertical Leap (inches)	34.33	32.75

Table 4. Mean Physical Values for Both Categories of Players

There was a small but distinct difference in every category, but a very noticeable difference in the vertical leap category. The drafted players jumped 1.58 inches higher on average then their undrafted counterparts. This prompted interest in the athletic abilities on drafted and undrafted players. The mean of measurements for athletic ability was taken for both classes of players, and the results are summarized in Table 5.

Variable ↓ Status →	Mean of Drafted	Mean of Undrafted
Bench Press (reps of 185 lb)	11.04	11.55
Agility (seconds)	11.40	11.59
Sprint (seconds)	3.28	3.30
Body Fat (%)	7.46	8.46

Table 5. Mean Athletic Values for Both Categories of Players

With these values in mind, it is still not clear that there is a distinction between those players that got drafted, and those did not. It is also apparent, however, that players that got drafted had a full 1 percent lower body fat percentage than those undrafted. Agility measurements were also 0.19 seconds faster, which is a significant difference. The rest of the values show no strong distinction, with undrafted players even getting almost a full rep more than drafted players. Thus it is worth looking at the combination of athletic results, to get a normalized measure of a player's athleticism and see this athletic score's correlation with draft status.

The 4 athletic measurements that need to be considered are running vertical leap, agility, sprint speed, and bench press strength. However, certain things must be taken into account in order to make the scores fairly measured. For example, we must derive formulas to adjust the vertical leap measurements so that small, light players aren't favored by having higher vertical leaps over bigger, and heavier players that still jump high in relation to their weight. The next set of derivations will give a formula for adjusted athletic values.

Derivations for the Athletic Formula

*refer to Appendix page 29 for all athletic formula code

Adjustment for Vertical Leap

To adjust for vertical leap, an accurate measurement would be to take the vertical leap, and divide it by the player's weight. This will simply give the amount of inches a player jumped per pound of body weight, which is also indirectly a measure of power.

Adjusted Vertical = $\frac{Vertical\ Leap}{Body\ Weight} = \frac{inches}{lbs}$

Adjustment for Sprint Speed

To adjust for sprint speed, it is necessary to account for differences in height between the various players. A short player is quicker than someone who is a few inches taller with a longer stride length, even if they have the same sprint times. The shorter player took more steps in the same amount of time, thus has quicker feet. To account for this, sprinting stride length must be taken into account. Various studies have been conducted[4] on this topic and 3 results have been obtained. It has been found that athlete height multiplied by coefficients of 1.14, 1.17, or 1.35 gives the athlete's sprinting stride length. The discrepancy in the measurements comes from differing running surfaces used in testing. For our purposes we will use an average of the 3 measurements, which comes out to 1.22. We will also use the length of the sprint test, which is 3/4ths of an NBA court, coming out to 75 feet.

$$Step\ Size = 1.22 \times Height$$

$$\frac{75\ feet}{Step\ Size\ (\frac{feet}{step})} = Steps\ Taken$$

$$\frac{Steps\ Taken}{Measured\ Time} = Foot\ Speed\ (\frac{steps}{second})$$

Adjusted Sprint Speed = $Foot\ Speed$

Adjustment for Bench Press

To adjust for bench press we simply must take into account the player's body weight. In addition we must take into account how many times the weight was lifted. The weight was 185 pounds in the test. The formula derived is the following:

If the weight was lifted 0 or 1 times, the default value is given of 0.

If the weight of the player was less than 185 pounds, and the raw amount of repetitions lifted is the score.

For all other players the following formula is used:

$$\text{Adjusted Bench Press} = \frac{185\ pounds}{weight} \times number\ of\ rep$$

<u>Adjustment for Agility</u>

The total distance ran in this test is 140 feet with 8 points of acceleration changes within the course. The same adjustment must be made for the height of the players as was made for sprint speed, but additional adjustments must be made for weight, as more weight makes it hard to switch directions. This means that the bigger the weight, the more of a decrease in score is needed for adjustment. Multiplying by a smaller number would work, thus we divide the number of cuts made by the weight as the final factor.

$$Step\ Size = 1.22 \times Height$$

$$\text{Adjusted Agility} = \frac{140\ feet}{Agility\ Test\ Score} \times \frac{1\ step}{Step\ Size\ (feet)} \times \frac{8}{weight}$$

Observations Involving Athletic Results

After taking every adjusted athletic result, an overall number was created called an *Athletic Score.* The results of each adjusted set was scaled from 0 to 10 to give each athlete a scaled score on the tests. The formula for this overall measure is the following:

$$Athletic\ Score = Sum\ of\ All\ Adjusted\ Athletic\ Scores/4$$

However, an adjustment had to be made for the really poor athletes that simply tested poorly even on the adjusted scores because their size made them too slow, and not explosive enough to get on the scale. To take into account these outliers, with athletic scores of 0, we simply made their athletic score be:

$$Athletic\ Score = Height/2$$

Giving for example a 7-foot player and athletic score of 3.5, making them an average athlete simply because of their height, which is their only significant athletic statistic, as the others were low. The following charts sum up the results of these adjusted scores.

Figure shows an elongated graph, plotting each player's athletic score on a number line, with the ones drafted being marked as red circles, and those that did not get into the NBA, as blue crosses.

Figure 8. Number Line of All Players

There was no clear distinction between a player's athletic score and their draft status. To investigate further in a more legible manner, the following histograms, Figures 9 and 10, display the same result.

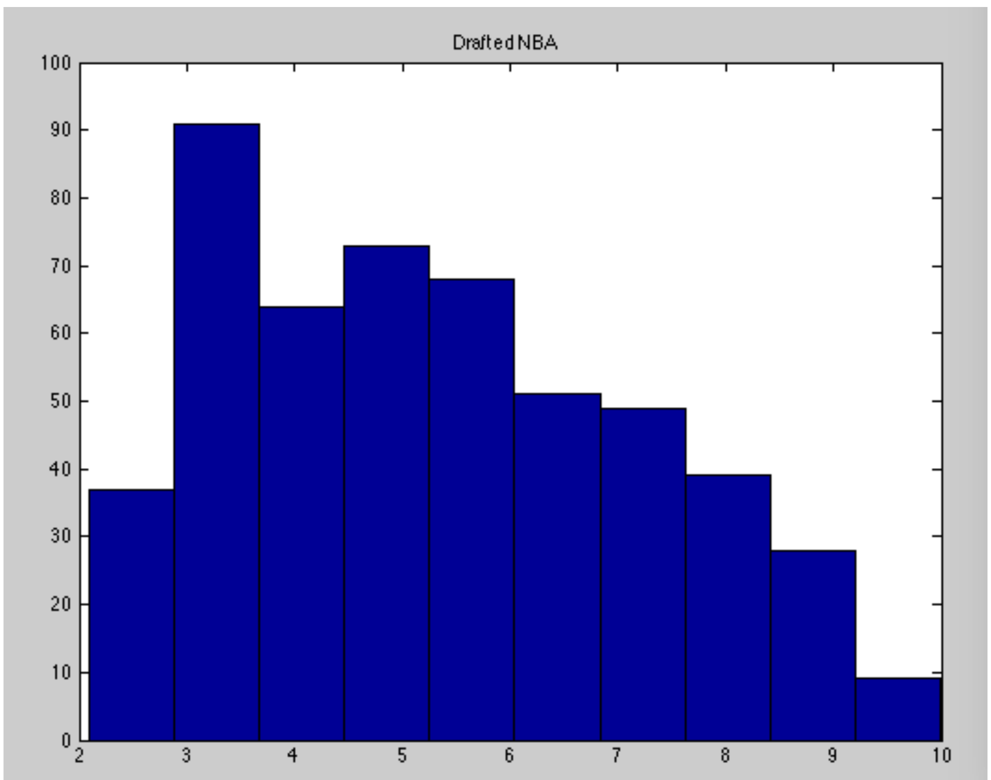


Figure 9. Histogram Of Athletic Scores of NBA players

Bin ->	1	2	3	4	5	6	7	8	9	10
Count ->	37	91	64	73	68	51	49	39	28	9

Table 6. Bin Counts for Drafted Players

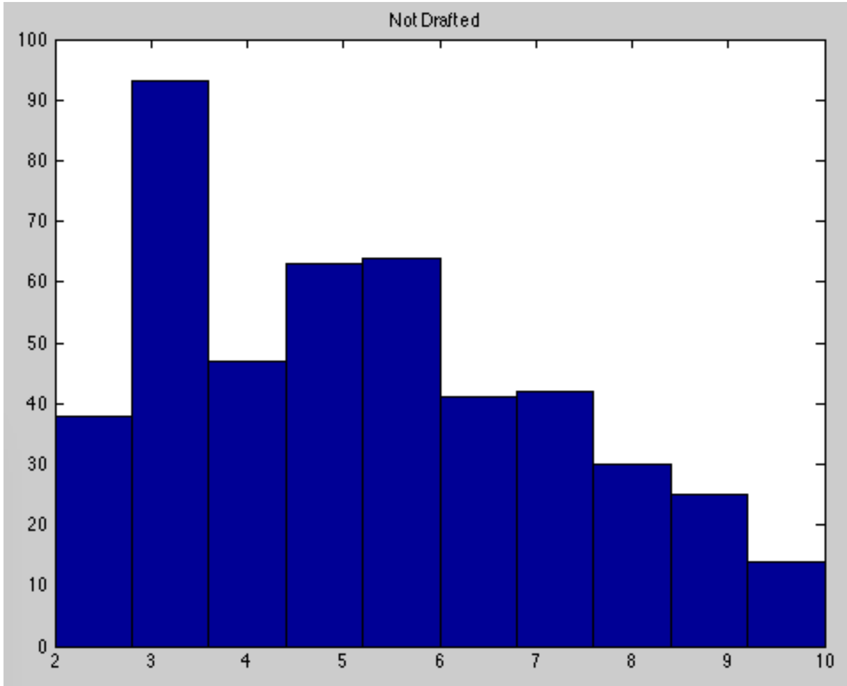


Figure 10. Histogram of Athletic Scores of Undrafted Players

Bin ->	1	2	3	4	5	6	7	8	9	10
Count ->	38	93	47	63	64	41	42	30	25	14

Table 7. Bin Counts for Undrafted Players

The distributions in both situations are very similar. There is a large spike around 3, because of the adjustment made for tall players without a high athletic score, giving them a default value of 3. Otherwise, the distributions are very similar. It was interesting to find that out of the athletes with the highest scores on the athletic tests, only 39% of them made it to the league. So unusual athleticism, interestingly enough, lowered the hopeful players' chances. Another observation is that there is a slight increase in the middle bins for those players that were drafted as opposed to those that were not.

Observations By Sorting Positions

*refer to Appendix page 33 for code

Taking into account all the different players with different desired characteristics was difficult. Some teams needed tall players; some teams needed short and quick players. Thus, the next logical step was to attempt to separate the players into groups based on their playing position. Since this data was not explicitly provided, an extraction technique had to be made to separate the players into their correct categories. The three categories allotted were guards, forwards, and centers. Table 8 shows the approximate height breakdown of each position.

Position 🡻 Height 🡺	Height (Feet-inches)
Guards	<= 6'5"
Forwards	6'5" > height <= 6'10"
Centers	>6'10"

Table 8. Height Chart

The first interesting

investigation was to take the Principal Component Analysis of each group of players, to see if any factor exhibits any sort of strong relation to draft position. Only the last column was extracted from the correlation matrix, to give the relations between all the variables and draft status. Tables 9-11 display the results of each correlation matrix analysis.

Variable ↓	Factor
Height	**0.1185**
Weight	-0.0274
Wingspan	**0.1409**
Reach	0.0136
Standing Vertical	**0.1282**
Standing Vertical Reach	**0.1441**
Running Vertical	0.0916
Running Vertical Reach	**0.1133**
Bench Press	-0.0313
Agility	**-0.1000**
Sprint Speed	-0.0397
Draft Status	1

Table 9. Guards Stats Correlation to Draft Status

Variable ↓	Factor
Height	-0.0310
Weight	-0.0140
Wingspan	-0.0596
Reach	-0.0846
Standing Vertical	0.0460
Standing Vertical Reach	-0.0254
Running Vertical	0.0573
Running Vertical Reach	-0.0061
Bench Press	-0.0211
Agility	**-0.1221**
Sprint Speed	-0.0318
Draft Status	1

Table 10. Forwards Stats Correlation to Draft Status

Variable ↓	Factor
Height	0.1237
Weight	0.0677
Wingspan	**0.1091**
Reach	**0.1456**
Standing Vertical	-0.0567
Standing Vertical Reach	0.0774
Running Vertical	0.0050
Running Vertical Reach	**0.1177**
Bench Press	-0.0406
Agility	**-0.0900**
Sprint Speed	0.0062
Draft Status	1

Table 11. Centers Stats Correlation to Draft Status

Further analysis of the above results leads us to conclude that all players, once broken down amongst positions, had a modest correlation between their measurements and draft status. The guards seem to have a positive correlation between draft status and height, wingspan, standing vertical, standing vertical reach, and running vertical reach. This means that the guards were favored if they were bigger, lankier, and jumped high. There was also a slight negative correlation with agility, meaning that the quicker and more agile the guard, the more it helped him on the draft. With forwards, the strongest correlation was a negative one with agility meaning that the quicker forwards were favored. For centers, it seems that wingspan, reach, running vertical reach, and quick agility were desired. This led to the natural transition into use of the PCA. The following charts (Tables 13-15) are the results of the 1st and 2nd Principal Components of each position.

Variable ↓	Component 1	Component 2
Height	-0.3837	0.2200
Weight	-0.3356	0.1752
Wingspan	**-0.4151**	0.1132
Reach	-0.3933	0.2405
Standing Vertical	-0.1469	**-0.4982**
Standing Vertical Reach	**-0.4263**	-0.1719
Running Vertical	-0.1208	**-0.5238f**
Running Vertical Reach	**-0.4039**	-0.2290
Bench Press	-0.1378	-0.1125
Agility	-0.0944	0.2734
Sprint Speed	-0.0134	0.3969
Draft Status	-0.0635	0.0285

Table 12. Guards PCA Analysis

Variable ↓	Component 1	Component 2
Height	-0.2365	0.3051
Weight	-0.2857	0.1780
Wingspan	-0.1721	**0.4530**
Reach	-0.2771	**0.4111**
Standing Vertical	**0.4470**	0.1637
Standing Vertical Reach	0.1960	**0.4934**
Running Vertical	**0.4717**	0.1095
Running Vertical Reach	0.2752	**0.4355**
Bench Press	0.0284	0.0523
Agility	-0.3041	0.1471
Sprint Speed	-0.3515	-0.0543
Draft Status	0.0523	-0.0431

Table 13. Forwards PCA Analysis

Variable ➜	Component 1	Component 2
Height	-0.1956	0.3046
Weight	-0.2227	0.1354
Wingspan	-0.0884	**0.5005**
Reach	-0.2426	**0.4387**
Standing Vertical	**0.4394**	0.1136
Standing Vertical Reach	0.2748	0.4532
Running Vertical	**0.4558**	0.0545
Running Vertical Reach	0.3219	0.3963
Bench Press	0.1343	-0.0538
Agility	**-0.3212**	0.2249
Sprint Speed	**-0.3753**	0.0362
Draft Status	0.0583	**0.1129**

Table 14. Centers PCA Analysis

For guards the first axis is spread mostly along wingspan, and how high they can touch from standstill and running positions. This makes sense, because there is a high variation of skill level as well among guards. There are some who are un-athletic but great shooters, and thus they are still desired. The second axis is dominated by agility. This also makes sense because guards are highly valued for their speed. The forwards are spread amongst their jumping ability in the first component and their physical size in the second component. Centers are spread mostly along their jumping abilities, and also quickness. In the second component, centers are spread along their body size, and their draft status. To make the point more clear, the following graphs are the plots of the data along their 2 principal components, with a least squares fit line introduced for all 3 groups. The players who got drafted are marked in red; those that did not are in blue.

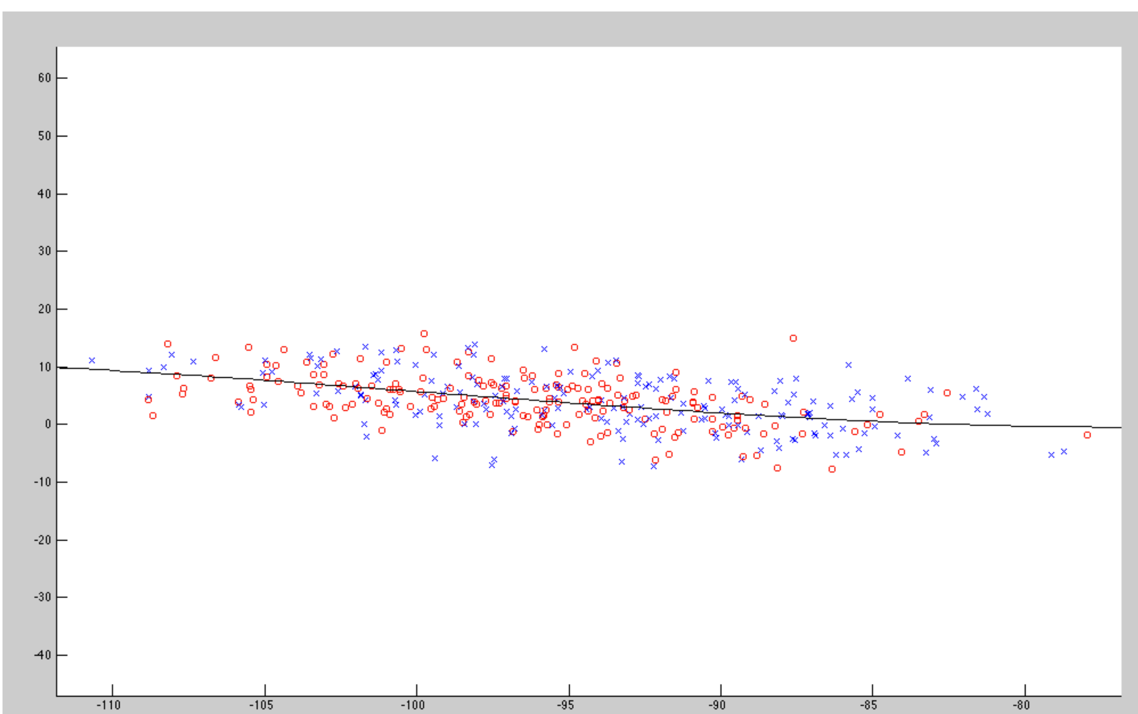


Figure 11. Guards PCA Plot

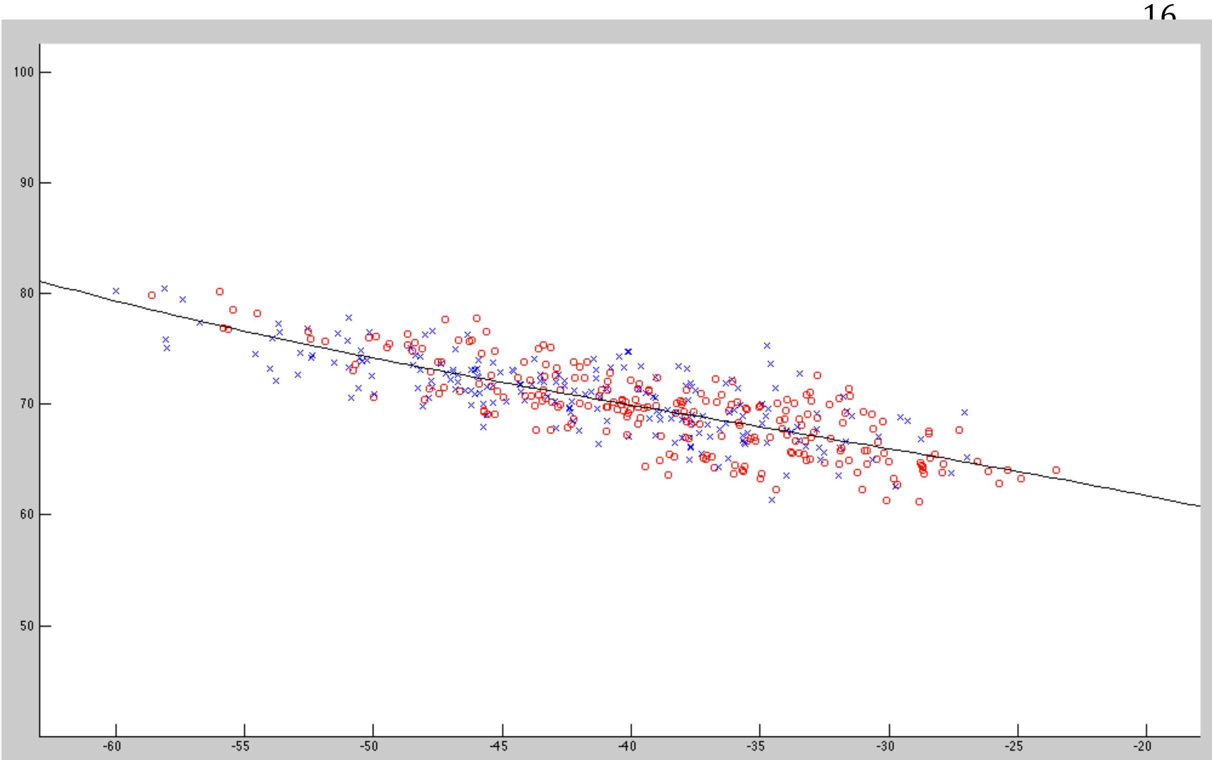


Figure 12. Forwards PCA Plot

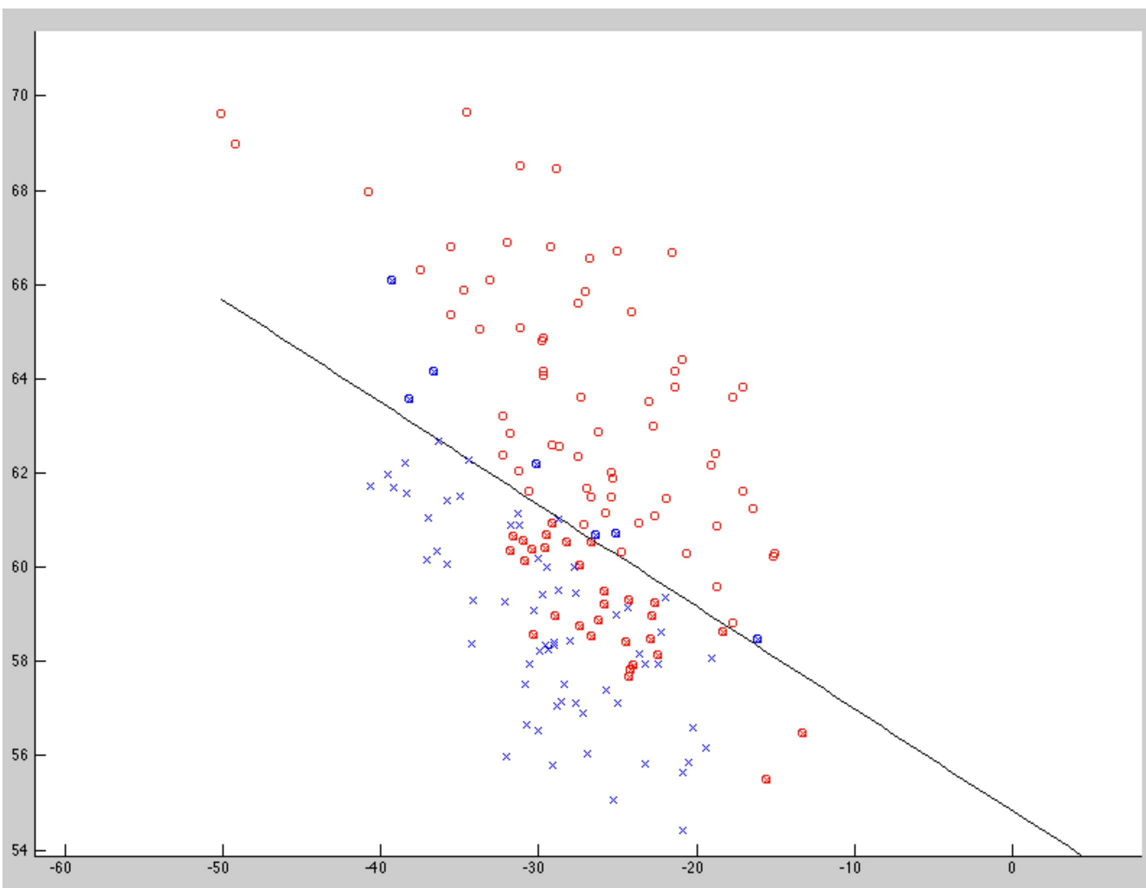


Figure 12. Forwards PCA Plot

From observing the PC graph of the guards, it isn't clear that there is a distribution of the guards in relation to their draft status. However, amongst the forwards there seems to be a slight correlation. If a player is below the fit line, it seems like there is more of a chance they would make it. Upon further calculations, when taking into account only the players below the fit line, a player underneath this fitted line has a **62%** chance of being drafted. This is significantly higher than the raw chances of any player at the combine of **46.2%** as presented in figure 7. Since the fit line is almost horizontal, it would appear that the preferred characteristics of a forward, to maximize their chances, would be to have a slightly lower vertical leap than their peers. This at first is counter-intuitive, but when you take into account the time players sometimes spend on improving things like vertical, it makes sense. This data might be analyzed to explain how for guards, and forwards alike skill is more important then physical ability. Lastly, the center chart is the only one so far that shows a clear trend between physical ability and draft status. If a center is above the fit line, there is a very high **70%** chance of getting drafted. This lets us know that in addition to a center having quick feet, and a good vertical leap, the distinctive characteristic is wingspan and reach, which is correlated to their draft status. Centers with extreme wingspans, and high reaches can be very useful defensively to any team to protect the rim from drives, and shorter players coming in close to the basket.

From observing all these characteristics, it became clear that all NBA players had something very interesting in common: it appeared that they all had longer wingspans than their heights. It is a commonly accepted fact, from Leonardo da Vinci's drawings and theories, that a man's wingspan is equal to a man's height. It appears that NBA players did not follow this ratio. One study quotes the average man as having a wingspan approximately 2.1 inches longer than their height[5]. This seemed to fit the NBA players as well making them tall, but averagely lanky men. However, upon plotting the following figure it was apparent that these athletes were anything but average. Figure 13 plots the expected average wingspan for each height in black, each NBA

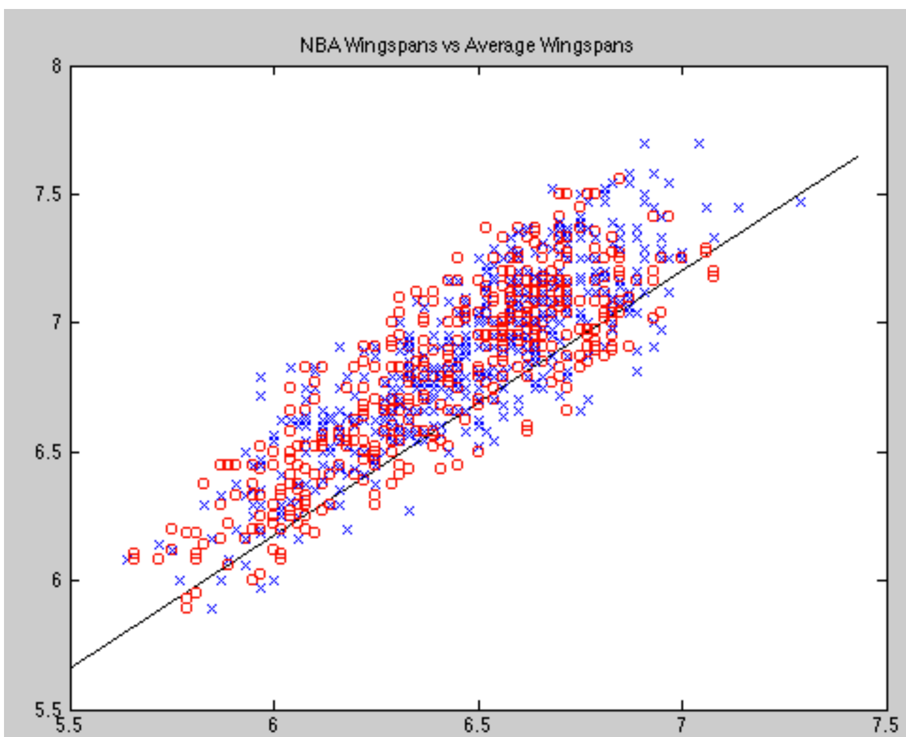


Figure 13. Wingspan Vs. Height

player that was drafted in red, and those who were not in blue. It is clear that the vast majority of all players trying out for the NBA have exaggeratedly longer wingspans than heights. The results actually revealed a 4.39 inch difference on average. This trend is very clear amongst all heights, positions, and ability levels. Wingspan is a crucial physical characteristic for basketball players because it simply allows them to play the game at a higher level. Consider two players of the same height but with different wingspans. The one with shorter arms will need to have a higher vertical leap in order to touch as high of a height as the player with longer arms can touch, which is good for shooting, blocking, playing defense in general, and dunking. Thus wingspan is a highly desired physical characteristic.

*refer to Appendix page 38 for code

Observations for NBA Hopefuls

*refer to Appendix page 40 for code

Even though there were few physical characteristics that were indicative of getting drafted, it is interesting to use historical data of all the drafted players to get a model for what a successfully drafted player has in terms of weight, height, and maximum vertical leap. This would be a fun tool to use for things such as to predict the necessary minimum leap a player would need to historically make the NBA at a given height and weight.

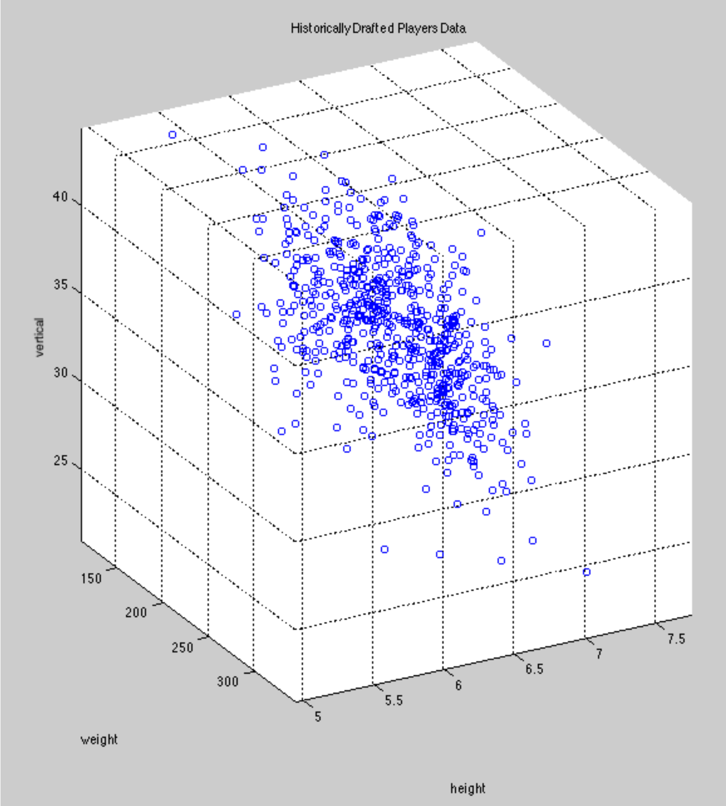


Figure 14. Overall Plot of All Players Drafted

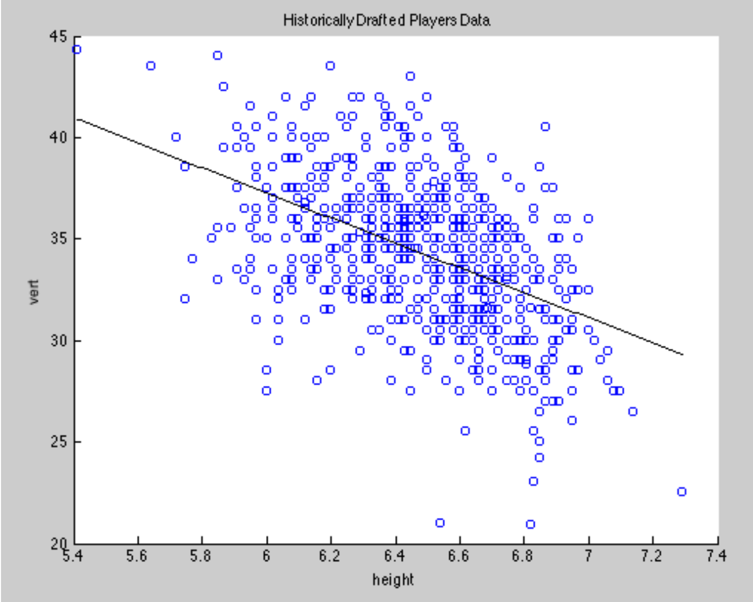


Figure 15. Height vs Vert for Drafted Players

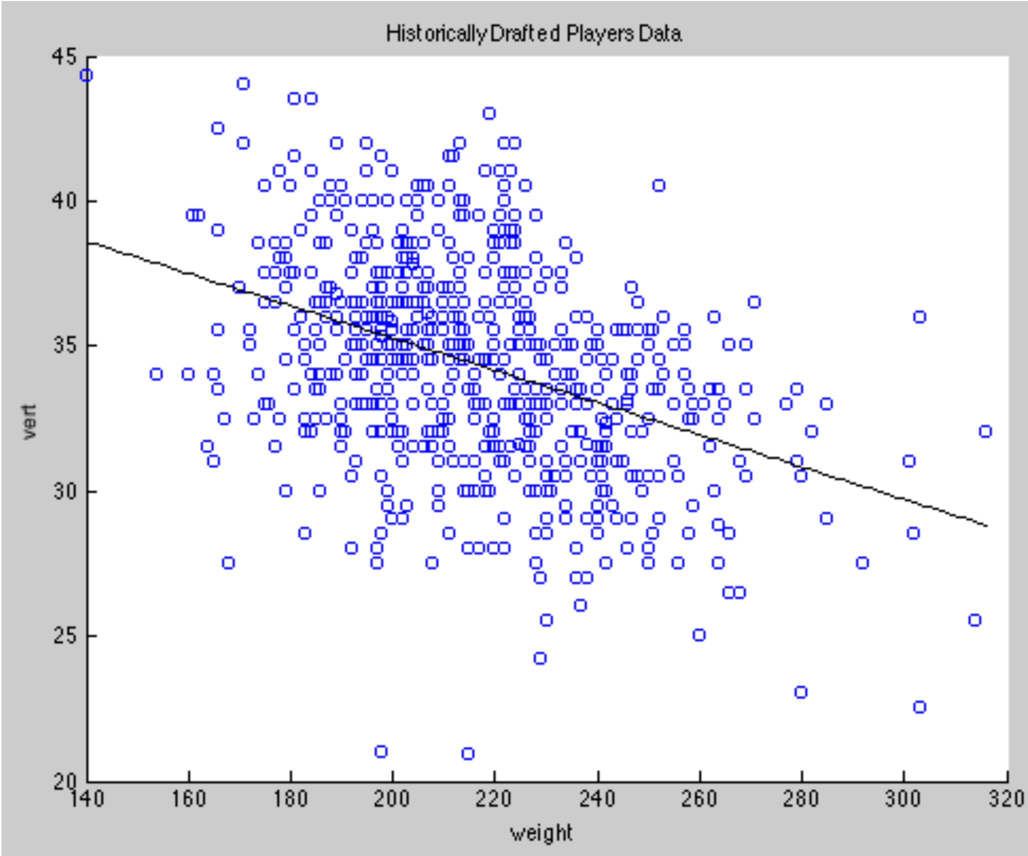


Figure 16. Weight against Vert

The following fit formula are obtained:

Height against Vertical

$$Y = -6.1485x + 74.1690$$

Weight against Vertical

$$Y = -0.0556x + 46.4022$$

Thus, someone who is 6 feet tall with a weight of 170 pounds would need the following vertical to historically be drafted.

$$Y1 = -6.1485 \times 6 + 74.1690 = 37.278\ inches$$

$$Y2 = -0.0556 \times 170 + 46.4022 = 36.9502\ inches$$

Average of the obtained guesses = 37.11 inches

According to this, an average sized high school basketball player needs to have a vertical leap of over 3 feet to be historically as athletic in the vertical leap measurement as previous NBA draftees. Using these two equations, one can get a pretty good estimate of what vertical leap an athlete needs to have to compete with previous NBA athletes.

Lastly, after observing so many athletes with exceptional body measurements, it is interesting to observe how often exceptional athletes, such as a Lebron James or Michael Jordan, are born. Considering an average size draft class to be 60 players each year, a simulation can be created to create 60 random players until one is created with the physical characteristics that are considered exceptional. We define an exceptionally gifted player as one with the following characteristics:

Variable ↓	Factor
Height	6.75 < height
Running Vertical	44 inches < vertical
Bench Press	10 < bench
Agility	10.3 > agility
Sprint Speed	3.05 > sprint

Table 15. Desired Characteristics

*refer to Appendix page 41 for code

After running the simulation a few times, it appears as though these types of players are very rare. Some of the results were the following:

Draft Classes Passed	Height (feet)	Running Vertical (inches)	Bench Press (reps)	Agility (seconds)	Sprint Speed (seconds)
18	6.78	44.25	25	10.23	2.98
5	7.01	49.40	19	10.09	3.00
20	6.93	47.49	26	9.8	2.99
28	6.84	45	21	10.29	2.82
4	6.90	46.27	20	9.42	2.99

Table 16. Simulation Results

Thus the average number of draft classes that passed until a Lebron James-type player was created was 15 years. This seems pretty accurate, as these players are very rare, with Michael Jordan comparable player to Lebron James in recent years, even though they were spaced about 19 years apart.

Conclusions

Considering all the results investigated, the following conclusion can be drawn. First, the data was valid, as the distribution of measured characteristics was all more or less normally distributed, which is natural. Next, taking into account all the players for which I had data and who participated in the NBA combine, I discovered only 46.2% of them got drafted. Therfore, I searched for a distinction between those drafted and those not drafted.

On average, the maximum vertical leap of the players drafted was 1.58 inches higher than of those not drafted. The drafted players' body fat was also 1% lower than their undrafted peers. Using various methods, such as deriving a formula which scaled and weighted all of the players' measurements in various athletic categories, such as vertical leap, sprint speed, agility, and bench press tests, no clear characteristics were found which led to a player being drafted. The formula resulted in an overall scaled athletic score, on a scale of 1-10 that rated each player, and both those drafted and undrafted had a similar distribution of scores. One interesting trend was that there were slightly more players in the medium score range that got drafted. The players that had scores of 10, interestingly enough, had a slightly lower chance of getting drafted, at 39%. This led to the conclusion that being a great athlete did not have a strong correlation with being drafted, at least as far as taking all the players at once and measuring them against one another.

After hitting a slight dead end with athletics, the next idea was to split the players up approximately according to position by taking into account their height. Using principal component analysis, this method revealed some interesting results. For forwards, their chances were increased to 62% by having a slightly lower vertical leap than the players they competed against, but a higher agility, and quickness. This makes sense, as a forward's key features are height, and ability to move quickly at that height. The other dominant characteristic of good forwards is great skill and intelligence with the basketball, which cannot be measured easily. The clearest result was with centers, which had a 70% chance of getting drafted if they had a comparatively higher vertical, and the longest of reaches amongst their peers. This also fits with basketball philosophy, as it is important for a center to have long arms and be able to jump in order to defend the rim from close shots, grab rebounds, make close offensive shots, and get offensive rebounds.

An interesting characteristic that all tested players had incredibly long arms. The average humans wingspan is 2.1 inches longer than their height. The average player difference was 4.39 inches. Thus, to even have a good chance of being a top-level basketball player, a long wingspan is crucial.

Since many players are interested in playing in the NBA, it was optimal to develop 2 formulas which allow a player to calculate their target vertical leap necessary to be on par with historically drafted players. The following two formulas can be used to get the vertical leap needed by an athlete to have average NBA athleticism.

Vertical Measure 1 =-6.1485 * (height in feet) +74.1690

Vertical Measure 2 = -0.0556*(weight in pounds) +46.4022

Finally, after observing all types of players in the NBA, some with outlying abilities, and others with average ones, it was interesting to find out how often very extreme outliers are born, such as Michael Jordan and Lebron James. After running a simulation to see how long it would take for a random player to be born that would possess the same abilities as Lebron James, an average of 15 years was obtained. Michael Jordan, who had very similar statistics to Lebron, was drafted exactly 19 years before Lebron, so the estimation is rather accurate.

Overall, the conclusion is that it is very difficult to predict an NBA player's draft potential solely based on their physical tests. This study proves that basketball is a game of skill. However athleticism is proven to come in handy in some cases on draft day, as having great athleticism in various categories specific to positions does slightly increase a player's chances of getting drafted.

Summary

1. Data familiarity/interest (Why is it that are you knowledgeable about/interested in this data?)

I have been playing basketball for over 10 years, and I love the sport. I played 3 years on varsity in high school, and was offered a spot on the NYU basketball team but declined to come to UCLA. I am interested in this particular data because I want to see how body measurements lead to a draft position in the NBA.

2. Data acquisition effort (How much effort is required to construct the dataset?)

5 hours of effort were required to construct the dataset as a parser had to be written in C++ in order to format the data in a usable format. The data was converted into a csv file to be worked with in MATLAB.

3. Data novelty (In what ways is looking at this kind of data new or innovative?)

This data analysis is innovative because after searching online, I couldn't find an analysis of NBA player's body statistics in relation to their draft position. I think it'll be a good analysis to predict a player's chance at making the league based on their bodily statistics.

4. Analysis effort (How much work is needed to analyze the data?)

There was a lot data processing needed to interpret the data. Over 50 hours were spent gathering data and perfecting the analysis tools I developed in order to extrapolate relevant information.

5. Analysis methods (Which methods are most unusual, novel or worthwhile?)

Several tools learned in class were used such as PCA, least squares fitting, in addition to a custom algorithm developed by me. I made an algorithm for deriving a scaled athletic score that rated each player based on 4 tests on a scale of 1-10 that adjusted for each player's height, and weight differences. In addition, I also created a draft class simulator that simulated years of draft classes in search for how long an ultra-athlete took to be born. I also created a 2 formula vertical jump predictor, which could predict vertical leap heigt a player would need on average to have athleticism that on par with drafted NBA players.

6. Project difficulty (In what ways is the project most challenging?)

This project was one of the most difficult CS projects I've had in a class at UCLA. It makes sense, as it was a final project, and I actually did not mind the many hours of work that were necessary to put this project together because I found it very enjoyable to be analyzing data I am very passionate about. It was a very cool project and I'm glad I did it.

7. Project relevance to the course (How does the project relate to matrix methods and modeling?)

The project relates highly to the course, as it required heavy MATLAB usage, matrix methods, and data modeling to find my resulting relations. I used several methods from the course notes in addition to my own, discussed above and below.

8. Project novelty (Which aspects of the project are most unusual or creative?)

The project's most interesting characteristics were the fact that this data set has never been publically analyzed (at least to my knowledge), and the results were novel. I created 3 new "algorithms" / formulas, one that measured how many years had to pass for an outlying athlete had to be born, one that helped athletes predict necessary vertical leaps, and an algorithm that scaled and weighted an athletes athleticism based on test scores.

9. Report (Which aspects of the report require the most effort?)

The report's most difficult component was interpreting the data analysis because there was a lot of data to go through, and thinking of how to analyze it was difficult as well. After looking at trends in the data I came up with the methods I u sed.

10. Overall effort (Which aspects of the project require the most effort?)

The project's overall most effort requiring part was the data analysis as stated above. It was a very fun project that took many hours to finish but it was worth it and the challenges that came with it. I learned how to create algorithms that adapt to what I am trying to analyze, such as the athletic score algorithm. I really liked this project.

APPENDIX

```
% //col 1 is height
% //col 2 is weight
% //col 3 is wingspan
% //col 4 is reach
% //col 5 is no step vert
% //col 6 is no step vert reach
% //col 7 is running vert
% //col 8 is running vert reach
% //col 9 is draft or not


%%%%% INTRO STUFF %%%%

%%%% 1st distributions

%%% height
%%% wingspan
%%% vertical standing
%%% vertical running

data = csvread('realFinalData.csv');

height=data(:,1);
wingspan=data(:,3);
vertStand=data(:,5);
vertRunning=data(:,7)

hist(height)
title('Distribution of Height');
figure;
hist(wingspan)
title('Distribution of Wingspan');
figure;
hist(vertStand)
title('Distribution of Standing Vertical Leap');
figure;
hist(vertRunning)
title('Distribution of Running Vertical Leap');
figure;

%%% end distributions
```

```
%%% see first of all what percantages of all draftees got into NBA %%%


drafted=data(:,9);
in=0;
for i=1:1384
	if(drafted(i)>0)
		in=in+1;
		boolDraft(i)=1;
	else
		boolDraft(i)=0;
	end
end
out=1384-in;
 %% percent chance of being in NBA if you get invited to combine

percOut=out/1384*100
percIn=in/1384*100

str=sprintf('%.1f',percIn);
s=strcat('Drafted %',str);
str1=sprintf('%.1f',percOut);
s1=strcat('Undrafted %',str1);

mat=[in;out];
labels = {s,s1};
explode=[1 0];
p=pie3(mat,explode,labels);

hp = findobj(p, 'Type', 'patch');

set(hp(2), 'FaceColor', 'g');




%%% now see if theres any relation between height and getting drafted
%%% since most people think being short you have no chance

corrcoef([height boolDraft'])
%% almost 0 correlation coefficient


%%% averages of key stats for groups like drafted or not drafted

mean(height)
```

```
mean(vertRunning)

%% now the same measurments of only the ones who are drafted
cur=1;
cur2=1;
for i=1:1384
	if(boolDraft(i)==1)
		onlyDrafted(cur,1:9)=data(i,:);
		cur=1+cur;
	else
		notDrafted(cur2,1:9)=data(i,:);
		cur2=cur2+1;
	end
end

draftedHeight=onlyDrafted(:,1);
draftedRunVert=onlyDrafted(:,7);
draftedWingspan=onlyDrafted(:,3);
draftedReach=onlyDrafted(:,4);

notDraftedHeight=notDrafted(:,1);
notDraftedRunVert=notDrafted(:,7);
notDraftedWingspan=notDrafted(:,3);
notDraftedReach=notDrafted(:,4);



mean(draftedHeight)
mean(draftedRunVert) %% higher by 1.58 inches
mean(draftedWingspan)
mean(draftedReach)

mean(notDraftedHeight)
mean(notDraftedRunVert)
mean(notDraftedWingspan)
mean(notDraftedReach)


%% the max running vert is higher for that of those not drafted

%%%%%%%%%%%%%% NOW MEASURE other athletic abilities in comparison
dataNBF=csvread('dataWithoutBodyFat.csv');
dataBF=csvread('dataWithBodyFat.csv');

%size 966
```

```
draft=dataNBF(:,12);
draftFat=dataBF(:,13);

cur=1;
cur2=1;
for i=1:966
        if(draft(i)>0)
                noBFdrafted(cur,1:12)=dataNBF(i,:);
                cur=1+cur;
        else
                noBFundrafted(cur2,1:12)=dataNBF(i,:);
                cur2=cur2+1;
        end
end

cur=1;
cur2=1;
for i=1:638
        if(draftFat(i)>0)
                BFdrafted(cur,1:13)=dataBF(i,:);
                cur=1+cur;
        else
                BFundrafted(cur2,1:13)=dataBF(i,:);
                cur2=cur2+1;
        end
end

vertDraft=noBFdrafted(:,7);
vertUnDraft=noBFundrafted(:,7);

mean(vertDraft)
mean(vertUnDraft)

benchDraft=noBFdrafted(:,9);
agilityDraft=noBFdrafted(:,10);
sprintDraft=noBFdrafted(:,11);P
draftDraft=noBFdrafted(:,12);

benchUnDraft=noBFundrafted(:,9);
agilityUnDraft=noBFundrafted(:,10);
sprintUnDraft=noBFundrafted(:,11);
draftUnDraft=noBFundrafted(:,12);

bodyFatDrafted=BFdrafted(:,5); %size 638
bodyFatUndrafted=BFundrafted(:,5);
```

```
mean(bodyFatDrafted)
mean(bodyFatUndrafted)


mean(benchDraft)
mean(benchUnDraft)

mean(agilityDraft)
mean(agilityUnDraft)

mean(sprintDraft)
mean(sprintUnDraft)


leapFatUndrafted=BFundrafted(:,8);
leapFatdrafted=BFdrafted(:,8);
mean(leapFatdrafted)
mean(leapFatUndrafted)

%%%%% END INTRO STUFF %%%%%%


%%%% BEGIN ATHLETIC FORMULA

data=csvread('dataWithoutBodyFat.csv');


height=data(:,1);
weight=data(:,2);
vert=data(:,7);
bench=data(:,9);
agility=data(:,10);
sprint=data(:,11);
draft=data(:,12);




%adjust vert first
for i=1:966
	adjVert(i)=vert(i)/weight(i);
end
%adjust sprint speed
for i=1:966
	stepSize=1.22*height(i);
	stepsTaken=75/stepSize;
```

```
        footSpeed=stepsTaken/sprint(i);
        adjSpeed(i)=footSpeed;
end

%adjusted bench press
for i=1:966
        if(bench(i)==0)
                adjBench(i)=0;
                continue;
        end
        if(bench(i)==1)
                adjBench(i)=0;
                continue;
        end
        if(weight(i)<=185)
                adjBench(i)=bench(i);
                continue;
        end
        if(weight(i)>185)
                adjBench(i)=(185/weight(i))*bench(i);
                continue;
        end
end

%adjusted agility
for i=1:966
        stepSize=1.22*height(i);
        mid=(140/agility(i))*(1/stepSize);
        adjAgility(i)=(8/weight(i))*mid;
        adjAgility(i)=mid;
end


newData=[height weight data(:,3) data(:,4) data(:,5) data(:,6) adjVert' data(:,8)
adjBench' adjAgility' adjSpeed' draft];

%now adjust each score to be a grade out 10 so we can add them all up

sortV=sort(adjVert);
sortSpeed=sort(adjSpeed);
sortAgility=sort(adjAgility);
sortBench=sort(adjBench);

minVert=sortV(50);
maxVert=sortV(900);
```

```
minSprint=sortSpeed(50);
maxSprint=sortSpeed(900);

minAgility=sortAgility(50);
maxAgility=sortAgility(900);

minBench=sortBench(50);
maxBench=sortBench(900);


for i=1:966
        benchPt=adjBench(i);
        agilityPt=adjAgility(i);
        sprintPt=adjSpeed(i);
        vertPt=adjVert(i);

        scaledBench(i)= (10*(benchPt-minBench))/(maxBench-minBench);
        if(scaledBench(i)<0)
                scaledBench(i)=0;
        end
        if(scaledBench(i)>10)
                scaledBench(i)=10;
        end

        scaledAgility(i)=(10*(agilityPt-minAgility))/(maxAgility-minAgility);
        if(scaledAgility(i)<0)
                scaledAgility(i)=0;
        end
        if(scaledAgility(i)>10)
                scaledAgility(i)=10;
        end


        scaledSprint(i)=(10*(sprintPt-minSprint))/(maxSprint-minSprint);
        if(scaledSprint(i)<0)
                scaledSprint(i)=0;
        end
        if(scaledSprint(i)>10)
                scaledSprint(i)=10;
        end

        scaledVert(i)=(10*(vertPt-minVert))/(maxVert-minVert);
        if(scaledVert(i)<0)
                scaledVert(i)=0;
        end
        if(scaledVert(i)>10)
```

```
        scaledVert(i)=10;
    end



    mid=scaledBench(i)+scaledVert(i)+(scaledAgility(i))+(scaledSprint(i));
    athleticScore(i)=mid/4;
    if(athleticScore(i)<2)
        athleticScore(i)=(height(i))/2;
    end
end


inAbove5=0;
inBelow5=0;

outAbove5=0;
outBelow5=0;
for i=1:966
    if(athleticScore(i)>5)
        if(draft(i)>0)
            inAbove5=inAbove5+1;
            scatter(athleticScore(i),1,'ro');
        else
            outAbove5=outAbove5+1;
            scatter(athleticScore(i),1,'bx');
        end

    hold on;
else
        if(draft(i)>0)
            inBelow5=inBelow5+1;
            scatter(athleticScore(i),1,'ro');
        else
            outBelow5=outBelow5+1;
            scatter(athleticScore(i),1,'bx');
        end

    hold on;
end
end
figure;

in=inAbove5+inBelow5;
out=outAbove5+outBelow5;
for i=1:966
```

```
	if(draft(i)>0)
		adjDraft(i)=1;
	else
		adjDraft(i)=0;
	end
end

cur=1;
cur1=1;
for i=1:966
	if(draft(i)>0)
		dataIn(cur)=athleticScore(i);
		cur=cur+1;
	else
		dataOut(cur1)=athleticScore(i);
		cur1=cur1+1;
	end
end
hIn=hist(dataIn)
title('Drafted NBA');
figure;
hOut=hist(dataOut)
title('Not Drafted');


hist(dataIn)
title('Drafted NBA');
figure;
hist(dataOut)
title('Not Drafted');
%%%%%%%%%%%% taking into accoutn all players no difference



%%%%% now looking into players by position
data=csvread('dataWithoutBodyFat.csv');

height=data(:,1);

cur=1;
for i=1:966
	if(height(i)<=6.41)
		guards(cur,1:12)=data(i,:);
		cur=cur+1;
	end
end
```

```
cG=corr(guards);
[U,S,V]=svd(cG);

firstG=U(:,1);
secondG=U(:,2);
plotFG1=guards*firstG;
plotFG2=guards*secondG;
guardDraft=guards(:,12);
%sze is 348
for i=1:384
    if(guardDraft(i)>0)
    scatter(plotFG1(i),plotFG2(i),'ro')
    hold on;
    else
    scatter(plotFG1(i),plotFG2(i),'bx')
    hold on;
    end
end
hold on;

%% least squares fit guards
A=[plotFG1.^3 plotFG1.^2 plotFG1.^1 plotFG1.^0];
b=plotFG2;
x=A\b;
p=[x'];
d=linspace(min(plotFG1),max(plotFG2),384);
y=polyval(p,d);
hold on;
plot(d,y,'k');

inAbove=0;
inBelow=0;

outAbove=0;
outBelow=0;

for i=1:384
    y=polyval(p,plotFG1(i));
    if(plotFG2(i)>y)%%ABOVE the fit line
        if(guardDraft(i)>0)
            inAbove=inAbove+1;
        else
            outAbove=outAbove+1;
        end
    else
```

```
        if(guardDraft(i)>0)
            inBelow=inBelow+1;
        else
            outBelow=outBelow+1;
        end
    end
end

inAboveG=inAbove;
inBelowG=inBelow;
outAboveG=outAbove;
outBelowG=outBelow;


figure;


%% forwards below 6'9.5" 489

cur=1;
for i=1:966
    if((height(i)>6.41) && (height(i)<6.79))
        forwards(cur,1:12)=data(i,:);
        cur=cur+1;
    end
end
cF=corr(forwards);
[UF,S1,V1]=svd(cF);

firstF=UF(:,1);
secondF=UF(:,2);
plotFF1=forwards*firstF;
plotFF2=forwards*secondF;
forDraft=forwards(:,12);
%sze is489
for i=1:453
    if(forDraft(i)>0)
    scatter(plotFF1(i),plotFF2(i),'ro')
    hold on;
    else
    scatter(plotFF1(i),plotFF2(i),'bx')
    hold on;
    end
end

hold on;
```

```
%%%%repeat forForwards
A=[plotFF1.^3 plotFF1.^2 plotFF1.^1 plotFF1.^0];
b=plotFF2;
x=A\b;
p=[x'];
d=linspace(min(plotFF1),max(plotFF2),453);
y=polyval(p,d);
hold on;
plot(d,y,'k');

inAbove=0;
inBelow=0;

outAbove=0;
outBelow=0;

for i=1:384
	y=polyval(p,plotFF1(i));
	if(plotFF2(i)>y)%%ABOVE the fit line
		if(forDraft(i)>0)
			inAbove=inAbove+1;
		else
			outAbove=outAbove+1;
		end
	else
		if(forDraft(i)>0)
			inBelow=inBelow+1;
		else
			outBelow=outBelow+1;
		end
	end
end

inAboveF=inAbove;
inBelowF=inBelow;
outAboveF=outAbove;
outBelowF=outBelow;

figure;



%% centers above 6 9.5 164
```

```
cur=1;
for i=1:966
	if((height(i)>=6.75))
		centers(cur,1:12)=data(i,:);
		cur=cur+1;
	end
end

centerDraft=centers(:,12);
cC=corr(centers);
[UC,S,V]=svd(cC);
firstC=UC(:,1);
secondC=UC(:,2);
plotC1=centers*firstC;
plotC2=centers*secondC;

for i=1:164
	if(centerDraft(i)>0)
	scatter(plotC1(i),plotC2(i),'ro')
	hold on;
	else
		scatter(plotC1(i),plotC2(i),'bx')
	hold on;
	end
end

hold on;
% plotFG1.^4

%%%%repeat for centers
A=[plotC1.^1 plotC1.^0];
b=plotC2;
x=A\b;
p=[x'];
d=linspace(min(plotC1),max(plotC2),164);
y=polyval(p,d);
hold on;
plot(d,y,'k');
hold on;

inAbove=0;
inBelow=0;

outAbove=0;
outBelow=0;
```

```
for i=1:164
	y=polyval(p,plotC1(i));
	if(plotC2(i)>y)%%ABOVE the fit line
		if(centerDraft(i)>0)
			inAbove=inAbove+1;
			scatter(plotC1(i),plotC2(i),'ro')
			hold on;
		else
			scatter(plotC1(i),plotC2(i),'bo')
			outAbove=outAbove+1;
			hold on;
		end
	else
		if(centerDraft(i)>0)
			scatter(plotC1(i),plotC2(i),'rx')
			inBelow=inBelow+1;
			hold on;
		else
			scatter(plotC1(i),plotC2(i),'bx')
			outBelow=outBelow+1;
			hold on;
		end
	end
end


inAboveC=inAbove;
inBelowC=inBelow;
outAboveC=outAbove;
outBelowC=outBelow;

%%%% wingspan plots



dataNoBf=csvread('dataWithoutBodyFat.csv');


data=dataNoBf;

draft=data(:,12);
height=data(:,1);
wingspan=data(:,3);

cur=5.5;
for i=1:966
```

```
        x(i)=cur;
        cur=cur+0.002;
end

plot(x,x*1.029,'k');
hold on;


in=0;
out=0;
for i=1:966
        if((draft(i)>0))
        scatter(height(i),wingspan(i),'bx');
        in=in+1;
        hold on;
        end
        if((draft(i)<=0))
        scatter(height(i),wingspan(i),'ro');
        out=out+1;
        hold on;
        end
end

title('NBA Wingspans vs Average Wingspans')


data = csvread('realFinalData.csv');
wingspan=data(:,3);
w=mean(wingspan);
h=mean(height);
w-h

%%%%%%%% plots of average draftees

data = csvread('realFinalData.csv');

draft=data(:,9);
cur=1;
for i=1:1384
        if(draft(i)>0)
                draftOnly(cur,1:9)=data(i,:);
                cur=cur+1;
        end
end
data=draftOnly;
```

```
height=data(:,1);
weight=data(:,2);
vert=data(:,7);


scatter3(height,weight,vert);

title('Historically Drafted Players Data');
xlabel('height');
ylabel('weight');
zlabel('vertical');
figure;
%%%%%% 3d plot of both variables

%%% go for weight against vertical

A=[weight.^1 weight.^0];
b=vert;
x=A\b;
p=[x'];
weightFactors=p;
d=linspace(min(weight),max(weight),164);
y=polyval(p,d);
scatter(weight,vert);
hold on;
plot(d,y,'k');
title('Historically Drafted Players Data');
xlabel('weight');
ylabel('vert');


%%% go for height against vertical
figure;

A=[height.^1 height.^0];
b=vert;
x=A\b;
p=[x'];
heightFactors=p;
d=linspace(min(height),max(height),164);
y=polyval(p,d);
scatter(height,vert);
hold on;
plot(d,y,'k');
title('Historically Drafted Players Data');
xlabel('height');
```

```
ylabel('vert');


%%% creating a random lebron james

players=0;
done=0;
while (done == 0)
% make random height
% between 5.5 and 7.5
%open interval a and b
players=players+1;

a = 5.5;
b = 7.5;
randomHeights = (b-a).*rand(1,1) + a;
	if(randomHeights>=6.75)
		% make random running vertical
% between 20 inches and 50 inches
%open interval a and b
		a = 20.0;
		b = 50.0;
		randomVert = (b-a).*rand(1,1) + a;
			if(randomVert>=44)
				% make random bench press
				% the least a player lifted is 0 and a
				% record is 30
				a = 0.0;
				b = 30.0;
				randomBench = (b-a).*rand(1,1) + a;
				if(randomBench>10)
					%make a random agility
						%2.91
						a = 9.0;
						b = 15.0;
						randomAgility = (b-a).*rand(1,1) + a;
						if(randomAgility<=10.3)
							%make a random sprint speed
								a = 2.8;
								b = 4.4
								randomSprint = (b-
a).*rand(1,1) + a;

								if(randomSprint<=3.05)
									done=1;
								end
```

end

end
end
end

end

years=players/60;

%% script to parse the data

```
#include <iostream>
#include <fstream>
#include <string>     // std::string, std::stoi
#include <vector>
#include <iomanip>      // std::setprecision

using namespace std;

//col 1 is height
//col 2 is weight
//col 3 is wingspan
//col 4 is reach
//col 5 is no step vert
//col 6 is no step vert reach
//col 7 is running vert
//col 8 is running vert reach
//col 9 is draft or not


string clean(string str)
{
    string str2 = "'";
    std::size_t found = str.find(str2);
    string feetStr = str.substr(0,found);
    double feet = stoi(feetStr);
    // cout << feet << endl;
```

```
    string str3 = "\"";
    std::size_t found2 = str.find(str3);

    string inchStr = str.substr(found+1,found2-2);
    double inches = stod(inchStr);
    inches = inches/12.0;
    // cout << inches << endl;

    double total=feet+inches;
    double scale = 0.01;  // i.e. round to nearest one-hundreth
    total = (int)(total / scale) * scale;

    string ret = to_string(total);
    return ret;

  }

int main() {


std::vector<string> myvector;
 ifstream myReadFile;
 myReadFile.open("runVertReach.txt");
 string file;
 if (myReadFile.is_open()) {
 while (!myReadFile.eof()) {


    myReadFile >> file;
    myvector.push_back(file);
 }
}
myReadFile.close();
int si = myvector.size();

cout << si;
ofstream myfile;
myfile.open ("runVertReachClean.txt");

for (std::vector<string>::iterator it = myvector.begin() ; it !=
myvector.end(); ++it)
  {
    // std::cout << *it << endl;
    string str = *it;
    string ret = clean(str);
    // cout << ret << endl;
    myfile << ret;
    myfile << endl;

  }
```

```
myfile.close();



return 0;
}
```

Works Cited

1. http://traintoball.com/what-odds-making-nba/
2. http://www.topendsports.com/testing/tests/agility-lane.htm
3.http://www.strath.ac.uk/aer/materials/4dataanalysisineducationalresearch/unit 4/correlationsdirectionandstrength/
4. http://www.brianmac.co.uk/economy.htm#ref
5. http://www.aafp.org/afp/2008/0901/p597.html