# American International University, Bangladesh

## Introduction to Data Science

**Project**

**Final Term**

**Topic:** Web Scrapping

**Section:** C

| No. | Name | ID |
|-----|------|-----|
| 1 | Saleheen, Mohammad Mushfiq Us | 19-39282-1 |
| 2 | Ahmed, Zisan | 19-39294-1 |
| 3 | Raidah, Fatema Nur | 20-43385-1 |
| 4 | Sohada, Khandaker Fatema Nur | 20-43386-1 |

Dr. AKINUL ISALAM JONY

Associate Professor
Computer Science

# Project Overview

The project consists with two datasets. These datasets are scrapped from a website called "fifa index" here they publish the ratings of the football clubs all over the world every year. Here in our data basically contains the Attacking, Midfield, Defense, and the overall ratings. Also, the corresponding teams name and the league name where they belong to are also included. First dataset was collected from the 2022 edition of top 30 clubs and the 2$^{nd}$ datasets are collected from 2021 edition. The attributes are common in the datasets, and they are,

- Name of the top 30 clubs
- Corresponding League names
- Attack, defense, midfield, and overall rating lists for 30 clubs of two datasets.

After the scrapping the data then we use necessary cleaning process and data handling methods as needed. Lastly, we will give a descriptive statistics analysis and a graph for final analysis.

# Project solution design

The solution for this project is, we scrapped the data from the desired website using selectorgadget tool and Rstudio. After importing the data to a csv file. Then we will clean and sorted data for our desired purpose through data pre-processing. This requires few steps of data pre-processing which are:

1. Data Cleaning
   - Smooth Noisy Data
   - Handling Missing Data
   - Data Wrangling or Munging
2. Data Reduction
3. Data Transformation
4. Data Integration
5. Data Discretization

These steps are required for processing the data to get a clean dataset for our desired work. Some data are dirty and noisy that is they are incomplete or is not consistence that is they have a different value set which might have resulted from typing error or any other mistake, data might be missing, or data might have extreme values that do not go with dataset ranges, data can have many more errors and outliers, negative values, small digits than its range, might have alphabetical or numerical errors. To overcome all these in our dataset, we will be preprocessing our data now.

Then by using quantitative analysis and simple graphics we will be able to give a descriptive Statistics report. At the very end, a plotting graph will be shown to visualize the data.
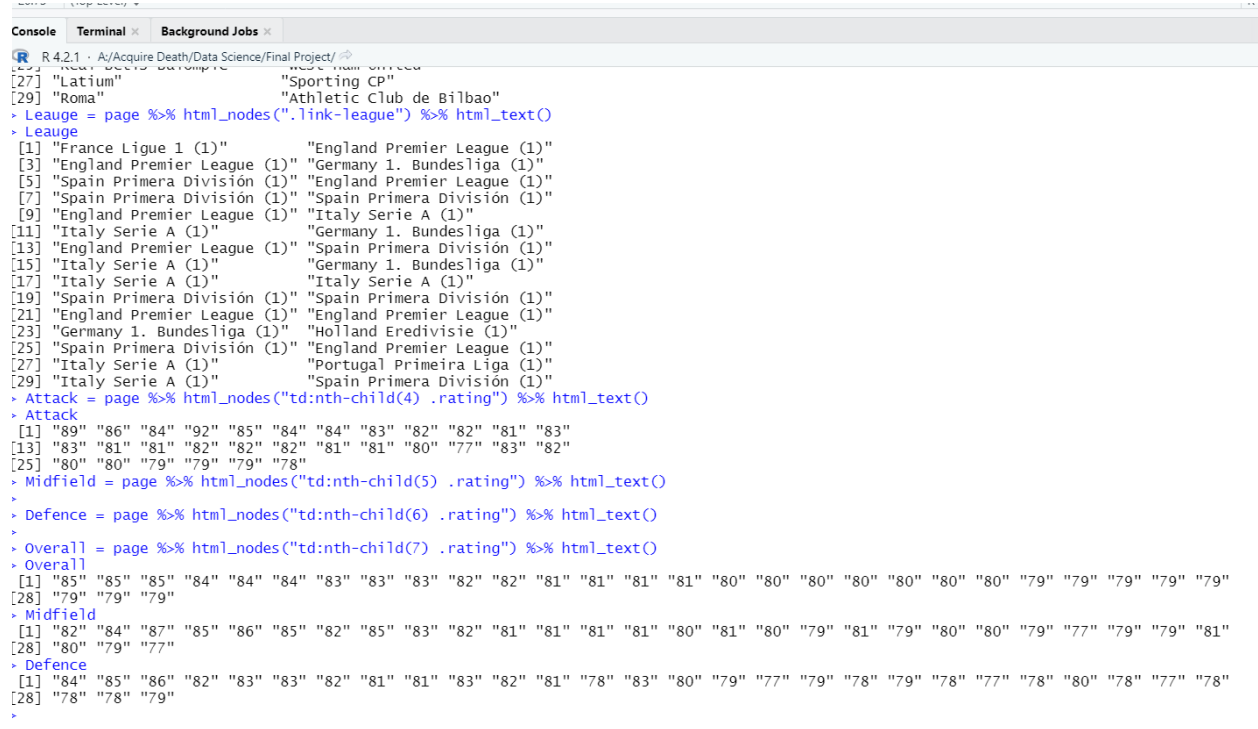
# Data Scrapping

Data is scrapped from "https:\\ifaindex.com/teams/" this website. The scrapping process is done with SelectorGadget.

**Code:**

**For FIFA 22 Dataset**

```
#install.packages("rvest")
#install.packages("dplyr")
library(rvest)
library(dplyr)
link = "https://www.fifaindex.com/teams/fifa22/"
page = read_html(link)

Club = page %>% html_nodes("td+ td .link-team") %>% html_text()
Leauge = page %>% html_nodes(".link-league") %>% html_text()
Attack = page %>% html_nodes("td:nth-child(4) .rating") %>% html_text()
Midfield = page %>% html_nodes("td:nth-child(5) .rating") %>% html_text()
Defence = page %>% html_nodes("td:nth-child(6) .rating") %>% html_text()
Overall = page %>% html_nodes("td:nth-child(7) .rating") %>% html_text()
FIFA22 = data.frame(Club, Leauge,Attack,Midfield,Defence,Overall, stringsAsFactors = FALSE)
write.csv(FIFA22, "fifa22.csv")
```

**For FIFA 21 Dataset**
library(rvest)
library(dplyr)

```
link2 = "https://www.fifaindex.com/teams/fifa21_486/"
page2 = read_html(link2)
Club = page2 %>% html_nodes("td+ td .link-team") %>% html_text()
Leauge = page2 %>% html_nodes(".link-league") %>% html_text()
Attack = page2 %>% html_nodes("td:nth-child(4) .rating") %>% html_text()
Midfield = page2 %>% html_nodes("td:nth-child(5) .rating") %>% html_text()
Defence = page2 %>% html_nodes("td:nth-child(6) .rating") %>% html_text()
Overall = page2 %>% html_nodes("td:nth-child(7) .rating") %>% html_text()
FIFA21 = data.frame(Club, Leauge,Attack,Midfield,Defence,Overall, stringsAsFactors = FALSE)
write.csv(FIFA21, "fifa21.csv")
```

## Datasets

| | Club | Leauge | Attack | Midfield | Defence | Overall |
|---|---|---|---|---|---|---|
| 1 | Paris Saint-Germai | France Ligue 1 (1) | 89 | 82 | 84 | 85 |
| 2 | Liverpool | England Premier League ( | 86 | 84 | 85 | 85 |
| 3 | Manchester City | England Premier League ( | 84 | 87 | 86 | 85 |
| 4 | Bayern MÃ¼nchen | Germany 1. Bundesliga (1 | 92 | 85 | 82 | 84 |
| 5 | Real Madrid | Spain Primera DivisiÃ³n (1 | 85 | 86 | 83 | 84 |
| 6 | Chelsea | England Premier League ( | 84 | 85 | 83 | 84 |
| 7 | AtlÃ©tico de Madi | Spain Primera DivisiÃ³n (1 | 84 | 82 | 82 | 83 |
| 8 | FC Barcelona | Spain Primera DivisiÃ³n (1 | 83 | 85 | 81 | 83 |
| 9 | Manchester Unite( | England Premier League ( | 82 | 83 | 81 | 83 |
| 10 | Inter | Italy Serie A (1) | 82 | 82 | 83 | 82 |
| 11 | Juventus | Italy Serie A (1) | 81 | 81 | 82 | 82 |
| 12 | Borussia Dortmun( | Germany 1. Bundesliga (1 | 83 | 81 | 81 | 81 |
| 13 | Tottenham Hotspu | England Premier League ( | 83 | 81 | 78 | 81 |
| 14 | Sevilla FC | Spain Primera DivisiÃ³n (1 | 81 | 81 | 83 | 81 |
| 15 | Milan | Italy Serie A (1) | 81 | 80 | 80 | 81 |
| 16 | RB Leipzig | Germany 1. Bundesliga (1 | 82 | 81 | 79 | 80 |
| 17 | Napoli | Italy Serie A (1) | 82 | 80 | 77 | 80 |
| 18 | Bergamo Calcio | Italy Serie A (1) | 82 | 79 | 79 | 80 |
| 19 | Real Sociedad | Spain Primera DivisiÃ³n (1 | 81 | 81 | 78 | 80 |
| 20 | Villarreal CF | Spain Primera DivisiÃ³n (1 | 81 | 79 | 79 | 80 |
| 21 | Leicester City | England Premier League ( | 80 | 80 | 78 | 80 |
| 22 | Arsenal | England Premier League ( | 77 | 80 | 77 | 80 |
| 23 | Bayer 04 Leverkus | Germany 1. Bundesliga (1 | 83 | 79 | 78 | 79 |
| 24 | Ajax | Holland Eredivisie (1) | 82 | 77 | 80 | 79 |
| 25 | Real Betis Balomp | Spain Primera DivisiÃ³n (1 | 80 | 79 | 78 | 79 |
| 26 | West Ham United | England Premier League ( | 80 | 79 | 77 | 79 |
| 27 | Latium | Italy Serie A (1) | 79 | 81 | 78 | 79 |
| 28 | Sporting CP | Portugal Primeira Liga (1) | 79 | 80 | 78 | 79 |
| 29 | Roma | Italy Serie A (1) | 79 | 79 | 78 | 79 |
| 30 | Athletic Club de Bi | Spain Primera DivisiÃ³n (1 | 78 | 77 | 79 | 79 |

FIG: Dataset 1 (fifa22)

| | Club | Leauge | Attack | Midfield | Defence | Overall | |
|---|---|---|---|---|---|---|---|
| 1 | Liverpool | Premier League | 86 | 83 | 80 | 85 | |
| 2 | Manchester City | Premier League | 85 | 86 | 83 | 85 | |
| 3 | Real Madrid | LaLiga Santander | 84 | 86 | 84 | 85 | |
| 4 | FC Bayern MÃ¼nchen | Bundesliga | 92 | 85 | 82 | 84 | |
| 5 | Paris Saint-Germain | Ligue 1 Uber Eats | 87 | 82 | 82 | 84 | |
| 6 | Juventus | Serie A TIM | 87 | 81 | 83 | 83 | |
| 7 | AtlÃ©tico de Madrid | LaLiga Santander | 85 | 82 | 83 | 83 | |
| 8 | FC Barcelona | LaLiga Santander | 84 | 82 | 81 | 83 | |
| 9 | Tottenham Hotspur | Premier League | 89 | 82 | 81 | 82 | |
| 10 | Manchester United | Premier League | 83 | 82 | 82 | 82 | |
| 11 | Inter | Serie A TIM | 83 | 81 | 83 | 82 | |
| 12 | Chelsea | Premier League | 82 | 84 | 82 | 82 | |
| 13 | Borussia Dortmund | Bundesliga | 87 | 81 | 81 | 81 | |
| 14 | Sevilla FC | LaLiga Santander | 80 | 81 | 82 | 81 | |
| 15 | Arsenal | Premier League | 83 | 79 | 79 | 80 | |
| 16 | Napoli | Serie A TIM | 82 | 80 | 80 | 80 | |
| 17 | Lazio | Serie A TIM | 82 | 80 | 79 | 80 | |
| 18 | Villarreal CF | LaLiga Santander | 82 | 79 | 79 | 80 | |
| 19 | Leicester City | Premier League | 80 | 81 | 79 | 80 | |
| 20 | RB Leipzig | Bundesliga | 78 | 81 | 80 | 80 | |
| 21 | Milan | Serie A TIM | 84 | 78 | 80 | 79 | |
| 22 | Atalanta | Serie A TIM | 82 | 79 | 78 | 79 | |
| 23 | Real Sociedad | LaLiga Santander | 81 | 81 | 77 | 79 | |
| 24 | Everton | Premier League | 81 | 79 | 80 | 79 | |
| 25 | Athletic Club | LaLiga Santander | 80 | 78 | 79 | 79 | |
| 26 | Borussia MÃ¶nchengladt | Bundesliga | 78 | 79 | 80 | 79 | |
| 27 | Bayer 04 Leverkusen | Bundesliga | 78 | 78 | 78 | 79 | |
| 28 | Wolverhampton Wandere | Premier League | 77 | 81 | 78 | 79 | |
| 29 | SL Benfica | Liga NOS | 77 | 79 | 78 | 79 | |
| 30 | Ajax | Eredivisie | 80 | 77 | 76 | 78 | |

FIG: Dataset 2 (fifa21)

## Data pre-processing:

Data preprocessing is one of the most crucial parts for data analysis. Without data preprocessing data scientists can never produce an accurate data for their research. Hence, data preprocessing is an integral part data science or any field. There are various processes to clean and process data before they are utilized for data analysis. They are the following:

1. **Data Cleaning**
- Smooth Noisy Data:

Now smoothy noisy data mean if there is any negative value in the Dataset then it needed to be handled. But in our case the dataset contains no such data. So, we can simply skip this part of the process.

- Handling Missing Data:

Though we don't have missing value here in our datasets yet we are writing missing data handling code so that there is any missing data then it will be handled using mean function.

**Code:**
**#Missing Data handle**
**mean(fifa22$Attack)**
**meanAs <- mean(fifa22$Attack, na.rm = TRUE)**
**print(meanAs)**

**fifa22[is.na(fifa22$Attack), "Attack"] <- meanAs**
**print(fifa22)**

## 2. Data Integration

Here no need for data integration but we write code to classify the teams give them a tag if above 86 the S means super class, 82 then A class then 80 for B and below 80 is C class. Though it is unnecessary for our datasets.

Code:
```
#Data Integration
for(i in 1:nrow(fifa22)){
  if(fifa22$overall[i] <= 86 ){
    fifa22$Type[i] <- "S"
  }else if(fifa22$overall [i] <= 82){
    fifa22$Type [i] <- "A"
  }else if(fifa22$overall [i] <= 80){
    fifa22$Type [i] <- "B"
  }else { fifa22$Type [i] <- "C"}
}
```

## 3. Data Reduction: We do data reduction to reduce unnecessary data from the dataset that can be used to attain the same results. In our dataset,

fifa22, there were additional (1). We have reduced the additional 1s in brackets through data reduction.

**Code:** fifa22$Leauge <- gsub("[(1)]","",fifa22$Leauge)

view(fifa22)

RStudio Source Editor

fifa22 ×

Filter

| | X | Club | Leauge | Attack | Midfield | Defence | Overall |
|---|---|------|--------|--------|----------|---------|---------|
| 1 | 1 | Paris Saint-Germain | France Ligue | 89 | 82 | 84 | 85 |
| 2 | 2 | Liverpool | England Premier League | 86 | 84 | 85 | 85 |
| 3 | 3 | Manchester City | England Premier League | 84 | 87 | 86 | 85 |
| 4 | 4 | Bayern München | Germany . Bundesliga | 92 | 85 | 82 | 84 |
| 5 | 5 | Real Madrid | Spain Primera División | 85 | 86 | 83 | 84 |
| 6 | 6 | Chelsea | England Premier League | 84 | 85 | 83 | 84 |
| 7 | 7 | Atlético de Madrid | Spain Primera División | 84 | 82 | 82 | 83 |
| 8 | 8 | FC Barcelona | Spain Primera División | 83 | 85 | 81 | 83 |
| 9 | 9 | Manchester United | England Premier League | 82 | 83 | 81 | 83 |
| 10 | 10 | Inter | Italy Serie A | 82 | 82 | 83 | 82 |
| 11 | 11 | Juventus | Italy Serie A | 81 | 81 | 82 | 82 |
| 12 | 12 | Borussia Dortmund | Germany . Bundesliga | 83 | 81 | 81 | 81 |
| 13 | 13 | Tottenham Hotspur | England Premier League | 83 | 81 | 78 | 81 |
| 14 | 14 | Sevilla FC | Spain Primera División | 81 | 81 | 83 | 81 |
| 15 | 15 | Milan | Italy Serie A | 81 | 80 | 80 | 81 |
| 16 | 16 | RB Leipzig | Germany . Bundesliga | 82 | 81 | 79 | 80 |
| 17 | 17 | Napoli | Italy Serie A | 82 | 80 | 77 | 80 |
| 18 | 18 | Bergamo Calcio | Italy Serie A | 82 | 79 | 79 | 80 |
| 19 | 19 | Real Sociedad | Spain Primera División | 81 | 81 | 78 | 80 |
| 20 | 20 | Villarreal CF | Spain Primera División | 81 | 79 | 79 | 80 |
| 21 | 21 | Leicester City | England Premier League | 80 | 80 | 78 | 80 |
| 22 | 22 | Arsenal | England Premier League | 77 | 80 | 77 | 80 |
| 23 | 23 | Bayer 04 Leverkusen | Germany . Bundesliga | 83 | 79 | 78 | 79 |
| 24 | 24 | Ajax | Holland Eredivisie | 82 | 77 | 80 | 79 |
| 25 | 25 | Real Betis Balompié | Spain Primera División | 80 | 79 | 78 | 79 |
| 26 | 26 | West Ham United | England Premier League | 80 | 79 | 77 | 79 |
| 27 | 27 | Latium | Italy Serie A | 79 | 81 | 78 | 79 |
| 28 | 28 | Sporting CP | Portugal Primeira Liga | 79 | 80 | 78 | 79 |
| 29 | 29 | Roma | Italy Serie A | 79 | 79 | 78 | 79 |
| 30 | 30 | Athletic Club de Bilbao | Spain Primera División | 78 | 77 | 79 | 79 |

4. **Data Discretization:** In our dataset, fifa22, we categorized the league row into one category. Earlier it was about all data individually written which looked unclear and messy, so we categorized it into whole rows.

**Code:**

```
class(fifa22$Leauge)
fifa22$Leauge[fifa22$Leauge=="France Ligue  "] <- "1"
fifa22$Leauge[fifa22$Leauge=="England Premier League "] <- "2"
fifa22$Leauge[fifa22$Leauge=="Germany . Bundesliga "] <- "3"
fifa22$Leauge[fifa22$Leauge=="Spain Primera División "] <- "4"
fifa22$Leauge[fifa22$Leauge=="Italy Serie A "] <- "5"
fifa22$Leauge[fifa22$Leauge=="Holland Eredivisie "] <- "6"
fifa22$Leauge[fifa22$Leauge=="Portugal Primeira Liga "] <- "7"


class(fifa21$Leauge)
fifa21$Leauge[fifa21$Leauge=="Ligue 1 Uber Eats"] <- "1"
fifa21$Leauge[fifa21$Leauge=="Premier League"] <- "2"
fifa21$Leauge[fifa21$Leauge=="Bundesliga"] <- "3"
fifa21$Leauge[fifa21$Leauge=="LaLiga Santander"] <- "4"
fifa21$Leauge[fifa21$Leauge=="Serie A TIM"] <- "5"
fifa21$Leauge[fifa21$Leauge=="Eredivisie"] <- "6"
fifa21$Leauge[fifa21$Leauge=="Liga NOS"] <- "7"
```

Filter

| X | | Club | Leauge | Attack | Midfield | Defence | Overall |
|---|---|---|---|---|---|---|---|
| 1 | 1 | Paris Saint-Germain | 1 | 89 | 82 | 84 | 85 |
| 2 | 2 | Liverpool | 2 | 86 | 84 | 85 | 85 |
| 3 | 3 | Manchester City | 2 | 84 | 87 | 86 | 85 |
| 4 | 4 | Bayern München | 3 | 92 | 85 | 82 | 84 |
| 5 | 5 | Real Madrid | 4 | 85 | 86 | 83 | 84 |
| 6 | 6 | Chelsea | 2 | 84 | 85 | 83 | 84 |
| 7 | 7 | Atlético de Madrid | 4 | 84 | 82 | 82 | 83 |
| 8 | 8 | FC Barcelona | 4 | 83 | 85 | 81 | 83 |
| 9 | 9 | Manchester United | 2 | 82 | 83 | 81 | 83 |
| 10 | 10 | Inter | 5 | 82 | 82 | 83 | 82 |
| 11 | 11 | Juventus | 5 | 81 | 81 | 82 | 82 |
| 12 | 12 | Borussia Dortmund | 3 | 83 | 81 | 81 | 81 |
| 13 | 13 | Tottenham Hotspur | 2 | 83 | 81 | 78 | 81 |
| 14 | 14 | Sevilla FC | 4 | 81 | 81 | 83 | 81 |
| 15 | 15 | Milan | 5 | 81 | 80 | 80 | 81 |
| 16 | 16 | RB Leipzig | 3 | 82 | 81 | 79 | 80 |
| 17 | 17 | Napoli | 5 | 82 | 80 | 77 | 80 |
| 18 | 18 | Bergamo Calcio | 5 | 82 | 79 | 79 | 80 |
| 19 | 19 | Real Sociedad | 4 | 81 | 81 | 78 | 80 |
| 20 | 20 | Villarreal CF | 4 | 81 | 79 | 79 | 80 |
| 21 | 21 | Leicester City | 2 | 80 | 80 | 78 | 80 |
| 22 | 22 | Arsenal | 2 | 77 | 80 | 77 | 80 |
| 23 | 23 | Bayer 04 Leverkusen | 3 | 83 | 79 | 78 | 79 |
| 24 | 24 | Ajax | 6 | 82 | 77 | 80 | 79 |
| 25 | 25 | Real Betis Balompié | 4 | 80 | 79 | 78 | 79 |
| 26 | 26 | West Ham United | 2 | 80 | 79 | 77 | 79 |
| 27 | 27 | Latium | 5 | 79 | 81 | 78 | 79 |
| 28 | 28 | Sporting CP | 7 | 79 | 80 | 78 | 79 |
| 29 | 29 | Roma | 5 | 79 | 79 | 78 | 79 |
| 30 | 30 | Athletic Club de Bilbao | 4 | 78 | 77 | 79 | 79 |

And same goes for dataset2 (fifa21)

| | X | Club | Leauge | Attack | Midfield | Defence | Overall |
|---|---|---|---|---|---|---|---|
| 1 | 1 | Liverpool | 2 | 86 | 83 | 80 | 85 |
| 2 | 2 | Manchester City | 2 | 85 | 86 | 83 | 85 |
| 3 | 3 | Real Madrid | 4 | 84 | 86 | 84 | 85 |
| 4 | 4 | FC Bayern München | 3 | 92 | 85 | 82 | 84 |
| 5 | 5 | Paris Saint-Germain | 1 | 87 | 82 | 82 | 84 |
| 6 | 6 | Juventus | 5 | 87 | 81 | 83 | 83 |
| 7 | 7 | Atlético de Madrid | 4 | 85 | 82 | 83 | 83 |
| 8 | 8 | FC Barcelona | 4 | 84 | 82 | 81 | 83 |
| 9 | 9 | Tottenham Hotspur | 2 | 89 | 82 | 81 | 82 |
| 10 | 10 | Manchester United | 2 | 83 | 82 | 82 | 82 |
| 11 | 11 | Inter | 5 | 83 | 81 | 83 | 82 |
| 12 | 12 | Chelsea | 2 | 82 | 84 | 82 | 82 |
| 13 | 13 | Borussia Dortmund | 3 | 87 | 81 | 81 | 81 |
| 14 | 14 | Sevilla FC | 4 | 80 | 81 | 82 | 81 |
| 15 | 15 | Arsenal | 2 | 83 | 79 | 79 | 80 |
| 16 | 16 | Napoli | 5 | 82 | 80 | 80 | 80 |
| 17 | 17 | Lazio | 5 | 82 | 80 | 79 | 80 |
| 18 | 18 | Villarreal CF | 4 | 82 | 79 | 79 | 80 |
| 19 | 19 | Leicester City | 2 | 80 | 81 | 79 | 80 |
| 20 | 20 | RB Leipzig | 3 | 78 | 81 | 80 | 80 |
| 21 | 21 | Milan | 5 | 84 | 78 | 80 | 79 |
| 22 | 22 | Atalanta | 5 | 82 | 79 | 78 | 79 |
| 23 | 23 | Real Sociedad | 4 | 81 | 81 | 77 | 79 |
| 24 | 24 | Everton | 2 | 81 | 79 | 80 | 79 |
| 25 | 25 | Athletic Club | 4 | 80 | 78 | 79 | 79 |
| 26 | 26 | Borussia Mönchengladbach | 3 | 78 | 79 | 80 | 79 |
| 27 | 27 | Bayer 04 Leverkusen | 3 | 78 | 78 | 78 | 79 |
| 28 | 28 | Wolverhampton Wanderers | 2 | 77 | 81 | 78 | 79 |
| 29 | 29 | SL Benfica | 7 | 77 | 79 | 78 | 79 |
| 30 | 30 | Ajax | 6 | 80 | 77 | 76 | 78 |

## Descriptive Statistics:

Descriptive statistics applies the concepts, measures, and terms that are used to describe the basic features of the samples in a study. Here these procedures are followed for 2 of our datasets (fifa22, fifa21) as an approximation result.

First the quantitative analysis then the simple graphics. For the dataset (fifa22) all statistics are done separately with specific code on the other hand for fifa21 dataset the summaries () function used for direct analysis.

Dataset 1 (fifa22)

```
> mean(fifa22$Overall)
[1] 81.2
> median(fifa22$Overall)
[1] 80.5
>    uv <- unique(fifa22$Overall)
> mode(fifa22$Overall)
[1] "numeric"
> max(fifa22$Overall)-min(fifa22$Overall)
[1] 6
> var(fifa22$Overall)
[1] 4.372414
> sd(fifa22$Overall)
[1] 2.091032
> quantile(fifa22$Overall)
    0%    25%    50%    75%   100%
79.00 79.25 80.50 83.00 85.00
> IQR(fifa22$Overall)
[1] 3.75
>
```

Dataset 2 (fifa21)

```
> summary(fifa21)
       X              Club              Leauge              Attack         Midfield        Defence
 Min.   : 1.00   Length:30          Length:30          Min.   :77.00   Min.   :77.0   Min.   :76.0
 1st Qu.: 8.25   Class :character   Class :character   1st Qu.:80.00   1st Qu.:79.0   1st Qu.:79.0
 Median :15.50   Mode  :character   Mode  :character   Median :82.00   Median :81.0   Median :80.0
 Mean   :15.50                                         Mean   :82.63   Mean   :80.9   Mean   :80.3
 3rd Qu.:22.75                                         3rd Qu.:84.75   3rd Qu.:82.0   3rd Qu.:82.0
 Max.   :30.00                                         Max.   :92.00   Max.   :86.0   Max.   :84.0
    Overall
 Min.   :78.00
 1st Qu.:79.00
 Median :80.00
 Mean   :81.03
 3rd Qu.:82.75
 Max.   :85.00
> var(fifa21$Overall)
[1] 4.585057
> sd (fifa21$Overall)
[1] 2.141275
> quantile(fifa21$Overall)
    0%    25%    50%    75%   100%
78.00 79.00 80.00 82.75 85.00
> IQR (fifa21$Overall)
[1] 3.75
>
```

Mean: The mean value for fifa22 overall is 81.2 where is n the dataset 2 the fifa21 mean values of attack, Mid and defense is respectively 82.63, 80.9 and 80.3.

Median: using R median () function to determine the overall median value of fifa22 is 80.5 where at fifa21 82, 81 and 80 are the median value for the attributes.

Mode: R does not have a built-in function to calculate the mode. So, using R with a user-defined function to find the modes of the values,

For fifa21 the mode is 79 and for fifa22 the value is also 79

```
> my_mode(x1)      > my_mode(x2)
[1] 79             [1] 79
```

**Fifa21 dataset Code:**

```
x1 <- c(fifa21$Overall)
my_mode <- function(x) {
  unique_x <- unique(x)
  tabulate_x <- tabulate(match(x, unique_x))
  unique_x[tabulate_x == max(tabulate_x)]
}
my_mode(x1)
```

**Fifa22 dataset Code:**

```
x2 <- c(fifa22$Overall)
my_mode <- function(x) {
  unique_x <- unique(x)
  tabulate_x <- tabulate(match(x, unique_x))
  unique_x[tabulate_x == max(tabulate_x)]
}
my_mode(x2)
```

Range: Use the R min() and max() functions to find the range of the values, here for fifa22 the range is 6 and for fifa21 dataset it is 85-78= 7.

Variance: now for fifa22 the variance is 4.372414 and for fifa21 the variance is 4.585057.

Standard Deviation: The standard deviation is simply the square root of the variance. Standard deviation measures how far a 'typical' observation is from the average of the data. Use the R sd() function to find the sample standard deviation of fifa22 is 2.091032 and for fifa21 it is 2.141575.

Quartiles: Quartiles are values that separate the data into four equal parts.

```
> quantile(fifa21$Overall)        > quantile(fifa22$Overall)
   0%    25%    50%    75%   100%      0%    25%    50%    75%   100%
78.00 79.00 80.00 82.75 85.00  79.00 79.25 80.50 83.00 85.00
```

Interquartile Range: Interquartile range is the difference between the first and third quartiles (Q1 and Q3).
Fifa22 IQR() = 3.75 and fifa21 IQR() = 3.75

Data Distributor: A histogram is a widely used graph to show the distribution of quantitative (numerical) data. It shows the frequency of rating values in the data, usually in intervals of values. Frequency is the number of times that rating value appeared in the data. We try to display all the attributes data using histogram and summaries the frequency rate.
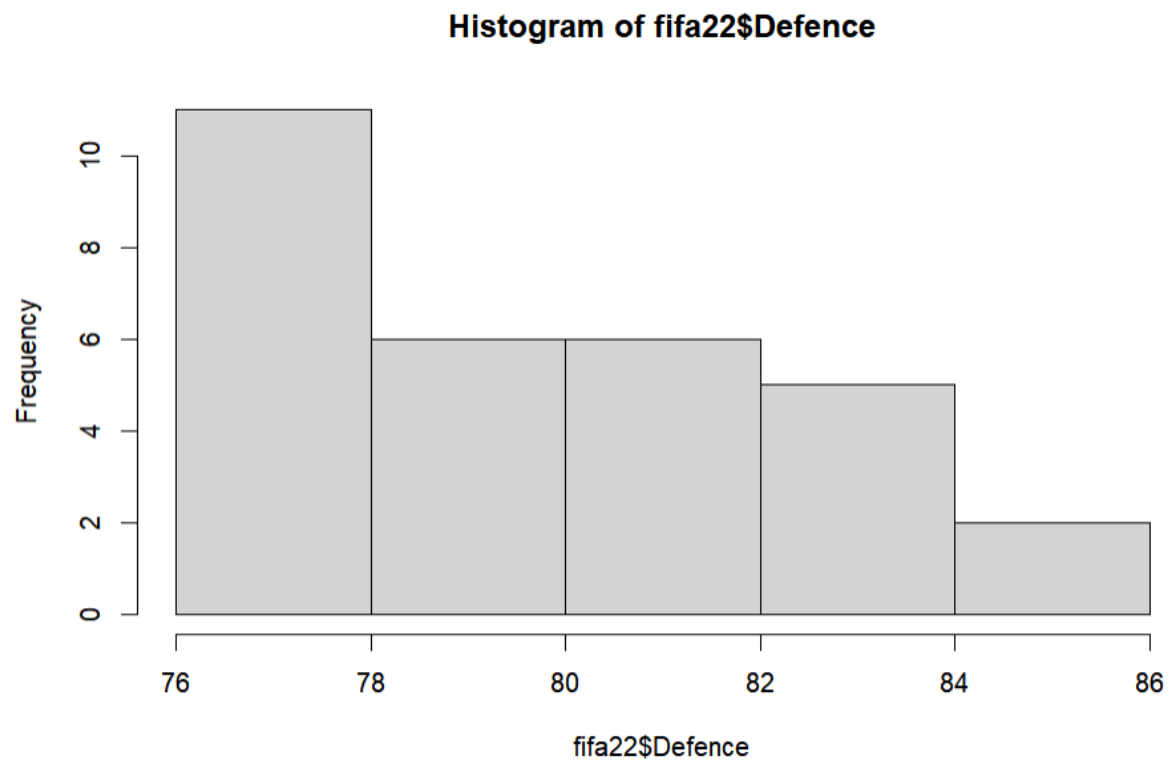

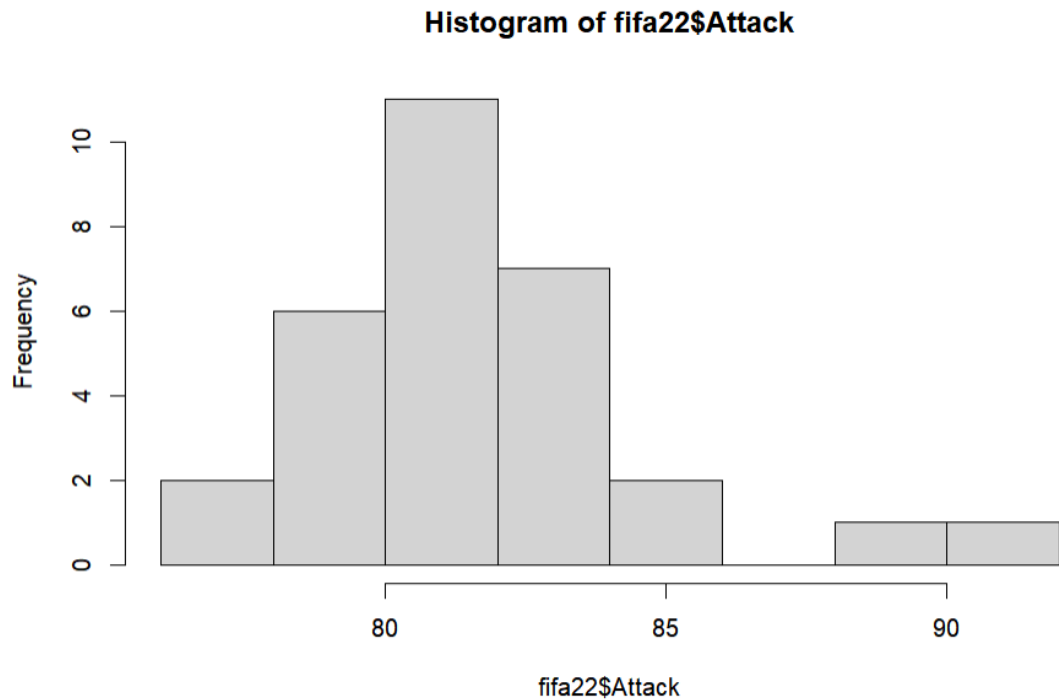**For fifa22 dataset**

**Code:**
**#Histogram**
**hist(fifa22$Overall)**
**hist(fifa22$Attack)**
**hist(fifa22$Defence)**
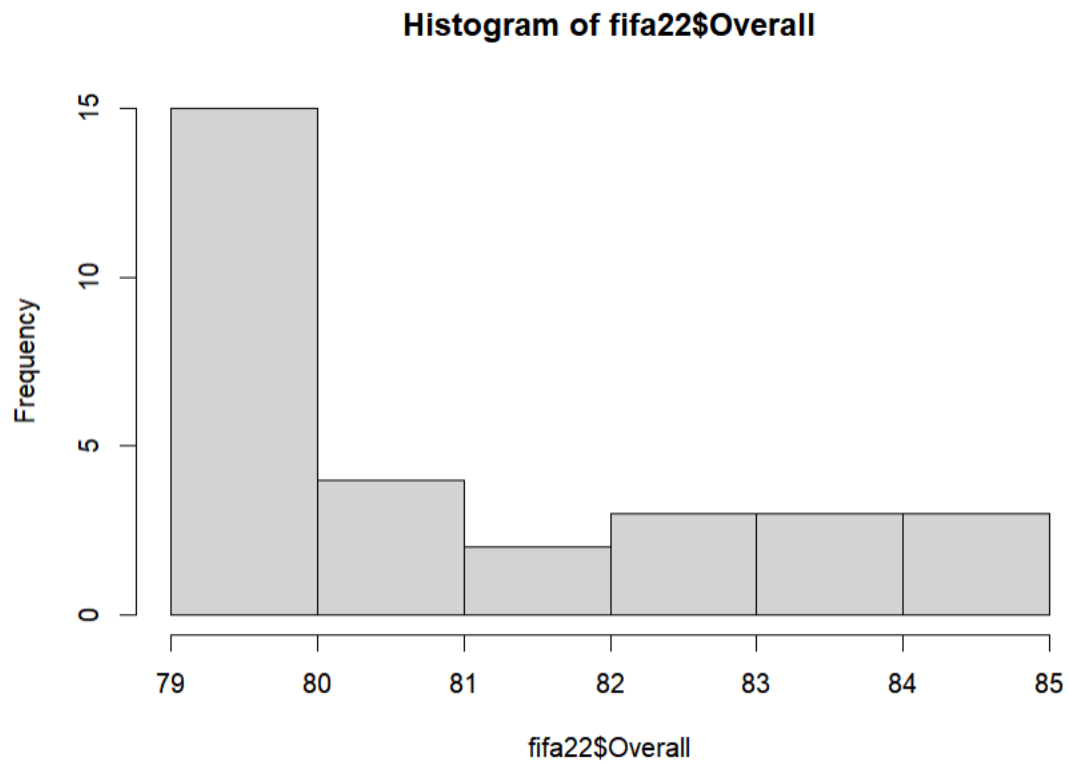**hist(fifa22$Midfield)**

## Histogram of fifa22$Defence



fifa22$defence the most teams have defense rating around 76-78 which is very low in case of defense. Defenders are not very highly rated as the attacker.

## Histogram of fifa22$Attack



fifa22$Attack

fifa22$Attack graph we can notice that most of the team's attack rating is around 80-82. Very few club possess high class striker.

## Histogram of fifa22$Midfield
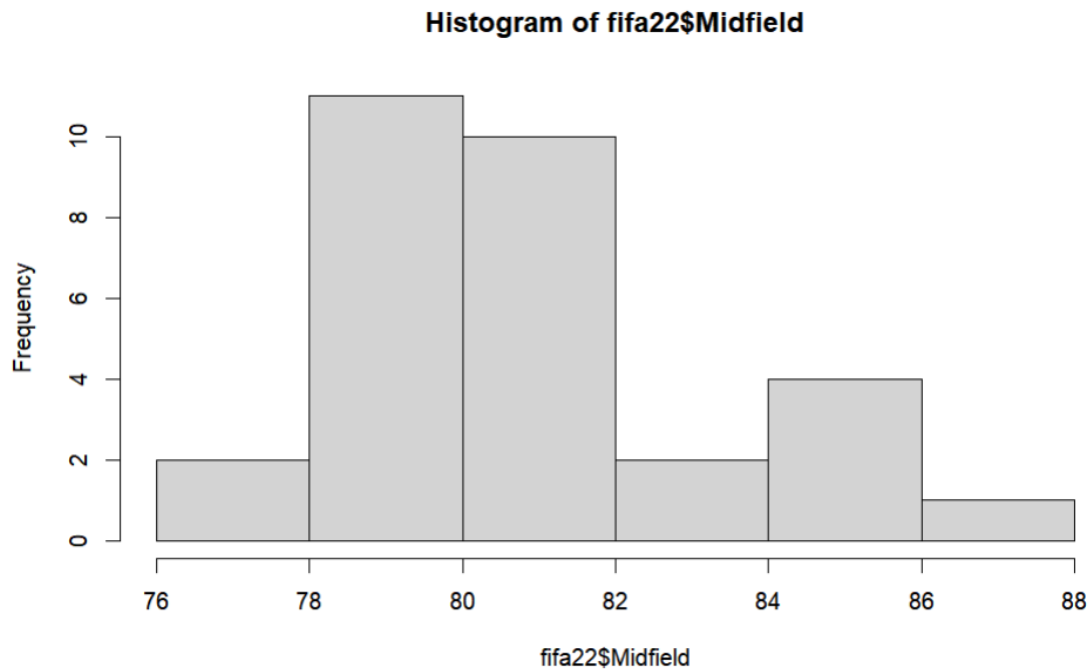


fifa22$Midfield

Where The midfielder apart from one team most of them are having 78 – 82 rating status.

**Histogram of fifa22$Overall**

. So the overall view is that most teams are around 79 to 80. Only top teams have the best overall ratings no doubts.
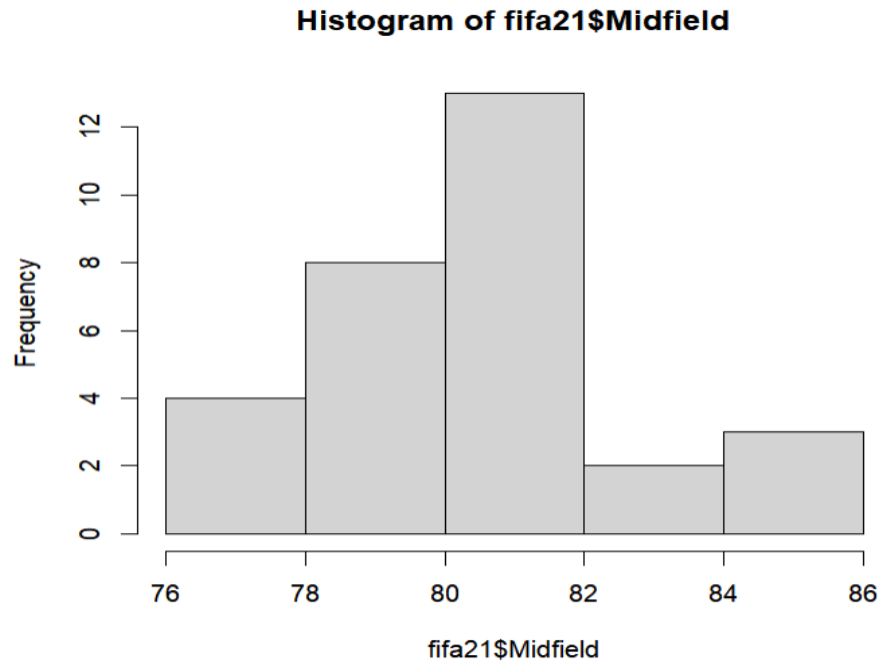
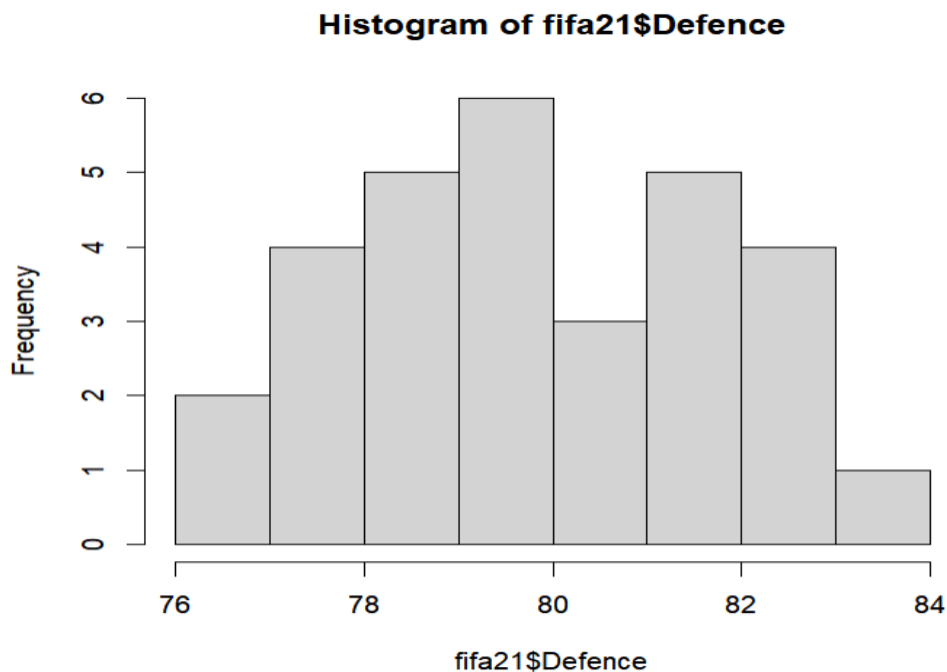**For fifa21 dataset**
**Code:**
#Histogram
hist(fifa21$Overall)
hist(fifa21$Attack)
hist(fifa21$Defence)
hist(fifa21$Midfield)

**Histogram of fifa21$Midfield**



The midfielder apart from one team most of them are having 80 – 82 rating status. In 21 clubs are mainly focused on strengthening midfield for their tactics purpose.

**Histogram of fifa21$Defence**



Where fifa21$defence the most teams have defense rating around 80 and the defense is very balanced according orderly. Though the defenders are not very high rated as the attacker.

## Histogram of fifa21$Attack
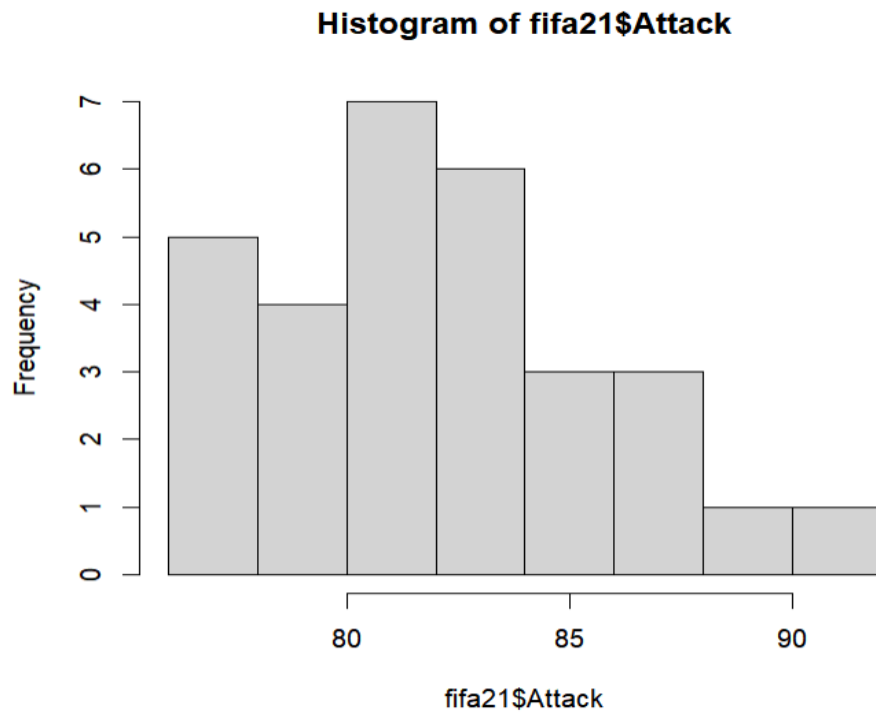


fifa21$Attack

fifa21$Attack graph we can notice that most of the team's attack rating is around 82-85. Which is higher than the previous dataset.
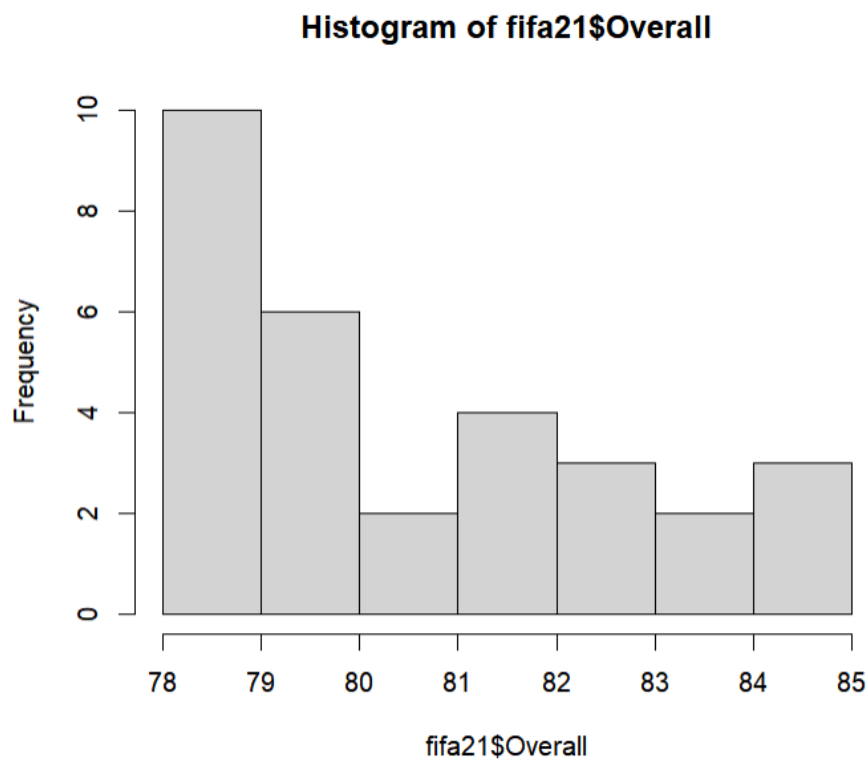
## Histogram of fifa21$Overall



fifa21$Overall

So, the overall view is that most teams are around 79 to 80. Only top teams have the best overall ratings no doubt.

# Data Visualization

Creating a graph plot using ggplot2.

**Dataset 1 (FIFA22)**

**Code:**

```
#ggplot visualize
data22=data.frame(fifa22$Overall)
str(data22)
#install.packages("ggplot2")
library(ggplot2)
ggplot(data22, aes(fifa22$Club, fifa22$Overall))+
  geom_point(size = 3)
```

Here in this scatter diagram, we can observe that most of the clubs are at below 82 rating. Only three of them are above 84. The players are expensive that is why most of the club could not afford the best rated players for their team. And top-level clubs have so much rating compared to others club which are below 82. The ideal rating for most of the teams is around 82-79.

**Dataset 2 (FIFA21)**
**Code:**

```
data21=data.frame(fifa21$Overall)
str(data21)
#install.packages("ggplot2")
```

```
library(ggplot2)
ggplot(data21, aes(fifa21$Club, fifa21$Overall))+
  geom_point(size = 3)
```
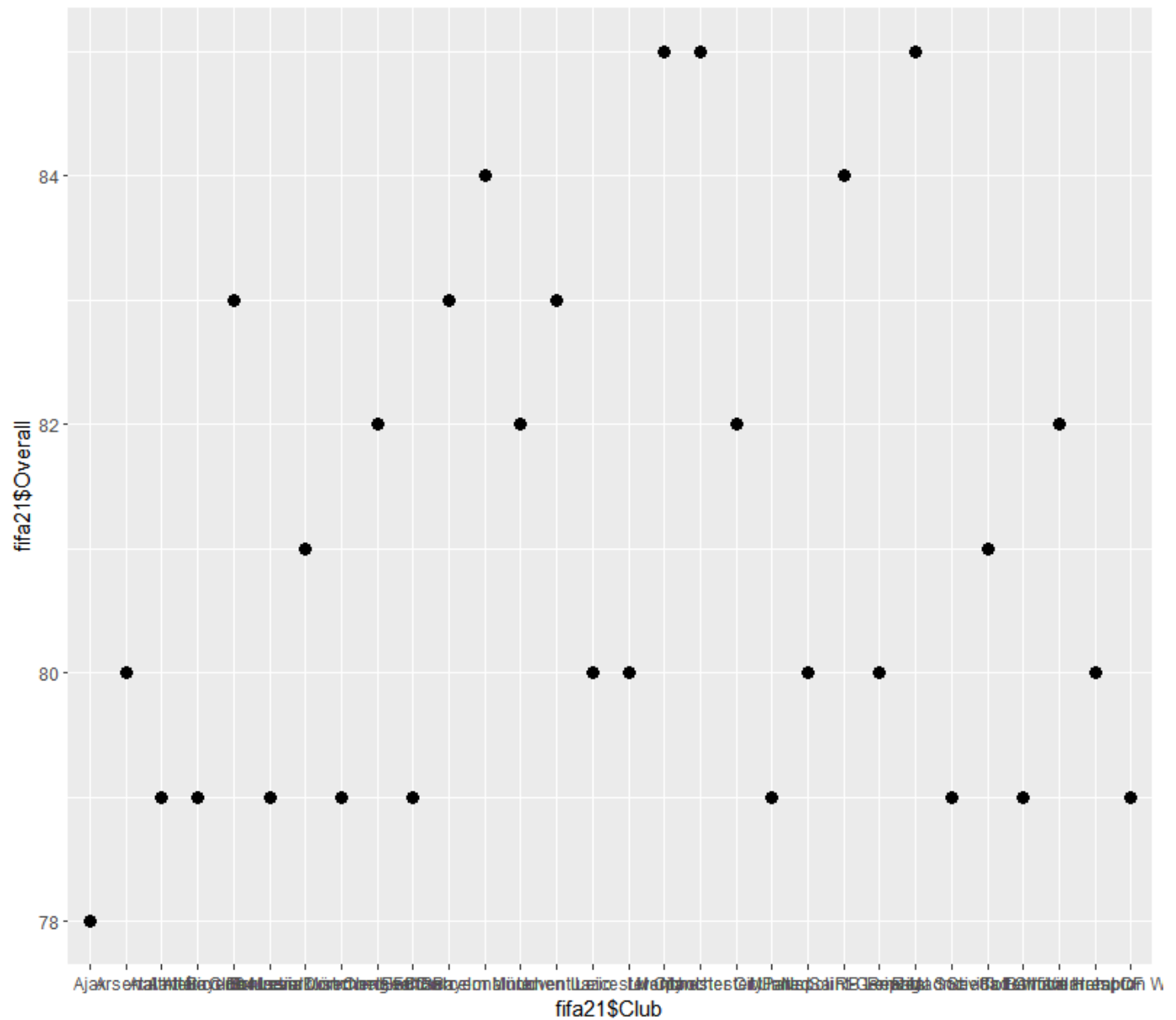


Here in this scatter diagram, all teams are almost balanced. Though most of the clubs' ratings got reduced from the previous diagram which is around 81-79. And also, clubs between 82-84 have increased in great number. It means the clubs

have improved themselves in the competition. Also, a huge number of talented players may have joined the clubs. The 3 top teams with 85 rating like previous.

## Discussion & Conclusion

After all the necessary process finally, our data set complete and ready to be used for further analysis. From the descriptive statistics and the graph plot we can say that in 2021 the club were balanced in every position aspect because all the positions ratings are well balanced but in the 2022 clubs are focusing more on defense rather than attacking. There is also a reason that a handful of good players are retiring so the defense player gets more ratings. This difficult thing is easily done by using Rstudio with R language thanks to the unlimited library makes our work easy. I also use Selectorgadget for scrapping data table from the website.