

Week13_FinalSubmission

lok

2023-11-14

Week 9

(1) What is the topic that you have finalized?

The topic chosen is football players' rating, the website should allow user to search and select player they want, display the the player's overall rating as well as rating of categories including pace, strength, shooting etc. (if possible, allow user to compare between different players)

(2) What are the data sources that you have curated so far?

data is available from EA FC24 (the football game) online, I'm still trying to find a way to export or download all the data. (<https://www.ea.com/games/ea-sports-fc/ratings> (<https://www.ea.com/games/ea-sports-fc/ratings>)) update: found on (<https://www.kaggle.com/datasets/nyagami/fc-24-players-database-and-stats-from-easports/data> (<https://www.kaggle.com/datasets/nyagami/fc-24-players-database-and-stats-from-easports/data>))

Week 10

(1) What is the question that you are going to answer?

The question I'm going to answer is "what are the insights that can be drawn from the football players' rating in 2024?" This can be broken down into various questions: 1. Which club has the highest average rating of players? 2. Which country has the highest count of professional players? 3. What percentage of players have a overall rating higher than 80, what does that imply?

(2) Why is this an important question?

football is one of the mainstream entertainment / sports, and by presenting the data about players in an comprehensive way, one can search and learn more about the player or teams of their interest, as well as the detail value after different categories such as pace, shooting, strength etc.

(3) Which rows and columns of the dataset will be used to answer this question? (Answer: Actual names of the variables in the dataset that you

plan to use).

the column of name, nation, club and position can be used to filter or search players, where the overall rating, and the categorical rating (pace, shooting, passing, dribbling, defending, physicality, acceleration sprint, positioning, finishing, etc) can be display to the user.

```
# data
library(tidyverse)
```

```
## — Attaching core tidyverse packages — tidyverse 2.0.0 —
## ✓ dplyr      1.1.3      ✓ readr      2.1.4
## ✓ forcats    1.0.0      ✓ stringr    1.5.0
## ✓ ggplot2    3.4.3      ✓ tibble     3.2.1
## ✓ lubridate  1.9.2      ✓ tidyr      1.3.0
## ✓ purrr      1.0.2
## — Conflicts — tidyverse_conflicts() —
## ✖ dplyr::filter() masks stats::filter()
## ✖ dplyr::lag()     masks stats::lag()
## i Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to become errors
```

```
eaafc_data <- read_csv("male_players.csv")
```

```
## New names:
## Rows: 15845 Columns: 47
## — Column specification
## ————— Delimiter: "," chr
## (9): Name, Nation, Club, Position, Att work rate, Def work rate, Prefer... dbl
## (38): ...1, Age, Overall, Pace, Shooting, Passing, Dribbling, Defending,...
## i Use `spec()` to retrieve the full column specification for this data. i
## Specify the column types or set `show_col_types = FALSE` to quiet this message.
## • `` -> `...1`
```

```
eaafc_data
```

```
## # A tibble: 15,845 × 47
##   ...1 Name      Nation Club Position Age Overall Pace Shooting Passing
##   <dbl> <chr>      <chr> <chr> <chr>   <dbl>   <dbl> <dbl>   <dbl>   <dbl>
## 1      0 Kylian Mbap... France Pari... ST      24      91    97      90      80
## 2      1 Erling Haal... Norway Manc... ST      23      91    89      93      66
## 3      2 Kevin De Br... Belgi... Manc... CM      32      91    72      88      94
## 4      3 Lionel Messi Argen... Inte... CF      36      90    80      87      90
## 5      4 Karim Benze... France Al I... CF      35      90    79      88      83
## 6      5 Thibaut Cou... Belgi... Real... GK      31      90    85      89      76
## 7      6 Harry Kane Engla... FC B... ST      30      90    69      93      84
## 8      7 Robert Lewa... Poland FC B... ST      35      90    75      91      80
## 9      8 Mohamed Sal... Egypt Live... RW      31      89    89      87      81
## 10     9 Rúben Dias Portu... Manc... CB      26      89    62      39      66
## # i 15,835 more rows
## # i 37 more variables: Dribbling <dbl>, Defending <dbl>, Physicality <dbl>,
## # Acceleration <dbl>, Sprint <dbl>, Positioning <dbl>, Finishing <dbl>,
## # Shot <dbl>, Long <dbl>, Volleys <dbl>, Penalties <dbl>, Vision <dbl>,
## # Crossing <dbl>, Free <dbl>, Curve <dbl>, Agility <dbl>, Balance <dbl>,
## # Reactions <dbl>, Ball <dbl>, Composure <dbl>, Interceptions <dbl>,
## # Heading <dbl>, Def <dbl>, Standing <dbl>, Sliding <dbl>, Jumping <dbl>, ...
```

```
# Attempting to create a search engine using function
searchData <- function(data, query, column_name) {
  result <- data[grep(query, data[["Name"]], ignore.case = TRUE), ]
  return(result)
}

user_query <- readline("Player name: ")
```

```
## Player name:
```

```
search_result <- searchData(eafc_data, user_query, "Name")
print(search_result[, ])
```

```
## # A tibble: 15,845 × 47
##   ...1 Name      Nation Club Position Age Overall Pace Shooting Passing
##   <dbl> <chr>      <chr> <chr> <chr>   <dbl>   <dbl> <dbl>   <dbl>   <dbl>
## 1      0 Kylian Mbap... France Pari... ST      24      91     97      90      80
## 2      1 Erling Haal... Norway Manc... ST      23      91     89      93      66
## 3      2 Kevin De Br... Belgi... Manc... CM      32      91     72      88      94
## 4      3 Lionel Messi Argen... Inte... CF      36      90     80      87      90
## 5      4 Karim Benze... France Al I... CF      35      90     79      88      83
## 6      5 Thibaut Cou... Belgi... Real... GK      31      90     85      89      76
## 7      6 Harry Kane Engla... FC B... ST      30      90     69      93      84
## 8      7 Robert Lewa... Poland FC B... ST      35      90     75      91      80
## 9      8 Mohamed Sal... Egypt Live... RW      31      89     89      87      81
## 10     9 Rúben Dias Portu... Manc... CB      26      89     62      39      66
## # i 15,835 more rows
## # i 37 more variables: Dribbling <dbl>, Defending <dbl>, Physicality <dbl>,
## # Acceleration <dbl>, Sprint <dbl>, Positioning <dbl>, Finishing <dbl>,
## # Shot <dbl>, Long <dbl>, Volleys <dbl>, Penalties <dbl>, Vision <dbl>,
## # Crossing <dbl>, Free <dbl>, Curve <dbl>, Agility <dbl>, Balance <dbl>,
## # Reactions <dbl>, Ball <dbl>, Composure <dbl>, Interceptions <dbl>,
## # Heading <dbl>, Def <dbl>, Standing <dbl>, Sliding <dbl>, Jumping <dbl>, ...
```

```
#integrating into shiny
```

```
library(shiny)
```

```
library(dplyr)
```

```
eafc_data <- read.csv("male_players.csv") %>% select(Name, Nation, Club, Position, Ov
erall, Pace, Shooting, Passing, Dribbling, Defending, Physicality)
```

```
ui <- fluidPage(
  sidebarLayout(
    sidebarPanel(
      textInput("search_query", "Search:")
    ),
    mainPanel(
      DT::dataTableOutput("search_result_table")
    )
  )
)
```

```
server <- function(input, output) {
  searchData <- reactive({
    result <- eafc_data[grepl(input$search_query, eafc_data$"Name", ignore.case = TRU
E), ]
    return(result)
  })

  output$search_result_table <- DT::renderDataTable({
    searchData()
  })
}
```

```
# Run the application
```

```
shinyApp(ui = ui, server = server)
```

##

Listening on http://127.0.0.1:3560

Search:

Show 10 entries

Search:

	Name	Nation	Club	Position	Overall	Pace	Shooting	Pas
1	Kylian Mbappé	France	Paris SG	ST	91	97	90	
2	Erling Haaland	Norway	Manchester City	ST	91	89	93	
3	Kevin De Bruyne	Belgium	Manchester City	CM	91	72	88	
4	Lionel Messi	Argentina	Inter Miami CF	CF	90	80	87	
5	Karim Benzema	France	Al Ittihad	CF	90	79	88	
6	Thibaut Courtois	Belgium	Real Madrid	GK	90	85	89	
7	Harry Kane	England	FC Bayern München	ST	90	69	93	
8	Robert Lewandowski	Poland	FC Barcelona	ST	90	75	91	
9	Mohamed Salah	Egypt	Liverpool	RW	89	89	87	
10	Rúben Dias	Portugal	Manchester City	CB	89	62	39	

Showing 1 to 10 of 15,845 entries

Previous **1** 2 3 4 5 ... 1,585 Next

Week 11

(1) List the visualizations that you are going to use in your project (Answer: What are the variables that you are going to plot? How will it answer your larger question?)

visualizations will consist of a few graph to answer the sub-questions, 1. Which club has the highest average rating of players? the data will be grouped by club, take the average of the overall rating of each club, then plot x = club and y = average overall rating 2. Which country has the highest count of professional players? data will be arranged by country, count number of data for each country, then plot x = country and y = number of players (number of data entry) 3. What percentage of players have a overall rating higher than 80, what does that imply? this can be calculated by the number of entries with rating higher than 80 divided by the total number of data, this will be illustrated with a Pie chart.

furthermore, an interactive shiny app will be embedded for user to search any player they interested in, ideally, will plot the chosen player's rating in radar plot.

(2) How do you plan to make it interactive?

(Answer: features of ggplot2/shiny/markdown do you plan to use to make the story interactive)

they interactive part will be illustrated by the shiny app, by allowing the user to search any player they interested in, as well as plotting the chosen player's specific rating with radar plot

(3) What concepts incorporated in your project were taught in the course and which ones were self-learnt? (Answer: Create a table with topics in one column and Weeks in the other to indicate which concept taught in which week is being used. Leave the entry of the Week column empty for self-learnt concepts)

```
week <- c("4", "7", "8", "NA", "NA")
concept <- c("manipulating data (select column)", "ggplot2", "shiny app", "radar plot", "building search engine")

data.frame(concept, week)
```

```
##               concept week
## 1 manipulating data (select column)    4
## 2                      ggplot2      7
## 3                      shiny app     8
## 4                      radar plot   NA
## 5          building search engine   NA
```

Week 12

```
library(shiny)
library(dplyr)

eafc_data <- read.csv("male_players.csv") %>% select(Name, Nation, Club, Position, Overall, Pace, Shooting, Passing, Dribbling, Defending, Physicality)

ui <- fluidPage(
  sidebarLayout(
    sidebarPanel(
      textInput("search_query", "Search:"),
      downloadButton("download_chart", "Download chart") # this is new addition in week 13
    ),
    mainPanel(
      DT::dataTableOutput("search_result_table")
    )
  )
)

server <- function(input, output) {
  searchData <- reactive({
    result <- eafc_data[grep(input$search_query, eafc_data$"Name", ignore.case = TRUE), ]
    return(result)
  })

  output$search_result_table <- DT::renderDataTable({
    searchData()
  })
}

# Run the application
shinyApp(ui = ui, server = server)
```

```
##
## Listening on http://127.0.0.1:5360
```

Search:

Download chart

Show 10 entries

Search:

	Name	Nation	Club	Position	Overall	Pace	Shooting	Pas
1	Kylian Mbappé	France	Paris SG	ST	91	97	90	
2	Erling Haaland	Norway	Manchester City	ST	91	89	93	
3	Kevin De Bruyne	Belgium	Manchester City	CM	91	72	88	
4	Lionel Messi	Argentina	Inter Miami CF	CF	90	80	87	
5	Karim Benzema	France	Al Ittihad	CF	90	79	88	
6	Thibaut Courtois	Belgium	Real Madrid	GK	90	85	89	
7	Harry Kane	England	FC Bayern München	ST	90	69	93	
8	Robert Lewandowski	Poland	FC Barcelona	ST	90	75	91	
9	Mohamed Salah	Egypt	Liverpool	RW	89	89	87	
10	Rúben Dias	Portugal	Manchester City	CB	89	62	39	

Showing 1 to 10 of 15,845 entries

Previous

1

2

3

4

5

...

1,585

Next


```
# attempting to create radar chart
```

```
library(fmsb)
```

```
scores <- data.frame(  
  row.names = c(eafc_data$X),  
  Pace = c(eafc_data$Pace),  
  Shooting = c(eafc_data$Shooting),  
  Passing = c(eafc_data$Passing),  
  Dribbling = c(eafc_data$Dribbling),  
  Defending = c(eafc_data$Defending),  
  Physicality = c(eafc_data$Physicality)  
)
```

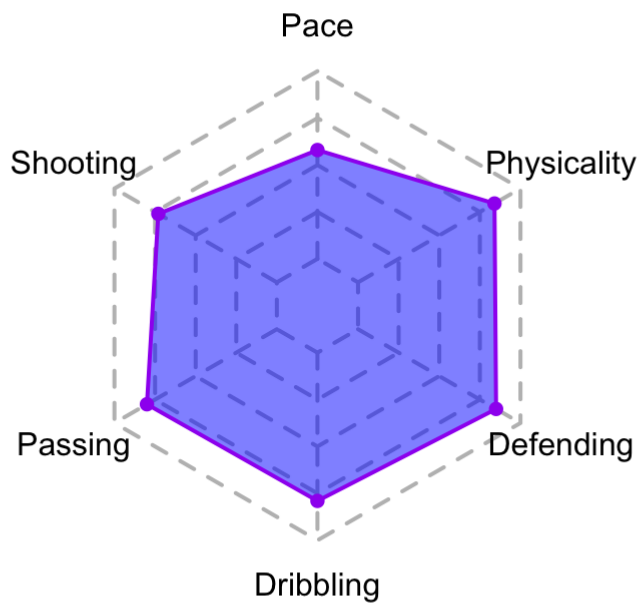
```
max_min <- data.frame(  
  Pace = c(100, 0), Shooting = c(100, 0), Passing = c(100, 0),  
  Dribbling = c(100, 0), Defending = c(100, 0), Physicality = c(100, 0)  
)
```

```
rownames(max_min) <- c("Max", "Min")
```

```
df <- rbind(max_min, scores)
```

```
player_radarchart <- df[c("Max", "Min", "12"), ]  
radarchart(  
  player_radarchart,  
  pcol = "purple", pfc col = scales::alpha("blue", 0.5), plwd = 2, plty = 1,  
  cglcol = "grey", cglty = 2, cglwd = 2,  
  title = "Player's Rating"  
)
```

Player's Rating



```

# integrating into shiny
library(shiny)
library(fmsb)

eafc_data <- read.csv("male_players.csv") %>% select(Name, Nation, Club, Position, Overall, Pace, Shooting, Passing, Dribbling, Defending, Physicality)

# Define UI
ui <- fluidPage(
  titlePanel("Radar Chart of FC24 Players"),
  sidebarLayout(
    sidebarPanel(
      numericInput("player_number", "Player Number:", value = 12, min = 1, max = nrow(eafc_data))
    ),
    mainPanel(
      plotOutput("radarChart")
    )
  )
)

# Define server
server <- function(input, output) {

  scores <- data.frame(
    row.names = c(eafc_data$X),
    Pace = c(eafc_data$Pace),
    Shooting = c(eafc_data$Shooting),
    Passing = c(eafc_data$Passing),
    Dribbling = c(eafc_data$Dribbling),
    Defending = c(eafc_data$Defending),
    Physicality = c(eafc_data$Physicality)
  )

  max_min <- data.frame(
    Pace = c(100, 0), Shooting = c(100, 0), Passing = c(100, 0),
    Dribbling = c(100, 0), Defending = c(100, 0), Physicality = c(100, 0)
  )

  rownames(max_min) <- c("Max", "Min")

  df <- rbind(max_min, scores)

  output$radarChart <- renderPlot({
    player_num <- input$player_number

    player_radarchart <- df[c("Max", "Min", as.character(player_num)), ]

    radarchart(
      player_radarchart,
      pcol = "purple", pfc col = scales::alpha("blue", 0.5), plwd = 2, plty = 1,
      cglcol = "grey", cglty = 2, cglwd = 2,
      title = "Player's Rating"
    )
  })
}

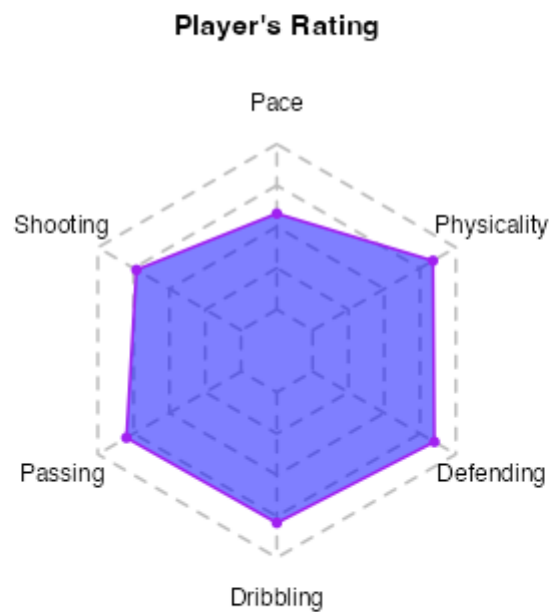
```

```
}  
  
# Run the app  
shinyApp(ui = ui, server = server)
```

```
##  
## Listening on http://127.0.0.1:7553
```

Radar Chart of FC24 Players

Player Number:



#1. Which club has the highest average rating of players?

```
library(ggplot2)
library(tidyverse)
```

```
eaafc_data <- read.csv("male_players.csv") %>% select(Name, Nation, Club, Position, Overall, Pace, Shooting, Passing, Dribbling, Defending, Physicality)
```

```
club_average_rating <- eaafc_data %>%
  group_by(Club) %>%
  summarise(mean_rating = mean(Overall))
```

```
club_with_highest_average_rating <- club_average_rating %>%
  arrange(desc(mean_rating))
```

```
club_with_highest_average_rating
```

```
## # A tibble: 654 × 2
##   Club                mean_rating
##   <chr>                <dbl>
## 1 FC Bayern München    81.4
## 2 Inter                 80.0
## 3 Paris SG             79.9
## 4 Manchester City      78.6
## 5 Real Madrid          78.1
## 6 OL Reign             78
## 7 Spurs                77.9
## 8 Manchester Utd        77.9
## 9 Atlético de Madrid   77.7
## 10 FC Barcelona         77.6
## # i 644 more rows
```

```
top_20_clubs <- club_with_highest_average_rating %>% slice_head(n=20)
```

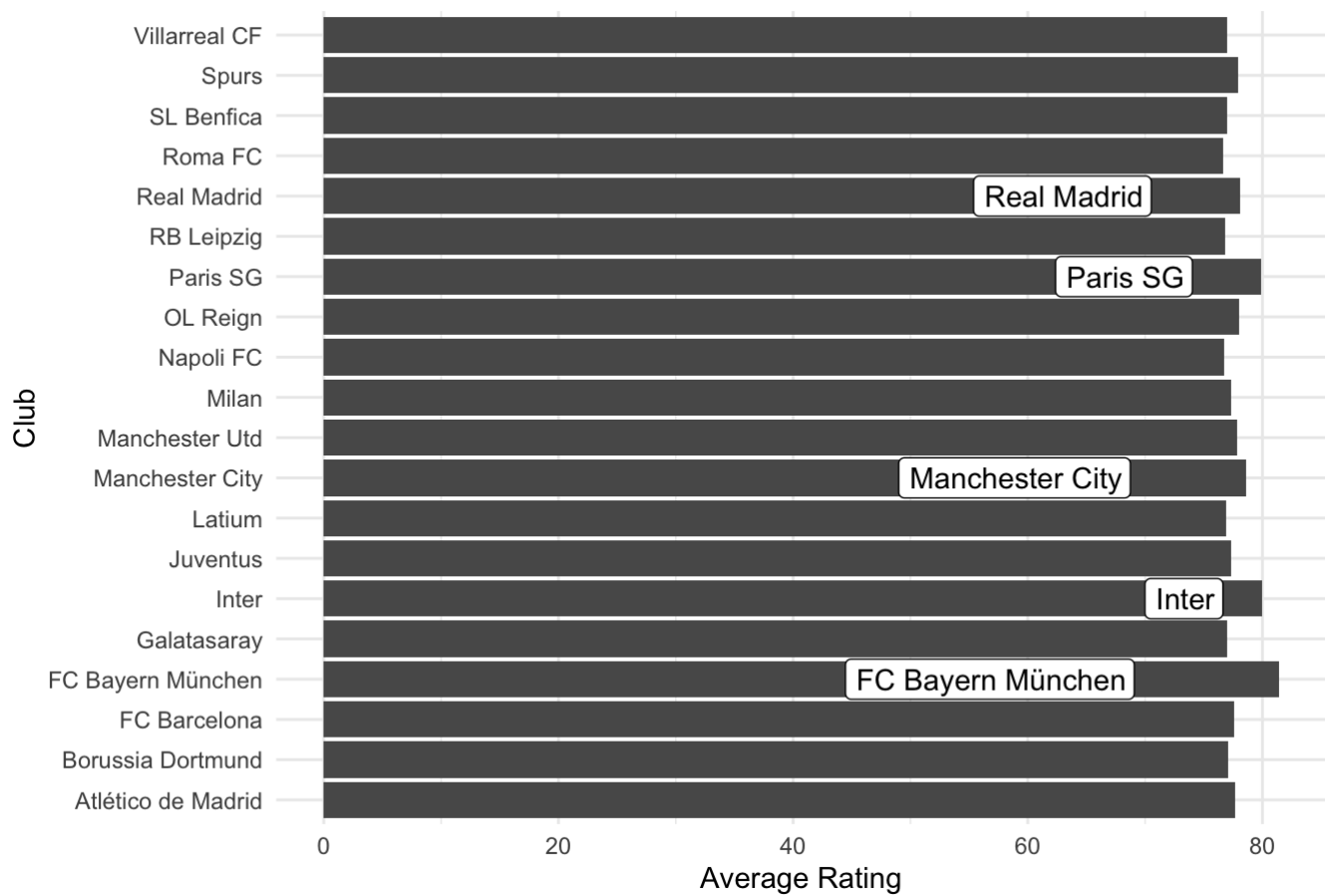
```
is_top_five <- club_with_highest_average_rating %>% slice_head(n=5)
```

```
# Plot the average rating for the top 50 clubs
```

```
# Add labels to the top three clubs
```

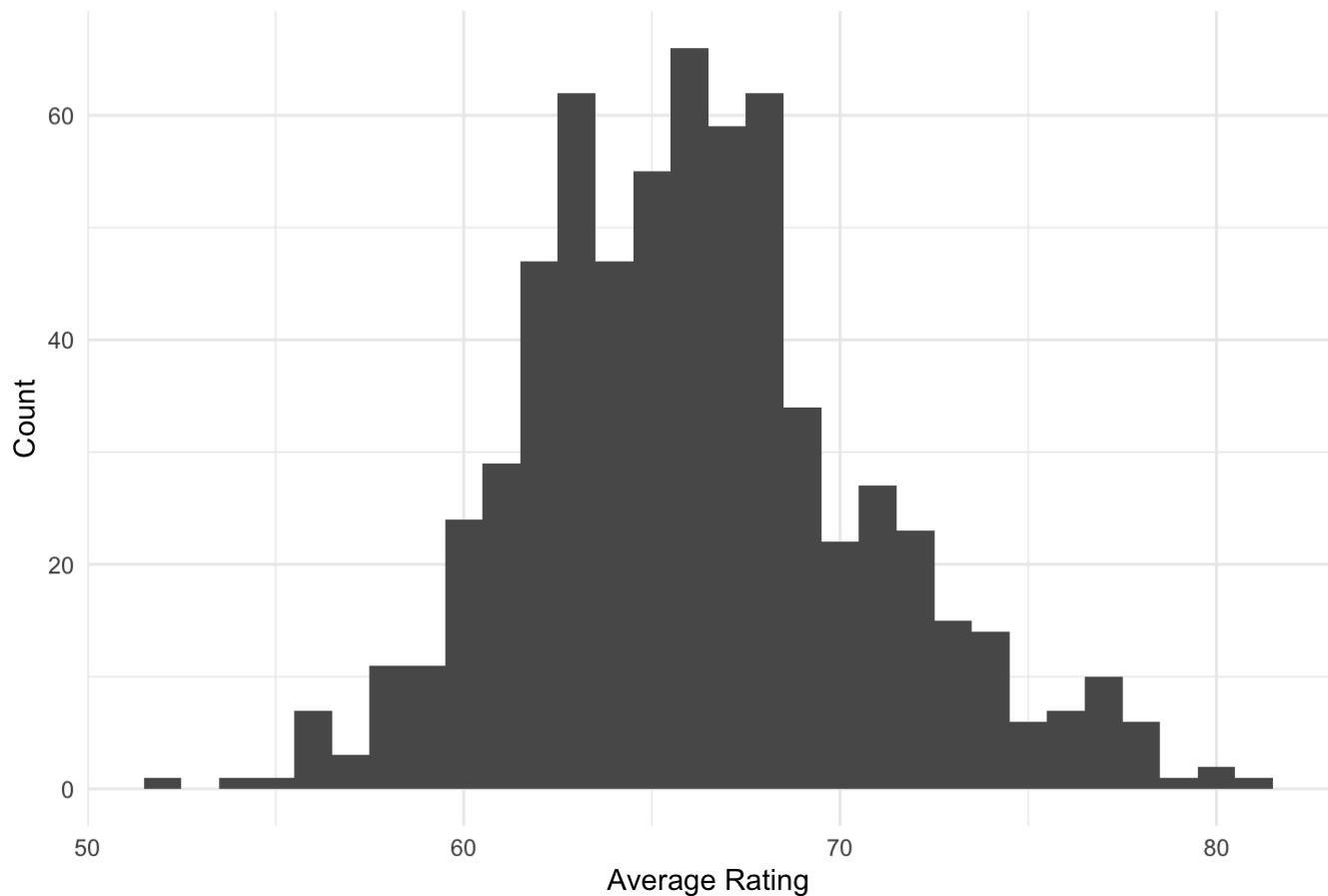
```
ggplot(top_20_clubs, aes(x = Club, y = mean_rating)) +
  geom_bar(stat = "identity")+
  labs(title = "Average Rating of FC-24 Players by Club (Top 20)", x = "Club", y = "Average Rating") +
  theme_minimal() + coord_flip() +
  geom_label(aes(label = Club), data = (is_top_five), hjust = 1.5, vjust = 0.5, fill = "white")
```

Average Rating of FC-24 Players by Club (Top 20)



```
ggplot(club_average_rating, aes(x = mean_rating)) +
  geom_histogram(binwidth = 1) +
  labs(title = "Average Rating of FC-24 Players by Club Distribution", x = "Average Rating", y = "Count") +
  theme_minimal()
```

Average Rating of FC-24 Players by Club Distribution



#2. Which country has the highest count of professional players?

```
library(tidyverse)
```

```
library(ggplot2)
```

```
# Count the number of players from each country
```

```
country_player_count <- eafc_data %>%
```

```
  group_by(Nation) %>%
```

```
  summarise(player_count = n())
```

```
# Identify the country with the highest player count
```

```
country_with_highest_player_count <- country_player_count %>%
```

```
  arrange(desc(player_count)) %>%
```

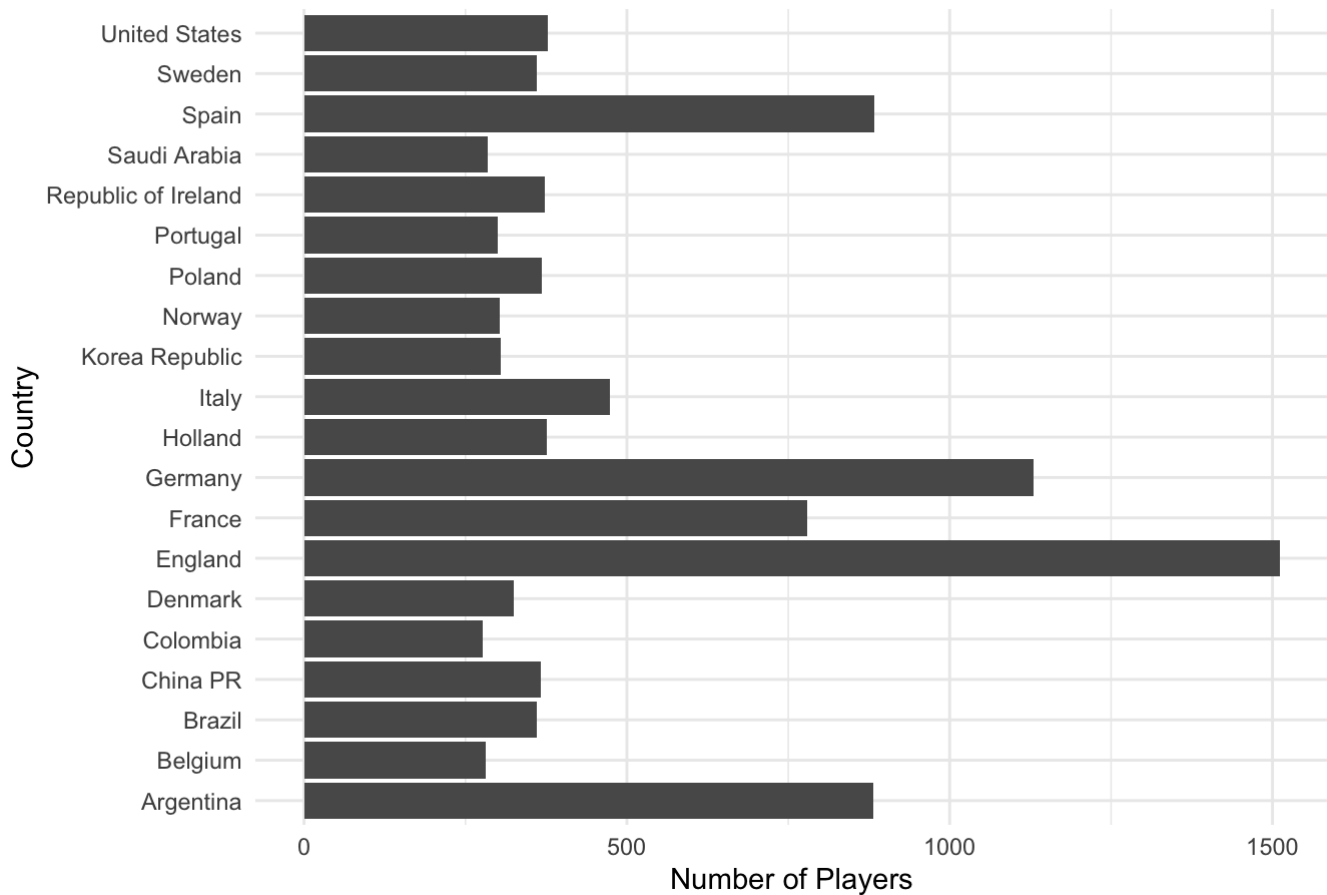
```
  slice_head(n=20)
```

```
country_with_highest_player_count
```

```
## # A tibble: 20 × 2
##   Nation                player_count
##   <chr>                  <int>
## 1 England                1511
## 2 Germany                1130
## 3 Spain                  883
## 4 Argentina              881
## 5 France                 780
## 6 Italy                  474
## 7 United States          378
## 8 Holland                376
## 9 Republic of Ireland    373
## 10 Poland                368
## 11 China PR              367
## 12 Brazil                360
## 13 Sweden                360
## 14 Denmark               325
## 15 Korea Republic        305
## 16 Norway                303
## 17 Portugal              300
## 18 Saudi Arabia          284
## 19 Belgium               282
## 20 Colombia              276
```

```
# Plot the player count for each country
ggplot(country_with_highest_player_count, aes(x = Nation, y = player_count)) +
  geom_bar(stat = "identity") +
  labs(title = "Number of FC-24 Players by Country (Top 20)", x = "Country", y = "Number of Players") +
  theme_minimal() + coord_flip()
```


Number of FC-24 Players by Country (Top 20)



#3. What percentage of players have a overall rating higher than 80, what does that imply?

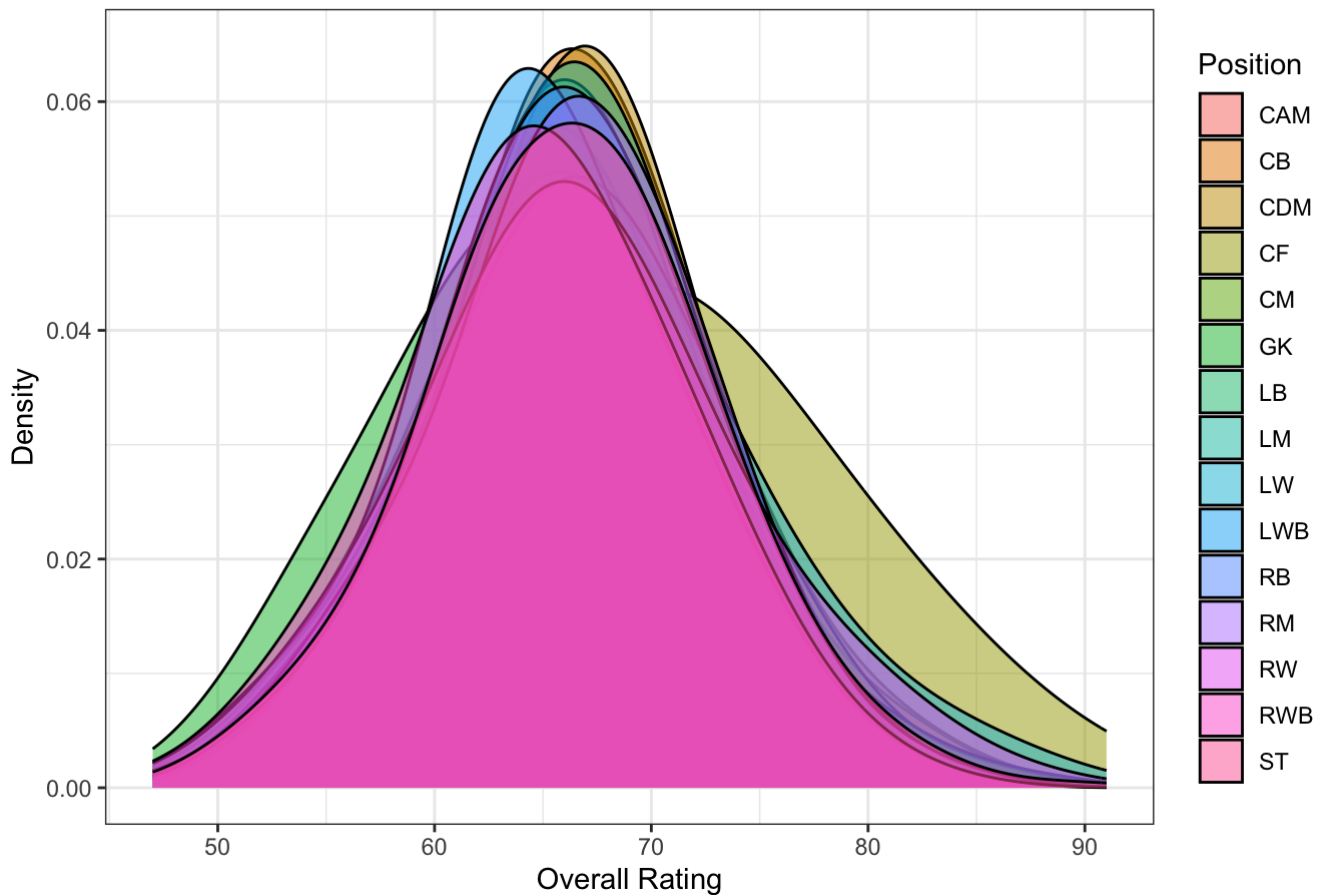
```
number_of_players_with_rating_above_80 <- length(which(eafc_data$Overall > 80))
total_number_of_players <- nrow(eafc_data)
percentage_of_players_with_rating_above_80 <- number_of_players_with_rating_above_80
/ total_number_of_players * 100
```

```
# Print the percentage
print(percentage_of_players_with_rating_above_80)
```

```
## [1] 2.240454
```

```
ggplot(eafc_data, aes(x = Overall, fill = Position)) +
  geom_density(adjust = 2, alpha = 0.5) +
  labs(title = "Distribution of Overall Ratings in FC-24", x = "Overall Rating", y =
"Density") +
  theme_bw()
```

Distribution of Overall Ratings in FC-24



Week 13 Final submission

(1) What is the theme of your data story?

The theme of the data story is to investigate the data set of football players' rating from EAFC24, and to provide insights about the data. The main question that would be answered is "what are the insights that can be drawn from the football players' rating in 2024?" This can be broken down into various questions: 1. Which club has the highest average rating of players? 2. Which country has the highest count of professional players? 3. What percentage of players have a overall rating higher than 80, what does that imply?

(2) Why is it important to address this question?

Football is one of the mainstream entertainment / sports, and by presenting the data about players in an comprehensive way, one can search and learn more about the player or teams of their interest, as well as the detail value after

different categories such as pace, shooting, strength etc. The specific questions are chosen because by answering these question, it provides insight into topic in various aspects, which explore the data from the view of club teams, national teams, and individual players.

(3) Why do you think the data sources that you have curated can help you answer the question?

the data set used in this project is a csv file that contain all ratings of the male football players registered in the EAFC24 game, the raw data set contains 15845 rows of observations with 40 categories (columns) of data for each observation. By filtering the desired categories, this detailed data set is sufficient to answer the questions constructed.

(4) What are the insights from the data and how are they depicted in plots?

From exploring the data and answering the designed questions, we were able to observe various insights. Firstly from the first question, we had a rough idea of teams rating distribution, that majority of the clubs have a average player rating from 60 to 70, this was observed from the distribution graph. This distribution was also compared with the top teams, where the Top 20 club was listed out and plot a separate graph, showing the competetiveness among the top teams. From the second question, the players were viewed from the point of national team, where the top nations are listed out and a graph was plotted, showing that the number of players differ significantly among the top countries. From the third question, viewing from the perspective of individual player, the point was stressed on the extreme difficulty to be a player of the top level, as indicated by the rating distribution graph, dissected by different playing position.

(5) How did you implement this entire project? Were there any new concepts that you learnt to implement some aspects of it?

For all of the three questions, grouping and arranging was the main steps involved in interpreting the data, as well as labelling and plotting the graphs in a visual appealing way, which involves using function like `coord_flip()` so the axis label will not overlap with one each other. In addition to answer the questions, two shiny apps are employed at the end, one allowing the user to search any player they want to, and one allowing user to display the player's detail rating with radar chart. Both shiny app were quite challenging to me, especially the radar chart one which involve using package "fmsb" that is not taught in class and prepare the data so that it could fit with the function.

Reference List

Coder, R. (2021, March 24). Radar chart (spider plot) in R with fmsb. R CHARTS | a Collection of Charts and Graphs Made With the R Programming Language. <https://r-charts.com/ranking/radar-chart/> (<https://r-charts.com/ranking/radar-chart/>)

A. (2020, December 12). Beautiful Radar Chart in R using FMSB and GGPlot Packages. Datanovia. <https://www.datanovia.com/en/blog/beautiful-radar-chart-in-r-using-fmsb-and-ggplot-packages/#prerequisites> (<https://www.datanovia.com/en/blog/beautiful-radar-chart-in-r-using-fmsb-and-ggplot-packages/#prerequisites>)

Horst, J. L. A. (2021, January 8). Chapter 3 R & RStudio, RMarkdown | R for Excel Users. <https://rstudio-conf-2020.github.io/r-for-excel/rstudio.html> (<https://rstudio-conf-2020.github.io/r-for-excel/rstudio.html>)

Holtz, Y. (n.d.). Spider Chart | the R Graph Gallery. <https://r-graph-gallery.com/spider-or-radar-chart.html> (<https://r-graph-gallery.com/spider-or-radar-chart.html>)

10+ EA SPORTS FC 24 HD Wallpapers and Backgrounds. (n.d.). Wallpaper Abyss. https://wall.alphacoders.com/by_sub_category.php?id=364319&name=EA+SPORTS+FC+24+Wallpapers (https://wall.alphacoders.com/by_sub_category.php?id=364319&name=EA+SPORTS+FC+24+Wallpapers)

EA SPORTS FC 24 FULL PLAYERS DATABASE AND STATS. (2023, October 9). Kaggle. <https://www.kaggle.com/datasets/nyagami/fc-24-players-database-and-stats-from-easports/data> (<https://www.kaggle.com/datasets/nyagami/fc-24-players-database-and-stats-from-easports/data>)