TD 1

Axel-Cleris Gailloty

2023-02-10

# 1) Importer la base de donnée

C’est le code qui permet d’importer la base de donnée.

library(readxl)  
fifa22 <- read\_excel("fifa22.xlsx")

# 2) Afficher les noms des colonnes du jeu de données actif

Dans R il y a la fonction colnames() qui permet d’afficher les colonnes d’une base de données.

colnames(fifa22)

## [1] "short\_name" "player\_positions"   
## [3] "overall" "potential"   
## [5] "value\_eur" "wage\_eur"   
## [7] "age" "dob"   
## [9] "height\_cm" "weight\_kg"   
## [11] "club\_team\_id" "club\_name"   
## [13] "league\_name" "league\_level"   
## [15] "club\_position" "club\_jersey\_number"   
## [17] "club\_loaned\_from" "club\_joined"   
## [19] "club\_contract\_valid\_until" "nationality\_id"   
## [21] "nationality\_name" "nation\_team\_id"   
## [23] "nation\_position" "nation\_jersey\_number"   
## [25] "preferred\_foot" "weak\_foot"   
## [27] "skill\_moves" "international\_reputation"   
## [29] "work\_rate" "body\_type"   
## [31] "real\_face" "release\_clause\_eur"   
## [33] "player\_traits" "pace"   
## [35] "shooting" "passing"   
## [37] "dribbling" "defending"   
## [39] "physic" "attacking\_crossing"   
## [41] "attacking\_finishing" "attacking\_heading\_accuracy"   
## [43] "attacking\_short\_passing" "attacking\_volleys"   
## [45] "skill\_dribbling" "skill\_curve"   
## [47] "skill\_fk\_accuracy" "skill\_long\_passing"   
## [49] "skill\_ball\_control" "movement\_acceleration"   
## [51] "movement\_sprint\_speed" "movement\_agility"   
## [53] "movement\_reactions" "movement\_balance"   
## [55] "power\_shot\_power" "power\_jumping"   
## [57] "power\_stamina" "power\_strength"   
## [59] "power\_long\_shots" "mentality\_aggression"   
## [61] "mentality\_interceptions" "mentality\_positioning"   
## [63] "mentality\_vision" "mentality\_penalties"   
## [65] "mentality\_composure" "defending\_marking\_awareness"  
## [67] "defending\_standing\_tackle" "defending\_sliding\_tackle"   
## [69] "goalkeeping\_diving" "goalkeeping\_handling"   
## [71] "goalkeeping\_kicking" "goalkeeping\_positioning"   
## [73] "goalkeeping\_reflexes" "goalkeeping\_speed"

# 3) Afficher les types de données de chaque colonnes

str(fifa22)

## tibble [19,239 × 74] (S3: tbl\_df/tbl/data.frame)  
## $ short\_name : chr [1:19239] "L, Messi" "R, Lewandowski" "Cristiano Ronaldo" "Neymar Jr" ...  
## $ player\_positions : chr [1:19239] "RW, ST, CF" "ST" "ST, LW" "LW, CAM" ...  
## $ overall : num [1:19239] 93 92 91 91 91 91 91 90 90 90 ...  
## $ potential : num [1:19239] 93 92 91 91 91 93 95 90 92 90 ...  
## $ value\_eur : num [1:19239] 7.80e+07 1.20e+08 4.50e+07 1.29e+08 1.26e+08 ...  
## $ wage\_eur : num [1:19239] 320000 270000 270000 270000 350000 130000 230000 86000 250000 240000 ...  
## $ age : num [1:19239] 34 32 36 29 30 28 22 35 29 27 ...  
## $ dob : POSIXct[1:19239], format: "1987-06-24" "1988-08-21" ...  
## $ height\_cm : num [1:19239] 170 185 187 175 181 188 182 193 187 188 ...  
## $ weight\_kg : num [1:19239] 72 81 83 68 70 87 73 93 85 89 ...  
## $ club\_team\_id : num [1:19239] 73 21 11 73 10 240 73 21 241 18 ...  
## $ club\_name : chr [1:19239] "Paris Saint-Germain" "FC Bayern München" "Manchester United" "Paris Saint-Germain" ...  
## $ league\_name : chr [1:19239] "French Ligue 1" "German 1, Bundesliga" "English Premier League" "French Ligue 1" ...  
## $ league\_level : num [1:19239] 1 1 1 1 1 1 1 1 1 1 ...  
## $ club\_position : chr [1:19239] "RW" "ST" "ST" "LW" ...  
## $ club\_jersey\_number : num [1:19239] 30 9 7 10 17 13 7 1 1 10 ...  
## $ club\_loaned\_from : chr [1:19239] NA NA NA NA ...  
## $ club\_joined : POSIXct[1:19239], format: "2021-08-10" "2014-07-01" ...  
## $ club\_contract\_valid\_until : num [1:19239] 2023 2023 2023 2025 2025 ...  
## $ nationality\_id : num [1:19239] 52 37 38 54 7 44 18 21 21 14 ...  
## $ nationality\_name : chr [1:19239] "Argentina" "Poland" "Portugal" "Brazil" ...  
## $ nation\_team\_id : num [1:19239] 1369 1353 1354 NA 1325 ...  
## $ nation\_position : chr [1:19239] "RW" "RS" "ST" NA ...  
## $ nation\_jersey\_number : num [1:19239] 10 9 7 NA 7 NA 10 1 NA 9 ...  
## $ preferred\_foot : chr [1:19239] "Left" "Right" "Right" "Right" ...  
## $ weak\_foot : num [1:19239] 4 4 4 5 5 3 4 4 4 5 ...  
## $ skill\_moves : num [1:19239] 4 4 5 5 4 1 5 1 1 3 ...  
## $ international\_reputation : num [1:19239] 5 5 5 5 4 5 4 5 4 4 ...  
## $ work\_rate : chr [1:19239] "Medium/Low" "High/Medium" "High/Low" "High/Medium" ...  
## $ body\_type : chr [1:19239] "Unique" "Unique" "Unique" "Unique" ...  
## $ real\_face : chr [1:19239] "Yes" "Yes" "Yes" "Yes" ...  
## $ release\_clause\_eur : num [1:19239] 1.44e+08 1.97e+08 8.33e+07 2.39e+08 2.32e+08 ...  
## $ player\_traits : chr [1:19239] "Finesse Shot, Long Shot Taker (AI), Playmaker (AI), Outside Foot Shot, One Club Player, Chip Shot (AI), Technic"| \_\_truncated\_\_ "Solid Player, Finesse Shot, Outside Foot Shot, Chip Shot (AI)" "Power Free-Kick, Flair, Long Shot Taker (AI), Speed Dribbler (AI), Outside Foot Shot" "Injury Prone, Flair, Speed Dribbler (AI), Playmaker (AI), Outside Foot Shot, Technical Dribbler (AI)" ...  
## $ pace : num [1:19239] 85 78 87 91 76 NA 97 NA NA 70 ...  
## $ shooting : num [1:19239] 92 92 94 83 86 NA 88 NA NA 91 ...  
## $ passing : num [1:19239] 91 79 80 86 93 NA 80 NA NA 83 ...  
## $ dribbling : num [1:19239] 95 86 88 94 88 NA 92 NA NA 83 ...  
## $ defending : num [1:19239] 34 44 34 37 64 NA 36 NA NA 47 ...  
## $ physic : num [1:19239] 65 82 75 63 78 NA 77 NA NA 83 ...  
## $ attacking\_crossing : num [1:19239] 85 71 87 85 94 13 78 15 18 80 ...  
## $ attacking\_finishing : num [1:19239] 95 95 95 83 82 11 93 13 14 94 ...  
## $ attacking\_heading\_accuracy : num [1:19239] 70 90 90 63 55 15 72 25 11 86 ...  
## $ attacking\_short\_passing : num [1:19239] 91 85 80 86 94 43 85 60 61 85 ...  
## $ attacking\_volleys : num [1:19239] 88 89 86 86 82 13 83 11 14 88 ...  
## $ skill\_dribbling : num [1:19239] 96 85 88 95 88 12 93 30 21 83 ...  
## $ skill\_curve : num [1:19239] 93 79 81 88 85 13 80 14 18 83 ...  
## $ skill\_fk\_accuracy : num [1:19239] 94 85 84 87 83 14 69 11 12 65 ...  
## $ skill\_long\_passing : num [1:19239] 91 70 77 81 93 40 71 68 63 86 ...  
## $ skill\_ball\_control : num [1:19239] 96 88 88 95 91 30 91 46 30 85 ...  
## $ movement\_acceleration : num [1:19239] 91 77 85 93 76 43 97 54 38 65 ...  
## $ movement\_sprint\_speed : num [1:19239] 80 79 88 89 76 60 97 60 50 74 ...  
## $ movement\_agility : num [1:19239] 91 77 86 96 79 67 92 51 39 71 ...  
## $ movement\_reactions : num [1:19239] 94 93 94 89 91 88 93 87 86 92 ...  
## $ movement\_balance : num [1:19239] 95 82 74 84 78 49 83 35 43 70 ...  
## $ power\_shot\_power : num [1:19239] 86 90 94 80 91 59 86 68 66 91 ...  
## $ power\_jumping : num [1:19239] 68 85 95 64 63 78 78 77 79 79 ...  
## $ power\_stamina : num [1:19239] 72 76 77 81 89 41 88 43 35 83 ...  
## $ power\_strength : num [1:19239] 69 86 77 53 74 78 77 80 78 85 ...  
## $ power\_long\_shots : num [1:19239] 94 87 93 81 91 12 82 16 10 86 ...  
## $ mentality\_aggression : num [1:19239] 44 81 63 63 76 34 62 29 43 80 ...  
## $ mentality\_interceptions : num [1:19239] 40 49 29 37 66 19 38 30 22 44 ...  
## $ mentality\_positioning : num [1:19239] 93 95 95 86 88 11 92 12 11 94 ...  
## $ mentality\_vision : num [1:19239] 95 81 76 90 94 65 82 70 70 87 ...  
## $ mentality\_penalties : num [1:19239] 75 90 88 93 83 11 79 47 25 91 ...  
## $ mentality\_composure : num [1:19239] 96 88 95 93 89 68 88 70 70 91 ...  
## $ defending\_marking\_awareness: num [1:19239] 20 35 24 35 68 27 26 17 25 50 ...  
## $ defending\_standing\_tackle : num [1:19239] 35 42 32 32 65 12 34 10 13 36 ...  
## $ defending\_sliding\_tackle : num [1:19239] 24 19 24 29 53 18 32 11 10 38 ...  
## $ goalkeeping\_diving : num [1:19239] 6 15 7 9 15 87 13 88 88 8 ...  
## $ goalkeeping\_handling : num [1:19239] 11 6 11 9 13 92 5 88 85 10 ...  
## $ goalkeeping\_kicking : num [1:19239] 15 12 15 15 5 78 7 91 88 11 ...  
## $ goalkeeping\_positioning : num [1:19239] 14 8 14 15 10 90 11 89 88 14 ...  
## $ goalkeeping\_reflexes : num [1:19239] 8 10 11 11 13 90 6 88 90 11 ...  
## $ goalkeeping\_speed : num [1:19239] NA NA NA NA NA 50 NA 56 43 NA ...

# 4) Est-ce que R a correctement reconnu le type de chaque colonne ?

Pour la plupart des données, R semble avoir bien reconnu le type des données. Par contre toutes les données qui comportent du texte ont été lus comme chr alors que certaines données texte peuvent représenter des catégories. Par exemple la variable work\_rate représente une catégorie (plusieurs joueurs peuvent être caractérisés pour un même work\_rate, cette donnée n’est pas unique à chaque joueur).

# 5) Convertir les colonnes dans le bon type si besoin

fifa22$nationality\_name <- as.factor(fifa22$nationality\_name)  
fifa22$work\_rate <- as.factor(fifa22$work\_rate)

# 6) Afficher le salaire mensuel moyen des joueurs par ligue. Commenter quelques uns

Dans la base de donnée la variable wage\_eur représente le salaire mensuel des joueurs en euro.

salaire\_moyen\_league <- aggregate(wage\_eur ~ league\_name, data = fifa22, FUN = mean)  
salaire\_moyen\_league

## league\_name wage\_eur  
## 1 Argentina Primera División 6140.9904  
## 2 Australian Hyundai A-League 1638.4477  
## 3 Austrian Football Bundesliga 5206.2121  
## 4 Belgian Jupiler Pro League 7422.7459  
## 5 Campeonato Brasileiro Série A 15263.8889  
## 6 Chilian Campeonato Nacional 2016.4516  
## 7 Chinese Super League 4255.8296  
## 8 Colombian Liga Postobón 951.0471  
## 9 Croatian Prva HNL 591.0714  
## 10 Cypriot First Division 601.7857  
## 11 Czech Republic Gambrinus Liga 739.2405  
## 12 Danish Superliga 4234.2767  
## 13 Ecuadorian Serie A 550.0000  
## 14 English League Championship 11440.9344  
## 15 English League One 2879.3131  
## 16 English League Two 2592.9853  
## 17 English National League 500.0000  
## 18 English Premier League 50847.6994  
## 19 Finnish Veikkausliiga 500.0000  
## 20 French Ligue 1 21462.7383  
## 21 French Ligue 2 2242.2495  
## 22 German 1, Bundesliga 24407.7132  
## 23 German 2, Bundesliga 6136.1765  
## 24 German 3, Bundesliga 1229.4280  
## 25 Greek Super League 804.4643  
## 26 Holland Eredivisie 5058.0285  
## 27 Hungarian Nemzeti Bajnokság I 617.8571  
## 28 Indian Super League 606.6308  
## 29 Italian Serie A 31004.5293  
## 30 Italian Serie B 5298.1061  
## 31 Japanese J, League Division 1 2664.6010  
## 32 Korean K League 1 2326.1905  
## 33 Liga de Fútbol Profesional Boliviano 516.8449  
## 34 Mexican Liga MX 12220.0820  
## 35 Norwegian Eliteserien 1415.7107  
## 36 Paraguayan Primera División 575.5747  
## 37 Peruvian Primera División 511.0429  
## 38 Polish T-Mobile Ekstraklasa 2034.6774  
## 39 Portuguese Liga ZON SAGRES 6335.8416  
## 40 Rep, Ireland Airtricity League 619.7674  
## 41 Romanian Liga I 2616.7421  
## 42 Russian Premier League 24213.4146  
## 43 Saudi Abdul L, Jameel League 8381.0897  
## 44 Scottish Premiership 6782.5000  
## 45 South African Premier Division 552.6786  
## 46 Spain Primera Division 31128.8310  
## 47 Spanish Segunda División 4734.3200  
## 48 Swedish Allsvenskan 1553.5623  
## 49 Swiss Super League 5498.7037  
## 50 Turkish Süper Lig 11915.1013  
## 51 UAE Arabian Gulf League 580.3571  
## 52 Ukrainian Premier League 716.9643  
## 53 Uruguayan Primera División 540.5303  
## 54 USA Major League Soccer 3255.3178  
## 55 Venezuelan Primera División 500.9317

J’utilise la fonction aggregate, je passe en paramètre la variable cible qui est wage\_eur et la variable qui sert à aggréger. Je précise à R de retrouver ces variables dans le jeu de données fifa22, puis la fonction à utiliser est mean. On peut utiliser median comme fonction aussi

library(dplyr)

##   
## Attachement du package : 'dplyr'

## Les objets suivants sont masqués depuis 'package:stats':  
##   
## filter, lag

## Les objets suivants sont masqués depuis 'package:base':  
##   
## intersect, setdiff, setequal, union

arrange(salaire\_moyen\_league, -wage\_eur)

## league\_name wage\_eur  
## 1 English Premier League 50847.6994  
## 2 Spain Primera Division 31128.8310  
## 3 Italian Serie A 31004.5293  
## 4 German 1, Bundesliga 24407.7132  
## 5 Russian Premier League 24213.4146  
## 6 French Ligue 1 21462.7383  
## 7 Campeonato Brasileiro Série A 15263.8889  
## 8 Mexican Liga MX 12220.0820  
## 9 Turkish Süper Lig 11915.1013  
## 10 English League Championship 11440.9344  
## 11 Saudi Abdul L, Jameel League 8381.0897  
## 12 Belgian Jupiler Pro League 7422.7459  
## 13 Scottish Premiership 6782.5000  
## 14 Portuguese Liga ZON SAGRES 6335.8416  
## 15 Argentina Primera División 6140.9904  
## 16 German 2, Bundesliga 6136.1765  
## 17 Swiss Super League 5498.7037  
## 18 Italian Serie B 5298.1061  
## 19 Austrian Football Bundesliga 5206.2121  
## 20 Holland Eredivisie 5058.0285  
## 21 Spanish Segunda División 4734.3200  
## 22 Chinese Super League 4255.8296  
## 23 Danish Superliga 4234.2767  
## 24 USA Major League Soccer 3255.3178  
## 25 English League One 2879.3131  
## 26 Japanese J, League Division 1 2664.6010  
## 27 Romanian Liga I 2616.7421  
## 28 English League Two 2592.9853  
## 29 Korean K League 1 2326.1905  
## 30 French Ligue 2 2242.2495  
## 31 Polish T-Mobile Ekstraklasa 2034.6774  
## 32 Chilian Campeonato Nacional 2016.4516  
## 33 Australian Hyundai A-League 1638.4477  
## 34 Swedish Allsvenskan 1553.5623  
## 35 Norwegian Eliteserien 1415.7107  
## 36 German 3, Bundesliga 1229.4280  
## 37 Colombian Liga Postobón 951.0471  
## 38 Greek Super League 804.4643  
## 39 Czech Republic Gambrinus Liga 739.2405  
## 40 Ukrainian Premier League 716.9643  
## 41 Rep, Ireland Airtricity League 619.7674  
## 42 Hungarian Nemzeti Bajnokság I 617.8571  
## 43 Indian Super League 606.6308  
## 44 Cypriot First Division 601.7857  
## 45 Croatian Prva HNL 591.0714  
## 46 UAE Arabian Gulf League 580.3571  
## 47 Paraguayan Primera División 575.5747  
## 48 South African Premier Division 552.6786  
## 49 Ecuadorian Serie A 550.0000  
## 50 Uruguayan Primera División 540.5303  
## 51 Liga de Fútbol Profesional Boliviano 516.8449  
## 52 Peruvian Primera División 511.0429  
## 53 Venezuelan Primera División 500.9317  
## 54 English National League 500.0000  
## 55 Finnish Veikkausliiga 500.0000

Pour trier les données par ordre décroissant, on écrit beaucoup de code. J’ai utilisé la librairie dplyr qui contient la fonction arrange qui permet de trier facilement le résultat.

Cette fonction prend en argument un tableau de donnée, et la colonne qui sert de tri. Dans notre cas on veut que wage\_eur soit trié par ordre décroissant d’où le -wage\_eur.

En moyenne les joueurs du English Premier League gagnent 50847€ par mois. Les joueurs de la Ligue 1 en France gagnent en moyenne 21462€.

# 7) Afficher le potentiel moyen des joueurs par ligue

aggregate(potential ~ league\_name, data = fifa22, FUN = mean)

## league\_name potential  
## 1 Argentina Primera División 72.74553  
## 2 Australian Hyundai A-League 67.22383  
## 3 Austrian Football Bundesliga 70.70606  
## 4 Belgian Jupiler Pro League 73.25410  
## 5 Campeonato Brasileiro Série A 71.78611  
## 6 Chilian Campeonato Nacional 69.74194  
## 7 Chinese Super League 62.94170  
## 8 Colombian Liga Postobón 69.70681  
## 9 Croatian Prva HNL 74.64286  
## 10 Cypriot First Division 69.82143  
## 11 Czech Republic Gambrinus Liga 75.27848  
## 12 Danish Superliga 69.98742  
## 13 Ecuadorian Serie A 69.50877  
## 14 English League Championship 72.62343  
## 15 English League One 68.96326  
## 16 English League Two 66.48287  
## 17 English National League 64.51852  
## 18 English Premier League 78.99693  
## 19 Finnish Veikkausliiga 68.12000  
## 20 French Ligue 1 76.69671  
## 21 French Ligue 2 69.96219  
## 22 German 1, Bundesliga 77.14519  
## 23 German 2, Bundesliga 71.25882  
## 24 German 3, Bundesliga 67.97048  
## 25 Greek Super League 75.00893  
## 26 Holland Eredivisie 73.59959  
## 27 Hungarian Nemzeti Bajnokság I 72.60714  
## 28 Indian Super League 61.63082  
## 29 Italian Serie A 77.19361  
## 30 Italian Serie B 71.95455  
## 31 Japanese J, League Division 1 68.09847  
## 32 Korean K League 1 66.52976  
## 33 Liga de Fútbol Profesional Boliviano 67.77005  
## 34 Mexican Liga MX 72.20492  
## 35 Norwegian Eliteserien 68.99501  
## 36 Paraguayan Primera División 69.23563  
## 37 Peruvian Primera División 68.88957  
## 38 Polish T-Mobile Ekstraklasa 68.18952  
## 39 Portuguese Liga ZON SAGRES 74.56634  
## 40 Rep, Ireland Airtricity League 63.31008  
## 41 Romanian Liga I 67.32353  
## 42 Russian Premier League 76.34146  
## 43 Saudi Abdul L, Jameel League 66.80769  
## 44 Scottish Premiership 69.62188  
## 45 South African Premier Division 69.57143  
## 46 Spain Primera Division 78.69510  
## 47 Spanish Segunda División 72.79040  
## 48 Swedish Allsvenskan 67.35369  
## 49 Swiss Super League 71.68519  
## 50 Turkish Süper Lig 71.78269  
## 51 UAE Arabian Gulf League 69.75000  
## 52 Ukrainian Premier League 77.57143  
## 53 Uruguayan Primera División 69.27273  
## 54 USA Major League Soccer 70.61479  
## 55 Venezuelan Primera División 68.15528

# 8) Afficher le salaire moyen des joueurs par niveau de réputation

aggregate(wage\_eur ~ international\_reputation, data = fifa22, FUN = mean)

## international\_reputation wage\_eur  
## 1 1 5948.719  
## 2 2 34924.396  
## 3 3 73034.629  
## 4 4 153368.421  
## 5 5 191500.000

# 9) Le salaire moyen des joueurs par ligue et par réputation internationale

aggregate(wage\_eur ~ league\_name + international\_reputation, data = fifa22, FUN = mean)

## league\_name international\_reputation wage\_eur  
## 1 Argentina Primera División 1 5883.0703  
## 2 Australian Hyundai A-League 1 1637.1377  
## 3 Austrian Football Bundesliga 1 5119.4190  
## 4 Belgian Jupiler Pro League 1 7104.4492  
## 5 Campeonato Brasileiro Série A 1 15263.8889  
## 6 Chilian Campeonato Nacional 1 2016.4516  
## 7 Chinese Super League 1 3761.8056  
## 8 Colombian Liga Postobón 1 951.0471  
## 9 Croatian Prva HNL 1 573.5849  
## 10 Cypriot First Division 1 600.0000  
## 11 Czech Republic Gambrinus Liga 1 732.4675  
## 12 Danish Superliga 1 4235.0158  
## 13 Ecuadorian Serie A 1 550.0000  
## 14 English League Championship 1 11311.3861  
## 15 English League One 1 2870.9135  
## 16 English League Two 1 2593.9542  
## 17 English National League 1 500.0000  
## 18 English Premier League 1 27107.5294  
## 19 Finnish Veikkausliiga 1 500.0000  
## 20 French Ligue 1 1 11806.2954  
## 21 French Ligue 2 1 2195.6480  
## 22 German 1, Bundesliga 1 12885.1389  
## 23 German 2, Bundesliga 1 5948.6948  
## 24 German 3, Bundesliga 1 1225.1391  
## 25 Greek Super League 1 758.5859  
## 26 Holland Eredivisie 1 4386.9612  
## 27 Hungarian Nemzeti Bajnokság I 1 612.0000  
## 28 Indian Super League 1 606.6308  
## 29 Italian Serie A 1 17530.3161  
## 30 Italian Serie B 1 4523.5537  
## 31 Japanese J, League Division 1 1 2573.1643  
## 32 Korean K League 1 1 2248.7654  
## 33 Liga de Fútbol Profesional Boliviano 1 516.8449  
## 34 Mexican Liga MX 1 12239.5062  
## 35 Norwegian Eliteserien 1 1415.7107  
## 36 Paraguayan Primera División 1 575.5747  
## 37 Peruvian Primera División 1 511.0429  
## 38 Polish T-Mobile Ekstraklasa 1 2020.6897  
## 39 Portuguese Liga ZON SAGRES 1 5616.3180  
## 40 Rep, Ireland Airtricity League 1 619.7674  
## 41 Romanian Liga I 1 2616.7421  
## 42 Russian Premier League 1 21711.2676  
## 43 Saudi Abdul L, Jameel League 1 6768.9773  
## 44 Scottish Premiership 1 6209.6154  
## 45 South African Premier Division 1 553.6364  
## 46 Spain Primera Division 1 18324.0042  
## 47 Spanish Segunda División 1 4627.5947  
## 48 Swedish Allsvenskan 1 1553.5623  
## 49 Swiss Super League 1 5145.9302  
## 50 Turkish Süper Lig 1 9923.1915  
## 51 UAE Arabian Gulf League 1 580.3571  
## 52 Ukrainian Premier League 1 681.0000  
## 53 Uruguayan Primera División 1 540.5303  
## 54 USA Major League Soccer 1 2965.3103  
## 55 Venezuelan Primera División 1 500.9317  
## 56 Argentina Primera División 2 12153.8462  
## 57 Australian Hyundai A-League 2 2000.0000  
## 58 Austrian Football Bundesliga 2 14666.6667  
## 59 Belgian Jupiler Pro League 2 14384.6154  
## 60 Chinese Super League 2 17727.2727  
## 61 Croatian Prva HNL 2 900.0000  
## 62 Cypriot First Division 2 650.0000  
## 63 Czech Republic Gambrinus Liga 2 1000.0000  
## 64 Danish Superliga 2 4000.0000  
## 65 English League Championship 2 20666.6667  
## 66 English League One 2 5500.0000  
## 67 English League Two 2 2000.0000  
## 68 English Premier League 2 68751.6779  
## 69 French Ligue 1 2 34359.3750  
## 70 French Ligue 2 2 4181.8182  
## 71 German 1, Bundesliga 2 36575.1634  
## 72 German 2, Bundesliga 2 13916.6667  
## 73 German 3, Bundesliga 2 2000.0000  
## 74 Greek Super League 2 1200.0000  
## 75 Holland Eredivisie 2 13772.7273  
## 76 Hungarian Nemzeti Bajnokság I 2 666.6667  
## 77 Italian Serie A 2 42734.6939  
## 78 Italian Serie B 2 14750.0000  
## 79 Japanese J, League Division 1 2 5700.0000  
## 80 Korean K League 1 2 4416.6667  
## 81 Mexican Liga MX 2 7500.0000  
## 82 Polish T-Mobile Ekstraklasa 2 3500.0000  
## 83 Portuguese Liga ZON SAGRES 2 20176.4706  
## 84 Russian Premier League 2 38750.0000  
## 85 Saudi Abdul L, Jameel League 2 31960.0000  
## 86 Scottish Premiership 2 29125.0000  
## 87 South African Premier Division 2 500.0000  
## 88 Spain Primera Division 2 43931.3725  
## 89 Spanish Segunda División 2 8333.3333  
## 90 Swiss Super League 2 13090.9091  
## 91 Turkish Süper Lig 2 20596.4912  
## 92 Ukrainian Premier League 2 1016.6667  
## 93 USA Major League Soccer 2 6818.1818  
## 94 Argentina Primera División 3 12000.0000  
## 95 Belgian Jupiler Pro League 3 27333.3333  
## 96 Chinese Super League 3 26000.0000  
## 97 English League Championship 3 20000.0000  
## 98 English Premier League 3 127306.4516  
## 99 French Ligue 1 3 65172.4138  
## 100 French Ligue 2 3 5000.0000  
## 101 German 1, Bundesliga 3 67428.5714  
## 102 Greek Super League 3 1000.0000  
## 103 Holland Eredivisie 3 24200.0000  
## 104 Italian Serie A 3 72672.1311  
## 105 Italian Serie B 3 8000.0000  
## 106 Japanese J, League Division 1 3 3900.0000  
## 107 Polish T-Mobile Ekstraklasa 3 6000.0000  
## 108 Portuguese Liga ZON SAGRES 3 17555.5556  
## 109 Russian Premier League 3 44666.6667  
## 110 Saudi Abdul L, Jameel League 3 48333.3333  
## 111 Spain Primera Division 3 82810.8108  
## 112 Swiss Super League 3 13000.0000  
## 113 Turkish Süper Lig 3 39785.7143  
## 114 USA Major League Soccer 3 9727.2727  
## 115 English Premier League 4 215000.0000  
## 116 French Ligue 1 4 126000.0000  
## 117 German 1, Bundesliga 4 121250.0000  
## 118 Holland Eredivisie 4 29000.0000  
## 119 Italian Serie A 4 98166.6667  
## 120 Italian Serie B 4 18000.0000  
## 121 Japanese J, League Division 1 4 10000.0000  
## 122 Portuguese Liga ZON SAGRES 4 14000.0000  
## 123 Spain Primera Division 4 210266.6667  
## 124 Turkish Süper Lig 4 37500.0000  
## 125 USA Major League Soccer 4 14000.0000  
## 126 English Premier League 5 270000.0000  
## 127 French Ligue 1 5 295000.0000  
## 128 German 1, Bundesliga 5 178000.0000  
## 129 Italian Serie A 5 51000.0000  
## 130 Spain Primera Division 5 132500.0000

# 10) Afficher les statistiques générales de toutes les données numériques

On sélectionne les colonnes numériques qui sont dans le jeu de données

Nous utilisons la fonction select de la librairie dplyr pour sélectionner les colonnes par types.

num\_cols <- select(fifa22, where(is.numeric))  
summary(num\_cols)

## overall potential value\_eur wage\_eur   
## Min. :47.00 Min. :49.00 Min. :9.00e+03 Min. : 500   
## 1st Qu.:61.00 1st Qu.:67.00 1st Qu.:4.75e+05 1st Qu.: 1000   
## Median :66.00 Median :71.00 Median :9.75e+05 Median : 3000   
## Mean :65.77 Mean :71.08 Mean :2.85e+06 Mean : 9018   
## 3rd Qu.:70.00 3rd Qu.:75.00 3rd Qu.:2.00e+06 3rd Qu.: 8000   
## Max. :93.00 Max. :95.00 Max. :1.94e+08 Max. :350000   
## NA's :74 NA's :61   
## age height\_cm weight\_kg club\_team\_id   
## Min. :16.00 Min. :155.0 Min. : 49.00 Min. : 1   
## 1st Qu.:21.00 1st Qu.:176.0 1st Qu.: 70.00 1st Qu.: 479   
## Median :25.00 Median :181.0 Median : 75.00 Median : 1938   
## Mean :25.21 Mean :181.3 Mean : 74.94 Mean : 50581   
## 3rd Qu.:29.00 3rd Qu.:186.0 3rd Qu.: 80.00 3rd Qu.:111139   
## Max. :54.00 Max. :206.0 Max. :110.00 Max. :115820   
## NA's :61   
## league\_level club\_jersey\_number club\_contract\_valid\_until nationality\_id   
## Min. :1.000 Min. : 1.00 Min. :2021 Min. : 1.0   
## 1st Qu.:1.000 1st Qu.: 9.00 1st Qu.:2022 1st Qu.: 21.0   
## Median :1.000 Median :18.00 Median :2022 Median : 45.0   
## Mean :1.354 Mean :20.95 Mean :2023 Mean : 58.6   
## 3rd Qu.:1.000 3rd Qu.:27.00 3rd Qu.:2024 3rd Qu.: 60.0   
## Max. :5.000 Max. :99.00 Max. :2031 Max. :219.0   
## NA's :61 NA's :61 NA's :61   
## nation\_team\_id nation\_jersey\_number weak\_foot skill\_moves   
## Min. : 1318 Min. : 1.00 Min. :1.000 Min. :1.000   
## 1st Qu.: 1338 1st Qu.: 7.00 1st Qu.:3.000 1st Qu.:2.000   
## Median : 1357 Median :12.00 Median :3.000 Median :2.000   
## Mean : 14481 Mean :12.57 Mean :2.946 Mean :2.352   
## 3rd Qu.: 1386 3rd Qu.:19.00 3rd Qu.:3.000 3rd Qu.:3.000   
## Max. :111473 Max. :28.00 Max. :5.000 Max. :5.000   
## NA's :18480 NA's :18480   
## international\_reputation release\_clause\_eur pace shooting   
## Min. :1.000 Min. : 16000 Min. :28.00 Min. :18.00   
## 1st Qu.:1.000 1st Qu.: 806000 1st Qu.:62.00 1st Qu.:42.00   
## Median :1.000 Median : 1600000 Median :69.00 Median :54.00   
## Mean :1.094 Mean : 5374044 Mean :68.21 Mean :52.35   
## 3rd Qu.:1.000 3rd Qu.: 3700000 3rd Qu.:76.00 3rd Qu.:63.00   
## Max. :5.000 Max. :373500000 Max. :97.00 Max. :94.00   
## NA's :1176 NA's :2132 NA's :2132   
## passing dribbling defending physic   
## Min. :25.00 Min. :27.00 Min. :14.0 Min. :29.00   
## 1st Qu.:51.00 1st Qu.:57.00 1st Qu.:37.0 1st Qu.:59.00   
## Median :58.00 Median :64.00 Median :56.0 Median :66.00   
## Mean :57.31 Mean :62.56 Mean :51.7 Mean :64.82   
## 3rd Qu.:64.00 3rd Qu.:69.00 3rd Qu.:64.0 3rd Qu.:72.00   
## Max. :93.00 Max. :95.00 Max. :91.0 Max. :90.00   
## NA's :2132 NA's :2132 NA's :2132 NA's :2132   
## attacking\_crossing attacking\_finishing attacking\_heading\_accuracy  
## Min. : 6.00 Min. : 2.00 Min. : 5.00   
## 1st Qu.:38.00 1st Qu.:30.00 1st Qu.:44.00   
## Median :54.00 Median :50.00 Median :55.00   
## Mean :49.58 Mean :45.89 Mean :51.78   
## 3rd Qu.:63.00 3rd Qu.:62.00 3rd Qu.:64.00   
## Max. :94.00 Max. :95.00 Max. :93.00   
##   
## attacking\_short\_passing attacking\_volleys skill\_dribbling skill\_curve   
## Min. : 7.00 Min. : 3.00 Min. : 4.00 Min. : 6.00   
## 1st Qu.:54.00 1st Qu.:30.00 1st Qu.:50.00 1st Qu.:35.00   
## Median :62.00 Median :43.00 Median :61.00 Median :49.00   
## Mean :58.87 Mean :42.46 Mean :55.66 Mean :47.27   
## 3rd Qu.:68.00 3rd Qu.:56.00 3rd Qu.:68.00 3rd Qu.:61.00   
## Max. :94.00 Max. :90.00 Max. :96.00 Max. :94.00   
##   
## skill\_fk\_accuracy skill\_long\_passing skill\_ball\_control movement\_acceleration  
## Min. : 4.00 Min. : 9.00 Min. : 8.00 Min. :14.00   
## 1st Qu.:31.00 1st Qu.:44.00 1st Qu.:55.00 1st Qu.:57.00   
## Median :41.00 Median :56.00 Median :63.00 Median :67.00   
## Mean :42.25 Mean :53.07 Mean :58.47 Mean :64.65   
## 3rd Qu.:55.00 3rd Qu.:64.00 3rd Qu.:69.00 3rd Qu.:75.00   
## Max. :94.00 Max. :93.00 Max. :96.00 Max. :97.00   
##   
## movement\_sprint\_speed movement\_agility movement\_reactions movement\_balance  
## Min. :15.00 Min. :18.0 Min. :25.00 Min. :15.00   
## 1st Qu.:58.00 1st Qu.:55.0 1st Qu.:56.00 1st Qu.:56.00   
## Median :68.00 Median :66.0 Median :62.00 Median :66.00   
## Mean :64.71 Mean :63.5 Mean :61.45 Mean :64.07   
## 3rd Qu.:75.00 3rd Qu.:74.0 3rd Qu.:67.00 3rd Qu.:74.00   
## Max. :97.00 Max. :96.0 Max. :94.00 Max. :96.00   
##   
## power\_shot\_power power\_jumping power\_stamina power\_strength   
## Min. :20.00 Min. :22.00 Min. :12.00 Min. :19.00   
## 1st Qu.:48.00 1st Qu.:57.00 1st Qu.:56.00 1st Qu.:57.00   
## Median :59.00 Median :65.00 Median :66.00 Median :66.00   
## Mean :57.78 Mean :64.81 Mean :63.08 Mean :65.01   
## 3rd Qu.:68.00 3rd Qu.:73.00 3rd Qu.:74.00 3rd Qu.:74.00   
## Max. :95.00 Max. :95.00 Max. :97.00 Max. :97.00   
##   
## power\_long\_shots mentality\_aggression mentality\_interceptions  
## Min. : 4.00 Min. :10.00 Min. : 3.00   
## 1st Qu.:32.00 1st Qu.:44.00 1st Qu.:26.00   
## Median :51.00 Median :58.00 Median :53.00   
## Mean :46.64 Mean :55.54 Mean :46.61   
## 3rd Qu.:62.00 3rd Qu.:68.00 3rd Qu.:64.00   
## Max. :94.00 Max. :95.00 Max. :91.00   
##   
## mentality\_positioning mentality\_vision mentality\_penalties mentality\_composure  
## Min. : 2.00 Min. :10.00 Min. : 7.00 Min. :12.00   
## 1st Qu.:40.00 1st Qu.:45.00 1st Qu.:38.00 1st Qu.:50.00   
## Median :56.00 Median :55.00 Median :49.00 Median :59.00   
## Mean :50.33 Mean :53.96 Mean :47.86 Mean :57.93   
## 3rd Qu.:64.00 3rd Qu.:64.00 3rd Qu.:60.00 3rd Qu.:66.00   
## Max. :96.00 Max. :95.00 Max. :93.00 Max. :96.00   
##   
## defending\_marking\_awareness defending\_standing\_tackle defending\_sliding\_tackle  
## Min. : 4.0 Min. : 5.00 Min. : 5.00   
## 1st Qu.:29.0 1st Qu.:28.00 1st Qu.:25.00   
## Median :52.0 Median :56.00 Median :53.00   
## Mean :46.6 Mean :48.05 Mean :45.91   
## 3rd Qu.:63.0 3rd Qu.:65.00 3rd Qu.:63.00   
## Max. :93.0 Max. :93.00 Max. :92.00   
##   
## goalkeeping\_diving goalkeeping\_handling goalkeeping\_kicking  
## Min. : 2.00 Min. : 2.00 Min. : 2.00   
## 1st Qu.: 8.00 1st Qu.: 8.00 1st Qu.: 8.00   
## Median :11.00 Median :11.00 Median :11.00   
## Mean :16.41 Mean :16.19 Mean :16.06   
## 3rd Qu.:14.00 3rd Qu.:14.00 3rd Qu.:14.00   
## Max. :91.00 Max. :92.00 Max. :93.00   
##   
## goalkeeping\_positioning goalkeeping\_reflexes goalkeeping\_speed  
## Min. : 2.00 Min. : 2.00 Min. :15.00   
## 1st Qu.: 8.00 1st Qu.: 8.00 1st Qu.:27.00   
## Median :11.00 Median :11.00 Median :36.00   
## Mean :16.23 Mean :16.49 Mean :36.44   
## 3rd Qu.:14.00 3rd Qu.:14.00 3rd Qu.:45.00   
## Max. :92.00 Max. :90.00 Max. :65.00   
## NA's :17107

La fonction summary() est utilisée pour résumer un jeu de donnée. Si les données sont numériques elle affiche le minimum (min), le premier quartile, la médiane, la moyenne, le 3e quartiel et le maximum de la variable. Si la colonne contient des données manquantes, la fonction affiche le nombre de données manquantes via la ligne NA’s.

# 11) Afficher la matrice de corrélation entre les variables suivantes :

cor(fifa22$height\_cm, fifa22$weight\_kg)

## [1] 0.765465

La corrélation entre le poids d’un joueur et sa taille de de 0.765.

cor(num\_cols[, c("height\_cm", "weight\_kg", "skill\_dribbling", "skill\_ball\_control")])

## height\_cm weight\_kg skill\_dribbling skill\_ball\_control  
## height\_cm 1.0000000 0.7654650 -0.4772269 -0.4004402  
## weight\_kg 0.7654650 1.0000000 -0.3973634 -0.3238045  
## skill\_dribbling -0.4772269 -0.3973634 1.0000000 0.9435056  
## skill\_ball\_control -0.4004402 -0.3238045 0.9435056 1.0000000

La corrélation entre deux variable va de -1 et 1. SI la corrélation est de 1, c’est que les deux variables vont parfaitement dans le même sens. Si le coefficient de corrélation linéaire est négatif c’est que les deux variables évoluent dans un sens opposé. Une augmentation dans x induit une baisse de y.

La variable skill\_dribbling s’oppose à la variable weight\_kg (elles évoluent dans un sens inverse) car leur corrélation est de -0.3973634.

# PARTIE 3 : Réaliser une ACP sur le jeu de données

Réaliser une ACP sur les 100 premiers joueurs en prenant comme variables quantitatives actives les caractéristiques des joueurs suivant :

* “height\_cm”, “weight\_kg”, “skill\_dribbling”, “skill\_ball\_control”, “movement\_acceleration”, “movement\_agility”, “power\_shot\_power”, “power\_jumping”, “power\_stamina”, “power\_strength”, “mentality\_aggression”, “mentality\_interceptions”, “defending\_marking\_awareness”, “defending\_standing\_tackle”, “defending\_sliding\_tackle”, “goalkeeping\_diving”, “goalkeeping\_handling”, “goalkeeping\_kicking”, “goalkeeping\_speed”

comme variables quantitatives illustratives

* “overall”, “potential”, “value\_eur”, “pace”, “shooting”, “passing”, “dribbling”, “defending”,“physic” et comme variable qualitative illustrative
* “body\_type”.

## Séparer les variables dans des objets R pour une meilleure visibilité.

Nous allons mettre les noms des colonnes dans des variables R pour mieux comprendre le code.

### A quoi servent les variables quanti actives ?

Les variables quantitatives illustratives servent à calculer les composantes de l’ACP. Elles entrent dans le calcul de la matrice de variance-covariance.

var\_quanti\_actives <- c("height\_cm", "weight\_kg", "skill\_dribbling", "skill\_ball\_control",  
"movement\_acceleration", "movement\_agility", "power\_shot\_power",  
"power\_jumping", "power\_stamina", "power\_strength", "mentality\_aggression",  
"mentality\_interceptions", "defending\_marking\_awareness",  
"defending\_standing\_tackle", "defending\_sliding\_tackle", "goalkeeping\_diving",  
"goalkeeping\_handling", "goalkeeping\_kicking", "goalkeeping\_speed")

### A quoi servent les variables quantitatives illustratives ?

Elles nous servent à mieux interpréter les résultats. Elles n’entrent pas dans le calcul des composantes.

var\_quanti\_illustratives <- c("overall", "potential", "value\_eur", "pace", "shooting", "passing", "dribbling","defending","physic")

### A quoi servent les variables qualitatives illustratives

Comme les variables quantitatives illustratives, elles aident à l’interprétation des composantes. Elles sont qualitatives (modalités discrètes) : par club, par league, par nationalité …

l’ACP ne se base que sur des données numériques.

var\_quali\_illustratives <- c("body\_type")

# Filtrer le jeu de données

Réaliser une ACP sur les 100 premiers joueurs en prenant comme variables quantitatives actives les caractéristiques des joueurs suivant :

## Prendre les 100 premiers joueurs

fifa\_100 <- head(fifa22, 100)

## Ne prendre que les colonnes (variables) qui nous intéressent

variables\_acp <- c(var\_quanti\_actives, var\_quanti\_illustratives, var\_quali\_illustratives)

fifa\_100 <- fifa\_100[, variables\_acp]