

Preliminary EDA

Kush Lalwani

2024-04-15

```
library(tidyverse)
```

```
## -- Attaching core tidyverse packages ----- tidyverse 2.0.0 --
## v dplyr      1.1.4      v readr      2.1.5
## v forcats    1.0.0      v stringr   1.5.1
## v ggplot2    3.4.4      v tibble    3.2.1
## v lubridate  1.9.3      v tidyr     1.3.0
## v purrr      1.0.2
## -- Conflicts ----- tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()     masks stats::lag()
## i Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to become errors
```

```
library(rvest)
```

```
##
## Attaching package: 'rvest'
##
## The following object is masked from 'package:readr':
##
##     guess_encoding
```

```
library(chromote)
```

```
library(ggpubr)
```

Research Question

I will be analyzing the relationship between a player's wages and their in game performance. The question is: Does a player's wages affect their performance in game?

Primary Data Source

- I found my data sources from FBref.com, also known as Football Reference.
- This data is put into table form by Football Reference, but the original data is from Capology. Which is a website that tracks and stores all financial information about players and clubs.
- The data was originally collected by Capology. "Capology is used by professional football clubs, agencies, as well as technology platforms, to analyze trends, improve scouting and contract negotiations, and build complex software solutions" which is quoted directly from their website.

- Here each case in the table represents a player in the Premier League. It contains their name, nationality, team, position, age, and weekly/annual wage in pounds, euros, and dollars. It contains the data of 592 players
- I plan to use the name and team variables to join this table to a secondary data source. I will also use the annual player salary to make my observations. Finally, I will use the position variable to determine which secondary table I join with. For example, I cannot judge a goalkeepers performance based on goals and assists, so I will use the position to determine how statistics are evaluated.

```
wage_link<-"https://fbref.com/en/comps/9/wages/Premier-League-Wages"
wage_stats <- wage_link %>%
  read_html() %>%
  html_elements(css ="table")%>%
  html_table()

wage_stats <- wage_stats[[2]]
head(wage_stats)
```

```
## # A tibble: 6 x 9
##      Rk Player      Nation Pos   Squad   Age 'Weekly Wages' 'Annual Wages' Notes
##   <int> <chr>      <chr> <chr> <chr> <int> <chr>          <chr>      <chr>
## 1     1 Kevin De B~ be BEL MF   Manc~   32 £ 400,000 (€ ~ £ 20,800,000 ~ ""
## 2     2 Erling Haa~ no NOR FW   Manc~   23 £ 375,000 (€ ~ £ 19,500,000 ~ ""
## 3     3 Casemiro    br BRA MF   Manc~   31 £ 350,000 (€ ~ £ 18,200,000 ~ ""
## 4     4 Mohamed Sa~ eg EGY FW   Live~   31 £ 350,000 (€ ~ £ 18,200,000 ~ ""
## 5     5 Raphaël Va~ fr FRA DF   Manc~   30 £ 340,000 (€ ~ £ 17,680,000 ~ ""
## 6     6 Raheem Ste~ eng E~ FW,MF Chel~   28 £ 325,000 (€ ~ £ 16,900,000 ~ ""
```

Here we have the raw data directly from the webpage. We see that is not in tidy data form. The rank variable is not needed and neither is nation or notes, so they can be selected out. Also the wages are a character string in three different currencies, we want it in only dollars.

```
extract_dollar <- function(value) {
  dollar_amount <- gsub(".*\\$([0-9,]+).*", "\\1", value)
  dollar_amount <- as.numeric(gsub(",", "", dollar_amount))
  return(dollar_amount)
}
```

This function is used to convert any of the wage amounts into a numeric dollar amount

```
player_wages <-
  wage_stats %>%
  select(!c(Rk,Notes,Nation,Age,`Weekly Wages`)) %>%
  rename(name = Player, position = Pos, team = Squad, pay_year = `Annual Wages`) %>%
  mutate(pay_year = extract_dollar(pay_year))
head(player_wages)
```

```
## # A tibble: 6 x 4
##   name      position team      pay_year
##   <chr>      <chr>   <chr>      <dbl>
## 1 Kevin De Bruyne MF      Manchester City 26253113
## 2 Erling Haaland FW      Manchester City 24612294
## 3 Casemiro    MF      Manchester Utd 22971474
```

```
## 4 Mohamed Salah    FW      Liverpool    22971474
## 5 Raphaël Varane   DF      Manchester Utd 22315146
## 6 Raheem Sterling FW,MF    Chelsea      21330654
```

Now the data is in tidy data form, and has been converted to US dollars # Secondary Data Sources - For my secondary data sources I will use different tables from the same website. I will use four different sets based on the four different positions a player could be. I will use Player Goalkeeping for the goalkeepers. Defenders will be judged based on Player Defensive Action. Midfielders will be judged based on Player Passing. Finally Attackers will be judged based on Player Shooting. - This data is collected by Opta, which is an organization in the UK that collects all kinds of stats related to the Premier League. - These stats were collected to analyze a players performance during a game. - I only want to analyze players that have played at least 5 games

```
get_table <- function(link){
  page <- link %>%
    read_html_live() %>%
    html_elements(css = "table") %>%
    html_table()
  out <- page[[3]]
  return(out)
}
```

Making function to read the live html. I had trouble reading the table I needed from the website. I found out that the table was loaded using javascript and not directly in the HTML. So to solve this issue, I used a new function in rvest called read_html_live(). This function uses the chromote package, which uses google chrome to load the entire website and read it directly. Unfortunately, this function doesn't work when trying to knit into a pdf. So I will use csv files that were copy and pasted instead. Now the data needs to be put into tidy data form and I will only select the stats that I am interested in.

Goalkeepers

For goalkeepers, performance will be rated on their save percentage, goals allowed, and penalty saves

```
gk_link<-"https://fbref.com/en/comps/9/keepers/Premier-League-Stats"
#gk_stats <- get_table(gk_link)

gk_stats <- read.csv("goalkeeper stats.csv")

#colnames(gk_stats) <- as.character(gk_stats[1,])
colnames(gk_stats)[26] <- "PKSave%"
head(gk_stats)
```

##	Rk	Player	Nation	Pos	Squad	Age	Born	MP	Starts	Min						
## 1	1	Alisson	br BRA	GK	Liverpool	31-197	1992	22	22	1,980						
## 2	2	Alphonse Areola	fr FRA	GK	West Ham	31-049	1993	27	27	2,339						
## 3	3	Daniel Bentley	eng ENG	GK	Wolves	30-278	1993	4	2	293						
## 4	4	Martin Dúbravka	sk SVK	GK	Newcastle Utd	35-092	1989	18	17	1,535						
## 5	5	Ederson	br BRA	GK	Manchester City	30-243	1993	28	28	2,402						
## 6	6	Łukasz Fabiański	pl POL	GK	West Ham	38-364	1985	9	6	631						
##	X90s	GA	GA90	SoTA	Saves	Save.	W	D	L	CS	CS.	PKatt	PKA	PKsv	PKm	PKSave%
## 1	22	20	0.91	77	58	75.3	13	6	3	7	31.8	1	1	0	0	0
## 2	26	42	1.62	153	111	75.8	10	8	9	4	14.8	7	5	2	0	28.6

```
## 3 3.3 4 1.23 12 8 66.7 1 0 1 1 50 0 0 0 0
## 4 17.1 34 1.99 102 67 70.6 7 3 7 5 29.4 5 4 1 0 20
## 5 26.7 26 0.97 74 49 68.9 19 6 3 8 28.6 3 3 0 0 0
## 6 7 16 2.28 46 30 71.7 3 1 2 1 16.7 3 3 0 0 0
## Matches
## 1 Matches
## 2 Matches
## 3 Matches
## 4 Matches
## 5 Matches
## 6 Matches
```

```
goalkeeper_stat <-
  gk_stats %>%
  select(c(Player,Pos,Squad,`X90s`,GA90,`Save.`,`PKSave%`)) %>%
  rename(name = Player, position = Pos, team = Squad, games = `X90s`, goal_against_per90 = GA90, save_percentage = `Save.`, pk_save_percentage = `PKSave%`)
  filter(team != "Squad") %>%
  mutate(games = as.numeric(games),goal_against_per90 = as.numeric(goal_against_per90),save_percentage = as.numeric(save_percentage),pk_save_percentage = as.numeric(pk_save_percentage))
  filter(games >= 5)

head(goalkeeper_stat)
```

```
##           name position           team games goal_against_per90
## 1      Alisson      GK      Liverpool  22.0             0.91
## 2 Alphonse Areola      GK      West Ham  26.0             1.62
## 3 Martin Dúbravka      GK Newcastle Utd  17.1             1.99
## 4      Ederson      GK Manchester City  26.7             0.97
## 5 Łukasz Fabiański      GK      West Ham   7.0             2.28
## 6      Mark Flekken      GK      Brentford 31.5             1.78
## save_percentage penalty_save_percentage
## 1           75.3              0.0
## 2           75.8             28.6
## 3           70.6             20.0
## 4           68.9              0.0
## 5           71.7              0.0
## 6           65.8              0.0
```

```
#Now we need to join this plot with the wage plot
goalkeeper_stat <-
  goalkeeper_stat %>%
  left_join(player_wages, by = c("name","position","team"))

head(goalkeeper_stat)
```

```
##           name position           team games goal_against_per90
## 1      Alisson      GK      Liverpool  22.0             0.91
## 2 Alphonse Areola      GK      West Ham  26.0             1.62
## 3 Martin Dúbravka      GK Newcastle Utd  17.1             1.99
## 4      Ederson      GK Manchester City  26.7             0.97
## 5 Łukasz Fabiański      GK      West Ham   7.0             2.28
## 6      Mark Flekken      GK      Brentford 31.5             1.78
## save_percentage penalty_save_percentage pay_year
## 1           75.3              0.0 9844917
```

```
## 2          75.8          28.6 7875934
## 3          70.6          20.0 2625311
## 4          68.9           0.0 6563278
## 5          71.7           0.0 4266131
## 6          65.8           0.0 1968983
```

Defenders

For defenders, performance will be rated on their tackle success percentage, blocks, interceptions, and errors leading to goals.

```
df_link<-"https://fbref.com/en/comps/9/defense/Premier-League-Stats"
#df_stats <- get_table(df_link)

df_stats <- read.csv("defender_stats.csv")

colnames(df_stats) <- as.character(df_stats[1,])
head(df_stats)
```

```
##   Rk      Player Nation Pos      Squad      Age Born  90s Tkl TklW
## 1 Rk      Player Nation Pos      Squad      Age Born  90s Tkl TklW
## 2  1      Max Aarons eng ENG   DF   Bournemouth 24-104 2000 12.3 28  18
## 3  2 Bénie Adama Traore ci CIV FW,MF Sheffield Utd 21-139 2002  4.3  4   2
## 4  3      Tyler Adams us  USA   MF   Bournemouth 25-063 1999  1.2  4   3
## 5  4      Tosin Adarabioyo eng ENG   DF      Fulham 26-206 1997  17 20  11
## 6  5      Elijah Adebayo eng ENG   FW   Luton Town 26-101 1998 12.9  1   0
##   Def 3rd Mid 3rd Att 3rd Tkl Att Tkl% Lost Blocks Sh Pass Int Tkl+Int Clr Err
## 1 Def 3rd Mid 3rd Att 3rd Tkl Att Tkl% Lost Blocks Sh Pass Int Tkl+Int Clr Err
## 2      20      6      2 19 30 63.3  11      9  5  4  6      34 23  0
## 3      1      2      1  0  4  0      4      4  1  3  1      5  1  0
## 4      1      3      0  2  3 66.7  1      2  0  2  4      8  2  0
## 5     15      5      0 11 18 61.1  7     15 11  4 23     43 75  0
## 6      1      0      0  0  9  0      9      8  1  7  5      6 23  0
##   Matches
## 1 Matches
## 2 Matches
## 3 Matches
## 4 Matches
## 5 Matches
## 6 Matches
```

```
defender_stats <-
  df_stats %>%
  select(c(Player,Pos,Squad,`90s`,`Tkl%`,Blocks,Int,Err)) %>%
  rename(name=Player,position=Pos,team=Squad,games = `90s`,
         tackle_percent=`Tkl%`,block=Blocks,interceptions=Int,errors=Err) %>%
  filter(team != "Squad") %>%
  filter(grepl("DF",position)) %>%
  mutate(tackle_percent=as.numeric(tackle_percent),block=as.numeric(block), games = as.numeric(games),
         interceptions=as.numeric(interceptions),errors=as.numeric(errors)) %>%
  filter(games >= 5)

head(defender_stats)
```

```
##           name position           team games tackle_percent block
## 1      Max Aarons      DF      Bournemouth  12.3         63.3     9
## 2 Tosin Adarabioyo      DF          Fulham  17.0         61.1    15
## 3      Nayef Aguerd      DF       West Ham  20.6         72.7    33
## 4 Anel Ahmedhodžić      DF Sheffield Utd  24.4         57.1    36
## 5      Ola Aina      DF Nott'ham Forest  14.8         73.1    14
## 6 Rayan Aït-Nouri    DF,MF        Wolves  21.2         58.3    25
##  interceptions errors
## 1           6      0
## 2          23      0
## 3          17      2
## 4          28      2
## 5          13      1
## 6          12      0
```

```
defender_stats <-
  defender_stats %>%
  left_join(player_wages, by = c("name", "position", "team"))
head(defender_stats)
```

```
##           name position           team games tackle_percent block
## 1      Max Aarons      DF      Bournemouth  12.3         63.3     9
## 2 Tosin Adarabioyo      DF          Fulham  17.0         61.1    15
## 3      Nayef Aguerd      DF       West Ham  20.6         72.7    33
## 4 Anel Ahmedhodžić      DF Sheffield Utd  24.4         57.1    36
## 5      Ola Aina      DF Nott'ham Forest  14.8         73.1    14
## 6 Rayan Aït-Nouri    DF,MF        Wolves  21.2         58.3    25
##  interceptions errors pay_year
## 1           6      0  2297147
## 2          23      0  2625311
## 3          17      2  3281639
## 4          28      2  1968983
## 5          13      1  2625311
## 6          12      0   656328
```

Midfielders

For midfielders, performance will be based on their pass completion percentage, assists (actual and expected), and progressive passes.

```
md_link<-"https://fbref.com/en/comps/9/passing/Premier-League-Stats"
#md_stats <- get_table(md_link)

md_stats <- read.csv("midfield_stats.csv")

colnames(md_stats) <- as.character(md_stats[1,])
colnames(md_stats)[16] <- "short_cmp%"
colnames(md_stats)[19] <- "med_cmp%"
colnames(md_stats)[22] <- "long_cmp%"
head(md_stats)
```

```
##  Rk           Player Nation   Pos           Squad   Age Born  90s Cmp  Att
```

```
## 1 Rk      Player Nation Pos      Squad Age Born 90s Cmp Att
## 2 1      Max Aarons eng ENG   DF   Bournemouth 24-104 2000 12.3 394 516
## 3 2 B  nie Adama Traore ci CIV FW,MF Sheffield Utd 21-139 2002 4.3 55 71
## 4 3      Tyler Adams us USA   MF   Bournemouth 25-063 1999 1.2 50 62
## 5 4      Tosin Adarabioyo eng ENG   DF   Fulham 26-206 1997 17 983 1156
## 6 5      Elijah Adebayo eng ENG   FW   Luton Town 26-101 1998 12.9 128 177
##      Cmp% TotDist PrgDist Cmp Att short_cmp% Cmp Att med_cmp% Cmp Att long_cmp%
## 1 Cmp% TotDist PrgDist Cmp Att      Cmp% Cmp Att      Cmp% Cmp Att      Cmp%
## 2 76.4      6518      2525 193 220      87.7 163 209      78 32 57      56.1
## 3 77.5      775      185 34 38      89.5 19 23      82.6 1 1      100
## 4 80.6      841      242 30 34      88.2 14 20      70 6 8      75
## 5 85      20174      7607 260 289      90 601 667      90.1 113 168      67.3
## 6 72.3      1627      336 84 105      80 35 52      67.3 3 6      50
##      Ast xAG xA A-xAG KP 3-Jan PPA CrsPA PrgP Matches
## 1 Ast xAG xA A-xAG KP 3-Jan PPA CrsPA PrgP Matches
## 2 1 0.8 0.9 0.2 7 22 13 2 41 Matches
## 3 0 0.5 0.5 -0.5 4 2 7 1 9 Matches
## 4 0 0.1 0 -0.1 1 8 0 0 4 Matches
## 5 0 0.1 0.3 -0.1 1 46 2 0 60 Matches
## 6 0 0.7 0.4 -0.7 14 7 3 0 19 Matches
```

```
midfield_stats <- md_stats %>%
  select(c(Player, Pos, Squad, `90s`, `Cmp%`, Ast, xA, PrgP)) %>%
  rename(name = Player, position = Pos, team = Squad, games = `90s`, cmp_perc = `Cmp%`, assists = Ast,
         xAssists = xA, prog_pass = PrgP) %>%
  filter(name != "Player") %>%
  filter(grepl("MF", position)) %>%
  mutate(games = as.numeric(games), cmp_perc = as.numeric(cmp_perc), assists = as.numeric(assists),
         xAssists = as.numeric(xAssists), prog_pass = as.numeric(prog_pass)) %>%
  filter(games >= 5)

head(midfield_stats)
```

```
##      name position      team games cmp_perc assists xAssists
## 1 Rayan Ait-Nouri DF,MF Wolves 21.2 84.8 1 1.5
## 2 Manuel Akanji DF,MF Manchester City 22.1 93.2 0 1.4
## 3 Edson   lvarez MF West Ham 23.9 85.6 1 0.7
## 4 Juli  n   lvarez MF,FW Manchester City 27.2 79.8 8 6.2
## 5 Sofyan Amrabat MF,DF Manchester Utd 7.2 86.1 0 0.2
## 6 Elliot Anderson MF,FW Newcastle Utd 7.5 79.2 1 0.6
##      prog_pass
## 1 90
## 2 124
## 3 79
## 4 97
## 5 52
## 6 41
```

```
midfield_stats <-
  midfield_stats %>%
  left_join(player_wages, by = c("name", "position", "team"))

head(midfield_stats)
```

```
##           name position           team games cmp_perc assists xAssists
## 1 Rayan Aït-Nouri    DF,MF         Wolves  21.2    84.8        1      1.5
## 2  Manuel Akanji    DF,MF Manchester City  22.1    93.2        0      1.4
## 3  Edson Álvarez      MF         West Ham  23.9    85.6        1      0.7
## 4  Julián Álvarez    MF,FW Manchester City  27.2    79.8        8      6.2
## 5  Sofyan Amrabat    MF,DF Manchester Utd   7.2    86.1        0      0.2
## 6  Elliot Anderson   MF,FW Newcastle Utd   7.5    79.2        1      0.6
##   prog_pass pay_year
## 1         90  656328
## 2        124 11813901
## 3         79  6563278
## 4         97  6563278
## 5         52  4266131
## 6         41  1968983
```

Forwards

For forwards, performance will be based on their goals (actual and expected) and shot on target percentage.

```
fw_link<-"https://fbref.com/en/comps/9/shooting/Premier-League-Stats"
#fw_stats <- get_table(fw_link)

fw_stats <- read.csv("attacker stats.csv")

colnames(fw_stats) <- as.character(fw_stats[1,])

head(fw_stats)
```

```
##   Rk           Player Nation Pos      Squad      Age Born  90s Gls Sh SoT
## 1 Rk           Player Nation Pos      Squad      Age Born  90s Gls Sh SoT
## 2 1      Max Aarons eng ENG   DF   Bournemouth 24-104 2000 12.3  0  2  0
## 3 2  Béné Adama Traore ci CIV FW,MF Sheffield Utd 21-139 2002  4.3  0  1  1
## 4 3      Tyler Adams us USA   MF   Bournemouth 25-063 1999  1.2  0  0  0
## 5 4  Tosin Adarabioyo eng ENG   DF           Fulham 26-206 1997  17   2 14  5
## 6 5  Elijah Adebayo eng ENG   FW   Luton Town 26-101 1998 12.9  9 28 12
##   SoT% Sh/90 SoT/90 G/Sh G/SoT Dist FK PK PKatt  xG npG npG/Sh G-xG np:G-xG
## 1 SoT% Sh/90 SoT/90 G/Sh G/SoT Dist FK PK PKatt  xG npG npG/Sh G-xG np:G-xG
## 2  0  0.16      0      0      23.9 0 0      0 0 0      0.02  0      0
## 3 100  0.23  0.23  0      0 15.3 0 0      0 0.3 0.3      0.27 -0.3 -0.3
## 4      0      0      0      0      0 0 0      0 0 0      0      0
## 5 35.7  0.83  0.29 0.14  0.4 14.2 0 0      0 0.7 0.7      0.05 1.3  1.3
## 6 42.9  2.17  0.93 0.32  0.75  9.5 0 0      0 5.6 5.6      0.2  3.4  3.4
##   Matches
## 1 Matches
## 2 Matches
## 3 Matches
## 4 Matches
## 5 Matches
## 6 Matches
```

```
attack_stats <- fw_stats %>%
  select(c(Player, Pos, Squad, `90s`, Gls, `SoT%`, xG)) %>%
```



```

rename(name = Player, position = Pos, team = Squad, games = `90s`, goals = Gl,
       shot_target_perc = `SoT%`, xGoal = xG) %>%
filter(name != "Player") %>%
filter(grepl("FW", position)) %>%
mutate(games = as.numeric(games), goals = as.numeric(goals),
       shot_target_perc = as.numeric(shot_target_perc), xGoal = as.numeric(xGoal)) %>%
filter(games >= 5)

head(attack_stats)

```

```

##           name position           team games goals shot_target_perc xGoal
## 1  Elijah Adebayo      FW      Luton Town  12.9     9           42.9    5.6
## 2   Simon Adingra      FW      Brighton   19.6     6           45.0    3.5
## 3 Miguel Almirón      FW Newcastle Utd  20.5     3           27.5    4.5
## 4 Julián Álvarez  MF,FW Manchester City  27.2     8           38.6   11.3
## 5     Zeki Amdouni      FW      Burnley   20.3     4           51.2    4.4
## 6 Elliot Anderson  MF,FW Newcastle Utd   7.5     0           23.1    1.2

```

```

attack_stats <-
  attack_stats %>%
  left_join(player_wages, by = c("name", "position", "team"))

head(attack_stats)

```

```

##           name position           team games goals shot_target_perc xGoal
## 1  Elijah Adebayo      FW      Luton Town  12.9     9           42.9    5.6
## 2   Simon Adingra      FW      Brighton   19.6     6           45.0    3.5
## 3 Miguel Almirón      FW Newcastle Utd  20.5     3           27.5    4.5
## 4 Julián Álvarez  MF,FW Manchester City  27.2     8           38.6   11.3
## 5     Zeki Amdouni      FW      Burnley   20.3     4           51.2    4.4
## 6 Elliot Anderson  MF,FW Newcastle Utd   7.5     0           23.1    1.2
##   pay_year
## 1    820410
## 2    820410
## 3   3937967
## 4   6563278
## 5   1968983
## 6   1968983

```

Plots

these plots are not final, they are just to get a good idea of how the data looks, I will add more layers to them and revise them when it is time to submit the final report.

GK plot

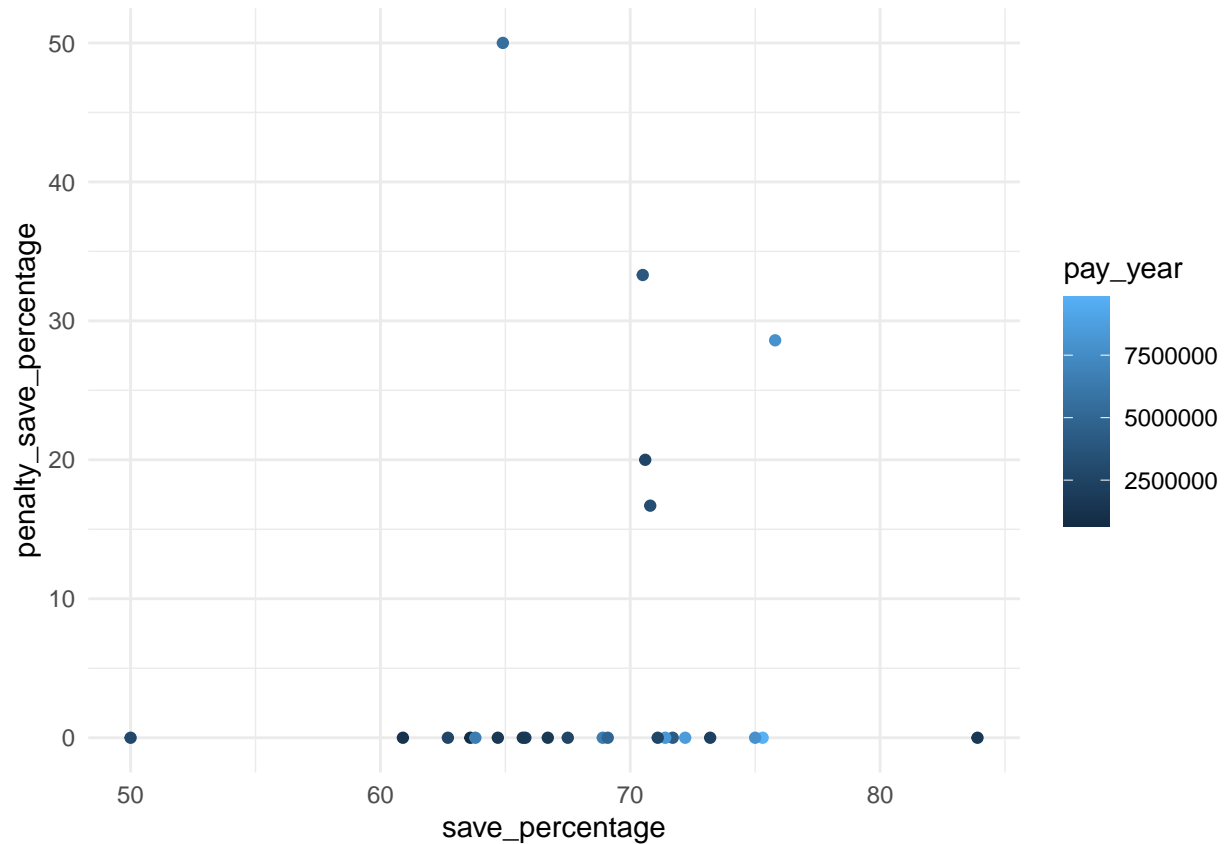
```

ggplot(goalkeeper_stat) +
  aes(x = save_percentage, y = penalty_save_percentage, colour = pay_year) +
  geom_point(shape = "circle", size = 1.5) +

```

```
scale_color_gradient() +  
theme_minimal()
```

```
## Warning: Removed 7 rows containing missing values ('geom_point()').
```

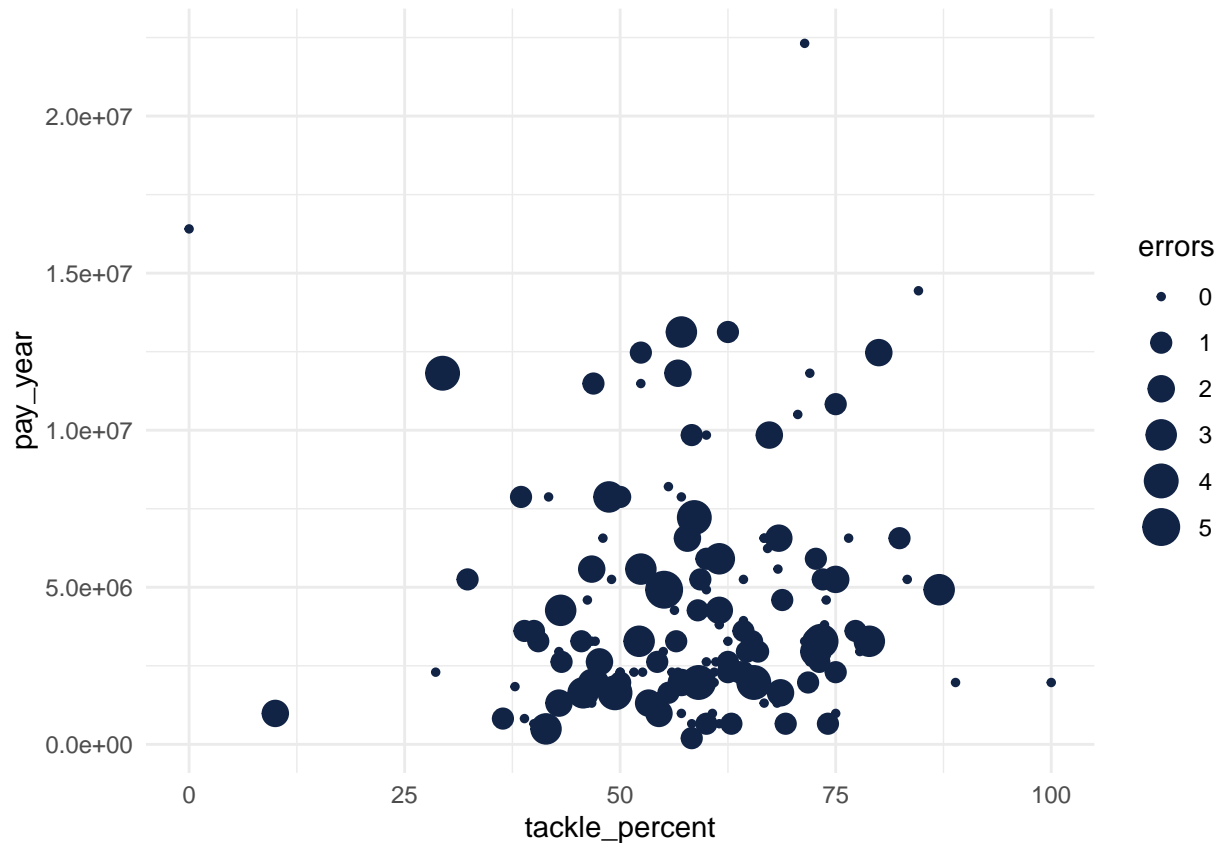


This plot shows save percentage vs penalty save percentage, with yearly pay being the color. as we can see there is not really a relation between both the save percentages, but there is a slight relation between pay and save percentage, with the higher paid goalies having a higher save percentage.

DF Plot

```
ggplot(defender_stats) +  
  aes(x = tackle_percent, y = pay_year, size = errors) +  
  geom_point(shape = "circle", colour = "#112446") +  
  theme_minimal()
```

```
## Warning: Removed 8 rows containing missing values ('geom_point()').
```

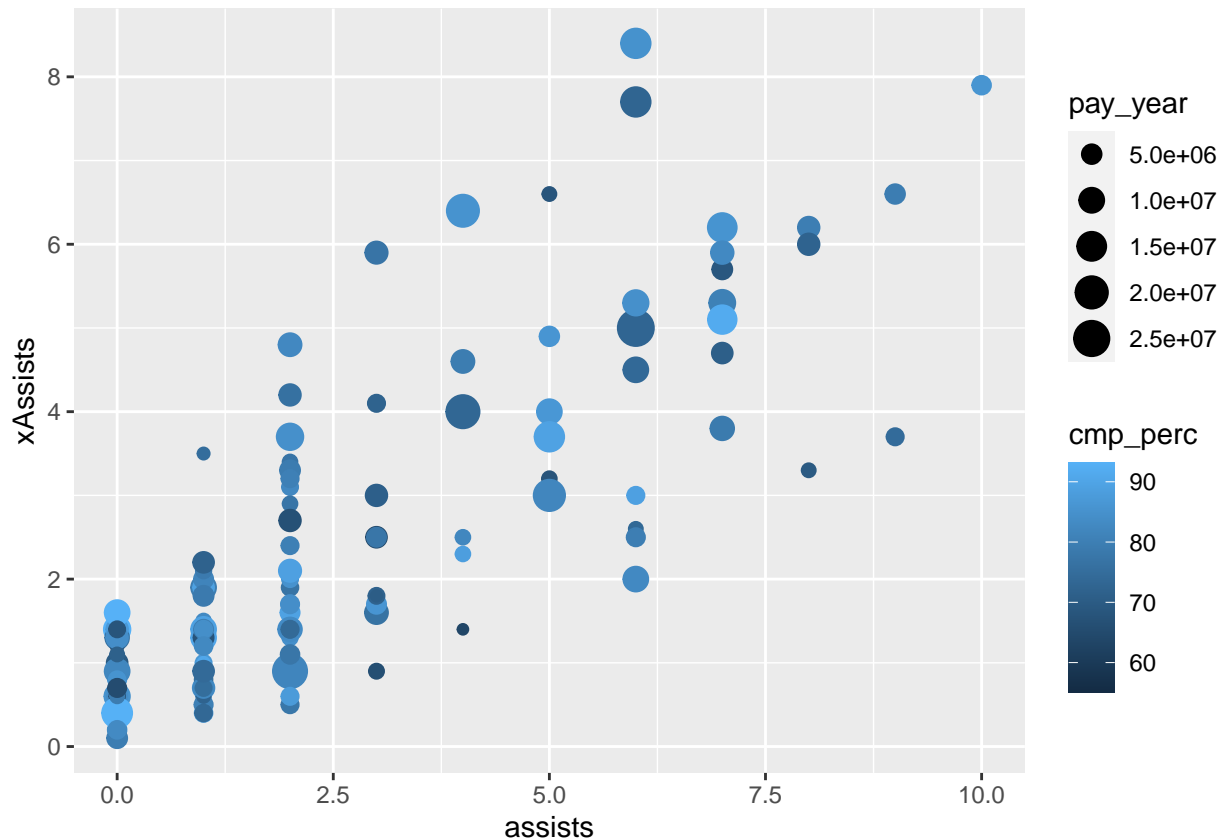


This plots tackle percentage vs the yearly pay, with the errors leading to goal being the size. We can see there is a relation between errors leading to goal and the yearly pay, with most of the players with many errors having a relatively low pay. There is also a slight positive correlation between tackle success percentage and pay.

MF Plot

```
ggplot(midfield_stats) +
  aes(x = assists, y = xAssists, colour = cmp_perc) +
  geom_point(shape = "circle", aes(size = pay_year)) +
  scale_color_gradient()
```

```
## Warning: Removed 15 rows containing missing values ('geom_point()').
```



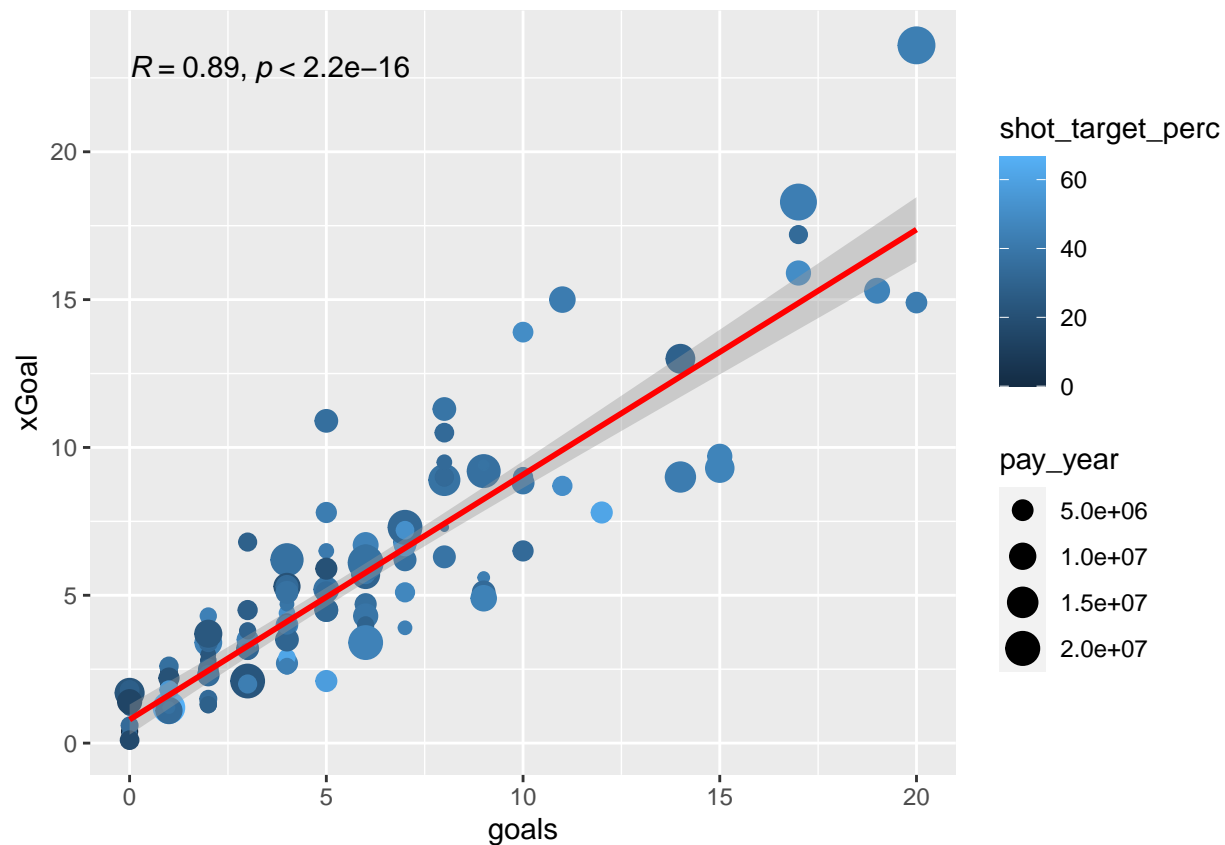
Here we have expected assists vs actual assists. As expected, we can see that there is a relation between expected assists and actual assists. We also have the pass completion percentage as the size and we have the annual pay as the color. As we can see the players that outperform and underperform their expected assists, are from the lower half of the pay spectrum. On the other hand, a lot of the players that are highly paid have expected assists similar to actual assists; so with these players teams are getting more consistent players

FW Plot

```
ggplot(attack_stats) +
  aes( x = goals, y = xGoal, colour = shot_target_perc) +
  geom_point(shape = "circle", aes(size = pay_year)) +
  scale_color_gradient() +
  stat_smooth(method = "lm", formula = y ~ x, geom = "smooth", color="red") +
  stat_cor(method = "spearman")
```

```
## Warning: The following aesthetics were dropped during statistical transformation: colour
## i This can happen when ggplot fails to infer the correct grouping structure in
## the data.
## i Did you forget to specify a 'group' aesthetic or to convert a numerical
## variable into a factor?
```

```
## Warning: Removed 9 rows containing missing values ('geom_point()').
```



This plot is similar to the midfielders plot, but instead of assists, it plots goals vs expected goals. It also has pay as the size and shot on target percentage as the color. As we see, the darker colors are more concentrated to the bottom left of the graph. This shows the relationship between goals/expected goals and shooting percentage. We can come to the conclusion that: the less accurate a player shoots, the less goals they are expected to score.