An initial analysis and exploration into the world of the National Basketball Association

September 27, 2018

Kenneth Reed
Michael Oristano
Fardi Yeung
John Michals

Core Message: Change is the only constant, especially in basketball

A brief yet involved data visualization project centered on the players of the National Basketball Association. This project focused on the players and some of the trends and changes that the league has undergone from inception in 1946 to 2017.

Summary:

While limited on time, our group focused on the National Basketball Association, its players and some of the trends that have developed, over time, as a result of shifts in strategy and development of coaching and playing strategies. In this analysis, the reader will find questions that were asked by the researchers that range from broad questions such as where and how does the NBA find its players to more basketball specific questions such as how did the implementation of a three point line implementation affect offensive strategy. This analysis used two specific third party csv files that contained NBA data on players, heights, weights and scoring statistics. Ideally, the data would have been retrieved directly from the National Basketball Association. However, budgetary requirements meant that data purchases were not an option for this analysis. Instead, these third party data sets, which were vetted and examined for accuracy, proved to be more useful. Additionally, APIs were utilized to collect a number of different data points early on in the project. These API requests collected the latitudes and longitudes of the colleges within our study.

Ultimately, the researchers in this group feel that we led an interesting and intelligent discussion surrounding the data. Not only were significant and relevant questions raised and answered, but also data cleanup and manipulation techniques, learned in coursework, were implemented to not only create usable data, but also to create better analysis. Overall the project yielded some interesting results and we do hope you enjoy learning about what it was we uncovered. Some of it was expected while some of it was quite surprising. We hope you enjoy.

Questions Posed and Answered

As previously stated, this analysis was performed on National Basketball Association data that dated from inception to present day, 1946 to 2017. Prior to any analytical work, the data from our csv files was examined for accuracy. Once completed, the following questions were established by the research team:

1. An examination of where NBA talent comes from and some analysis of those colleges
    a. Where do these future NBA players play college basketball?
    b. How has spending on those college programs developed over time?
    c. What does the revenue at the college level look like?
    d. What has happened over time to the average age of player entering the league?
    e. How does student aid compare to other expenses?
    f. How does expense compare with revenue for these programs?
    g. For the top college by college players, what does the head coach salary look like over time?

2. How have NBA players physically evolved themselves?
3. Does the frequency of NBA players from a given school have any relationship to the career length?
4. How has the introduction and strategical movement of the three point line shifted NBA strategy over time?
5. Is the NBA center as we know it dead? How has that position changed?
6. Who is the greatest player of all time?

The answers and analysis to each of those questions are presented in various subsections to follow.

Data Used

NBA data, two csv files:

(Both were accessed from the below website)

https://www.kaggle.com/drgilermo/nba-players-stats

US Department of Education Data: https://ope.ed.gov/athletics/#/datafile/list

- Data available from 2003-2017 in Excel. Excel from 2010-2017 was evenly normalized, so these spreadsheets were downloaded and save as .csv for the project.

Highlighting some of the Data Cleanup Techniques:

There were various points throughout the analysis that required different methodology and coding to "fix" data in an effort to make it usable. This section will expand on three of those techniques. Their code and explanation will be presented as proof of work. These three representations do not account the total amount of manipulation necessary. These are just some of the highlights.

Manipulation One: *Column Name change for the purpose of a merge*

In an attempt to merge the two csv files, the question of how to merge arose. In one set, there was data that covered players' colleges, heights, weights and positions. In another set, there was player statistical data. The decision was made to merge on name. However, one data set had the column named "name", while the other had the column named "Player".

To merge the data, we needed to manipulate one of the column headings to match the name of the other one. Once the renaming was completed, we merged the data on name and also completed a groupby on those names to collect all the entries for a specific player into one total name. Please see the below code:

```
#Renamed for the merge, a new data frame with name not player

bigmencleanformerge_df = bigmenclean_df.rename(columns={"Player": "name"})

notbigmencleanformerge_df = shortiesclean_df.rename(columns={"Player": "name"})
```

```
nba_merge = pd.merge(bigmencleanformerge_df, only_bigs, on="name", how="inner")

nba_otherposmerge = pd.merge(notbigmencleanformerge_df, other_than_bigs, on="name", how="inner")
```

Manipulation Two: *Manipulating the Height Column*

Our group was interested in examining player weight and height over the course of the project. While weight was not an issue since the values were integers, the height variable presented a dilemma for us. The data was listed in feet and inches and presented in a string, with a hyphen between the feet and inches. Several steps were undergone to treat the data. Firstly, we ran a delimiter on the column and separated the data into two separate columns, one for feet and one for inches. Secondly, we manipulated feet into inches and lastly, added the columns together to get a new variable "Total Height" that accurately presented the players' heights in inches. Additionally, a new column named "Career Length" was created. This took a players' last season, subtracted it from his initial season, and returned an integer value that could be analyzed. Please see the below code:

```
#NEXT FEW LINES OF CODE DROP NAs, CHANGE HEIGHT TO TOTAL INCHES, CREATE CAREER LENGTH COLUMN

nba_df = nba_df.dropna(how="any")

nba_df[["height","inches"]]=nba_df["height"].str.split('-',expand=True)

nba_df["height"] = pd.to_numeric(nba_df["height"])

nba_df["inches"] = pd.to_numeric(nba_df["inches"])

nba_df['height'] = nba_df['height'].apply(lambda x: x*12)

nba_df["total height"] = nba_df["height"] + nba_df["inches"]

nba_df["Career Length"]= nba_df["year_end"] - nba_df["year_start"]

nba_df = nba_df[["name", "Career Length", 'position', 'total height', 'weight', 'birth_date', 'college']]

nba_df.head()
```

Manipulation Three:

A third manipulation to highlight was when we used the pandas feature "unstack". This was used during the analysis of the three point line and its implications on basketball strategy. Unstacking allows the coder to pivot a level of the index labels, returning a data frame that has a new level of column labels.

```
Year_3PAvr = Seasons_Stats_DF[Seasons_Stats_DF['Year'] >= 1980][['Year','Player','3PAr']]

df_Year_3PAvr = Year_3PAvr.groupby(['Player', 'Year'])[['3PAr']].mean()

new_df = df_Year_3PAvr.unstack()
```

The first line of the code filtered all the applicable data by the year. Since the three point line was ushered in after 1980, the filter was done to return data from 1980 to present. We also wanted the columns associated with that filter. The second line groups the data by player and year while simultaneously calculating the three point average mean. Lastly, the "cool" part of the code was the unstack method that pulls the year index to become the column head.

Manipulation Four:

A fourth manipulation was used to pull in the data from the U. S. Department of Education (Equity in Athletics). There are 4200+ columns with data for over 2000 institutions. The columns had all sports information broken up in various categories. There was a data key in a separate Word document that described all the column names, so only basketball-specific information was used.

```
# Read the Budget Data From 2008-2017; reduce 4200+ columns to 17

budget2010_data_df = pd.read_csv(budgets_2010_data_to_load, encoding="UTF-8",

usecols=[1, 4, 8, 10, 15, 26, 50, 143, 152, 158, 168, 316, 390, 427, 945, 3609, 3942])
```

Additionally, to match with the player data set some clean-up was necessary on institution name, especially since the DOE data was more formalized (e.g., University of North Carolina at Chapel Hill instead of UNC).

Background of the NBA

The National Basketball Association (NBA) is a men's professional basketball league in North America; composed of 30 teams (29 in the United States and 1 in Canada). It is widely considered to be the premier men's professional basketball league in the world. The NBA is an active member of USA Basketball (USAB), which is recognized by FIBA (also known as the International Basketball Federation) as the national governing body for basketball in the United States. The NBA is one of the four major professional sports leagues in the United States and Canada. NBA players are the world's best paid athletes by average annual salary per player.

The league was founded in New York City on June 6, 1946, as the Basketball Association of America (BAA). The league adopted the name National Basketball Association on August 3, 1949, after merging with the competing National Basketball League (NBL).

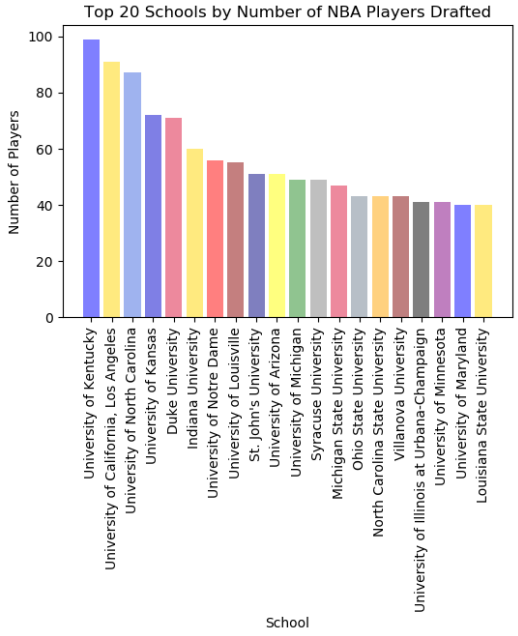
Data Analysis

Question 1 – An examination into the annual incoming NBA talent and the schools from where they hail

We began our analysis by first asking, where do NBA players actually come from? Any professional sporting league needs to have the best possible

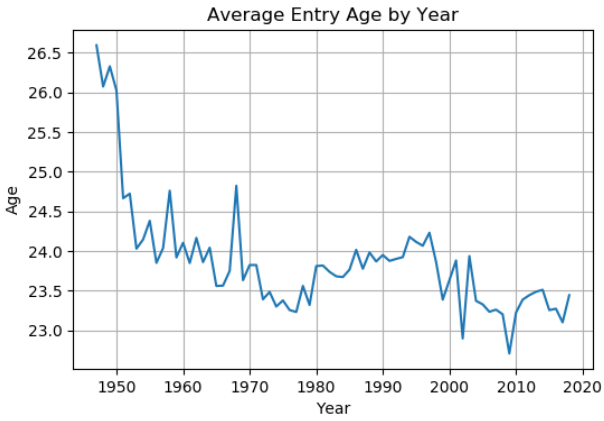
athletes involved or the final product suffers. Basketball is no exception. Our analysis of the NBA, therefore, had to start not at the NBA level, but at the collegiate level. However, our data sets were all on professional basketball players, not college amateurs. Luckily, we had the players' colleges and could construct some very interesting analysis on not only where players attended but also how many NBA players hailed from a given school.

After uploading the csv files to our workbooks, the group grouped players by counts of players from each college. We then took the top twenty-five schools and finally, through the utilization of a Google maps API, we plotted the count of players from schools by the school locations, latitudes and longitudes, on a map of the continental United States. The heat map, above, is what was developed. The map shows that while there is a few schools in the south west that produce NBA talent, overwhelmingly, the "best" schools for creating NBA talent are in the east and the greatest concentration of them is North Carolina. This makes some sense given that within the state of North Carolina, there are at least three major basketball programs within a short drive of each other: Duke University, North Carolina State University and the University of North Carolina at Chapel Hill.



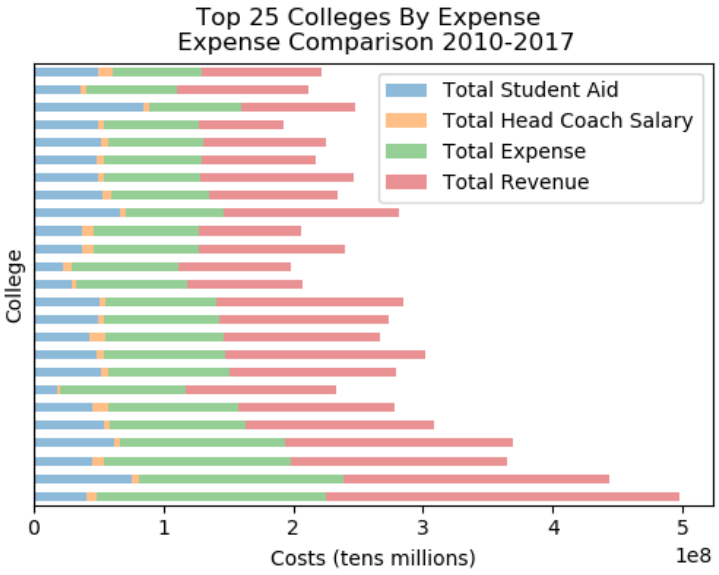
In addition to the heat map above, we thought it would be interesting to examine where most of the NBA talent play their college basketball. We grouped the player data by college attended and then performed a count function on the result. We then looked at the top twenty schools and saw some interesting results. On the left is a color coded graph of the results. The top twenty ranking schools for NBA talent really come as no surprise as it is generally large, traditional powerhouse basketball programs. These programs have always and continue to recruit well. Mixed in the results were a few names that while large schools, were somewhat surprising. The University of Minnesota was one that fit this surprising mold.

The next question that followed was of the players in the NBA, what age were they when they entered? Furthermore, with the increases in player salaries over time in the NBA, had that age of entry into the league been decreasing with time? Unfortunately, we didn't have salary data, but we did have a way to calculate entry age. To calculate this player age variable, we took the player birth dates, subtracted from their league entry dates and calculated an age as to when they became professional players.



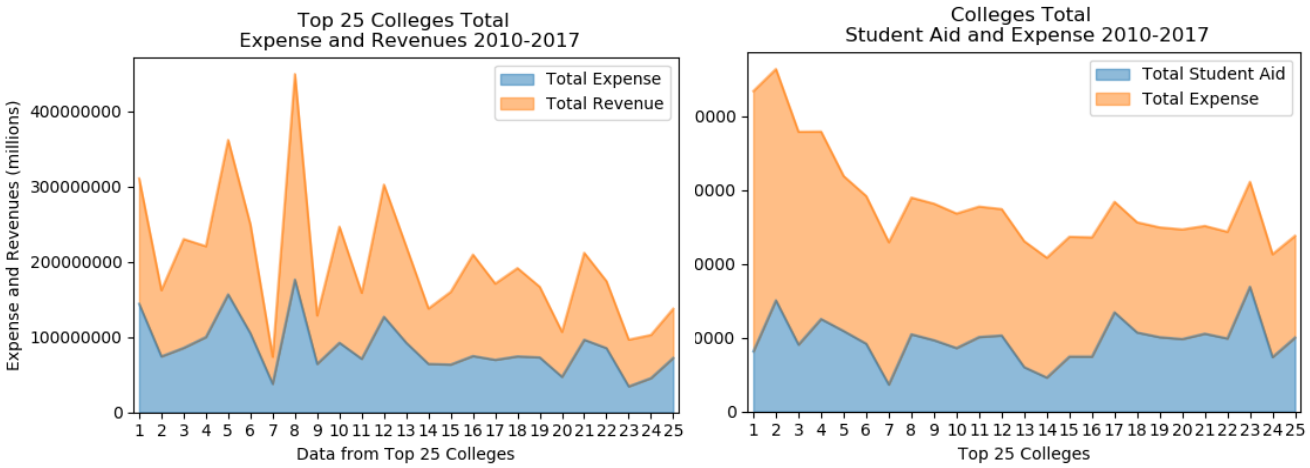
What we found was that "yes", in fact it had. As we can see from the plot on the right, when the NBA first formed, the

average player entering the league was roughly 26.5 years old. An old man by today's NBA which has an average entry age of approximately 23.5 years old. A few things accounted for this change over time. Firstly, when the league officially began, the year was 1946. Given that World War II had just concluded, most young men that would have been of professional basketball age were off serving the nation. Given the casualty rates of that war mixed with the average age of enlisted of roughly 23.5 years, it becomes very plausible to consider that by the time the NBA formed, "young men" were in fact in their middle twenties. From the graph, it is observable that within ten years, the average age fell dramatically to roughly 24 years of age. Normalization of daily life set in and as young boys opted for college and not the military, the eligible age they would be when it was time for professional sports decreased. The graph stays somewhat flat, rising slightly up until 2000 when suddenly, entry age dips sharply to 23.0 years old. It has since oscillated between 23 and 23.5 years old and most likely will for years to come.

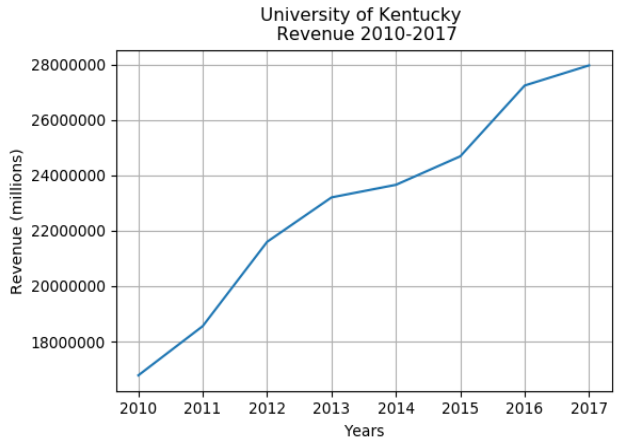


That is because the NBA instituted a "one and done" rule where-by to qualify for the NBA, a prospective player has to be nineteen years old. This has produced a phenomenon known as "one and done", which means one year of college basketball, then right to the professional ranks. For context, during the early 2000s many eligible basketball players were opting to bypass college and go straight to the professional ranks. This was in an attempt to rise to the professional ranks quicker and reap the benefits of the large NBA contracts. What began to happen though was a reduction of talent at the collegiate level as well as a number of players who washed out with little option after failing to stay at the professional level. The NBA had a moral dilemma, mixed with NCAA college resentment toward the NBA for "stealing" its talent. The executives at the NBA attempted to rectify the situation by implementing a "one and done" policy. We observe that graphically with the rebounding of the age from below 23 to where it is currently.

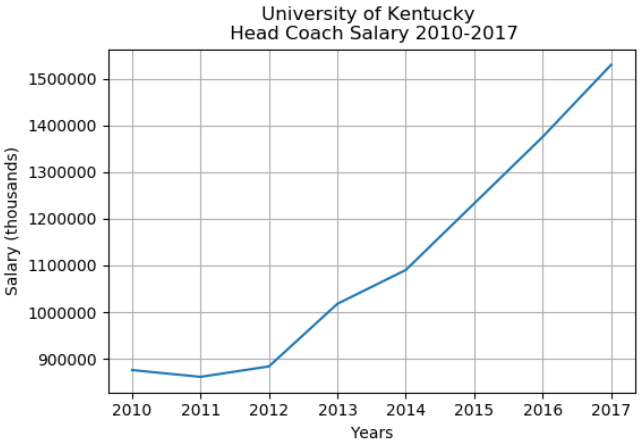
The third portion of the examination brought us to the question of how these top basketball



institutions spend their money. Specifically, we wanted to evaluate the various expenses and budgets that these colleges have when it comes to basketball. This data was found on the US Department of Education website. We were able to obtain the data for 2010 through 2017. Once we had the data imported, we took the top twenty-five schools by total expense and revenue to perform our comparisons. We also analyzed the head coaching salary for the top school; specifically how this has evolved over time. We plotted the data by utilizing a stacked horizontal bar graph as well as some various line graphs. The stacked bar graph is to the left and the line charts showing student aid and expenses as well as expenses and revenues are just below. The results are not necessarily surprising, but they do show that at the heart of it, these are real businesses. The term "student-athlete", when you observe that all of these top twenty-five programs are e profitable, gets very opaque.



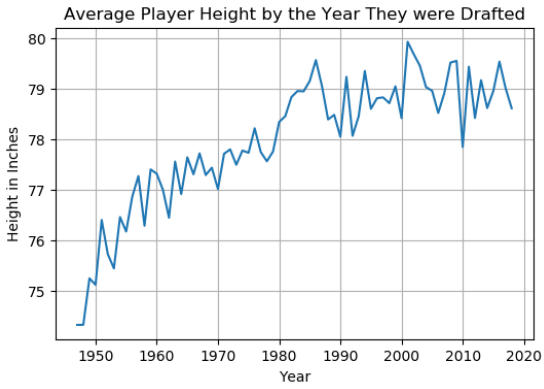
The last of the first examination questions brought us to the age old question, "How much are you making, Coach?" Deion Sanders, a former two sport professional athlete, had a very famous, and very telling quote during his playing career. When asked a question about compensation, Deion replied to the reporter "If you look good, you feel good. If you feel good, you play good. If you play good, they PAY GOOD!" While Deion doesn't coach basketball, or any college sports, the sentiment transcends as we observed that from 2010 to 2017, our top college coach has seen his salary increase, each and every year. As if generating $900,000 per year could ever be considered baseline, the coach of the University of Kentucky has seen a quite linear relationship between his tenure and his salary. Also of interest was just how this salary compared with total revenue. After all, there should be a relationship. An executive's salary shouldn't continue to grow if the revenue is stagnant or declining. In fact, it is as we expected. Our coach's salary has grown as revenues have continued to grow.

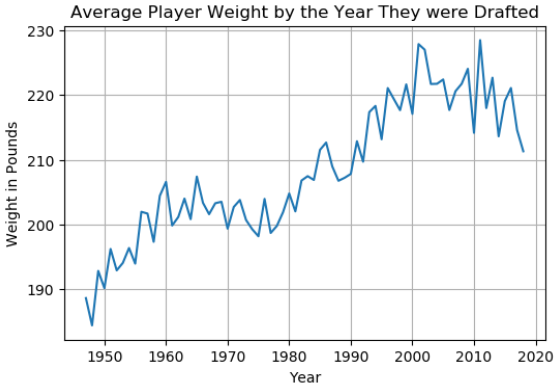


Question 2 – The evolution of the NBA player

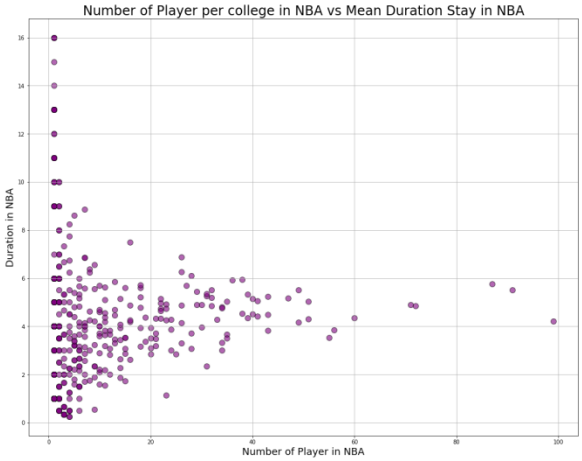
After our analysis of the college backgrounds of NBA players, it was time to examine the players themselves. How has the typical ballplayer changed over time? As human beings, in first world countries, we have gotten somewhat used to the concept of growing bigger and stronger over time. What about NBA players? Typically, NBA players have always been the tallest and the most athletic. Will the data show us that result? Turns out, "yes" it will.

To analyze player physicality, we had height and weight variables to use. As mentioned in the data manipulation, weight was straight forward yet height took some work to bend the variable into compliance. Once we were able to get workable columns of data, we grouped the data on the players starting year, in order to calculate an average weight per draft class. The same technique was used for height. Once completed, it was time to graph. Both graphs below show results that appear to trend upwards ad a similar slope. Height and weight are fairly well correlated so it does make intuitive sense that as a player grows in inches, he too will grow in pounds. If only the reverse was true.





Question 3 – Examining the potential relationship between current NBA players, the colleges they attend, and if there is a relationship between the two as measured by career length



Through analysis of the college to professional data, it became apparent that there are definitely some colleges that attract high quality and high caliber basketball players. That statement can be made based upon the numbers of players that successfully make it to the professional ranks. But does attending a given school "guarantee" a prospective professional player any pathway to the professional ranks? That is the question we really wanted to answer and our initial thought was, absolutely it does.

Good, successful college athletic programs have one major commonality, quality athletes. It was felt that because successful programs recruit successful athletes, that it would be logical that more of those successful athletes would make it to the professional ranks. Therefore, if for example, you were a

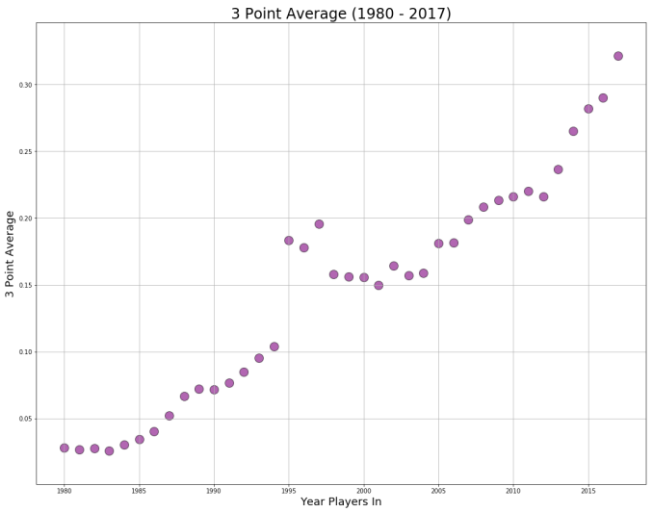
high school basketball player, and you were recruited to play at the University of Kentucky, you would have an edge in terms of successfully making it to the NBA and having a lengthy career, since Kentucky has sent a large number of players to the NBA.

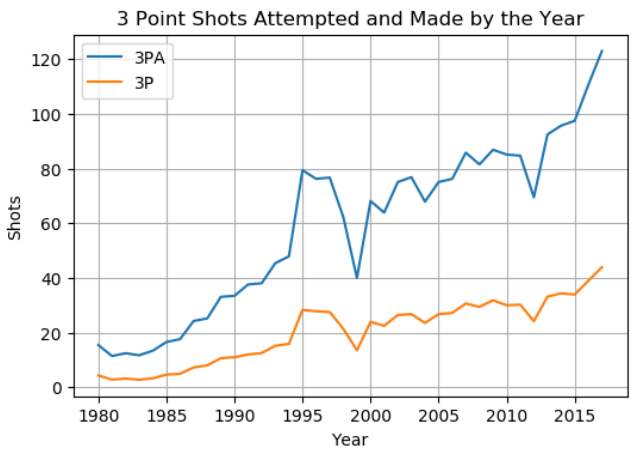
To measure this, we took the csv data and extracted the number of professional players per college and the duration of those players' professional careers. We then placed the data on a scatter plot, above, and to our amazement, found relatively no meaningful relationship. The plot suggests that there is a definite average career length in the NBA, but what college a player attends has little to no bearing on that. The school that sends ninety-nine players to the NBA, Kentucky, has the same player average career length as schools that send one or two. In fact, schools that send one or two players seem to have more cases where players are enjoying quite lengthy careers. While we were surprised to find no relationship, the latter part did make intuitive sense to us given that if you are the single player from a given college to make it to the professional ranks, you must be extremely talented. Thus, that same extreme talent could propel you to have tremendous success in the NBA. An example would be Dwayne Wade who hails from Marquette, a good basketball school but certainly not a modern day powerhouse. While Wade is one of the few Golden Eagles to successfully rise to the NBA, he is one of the best players in the league to this day and is about to enter his 15th season in the league. A major accomplishment especially when compared to the league career length average nearly 4.8 years.

Additionally, the heat map, on the previous page, displays the average career length of NBA players from a given school. The intensity corresponds to the frequency of players from that given school. We see quite a bit of color on the map with the east well populated and "hot" with activity.

Question 4 – Examining the evolution of the three point shot in the NBA.

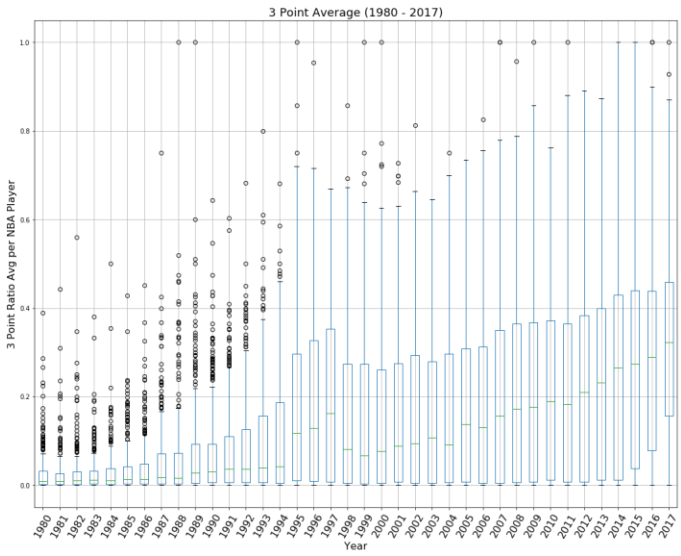
This was a particularly interesting question because it dealt directly with analyzing how offensive strategy changed with the introduction of a new tool to utilize. Prior to the three point line, every shot on the court, no matter the proximity to the basket, was worth the same, two points. The lone exception to this was the free throw, which was and still is worth one point per successful make. To recall, in 1980 in an attempt to boost excitement, the NBA ushered in a new concept, rewarding an extra point to shooters that successfully made baskets from a distance of 23 feet 9 inches around the basket. This distance remained the official distance until the 1994-1995 season when for a two





year period, the NBA rules committee shifted the three point line up, to 22 feet around the entire basket. Following the conclusion of the 1997 season, the league moved the three point line back 21 inches to its original distance. It remains at 23 feet 9 inches to this day.

To analyze the question, we developed three graphs. The first step of our analysis was to take the imported csv data and filter it by three-point scoring for each player. A mean was then calculated. Of important note, since three-point data was only available for players who played between 1980 and 2017, all of the data prior to 1980 was unusable for this analysis. Additional manipulation was undertaken to rearrange the average score on a yearly basis using the python data frame method unstack. This was done after the filtered data was grouped by year and player. We were left with two columns of data, the year and the three point scoring. The following plots correspond with the findings. Let's talk about the plots. In general, we can see from the scatter plot that from inception to 1995, the three point average increases. The relationship seems quite linear in fact. This is somewhat expected since in the beginning, players would not necessarily be proficient in attempting the three point shot since prior to 1980, that same shot would be worth two points only. What is particularly interesting is that we see 1995, the averages spike up almost nine percentage points. As previously discussed, in 1995, the three point line was moved closer. You can deduce what would happen then, a spike in proficiency. Players, who were building proficiency in shooting this long shot, were now being given a gift, a twenty-one inch closer shot for the same value. Following the 1997 season, when the league moved the line back to its original location, we see a drop in the average. Just as the span from 1980 to 1995, the averages have continued to increase, suggesting that players are becoming extremely proficient at draining this long shot. Perhaps it is time to shift the line, this time backwards, to help keep players honest and in an effort to manage the scoring? That is for the league to decide.



Now average is a fair variable to use, but an average can be manipulated by either becoming more efficient as a shooter, making more shots, or becoming more selective as a shooter, taking less attempts. To determine if what we saw from the average graph was actually more efficient three-point shooting, we graphed the three point attempts as well as the successful three point shots. The trend line for both makes and attempts points higher over time, with attempts spiking as the three point line moved closer and then dipping back down after the line was moved back. What is interesting from the graph is that there is a steep incline in attempts after 2015. While we understand what happened in the graph between 1995 and 2000, why are attempts and makes decoupling somewhat after 2015? We believe the answer lies in the more statistical approach teams are taking in terms of strategy. Beginning in 2015, teams began to recognize that it was more efficient

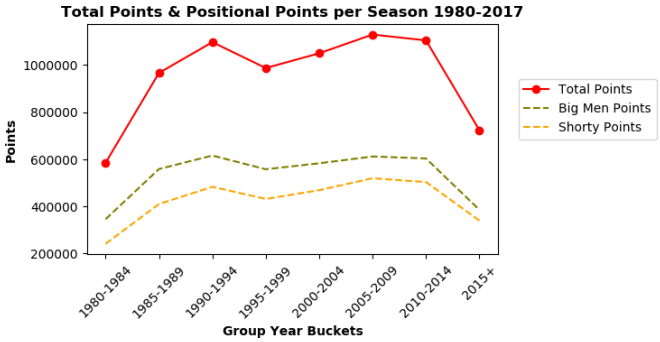
to shoot a higher frequency of three pointers than it was trying to get two pointers. No team personifies this concept more than the Golden State Warriors who are consistently atop the league in attempts. They also have won three of the last four NBA championships.

The box plot on the right is an interesting look at how the three point shot has evolved over time. As we noted, the three point average has grown and increased nearly every year since 1980. What we see on the box plot is that especially over the last 5 years, the mean of the league has really increased and the bottom quartile of shooters are actually much better shooters than the top quartile from fifteen years prior. Furthermore, we see a massive amount of outliers during the development of the three point line, when compared to today's almost non-existent outlier. Over time, the pure amount of NBA shooters attempting three pointers has grown, so this has some direct implications to the reduction of outliers over time.

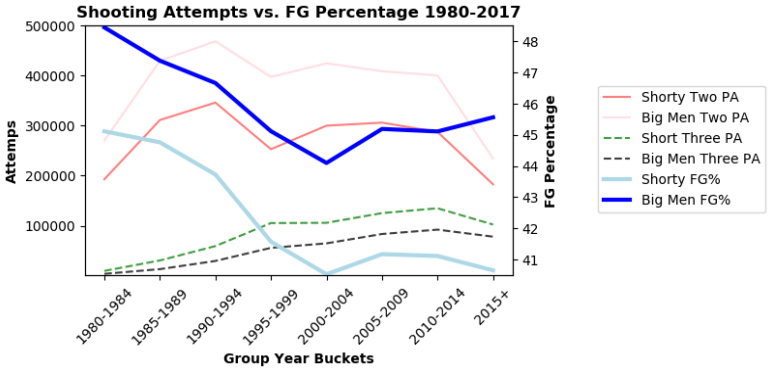
Question 5 – Is the center position dead?

For all my friends in the media who like quotes, mark this quote down. From this day on I'd like to be known as 'The Big Aristotle' because Aristotle once said, 'Excellence is not a singular act; it's a habit. You are what you repeatedly do.' – Shaquille O'Neal

-Shaquille O'Neal, arguably one of the greatest centers to play the game of basketball, declared that at the end of his career, the position he played, Center, was "dead'. We examined whether or not Shaq was accurate with his statement. We concluded that while Shaq was a dominate center in the league, his forecasting of the life of his position was off.

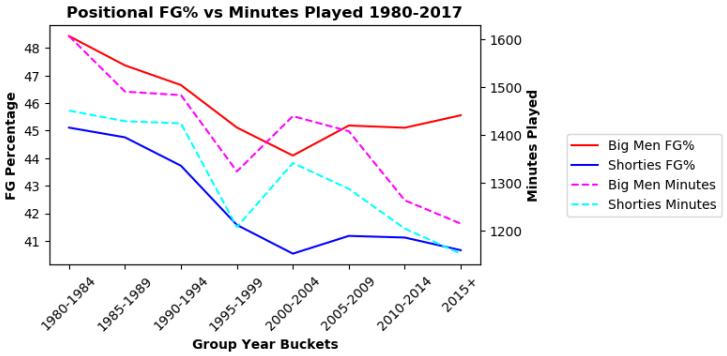


How did we start? The first step was to create usable, relevant variables from our raw data. This was done by parsing the dataset by position played. In the NBA, players who typically participate in the game at the center position are the tallest figures on the court, but often times, there are multiple tall players so positions can and often do rotate. Conventional basketball wisdom places the tallest players closest to the basket with the shorter players further away. However, sometimes Centers play Forward, Forwards play Centers. Given the overlap, we broke the data into two categories, those likely to play center, and those unlikely to play center. The variable names were "Big Men" and "Shorty" respectively.

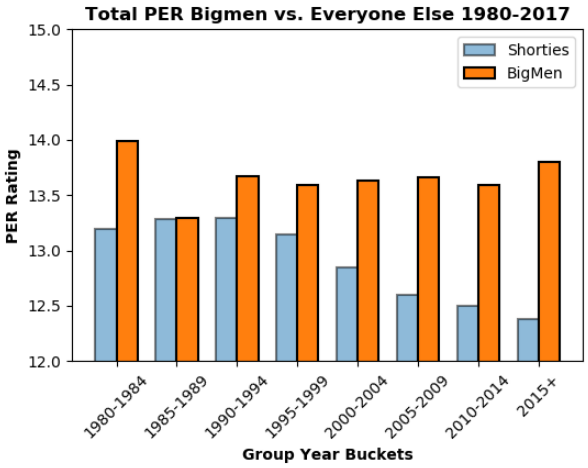


Now that the variables were created, the first question was, "Who scores more?" As the above graph indicates, in general Big Men score more than Short players. Why is this? For starters, as conventional basketball wisdom goes, taller players are closer to the basket, therefore, it is certainly reasonable to deduce that the opportunities to score are greatest for players that are closer in proximity to the basket. Note that while the "2015+" data point drops off from the "2010-2014" data point, it is because the binning of the data is measuring two points, one with five years of aggregated data and one with only three years of data.

So what about this question of attempts? Given that Big Men are still outscoring their shorter counterparts, are the opportunities they are getting shrinking? Again, Shaq is wrong, big men aren't dying, they are flourishing! The graph above tells a few differnet tales. Firstly, for reasons already stated, Big Men traditionally shoot less three pointers. That is measured in the dashed lines indicating attempts for shorties being greater than that of big men. However, it should be noted that both shorty and big men variables have postive slopes. Secondly, big men two point attempts are, historically, higher than that of their shorter counterparts. Thirdly, and most interestingly, is that Big Men are much more efficient than their shorter counterparts. When measured, the shorty FG% is downward sloping, completely opposite the case of the Big Men FG%. While this makes some intuitive sense in terms of big men taking more high percentage shots to begin with, what else could lead to this uptick in FG%?



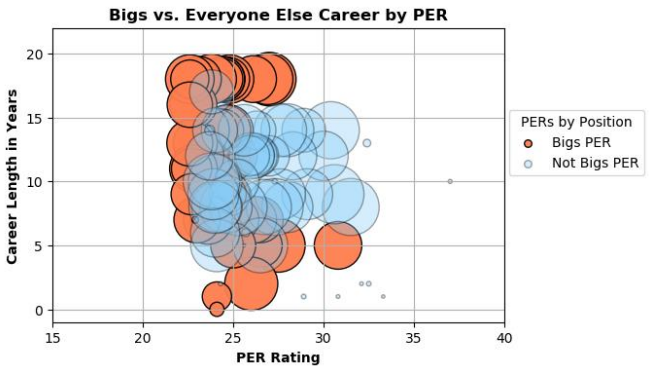
The answer we arrived at was minutes played. The graph above shows negative correlation between Field goal percentage and minutes played. Why? Simple answer: like with most things in life, when you are tired and stressed, you tend to underperform. While this seems somewhat logical, looking back to 1980, we see that players in general were logging a tremendous amount of minutes. Those minutes began to go down, and then around the time of the 1995 NBA expansion, minutes expanded again. Then around 2004 and trending down, before and after the departure of Oneal, minutes trend down and Field goal percentage rises.

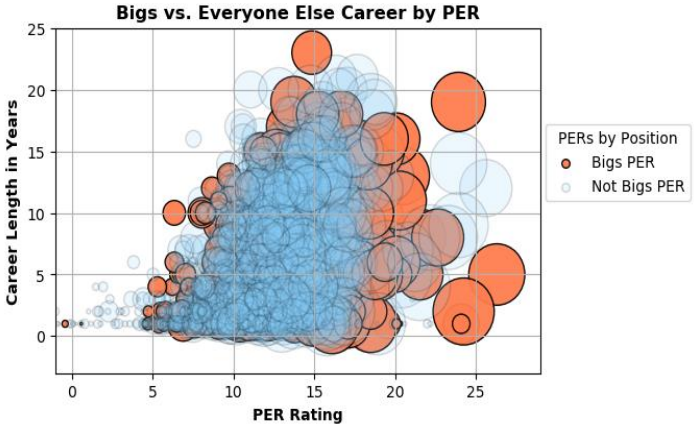


Lastly, to put the proverbial nail in Shaq's coffin, we wanted to look at the variable PER. PER is a player efficieny rating that is calculated by the NBA in an attempt to answer the quesiton of who is the greatest of all time. The equation, often critisized for its offensive data skew, is a proprietary blend of offensive categories. What we see when we analyze the data is that in general, Big Men PER stays constant over time. The same cannot be said for the shorties.

The following graphs further represent this PER variable. The bubble plots are particularly interesting because they show some trends. The top plot shows a bubble plot of the top 50 PERs for Big Men versus Shorter players. The below plot shows the entire sample size of PERs, Big men versus Shorter players. In general,

shorter players, on both charts, tend to gravitate around the center of the graph while the taller players measured by Centers tend to dominate the exterior of the chart. What this basically says is that Big Men are eccentric. They don't always play for long, but on a PER scale, they tend to have extremely efficient and productive seasons. Furthermore, shorter players, the point guards and shooting guards, tend to play roughly the same length of time and tend to be fairly productive. The career lengths for big men varies greatly. This can be somewhat expected given their sizes and the potential health complicaitons that come with both being that large and playing such a physical game.



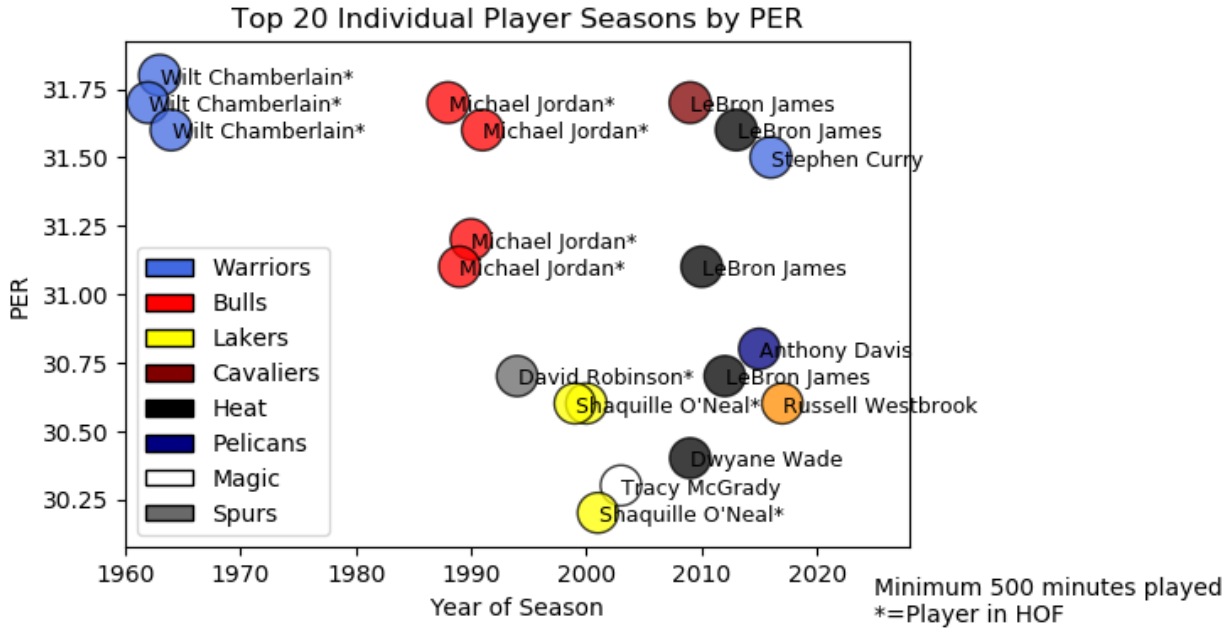


Question 6: Who is the GOAT (Greatest of All Time?)

The last question, and to basketball junkies, the most relevant of all our questions, can you determine who the greatest player of all time is? As already explained, the PER statistic is the NBA statisticans answer to how to solve this riddle. However, PER will not yield a name, it yields a season. The PER is a rating on how a player has performed over a given season, therefore, it is more appropriate when answering the question of who is the greatest, by instead answering, who has had the greatest season of all time?

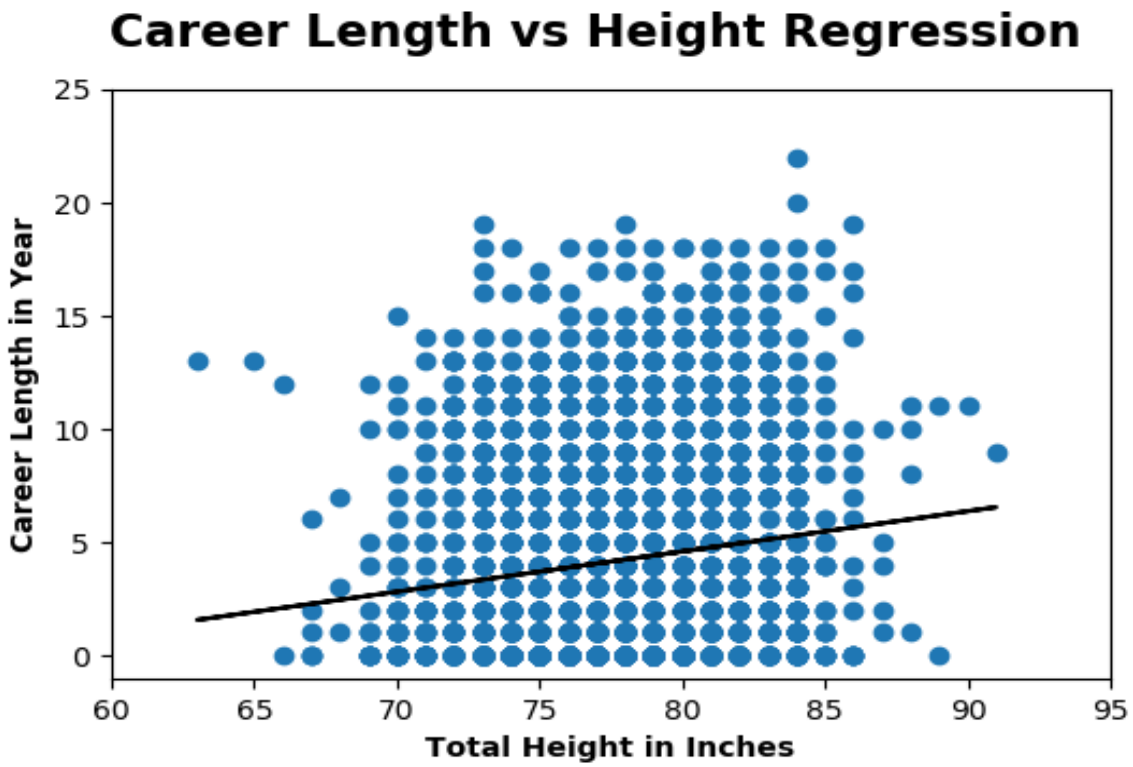
The best way we felt to present this was with a scatter plot. The data from the season stats was grouped by player and by season. We then set a minimum of 500 minutes played to remove any outliers. We then plotted the PER statistic by the season year. What resulted was a pretty interesting graph of players. Some are well known, some are obscure. One thing that jumped out at us was the amount of data points that came after 1985. Nearly all the points were plotted for players who had high PER seasons after 1985. We believe this is more evidence of the three point shot and its impact on the game. Another interesting and surprising result was that of all the players that were listed, Kobe Bryant,

often regarded as one of the top ten greatest players, never had a season with a PER high enough to crack the top twenty.



Conclusion & The Next Steps

Ideally, the next steps in data would be two fold. On the one hand, we have a solid analytical foundation to now build from. All the questions posed were quality quesitons that could be expanded upon. Akin to chasing the rabbit down the hole, we have more to dig through to attempt to develop



more realistic and significant relationships.

The second part of the two fold next step would be new analysis. We could, with time, find more available data on salaries, for example. Below, we performed a simple regression on height versus playing career. What we found was a slightly linear relationship between having a longer expected career based on being taller. This simple regression was easliy performed, but much more analysis could be attempted in order to further develop statistical relationships between dependent and independent variables. Obviously, there is more that goes into having an NBA career than being tall, and had we had more time to analyze and build out a regression model, we could assess different factors that impact career length. For instance, maybe a players socio economic background could have been incorporated to measure what a player's upbringing meant as far as his odds of not only making it to the NBA but also thriving. Obviously we would need new data for this but again, if time allowed, perhaps this would be plausible.

In the end, the analysis performed here, in our humble opinion, was accurate and true. We were able to test and answer the questions that arose and came away from the project with some confirmations and some new discoveries. While more is often better, for now, this will suffice.