Impact of Field Goal Percentage vs Free Throw Percentage

University of California, San Diego

ECON 5/POLI 5D: Data Analytics

Summer Session II 2020

Abstract: This report analyzes whether the relationship between field goal percentage and free throw percentage is more influential in a team's overall points scored in a game. Field goal percentage is measured as a decimal value from 0.0 - 1.0 and free throw percentage is measured the same way. This research question hopes to explore what makes more of an impact statistically on a team's overall points scored in a game. The data is presented per game and is analyzed on a per game basis. There are 9840 games present with 30 different teams with data ranging from 2014-2018. The results largely support the hypothesis that field goal percentage is more influential than free throw percentage when comparing the values to an overall team's points.

Introduction

Basketball is one of the most popular sports in the United States following only football in terms of national viewership. The NBA is the highest level of basketball played within the United States and has 30 teams. There are 4 12 minute quartersOf these 30 teams, there are 2 divisions and 3 subdivisions within each of these: The East [Atlantic, Central, SouthEast] and the West [Southwest, Northwest, Pacific] (nba.com). To better understand basketball, I will provide a few basic descriptions that will be useful to note before looking more in depth at the data. Specifically I will explain field goals, rebounds, free throws, steals, blocks, turnovers, fouls, and how the percentages are calculated. One of the most important being the definition of a field goal: "A field goal is considered any basket to be made during regular game play other than a free throw, worth two or three points" (masterclass.com). A rebound is granted to a player on the court if a shot "does not go into the net and bounces off the rim or backboard and into the hands of a player below" (rookieroad.com). A free throw is defined as "an unimpeded attempt at a basket (worth one point) awared to a player following a foul or other infringement" (Oxford Dictionary). According to sportslingo.com a steal is when a defensive player is responsible for a legal turnover due to his/her positive defensive actions (sportslingo.com). A block is defined as a "successful deflection of an attempted shot by a defender" (rookieroad.com).A turnover in basketball is when a team "loses possession of the ball and the other team gains possession" (rookieroad.com). A foul in basketball is defined as "a rule-breaking action that results in a loss of possession and possible free throws" (rookieroad.com). Lastly, the field goal and free throw percentages are calculated by dividing the number of made shots by the number of attempted shots and gathering a percentage value from 0-100. Despite many believing that the best way to score points is through field goals, these shots are contested and often have a lower percentage of going in. Because of this, my goal in pursuing analytics within this dataset is to find out whether a higher free throw percentage is more important in a team's overall points scored compared with a higher field goal percentage.

My key question is: Is a team's free throw percentage or field goal percentage more likely to predict their team points? In other words does a higher percentage of either of these shots contribute to a higher chance of scoring more in a game? One might initially think that field goals are going to be much more valuable in predicting a team's points; however, the percentage is much lower and that might play a larger factor into this equation. I chose these two variables because I believe that both will have a large effect on a team's points and it would be interesting to see as a basketball player and fan, which percentage is more important in this dataset, and how it could potentially be extrapolated outward.

I hypothesize that field goal percentage will be more influential in a team's overall points because there are more opportunities for a player to score a field goal despite having a lower percentage. In addition to this, field goals are also worth 2 or 3 points compared to a free throw only being worth 1. In contrast to this, it is important to consider that free throws in general should be a higher percentage shot since there is less pressure involved and it is a mid-distance

shot. A null hypothesis for this equation would be that field goal percentage is not more predictive of a team's points compared to free throw percentage.

Data

My data frame consists of the NBA dataset from 2014-2018 that contains 9840 games of basketball amongst 30 different teams. Each individual row is a game listed that has 41 columns of information, a few of note that will be used in this project are field goals, rebounds, free throws, steals, blocks, turnovers, fouls, and the percentages within field goal attempts and free throw attempts. My main dependent variables that I will be comparing are field goal percentage and free throw percentage. My independent variable will be a team's overall points. I chose to work with the percentages of my dependent variables instead of attempts overall or shots made because it would help control for differences in attempts being shot, and would allow for the reader to better visualize the information. Both percentages are ratio values that are decimal values between 0 and 1.0 which correspond with a percentage value when multiplied by 100 from 0-100. The independent variable is an interval value that ranges from 0-149 points.

Table 1: Data Dictionary

Variable Name	Description
TeamPoints	Total points scored by the team listed in the row
FieldGoals.	Field Goal percentage calculated by dividing shots made by attempted (DV)
FreeThrows.	Free Throws percentage calculated by dividing free throws made by attempted (DV)
TotalRebounds	Total rebounds gathered by a team
Assists	Total assists acquired by a team
Steals	Total steals acquired by a team
Blocks	Total blocks acquired by a team
Turnovers	Total turnovers acquired by a team

Listed above in table 1 is my data dictionary that lists my useful variables and a brief description.

One thing to note from this dataset is that the games are from 2014-2018 which contains a large sample size in terms of number wise (9048); however, in the grand scheme of the NBA's

history this is not very representative of the NBA. Also, more recently the game of basketball has become much more of a "small man's game" where centers are now less valuable and the focus has become much more centered around being a good three-point shooter. Because of this, this information would be useful to consider when doing this analysis.

Table 2: Summary Statistics for Numerical Values in NBA Dataset

Variable	N	Mean	St. Dev.	Min	Pctl(25)	Pctl(75)	Max
X	9,840	11,960.160	18,654.870	1	2,537.8	7,715.2	82,119
Game	9,840	41.500	23.671	1	21	62	82
TeamPoints	9,840	103.652	12.188	64	95	112	149
OpponentPoints	9,840	103.652	12.188	64	95	112	149
FieldGoals	9,840	38.602	5.030	19	35	42	58
FieldGoalsAttempted	9,840	84.902	7.130	60	80	89	129
FieldGoals.	9,840	0.456	0.055	0.271	0.418	0.494	0.684
X3PointShots	9,840	9.127	3.598	0	7	11	25
X3PointShotsAttempted	9,840	25.624	7.103	4	21	30	61
X3PointShots.	9,840	0.354	0.098	0.000	0.290	0.419	0.750
FreeThrows	9,840	17.321	6.003	1	13	21	44
FreeThrowsAttempted	9,840	22.749	7.390	1	18	27	64
FreeThrows.	9,840	0.762	0.104	0.143	0.696	0.833	1.000
OffRebounds	9,840	10.288	3.807	0	8	13	38
TotalRebounds	9,840	43.521	6.410	20	39	48	81
Assists	9,840	22.547	5.123	6	19	26	47
Steals	9,840	7.751	2.959	0	6	10	21
Blocks	9,840	4.828	2.537	0	3	6	18
Turnovers	9,840	13.639	3.870	2	11	16	29
TotalFouls	9,840	20.059	4.318	7	17	23	42
Opp.FieldGoals	9,840	38.602	5.030	19	35	42	58
Opp.FieldGoalsAttempted	9,840	84.902	7.130	60	80	89	129
Opp.FieldGoals.	9,840	0.456	0.055	0.271	0.418	0.494	0.684
Opp.3PointShots	9,840	9.127	3.598	0	7	11	25
Opp.3PointShotsAttempted	9,840	25.624	7.103	4	21	30	61
Opp.3PointShots.	9,840	0.354	0.098	0.000	0.290	0.419	0.750
Opp.FreeThrows	9,840	17.321	6.003	1	13	21	44
Opp.FreeThrowsAttempted	9,840	22.749	7.390	1	18	27	64
Opp.FreeThrows.	9,840	0.762	0.104	0.143	0.696	0.833	1.000
Opp.OffRebounds	9,840	10.288	3.807	0	8	13	38
Opp.TotalRebounds	9,840	43.521	6.410	20	39	48	81
Opp.Assists	9,840	22.547	5.123	6	19	26	47

Opp.Steals	9,840	7.751	2.959	0	6	10	21
Opp.Blocks	9,840	4.828	2.537	0	3	6	18
Opp.Turnovers	9,840	13.639	3.870	2	11	16	29
Opp.TotalFouls	9,840	20.059	4.318	7	17	23	42

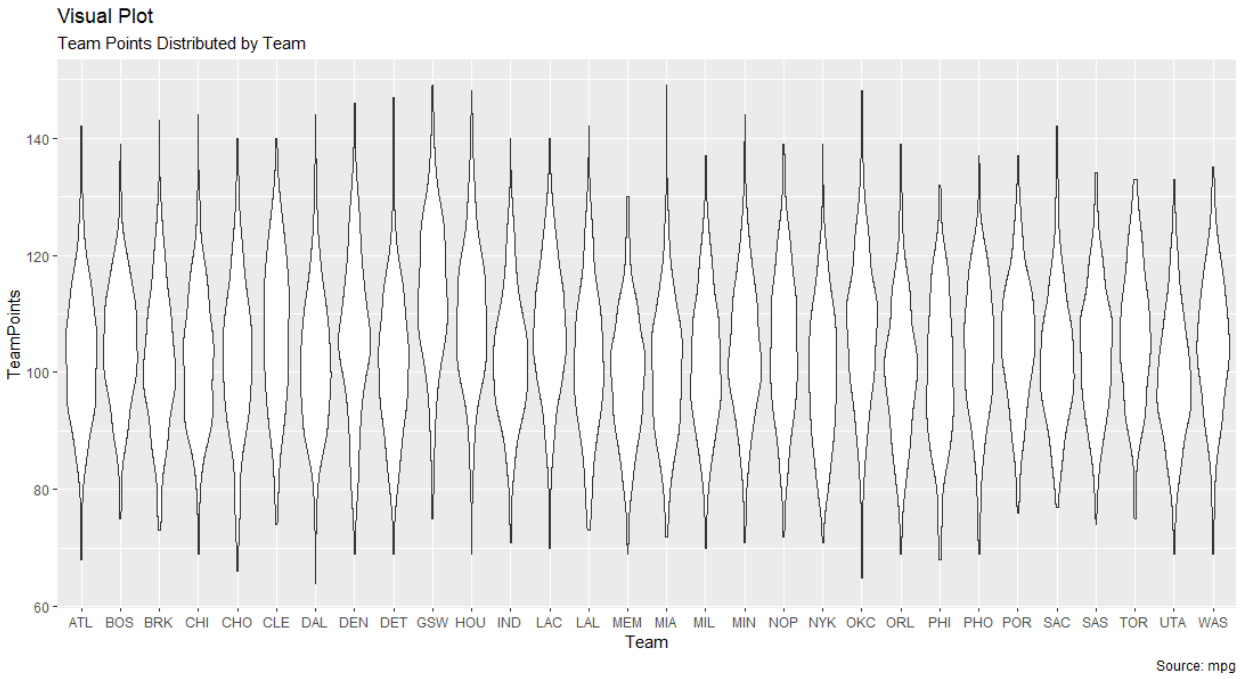
Table 2 has all of the summary statistics from the total dataset. The bolded variables are the main focus of this project and the ones I will look at briefly.

Table 3: Useful Summary Statistics

Variable	N	Mean	Median
FieldGoals.	9,840	0.456	0.455
FreeThrows.	9,840	0.762	0.769
TotalRebounds	9,840	43.521	43
TeamPoints	9,840	103.652	103
Assists	9,840	22.547	22
Steals	9,840	7.751	8
Blocks	9,840	4.828	5
Turnovers	9,840	13.639	13
TotalFouls	9,840	20.059	20

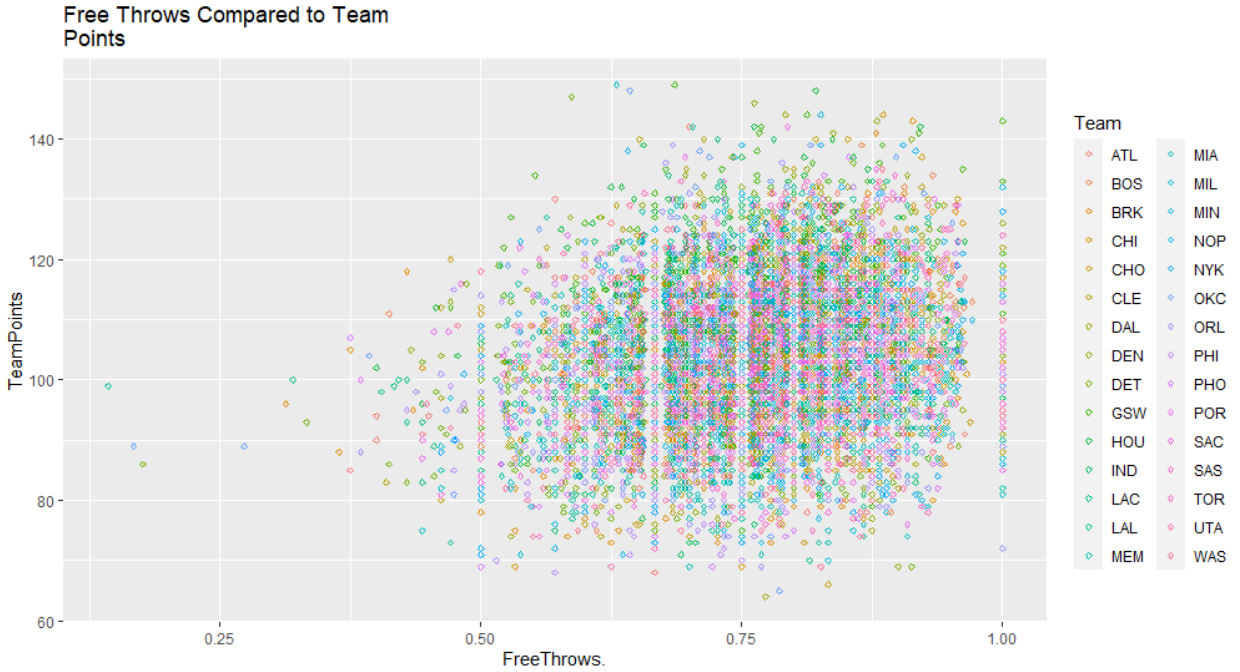
Table 3 provides us the important statistics that we will look at within our important variables. To note we see that the mean is greater than the median for field goal percentage, total rebounds, team points, assists, turnovers and total fouls (even if just slightly) could imply that there is a minor right skew. However, we see that the median is larger than the mean in free throw percentage, steals, and blocks which could imply that there is a slight left skew of the data. In table 2 we can see that the range of values in team points is quite large, and that this can later be seen in our graphs providing further visual analysis into our project.

Figure 4: Violin Plot on TeamPoint Distributions



The violin plot above shows a breakdown of the data specifically focusing on the TeamPoints variable within the NBA dataset.

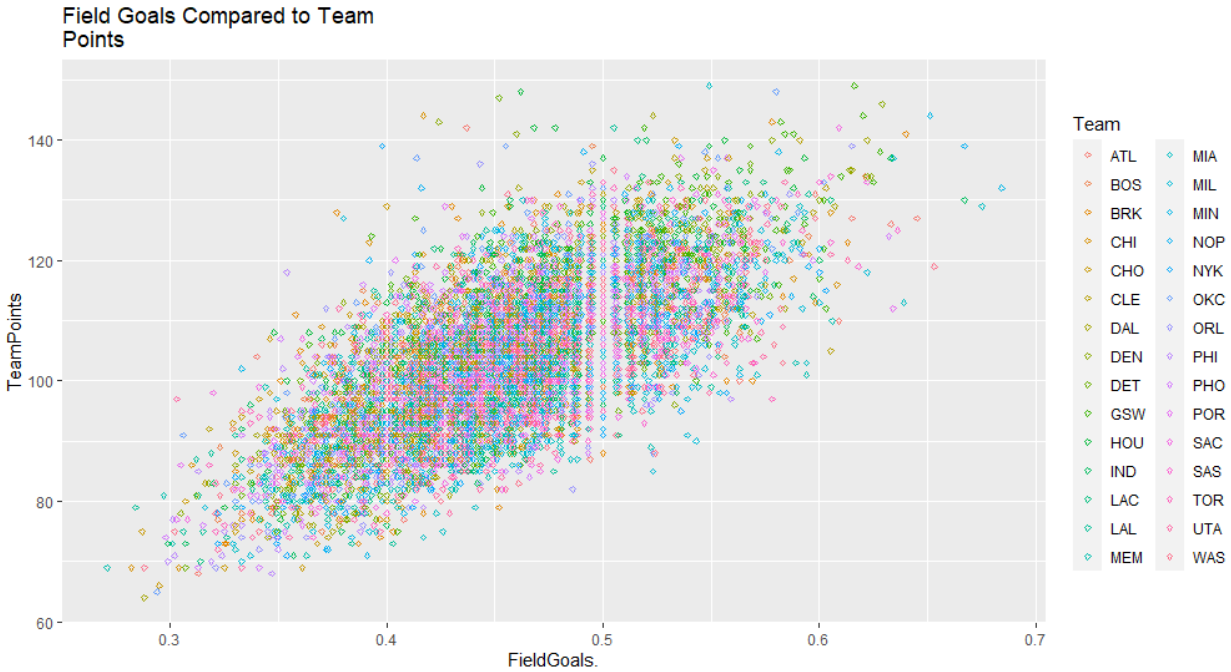
Figure 5:Free Throws



The scatterplot above shows the free throw percentage broken down by team color on the x axis and the team's points on the y axis. It shows that There is a positive correlation but the data is skewed left slightly because there are very few values that are low free throw percentage values

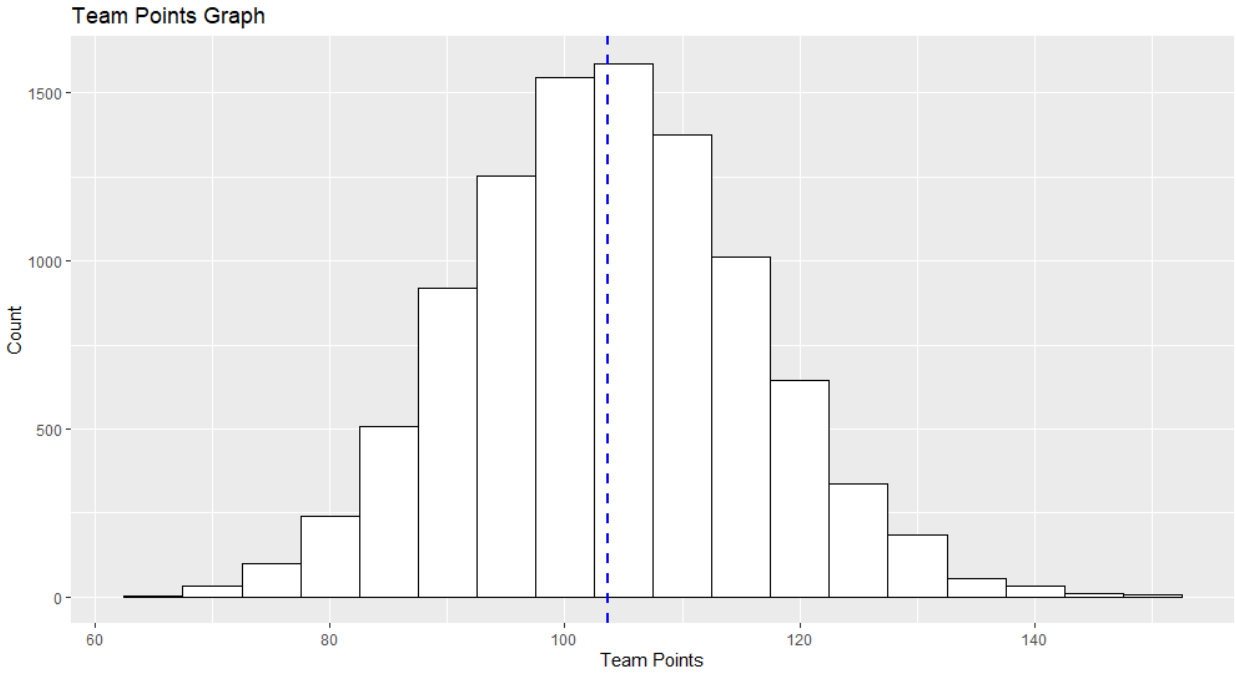
within the dataset mainly because the variable included in the analysis is considered to be very free points within a basketball consideration. There does not appear to be a linear correlation within this dataset.

Figure 6:



The figure above shows field goal percentage broken down by team in colors and field goal percentage on the x axis and team points on the y axis. This data does not appear to have any skew to it and it seems to be more tight fitting in a positive linear correlation. These two figures above help provide visual analysis towards the overall goal of finding out if free throws or field goal percentages are more influential.

Figure 7:



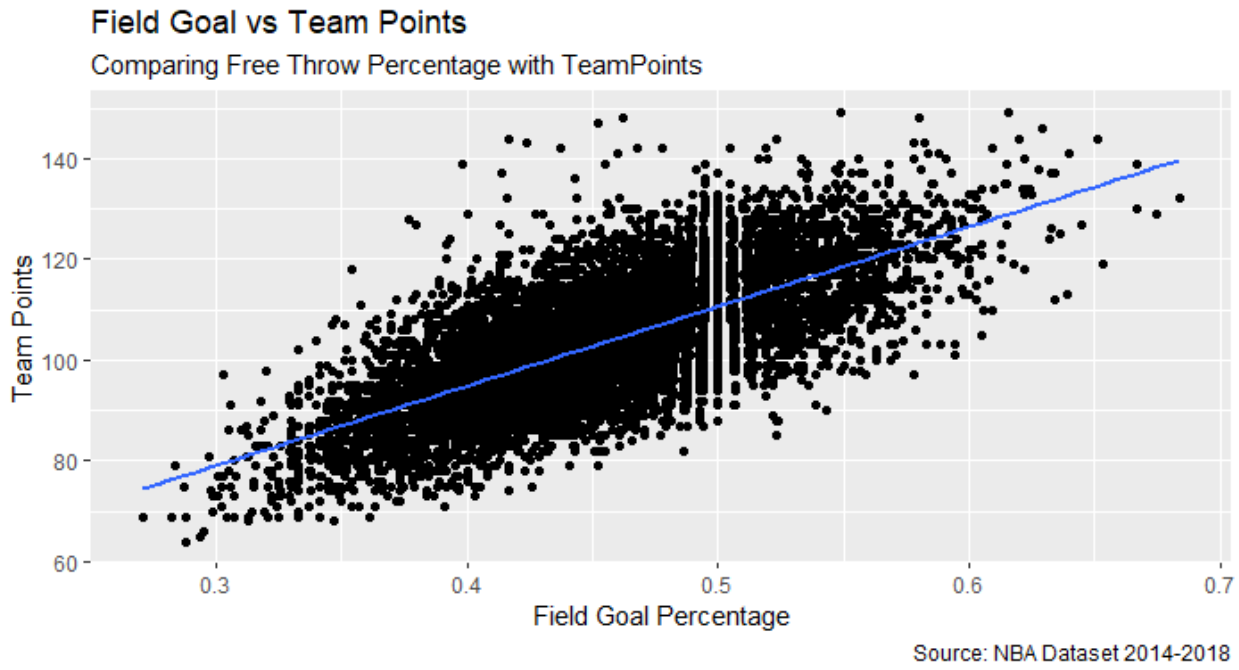
This graph is a histogram that shows the breakdown of team points and the counts. It shows that the team points variable appears to be normally distributed with no skews.

Methods and Results

	TeamPoints	FieldGoals.	FreeThrows.
TeamPoints	1.0000000	0.71018712	0.17570008
FieldGoals.	0.7101871	1.00000000	0.03072733
FreeThrows.	0.1757001	0.03072733	1.00000000

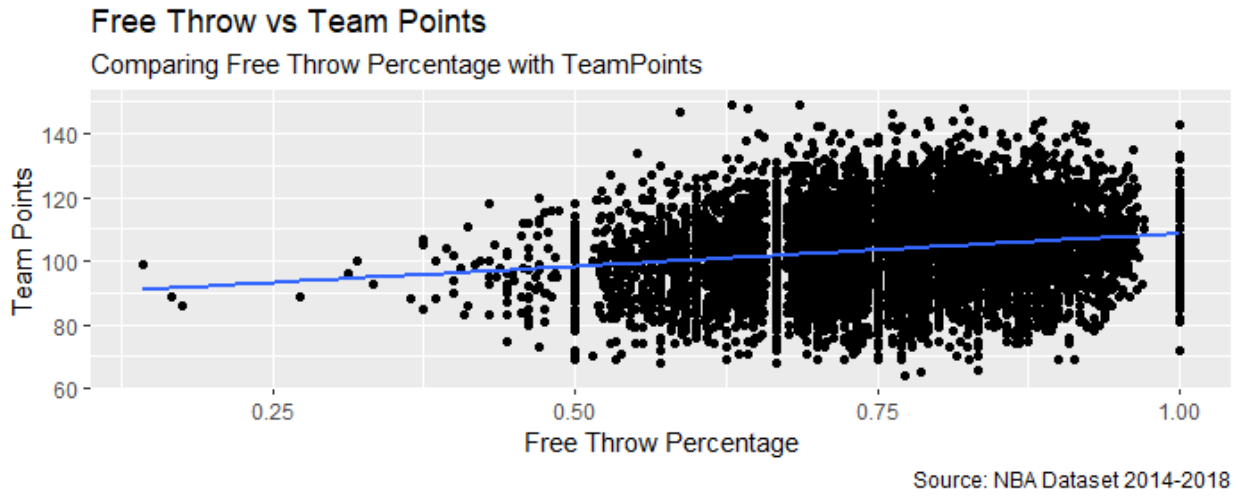
This small correlational matrix shows the correlation between the values of interest in my dataset. We can see that the field goal percentage is more influential.

Figure 8:



This figure above shows the scatterplot with a best fit line on field goal percentage and it shows a positive linear correlation. There are a lot of stacked values, so looking more in-depth at the R-squared value and p-values are going to be essential.

Figure 9:



This figure above shows the scatterplot with a best fit line of free throw percentage and it shows a slight positive correlation but does not appear to be very linear. The values do have correlation but linear does not appear to be the best fit for this graph.

Call:

```
lm(formula = TeamPoints ~ FreeThrows., data = data)

Residuals:
   Min     1Q  Median     3Q    Max
-39.870  -8.252  -0.287   8.027  48.066

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  87.9987     0.8925    98.6   <2e-16 ***
FreeThrows.  20.5321     1.1598    17.7   <2e-16 ***
---
Signif. codes:
0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 12 on 9838 degrees of freedom
Multiple R-squared:  0.03087,        Adjusted R-squared:  0.03077
F-statistic: 313.4 on 1 and 9838 DF,  p-value: < 2.2e-16
```

This regression output is for Team points and Free throw percentage which implies that the p-value is extremely low, which means that there is a statistically significant correlation between the variables. In addition to this we see that the R-squared value is 0.03 which states that 3% of the variation in our independent variable can be explained by our dependent variable, which is quite low.

```
Call:
lm(formula = TeamPoints ~ FieldGoals., data = data)

Residuals:
   Min     1Q  Median     3Q    Max
-29.302  -5.780  -0.473   5.355  46.469

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  31.5577     0.7257   43.48   <2e-16 ***
FieldGoals. 158.2101     1.5812  100.06   <2e-16 ***
---
Signif. codes:
0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 8.581 on 9838 degrees of freedom
```

Multiple R-squared: 0.5044, Adjusted R-squared: 0.5043
F-statistic: 1.001e+04 on 1 and 9838 DF, p-value: < 2.2e-16

This regression output is for Team points and field goal percentage which implies that the p-value is extremely low, which means that there is a statistically significant correlation between the variables. In addition to this we see that the R-squared value is 0.50 which states that 50% of the variation in our independent variable can be explained by our dependent variable, which is quite high. This regression output also proves that our initial hypothesis that field goal percentage is much more influential is true.

Call:
lm(formula = TeamPoints ~ FieldGoals. + FreeThrows. + TotalRebounds +
Assists + Steals + Blocks + Turnovers, data = data)

Residuals:

Min	1Q	Median	3Q	Max
-26.327	-4.868	-0.414	4.455	39.199

Coefficients:

	Estimate	Std. Error	t value	Pr(>|t|)	
(Intercept)	-7.14663	1.05920	-6.747	1.59e-11	***
FieldGoals.	147.64521	1.63293	90.417	< 2e-16	***
FreeThrows.	19.48983	0.68419	28.486	< 2e-16	***
TotalRebounds	0.52611	0.01194	44.080	< 2e-16	***
Assists	0.44127	0.01715	25.738	< 2e-16	***
Steals	0.52266	0.02469	21.169	< 2e-16	***
Blocks	-0.13719	0.02861	-4.795	1.65e-06	***
Turnovers	-0.55539	0.01887	-29.440	< 2e-16	***

Signif. codes:
0 ‘*’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1**

Residual standard error: 7.065 on 9832 degrees of freedom
Multiple R-squared: 0.6642, Adjusted R-squared: 0.664
F-statistic: 2778 on 7 and 9832 DF, p-value: < 2.2e-16

This regression output is the multivariate regression model and controls for other variables within our dataset. Because of this we can see that when adding in other variables there is a 66%

R-squared value which is much more predictive and adds more predictability. There are other variables within the dataset that could potentially make this value higher and allow for better analysis.

Discussion

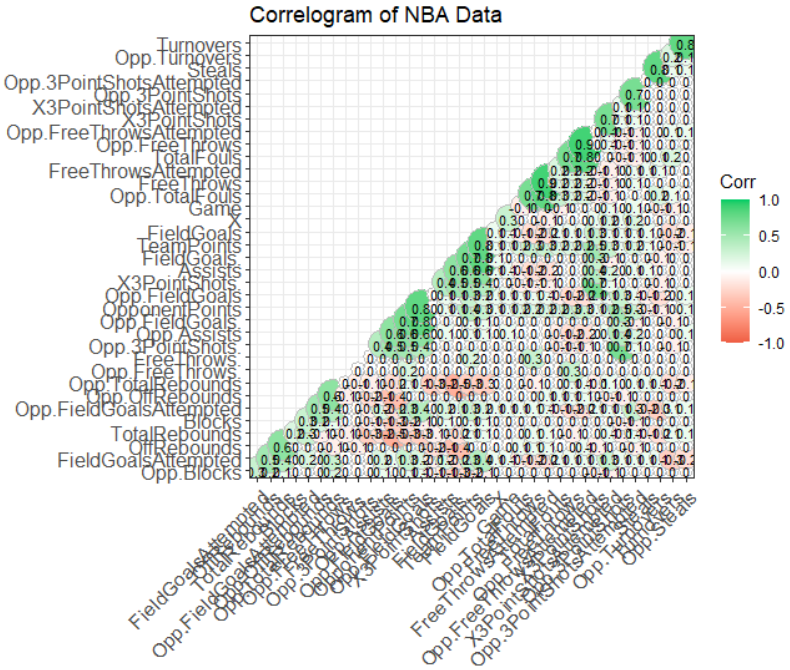
The results of my bivariate analysis between Teampoints and Free throw percentage and Teampoints and Field goal percentage supports my original hypothesis that field goal percentage is more influential on a team's overall scored points. We can see from both graphs that there is a positive trend upwards; however, the graph of field goal percentage is more closely fit around the line of best fit with a slightly stronger correlation. There appears to be no skews in my dataset heavily; however, one thing to consider would be to look at any potential outliers in variables. This would be interesting to see how many values would be considered outliers within the dataset, as this data is data that is gathered from real-life. In addition to this one consideration I had was to see if I could use winorloss percentage/ ratio as an independent variable; however, recoding the values was quite difficult and I also considered that if every win/loss is listed in the dataset there could be potential issues with analysis. I am unsure if including a win and a loss for both teams involved would cause potential problems within the regression model, thus I chose to focus on team points.

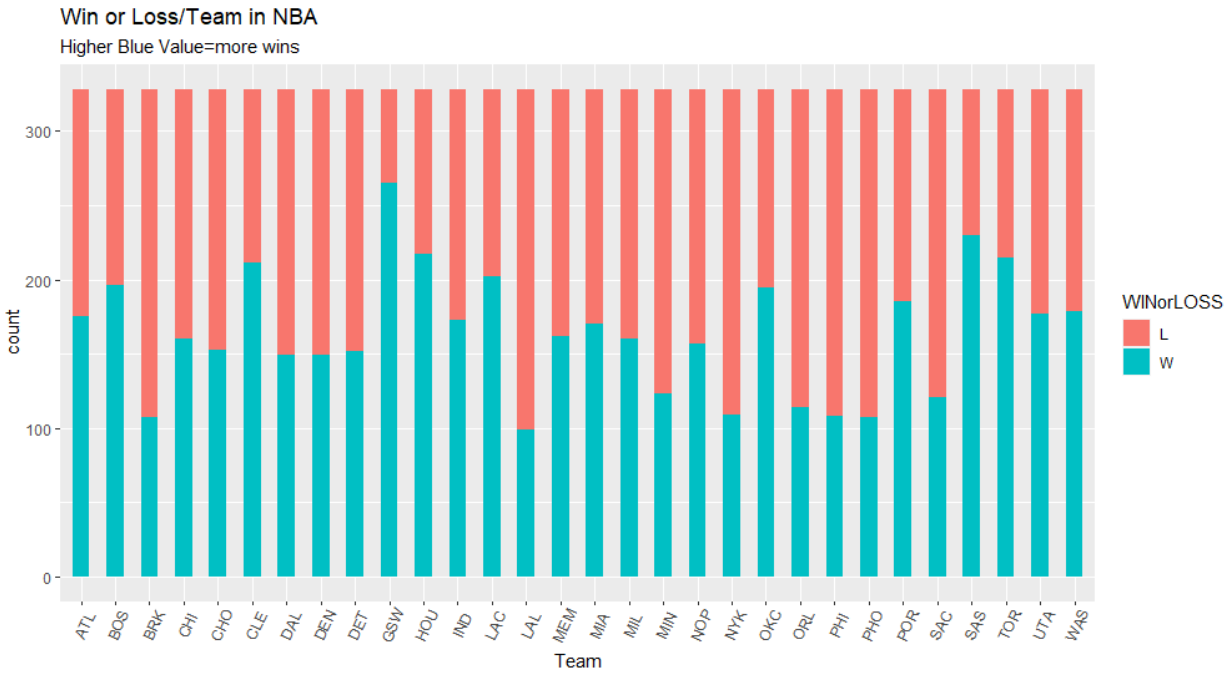
The results of my multivariate analysis are important because it helps allow control between external variables presented in the dataset and significantly improves the R-squared value in the regression model. I think that all of the variables used as control variables would help make the variables become more significant, which it ultimately did.

Conclusion

My goal in this research report is to find out what specific shooting perspective in a game of basketball is more predictive and better fit with a team's overall points scored. Through the use of regression models I was able to support the conclusion that field goal percentage was more predictive, even if very slightly, than free throw percentage. This can be important to know as an NBA player, fan, analyst of basketball, or young basketball player with aspirations of the NBA. This research report finds correlational information, but does not attempt to dig further as to why field goals may have more of an impact than free throws other than minor research and speculation. Future research could study whether a two-pointer or three-pointer would be more influential in a team's overall points. Also, future research could focus on how all of the variables used within this report could influence a win/loss percentage.

Appendix:


Correlogram of NBA Data
Turnovers
Opp.Turnovers
Steals
Opp.3PointShotsAttempted
Opp.3PointShots
X3PointShotsAttempted
X3PointShots
Opp.FreeThrowsAttempted
Opp.FreeThrows
TotalFouls
FreeThrowsAttempted
FreeThrows
Opp.TotalFouls
Game
X
FieldGoals
TeamPoints
FieldGoals.
Assists
X3PointShots.
Opp.FieldGoals
OpponentPoints
Opp.FieldGoals.
Opp.Assists
Opp.3PointShots.
FreeThrows.
Opp.FreeThrows.
Opp.TotalRebounds
Opp.OffRebounds
Opp.FieldGoalsAttempted
Blocks
TotalRebounds
OffRebounds
FieldGoalsAttempted
Opp.Blocks
Corr
1.0
0.5
0.0
-0.5
-1.0


Win or Loss/Team in NBA
Higher Blue Value=more wins
count
300
200
100
0
ATL
BOS
BRK
CHI
CHO
CLE
DAL
DEN
DET
GSW
HOU
IND
LAC
LAL
MEM
MIA
MIL
MIN
NOP
NYK
OKC
ORL
PHI
PHO
POR
SAC
SAS
TOR
UTA
WAS
Team
WINorLOSS
L
W

Call:
lm(formula = TeamPoints ~ FieldGoals, data = data)

Residuals:
Min 1Q Median 3Q Max
-20.4840 -4.5102 -0.4186 4.4636 29.4374

Coefficients:
Estimate Std. Error t value Pr(>|t|)
(Intercept) 25.4368 0.5215 48.77 <2e-16 *
FieldGoals 2.0262 0.0134 151.24 <2e-16 ***

Signif. codes:
0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1**

**Residual standard error: 6.685 on 9838 degrees of freedom
Multiple R-squared: 0.6992, Adjusted R-squared: 0.6992
F-statistic: 2.287e+04 on 1 and 9838 DF, p-value: < 2.2e-16**

This regression output is just for consideration into how field goals made might compare to field goal percentage.

Works Cited

"Basketball Block." *Rookie Road*, www.rookieroad.com/basketball/what-is-block/.

"Basketball Fouls." *Rookie Road*, www.rookieroad.com/basketball/fouls/.

"Basketball Rebounds." *Rookie Road*, www.rookieroad.com/basketball/stats/rebounds/.

"Basketball Turnover." *Rookie Road*, www.rookieroad.com/basketball/101/turnover/.

Brian. "What Is a Rebound in Basketball? Explained!" *Basketball Word!*, Basketball Word!, 5 June 2021, basketballword.com/what-is-a-rebound-in-basketball/.

MasterClass. "A Guide to Basketball Scoring: 3 Ways to Score in Basketball." *MasterClass*, MasterClass, 24 May 2021, www.masterclass.com/articles/basketball-scoring-guide.

"What Is a Steal in BASKETBALL? Definition & Meaning On Sportslingo." *What Is A Steal In Basketball? Definition & Meaning On SportsLingo*, www.sportslingo.com/sports-glossary/s/steal-basketball/.