Writing Research Reports Iteration 3

Eduardo Salvador

April 21, 2022

#Project Methodology

This project will be focusing on deciphering what position inside an NBA team is the most crucial for a team to be successful. It will also try to answer the questions as to which NBA team has the greatest 2pt and 3pt conversion rate as well asif there is a trend as to how the older the player the least effective he is.

To accomplish this goal, most of the visualizations will be scatter plots, histograms, linear regression and possibly clustering some key performance indicators to provide the most optimal solution to the questions above.

1 Introduction

This project is to be used as a way to research each player position to find correlation between the data of the 2020-2021 NBA season to see how has each position in the NBA has change from the traditional playstyle. I want to be able to look at trends in the data that suggests which position has the most and least point conversion where it be 3pt or 2pt and what position has on average the youngest and oldest players. Be aware to not use this model for any type of sports betting since, all models is wrong, only some are useful.

2 Description of the Data

1. I got the data from Kaggle. The link I used was: https://www.kaggle.com/datasets/umutalpaydn/nba-20202021-season-player-stats
2. This data frame has all the NBA players from 2020 and 2021
3. Content: 1.Per Game Stats: Basically dividing every individual stats by the played game . 2.Per 36 Minute Stats: In order to calculate per-36 minute stats, you divide 36 by the number of minutes the player actually played, then take that number and multiply all of the player's stats by it. 3.Advanced Stats: These are more focused on the players direct effect on winning the games or scoring a point. For example a key tenet for many modern basketball analysts is that basketball is best evaluated at the level of possessions. During a single game, both teams have approximately the same number of possessions, because they alternate possession. (A team can have slightly more if it begins and ends a quarter or half with possession.) However, over the course of the season, teams play at very different paces, which can dramatically color their points scored and points allowed per game. Therefore, these analysts favor use of points scored per 100 possessions (offensive rating) and points allowed per 100 possessions (defensive rating).

3 Methods

#Decided to use multiple boxplot method for first graph to get an idea of the data I'm looking at.

#For the second graph, I decided to use the scatterplot to see correlation between 2 variables

#For third graph, I decided to use Histogram+Scatterplot to see the relationship between 3 variables

#For the last graph I decided to use a stack area graph for a fancy way of seeing the relationshipt between 2

4 Results and Discussion

Multiple boxplot

An initial investigation I wanted to know the average player age based upon their position to see if there is a correlation between player position and the amount of experience. Based on the results, the graph shows that the center position in the NBA on average has the oldest players which makes sense since the center position most important attribute is the height of a player and it is really difficult to find tall younger players. The graph also shows the youngest position is the point guard and the role small forward and power forward players. This also makes sense since for this kind of positions the most important attribute is athlethisism and strength in which young players may have the advantage as compared to older players because age does affect this.

Average player age based on position
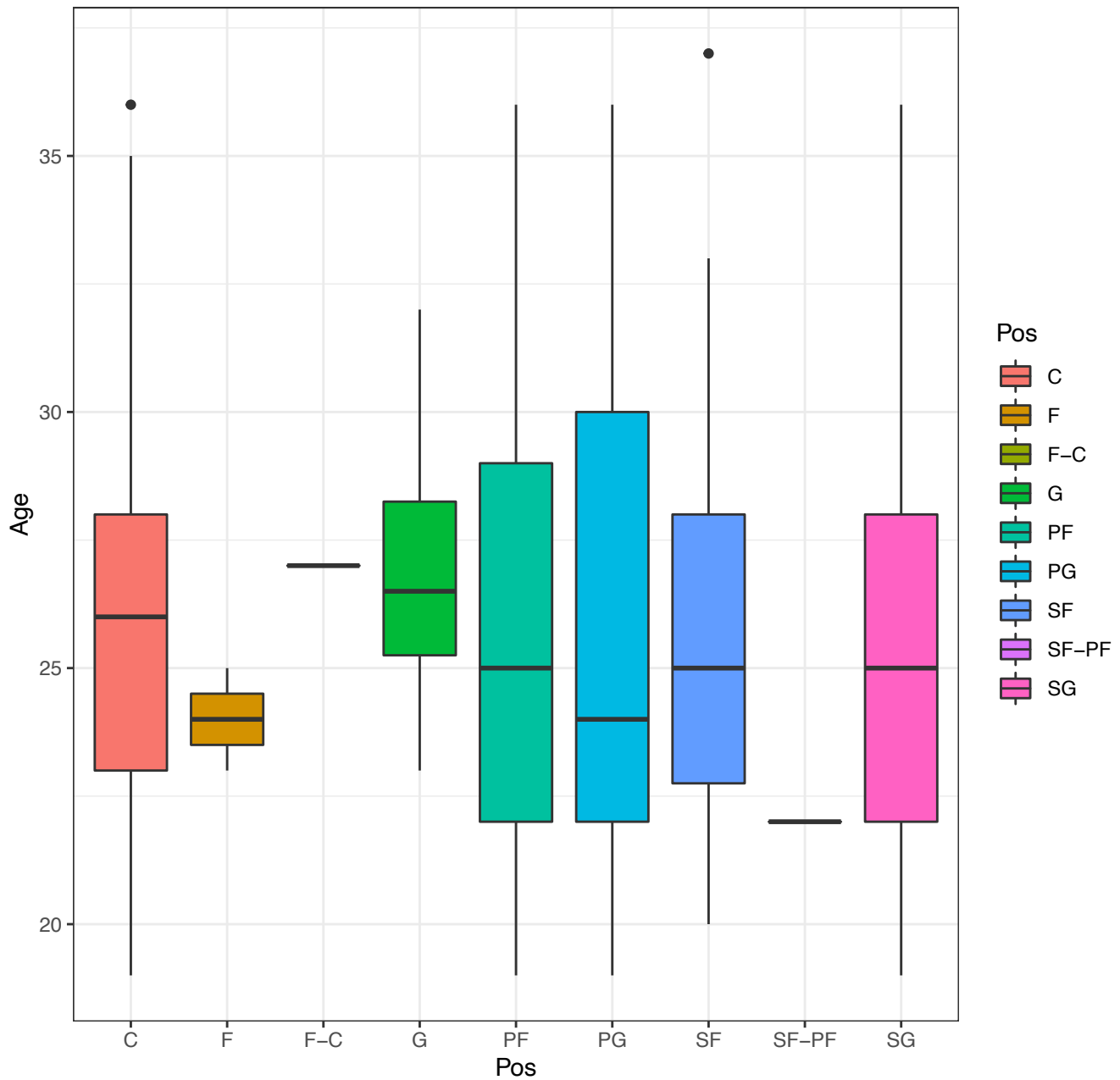
[1] 25.62374

Scatterplot

I wanted to see if there was a correlation between 3 pt attempts and the amount of points scored per position in the NBA. The results from this graph suggests that on average the Shooting guard has the highest percent 3 pt convertion out of all the positions which makes sense since their usual role in the team is to be able to score points from oustide the box. Even so, it is intresting to see the position with most attents and lowest convertion rate from 3 pt is Guard as wheras I thought it was going to be the Center position. This may be explained by the reason as to which centers don't attempt 3 pt as often as guards since they know that is not their role in the team. Further investigation needed for this claim.

Points scored vs 3pt attempt per position
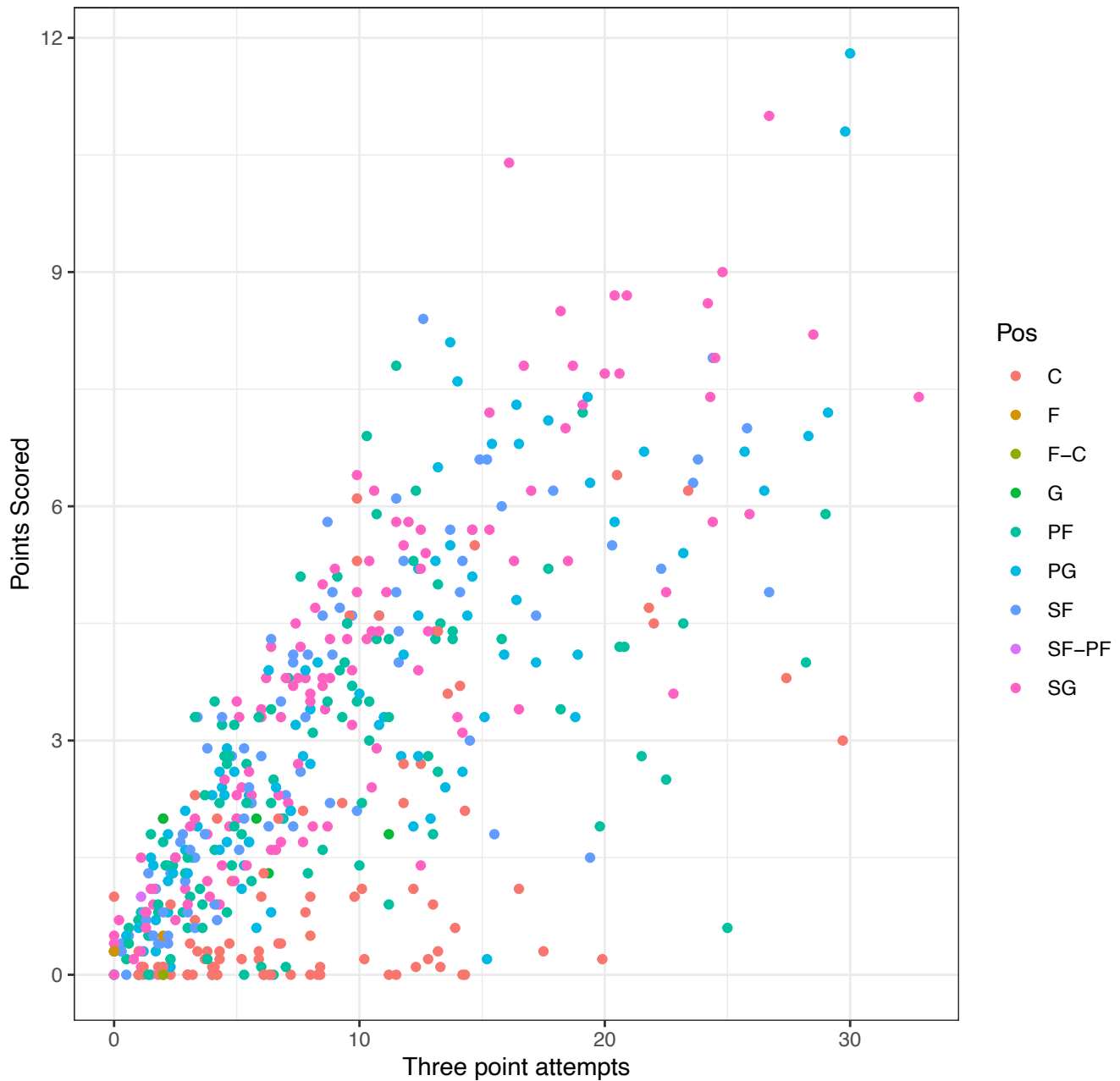

Histogram+Scatterplot

I wanted to see the relationship between 2 point attempts, 3 point attempts and the amount of points players average. The results for this graph is interesting since the histogram suggests that players do attempt more 2 pts than 3 pts but the scatterplot suggests that players who attempt more 3pt do score more points at the end. Further investigation needed for this claim.

```
             Player Pos Age  Tm  G GS   MP  FG  FGA   FG. X3P X3PA  X3P. X2P
1  Precious Achiuwa  PF  21 MIA 28  2 14.6 2.6  4.4 0.590 0.0  0.0 0.000 2.6
2      Jaylen Adams  PG  24 MIL  6  0  2.8 0.2  1.3 0.125 0.0  0.3 0.000 0.2
3      Steven Adams   C  27 NOP 27 27 28.1 3.5  5.8 0.603 0.0  0.0 0.000 3.5
4       Bam Adebayo   C  23 MIA 26 26 33.6 7.4 12.9 0.573 0.1  0.2 0.400 7.3
5 LaMarcus Aldridge   C  35 SAS 18 18 26.7 5.9 12.5 0.476 1.3  3.7 0.358 4.6
6 Ty-Shon Alexander  SG  22 PHO  3  0  2.7 0.0  1.0 0.000 0.0  0.3 0.000 0.0
  X2PA  X2P.  eFG.  FT FTA   FT. ORB DRB TRB AST STL BLK TOV  PF  PTS
1  4.4 0.590 0.590 1.3 2.4 0.561 1.3 2.7 4.0 0.6 0.4 0.5 1.0 1.9  6.5
2  1.0 0.167 0.125 0.0 0.0 0.000 0.0 0.5 0.5 0.3 0.0 0.0 0.0 0.2  0.3
```

```
3   5.7 0.606 0.603 1.1 2.3 0.468 4.3 4.6 8.9 2.1 1.0 0.6 1.7 1.9   8.0
4  12.7 0.576 0.576 5.1 6.0 0.841 1.9 7.3 9.2 5.3 1.0 1.0 3.0 2.6  19.9
5   8.8 0.525 0.529 0.9 1.2 0.762 0.8 3.5 4.3 1.9 0.4 0.9 0.9 1.5  14.1
6   0.7 0.000 0.000 0.0 0.0 0.000 0.0 0.3 0.3 0.3 0.0 0.0 0.0 0.3   0.0
```

Points convertion based on 2pt and 3pt attempts
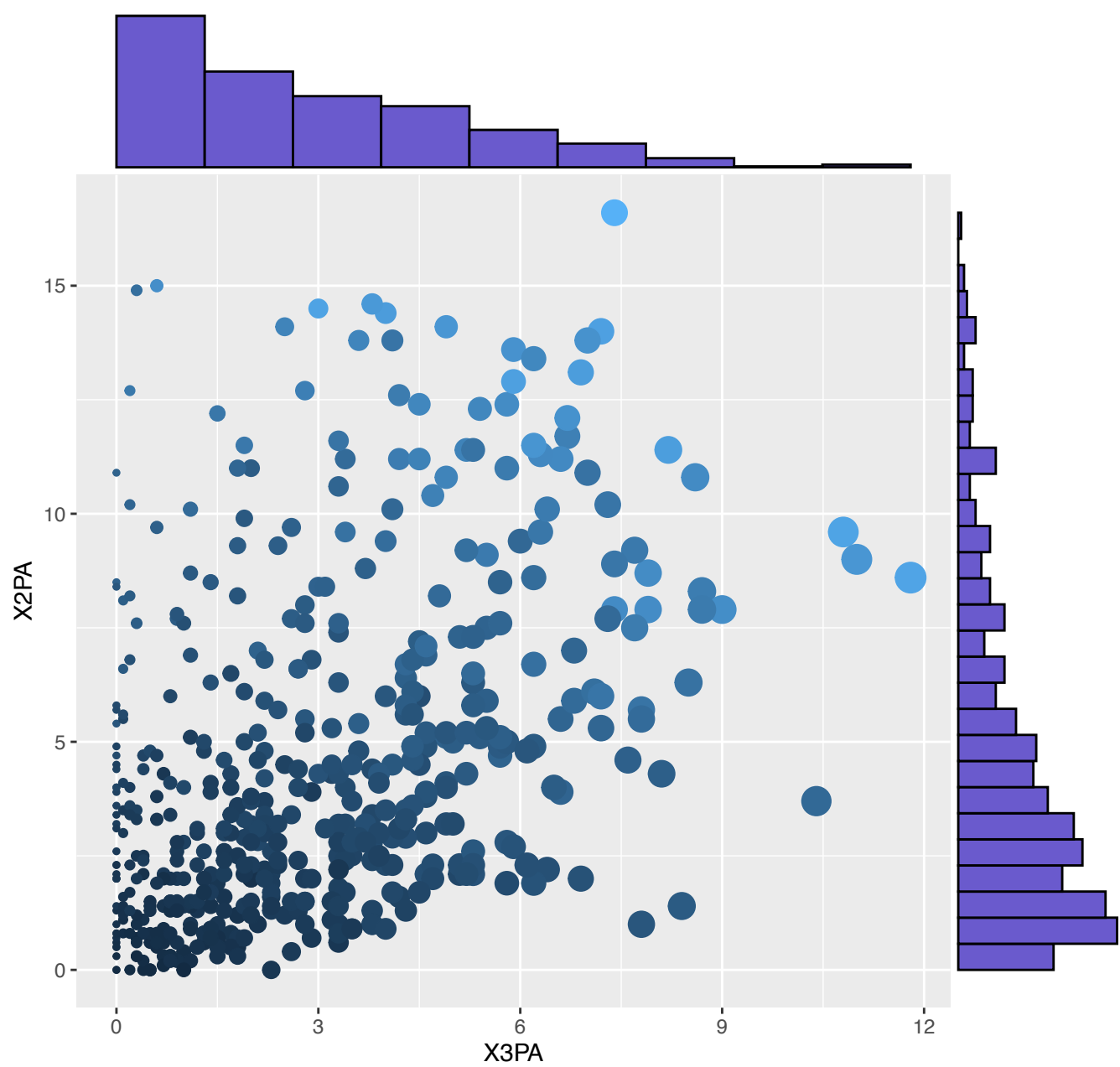
Stacked Area Graph

I did this without thinking I was going to find anything intresting but, I found out that the Center position for some reason creates more field goal attempts than any other position. The result from this graph is intresting since it suggests that the center position have usually more field goal attempts than any other position with the least converstion rate.

Number of points per field goal attempt
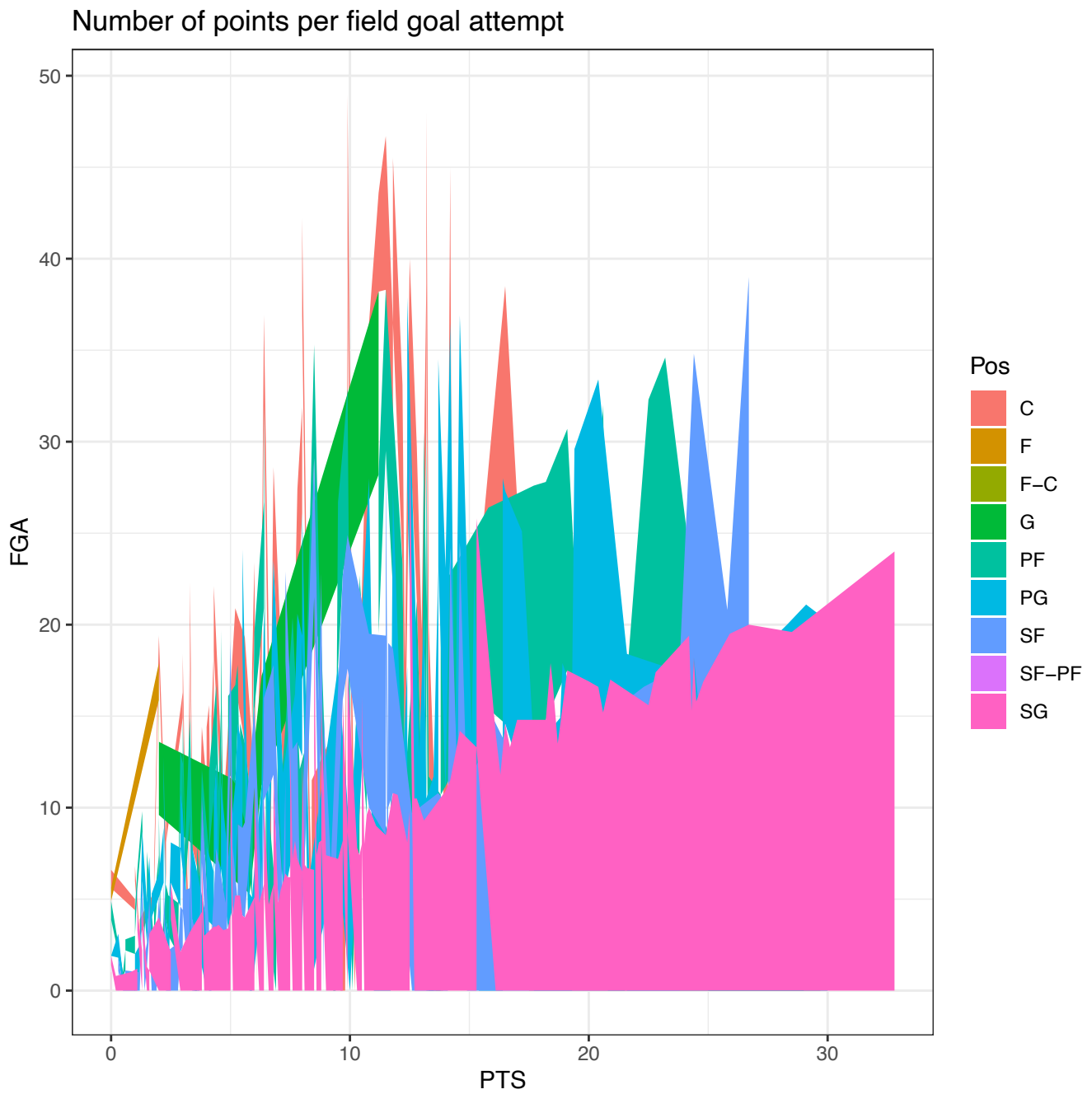
5 Conclusions

6 Citation

7 Appendix

##Graph 1

library(ggplot2)

library(tidyverse)

library(dplyr)

library(olsrr)

library(pander)

#Reading the file

dfadplayers<-read.csv("/Users/eduardosalvador/Desktop/FINAL Spring Semester 2021/CMDA /Assignments/Project 1/archive/nba
= T)

dfpergame<-read.csv("/Users/eduardosalvador/Desktop/FINAL Spring Semester 2021/CMDA /Assignments/Project 1/archive/nba2
= T)

ggplot(dfadplayers,aes(x=Pos,y=Age,fill=Pos))+theme_bw()+geom_boxplot()+ggtitle("Average player age based on position")

mean(dfadplayers$Age)

##Graph 2

library(GGally)

ggplot(dfpergame,aes(x=PTS,y=X3PA,color=Pos))+theme_bw()+ggtitle("Number of points per position based on field goal attemps")+geom_point()+theme_bw() + labs(x = "Three point percentage", y = "Field Goal percentage", title = "Points scored vs 3pt attempt")

##Graph 3

library(ggplot2)

library(ggExtra)

##The mtcars dataset is proposed in R

head(dfpergame)

##classic plot :

p <- ggplot(dfpergame, aes(x=PTS, y=X2PA, color=X3PA, size=X3PA)) + geom_point() + theme(legend.position="none")+ggtitle based on 2 pt and 3 pt attempts average per game")

##Custom marginal plots:

p2 <- ggMarginal(p, type="histogram", fill = "slateblue", xparams = list(bins=10))

p2

##Graph 4

ggplot(dfpergame,aes(y=FGA,x=PTS, fill=Pos))+geom_area()+theme_bw()+ggtitle("Number of points per field goal attempt")

" '

etc...