ANALYSIS OF HEIGHT AND WEIGHT OF NBA PLAYERS

December 14, 2022

CSE 270 Sport Analytic

Student: Anna Martirosyan

Instructor: Habet Madoyan

Date: 14/12/2022

1 ABSTRACT

In the basketball, five players play per team, each assigned to positions: Point guard, Shooting guard, Small forward, Power forward, Center. It is a common knowledge that each position requires specific phisical profile, such as height and weight. For example it is typically for the team's shortest player to be the Point guard player or the pivot (or the big man) to be the Center player. But during the previous years, those divisions are not so evident.

The main aim of the following project is to analyse and determin how the phisical forms of NBA players have changed from 1996 to 2019.

2 Literature Review

In order to have deep understanding about NBA players' phisical state some scientific papers were analysied.

One of them was "Body height, body weight, body mass index of elite basketball players in relation to the playing position and their importance for success in the game" written by Milan Andelic and others (see: https://bit.ly/3UJbcjW). The main aim of the paper was to determine the differences in morphological characteristics in professional basketball players in relation to the playing position in different league competitions and to determine the relevance of these characteristics for success in the game. The survey included a sample of 773 elite basketball players competing in five different leagues in Europe. The study concluded that centers are significantly taller and heavier than guards and forwards are. In addition to other components of basketball (technique, tactics, coach strategy), body height, body mass and body mass index play a major role in overall performance in a basketball game in all positions.

Also, besides the abovementioned paper another paper called "Body Height of Elite Basketball Players: Do Taller Basketball Teams Rank Better at the FIBA World Cup?" published by the International Journal of Environmental Research and Public Health (https://bit.ly/3PiC7BX) was also investigated. The paper analyzed the differences between the basketball players from the teams ranked 1-16 and those ranked below 16th place. The body heights of all players from the last three FIBA-WCs were collected and allocated according to the ranking at the FIBA-WC and analyzed by position in team. An independent sample t-test was conducted to analyze the difference in body height of players ranked 1-16 and players who ranked below 16th place. The research concluded that the players from the first 16 teams were significantly taller at three positions: point guards, shooting guards, and small forwards. Considering that all players at the FIBA-WC went through rigorous selection process to be in their national teams, body height of the higher-ranked players could be used as a reference value.

Comparison of these two scientific articles, their methods and conclusions, played significant role in analzing NBA players' phisical charachteristics.

For the process of analysing data, making some visualizations and conclusions at first we need to collect all the necessary data, filter it, transform and make convinient to use. During this project I will use data of basketball players, such as name, team, age, heught, weight, draft year, points per season, seasons played and so on (more details you can see in the dataframe below). Besides that, we also need a dataset with country names and coutry codes.

[3]:

	player_name	team_abbreviation	age	player_height	player_weight	college	country	draft_year	draft_round	draft_number	…	pts	reb	ast	net_rating	oreb_pct	dreb_pct	usg_pct	ts_pct	ast_pct	season
0	Dennis Rodman	CHI	36.0	198.12	99.790240	Southeastern Oklahoma State	USA	1986	2	27	…	5.7	16.1	3.1	16.1	0.186	0.323	0.100	0.479	0.113	1996-97
1	Dwayne Schintzius	LAC	28.0	215.90	117.933920	Florida	USA	1990	1	24	…	2.3	1.5	0.3	12.3	0.078	0.151	0.175	0.430		
2	Earl Cureton	TOR	39.0	205.74	95.254320	Detroit Mercy	USA	1979	3	58	…	0.8	1.0	0.4	-2.1	0.105	0.102	0.103	0.376		
3	Ed O'Bannon	DAL	24.0	203.20	100.697424	UCLA	USA	1995	1	9	…	3.7	2.3	0.6	-8.7	0.060	0.149	0.167	0.399		
4	Ed Pinckney	MIA	34.0	205.74	108.862080	Villanova	USA	1985	1	10	…	2.4	2.4	0.2	-11.2	0.109	0.179	0.127	0.611		

```
 1    0.048  1996-97
 2    0.148  1996-97
 3    0.077  1996-97
 4    0.040  1996-97

 [5 rows x 21 columns]
```

```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 11145 entries, 0 to 11144
Data columns (total 21 columns):
 #   Column             Non-Null Count  Dtype
---  ------             --------------  -----
 0   player_name        11145 non-null  object
 1   team_abbreviation  11145 non-null  object
 2   age                11145 non-null  float64
 3   player_height      11145 non-null  float64
 4   player_weight      11145 non-null  float64
 5   college            11145 non-null  object
 6   country            11145 non-null  object
 7   draft_year         11145 non-null  object
 8   draft_round        11145 non-null  object
 9   draft_number       11145 non-null  object
 10  gp                 11145 non-null  int64
 11  pts                11145 non-null  float64
 12  reb                11145 non-null  float64
 13  ast                11145 non-null  float64
 14  net_rating         11145 non-null  float64
 15  oreb_pct           11145 non-null  float64
 16  dreb_pct           11145 non-null  float64
 17  usg_pct            11145 non-null  float64
 18  ts_pct             11145 non-null  float64
 19  ast_pct            11145 non-null  float64
 20  season             11145 non-null  object
dtypes: float64(12), int64(1), object(8)
memory usage: 1.9+ MB
```

We need to change:

* drafted column type to boolean

* draft_year column type to integer

* season column type to integer

[25]:
```
                age  player_height  player_weight            gp           pts  \
count  11145.000000   11145.000000   11145.000000  11145.000000  11145.000000
mean      27.168686     200.812818     100.637868     52.005832      8.126487
std        4.344164       9.190973      12.576295     25.069495      5.935482
```

min	18.000000	160.020000	60.327736	1.000000	0.000000
25%	24.000000	195.580000	90.718400	32.000000	3.500000
50%	27.000000	200.660000	99.790240	58.000000	6.600000
75%	30.000000	208.280000	109.315672	74.000000	11.500000
max	44.000000	231.140000	163.293120	85.000000	36.100000

	reb	ast	net_rating	oreb_pct	dreb_pct	\
count	11145.000000	11145.000000	11145.000000	11145.000000	11145.000000	
mean	3.560036	1.801463	-2.153899	0.055593	0.141772	
std	2.495394	1.789940	12.150611	0.043889	0.063194	
min	0.000000	0.000000	-200.000000	0.000000	0.000000	
25%	1.800000	0.600000	-6.300000	0.022000	0.096000	
50%	3.000000	1.200000	-1.300000	0.043000	0.132000	
75%	4.700000	2.400000	3.200000	0.086000	0.182000	
max	16.300000	11.700000	300.000000	1.000000	1.000000	

	usg_pct	ts_pct	ast_pct	drafted
count	11145.000000	11145.000000	11145.000000	11145.000000
mean	0.185599	0.508099	0.131078	0.825751
std	0.053047	0.098879	0.095017	0.379340
min	0.000000	0.000000	0.000000	0.000000
25%	0.150000	0.478000	0.065000	1.000000
50%	0.182000	0.521000	0.102000	1.000000
75%	0.218000	0.557000	0.178000	1.000000
max	1.000000	1.500000	1.000000	1.000000

From the dataset it is clear that:

* Since each season consists of 82 games, but the maximum number of games played is 86 => we need to analyse the 4 additional ones. Some analyse showed us the players who played more than 82 games, and overall there are 14 of them. This is possible because the reason is that those players were traded in the mid of the season. Therefore the records will remain unchanged.

* Next is, that net rating ranges from 300 to -200. Strang! Isn't it? This will also be analysed! The formula to calculate net rating is the following: nr = 100((points)/(poss) - 100((opponents points)/(opponents poss)). Still confusing since Bowen could get rating of 300, but outliers are the players who had very light contribution(just one game played) or didn't contribute a team at all => we will get rid of those players (records).

[24]:

	player_name	team_abbreviation	college	country	draft_year	\
count	11145	11145	11145	11145	9203	
unique	2235	36	316	71	44	
top	Vince Carter	TOR	None	USA	1998-01-01 00:00:00	
freq	22	390	1684	9410	454	
first	NaN	NaN	NaN	NaN	1963-01-01 00:00:00	

```
last             NaN               NaN     NaN     NaN  2019-01-01 00:00:00

       draft_round draft_number               season
count        11145        11145                11145
unique           8           75                   24
top              1    Undrafted  2017-01-01 00:00:00
freq          6513         1959                  540
first          NaN          NaN  1996-01-01 00:00:00
last           NaN          NaN  2019-01-01 00:00:00
```

Let's summarize what we have at this point:

NBA is extremly competitive, because there are 2235 unique players for 24 unique seasons.

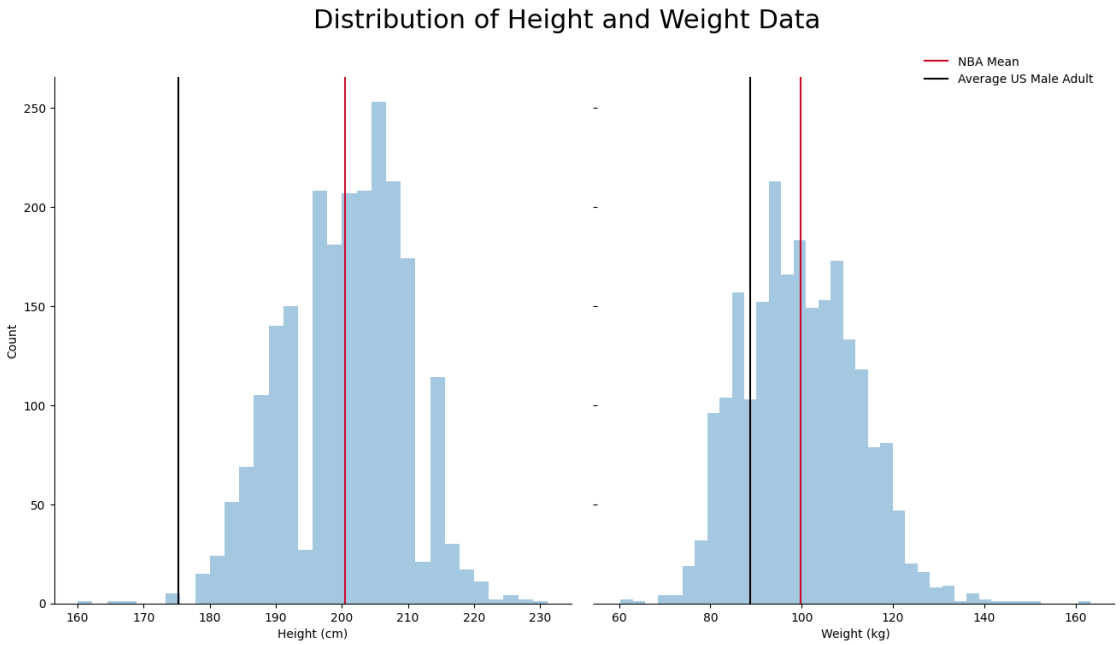
Now we see overall 36 unique teams, but in 2019 it was 30. So we will analyze additional 6 teams.

8 draft rounds are strange, beacuse in 2019 there were only 2.

* After some research I found out that before 1989 there were many rounds (6,7,...,21), and after 1989 the system was changed to 2 round.

* The 8th unique value is undrafted variable. The unique draft number column is 75, but since there are 2 rounds and 30 in each of them => again strange! Research found out that this is again the consequence of previous draft format.

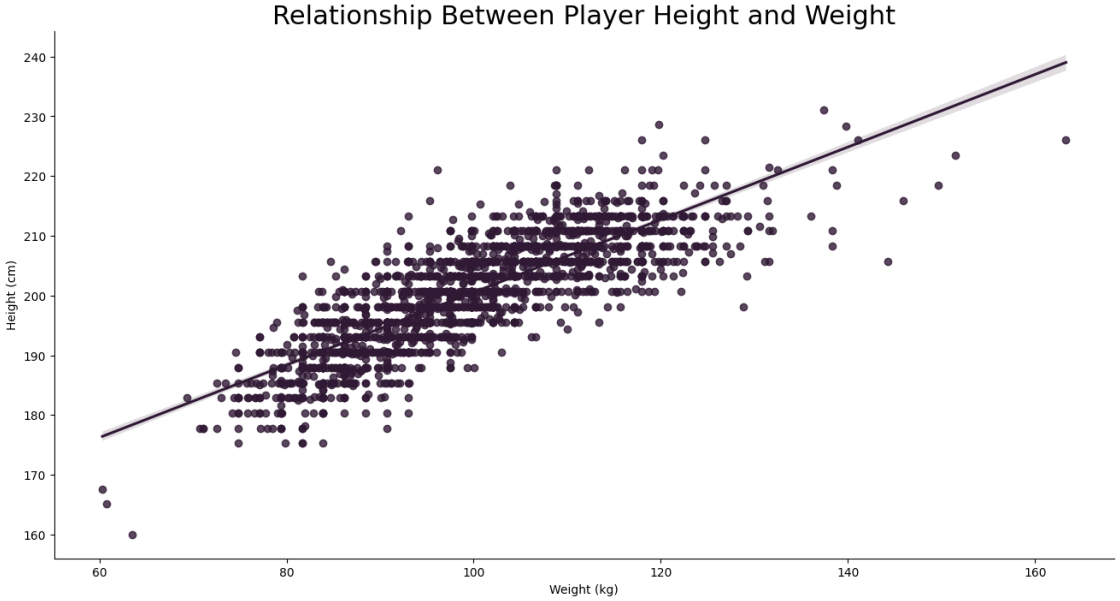
NOW, lets change the types of some columns, add additional new columns or change the content of excisting ones, to make it more convinient to work with them further.



From the above visualizations, with the red line we can see the mean of height and weight of NBA male players. And it is clear that there is a deviation compared with height and weight of average US man. Also both height and weight are distributted normally.

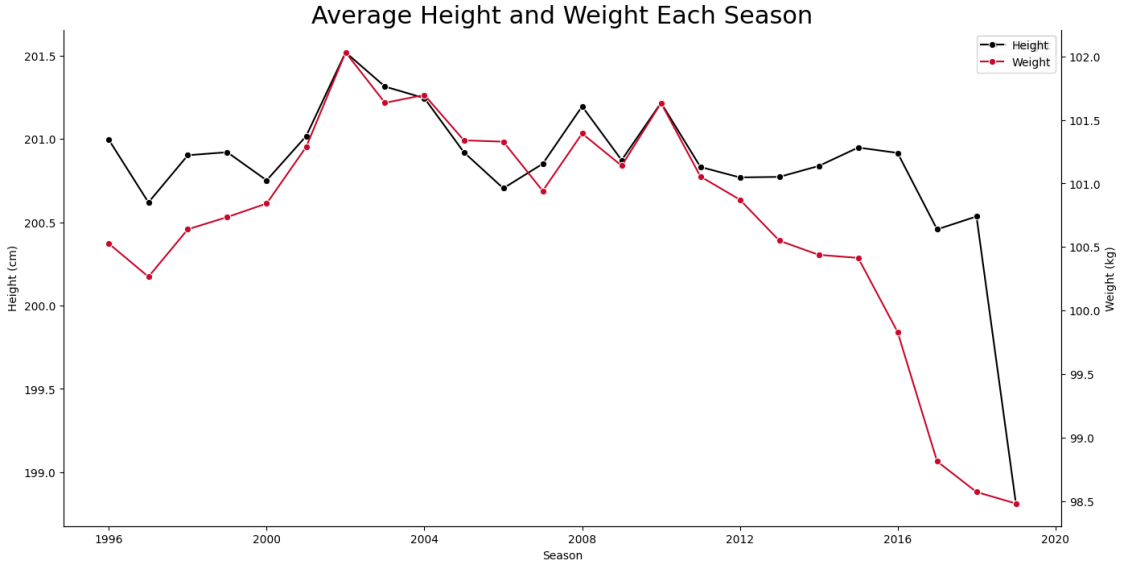
Now let's find the correlation between height and weight with a scatterplot.

Now lets plot a correlation line to see more clearly, how these 2 attributes are correlated.

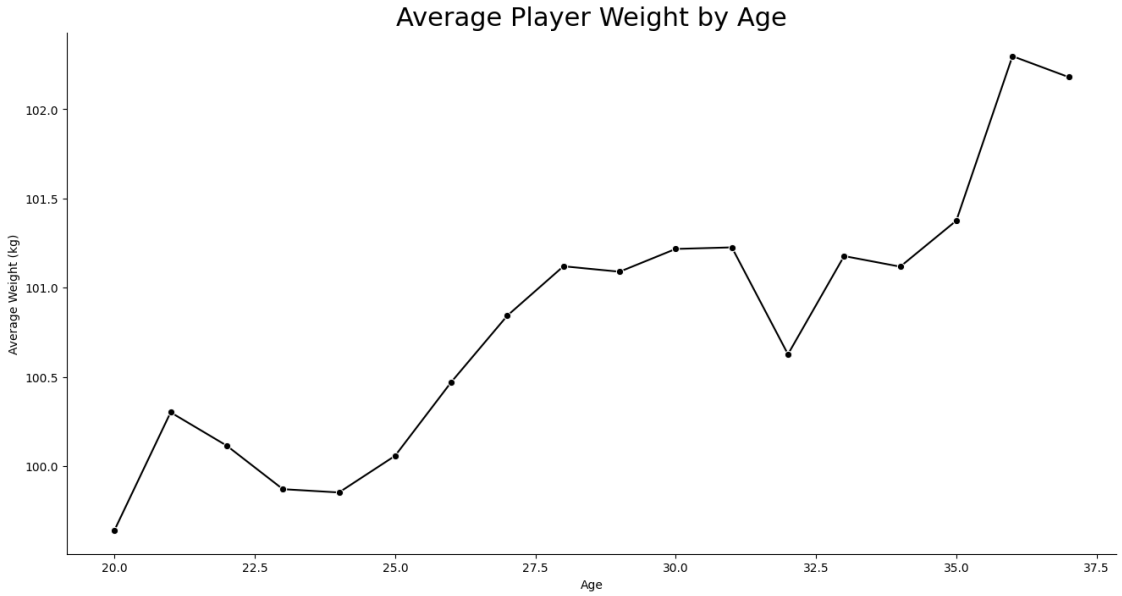


From the above graph we can conclude that height and weight are nearly linearly positively correlated. Which means that increase in height will lead to incrase in weight and vice versa.

This is an early indicator of the changing body types of NBA players. Now let's see the correlation results per season, so it will be more visible to understand whether it is increasing or decreasing. As as wee can see from the below graph phisical attributes(height and weight) and body types are changing over time. And after 2015 it seems that it tends to decrease.



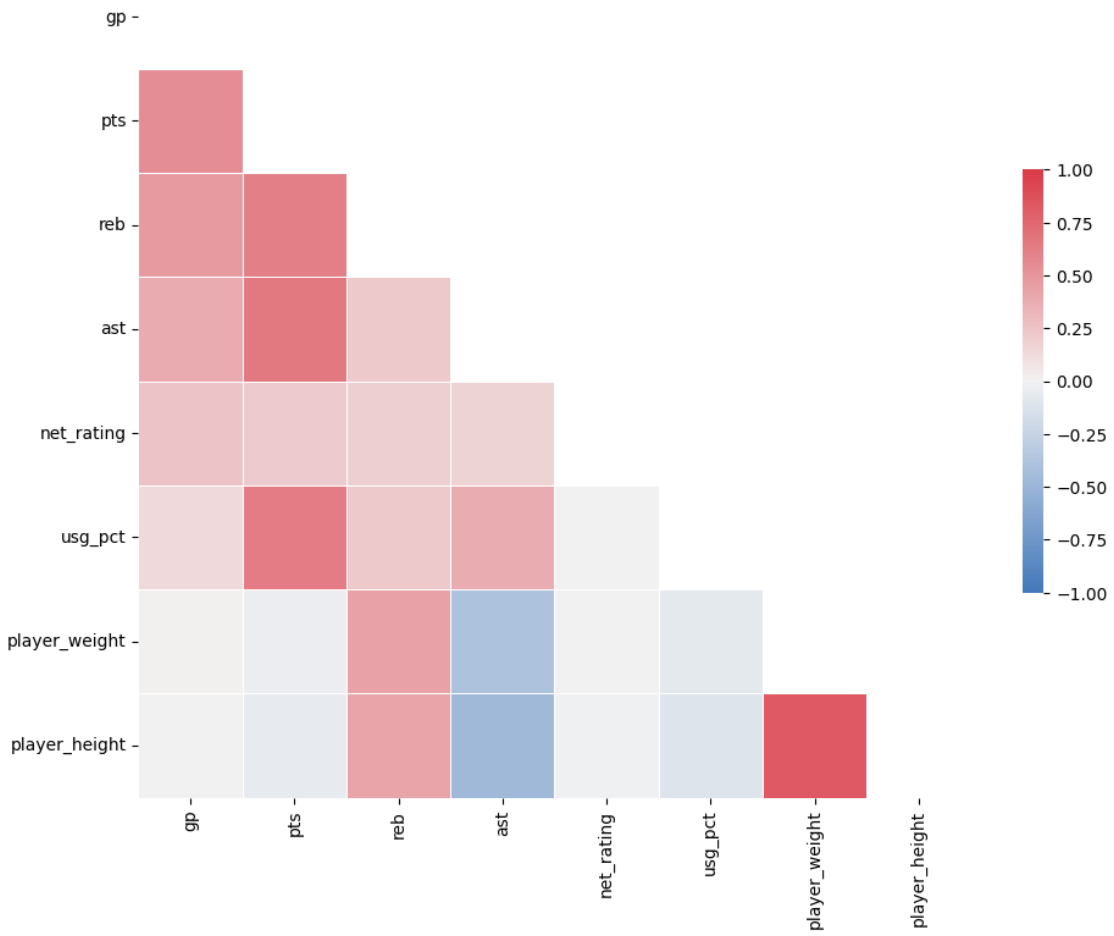
From the above graph we can see that after 2015-2016 season both the height and weight average is decreasing and in 2019 it reaches its lowest stage.



The above graph shows the average player weight by age. And as we can see by the increase of age the weight also increases. So we can conclude that younger NBA players weight less than older ones.

The below visualisation showcases average player height and weight (minimum 3 unique players) based on the country they were born in (or have citizenship of).

Now lets find the correlation of player height and weight with average assists and rebounds, points and other statistics. And as we can see from below graph height and weight are affecting average assist and rebound statistics.



In the above correlation heatmap, if the 2 variables have positive correlation (near 1) they are colored red.
If the correlation is negative the color is blue.

3 CONCLUSION

To conclude, the height and weight as well as other phisical attributes highly affect NBA players' performance and are highly correlated with average assists, rebounds, blocks and other statistics. The body structure of NBA players have changed through years and it seems that it tends to decrease through time. Also age is an important factor which affects player's weight. And finaly from the research we can conclude that the players with highest height and weight were born in China or have citizenship of China.