

# What Predicts Good 3PT Shooting?

Matt Saraceno

2024-07-15

## Introduction

Three point shooting(3PT) is a critical part of modern basketball. The best teams in the National Basketball Associate(NBA) are all in an arms race to load up on the best three point shooters in the world. With so many players being evaluated on their ability to shoot the three ball, what data can be used to help predict a players ability to make three pointers?

The goal of this report is to predict a players ability to hit three pointers by looking at three different factors: free-throw(FT) percentage, mid-range percentage, and the position they play.

## Question #1: Does the ability to make free-throws influence three-point shooting?

We have all heard this before in one way or another, "A player's ability to make free throws is an indicator that a player is, or can be, a great shooter". This line of thinking has been repeated so many times across sports media that it has become basketball common knowledge, but is it true?

I have decided to take a look at last years shooting totals from all 572 NBA players, and see if there truly is an correlation between the two statistics.

The code below gives a quick preview of what our data looks like.

```
library(tidyverse)
```

```
## — Attaching core tidyverse packages — tidyverse 2.0.0 —
## ✓ dplyr      1.1.4      ✓ readr      2.1.5
## ✓ forcats    1.0.0      ✓ stringr    1.5.1
## ✓ ggplot2    3.5.1      ✓ tibble     3.2.1
## ✓ lubridate  1.9.3      ✓ tidyr      1.3.1
## ✓ purrr      1.0.2
## — Conflicts — tidyverse_conflicts() —
## * dplyr::filter() masks stats::filter()
## * dplyr::lag()     masks stats::lag()
## i Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to become errors
```

```
library(readxl)
mydata1 <- read_excel("D:/Data Analyst/3pt shooting/3pt_shooting.xlsx")
head(mydata1)
```

```
## # A tibble: 6 × 30
##      Rk Player    Pos    Age Tm      G    GS    MP    FG    FGA FG_per three_P
##    <dbl> <chr>    <chr> <dbl> <chr> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl>
## 1     1 Precious... PF      24 TOT     74    18 1624   235   469 0.501     26
## 2     2 Bam Adeb... C       26 MIA     71    71 2416   530  1017 0.521     15
## 3     3 Ochai Ag... SG      23 TOT     78    28 1641   178   433 0.411     62
## 4     4 Santi Al... PF      23 MEM     61    35 1618   247   568 0.435    106
## 5     5 Nickeil ... SG      25 MIN     82    20 1921   236   538 0.439    131
## 6     6 Grayson ... SG      28 PHO     75    74 2513   340   682 0.499    205
## # i 18 more variables: three_PA <dbl>, threeP_percent <dbl>, two_P <dbl>,
## #   two_PA <dbl>, twoP_per <dbl>, eFG_per <dbl>, FT <dbl>, FTA <dbl>,
## #   FT_percent <dbl>, ORB <dbl>, DRB <dbl>, TRB <dbl>, AST <dbl>, STL <dbl>,
## #   BLK <dbl>, TOV <dbl>, PF <dbl>, PTS <dbl>
```

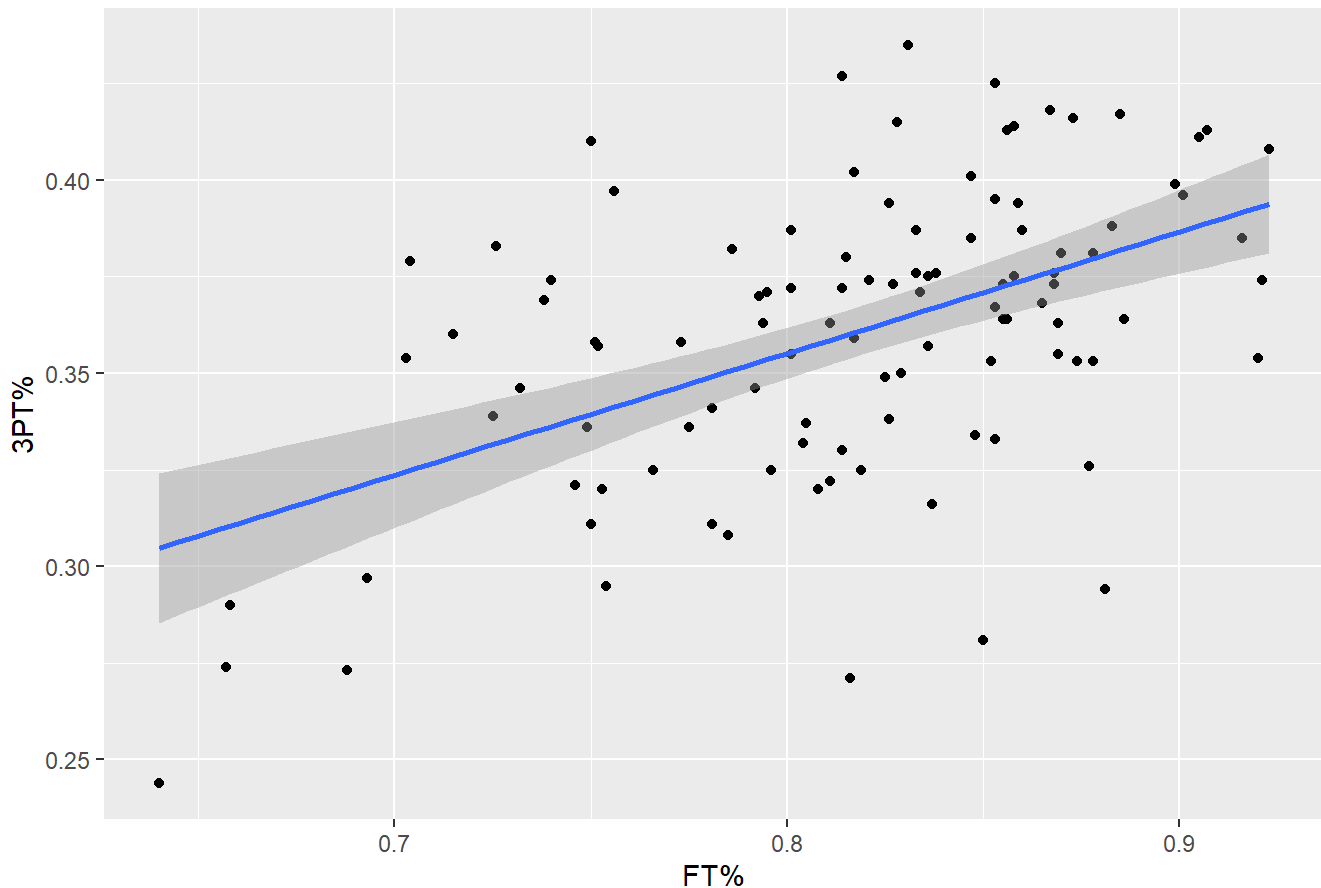
Before we can start to do any analysis, we must first filter out some data. Not every NBA player this season shot enough FTs or three-pointers to give us any meaningful data. This report will look at only players that have attempted at least 150 FTs and attempted at least 85 three pointers.

Below is a graph of charting the two statistics:

```
filtered_mydata1 <- mydata1[mydata1$FTA >= 150 & mydata1$three_PA >=85, ]
ggplot(filtered_mydata1, aes(x = FT_percent, y = threeP_percent))+
  geom_point()+
  ggtitle("3PT% based on FT%")+
  xlab("FT%")+
  ylab("3PT%")+
  geom_smooth(method = 'lm')
```

```
## `geom_smooth()` using formula = 'y ~ x'
```

## 3PT% based on FT%



Out of the 572 initial players that were observed, only 106 made the cut after our filtering. The median FT percentage from this group of players is 81.62% and their 3PT shooting percentage is 36.03%. Those are some pretty solid numbers from some of the best players in the world.

The initial thoughts from examining this graph, is that there seems to be a positive correlation between the two stats. A trend line is added to the graph to provide the audience the direction of the data. The data is not heavily correlated due to the random nature that the data presents. That is reinforced by our correlation coefficient.

```
cor(filtered_mydata1$FT_percent, filtered_mydata1$threeP_percent)
```

```
## [1] 0.5059518
```

Correlation coefficient provides a measurement to determine if there is any significant relationship between two variables. The closer the correlation coefficient is to 1, the stronger the relationship is. The correlation coefficient for this data is .505918. A correlation coefficient of less than .7 is not strong enough to infer that high FT percentage is linked to good three point shooting.

## Question #2: Does a player's ability to make mid-range jump shots influence a player ability to make three-pointers?

The next variable of interest is the mid range jump shots. For this study, shooting data by distance was used to help determine mid range shoots. Shoots taken between 10-16 feet was labeled as a mid-range. Just like with the free throw study above, the data used to answer this question was from shooting data of 572 NBA players from

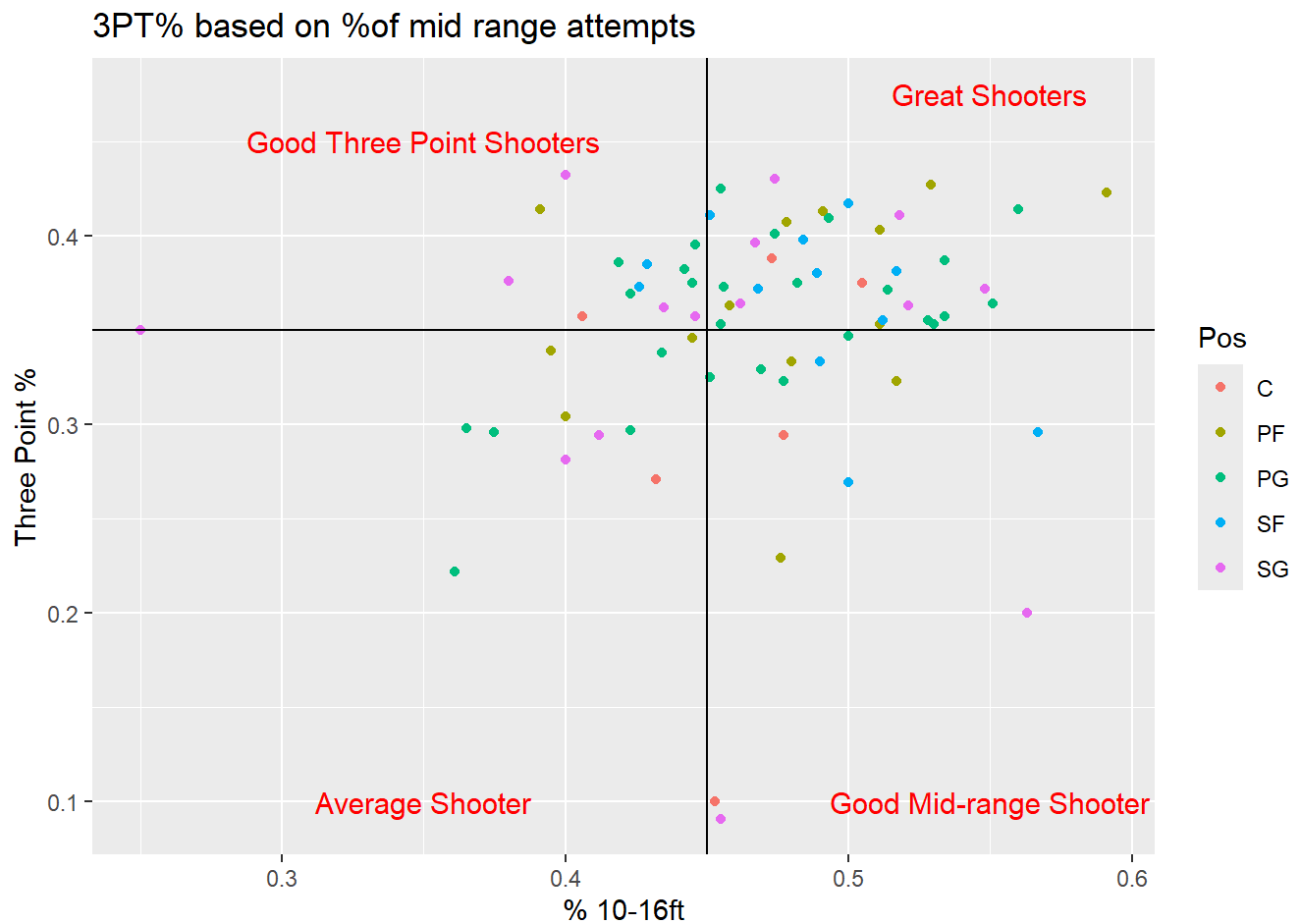
the 2023-24 season. Then the data was filtered to look at players who took at least 10% of the shots from 10-16 feet and played at least 25 games. Some outliers were removed to clean up the data as well.

```
mydata2 <- read_excel("D:/Data Analyst/3pt shooting/shooting_advanced.xlsx")
filtered_mydata2 <- mydata2[mydata2$ten_sixteenAperbyDistance >= .125 & mydata2$Games >=
25, ]
filtered_mydata3 <- filtered_mydata2 %>% filter(!row_number() %in% c(53, 68))
filtered_mydata4 <- subset(filtered_mydata3, !is.na(filtered_mydata3$threep_fgperbyDistance))
head(filtered_mydata4)
```

```
## # A tibble: 6 × 29
##      Rk Player      Pos      Age Tm      Games      MP `FG%` Dist. twoPA_perbyDistance
##    <dbl> <chr>      <chr> <dbl> <chr> <dbl> <dbl> <dbl> <dbl>
## 1      2 Bam Adeb... C        26 MIA      71  2416 0.521  8.5      0.959
## 2     10 Kyle Ande... PF        30 MIN      79  1782 0.46    9      0.889
## 3     13 Cole Anth... PG        23 ORL      81  1817 0.435 14.9      0.64
## 4     17 Deandre A... C        25 POR      55  1784 0.57    8.9      0.987
## 5     25 Paolo Ban... PF        21 ORL      80  2799 0.455 12.6      0.751
## 6     28 Dominick ... PF        20 SAS      33   420 0.496  7.4      0.92
## # i 19 more variables: zero_3AperbyDistance <dbl>,
## #   three_tenAperbyDistance <dbl>, ten_sixteenAperbyDistance <dbl>,
## #   sixteen_threepAperbyDistance <dbl>, threepAperbyDistance <dbl>,
## #   twop_fgperbyDistance <dbl>, zero_threefgperbyDistance <dbl>,
## #   three_tenfgperbyDistance <dbl>, ten_sixteenfgperbyDistance <dbl>,
## #   sixteen_threepfgperbyDistance <dbl>, threep_fgperbyDistance <dbl>,
## #   twop_perass <dbl>, threep_perass <dbl>, dunks_perFGA <dbl>, ...
```

After the filtering is applied, we are left with 74 players that will be elevated to determine if the ability to make mid range shots has any influence on a players ability to hit three pointers. We will replicate the same exercise that we did in question #1. First, we will plot the data and then we will look at the correlation coefficient to determine if there is any relationship between the two variables.

```
ggplot(filtered_mydata4, aes(x = ten_sixteenfgperbyDistance, y = threep_fgperbyDistance)) +
  geom_point(aes(color = Pos)) +
  ggtitle("3PT% based on % of mid range attempts") +
  xlab("% 10-16ft") +
  ylab("Three Point %") +
  geom_hline(yintercept = .35) +
  geom_vline(xintercept = .45) +
  annotate("text", x = .55, y = .475, label = "Great Shooters", col = "red", size = 4) +
  annotate("text", x = .35, y = .45, label = "Good Three Point Shooters", col = "red", size = 4) +
  annotate("text", x = .35, y = .1, label = "Average Shooter", col = "red", size = 4) +
  annotate("text", x = .55, y = .1, label = "Good Mid-range Shooter", col = "red", size = 4)
```



The graph above shows a players three point percentage in relation to their mid range percentage. The data points are random in nature with each color representing the position that that player plays. The plot is also broken up into four different quadrants to help viewers understand how well a player shoots based on their position on the plot.

The initial thoughts from the graph is there will be little to no correlation between a players ability to make a mid range shots and their ability to make three pointers.

```
cor(filtered_mydata4$ten_sixteenfgperbyDistance, filtered_mydata4$threep_fgperbyDistance)
```

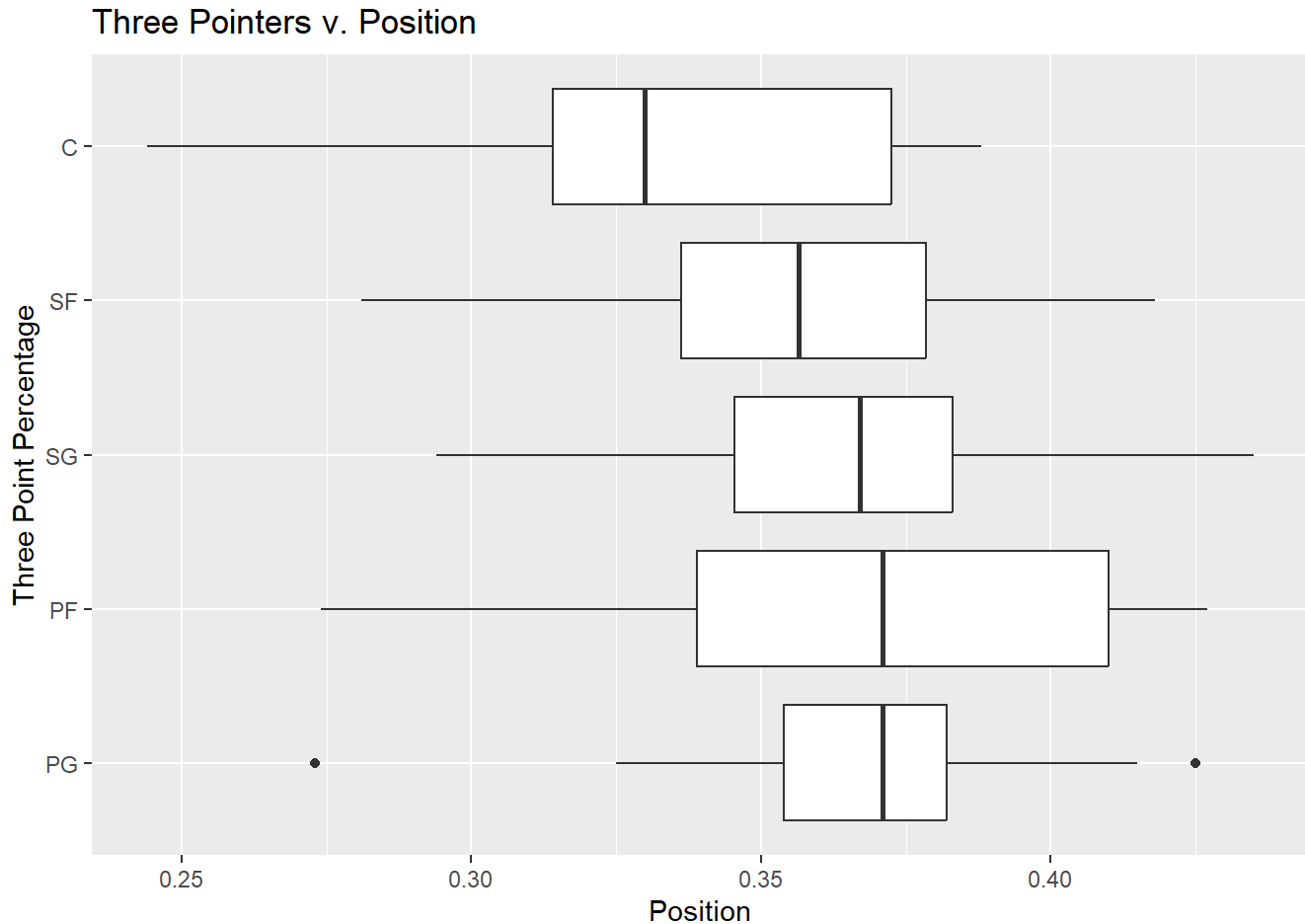
```
## [1] 0.1603543
```

The correlation coefficient is .160, way to low to infer that there is any association between the two variables.

### Question #3: Does a player's position have any influence on their ability to hit three pointers?

Finally, we will look at a player's position on the court, and if it has any influence on their ability to shoot three pointers. Does a player's position influence their training enough where one position would shoot much better than an other?

```
ggplot(filtered_mydata1, aes(x = reorder(Pos, -threeP_percent), y = threeP_percent))+
  geom_boxplot()+
  ggtitle("Three Pointers v. Position")+
  xlab("Three Point Percentage")+
  ylab("Position")+
  coord_flip()
```



The graph above uses the same filtered data from question #1. These box plots help demonstrate the spread of data, but ultimately this allows us to compare averages across each position.

Below is a table of the average three point percentage by position.

Position		Average Three point percentage
PG	.368	
PF	.368	
SG	.365	
SF	.356	
C	.335	

The dark lines within each box represents the average three point percentage from each position. The box represents the Interquartile range, or the 50% of the data. The the dark lines that look like tails, are the upper and lower limits of the data. To no surprise, the best shooting position is point-guard(PG), and Center(C) in last place.

The second best three point shooting position is PF, but they are a position with the most spread.

This plot does not however demonstrate whether there is an association between position and three point shooting. To do that we will use the Kruskal-Wallis test. This test allows to determine if there are statically significant differences between two or more groups. In our case the groups are positions on a basketball court, and the dependent variables are the shooting percentages of those positions.

```
kruskal.test(filtered_mydata1$threeP_percent~filtered_mydata1$Pos)
```

```
##  
##  Kruskal-Wallis rank sum test  
##  
## data:  filtered_mydata1$threeP_percent by filtered_mydata1$Pos  
## Kruskal-Wallis chi-squared = 6.5341, df = 4, p-value = 0.1627
```

From this test, we are returned several values but the one that is most important is the one titled “p-value”. The p-value is a number that describes how significantly likely a participle set of observations are to occur given the null hypothesis to be true. What we are looking for here is for our p-value to be less than .05. To keep things simple, the p-value for this experiment needs to be less than .05 for us to confidently say that the position of a player significantly impacts their ability to make three pointers. The p-value for our test is .1627, greater than .05, so no. There is no significant impact on three point shooting based on the position of a player

## Rebuttal

There are several limitations to this study that need to be addressed. First, there is a lack of data. There are limitations to using only one season. There could be contributing factors that influence the data. For example, the Covid lock down season or even shortened seasons. Secondly, correlation coefficients has several pitfalls and does not perfectly describe the relationship between two variables. As the famous saying goes, “correlation does not imply causation”.

## Conclusion

In conclusion, this report aimed to determine if any statistical data could evaluate a players ability to make three pointers at a high level. This report looked at common associations between variables such as free-trow and mid-range shooting, and uncommon associations like player positions.

Making three pointers at a high level is a difficult task. There is a lot that goes into it, and ultimately no data point can determine how well someone can make three pointers better than the eye test. To truly understand a player, you must watch and appreciate what they do. Basketball is a beautiful sport, and it is best enjoyed when one can be part of it.