NBA Player Features

Ian and Matteo

2022-12-14

Player Attributes vs. In game Stats

Since the introduction of statistics into sports, analysis's have always been trying to find out how to win more games. Many things about a player contribute to their effectiveness. Field goal percentage, passes per game, minutes per game etc. are all attributes that give some sort of insight into a players game. But something that has always interested us is how features that a player can't change effect their game. So the question we wanted to answer is, how do a players physical features effect their in game stats?

Data Wrangling

To answer this question, we need data that includes all the current NBA players physical features with their average points per game. We were not lucky enough to find this exact data in one place. NBA.com had two different datasets that when combined, would give us all the information we needed. The first dataset has every player in the current season, along with some of their stats. The second dataset contained information about the physical attributes of each player in a given NBA combine season. Dr. Hatfield informed us that the website that contained this data was dynamically written. To scrape the from the site, we needed to first save the website, and then provide the location of the saved website in our computer as the html. We chose to start our data collection at the 2003-2004 combine because the years prior do not have as many current NBA players. After we scraped all the years, we combined them into one data frame of all combine players physical data. This dataframe has 1327 cases while the total current players dataframe has only 289 cases. This is because not every player that was recorded during the combine made it to an NBA team. Also, many of the players who were drafted in seasons prior, no longer play in the NBA. To make sure that only the current players are kept, the two tables were inner joined by a players name so only the players whose names are present in the current seasons data would be present in the final dataset. This final table contains 182 cases.

Figure 1: Summary statistics of rebounds per game grouped by position

Figure 2: Summary statistics of blocks per game grouped by position

Figure 3: Summary statistics of points per game grouped by position

Figure 4: Summary statistics of assists per game grouped by position

Figure 5: Summary statistics of steals per game grouped by position

Before fitting a regression model to predict certain features, we want to make sure which features are correlated with each other. We don't want to train a model on features that have no correlation with each other. To see how features are related, we want to create a heat map showing our desired features correlation.

Table 1: NBA Reounds per Game (RPG) per position

POS	min	Q1	Median	Q3	Max	Mean	SD
	4.2	4.20	4.2	4.20	4.2	4.20	NA
C	2.8	5.50	7.4	9.00	12.1	7.11	2.64
PF	1.4	3.70	5.0	6.97	12.4	5.58	2.53
PG	0.0	2.15	2.8	4.43	6.6	3.17	1.61
SF	2.0	2.78	4.2	5.27	8.6	4.34	1.75
SG	0.0	2.70	3.1	4.00	8.3	3.38	1.53

Table 2: NBA Blocks per Game (BPG) per position

POS	min	Q1	Median	Q3	Max	Mean	SD
	0.5	0.50	0.50	0.50	0.5	0.50	NA
C	0.4	0.70	1.00	1.55	2.9	1.22	0.72
PF	0.1	0.30	0.50	0.88	2.2	0.64	0.43
PG	0.0	0.18	0.20	0.32	1.1	0.31	0.27
SF	0.1	0.20	0.45	0.58	1.7	0.48	0.33
SG	0.0	0.10	0.20	0.40	0.6	0.25	0.14

Table 3: NBA Points per Game (PPG) per position

POS	min	Q1	Median	Q3	Max	Mean	SD
	5.4	5.40	5.40	5.40	5.4	5.40	NA
C	3.8	6.60	8.40	12.65	17.5	9.39	3.95
PF	4.3	6.32	9.40	13.93	28.1	11.04	5.89
PG	0.0	6.50	12.00	19.60	31.2	13.53	8.14
SF	5.1	7.30	8.95	13.85	30.0	11.81	6.90
SG	0.0	7.00	10.60	15.70	29.0	11.95	6.88

Table 4: NBA Assists per Game (APG) per position

POS	min	Q1	Median	Q3	Max	Mean	SD
	1.4	1.40	1.40	1.40	1.4	1.40	NA
C	0.0	0.65	1.00	1.25	4.1	1.25	1.02
PF	0.2	0.80	1.35	2.48	6.8	1.79	1.32
PG	0.0	2.50	4.05	5.85	9.6	4.34	2.27
SF	0.4	1.22	1.85	3.30	6.5	2.32	1.52
SG	0.0	1.20	2.20	3.50	5.8	2.41	1.44

Table 5: NBA Steals per Game (SPG) per position

POS	min	Q1	Median	Q3	Max	Mean	SD
	0.5	0.50	0.50	0.50	0.5	0.50	NA
C	0.4	0.70	1.00	1.55	2.9	1.22	0.72
PF	0.1	0.30	0.50	0.88	2.2	0.64	0.43
PG	0.0	0.18	0.20	0.32	1.1	0.31	0.27
SF	0.1	0.20	0.45	0.58	1.7	0.48	0.33
SG	0.0	0.10	0.20	0.40	0.6	0.25	0.14

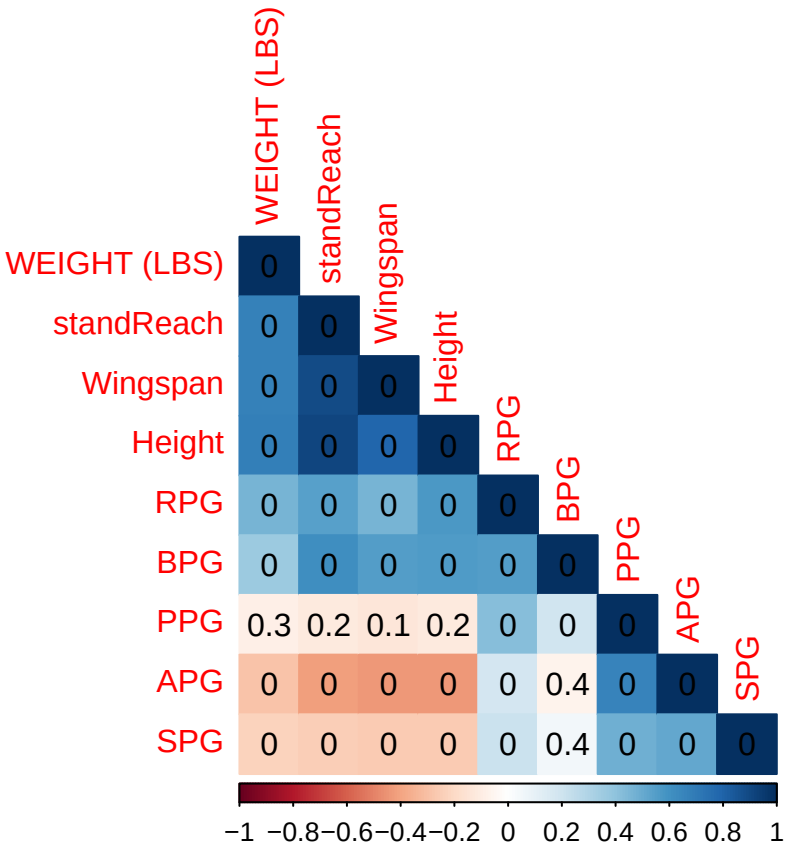
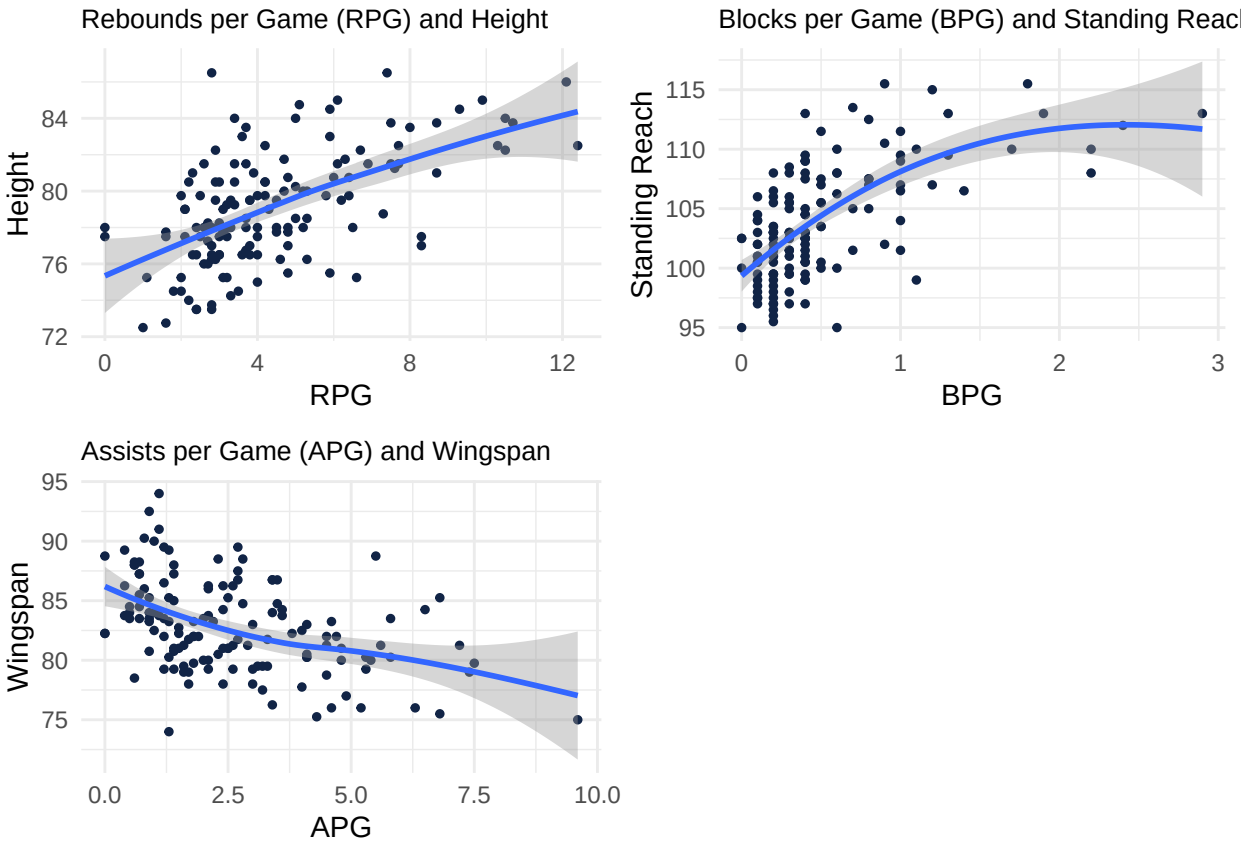


Figure 6: Correlation matrix showing the relations of all our desired features

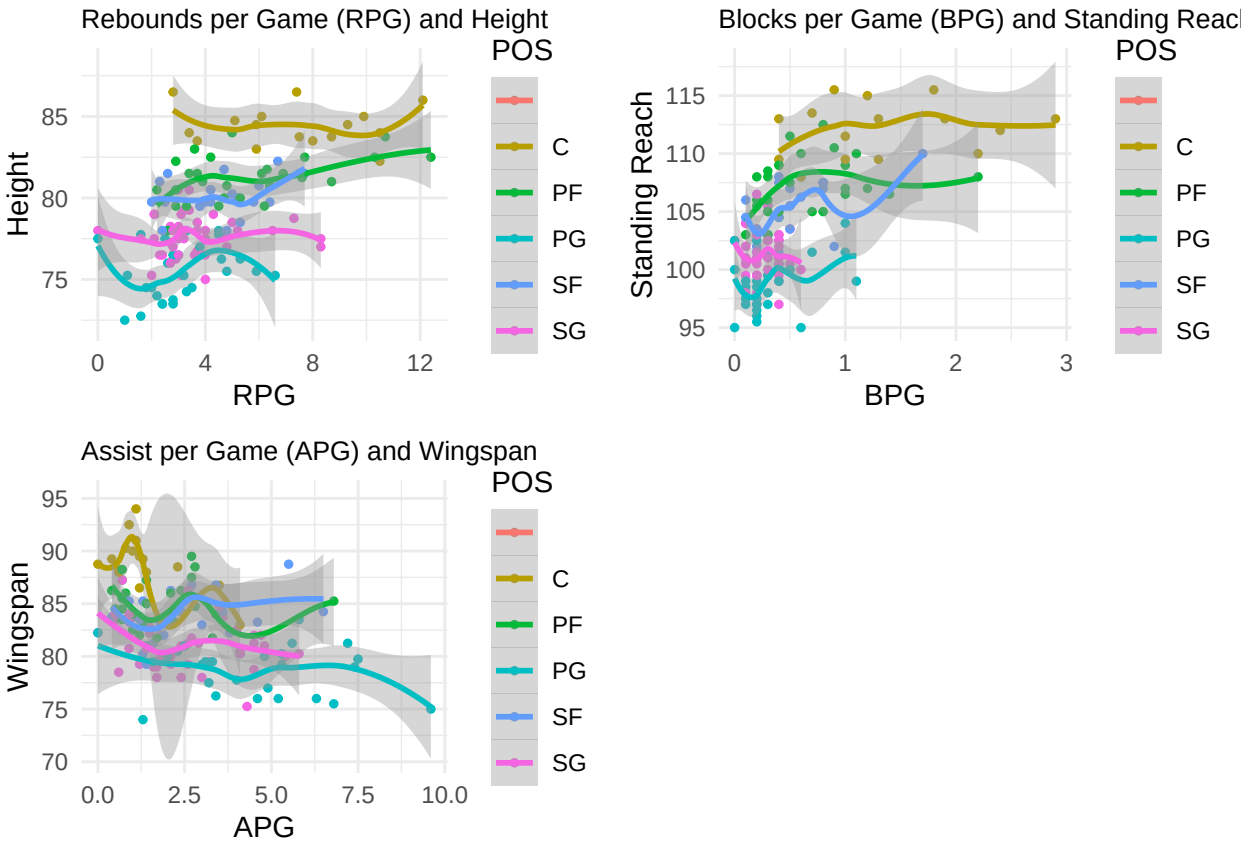
Based on the matrix, we decided to make three comparisons. Rebounds per game and height, blocks per game and standing reach, and assists per game and wingspan. We made a regression summary for each variables we are testing (RPG, BPG, PPG, ASP, SPG). From each, we can see the residual, as well as different coefficients. Something important to our test is the correlation and p-values. These coefficients can tell us if any of our NBA player measurements could determine a statistic.

From the regression summary, we can see that three of the five variables saw strong correlation between the NBA player measurements and statistics. Those were RPG, BPG, and ASP. The strongest measurement for each were RPG and height, BPG and standing reach, and APG and wingspan. We decided to then create a visualization for each.



Results

The fitting of the linear regression models is visualized by these three graphs. With the fitting acting as a line of best fit, the relationship between physical features and in game stats can be seen. For example, the taller a player, the more blocks per game. The longer a players standing reach, the more average blocks per game. The longer a players wingspan, the less assists per game they average.



These final visualizations contain the same data as the prior group of visualizations, but now, a players position is used as a filter. This allows us to see the importance of features depending on a players position. For example, in Rebounds per Game and Height, in the power forward (PF) position, the taller the player the more rebounds they get relationship is apparent and can be seen by the increasing green line. But if you look at the shooting guard (SG) position or the point guard (PG) position, there seems to be no correlation between the players height and the number of rebounds per game. Looking at the Blocks per Game and Standing Reach visualization it can be seen that standing reach might not matter as much for point guards as it does for small forwards. Lastly, in the Assist per Game and Wingspan visualization, a greater wingspan results in less assists per game throughout all positions.