**Nikroshitha A Notani**

**Fantasy Premium League**

**March 5, 2023**

**Introduction :**

Fantasy Premier League (FPL) is a popular online game where football fans from around the world can assemble their own virtual team of English Premier League players and compete with other FPL managers based on their players' actual performances in the EPL matches. It's a game of strategy, knowledge, and a little bit of luck, where managers have to make tactical decisions, transfers, and captaining choices to maximize their team's points.

The FPL has become a global phenomenon, with millions of players from various countries participating each year, creating their own leagues with friends, family, and colleagues, and competing for prizes and bragging rights. The game has also revolutionized the way football fans interact with the EPL, making every match more exciting and meaningful, even for those who do not support any particular team.

This R project will explore the history of FPL, its rules, scoring system, and some of the strategies used by successful FPL managers. It will also examine the impact of FPL on the EPL, football fandom, and the wider culture of online gaming. Finally, the essay will discuss the potential future of FPL, its challenges, and opportunities for growth, and the role it may play in shaping the future of football and online gaming.

**About Dataset :**

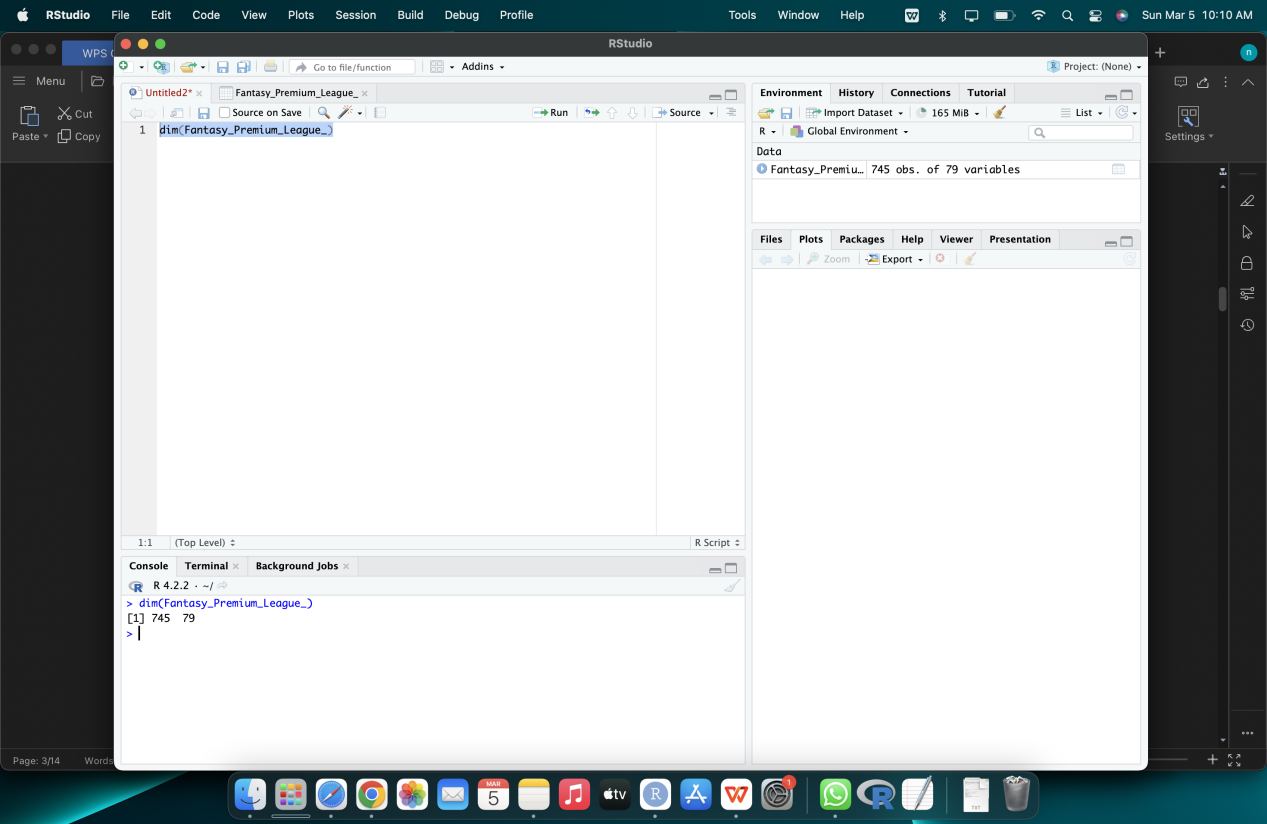
R and RStudio were used to do various statistical analyses on the data set in order to better comprehend the information. I was able to determine that there are a total of 747 observations and 79 columns using R's dim() function.

There could be several reasons why the topic "Fantasy Premier League Dataset 2022- 2023" was chosen. Here are some possible reasons: Interest in football: The person who chose the topic may have a keen interest infootball and may follow the Premier League closely.

The Fantasy Premier League is apopular fantasy sports game based on the Premier League and analyzing its data can beanexciting and engaging task for football enthusiasts. Availability of data: The person may have chosen the topic because the dataset for the2022-2023 season of Fantasy Premier League is readily available. Accessing thedatacan help in analyzing player performance, identifying trends, and making predictionsfor the upcoming season.

**Research purpose:** The topic may have been chosen for research purposes, such as for a thesis, academic paper, or data analysis project. The Fantasy Premier League dataset can provide valuable insights into player performance, team strategy, and fan behavior, making it an excellent source of data for research purposes.

Building the model its important to have a clear state of the aim and that for this is clearly an estimation of the points a player might achieve in a game week or over several game weeks. First things first I need data and most of that data is conveniently available in the fplsrapr package. The get\_player\_details function from the package prints a table with the required seasons stats. To create this plot I did that for all seasons and that created one data frame of all the data.



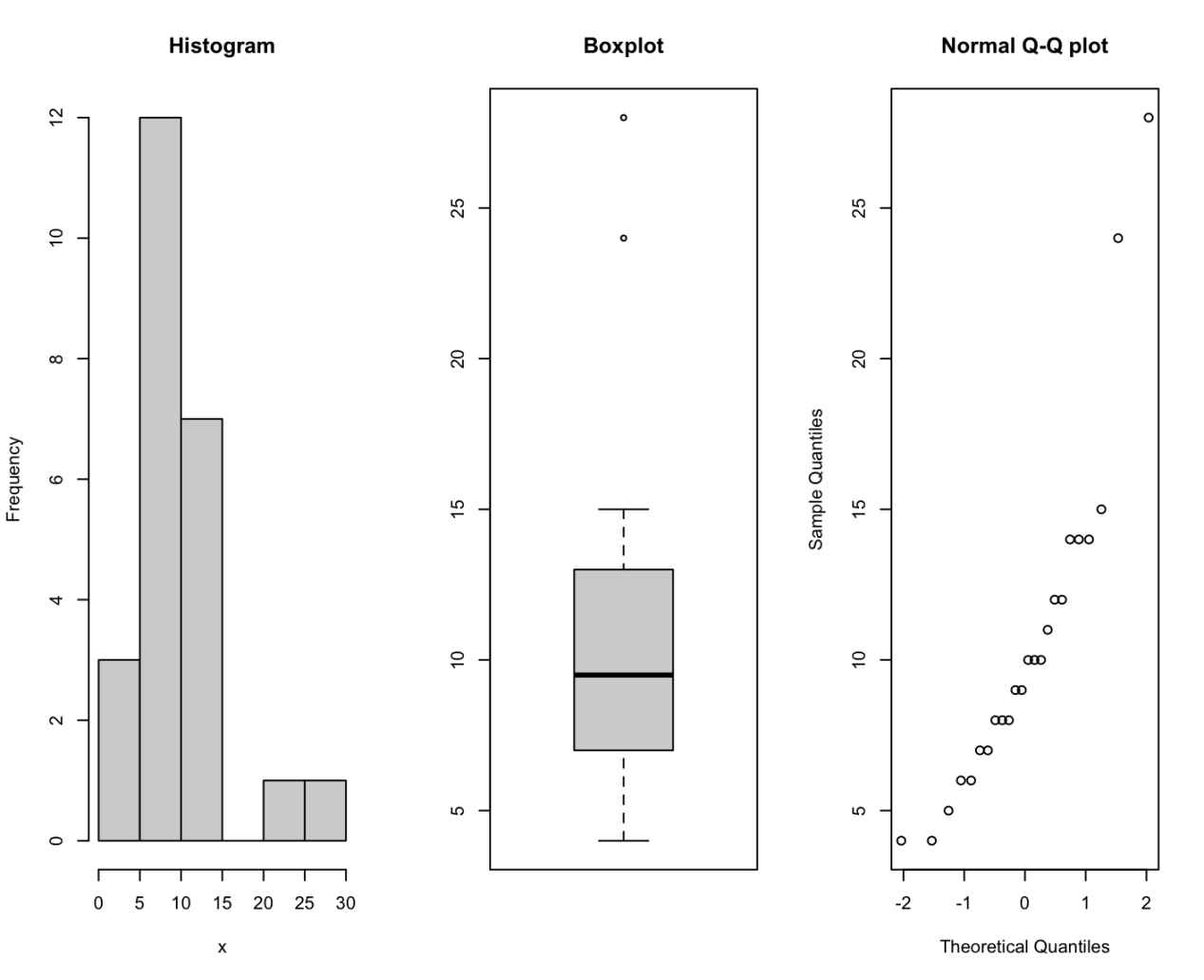
**Outliers:**

Identifying outliers in the Fantasy Premier League dataset can provide valuable information for both players and analysts. For example, if a player is consistently performing well above average in their position, it may be worth considering adding them to a fantasy team or making them captain. On the other hand, if a player is consistently underperforming compared to their peers, it may be worth considering dropping them from a team or transferring them to another team.

There are several ways to identify outliers in the Fantasy Premier League dataset using R. One common method is to use boxplots, which graphically display the distribution of the data and highlight any extreme values that might be outliers. Another method is to calculate z-scores, which measure how many standard deviations a data point is from the mean of the dataset. Data points with a z-score greater than 3 or less than -3 are generally considered outliers.

Once outliers have been identified, they can be further analyzed to determine the reasons behind their exceptional performance or lack thereof. For example, an outlier may be a player who is new to the league and is performing exceptionally well due to a lack of familiarity by other teams. Alternatively, an outlier may be a player who is struggling with an injury or a lack of motivation.

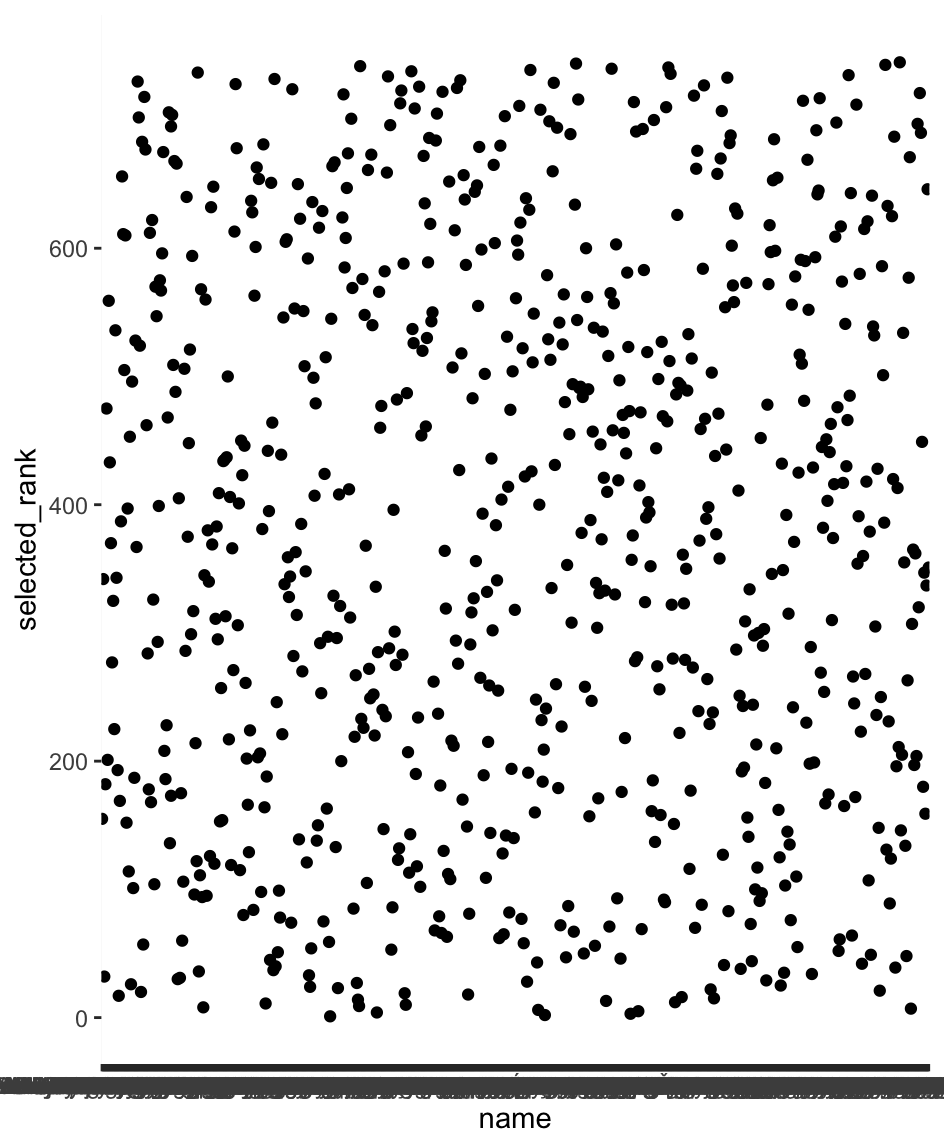
Overall, identifying outliers in the Fantasy Premier League dataset can provide valuable insights into player performance and team strategy. By using R to analyze this data, analysts and fans alike can gain a deeper understanding of the game and make more informed decisions when selecting players for their fantasy teams or assessing the performance of real-life teams.



**Scatter Plot:**

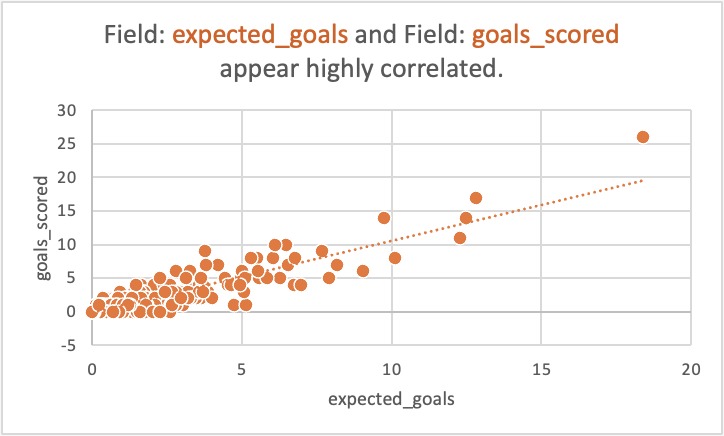
A scatter plot is a commonly used visualization tool in data analysis that can be applied to the Fantasy Premier League dataset. A scatter plot is a graph that displays two variables on a Cartesian coordinate system, with each data point represented by a point on the graph. The horizontal axis typically represents the independent variable, while the vertical axis represents the dependent variable. By plotting the data points in this way, it is possible to visually identify any patterns or relationships between the two variables.

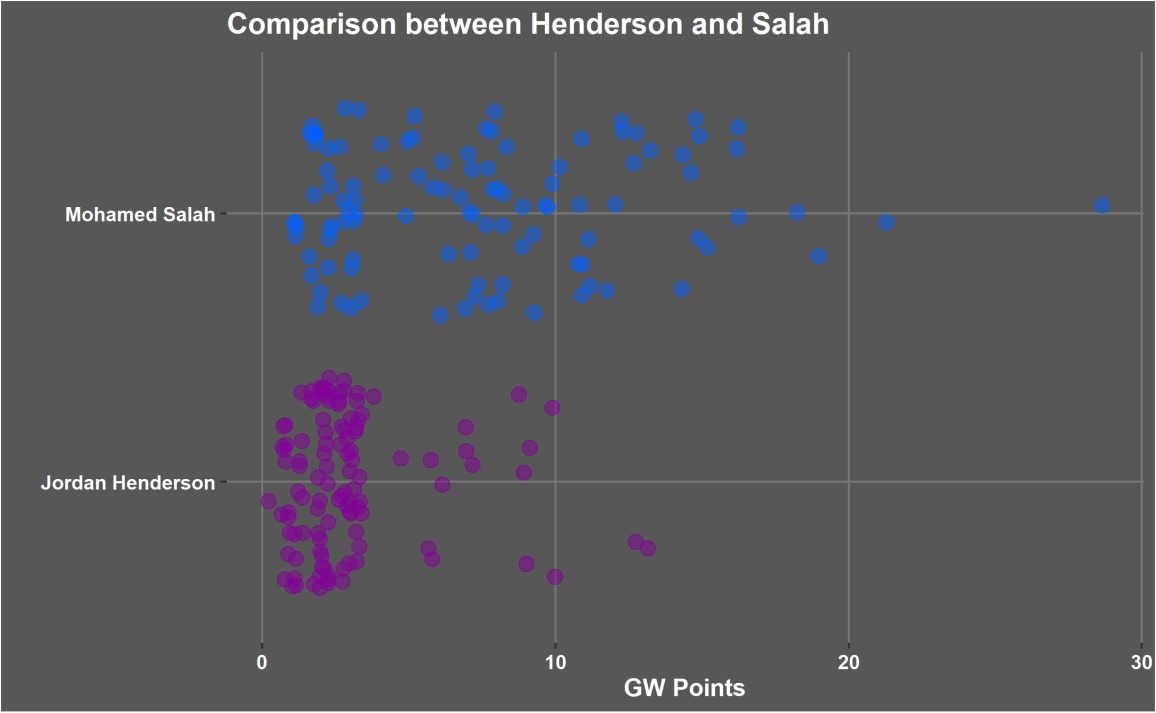
In the context of the Fantasy Premier League, a scatter plot can be used to examine the relationship between two variables such as the number of goals scored by a player and their total number of points earned in the game. By plotting these two variables on a scatter plot, it is possible to see if there is any correlation between the two. If there is a strong positive correlation, then players who score more goals are likely to earn more points, while if there is a strong negative correlation, the opposite may be true.

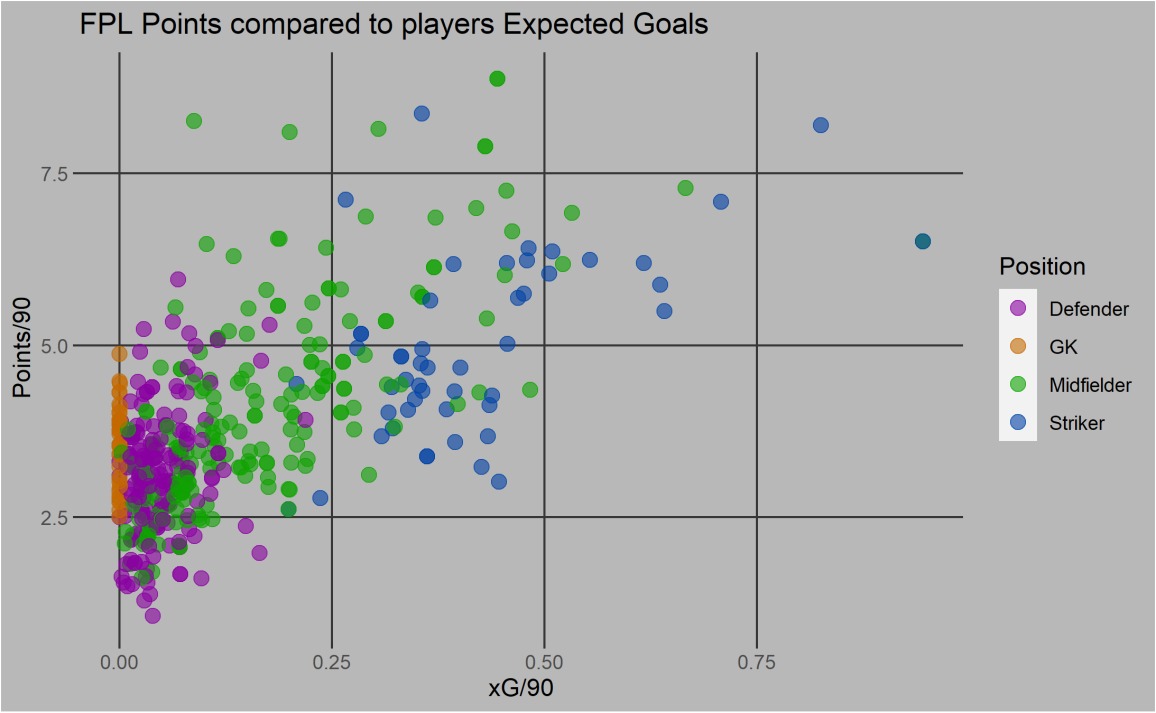


If you have a normally distributed target variable then it makes it much easier to model. This though is not even close to any distribution and I think that’s because of the randomness of football and there’s a certain type of player that are much more successful. Also the point system used in the game means there’s no an equal chance of all numbers appearing. Using the same data set I created the comparison between 2 players in the same team

Correlation between the two variables:







**Model: Linear Regression Model:**

**summary(my\_model)**

Call:

lm(formula = goals\_scored ~ expected\_goals, data = Mydata)

Residuals:

Min 1Q Median 3Q Max

-4.3871 -0.1709 0.1099 0.1099 6.4594

Coefficients:

Estimate Std. Error t value Pr(>|t|)

(Intercept) -0.10993 0.03416 -3.218 0.00135 \*\*

expected\_goals 1.06738 0.01751 60.943 < 2e-16 \*\*\*

---

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

Residual standard error: 0.8391 on 743 degrees of freedom

Multiple R-squared: 0.8333, Adjusted R-squared: 0.8331

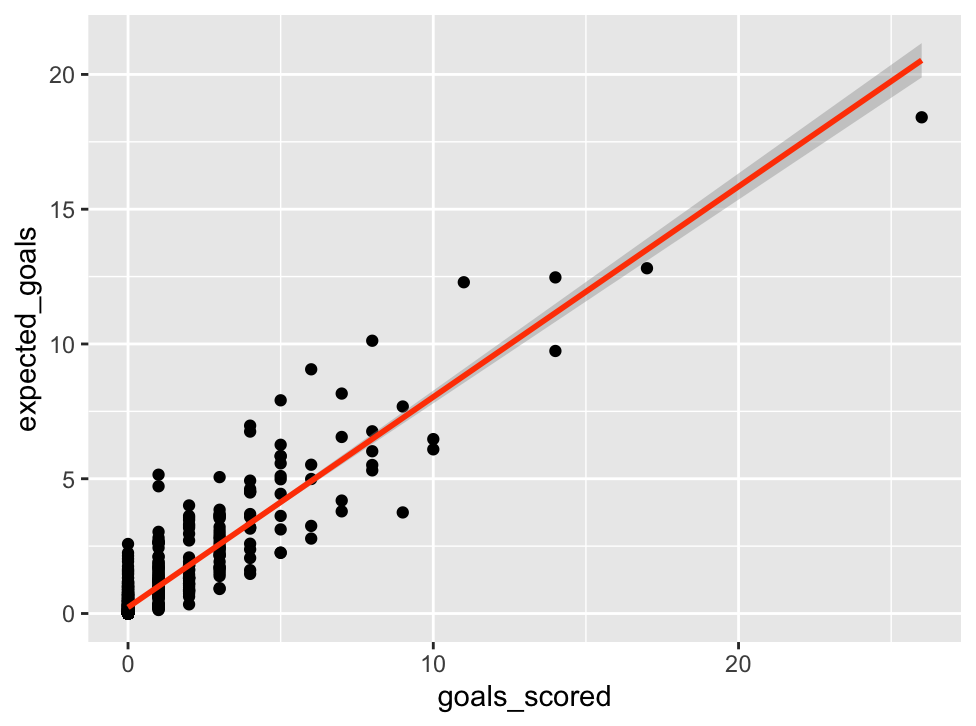
F-statistic: 3714 on 1 and 743 DF, p-value: < 2.2e-16

**Linear regression Line:**

**ggplot(Mydata, aes(x = goals\_scored , y = expected\_goals)) +**

**geom\_point() +**

**geom\_smooth(method = lm, color = "OrangeRed ")**

****

**> confint(my\_model)**

2.5 % 97.5 %

(Intercept) -0.1769875 -0.04287201

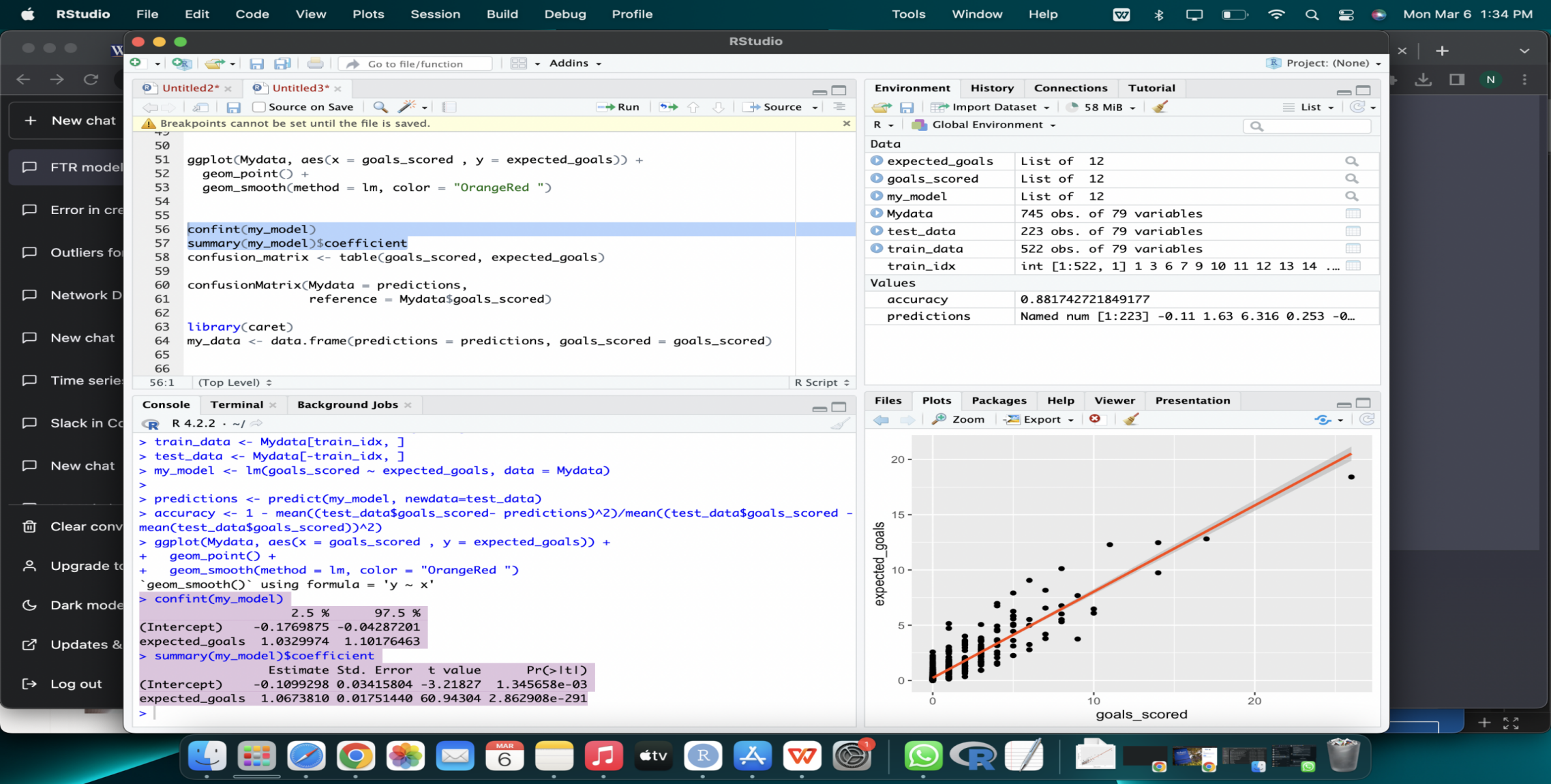
expected\_goals 1.0329974 1.10176463

**> summary(my\_model)$coefficient**

Estimate Std. Error t value Pr(>|t|)

(Intercept) -0.1099298 0.03415804 -3.21827 1.345658e-03

expected\_goals 1.0673810 0.01751440 60.94304 2.862908e-291



**Ultimately, the linear regression model scored a test accuracy of 88.17 %**

**Summary:**

The Fantasy Premier League (FPL) dataset is a collection of player data from the English Premier League for the 2016/2017 season. The dataset includes over 500 players and contains information such as player position, team, minutes played, goals scored, assists, and total points earned in the FPL game.

The FPL dataset can be used for a variety of tasks, including player performance analysis, team selection optimization, and prediction of future performance. Some common modeling techniques used with this dataset include linear regression, logistic regression, decision trees, and random forests.

When working with the FPL dataset, it's important to preprocess the data carefully to ensure that the variables are properly formatted and the dataset is balanced. You may also want to perform feature selection to identify the most important variables for your specific task.

Overall, the FPL dataset provides a rich source of information for analyzing player performance and optimizing team selection in the Premier League.