COMP 430 Project

David Gan, dwg2
Judson Webb, njw6
Kathy Yu, ky17

Abstract

Assuming that the three-point line is moved back by a non-trivial amount (~10 feet), we found that the San Antonio Spurs and taller, heavier players would benefit the most, while the Sacramento Kings would be hurt the most. Moving back the three point line hurts players proportionately to how many three pointers they attempt and make, thus players that do not shoot any three pointers are not hampered in any way, and are the most likely to benefit. It follows that teams in which a high percentage of their players attempt a low number of threes per game will benefit the most. The Spurs have very few players that attempt more than a couple of threes a game, while the Kings have several players that shoot more than three, three pointers a game.

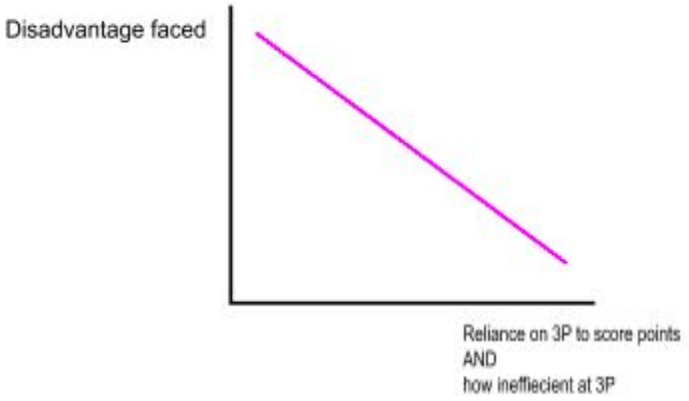
Reasoning

First, we will make the assumption that players will not just stop shooting three-pointers because of the rule change.

The consequences of moving the three-point line back are that three-pointers will become harder to make. This means that

a) The *percentage of successful three-pointers* for players and teams should go down
b) The *number of three-pointers made per game* for players and teams should go down

Given how the rule change just makes it harder for players to score in a certain fashion, we believe that players who take zero three pointers a game will have the biggest advantage, because they are not being disadvantaged in any way. Meanwhile the more three pointers a player takes the more disadvantaged, or hurt, they will be by the change.



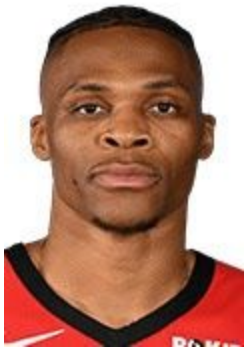
We believe it is beneficial to think about the effects in terms of how much the rule change disadvantages a certain player. This means that there will be an inverse relationship between how crucial a player's or team's reliance on three-pointers to score points is and how much they will be disadvantaged, and this disadvantage is only compounded with how inefficient/poor of a three-point shooter a player is. Meaning that efficient three-point shooters (have a very high 3P%) are less disadvantaged than inefficient three-point shooters (have a very low 3P%).

We thought it was necessary to include the caveat that efficient and inefficient three point shooters would be impacted differently, because watching the NBA today, lots of great three-point shooters are already shooting well behind the three-point line. It has become fashionable for players like Steph Curry, Trae Young, and Damien Lillard to shoot from well near mid-court. Thus these players are more capable of continuing to make a higher percentage of their three-pointers and face less of a disadvantage than inefficient three-point shooters.

For ease, we will take three players as examples and look at how they differ in the crucial three points stats of three-point-attempts-per-game (3PAG) and three-point-percentage (3P%).

Rudy Gobert
3PAG: 0.0
3P%: null
Least Disadvantaged

Steph Curry
3PAG: 11.7
3P%: 43.7%
Disadvantaged



Russell Westbrook
3PAG: 5.6
3P%: 29.0%
Most Disadvantaged

Reference: all data from Basketball-reference for the 2019-2019 season
https://www.basketball-reference.com/leagues/NBA_2019_per_game.html

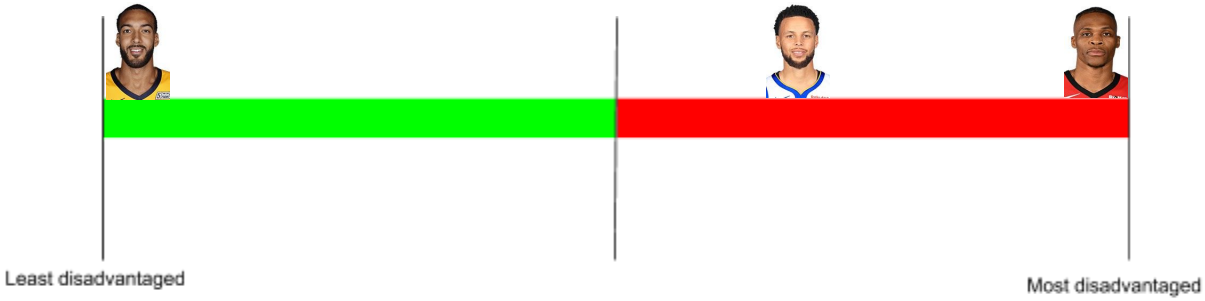
These players represent three different groups that we thought would be noticeably impacted (or not so) by moving the three-point line back.

First, Rudy Gobert is a tall Center that usually plays near the basket, and he represents players that do not take any three-pointers a game. Since his reliance on three-pointers to score points is 0, he is not negatively affected by the rule change and is thus the least disadvantaged (benefits the most from the rule change)

Russell Westbrook relies heavily on three-pointers, attempting a lot of them per game, but he is also widely inefficient from behind the arc, having the lowest 3P% in the league at 29%. He represents players that have a high reliance on threes to score points and are also inefficient at making three-pointers. Thus he will face the biggest disadvantage.

Finally, we have Steph Curry who is even more widely reliant on three-pointers but is extremely efficient from behind the arc (He has the fourth-highest 3P% in the league). He represents all of these types of players, and we believe will be the second-most-disadvantaged, as their efficiency in making three-pointers will drop, but likely not as much as the Russell Westbooks.

If one imagines a spectrum from “least disadvantaged” to “most disadvantaged” the following players would fall on it as such.



Data Input

We did run into some complications when trying to load in the original data the first time, mainly stemming from us assuming the data types for certain columns. For example, we thought that some number based columns would be integers--and thus tried to load them in as integers--only to find out that they had been designated as

character varying data types. Also, we thought percentage values (i.e. three-point percentage) would be between 0 and 100, but they were actually between 0 and 1.

Once we delved more in-depth in the SQL files provided and found exactly what data type each column was, we loaded the data without any roadblocks.

Data Transformation

Before we could do any sort of analysis we needed to cleanse our data. Given that our goal was in trying to predict what would happen if the NBA moved back the three-point line specifically, we thought a good starting point would be in looking at the various three-point statistics.

Knowing that having even a single NULL value could lead to incorrect calculations, we looked through 3P%, 3PA (attempts), 3P (made), and 3PAR looking for NULL values. We found that some players had NULL values for every one of these columns, and we found the reason to be that most of them played before 1979 before the three-point line was introduced to the NBA. Since these player's had NULL three-point data, they would not help us answer the questions (and could possibly lead to errors in our future calculations) so we decided to remove these tuples in stats in which all of these values were null.

Question 2 asks about how teams would be impacted by the rule change, so we also decided to look at the "team" column and see if there were any NULL values there. Once again we found NULL values for some players and since these rows would not contribute to our calculations we removed them.

Next, we had to do some updates to our data in order to standardize similar columns across the different tables we had as input. We noticed that in some of our tables a player's height was in feet and inches but in others, it was in centimeters. Similarly, in some tables, a player's weight was in pounds, and in others, it was in kilograms. So, we converted the "height" and "weight" columns in player_detail from feet and inches to centimeters and from pounds to kilograms, respectively.

Finally, since basketball teams, and the game itself, can change so much over a short period of time, we thought it would not be useful to take data from the 1970s in determining what team would be most benefited by moving the three point line back. Rather we think our analysis is more relevant if it were just looking at recent data

because that is where the implication of our findings can make a difference to how a team is composed and plays. For these reasons we only took data that came from a year later than 2012.

Created Attributes

We created three new attributes for our analysis. The first one is a statistic that already exists but was not in our data set: three-point attempts per game. We needed this statistic to better aggregate and find averages among teams since players could play in a different number of games a year due to injuries.

Another attribute we created was called “delta_percent”. This was just a player’s two-point percentage (2P%) subtracted by their three-point percentage (3P%). Our rationale for creating this attribute to find players that were efficient shooters that were less hampered by distance (and thus finding the players that would fall into the “Steph Curry” category instead of the “Russell Westbrook” category from the “Reasoning” section”). Great shooters, that are not as impacted by how far away they shoot from the basket, will have very similar 2P% and 3P%, and there is greater variance among bad shooters between these two statistics. Once again lets look at basketball-reference statistics from the 2018-2019 season for an example.

If you sort by highest 3P% you get the following players in the top 4. A player is considered to be a “good” or “efficient” shooter if their 2P% is >= 45% and their 3P% >= 40%. Clearly all four of these players would be considered not only “good” shooters but “great” shooters, and their percentages for these two stats are relatively similar, differing by as little as 1.6% in the case of Steph’s little brother Seth Curry.

Rk	Player	Pos	Age	Tm	G	GS	MP	FG	FGA	FG%	3P	3PA	3P% ▼	2P	2PA	2P%
1	Joe Harris	SG	27	BRK	76	76	30.2	4.9	9.8	.500	2.4	5.1	.474	2.5	4.8	.528
2	Danny Green	SG	31	TOR	80	80	27.7	3.7	7.9	.465	2.5	5.4	.455	1.2	2.4	.487
3	Seth Curry	SG	28	POR	74	2	18.9	2.9	6.3	.456	1.5	3.4	.450	1.3	2.9	.463
4	Stephen Curry	PG	30	GSW	69	69	33.8	9.2	19.4	.472	5.1	11.7	.437	4.0	7.7	.525

https://www.basketball-reference.com/leagues/NBA_2019_per_game.html

Now we can look at the players with the lowest 3P% in the league that actually attempt three-pointers.

Rk	Player	Pos	Age	Tm	G	GS	MP	FG	FGA	FG%	3P	3PA	3P% ▲	2P	2PA	2P%
1	Russell Westbrook	PG	30	OKC	73	73	36.0	8.6	20.2	.428	1.6	5.6	.290	7.0	14.5	.481
2	Kyle Kuzma	PF	23	LAL	70	68	33.1	7.1	15.5	.456	1.8	6.0	.303	5.3	9.5	.553
3	Nikola Jokić	C	23	DEN	80	80	31.3	7.7	15.1	.511	1.0	3.4	.307	6.7	11.7	.569
4	Austin Rivers	SG	26	TOT	76	15	26.7	3.1	7.5	.406	1.4	4.3	.318	1.7	3.2	.522

We can see that not only is their 3P% much lower (by 18% from Joe Harris to Russell Westbrook), but also the difference of their 2P% - 3P%.

What we took from this is that some “great” shooters are severely less affected by how far they shoot from the basket, in comparison to “bad” shooters. Thus, these “great” shooters might not actually face as severe of a disadvantage if the three-point line were to be moved back. So we calculated “player_percent” to be 2P% - 3P%, and used it to seperate players who shot a lot of three pointers (more than 5 a game) into either a “good shooter” (Steph Curry) group or a “bad shooter” (Russell Westbrook) group.

Finally, our last attribute was “player_score”. This was just a discrete variable from 1 to 4 that we would calculate that told how disadvantaged the rule change would be for that given player, with 4 representing “not disadvantaged” (i.e. Rudy Gobert), 3 representing “slightly disadvantaged”, 2 representing “more disadvantaged” (i.e. Steph Curry), and 1 representing “most disadvantaged” (i.e. Russell Westbrook).

Analysis

Our analysis and judgement of which players (and thus teams) were most affected was based on the ideas that

A) Most importantly, players who take less three-point attempts (3PA) per game face less of a disadvantage.
B) Secondarily, players with smaller deltas between their 2P% and 3P% were less disadvantaged than players with high deltas

We reasoned this because for players that do not take any 3PA, moving the three point line will not affect the way they score or their efficiency in scoring. Also, players with smaller deltas closer to 0% correlated to players that could shoot at distance with great efficiency and would be hindered less.

So we created 3 groups of players, those who did not attempt many 3-Pointers (<1.5 3PA per game), those who were in the middle range (between 1.5 and 3, 3PA per game), and those who attempted many 3-Pointers (>3, 3PA per game). We further segmented those who attempted many 3-Pointers by their delta values (2P% - 3P%), ending with a total of four groups.

These four groups represent segments of the population (that we reasoned about earlier) who would be impacted differently by the rule change. We Assigned each player's "player_score" column depending on which group he fell into as such

4. the low 3PA players (not disadvantaged)
3. mid 3PA players (slightly disadvantaged)
2. high 3PA players with high accuracy (more disadvantaged)
1. high 3PA players with low accuracy (most disadvantaged)

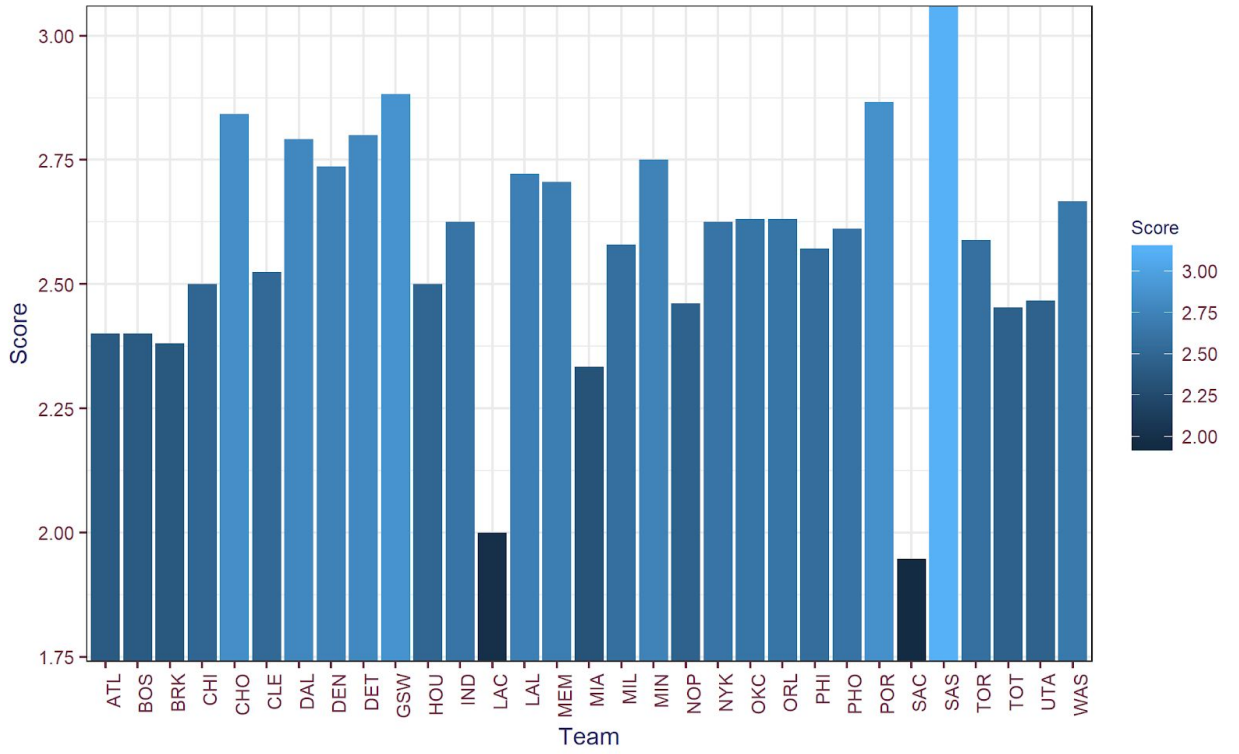
Where 1 represents most disadvantaged and 4 represents least disadvantaged (or also, most advantaged).

This allowed us to aggregate data related to players based on their "player_score" and determine correlating variables, allowing us to visualize what kind of players are benefitting based on their attributes in our displayed figures.

Furthermore, through joining our database tables and sorting our players into their respective teams, we could acquire our mean team score for each team. Mean was chosen over total score in order to accomodate teams that switched players frequently and thus had a greater number of players listed. With the mean team scores calculated, we were able to display clearly the teams that had the greatest prospective benefit from the proposed rule change.

Opinion

Evaluating on a team basis according to the steps shown in our analysis, we were able to produce this evaluation of each team's potential to benefit from the rule change.



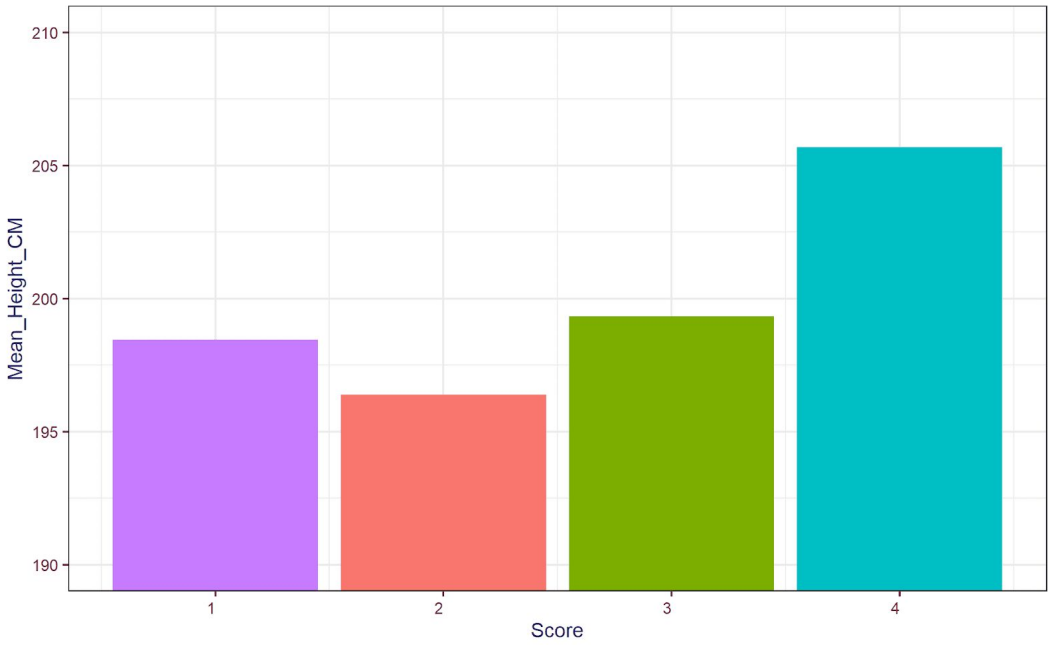
Our methodology shows that the current team most likely to benefit from the rule change is the San Antonio Spurs, with the Golden State Warriors and Portland Trail Blazers also benefiting a lot.

Initially, we were surprised to find the Warriors, a team that is known for shooting a lot of three pointers in our top three teams most likely to benefit. Upon further inspection of the data though we saw that although they attempted a lot of three pointers, those three pointers were concentrated among a couple of players, so there were far more players on the team taking less than 1.5 three-point attempts per game, causing the overall team score to be high.

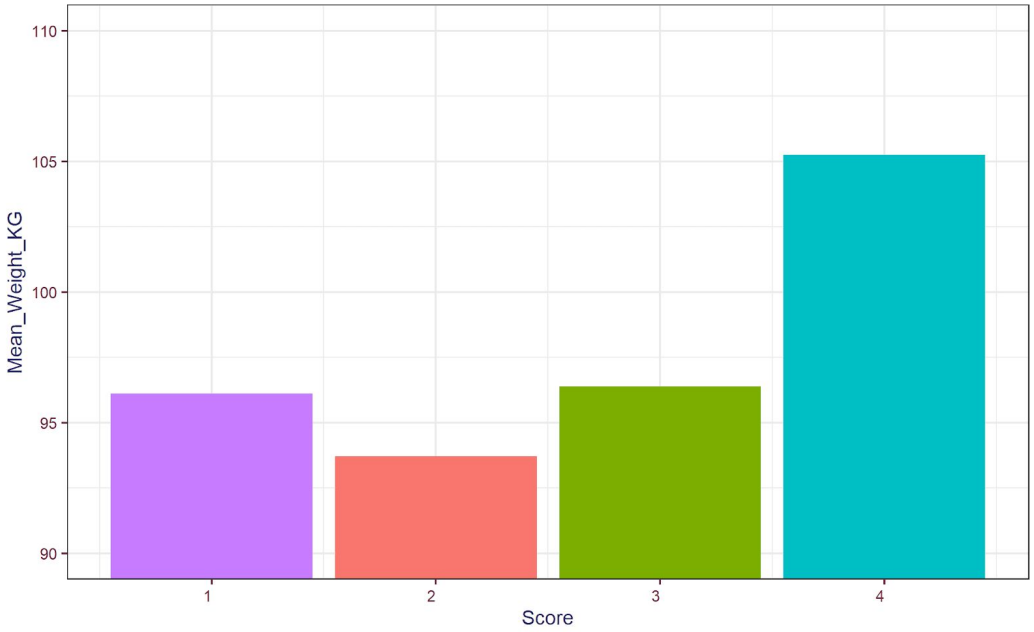
On the opposite side we see the Sacramento Kings followed by the Los Angeles Clippers as the teams most likely to be hurt by the rule change. This was reasonable for us because our player records indicated a well above average number of players on those teams who shot more than 3 three-point attempts per game and more than 5 three-point attempts per game. Thus these teams had a majority of their players with a 1 or 2 for their “player_score” column.

Evaluating on a player by player basis, there are also clear trends evident in the data to suggest what kinds of physical traits and age groups are going to be most benefited by this proposed change.

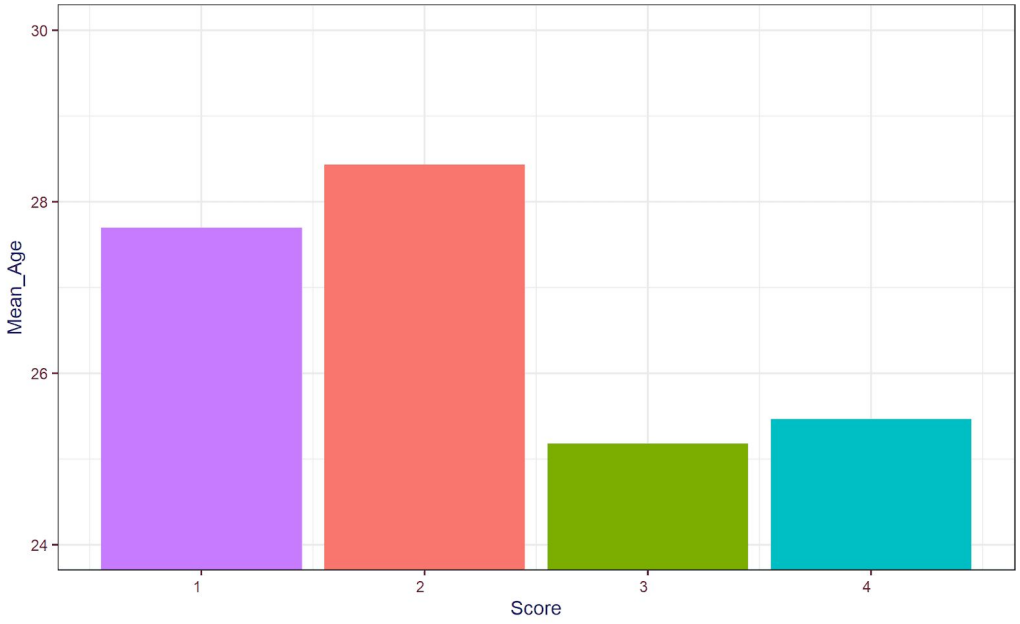
Mean Height in Centimeters by Score



Mean Weight in Kilograms by Score

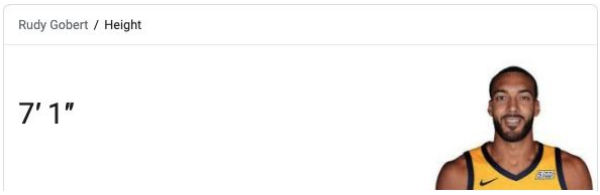


Mean Age by Score



From these figures we see first of all that taller and heavier players were way more likely to have a “player_score” of 4, or be in the not disadvantaged group. Thus players that are taller and heavier see the biggest advantage from the rule change.

We believe this to be the case because taller, heavier players are more likely to play the Center position (like Rudy Gobert). This is because their height gives them a significant advantage in scoring easier baskets closer to the hoop so they are likely to attempt far fewer three-point attempts per game than shorter players like Russell Westbrook.



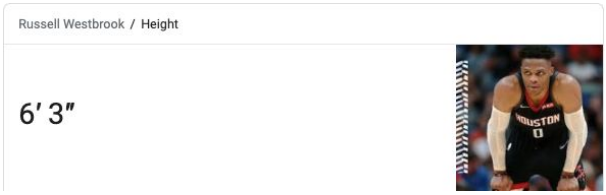


Table 1: Average Player Height per Position from 1952-2018

Year	All	PG	SG	SF	PF	C
2017/18	6'7"	6'2.8"	6'5"	6'7.5"	6'9.3"	6'11"
2016/17	6'7"	6'2.3"	6'5"	6'7.7"	6'9.4"	6'11.3"
2015/16	6'7"	6'2.4"	6'5"	6'7.5"	6'9.4"	6'11.4"
2014/15	6'7"	6'2.3"	6'5.3"	6'8"	6'9.4"	6'11.3"
2013/14	6'7"	6'2.1"	6'5.2"	6'7.8"	6'9.4"	6'11.2"
2012/13	6'6.9"	6'1.9"	6'4.9"	6'7.7"	6'9.2"	6'11"

Table 2: Average Player Weight per Position from 1952-2018

Year	All	PG	SG	SF	PF	C
2017/18	217 lbs	190 lbs	204 lbs	216 lbs	232 lbs	249 lbs
2016/17	218 lbs	189 lbs	203 lbs	220 lbs	236 lbs	251 lbs
2015/16	219 lbs	190 lbs	204 lbs	221 lbs	238 lbs	253 lbs
2014/15	221 lbs	190 lbs	208 lbs	221 lbs	241 lbs	254 lbs
2013/14	221 lbs	189 lbs	207 lbs	221 lbs	242 lbs	255 lbs
2012/13	221 lbs	188 lbs	207 lbs	222 lbs	240 lbs	254 lbs

https://www.thehoopsgeek.com/average-nba-height/

Trends amongst age were more simple to predict, with older players who are more experienced and less physically fit being more prone to taking further shots, making them disadvantaged were the proposed change to occur. Inversely, younger players more capable of dealing with the physical exertion of getting in the mix and taking close shots were visibly much more likely to benefit.

Next Steps

This question would benefit from a more focused study of players and how far away from the net they were able to successfully shoot during a game. Doing this study on just a few games would allow us to create a linear regression of accuracy by distance from the net which would better allow us to extrapolate accuracy at the new 3-Point line. We could also compare that data with our assumption of 2P% - 3P% for how much distance affects certain players.

In addition, if we operate under our current assumption that players that shoot fewer 3-Pointers would be less disadvantaged, data on the percentage of points scored from 3-Pointers from each team would be a good measure for how teams would be affected.

Reflection

It is very difficult to answer questions based on data that already exists that are not targeted towards the question. We had to make a lot of assumptions to get a starting point for our data analysis. We were very lucky because the data we were working with was very reliable and consistent, so we did not need to do a lot of data validation and we did not have much trouble importing.

One difficult part was compiling all the data for each player together from multiple sources. This could not be done simply with joins because we had to do some processing and conversion of data. The heights and weights were relatively easy because they are numeric and can be averaged, but compiling college data required taking either one data point or another. In addition, excluding a few columns from the data table required listing out all of the other columns, which was a bit tedious and easy to mistake.

Files

- Project4.ipynb - walkthrough of our pgSQL code
- Final Report.docx - final write up