d

Project: NBA Playoffs Teams

Professor: Stela Maris Salvatierra Galano

Group members: Marcelo Gomez Vives, Luis Álvarez-Cascos López,Alejandro Saiz Guerricagoitia and Santiago Uranga Gayol

Structure of the analysis

1. **Description of the project and dataset justification**

During this analysis, we will examine the most defining variables of the players who demonstrate their ability and skill on the court during the 2021-2022 playoffs. From established stars to emerging stars, our goal is to analyze and explain which are the most important variables for the team or for decision making . The dataset we have worked with is based on NBA statistics 2021-2022 and thus, can be found in:

https://www.kaggle.com/datasets/vivovinco/nba-player-stats?select=2021-2022+NBA+Player+Stats+-+Regular.csv

2. The **variables used** are defined as follows:
   - **Position**: refers to the specific role they typically play on the court during a game.
   - **Age**: the age of NBA players refers to the number of years elapsed since their birth date up to the current date.
   - **Team**: refers to the name of the club each player plays for.
   - **MP**: accounts for the number of minutes played by each one of the players.
   - **FG**%: stands for the Percentage of Field Goals per Game of each player.
   - **3P**%: "3P%" means "Three-Point Percentage," which is a metric used to assess how accurate a player or team is at making three-pointers.
   - **TRB**: total rebound variable (TRB) is a quantitative variable defined as the total number of rebounds after shooting in a game. This is a crucial aspect of basketball, as it involves gaining possession of the ball after a missed goal attempt.
   - **AST**:the variable **"Assists (AST)"** is a quantitative variable defined as the number of assists made by a player in a game for the NBA.
   - **PF**:"personal fouls" refer to the number of fouls committed by a player or a team during a game or a playoff series.
   - **PTS**:the National Basketball Association (NBA) uses the statistic "points per game" (PPG) to describe the average number of points a player scores in each game they participate in over a specific period of time.

3. **Univariate analysis**:

3.1 Position

The **variable `Position´** refers to the role of the player in the team. There are 5 positions:Center(C), Power Forward(PF), Point Guard(PG),Small Forward(SF),and Shooting Guard(SG). It is defined as a Qualitative variable as it has a non-numerical nature, also you only have a limit of options.

Position	Frecuency	Relative frecuency
C	40	0.184
PF	50	0.230
PG	39	0.180
SF	39	0.180
SG	49	0.226
TOTAL	217	1

We can see that the most repeated position**(mode)** in the list of players on the playoffs is the Point Forward. As expected, most of the positions incorporate a similar number of players, the relative frequency moves around mostly between 18% and 23%.

3.2 Age

The **variable `Age´** in this case is a quantitative variable defined as the age of the players during the 2021-2022 playoffs.

The **minimum** age is found at 19 years represented by 3 players out of 217.The **maximum** age based on the dataset an NBA player is playing now is 38 years with only one player. This is the range in which the Age variable moves.

The mean is 26.59. This tells us that the average player in the playoffs is about 27 years old. The most repeated age **(mode)** is 24, we have 21 players this old.

The **maximum** age based on the dataset an NBA player is playing now is 38 years with only one player.

Age	Total
19	3/217
20	7/217
21	12/217
22	19/217
23	18/217
24	21/217
25	20/217
27	15/217
28	13/217
29	20/217
30	7/217
31	11/217
32	10/217
33	10/217
34	3/217
35	7/217
36	3/217
38	1/217

(*The following table summarizes the total amount of players of each age*)

The range is defined as the totally different ages of all players, this is = [18,19,20,21,22,23,...,31,32,33,34,35,36,37,38].

(Note: these values represent the number of players of each of the quartiles compared to the total number of players)

```
> summary(data_set_nba_playoffs$Age)
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max. 
  19.00   23.00   26.00   26.59   29.00   38.00 
```

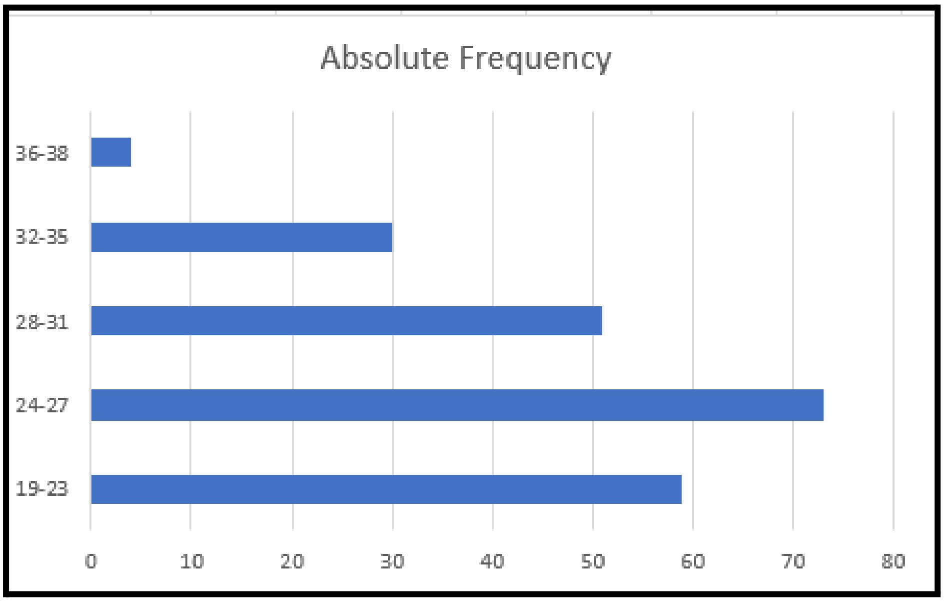
To calculate all the data mentioned above we have used the summary function in R Studio.

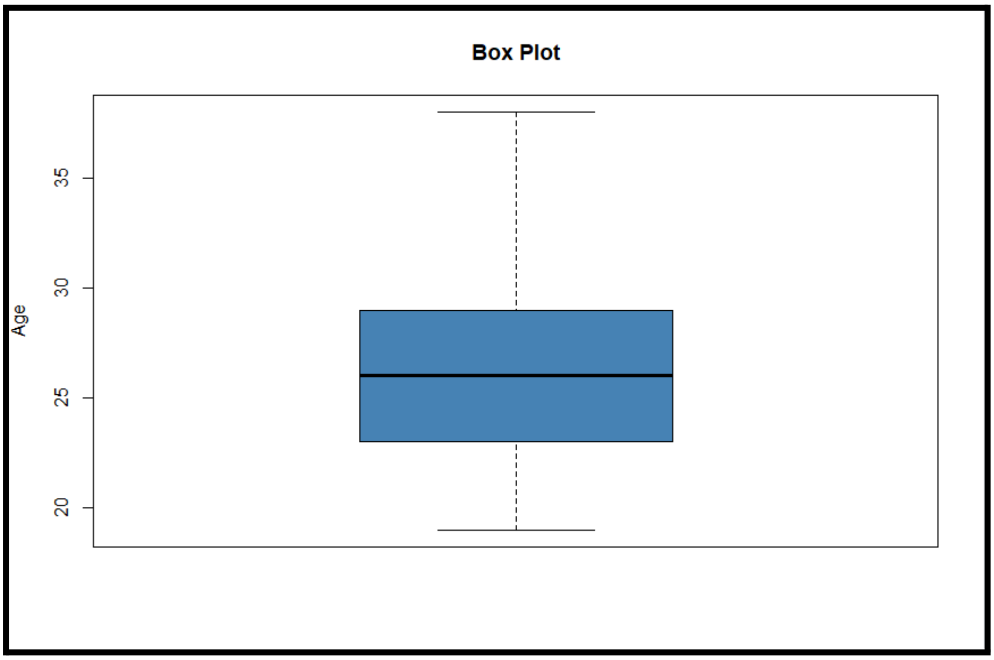
<u>Calculation and explanation of the variance and standard deviation</u>

The **variance** of the age variable is 17.96527 square years. The **standard deviation**, which is the square root of the variance is 4.238545 years. This tells us that most players are within approximately 4.24 years of the mean age. We can see that this variation is moderate. The **coefficient of variation** is the Standard deviation / Mean = 4.238545 / 26,59 = 0.159. In this case means that there is about 15.9% of variability from the mean.

To calculate the frequency table, we first ask R to count the number of players with the same age. Now, with this information we created the following graphs to understand the data better in a visual manner.

Age	Absolute Frequency	Relative Frequency
19-23	59	0.271889401
24-27	73	0.33640553
28-31	51	0.235023041
32-35	30	0.138248848
36-38	4	0.01843318
Total	**217**	**1**



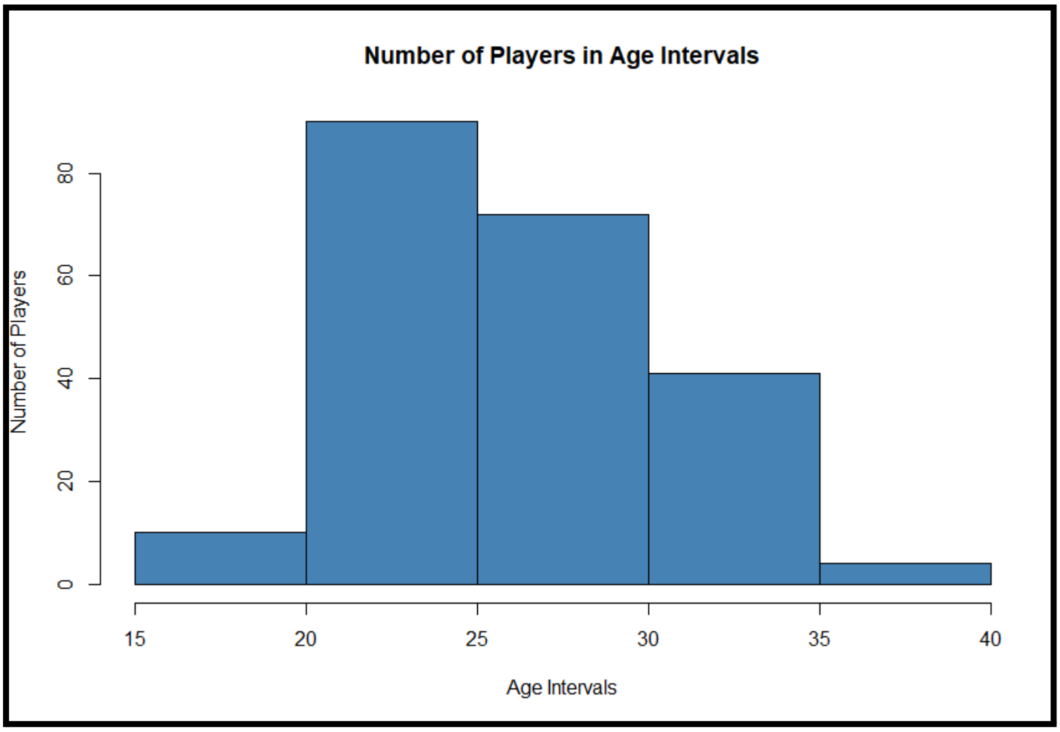


From the box plot we can see how the quartiles regarding players with same age are distributed among the dataset.

The value of 1st quartile(Q1) is 59/217 (less or equal to 23 years) and the 3rd quartile/(Q3) is 48/217 (less or equal to 29 years)

In the box plot we can see the distribution of the quartiles. The first one is the lower line of the blue box. The median is the Black line in the middle of the box. And the Third Quartile is the upper line.

This box plot shows us that the variable is not symmetrical, because the whiskers are not equally portioned.



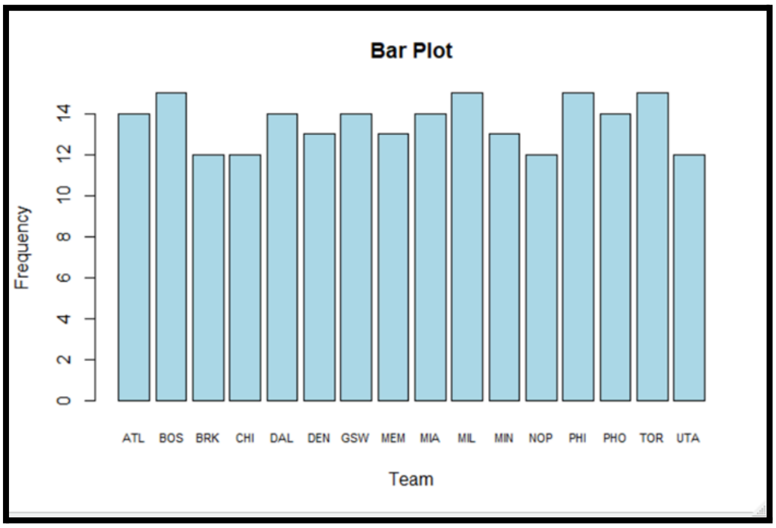
Explanation/interpretation of the histogram chart:

The following graph shows the intervals age in the Y-axis and the total number of teams and the y-axis the ages intervals based on the data expressed above.

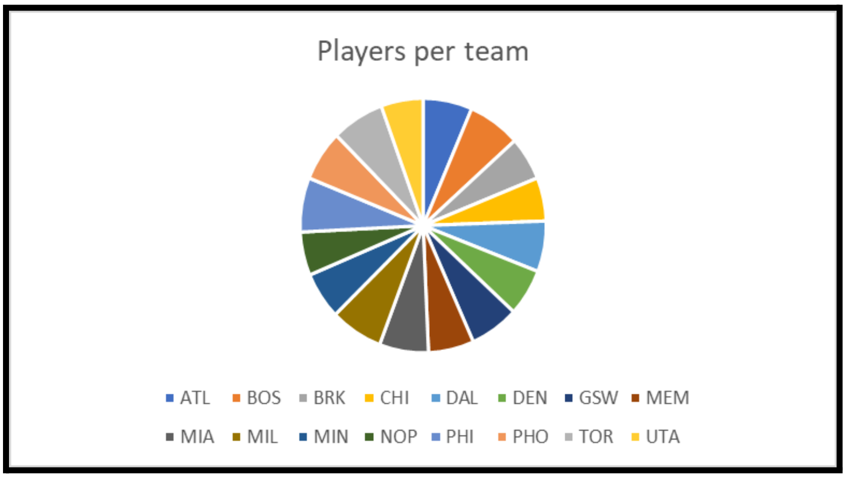
From the picture we can see that the higher number of players is concentrated in the intervals of ages 20-25 years. One possible interpretation of this data lies in the fact that NBA teams tend to look for players who have better physical conditions and performance in matches than the older ones. The reason behind this is because they can perform better in matches and give a competitive advantage to their teams.

3.3 Team

The **variable Team** is a qualitative variable that measures the teams of each one of the players of the NBA playoffs. There are 16 teams analyzed.



Team	#players	Relative frecuency
ATL	14	0.065
BOS	15	0.069
BRK	12	0.055
CHI	12	0.055
DAL	14	0.065
DEN	13	0.060
GSW	14	0.065
MEM	13	0.060
MIA	14	0.065
MIL	15	0.069
MIN	13	0.060
NOP	12	0.055
PHI	15	0.069
PHO	14	0.065
TOR	15	0.069
UTA	12	0.055
TOTAL	**217**	**1**



We made a frequency table of the variable, and we also can see a pie chart and a Bar Plot made out of it. The biggest number of players in the same team is 15. We do not have a visible **mode**, because there are multiple teams with this same value of players.

3.4 MP

The **variable `MP´** is a quantitative variable that measures the minutes played of each one of the players by match.

The minimum is found in 0 since there are players who do not play since they might not be worth enough to contribute to that specific match or are suffering from an injury. The maximum can be found in 44 minutes played.

The total number of minutes played by all players from the dataset by match is 4216.1 minutes (or nearly 70 hours).

The mean, in this case, is the average number of minutes played per game by each player which is 19.43 minutes. This means that most of the players play around 19.43 minutes

The variance is 165.8 square minutes. This means that the minutes played per match by NBA players vary widely from the average.

```
> summary(data_set_nba_playoffs$MP)
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
   0.00    7.60   18.40   19.43   31.20   44.00
```

The range is defined in the set of minutes S= {0,…,44}. (Assumption: all values included)

<u>Calculation and explanation of the variance and standard deviation</u>

The standard deviation is 12.87 minutes meaning teams might have a well-defined starting lineup and rotation pattern, leading to stable playing time for most players.

The coefficient of variation for the variable MP is 0.662 or 12.87 / 19.43. Since the c.v is high, this means that the dataset is less predictable or steady, with individual data points that might deviate greatly from the mean.

To calculate the frequency table, we first ask R to count the number of players with the same age.

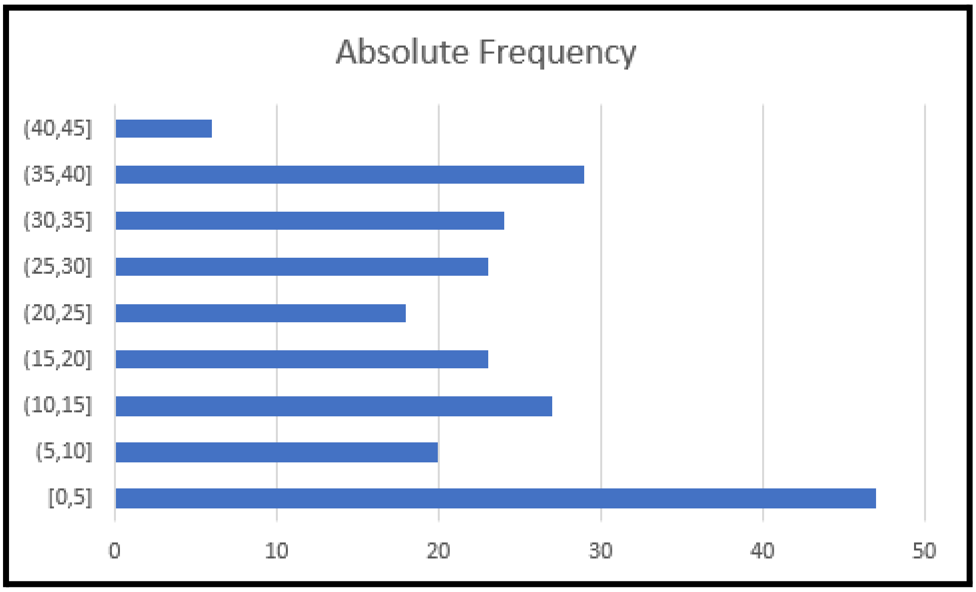
```
> # Define breaks for minutes played intervals
> minutes_breaks <- seq(0, 44, by = 5)  # Adjust the interval width as needed
>
> # Categorize minutes played into intervals
> data_set_nba_playoffs$MinutesInterval <- cut(data_set_nba_playoffs$MP, breaks = minutes_breaks, include.lowest = TRUE)
>
> # Create a frequency table
> minutes_frequency_table <- table(data_set_nba_playoffs$MinutesInterval)
>
> # Print the frequency table
> print(minutes_frequency_table)

  [0,5]  (5,10] (10,15] (15,20] (20,25] (25,30] (30,35] (35,40]
     47      20      27      23      18      23      24      29
```

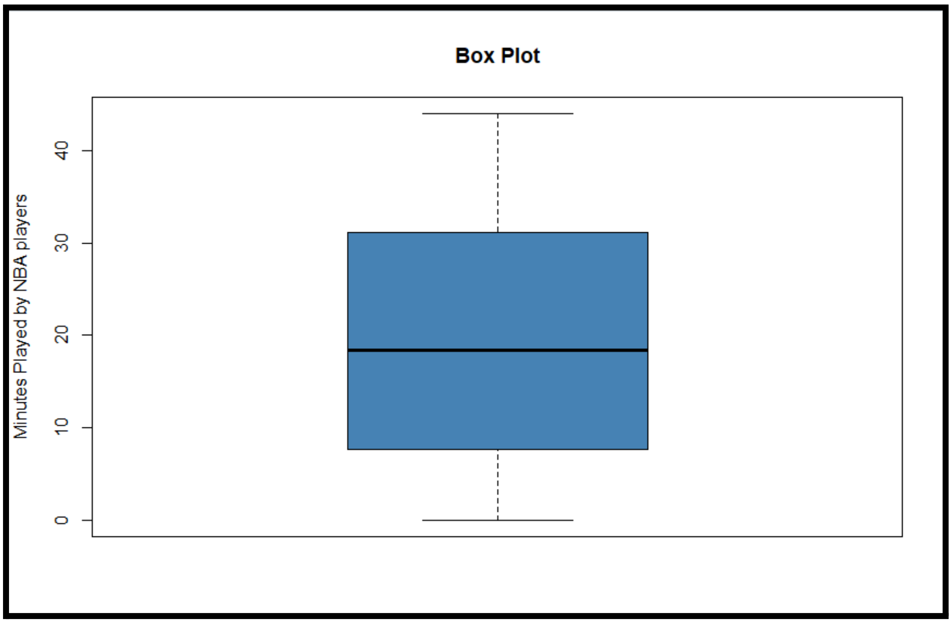
With this information we go to excel and plug the information but instead of working one by one we use intervals to simplify the example.

MP	Absolute Frequency	Relative Frequecy
[0,5]	47	0.216589862
(5,10]	20	0.092165899
(10,15]	27	0.124423963
(15,20]	23	0.105990783
(20,25]	18	0.082949309
(25,30]	23	0.105990783
(30,35]	24	0.110599078
(35,40]	29	0.133640553
(40,45]	6	0.02764977
	217	**1**

This table allows us to represent the Absolute Frequency graphically



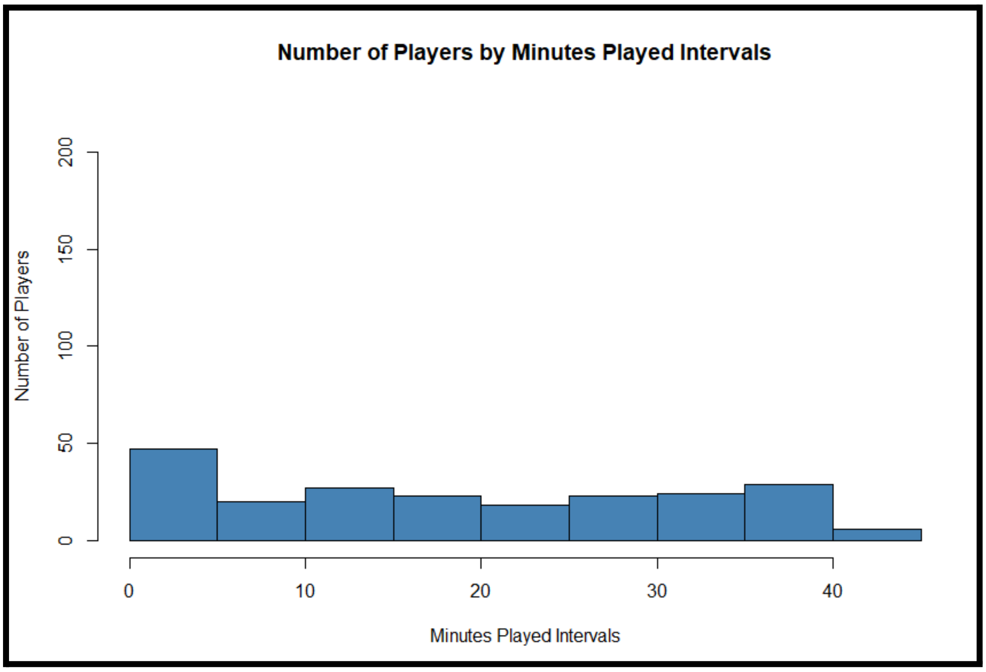
Now, we create the box plot and the histogram for the minutes played



The first quartile is found in 7,60 minutes meaning that the first 25% of all players play less or the same time.The median is located as the middle black line in the blue box, this means that from that The third quartile is found in 31,20 minutes meaning that the third 25% of all players play less or the same time.

This box plot indicates how the quartiles are divided, how the Minutes Played by NBA players are divided. This shows that because teams in Playoffs tend to be the best ones, players play a significant amount of time compared to other teams that do not reach the Playoffs. We can see this variable is not symmetrical, the whiskers and median line are not equally divided.

The following table summarizes the total number of minutes played and the ages of the players inside each minute's intervals.



From the histogram we can see that the highest number of minutes played is found between 0 and 5 minutes with 50 players.

```
> # Define breaks for minutes played intervals, ensuring it covers the range of MP values
> minutes_breaks <- seq(floor(min_minutes / 5) * 5, ceiling(max_minutes / 5) * 5, by = 5)
>
> # Create the histogram
> hist(data_set_nba_playoffs$MP,
+      breaks = minutes_breaks,
+      col = "steelblue",
+      xlab = "Minutes Played Intervals",
+      ylab = "Number of Players",
+      main = "Number of Players by Minutes Played Intervals",
+      xlim = c(0, max(minutes_breaks)),  # Set x-axis limits
+      ylim = c(0, 220),  # Set y-axis limits (adjust 220 based on your dataset),
+      axes = TRUE)  # Show axes
```

3.5 FG%

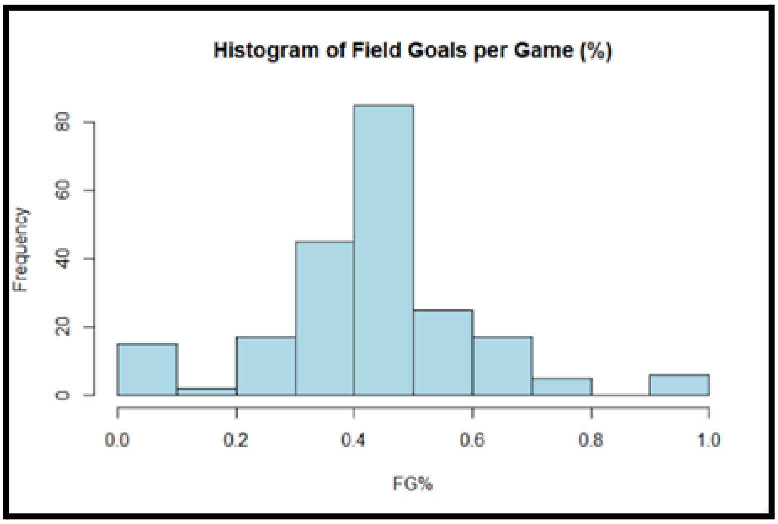
The variable **FG%** stands for the ***Percentage of Field Goals per Game*** in the playoffs. Field Goals are every point scored by a player such as three point and two point shots, with the exception of foul shots. This numerical variable is classified as a quantitative one.

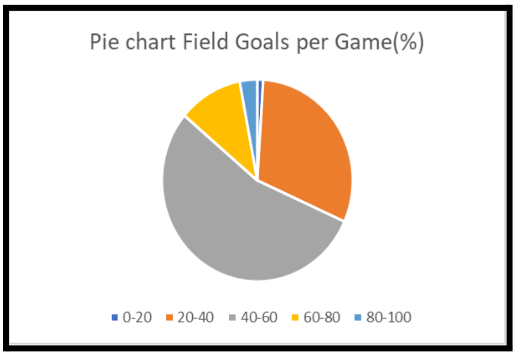
Then we calculated with Rstudio some important numbers to be more precise when analyzing this variable. The **mean** is 0.44, this tells us that the average of field goals a player gets per game is 44%. With regards to the **median** (0.44), we can see that roughly half of the players have a percentage of field goals of 44 or below and the other half 44% or more.

We calculated a **Standard Deviation** of 0.185 and a **Variance** of 0.03. This low number tells us that the majority of the field goals per game are very close to the **mean (0.44),** so a majority of the players in those playoffs have field goals per game very close to 44%. The **Coefficient of Variation** is 23.54%. This implies that the variability around the average is not too big, confirming what we said before.

The following is a **Frequency Table** we made by dividing the values of the FG% on intervals. Afterwards, we developed a Pie chart and a Histogram based on this table. It indicates the Absolute and Relative frequency of the variable.

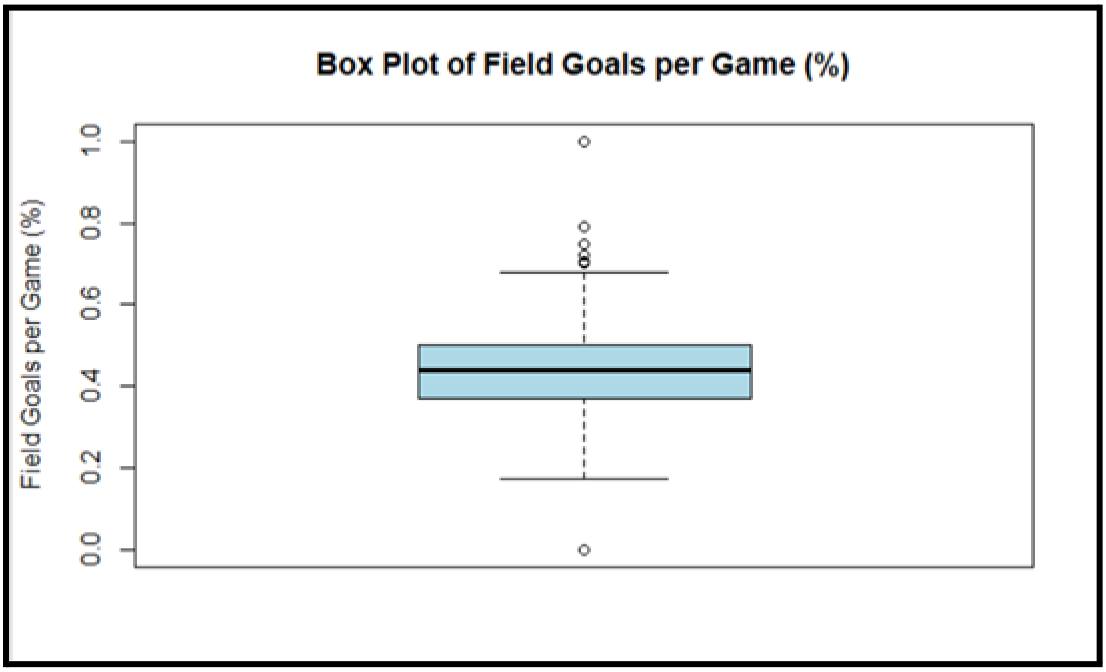
FG%	Absolute frecuency	relative frecuency
0-20	2	0.010
20-40	62	0.307
40-60	110	0.545
60-80	22	0.109
80-100	6	0.030
TOTAL	202	1





As we can see in the frequency table and more clearly in the Histogram and Pie Chart, the biggest concentration of players is between the 40 and 60 percent interval. That means that 110 players had between 40-60% of their total points being made by field goals. We can also see that only a few players actually get considerably more than half of their total points per game by field goals. This data tells us how important the foul shots surprisingly are.

Finally, we had Rstudio create us a **Box plot** of the variable, where we can see the mean and the quartiles of the Field Goals per Game(%).



Using the Summary Function we got most of our values for this analysis. The **first quartile** is 0.37, as we can also see in the box plot as the lower border of the light blue box. This means that the 25% of the players have below 37% of their Points per game being Field Goals. Next, we can see the **median** of 0.44 as the middle line in the box. Finally, we have a **third quartile** of 0.5, the upper border of the box. This means that only 25% of the players have 50% of their total points or more being Field Goals. This coincides with the analysis of the frequency table, as we can see that 75% of the players have below 50%, so field goals are not that important in the whole total. This variable looks symmetrical according to the box plot, because the median is right in the middle of the box and the whiskers are proportional.

This are the R formulas we used for this variable.

```
> boxplot(data_set_nba_playoffs$`FG%`,
+         main = "Box Plot of Field Goals per Game (%)",
+         ylab = "Field Goals per Game (%)",
+         col = "lightblue",
+         border = "black")
```

```
> summary(data_set_nba_playoffs$`FG%`)
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max. 
 0.0000  0.3700  0.4390  0.4375  0.5000  1.0000 
> sd(data_set_nba_playoffs$`FG%`)
[1] 0.1845805
> var(data_set_nba_playoffs$`FG%`)
[1] 0.03406996
```

```
> hist(data_set_nba_playoffs$FG_Intervals,
+      main = "Histogram of Field Goals per Game (%)",
+      xlab = "FG%",
+      ylab = "Frequency",
+      col = "lightblue",
+      border = "black")
```

```
> "Frecuency Table FG%"=table(data_set_nba_playoffs$FG_Intervals)
```

3.6 3P%

The variable 3P% is a continuous and quantitative variable as it takes values from 0 to 100 including decimal numbers.

When referring to the NBA playoffs, "3P%" means "Three-Point Percentage," which is a metric used to assess how accurate a player or team is at making three-pointers. The formula for calculating it is to divide the total number of three-pointers made by the total number of attempts, then compute the percentage.

This stat should be analyzed for several reasons such as Scoring Efficiency, Strategic Advantage, Comparative Analysis, and Spacing and Floor Stretching.

All things considered, 3P% is a useful metric in the NBA since it indicates a team's or a players's long-range shooting accuracy, which has a big influence on the results of games and series. It offers more insight into a player's scoring ability and a team's offensive potential than just field goal percentage.

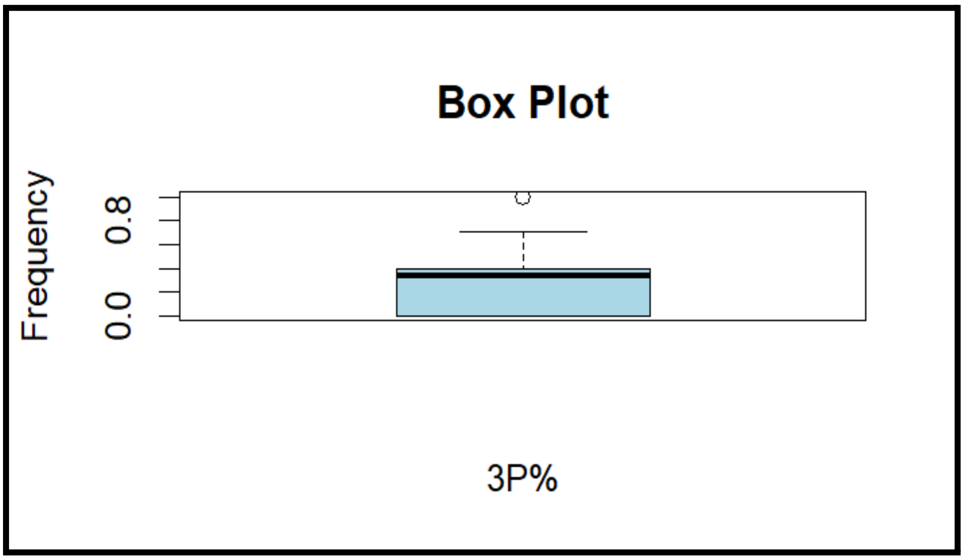
interval	frequency	Relative Frequency
[0,0.1)	57	0,262672811
[0.1,0.2)	8	0,036866359
[0.2,0.3)	34	0,156682028
[0.3,0.4)	68	0,313364055
[0.4,0.5)	27	0,124423963
[0.5,0.6)	17	0,078341014
[0.6,0.7)	2	0,00921659
[0.7,0.8)	1	0,004608295
[0.8,0.9)	0	0
[0.9,1)	3	0,013824885
total	217	1

```
summary(data_set_nba_playoffs_el_bueno$`3P%`)
  Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
0.0000  0.0000  0.3310  0.2733  0.3930  1.0000
sd(data_set_nba_playoffs_el_bueno$`3P%`)
[1] 0.2010606
var(data_set_nba_playoffs_el_bueno$`3P%`)
[1] 0.04042537
```

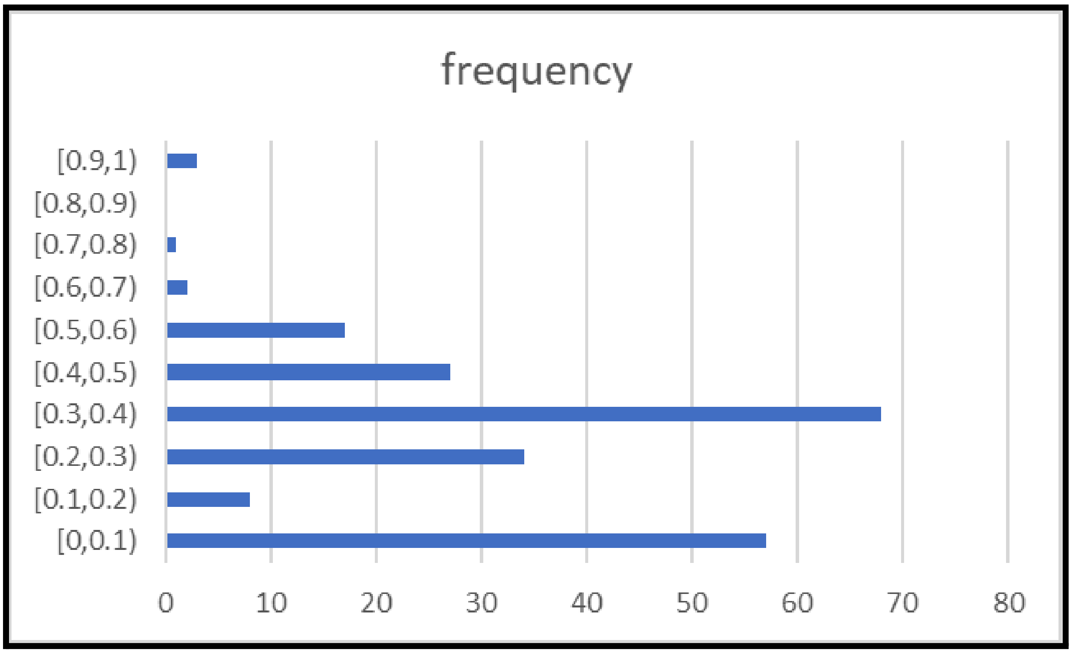
When analyzing 3P%, the variance helps to assess the variability of a player´s shooting performance, which has important implications for player evaluation, game strategy, and how the team performs. Players with low variance are often considered reliable threats from beyond the arc.

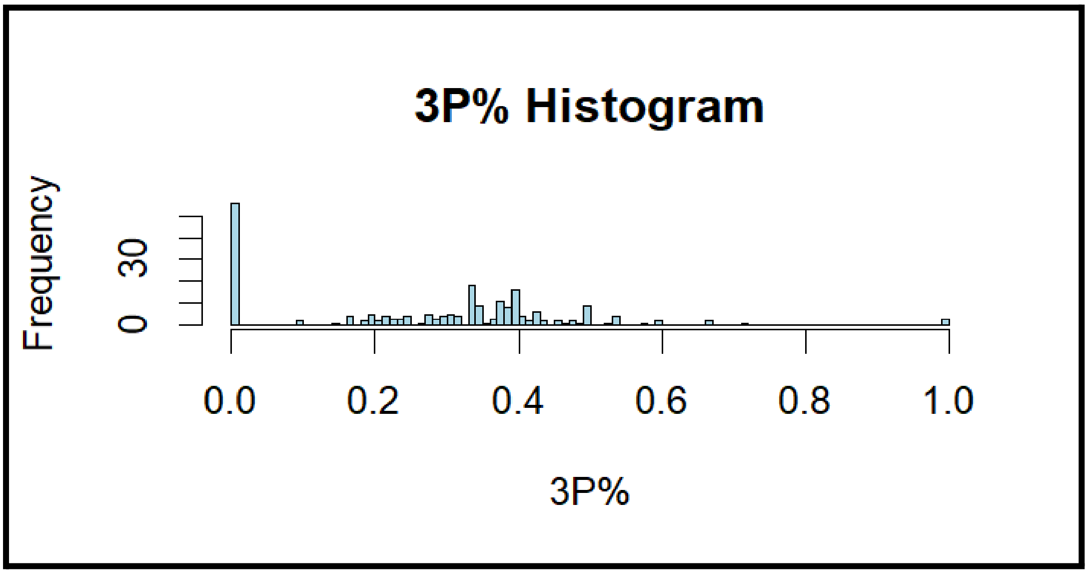
The variance for 3P% is equal to 0.04042537. This result suggests that the 3-point percentages in our data set are relatively consistent . Players with a low variance are more predictable and maintain a steady-level of performance in making three-pointers. The positive effect of this on a team is that it allows the management to build up strategies around their reliable shooting, as this player's shooting performance will have small significant fluctuations.

The standard deviation for 3P% is equal to 0.2010606. This result, as it is close to zero, suggests that the data points in the dataset are very close to the mean percentage.It is important to mention that low standard deviations are generally desirable. In the context of NBA 3P% , this would mean that the players in the dataset tend to have similar and consistent percentages when shooting three-pointers



From the box plot, we can see that the first quartile is found at 0. The black line in the blue box indicates the median, which in this case is 0.331 and the third quartile is located at 0.393. This makes sense since the distance between the first and third quartile is relatively low. This variable is not symmetrical, the median is almost at the top of the box.





As we can see it is very hard to have a 3P success percentage bigger than 40%, especially for those who take a bigger amount of shots. The players who average above 35% in their three point shots are considered a very valuable asset for their franchises taking into account the new tendency in the league which consists of maximizing volume shooting.

3.7 TRB

Total rebound variable (TRB) is a quantitative variable defined as the total number of rebounds after shooting in a game. This is a crucial aspect of basketball, as it involves gaining possession of the ball after a missed goal attempt.

The minimum of rebounds of a player in a game is 0, as there are some games where a player did not take the possession after a rebound in a game.

The average of total rebounds per game in the NBA is 3.304. Which is also called the mean of total rebounds of a player per game.

The maximum rebounds of a player per game is 14.300, where two players were able to achieve it.

Mode is the most repeated number of rebounds in a game, which is from 1 to 2, where 44 players achieved to gain possession of the ball after a throw in a game.

total rebounds	frequency	relative frequency
0 to 1	37	0,170506912
1 to 2	44	0,202764977
2 to 3	29	0,133640553
3 to 4	33	0,152073733
4 to 5	23	0,105990783
5 to 6	19	0,087557604
6 to 7	9	0,041474654
7 to 8	9	0,041474654
8 to 9	3	0,013824885
9 to 10	3	0,013824885
10 to 11	3	0,013824885
11 to 12	0	0
12 to 13	1	0,004608295
13 to 14	2	0,00921659
14 to 15	2	0,00921659
Total	217	1

The range is defined as the totally different number of rebounds per game per player. This is = [1,2,3,4,5,6,7,8,9,10,11,12,13,14,15]

Quartiles, defined as Q1, Q2, and Q3 are values that divide the dataset into four equal parts, each part represents 25% of the dataset, this is commonly used in box plots.

- Q1 = 1.200
- Q2 = 2.800
- Q3 = 4.800

Overall summary figures:

	Player	Pos	Age	Tm	MP	FG%	3P%	TRB	AST	PF	PTS	intervals
1	Precious Achiuwa	C	22	TOR	27.8	0.481	0.313	4.8	1.0	2.3	10.2	[4,5)
2	Steven Adams	C	28	MEM	16.3	0.429	0.000	6.4	2.1	1.7	3.4	[6,7)
3	Bam Adebayo	C	24	MIA	34.1	0.594	0.000	8.0	2.7	3.1	14.8	[8,9)
4	Nickeil Alexander-Walker	SG	23	UTA	5.0	1.000	0.000	1.0	1.0	0.0	5.0	[1,2)
5	Grayson Allen	SG	26	MIL	25.4	0.451	0.396	2.9	1.3	1.8	8.3	[2,3)
6	Jose Alvarado	PG	23	NOP	19.5	0.485	0.375	1.3	1.5	3.0	8.0	[1,2)
7	Kyle Anderson	PF	28	MEM	18.4	0.569	0.250	4.3	1.8	1.8	6.0	[4,5)
8	Giannis Antetokounmpo	PF	27	MIL	37.3	0.491	0.220	14.2	6.8	3.6	31.7	[14,15]
9	Thanasis Antetokounmpo	SF	29	MIL	2.5	0.667	0.000	0.5	0.1	0.5	0.6	[0,1)
10	OG Anunoby	SF	24	TOR	36.2	0.476	0.341	4.0	2.5	3.0	17.3	[4,5)
11	Deandre Ayton	C	23	PHO	30.5	0.640	0.500	8.9	1.7	2.6	17.9	[8,9)
12	Desmond Bane	SF	23	MEM	35.7	0.478	0.489	3.8	2.2	1.9	18.8	[3,4)
13	Dalano Banton	PG	22	TOR	2.0	1.000	0.000	0.5	0.3	0.0	1.8	[0,1)
14	Scottie Barnes	PF	20	TOR	33.3	0.429	0.167	9.0	4.3	3.0	12.8	[9,10)
15	Will Barton	SG	31	DEN	34.4	0.409	0.393	5.6	2.8	2.2	13.8	[5,6)
16	Charles Bassey	PF	21	PHI	4.0	0.500	0.000	1.7	0.3	0.3	0.7	[1,2)
17	Malik Beasley	SG	25	MIN	19.8	0.432	0.320	3.3	0.7	1.5	8.5	[3,4)
18	D?vis Bert?ns	PF	29	DAL	10.7	0.400	0.373	1.4	0.3	1.4	4.1	[1,2)

Showing 1 to 19 of 217 entries, 12 total columns

Console Terminal Background Jobs

R 4.3.1 · ~/

```
>
>
> summary(data_set_nba_playoffs_bueno$TRB)
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max. 
  0.000   1.200   2.800   3.404   4.800  14.300 
```

<u>Calculation with (R) and explanation of variance and standard deviation</u>

The variance of the TRB variable is 8.07901. This means that most players in the NBA take an amount of 8 rebounds per game. Which infers that a positive amount of rebounds happen in a game.

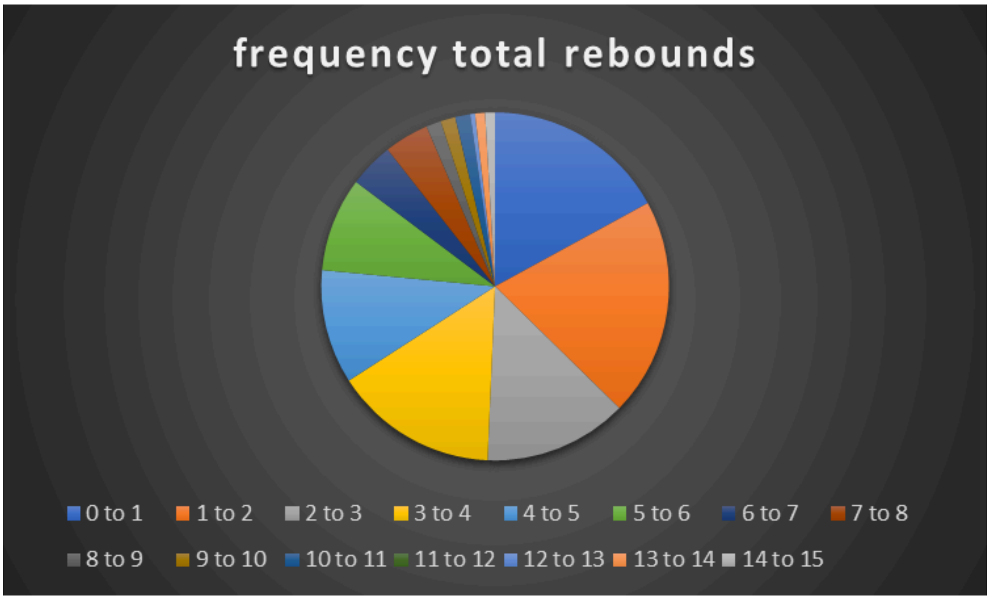
The standard deviation is 2.84236. with respect to the mean, the variance is lower. This implies that there is little variation in the number of rebounds per player per game, which in conclusion, means that it is common to see more or less the same amount of rebounds in all games.

```
> var(data_set_nba_playoffs_bueno$TRB)
[1] 8.07901
> sd(data_set_nba_playoffs_bueno$TRB)
[1] 2.84236
```

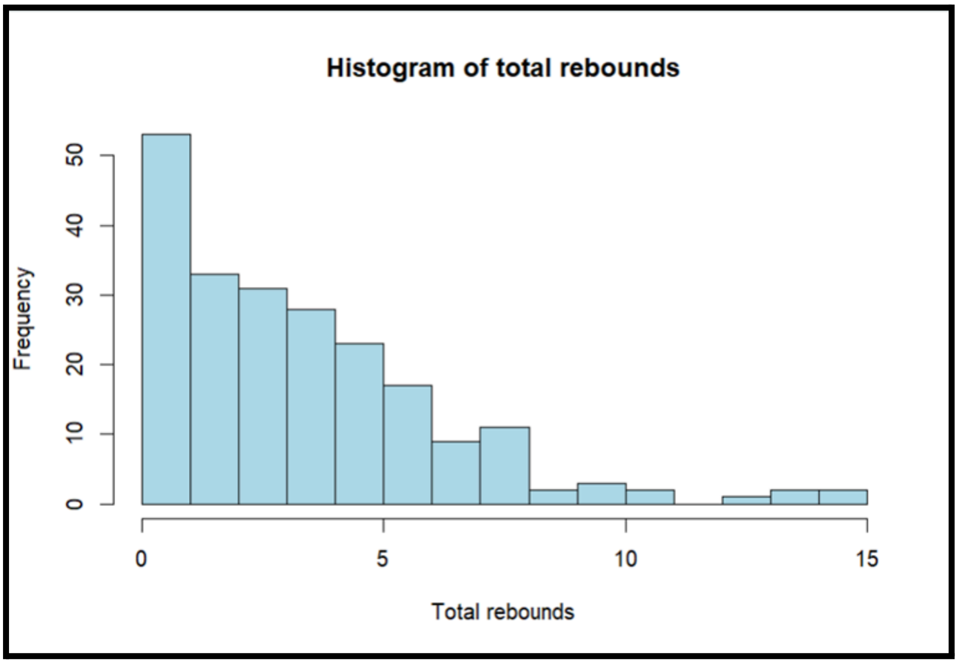
When we calculated the frequency table, we completed a pie chart of the absolute frequency. The information has been taken from the previous table, where we can see the total of rebounds, the frequency and the relative frequency.

The absolute frequency is the number of players that gained possession of the ball after a missed throw (rebound) in the interval 0-15.

The relative frequency is the relationship between the total number of players, and the absolute frequency from each interval.



The following bar chart shows in the Y-axis the frequency and in the X-axis the rebounds. And it shows graphically the relationship between both.



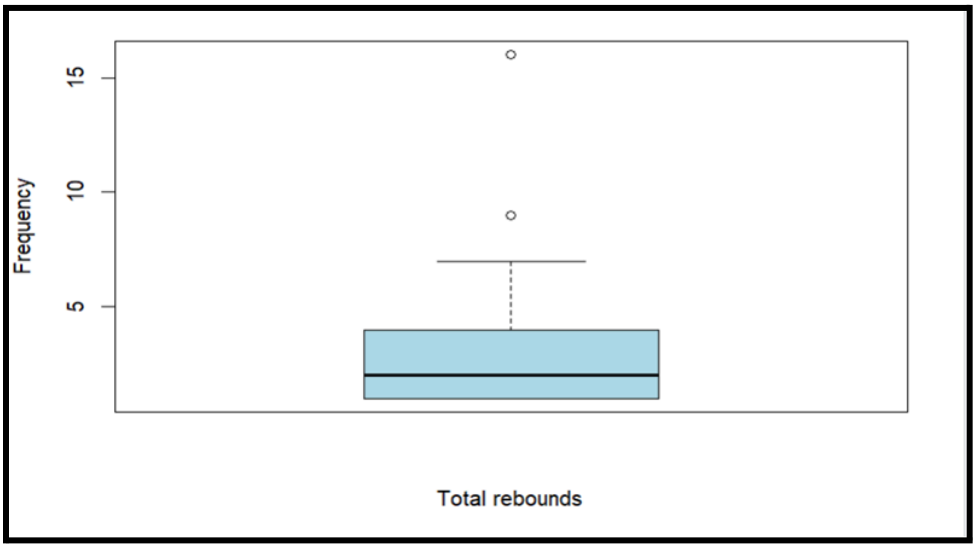
This histogram shows a graphical representation that shows the distribution of the number of rebounds collected by NBA players in a game. The graph shows that there is a great number of players that gain possession of the ball in the interval from 0 to 2 total rebounds. The histogram can help analyze players information, from the results recorded, we can see that only two players achieved to be between the intervals of 14 to

15, this means that these players could be the best options to place them strategically in the court. In conclusion, the histogram can help identify players who excel in this particular aspect of the game, contributing to their team's success.

The way we built up the graph was by using R-studio and plotting the different commands to get the values represented graphically. The commands we used are the following:

```
> hist(data_set_nba_playoffs_bueno$TRB,
+ main = "Histogram of total rebounds",
+ xlab = "Total rebounds",
+ ylab = "Frequency",
+ col = "lightblue",
+ Border = "black",
+ breaks = 15)
```

Box plot of total rebounds



A box plot of total rebounds per player per game provides a graphical representation of the distribution and summary of the data. It allows you to visualize key characteristics and the spread of rebounding performance among NBA players.

The box in the plot represents the interquartile range. The bottom edge of the box indicates the first quartile (Q1), representing 25% of the data. The top edge of the box indicates the third quartile (Q3) representing 75% of the data. The middle line in the box indicates the median of the data, also called the second quartile (Q2), where it separates the box in two parts.

The plot also includes whiskers, which extend the box to the minimum and maximum values (in this case maximum). They represent the data range. This bar plot also has outliers, individual data points that are usually high or low compared to most of the data (in this case, these points are above the rest data). With this we can see that the variable is not symmetrical.

In conclusion, box plotting is a useful key to assess the distribution of a dataset. It can help identify variations in rebounding performance and highlight standout players with exceptional performance (the two individual points above the whiskers)

We used RStudio, which helped us get a visual representation of the information by using different commands. These commands are:

```
> boxplot(table(data_set_nba_playoffs_bueno$TRB),
+ main = "Bar Plot",
+ xlab = "Total rebounds",
+ ylab = "Frequency",
+ col = "lightblue",
+ Border = "black",
+ names.arg = names(table(data_set_nba_playoffs_bueno$TRB)),
+ cex.names = 2.8)
```

3.8 AST

The variable **"Assists (AST)"** is a quantitative variable defined as the number of assists made by a player in a game for the NBA.

The minimum of assists of a player per game is 0, as there are times where NBA players do not make any assists in a game.

The average assists of a player per game in the NBA is: 1.829, which is also called the mean of assists per game.

The maximum number of assists per game is: 9.8, with only one player to achieve it.

Mode: is the most repeated number of assists in a game, which is from 0 to 1 assists, made by 105 players.

ASISTS	Frequenc	relative frequency
0 to 1	105	0,483870968
1 to 2	41	0,188940092
2 to 3	24	0,110599078
3 to 4	15	0,069124424
4 to 5	10	0,046082949
5 to 6	7	0,032258065
6 to 7	11	0,050691244
7 to 8	1	0,004608295
8 to 9	2	0,00921659
9 to 10	1	0,004608295
10 to 11	0	
total	217	

The range is defined as the totally different number of assists per game per player. This is = [1,2,3,4,5,6,7,8,9,10].

Quartiles, denoted as Q1, Q2 and Q3 are values that divide the dataset into four equal parts, each part represents 25% of the dataset. This is commonly used in box plots.

- Q1= 0.400
- Q2= 1.000
- Q3= 2.700

Overall summary figures:

	Player	Pos	Age	Tm	MP	FG%	3P%	TRB	AST	PF	PTS	intervals
206	Coby White	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
207	Derrick White	SG	27	BOS	25.4	0.364	0.313	3.0	2.7	2.0	8.5	[3,4)
208	Hassan Whiteside	C	32	UTA	10.8	0.417	0.000	5.2	0.0	1.5	1.8	[5,6)
209	Andrew Wiggins	SF	26	GSW	34.9	0.469	0.333	7.5	1.8	2.5	16.5	[7,8)
210	Grant Williams	PF	23	BOS	27.3	0.433	0.393	3.8	0.8	3.5	8.6	[3,4)
211	Patrick Williams	PF	20	CHI	30.6	0.468	0.333	5.4	0.8	3.0	11.8	[5,6)
212	Robert Williams	C	24	BOS	23.2	0.679	0.000	6.2	1.0	2.0	7.7	[6,7)
213	Ziaire Williams	SF	20	MEM	16.8	0.442	0.306	1.6	0.5	1.4	6.9	[1,2)
214	Delon Wright	SG	29	ATL	27.4	0.517	0.385	4.8	2.8	0.4	8.2	[4,5)
215	Thaddeus Young	PF	33	TOR	14.5	0.500	0.143	3.0	1.7	1.3	3.3	[3,4)
216	Trae Young	PG	23	ATL	37.2	0.319	0.184	5.0	6.0	2.6	15.4	[5,6)
217	Omer Yurtseven	C	23	MIA	4.2	0.667	0.000	0.8	0.3	0.2	2.8	[0,1)

Showing 206 to 217 of 217 entries, 12 total columns

Console Terminal Background Jobs

R 4.3.1 · ~/

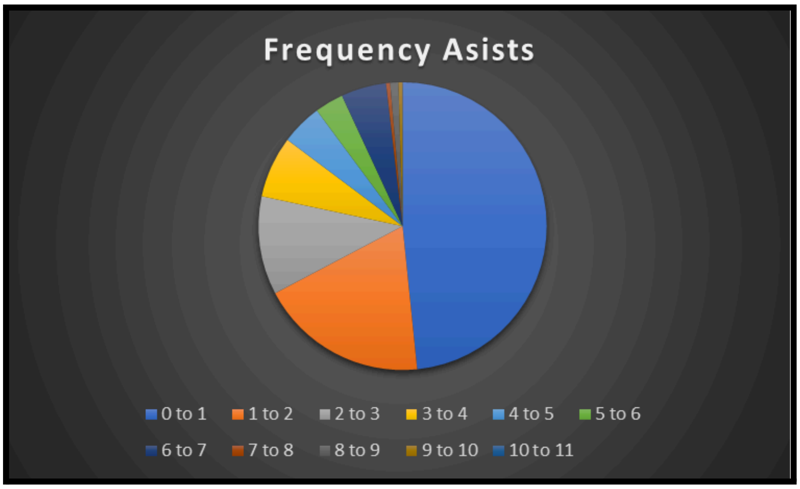
```
> library(readxl)
> data_set_nba_playoffs_bueno <- read_excel("C:/Users/alexs/Downloads/data_set_nba_playoffs_bueno.x
lsx")
> View(data_set_nba_playoffs_bueno)
> summary(data_set_nba_playoffs_bueno$AST)
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max. 
  0.000   0.400   1.000   1.829   2.700   9.800 
> table(data_set_nba_playoffs_bueno$AST)
```

__Calculation with (R) and explanation of variance and standard deviation__

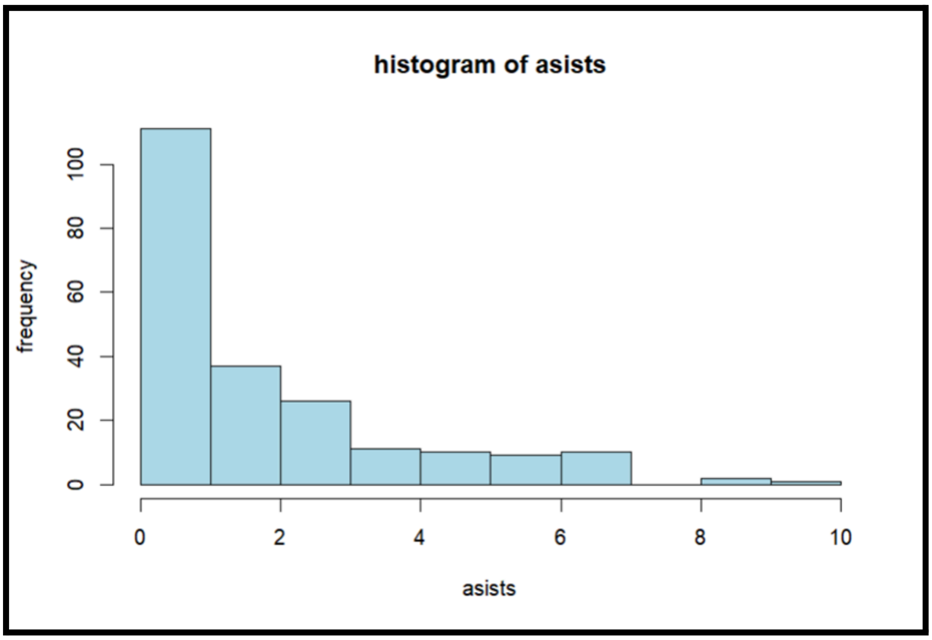
The variance of the AST variable is 4.028532 assists. Since the total assists per game per player is close to 4 assists, this means that most players are around a similar assist range, and this range is 4.

The standard deviation is 2.00712 assists. With respect to the mean, the variance is greater. This infers that there is a relative amount of variation in the number of assists per player in a game.

```
> var(data_set_nba_playoffs_bueno$AST)
[1] 4.028532
> sd(data_set_nba_playoffs_bueno$AST)
[1] 2.00712
```



The following bar chart shows the frequency on the Y-axis, and in the X-axis, the total number of assists.

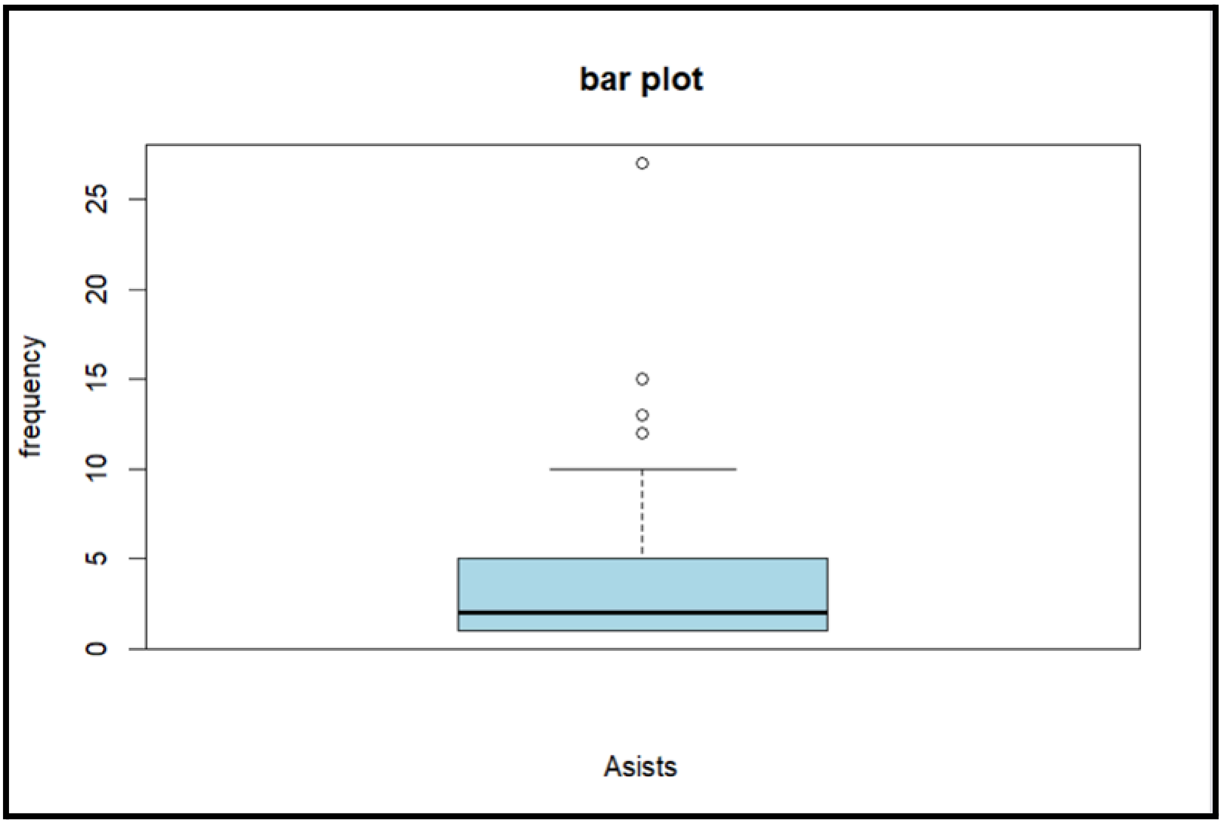


From the chart we can see that the highest frequency of assists it's between the intervals 0 to 1. This means that most players of the NBA complete from 0-1 assists in a game. From the chart we can infer that as higher the intervals of assists, the lower the frequency. This means that it is more difficult for a player to complete a higher rate of assists per game. We can also infer that the players above the interval from 4-5 assists per game, are players that play a greater number of minutes and have better conditions to be on the court.

The way we could build up the graph was by using R-studio and plotting the different commands to get the values represented graphically. The commands we used are the following:

```
> hist(data_set_nba_playoffs_bueno$AST,
+       main = "histogram of asists",
+       xlab = "asists",
+       ylab = "frequency",
+       col = "lightblue",
+       border = "black",
+       breaks = 10)
```

Bar plot for Assists



A box plot for assists per player per game provides a graphical representation of the distribution and summary of the data. It allows you to visualize the number of assists made by a player in an NBA game.

The box in the plot represents the interquartile range. The bottom edge of the box indicates the first quartile (Q1), representing 25% of the data. The top edge of the box indicates the third quartile (Q3) representing 75% of the data. The middle line in the box indicates the median of the data, also called the second quartile (Q2), where it separates the box in two parts. This variable is definitely not symmetrical, the median is close to the bottom of the box.

The plot also includes whiskers, which indicates the range of the data outside the interquartile range, showing maximum and minimum points. However, this bar plot includes outliers, which are individual points that are above the IQR and whiskers. These individual data points are higher compared to the rest of the data

In conclusion, box plotting is a useful key to assess the distribution of a dataset. It can help identify central tendencies, assessing effectiveness, and spotting outliers. It helps gain insight into the performance of players when it comes to creating scoring opportunities for their teammates.

3.9 PF

This variable PF is a discrete quantitative value as it takes values from 0 to 5.

In the context of the NBA playoffs, "personal fouls" refer to the number of fouls committed by a player or a team during a game or a playoff series. An infraction of the rules that results in forced physical contact with an opponent is known as a personal foul. Pushing, falling, or making excessive physical contact with an opponent are some examples of these fouls.

In the NBA playoffs, personal fouls is a significant statistic for various reasons:

Player impact: Personal fouls can have a big effect on a player's performance and playing time. A player may be kicked out of the game if they commit too many fouls, which could be a crucial setback for his side.
Team Defense: The number of personal fouls committed by a team reflects their defensive performance. High foul totals may be a sign of a forceful defensive approach, but they can also result in free-throw opportunities for the opposition, potentially changing the course of the game.

Player Effectiveness: Examining personal fouls can reveal information about a player's defensive focus and efficiency. Because they can stop the opponent's plays without sending them to the free-throw line, players who can defend without fouling are valued.

In the NBA, personal fouls are a crucial statistic since they provide important information on team and player performance, defensive tactics, and game outcomes. Coaches, analysts, and spectators can better comprehend playoff match dynamics and player contributions by analyzing personal fouls.

Here is the frequency table

Interval	Frequency	Relative Frequency
[0,1)	58	0,267281106
[1,2)	60	0,276497696
[2,3)	60	0,276497696
[3,4)	30	0,138248848
[4,5]	9	0,041474654
total	217	1

The maximum amount of personal fouls averaged by a player was 4.7 (which is rational taking into account that A player who commits six personal fouls over a match, out and is disqualified for the remainder of the game) and the minimum score averaged was 0, that is why I decided to do the frequency table of the variable personal fouls per game dividing the whole frequency table in the intervals shown in the screenshot. With this image we can determine the **mode**, which is the most repeated value. In this case we find it in the interval [1,3). In fact the exact frequency that was repeated the most was 0, which was averaged by 21 different players out of the 217 players that played the NBA 2021-2022 Play Offs.

Here is the evidence of how I managed to get this results using Rstudios:

```
> table(data_set_nba_playoffs_el_bueno$intervals)

[0,1) [1,2) [2,3) [3,4) [4,5]
  58    60    60    30     9
> summary(data_set_nba_playoffs_el_bueno$PF)
  Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
 0.000   0.800   1.800   1.784   2.800   4.700
> sd(data_set_nba_playoffs_el_bueno$PF)
[1] 1.175811
> var(data_set_nba_playoffs_el_bueno$PF)
[1] 1.382531
```

From these calculations made in Rstudios we can define the following values:
Mean, which is the average of personal fouls committed by a player , in this specific case is 1.784
Min and Max which as I mentioned before are 0 and 4.7 respectively, meaning that the **range** or all the possible values that the variable can take, goes from 0 to 4.

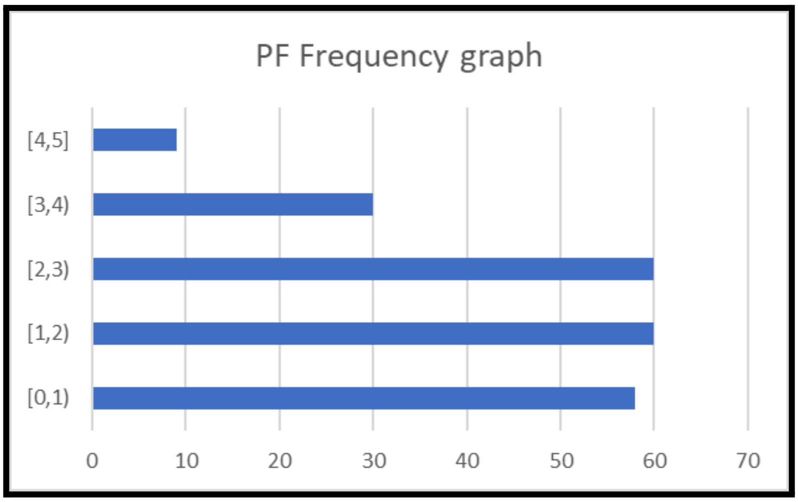
The standard deviation of 1.175811 for personal fouls in the 2021-2022 NBA Playoffs dataset is an interesting statistic. It provides valuable insight on the variation in player foul totals throughout the postseason this year. This low standard deviation indicates that, generally speaking, the amount of personal fouls committed during a game is quite constant. The real fouls, however, can deviate from this average by roughly 1.175811 in either way. This variability could be due to several factors, such as Game Intensity, Player Styles, Refereeing or Opponent Matchups.

In conclusion, a low standard deviation suggests that although personal fouls vary somewhat, the data are generally concentrated towards the mean. In order to plan plans, coaches and analysts need to be aware of this number because it allows them to estimate the normal range of fouls that could occur in postseason games.

The variance represents the degree of spread or dispersion in the data, specifically in this case, it tells us about the variability in the number of personal fouls committed by players during the playoffs. There is an average amount of personal fouls committed per game, but there is also a small degree of variability around that average, as indicated by the variance of 1.382531.

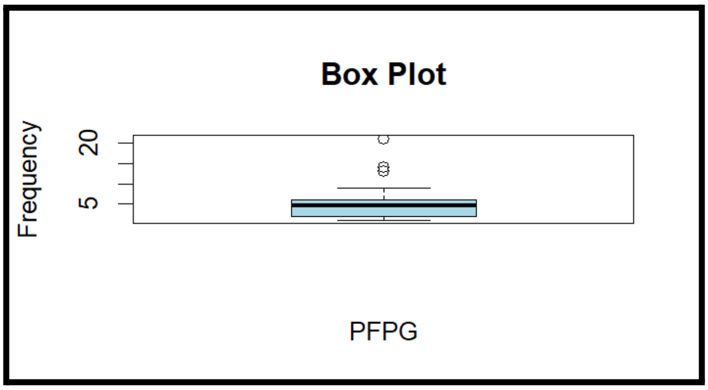
It is essential for coaches, analysts, and team plans to comprehend variance. They are able to plan player rotations and tactics by anticipating the variety of fouls that can happen in postseason games. A greater variation indicates more options and the necessity for flexible game plans.

This bar chart just helps making more visual the Absolute frequency



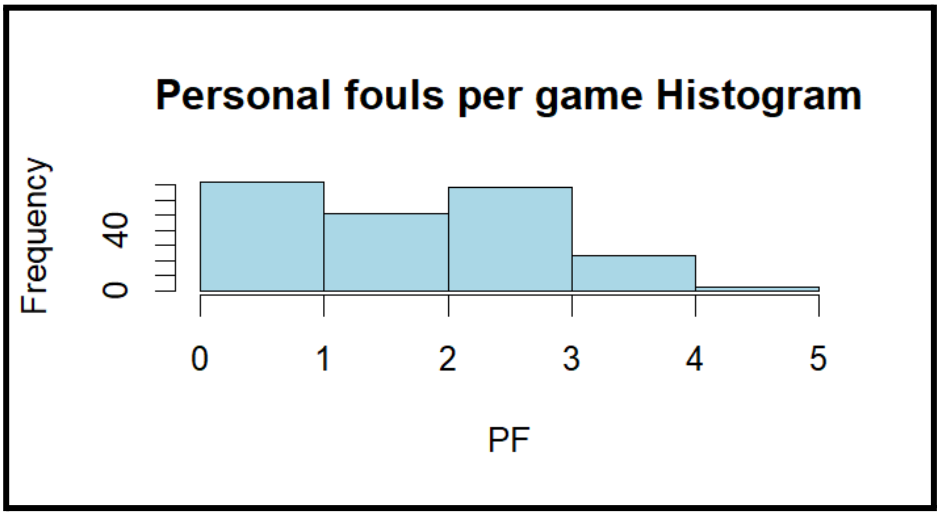
This bar chart just helps making more visual the Absolute frequency

Now, we create the box plot and the histogram for the “PF” per game.



This box plot indicates how the quartiles are divided. So based on the plot we can see that most players are distributed between personal fouls averaged per game in the play-offs. The value of 1^{st} quartile is 0.8(this means that the first 25% of players average this score per game or below), we can see it as the lower line of the blue box. The Median value is 1,784 (this means that the first 50% of players average this score per game or below),we can see it as the black line in the middle. The 3^{rd} quartile is 2.8 (this means that the first 75% of players average this score per game or below), we can see it as the upper line of the box. We can see that, as the median is almost at the border of the box, this is an asymmetrical variable.

Finally we made this histogram that show us the most frequent amount of fouls averaged by players in a single match



3.10 PTS

Definition: This variable is a continuous quantitative variable as it takes values from 0 to 31.7 including decimal numbers.

The National Basketball Association (NBA) uses the statistic "points per game" (PPG) to describe the average number of points a player scores in each game they participate in over a specific period of time. It is a frequently used statistic to evaluate a player's offensive performance and scoring ability.

A player's PPG is calculated by adding up all of their points over a specific amount of games, dividing that total by the number of games played. The following equation can be used to determine PPG:

PPG = (Total Points Scored) / (Number of Games Played)

For example, if a player has scored a total of 500 points over 50 games, their PPG would be:

PPG = 700 points / 70 games = 10 points per game

A player who scores frequently has a high PPG, whereas a player who scores less frequently may not be as involved in the scoring aspect of the game. It's an important indicator for evaluating a player's offensive contribution and consistency in scoring goals for their group.

Here is the frequency table

interval	frecuency	relative frecuency
[0,5)	95	0,437788018
[5,10)	46	0,211981567
[10,15)	36	0,165898618
[15,20)	18	0,082949309
[20,25)	11	0,050691244
[25,30)	8	0,036866359
[30,35]	3	0,013824885
total:	217	1

The maximum amount of points averaged by a player was 31.7 and the minimum score averaged was 0, that is why I decided to do the frequency table of the variable points per game dividing the whole frequency table in the intervals shown in the screenshot. With this image we can determine the **mode**, which is the most repeated value. In this case we find it in the interval [0,5). In fact the exact frequency that was repeated the most was 0, which was averaged by 12 players out of the 217 players that played the NBA 2021-2022 Play Offs.

Here is the evidence of how I managed to get this results using Rstudios:

Statistics:
summary(data_set_nba_playoffs_el_bueno$PTS)

```
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
  0.000   2.100   6.000   8.457  12.600  31.700
```

standard deviation:sd(data_set_nba_playoffs_el_bueno$PTS)
[1] 7.625418
variance: var(data_set_nba_playoffs_el_bueno$PTS)
[1] 58.147

From these calculations made in Rstudios we can define the following values:
Mean, which is the average of the points averaged per game, in this specific case is 8.457
Min and Max which as I mentioned before are 0 and 31.7 respectively, meaning that the **range** or all the possible values that the variable can take, goes from 0 to 31.7

The value of 1st quartile is 2.1(this means that the first 25% of players average this score per game or below), the 2nd quartile value is 6.0 (this means that the first 50% of players average this score per game or below), the 3rd quartile is 12.6 (this means that the first 75% of players average this score per game or below)

The **standard deviation** of the variable "points per game" is 7.625418, and it tells us how much individual data points typically deviate from the average points scored per game. In this context, it shows the level of dispersion or variability in player performance during the playoffs.

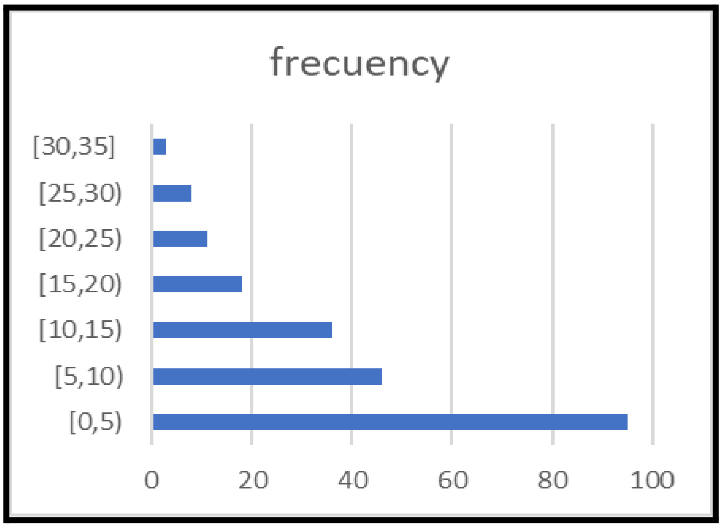
A higher standard deviation indicates that there is a broader range of point-scoring performances among the players. Variables including differing playing styles, matchups, and the unpredictable nature of postseason games may be responsible for this fluctuation. On the other side, a lower standard deviation would suggest a more consistent player performance. Understanding the standard deviation helps in determining the degree of consistency or volatility in NBA players' postseason point-scoring behaviors.Coaches, analysts, and spectators can all benefit from this statistic to determine whether player performance in this crucial matchup is unpredictable or stable.

The **variance** of the variable "points per game" has a value of 58.147, indicating the extent to which individual data points in the dataset deviate from the mean (average) points scored per game. A higher variance suggests a wider spread of data points around the mean, meaning greater variability in player performance. In the context of the NBA Playoffs, The variance may be explained by the players' varied skill sets and playing skills, which result in distinct point contributions in various games. Examining the variance can be helpful for determining whether player performance throughout the playoffs is consistent or unpredictable.

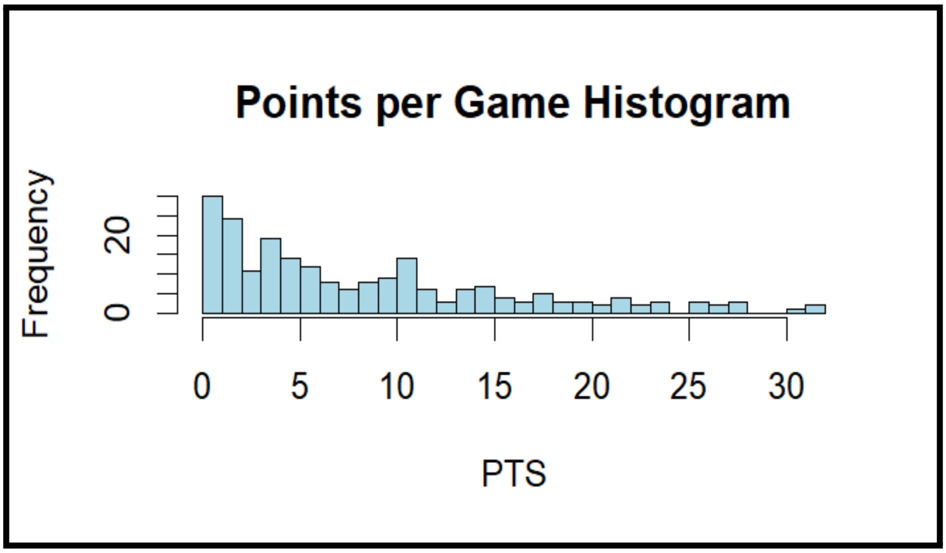
interval	frecuency	relative frecuency
[0,5)	95	0,437788018
[5,10)	46	0,211981567
[10,15)	36	0,165898618
[15,20)	18	0,082949309
[20,25)	11	0,050691244
[25,30)	8	0,036866359
[30,35]	3	0,013824885
total:	217	1

Now, with this information we create the bar chart and the bar plot to be able to interpret the graphs obtained.

This bar chart just helps to make the Absolute frequency be more visual..

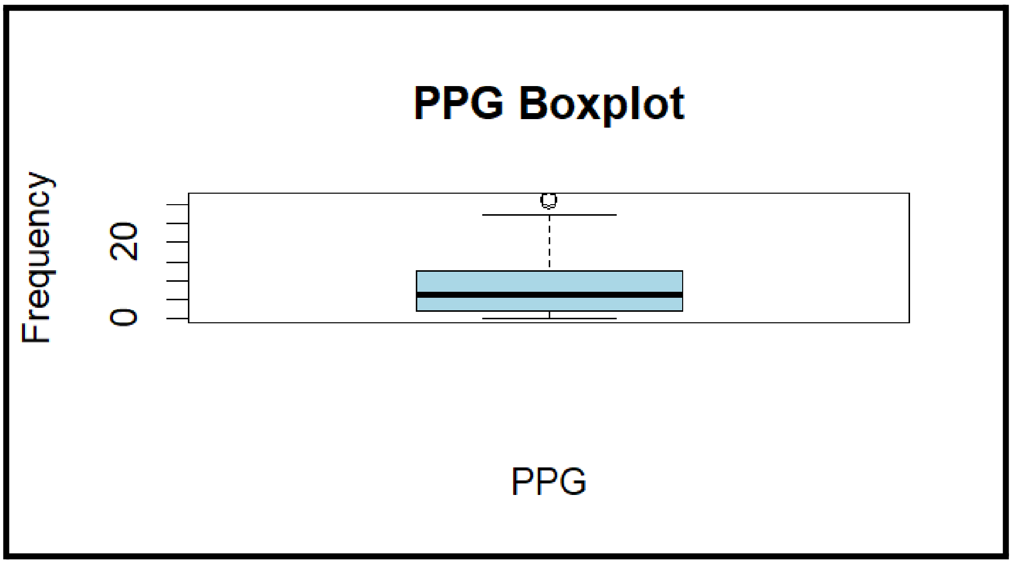


The histogram gives us a better visual of the distribution of the most frequent intervals of points averaged per game by all the players that participated on the play-offs.



As we can see only 11 players managed to average 25 or more points per game, which is considered the minimum amount of points required to be seen as a Elite player in the league

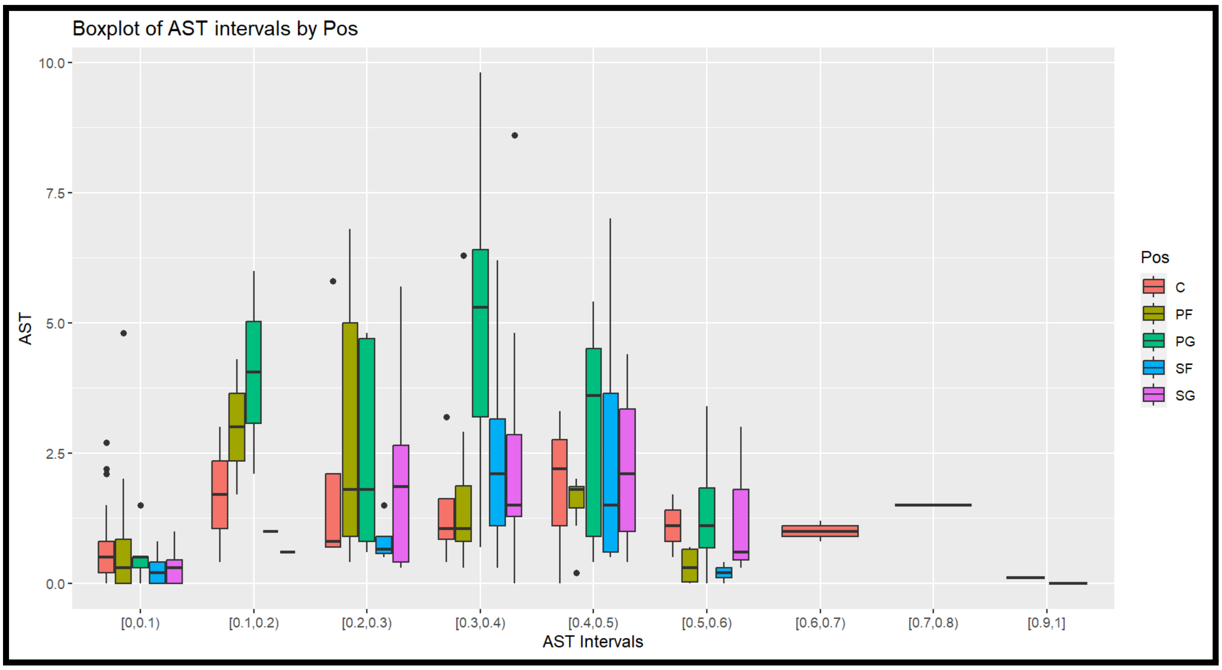
Finally, we had Rstudio create us a **Box plot** of the variable, where we can see the mean and the quartiles of the points per Game.

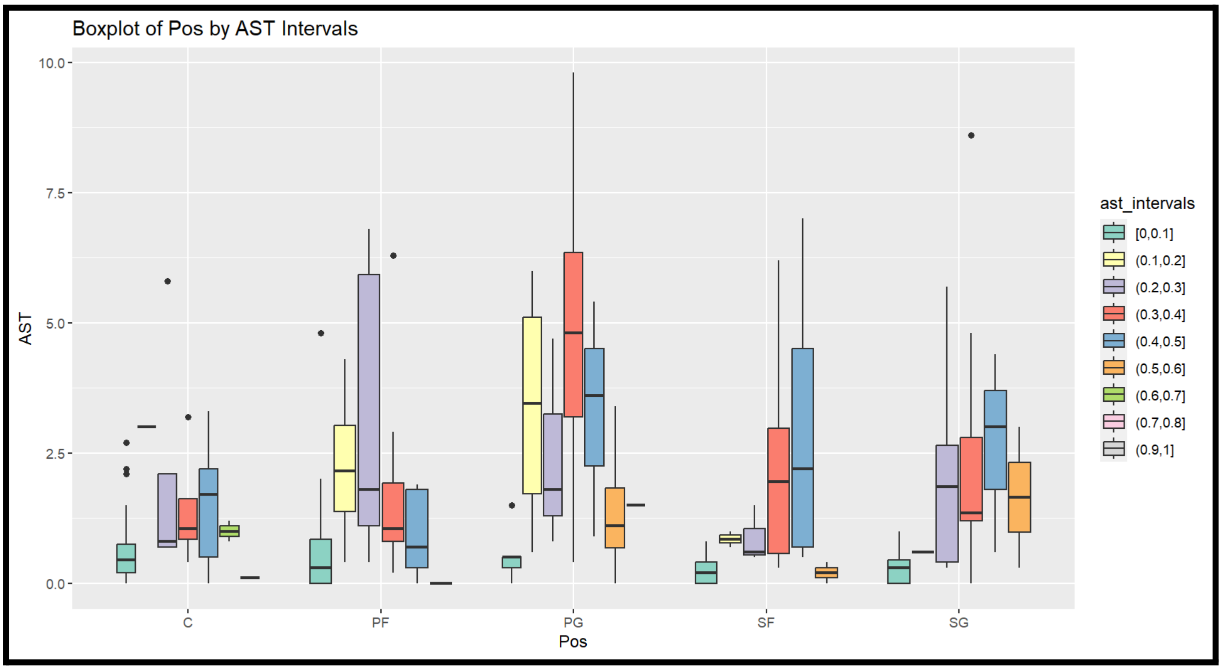


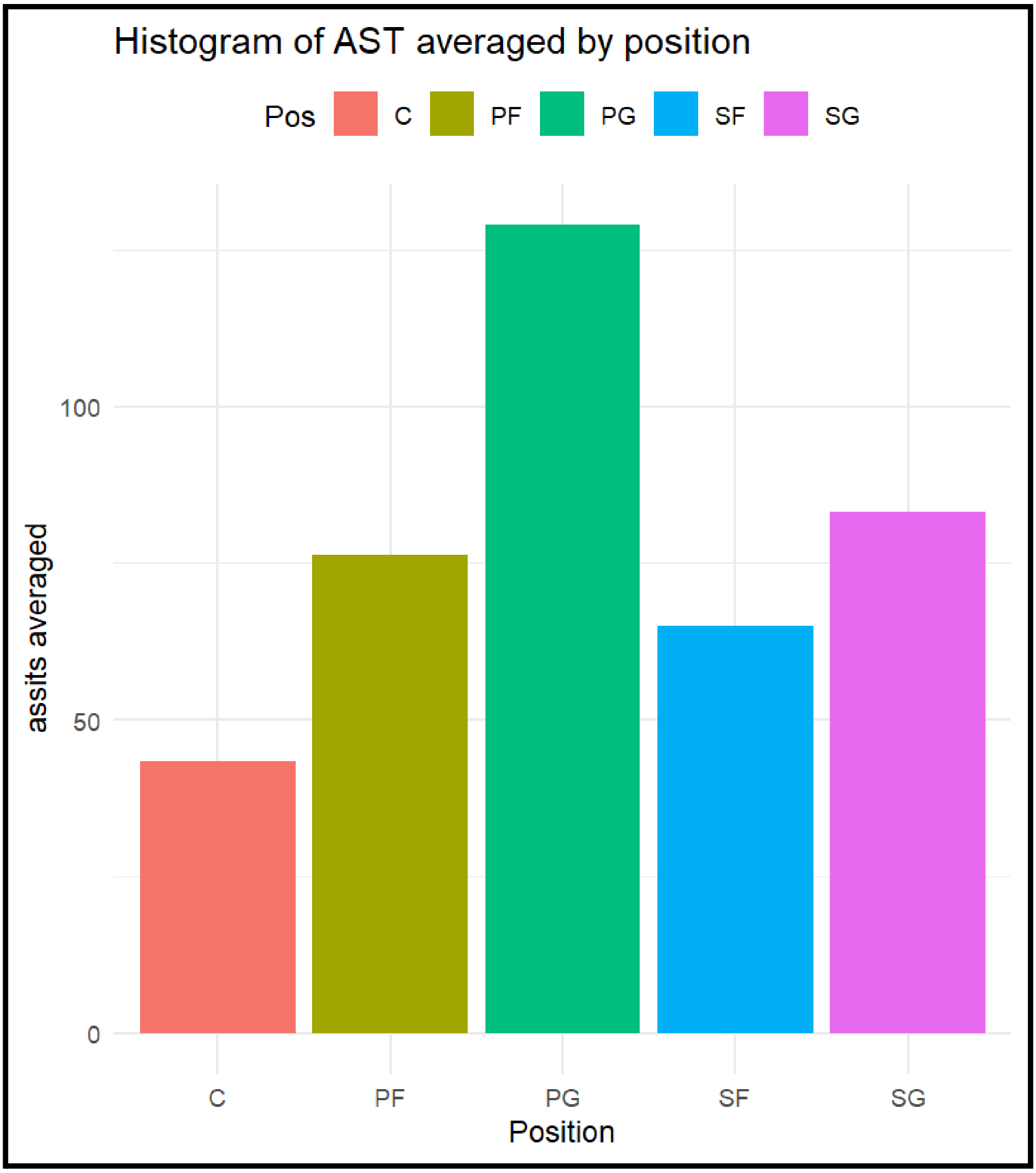
The value of 1st quartile is 2.1(this means that the first 25% of players average this score per game or below), we can see it as the lower line of the box. The **Median** value is 6.0 (this means that the first 50% of players average this score per game or below), it is located as the black line in the middle of the blue box. The 3rd quartile is 12.6 (this means that the first 75% of players average this score per game or below), we can see it as the upper line of the box. This variable is asymmetrical, because the whiskers are not proportional.

4. Bivariate analysis

Analyzing the relation between AST and POSITION:







As we can observe in both graphs both point guards (PG) and shooting guards (SF), tend to average more assists per game. This fact makes sense since guards traditionally are the players in charge of moving the ball and creating spaces for other players to score. This skill is more typical of point guards and translates to more assists on the court from their side.

For creating the previous graph, we have used the following code in R

```
library(ggplot2)
ggplot(data_set_nba_playoffs_el_bueno, aes(x = ast_intervals, y = AST, fill
= Pos)) +
  geom_boxplot() +
  labs(title = "Boxplot of AST intervals by Pos",
       x = "AST Intervals",
       y = "AST",
```

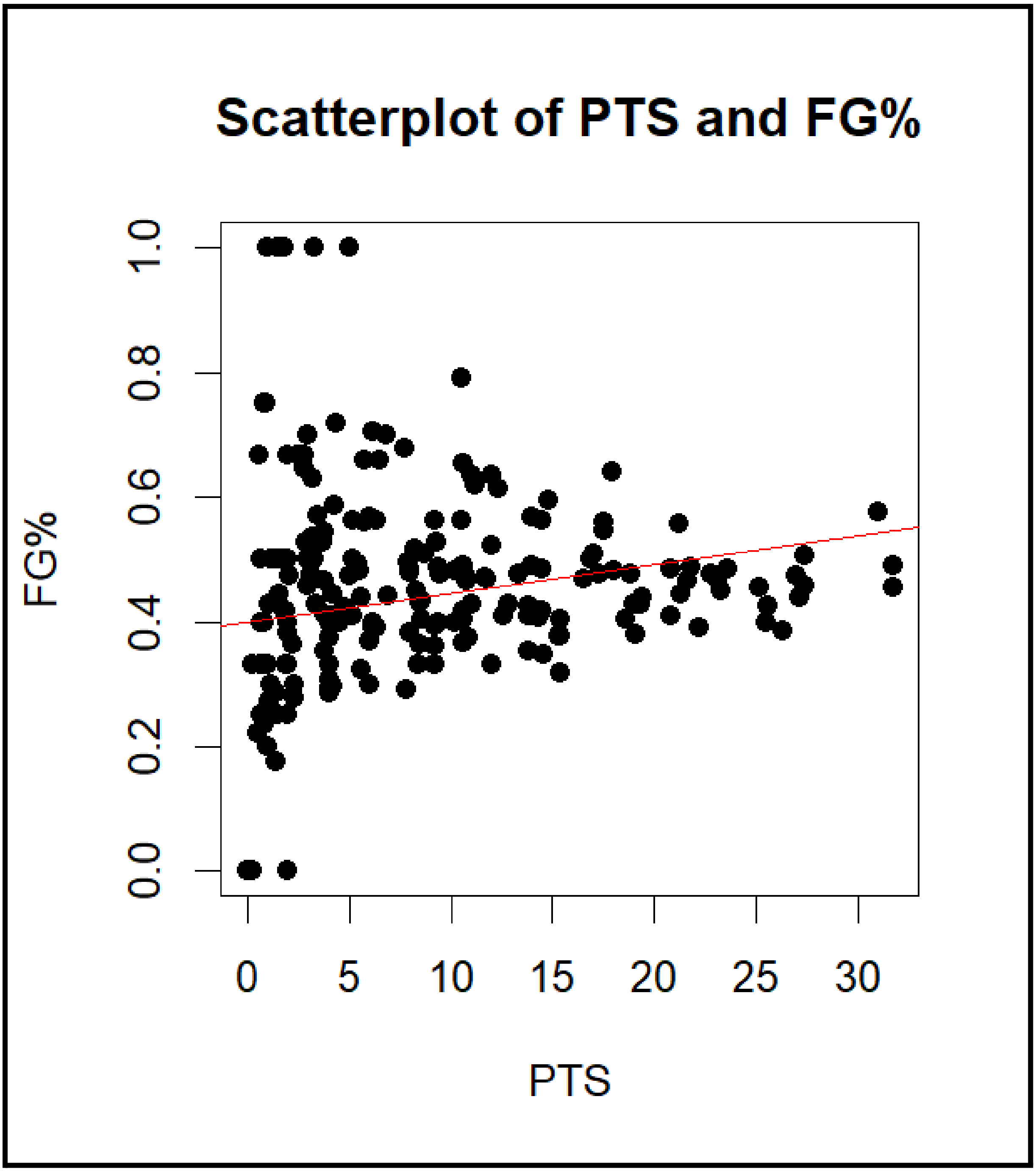
```
       border = "black",
       col = "lightblue")
  scale_fill_brewer(palette = "Set3")

 library(ggplot2)

  ggplot(data_set_nba_playoffs_el_bueno, aes(x = Pos, y = AST, fill =
ast_intervals)) +
    geom_boxplot() +
    labs(title = "Boxplot of Pos by AST Intervals",
         x = "Pos",
         y = "AST") +
    scale_fill_brewer(palette = "Set3")
```

Analyzing the relation between PTS and FG%:

Pearson Coefficient: 0.1910544



We wanted to know how crucial in the whole game points are the Field Goals, and with that see the level of importance foul goals have in the NBA playoffs.

We found out that the Pearson coefficient is 0.1910544. As it is considerably far from 1, it tells us that these two variables are not very much related.

After observing the graph we got to the following conclusion related to the slope of the trendline or the efficiency of each player

The trendline shows a positive slope which indicates that as the value of one variable increases, the other variable also tends to increase. In other words, there is a positive correlation between the two variables, meaning that they are directly related.

There is only one player among the players that averaged 25 points per game or more (which is considered the minimum amount of points required to be seen as a Elite player) whos FG% is above the trendline, and only 2 players There are only 2 players whose field goal shooting averages exceed 50%.", These two players (Jimmy Butler and Nikola Jokic) are therefore considered to be the two most efficient players of this play-offs, Being considered as valuable players since efficiency in the game is regarded as one of the most significant qualities a player can possess. In fact one of these two players, Nikola Jokic was awarded that season with the NBA MVP (Most Valuable Player).

Player	Pos	Age	Tm	MP	FG%	3P%	TRB	AST	PF	PTS
Nikola Jok	C	26	DEN	34,2	0,575	0,278	13,2	5,8	4	31
Jimmy But	SF	32	MIA	37	0,506	0,338	7,4	4,6	1,5	27,4

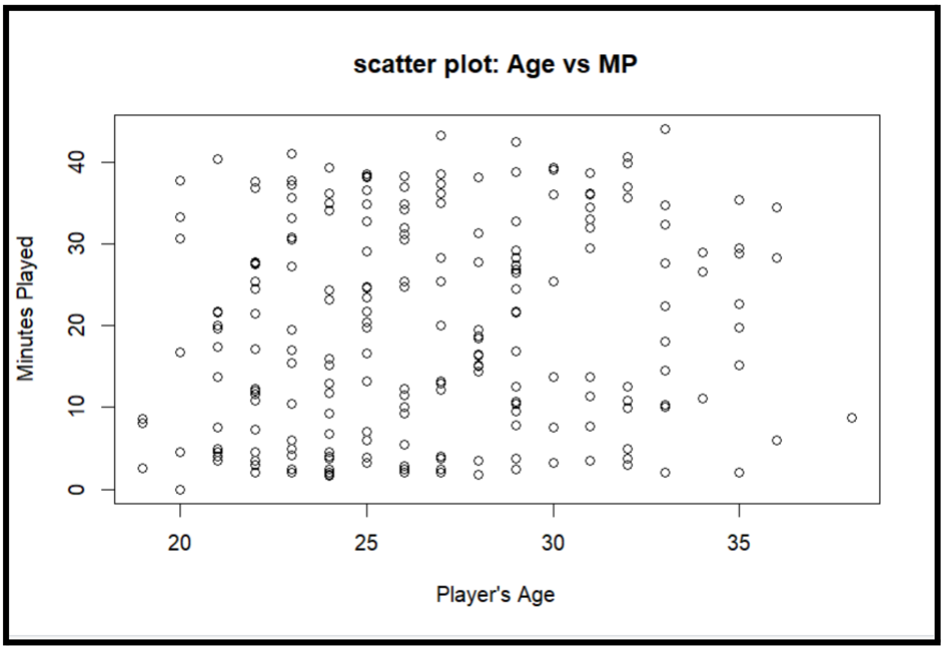
The commands used in Rstudio in order to create the scatter plot are the following:

```
> plot(data_set_nba_playoffs_el_bueno$PTS,
data_set_nba_playoffs_el_bueno$`FG%`, main="Scatterplot de PTS and FG%",
xlab="PTS", ylab="FG%", pch=19)
> abline(lm(`FG%` ~ PTS, data=data_set_nba_playoffs_el_bueno), col="red")
```

Pearson coefficient:

```
> correlation<- cor(data_set_nba_playoffs_el_bueno$PTS,
data_set_nba_playoffs_el_bueno$`3P%`)
> print(correlation)
```

Analyzing the relation between Age and MP:



The two variables (bivariables) were used in order to create the previous scatter plot where "players age" with "minutes played". We had the hypothesis that there was an existing relationship between them which is that, depending on the age player's have, the greater or fewer minutes they will play. We thought that if a player is too young or too old, they do less minutes and that provides effectiveness in the rotation of the players. Also that there is a significance of player development and impact of age on their performance and playing time in the NBA. Younger players often require more time to adapt to the league and gain experience, which may result in less playing time. Conversely, players between 25 and 32 years old are typically in their prime years, showcasing higher skill levels and, therefore, receiving more playing time.

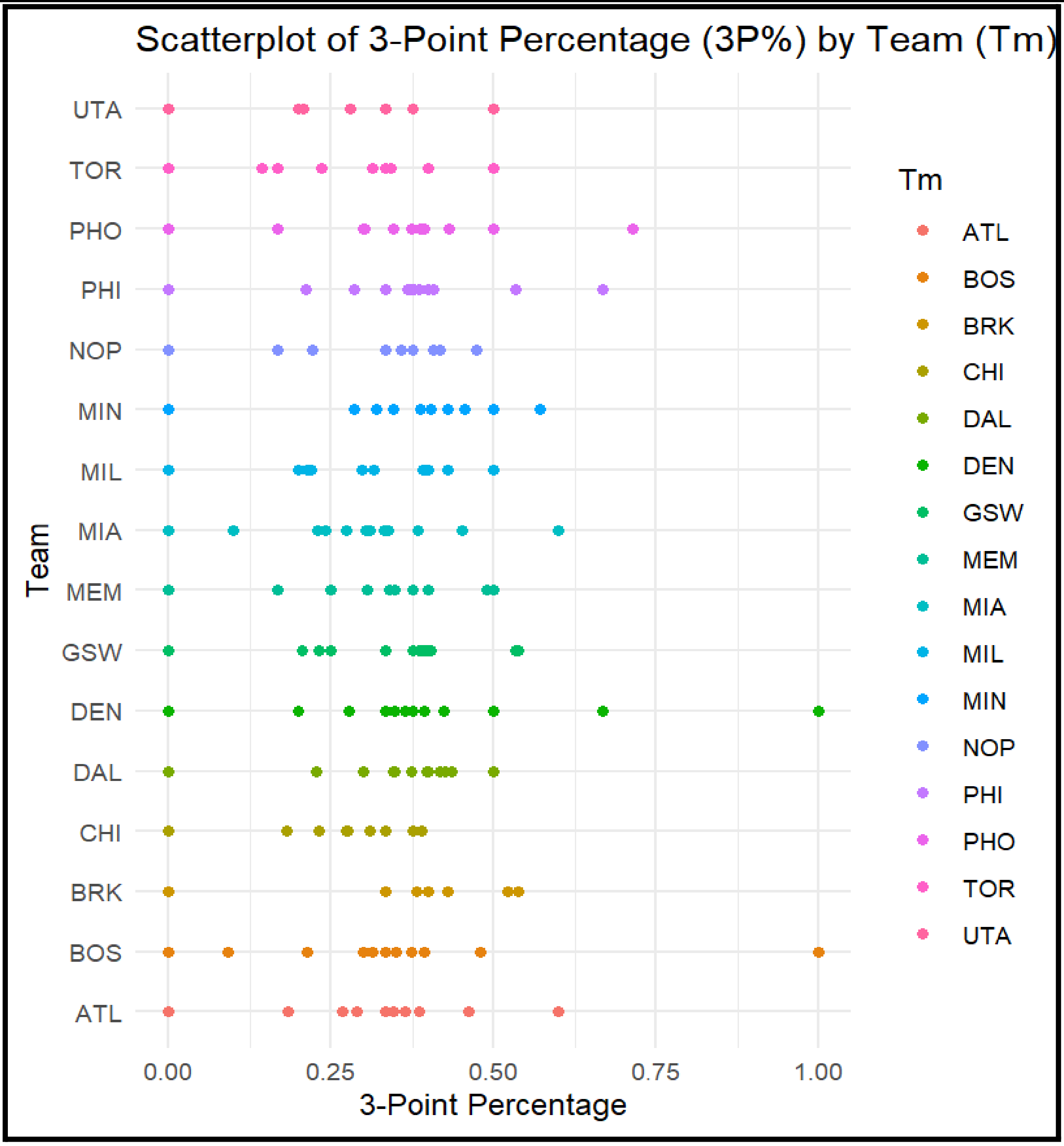
We encountered a problem when getting the data from Rstudio and the scatter plot. As you can see, this scatter plot looks different from the others, as the points are very dispersed and we cannot really se a trendline. What this means is that the relationship between this two variables is almost non-existent. This really affected our perspective because we were convinced that they affected each other widely. So in conclusion our prediction of this relationship was wrong, and the data made us confirm that.e

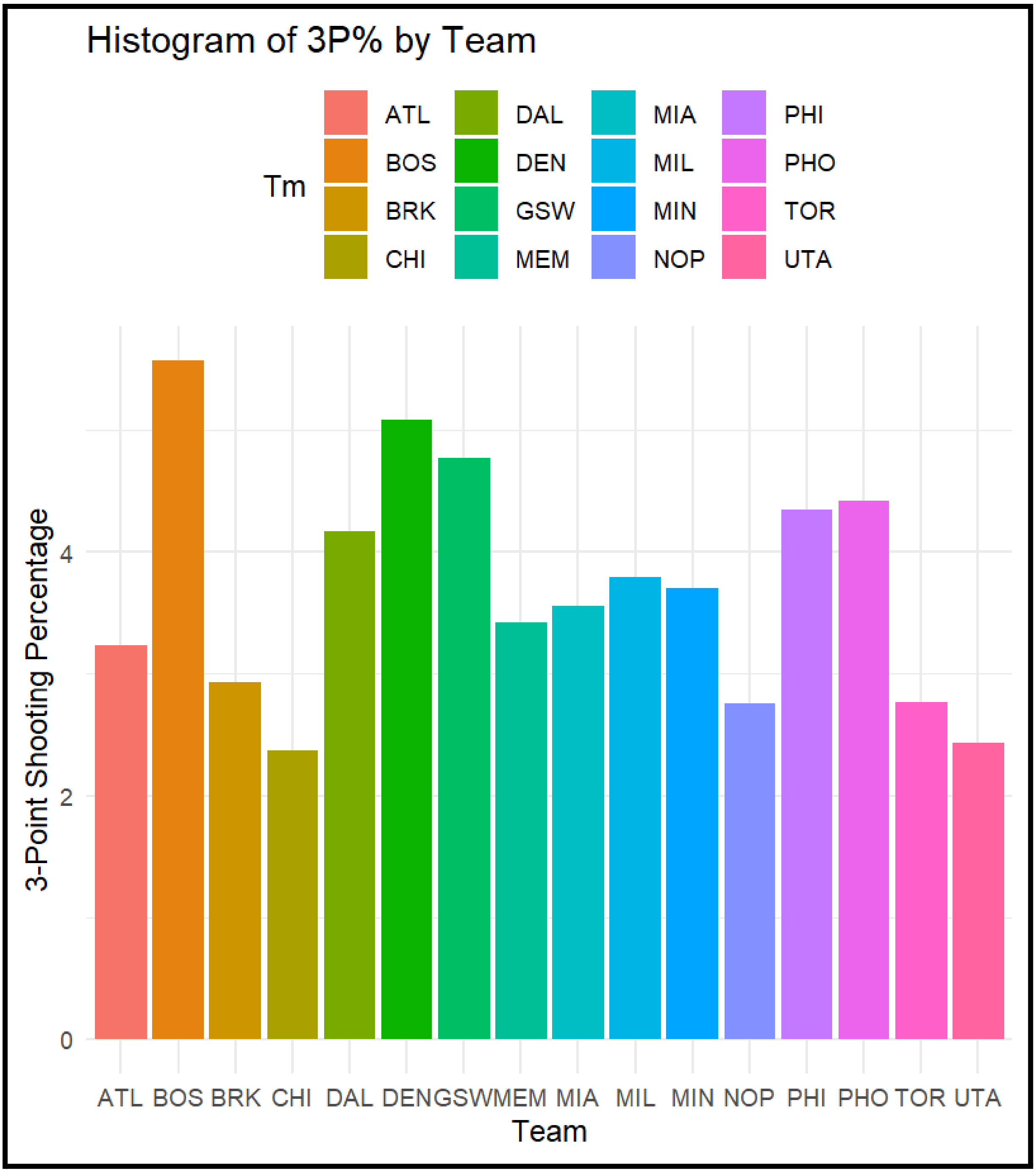
The commands used in Rstudio in order to create the scatter plot are the following:

```
> plot(data_set_nba_playoffs_bueno$Age, data_set_nba_playoffs_bueno$MP,
+ xlab = "Player's Age",
+ ylab = "Minutes Played",
+ main = "scatter plot: Age vs MP")
```

Pearson coefficient:

```
> correlation <- cor(data_set_nba_playoffs_bueno$Age, data_set_nba_playoffs_bueno$MP)
> print(correlation)
[1] 0.1561825
```





This graph shows the relationship between teams and 3P% and how important it is for the NBA Playoffs teams to have players who can score the maximum number of three-point shots possible. This makes sense since these two of the three teams with the highest percentages in the Playoffs were NBA finalists (being Boston and Golden States). This shows that being these two teams the best three-point shooters of Playoffs we can demonstrate the significant importance of this analysis between these two variables.

Because the percentage of 3-point shots was higher than other teams in Playoffs, resulting in more points scored by matches, Boston and Golden State were able to eliminate other teams and qualify to the following rounds.

On the opposite side we have teams like Utah or Chicago Bulls who had the lowest 3-point shooting Percentage successes and because of that, of not having players that score 3-points shots, got eliminated in the first rounds of NBA Playoffs.

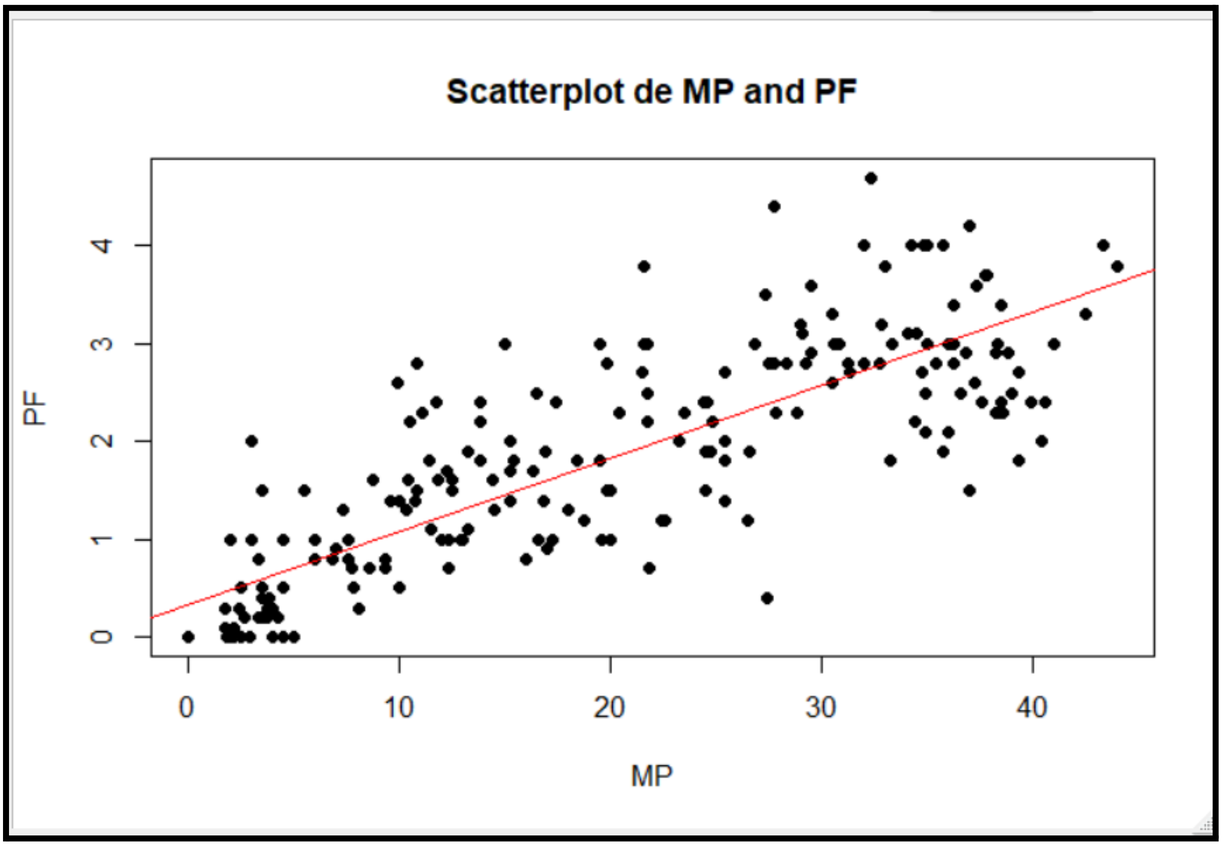
All in all, we can prove that when relating the quantitative variable of 3P% with the variable Team we can see in this case how the ability to score 3 points shots affected teams qualifying or not and getting first, second or the last positions.

For creating the previous graph, we have used the following code in R

```
> library(ggplot2)
> ggplot(data_set_nba_playoffs_el_bueno, aes(x = Tm, y = `3P%`, fill = Tm))
+
+  geom_bar(stat = "identity") +
+  labs(title = "Histogram of 3P% by Team", x = "Team", y = "3-Point
Shooting Percentage") +
+  theme_minimal() +
+  theme(legend.position = "top")
```

```
library(ggplot2)
ggplot(data_set_nba_playoffs_el_bueno, aes(x = `3P%`, y = Tm, color = Tm))
+
  geom_point() +
  labs(title = "Scatterplot of 3-Point Percentage (3P%) by Team (Tm)", x =
"3-Point Percentage", y = "Team") +
  theme_minimal()
```

Analyzing the relationship between MP and PF



Pearson coefficient.
`0.8216469`

We were curious to see how much the minutes played affect the amount of fouls committed by a player. Our hypothesis was that the more you play the more you foul. So, we made a scatter plot and calculated the coefficient of covariance to see how real this hypothesis is.

The Pearson coefficient is 0.8216469, considerably close to 1. This means that the relationship between these variables is pretty strong and that they affect each other widely.

Observing the scatter plot and the slope of the trendline we can see how as the minutes played per player goes up, so does the personal fouls per game, with some exceptions of course. The general trend of these bivariate analysis is that the variables have a positive correlation between each other.

5. <u>Conclussion</u>

In conclusion, this project explains how different variables affect the output on how NBA games succeed. The influence of variables on a team's success in the NBA can vary depending on the team's playing style, strategy, and overall goals. However, there are variables that are considered more influential when it comes to tactics and team success.

PTS: points are the most direct measure of an individual player's offensive contribution.
A high PTS suggests that the player has an ability to score compared to others, which is crucial for a successful NBA team. Offensive progress gives opportunities to develop tactics, especially in late-game situations.
Assists: Assists reflect a player's ability to develop the situation for scoring a point. This player gives the opportunity for teammates to score, and tactics are developed in order for this situations to happen more frequently in a game.
3P%: 3P% is a crucial aspect of basketball, having a player with the ability of scoring several three-pointers makes it possible to change the match leading score, and it is one of the most valuable skills a player can have.
FG%: This variable indicates a player's overall shooting accuracy. It's important for efficient scoring and offensive tactics, as higher shooting percentages can lead to more productive possessions.
Total Rebounds: It is important for both offensive and defensive tactics, as it provides second chance opportunities and limits the opponents opportunities. Teams with strong rebounders can control the pace of the game.
Minutes played: The better a player is, the greater the amount of minutes he will play. However, having a good record of minutes played, gives opportunities for tactics to be made depending on the physical condition a player might be faced to.
Personal Fouls: Foul trouble can limit the effectiveness of defensive tactics. This means that teams are looking for players that do not commit many fouls, in order to have more tactical flexibility.
Age: The age of a player can affect tactics, as old players may not be in the physical conditions in order to play at their best. Also, younger players do not have the sufficient experience to undertake the role of being a leader when games become difficult. This is why the variable age could be very important in order to make tactical decisions.
Position: Positioning of players takes one of the most important roles in basketball. for example, point guards are in charge of ball distribution, while centers are critical for gaining possession of the ball after rebounds.
Team: the composition of the team, including teammates, coaching and chemistry between players, is a fundamental factor, A well-coached team can execute tactics more efficiently.

Finally we made 5 Bivariate analysis in order to prove how these variables were related and the degree of importance regarding each variable.

<u>Analyzing the relation between AST and POSITION</u>: In first place we tried to analyze these two variables to demonstrate that guards (players that play outside of the paint) are the ball keepers and therefore they tend to produce more assists than those who play closer to the paint (centers, power forwards and shooting forwards). This has always been a very common thing in the game of basketball.

Analyzing the relation between PTS and FG%: In this case we wanted to explain how important efficiency as a skills nowadays in the nba, as every team would rather have a player who averages 17 points per game with a 60% success rate on their shots than a player who averages 40 point per game but with a low FG%. As we said in our analysis the MVP of that season was the only player with Jimmy Butler, who managed to score more than 25 ppg with a FG% above 50%.

Analyzing the relation between Age and MP: For this analysis we tried to see if there was any relation between this two variables, and after making a scatter plot and calculating the pearson coefficient we got to the conclusion that our whole hypothesis was wrong, there was close to no relationship between the Age and the Minutes played per player. So what we can get from this is that you always need to prove your thoughts on something, with facts and actual data, because you might get it wrong if not. In this case, it genuinely surprised us what we found after doing the bivariate analysis of this variables.

Analyzing the relationship between Team and 3P%: This is one of the most important points that we Analyzed, as the 3 point shooting has become one of the most important stats in the game of basketball in the last years. Every year the amount of three pointers attempted by all players in the league increases. In the analysis we proved this by saying that the two teams that got to the finals (GSW & BOSTON) were placed 3rd and 1st respectively on efficiency in 3P shot.

Analyzing the relationship between MP and PF: Finally we came up with this basic analysis where we compared the amount of minutes played with the amount of fouls committed and we found out that these two variables were very related after calculating the pearson coefficient, which is pretty obvious.