CS5801 Coursework Template Proforma

2144810

09 Enero, 2022

0. Instructions

1. Remove the (italicised) guidance text but keep the section headings.
2. Add as many chunks of R code as required.
3. Add descriptions of your analysis plans and explanations of your code and findings. Please be detailed and where you have made choices explain the rationale for them.
4. Write your report using RMarkdown. For guidance see a helpful blog (https://www.dataquest.io/blog/r-markdown-guide-cheatsheet/#tve-jump-17333da0719) or use the R Markdown cheatsheet which can be accessed from within RStudio by selecting `Help > Cheatsheets > R Markdown Cheat Sheet`.
5. Your report should be clearly and professionally presented with appropriate use of cited external sources. (5 marks)
6. It should also be easy to understand, with well-documented code following the principles of literate programming. (5 marks)

```
# Add code here to load any required libraries with `library()`.
# We suggest you use `install.package()` for any required packages externally to this document
# since installation only need be done once.
library(corrplot)
```

```
## corrplot 0.91 loaded
```

```
library(ggpubr)
```

```
## Loading required package: ggplot2
```

```
options(scipen = 999)
```

1. Organise and clean the data

1.1 Subset the data into the specific dataset allocated

A description of the data set provided, its contents and which subset you should select is documented in the assessment brief at ???.pdf
Use R code to correctly select the subset of data allocated. (5 marks)

```
# Assign your student id into the variable SID, for example:
SID <- 2144810                  # This is an example, replace 2101234 with your actual ID
SIDoffset <- (SID %% 25) + 1    # Your SID mod 25 + 1

load("CS5801_football_analysis.Rda")
```

```
## Warning in load("CS5801_football_analysis.Rda"): strings not representable in
## native encoding will be translated to UTF-8
```

```
# Now subset the football data set
# Pick every 25th observation starting from your offset
# Put into your data frame named mydf (you can rename it)
mydf <- football.analysis[seq(from=SIDoffset,to=nrow(football.analysis),by=25),]
write.csv(mydf, "data.csv")
```

1.2 Data quality analysis

Provide a description of a comprehensive plan to assess the quality of the data, and document your findings. Include all variables/columns (5 marks) from the data set and provide a full implementation (5 marks). NB even if no data quality issues are identified you should still check and report.

Here is my following plan to check the quality of the data for each column:

1. `sofifa_id` : First, it must be checked if the column is a string. Then it must be checked if each value is unique since it is an ID. It must be checked if missing values (NA) exist.
2. For columns `wage_eur` , `age` , `height_cm` and `weight_kg` , it must be checked that the columns are numerical. Then it must be checked that values are not below 0. It also must be checked for outliers and if they are reasonable or not. It must be checked if missing values (NA) exist.
3. For columns `potential` , `pace` , `shooting` , `passing` , `dribbling` , `defending` , `physic` , `power_strength` and `power_long_shots` , it must be checked that the column is numerical. Then it must be checked that the maximum value is not above 100 and that the minimum value is not below 0. It must be checked if missing values (NA) exist.
4. `club_name` : First, it must be checked that the column is a string. This column won't be used for analysis so not further studying will be done.
5. `preferred_foot` : First, it must be checked that the column is categorical (factor). This column must be constraint to two unique values: `Right` and `Left` . It exist the possibility that values like `left` or `LEft` could be encountered and if they do, they need to be changed to the 2 unique values. It must be checked if missing values (NA) exist.

6. `high.wage.ind` : First, it must be checked that the column is categorical (factor). This column must be constraint to two unique values: `1` and `0` . It must be checked if missing values (NA) exist. It also must be checked that it has a value of `1` if weekly wage is above 8000 and `0` otherwise.

Note: a) NA refers to Not Available. b) There are many methods for calculating outliers. However, for this lab, 3 will be used. The first is by using a boxplot. Where points are visualized, outliers exist. The second method is by using the standard deviation (std). Values that are beyond 3 stds from the mean are considered outliers. The final method is using interquartile range (IQR). Values that are beyond quartile 3 plus 1.5 times the IQR and values that are below quartile 1 minus 1.5 times the IQR are considered outliers.

Checking for quality issues:

a. Checking data type of each column:

```
str(mydf)
```

```
## 'data.frame':    514 obs. of  17 variables:
##  $ sofifa_id       : int  155862 218667 189513 146536 212194 186345 208574 176635 210665 179846 ...
##  $ potential       : int  89 88 85 84 88 83 83 82 82 81 ...
##  $ wage_eur        : num  300000 230000 59000 43000 75000 71000 45000 100000 65000 87000 ...
##  $ age             : int  34 25 31 34 24 29 27 31 28 33 ...
##  $ height_cm       : int  184 173 182 172 185 173 184 180 188 189 ...
##  $ weight_kg       : int  82 64 74 60 83 71 82 71 87 90 ...
##  $ club_name       : chr  "Real Madrid CF" "Manchester City" "Villarreal CF" "Sevilla FC" ...
##  $ preferred_foot  : chr  "Right" "Left" "Right" "Right" ...
##  $ pace            : int  71 81 50 87 81 73 88 64 66 45 ...
##  $ shooting        : int  70 77 83 69 78 60 76 69 66 74 ...
##  $ passing         : int  76 83 90 79 82 81 79 85 73 77 ...
##  $ dribbling       : int  73 92 81 83 87 77 83 83 72 77 ...
##  $ defending       : int  88 51 71 80 51 79 68 24 81 79 ...
##  $ physic          : int  85 67 69 58 66 72 80 55 80 78 ...
##  $ power_strength  : int  85 54 70 39 71 64 77 56 79 85 ...
##  $ power_long_shots: int  64 76 88 70 80 70 78 67 69 78 ...
##  $ high.wage.ind   : int  1 1 1 1 1 1 1 1 1 1 ...
```

- `power_long_shots` , `power_strength` , `physic` , `defending` , `dribbling` , `passing` , `shooting` , `pace` and `potential` have the right data type. They are integers (numerical).
- `high.wage.ind` is integer (numerical) and therefore it needs to be changed to categorical (factor).
- `preferred_foot` is a string (character) and needs to be changed to categorical (factor).
- `sofifa_id` is a integer (numerical) and needs to be changed to string (character).
- `wage_eur` is numerical therefore it is correct.
- `age` , `height_cm` , `weight_kg` are integer (numerical) and therefore they are correct.
- `club_name` is a string (character) and therefore it is correct.

b. Checking for missing values (NA) for each column:

```
# df_col_names is a list of all the column names of our dataframe.
df_col_names <- c(colnames(mydf))
# iteration over each column to find missing values.
for (col_name in df_col_names){
  message = paste("Column", col_name, "has", sum(is.na(mydf$col_name)), "missing values (NA)")
  print(message)
}
```

```
## [1] "Column sofifa_id has 0 missing values (NA)"
## [1] "Column potential has 0 missing values (NA)"
## [1] "Column wage_eur has 0 missing values (NA)"
## [1] "Column age has 0 missing values (NA)"
## [1] "Column height_cm has 0 missing values (NA)"
## [1] "Column weight_kg has 0 missing values (NA)"
## [1] "Column club_name has 0 missing values (NA)"
## [1] "Column preferred_foot has 0 missing values (NA)"
## [1] "Column pace has 0 missing values (NA)"
## [1] "Column shooting has 0 missing values (NA)"
## [1] "Column passing has 0 missing values (NA)"
## [1] "Column dribbling has 0 missing values (NA)"
## [1] "Column defending has 0 missing values (NA)"
## [1] "Column physic has 0 missing values (NA)"
## [1] "Column power_strength has 0 missing values (NA)"
## [1] "Column power_long_shots has 0 missing values (NA)"
## [1] "Column high.wage.ind has 0 missing values (NA)"
```

No missing values were found in any column.

c. Check maximum and minimum values for numerical columns:

```
summary(mydf)
```

```
##    sofifa_id        potential        wage_eur           age
##  Min.   :143809   Min.   :54.00   Min.   :     4   Min.   :16.00
##  1st Qu.:211561   1st Qu.:67.00   1st Qu.:  1000   1st Qu.:22.00
##  Median :232602   Median :72.00   Median :  3000   Median :25.00
##  Mean   :227354   Mean   :71.67   Mean   : 10821   Mean   :25.22
##  3rd Qu.:247061   3rd Qu.:76.00   3rd Qu.: 10000   3rd Qu.:28.00
##  Max.   :258923   Max.   :89.00   Max.   :300000   Max.   :77.00
##    height_cm       weight_kg       club_name         preferred_foot
##  Min.   :160.0   Min.   : 58.00   Length:514         Length:514
##  1st Qu.:175.0   1st Qu.: 70.00   Class :character   Class :character
##  Median :180.0   Median : 75.00   Mode  :character   Mode  :character
##  Mean   :180.1   Mean   : 74.54
##  3rd Qu.:185.0   3rd Qu.: 79.00
##  Max.   :220.0   Max.   :188.00
##       pace            shooting        passing        dribbling
##  Min.   :-66.00   Min.   :21.00   Min.   :26.00   Min.   :-66.0
##  1st Qu.: 62.00   1st Qu.:42.00   1st Qu.:51.00   1st Qu.: 57.0
##  Median : 69.00   Median :55.00   Median :58.00   Median : 64.0
##  Mean   : 68.13   Mean   :53.04   Mean   :57.81   Mean   : 62.8
##  3rd Qu.: 76.00   3rd Qu.:65.00   3rd Qu.:65.00   3rd Qu.: 70.0
##  Max.   : 94.00   Max.   :83.00   Max.   :90.00   Max.   : 92.0
##    defending         physic      power_strength  power_long_shots
##  Min.   :18.00   Min.   :35.00   Min.   :29.00   Min.   :16.00
##  1st Qu.:36.00   1st Qu.:58.00   1st Qu.:57.00   1st Qu.:40.00
##  Median :57.00   Median :66.00   Median :67.00   Median :54.00
##  Mean   :51.53   Mean   :64.61   Mean   :65.31   Mean   :51.83
##  3rd Qu.:65.00   3rd Qu.:72.00   3rd Qu.:75.00   3rd Qu.:64.00
##  Max.   :88.00   Max.   :85.00   Max.   :93.00   Max.   :88.00
##  high.wage.ind
##  Min.   :0.0000
##  1st Qu.:0.0000
##  Median :0.0000
##  Mean   :0.2899
##  3rd Qu.:1.0000
##  Max.   :1.0000
```

- Column `wage_eur` has minimum value of 4 and maximum value of 300000. All values are positive. 4 euros for weekly salary is very low. Low values in this column need further investigation. Also, I need to look for outliers.
- Column `age` has a minimum value of 16 and maximum value of 77. All values are positive. Now, I need to look for outliers.
- Column `height_cm` has a minimum value of 160 and a max value of 220. All values are positive. Now, I need to look for outliers.
- Column `weight_kg` has a minimum value of 56 and a maximum value of 188. All values are positive. Now, I need to look for outliers.
- Column `potential` has a minimum value of 54 and has a maximum value of 89 . Values are between 0 and 100 so no changes are needed.
- Column `pace` has a minimum value of -66 and has a maximum value of 94 . Values must be between 0 and 100. Therefore, changes are needed
- Column `shooting` has a minimum value of 21 and has a maximum value of 83 . Values are between 0 and 100 so no changes are needed.
- Column `passing` has a minimum value of 26 and has a maximum value of 90 . Values are between 0 and 100 so no changes are needed.
- Column `dribbling` has a minimum value of -66 and has a maximum value of 92 . Values must be between 0 and 100. Therefore, changes are needed
- Column `defending` has a minimum value of 18 and has a maximum value of 88 . Values are between 0 and 100 so no changes are needed.
- Column `physic` has a minimum value of 35 and has a maximum value of 85 . Values are between 0 and 100 so no changes are needed.
- Column `power_strength` has a minimum value of 29 and has a maximum value of 93 . Values are between 0 and 100 so no changes are needed.
- Column `power_long_shots` has a minimum value of 16 and has a maximum value of 88 . Values are between 0 and 100 so no changes are needed.

Checking for low values in the `wage_eur` column.

```
mydf[mydf["wage_eur"]<500,]
```

	sofifa_id <int>	potential <int>	wage_eur <dbl>	a… <int>	height_cm <int>	weight_kg <int>	club_name <chr>	preferred_foot <chr>	pace <int>
5111	222993	68	4.0001	30	177	68	Cerezo Osaka	Right	75

1 row | 1-10 of 18 columns

- There is one player which weekly wage in euros is 4. This player belongs to a team in Japan (Cerezo Osaka). This is consider an error since the minimum weekly wage for someone in Japan is around 340 euros. [1]
- Columns `pace` and `dribbling` need further investigation since their minimum value is -66.

```
mydf[mydf["pace"] < 0,]
```

	sofifa_id <int>	potential <int>	wage_eur <dbl>	a… <int>	height_cm <int>	weight_kg <int>	club_name <chr>
6011	237742	70	2000	25	179	80	Hokkaido Consadole Sapporo

1 row | 1-8 of 18 columns

Column `pace` has one row which value is -66 which is below 0.

```
mydf[mydf["dribbling"] < 0,]
```

	sofifa_id <int>	potential <int>	wage_eur <dbl>	… <int>	height_cm <int>	weight_kg <int>	club_name <chr>	preferred_foot <chr>
7561	248011	73	3000	22	180	76	Sepsi OSK Sf. Gheorghe	Right

1 row | 1-9 of 18 columns

Column `dribbling` has one row which value is -66 which is below 0.

d. Looking for outliers in columns `wage_eur` , `age` , `height_cm` and `weight_kg` .

Outliers using IQR method:

```
c_col <- c("wage_eur", "age", "height_cm", "weight_kg")

# Outliers using IQR
for (col in c_col){
  # calculating iqr method
  iqr <- IQR(mydf[, col])
  # calculatating q1 and q3
  quant <- quantile(mydf[, col], c(0.25, 0.75))
  # numbers below min limit will be considered outliers
  min_limit <- quant[1] - 1.5*iqr
  # numbers above max limit will be considered outliers
  max_limit <- quant[2] + 1.5*iqr
  print(mydf[mydf[col] > max_limit, ])
  print(mydf[mydf[col] < min_limit, ])
}
```

```
##      sofifa_id potential wage_eur age height_cm weight_kg
## 11      155862        89   300000  34       184        82
## 36      218667        88   230000  25       173        64
## 61      189513        85    59000  31       182        74
## 86      146536        84    43000  34       172        60
## 111     212194        88    75000  24       185        83
## 136     186345        83    71000  29       173        71
## 161     208574        83    45000  27       184        82
## 186     176635        82   100000  31       180        71
## 211     210665        82    65000  28       188        87
## 236     179846        81    87000  33       189        90
## 261     204884        84   115000  25       185        85
## 286     225850        85    69000  24       189        80
## 311     159145        80    53000  34       184        77
## 336     189805        80    38000  29       188        86
## 361     202750        80    84000  29       195        97
## 386     219683        83    74000  25       181        81
## 411     232656        85    37000  22       184        82
## 436     178509        79    95000  33       193        91
## 461     193983        79    43000  28       181        83
## 486     206511        80    46000  26       184        74
## 536     226456        83    41000  24       178        67
## 561     234943        86    37000  23       183        74
## 586     170733        78    38000  33       179        65
## 611     192045        78    46000  29       177        76
## 636     199915        80    56000  28       192        88
## 661     207421        81    56000  25       172        67
## 686     213160        81    51000  26       182        78
## 711     223816        79    68000  26       177        79
## 736     230442        78    33000  20       187        85
## 786     183427        77    65000  30       174        78
## 811     189681        77    36000  29       172        78
## 886     208509        81    24000  25       177        67
## 936     223627        77    42000  31       190        86
## 961     230331        77    25000  36       177        77
## 1011    164859        76    66000  31       176        68
## 1036    184826        76    66000  31       175        69
## 1086    201939        77    29000  27       189        86
## 1111    208451        77    29000  26       183        77
## 1161    219733        79    30000  26       197        82
## 1186    225782        84    52000  22       177        70
## 1211    230373        76    30000  24       172        69
## 1236    235424        82    25000  24       176        68
## 1286    162131        75    41000  35       193        92
## 1361    191053        75    38000  32       177        76
## 1386    194957        75    79000  28       185        71
## 1411    200478        75    37000  28       183        84
## 1436    205193        79    33000  25       186        84
## 1536    222148        79    24000  25       181        75
## 1586    229942        81    33000  22       190        86
## 1661    236947        85    30000  22       191        80
## 1811    199439        77    26000  26       193        87
## 2086    247028        84    34000  22       186        76
## 2111    163050        73    30000  34       175        70
## 2361    224656        80    28000  23       184        82
## 2386    229348        82    52000  22       183        70
## 2711    208502        73    29000  27       178        73
## 3036    241390        84    28000  20       184        74
## 3411    227678        83    25000  22       183        77
## 3486    234579        79    26000  23       177        78
## 3786    209610        70    24000  32       166        64
## 3811    212680        73    26000  25       174        61
## 4636    234679        77    34000  23       190        82
## 6186    253163        81    24000  21       188        79
##                      club_name preferred_foot pace shooting passing dribbling
## 11              Real Madrid CF          Right   71       70      76        73
## 36              Manchester City           Left   81       77      83        92
## 61                Villarreal CF          Right   50       83      90        81
## 86                   Sevilla FC          Right   87       69      79        83
## 111           Borussia Dortmund          Right   81       78      82        87
## 136          Atlético de Madrid          Right   73       60      81        77
## 161         Eintracht Frankfurt           Left   88       76      79        83
## 186                     Arsenal           Left   64       69      85        83
## 211                  RB Leipzig           Left   66       66      73        72
## 236                    Juventus          Right   45       74      77        77
## 261             Manchester City           Left   80       54      77        77
## 286         Paris Saint-Germain           Left   73       39      67        67
## 311                    Al Hilal          Right   67       82      67        76
## 336                  Sevilla FC          Right   55       78      70        69
## 361     Wolverhampton Wanderers          Right   62       36      64        64
## 386           FC Bayern München          Right   72       77      78        78
## 411                    AC Milan           Left   92       67      73        77
## 436                     Chelsea           Left   39       79      70        71
```

```
## 461      Montpellier Hérault SC         Right   84       80      70        75
## 486                VfL Wolfsburg          Left   66       80      78        77
## 536              West Ham United         Right   68       76      78        80
## 561    Borussia Mönchengladbach         Right   67       69      76        81
## 586                        Parma         Right   90       70      71        80
## 611              Tigres U.A.N.L.         Right   84       64      72        75
## 636      Brighton & Hove Albion          Right   51       49      62        63
## 661      Brighton & Hove Albion          Right   82       77      76        81
## 686              Spartak Moskva          Right   76       62      76        77
## 711                   Cruz Azul          Right   87       78      68        81
## 736                   São Paulo          Right   73       82      62        69
## 786                     Everton           Left   66       68      74        76
## 811      Istanbul Basaksehir FK          Right   76       60      74        78
## 886               Stade de Reims         Right   75       57      70        72
## 936                     Al Ahli          Right   69       79      69        72
## 961                   Palmeiras          Right   67       72      75        73
## 1011                    Everton          Right   86       70      69        78
## 1036             Leicester City          Right   53       73      77        75
## 1086                   Atalanta          Right   76       42      64        70
## 1111           1. FSV Mainz 05           Right   75       77      68        79
## 1161       RC Strasbourg Alsace           Left   63       76      58        68
## 1186                    Arsenal          Right   83       59      69        76
## 1211                  São Paulo          Right   72       69      76        76
## 1236             TSG Hoffenheim          Right   60       49      67        72
## 1286                     Napoli          Right   31       76      61        66
## 1361                Torino F.C.          Right   66       59      71        75
## 1386          Manchester United          Right   53       50      60        56
## 1411           Newcastle United          Right   67       68      74        76
## 1436                 Hertha BSC           Left   69       39      58        62
## 1536                 1. FC Köln          Right   68       73      72        79
## 1586                  AS Monaco          Right   55       28      49        56
## 1661                 Hertha BSC           Left   74       34      64        63
## 1811                FC Augsburg           Left   57       74      72        70
## 2086             Spartak Moskva          Right   75       68      67        74
## 2111           Sheffield United          Right   57       74      62        64
## 2361                     Fulham          Right   86       61      68        74
## 2386                     Fulham           Left   87       39      50        72
## 2711              Santos Laguna          Right   77       39      63        68
## 3036                     Napoli          Right   67       69      71        76
## 3411                Aston Villa          Right   74       38      58        68
## 3486             Tigres U.A.N.L.          Right   88       69      60        72
## 3786 Club Deportivo Guadalajara          Right   72       53      68        66
## 3811                  Rangers FC          Right   91       57      66        72
## 4636             Manchester City          Right   64       39      62        69
## 6186                FC Barcelona          Right   61       30      50        48
##      defending physic power_strength power_long_shots high.wage.ind
## 11          88     85              85               64             1
## 36          51     67              54               76             1
## 61          71     69              70               88             1
## 86          80     58              39               70             1
## 111         51     66              71               80             1
## 136         79     72              64               70             1
## 161         68     80              77               78             1
## 186         24     55              56               67             1
## 211         81     80              79               69             1
## 236         79     78              85               78             1
## 261         77     77              79               65             1
## 286         82     84              84               39             1
## 311         36     74              83               75             1
## 336         50     76              84               69             1
## 361         80     83              93               23             1
## 386         76     77              75               81             1
## 411         74     77              78               69             1
## 436         42     77              88               66             1
## 461         33     81              84               74             1
## 486         73     73              71               87             1
## 536         70     67              63               80             1
## 561         64     67              65               81             1
## 586         35     57              52               64             1
## 611         72     77              75               68             1
## 636         79     78              83               67             1
## 661         30     62              60               82             1
## 686         76     72              66               62             1
## 711         38     76              76               72             1
## 736         31     75              79               80             1
## 786         73     71              64               74             1
## 811         74     76              72               62             1
## 886         73     76              70               62             1
## 936         39     80              88               76             1
## 961         74     76              78               80             1
## 1011        39     60              55               65             1
## 1036        69     67              63               77             1
## 1086        75     76              79               35             1
## 1111        44     69              72               76             1
```

```
## 1161        26     79             89              68             1
## 1186        69     68             66              62             1
## 1211        57     45             42              71             1
## 1236        75     74             68              54             1
## 1286        33     68             86              64             1
## 1361        74     78             72              72             1
## 1386        76     73             77              49             1
## 1411        70     75             73              68             1
## 1436        75     77             78              31             1
## 1536        54     61             55              68             1
## 1586        74     82             91              17             1
## 1661        77     74             78              25             1
## 1811        41     70             79              73             1
## 2086        72     79             76              72             1
## 2111        42     68             74              64             1
## 2361        70     73             75              68             1
## 2386        70     69             67              38             1
## 2711        68     71             66              40             1
## 3036        58     67             64              75             1
## 3411        71     72             72              32             1
## 3486        25     71             76              61             1
## 3786        71     70             63              64             1
## 3811        40     46             34              46             1
## 4636        70     68             73              52             1
## 6186        66     68             73              35             1
##  [1] sofifa_id        potential        wage_eur         age
##  [5] height_cm        weight_kg        club_name        preferred_foot
##  [9] pace             shooting         passing          dribbling
## [13] defending        physic           power_strength   power_long_shots
## [17] high.wage.ind
## <0 rows> (or 0-length row.names)
##      sofifa_id potential wage_eur age height_cm weight_kg          club_name
## 4311    201965        69     4000  77       180        76     IFK Norrköping
## 6986    156478        65     2000  38       178        77 Urawa Red Diamonds
##      preferred_foot pace shooting passing dribbling defending physic
## 4311          Right   76       65      67        71        58     70
## 6986          Right   32       66      70        63        66     60
##      power_strength power_long_shots high.wage.ind
## 4311             66               73             0
## 6986             75               72             0
##  [1] sofifa_id        potential        wage_eur         age
##  [5] height_cm        weight_kg        club_name        preferred_foot
##  [9] pace             shooting         passing          dribbling
## [13] defending        physic           power_strength   power_long_shots
## [17] high.wage.ind
## <0 rows> (or 0-length row.names)
##      sofifa_id potential wage_eur age height_cm weight_kg   club_name
## 9011    255592        69     1000  23       220        61 Yokohama FC
##      preferred_foot pace shooting passing dribbling defending physic
## 9011          Right   69       60      60        68        42     42
##      power_strength power_long_shots high.wage.ind
## 9011             36               64             0
##  [1] sofifa_id        potential        wage_eur         age
##  [5] height_cm        weight_kg        club_name        preferred_foot
##  [9] pace             shooting         passing          dribbling
## [13] defending        physic           power_strength   power_long_shots
## [17] high.wage.ind
## <0 rows> (or 0-length row.names)
##      sofifa_id potential wage_eur age height_cm weight_kg
## 361     202750        80    84000  29       195        97
## 8511    211095        64     3000  27       191       188
## 9736    210820        66      550  25       189        93
##                    club_name preferred_foot pace shooting passing dribbling
## 361  Wolverhampton Wanderers          Right   62       36      64        64
## 8511              AC Horsens          Right   60       60      48        57
## 9736        Varbergs BoIS FC          Right   57       51      48        50
##      defending physic power_strength power_long_shots high.wage.ind
## 361         80     83             93               23             1
## 8511        36     75             87               57             0
## 9736        59     71             82               49             0
##  [1] sofifa_id        potential        wage_eur         age
##  [5] height_cm        weight_kg        club_name        preferred_foot
##  [9] pace             shooting         passing          dribbling
## [13] defending        physic           power_strength   power_long_shots
## [17] high.wage.ind
## <0 rows> (or 0-length row.names)
```

Outliers using Standard Deviation method:

```
# Outliers using sd
for (col in c_col){
  sd_col <- 3*sd(mydf[, col])
  mean_col <- mean(mydf[, col])
  min_limit <- mean_col - sd_col
  max_limit <- mean_col + sd_col
  print(mydf[mydf[col] > max_limit, ])
  print(mydf[mydf[col] < min_limit, ])
}
```

```
##     sofifa_id potential wage_eur age height_cm weight_kg
## 11     155862        89   300000  34       184        82
## 36     218667        88   230000  25       173        64
## 186    176635        82   100000  31       180        71
## 236    179846        81    87000  33       189        90
## 261    204884        84   115000  25       185        85
## 361    202750        80    84000  29       195        97
## 436    178509        79    95000  33       193        91
##                   club_name preferred_foot pace shooting passing dribbling
## 11           Real Madrid CF          Right   71       70      76        73
## 36           Manchester City           Left   81       77      83        92
## 186                  Arsenal           Left   64       69      85        83
## 236                 Juventus          Right   45       74      77        77
## 261          Manchester City           Left   80       54      77        77
## 361 Wolverhampton Wanderers          Right   62       36      64        64
## 436                  Chelsea           Left   39       79      70        71
##     defending physic power_strength power_long_shots high.wage.ind
## 11         88     85             85               64             1
## 36         51     67             54               76             1
## 186        24     55             56               67             1
## 236        79     78             85               78             1
## 261        77     77             79               65             1
## 361        80     83             93               23             1
## 436        42     77             88               66             1
##  [1] sofifa_id        potential        wage_eur         age
##  [5] height_cm        weight_kg        club_name        preferred_foot
##  [9] pace             shooting         passing          dribbling
## [13] defending        physic           power_strength   power_long_shots
## [17] high.wage.ind
## <0 rows> (or 0-length row.names)
##      sofifa_id potential wage_eur age height_cm weight_kg      club_name
## 4311    201965        69     4000  77       180        76 IFK Norrköping
##      preferred_foot pace shooting passing dribbling defending physic
## 4311          Right   76       65      67        71        58     70
##      power_strength power_long_shots high.wage.ind
## 4311             66               73             0
##  [1] sofifa_id        potential        wage_eur         age
##  [5] height_cm        weight_kg        club_name        preferred_foot
##  [9] pace             shooting         passing          dribbling
## [13] defending        physic           power_strength   power_long_shots
## [17] high.wage.ind
## <0 rows> (or 0-length row.names)
##      sofifa_id potential wage_eur age height_cm weight_kg   club_name
## 9011    255592        69     1000  23       220        61 Yokohama FC
##      preferred_foot pace shooting passing dribbling defending physic
## 9011          Right   69       60      60        68        42     42
##      power_strength power_long_shots high.wage.ind
## 9011             36               64             0
##  [1] sofifa_id        potential        wage_eur         age
##  [5] height_cm        weight_kg        club_name        preferred_foot
##  [9] pace             shooting         passing          dribbling
## [13] defending        physic           power_strength   power_long_shots
## [17] high.wage.ind
## <0 rows> (or 0-length row.names)
##      sofifa_id potential wage_eur age height_cm weight_kg  club_name
## 8511    211095        64     3000  27       191       188 AC Horsens
##      preferred_foot pace shooting passing dribbling defending physic
## 8511          Right   60       60      48        57        36     75
##      power_strength power_long_shots high.wage.ind
## 8511             87               57             0
##  [1] sofifa_id        potential        wage_eur         age
##  [5] height_cm        weight_kg        club_name        preferred_foot
##  [9] pace             shooting         passing          dribbling
## [13] defending        physic           power_strength   power_long_shots
## [17] high.wage.ind
## <0 rows> (or 0-length row.names)
```

- The first outlier found was a player which age is 77. This number I believe is incorrect because it is too old to be a football player at professional football. [2]
- The second outlier found was a player which height is 220cm. This number I believe is incorrect. [3]
- The third outlier found was a player which weight is 188kg. This number is to high to be correct. [4]

e. Checking for duplicates on the `sofifa_id`

```
# Check for duplicated values
num_duplic = sum(duplicated(mydf$sofifa_id))
message = paste("There are", num_duplic, "duplicated values")
print(message)
```

```
## [1] "There are 1 duplicated values"
```

```
mydf[duplicated(mydf$sofifa_id),]
```

	sofifa_id <int>	potential <int>	wage_eur <dbl>	… <int>	height_cm <int>	weight_kg <int>	club_name <chr>	preferred_foot <chr>	pace <int>
4011	230400	70	8000	31	185	84	Atlético Tucumán	Right	68

1 row | 1-10 of 18 columns

There is 1 duplicated row on the `sofida_id` column.

f. Checking for unique values on the `high.wage.ind` and `preferred_foot` columns.

```
table(mydf$"high.wage.ind")
```

```
##
##   0   1
## 365 149
```

There are 365 values for “0” and 149 values for “1”

```
table(mydf$"preferred_foot")
```

```
##
##  Left right Right
##   121     1   392
```

Values must be constrained to 2 unique values: “Left”, “Right”, however, 1 “right” value was found and it must be changed. There are 121 values for “Left” and 392 values for “Right”.

g. Checking for binary value to be correct. It should be “1” if weekly wage is above 8000 and “0” otherwise.

```
a <- mydf[mydf["wage_eur"] > 8000,]
sum(mydf[mydf["wage_eur"] > 8000,]["high.wage.ind"])
```

```
## [1] 149
```

```
sum(mydf[mydf["wage_eur"] <= 8000,]["high.wage.ind"])
```

```
## [1] 0
```

All values above 8000 are “1” and “0” otherwise. Not mistakes were found.

1.3 Data cleaning

Explain any data quality issues found in 1.2 (5 marks), justify and document the responses made (if any) (5 marks).

Data quality issues found:

1. `preferred_foot` column has 1 value which is “right”, however, it should be “Right”.
2. `sofifa_id` column is int and it should be changed to string (char). `preferred_foot` column is a string (char) and should be changed to factor. `high.wage.ind` is a string (char) and should be changed to factor.
3. `pace` and `dribbling` columns both have 1 row which value is below 0. They need to be changed.
4. The ideal weight for a football player is 48kg for the first 152cm and then 2.7kg for every extra 2.5cm. Therefore, if we have a player that has an altitude of 191cm, it is impossible that his weight is 188kg. Therefore this outlier is considered a mistake. [4]
5. The tallest football player according to FIFA has a height of 206cm. However, one football player whose height is 220cm was found. Therefore, this outlier will be considered as a mistake. [3]
6. `sofifa_id` has 1 duplicated value. Id values can’t have duplicate values since it is an identifier of a person.
7. It was found that 1 player has an age of 77. The oldest football player in professional football has an age of 54. Therefore, this outlier is consider a mistake. [2]
8. There is one football player in a football club in Japan which weekly wage is 4 euros. This does not make sense since the minimum weekly salary in Japan is around 340 euros. This was not identified as an outlier, however this is considered a mistake aswell. [1]

Data Cleaning Implementation:

```
df_clean <- data.frame(mydf)
```

1. Fixing the “right” value to “Right”

```
df_clean[df_clean["preferred_foot"]=="right",]["preferred_foot"] <- "Right"
table(df_clean$"preferred_foot")
```

```
##
##  Left Right
##   121   393
```

2. Converting `sofifa_id` , `preferred_foot` and `high.wage.ind` to their factor data type.

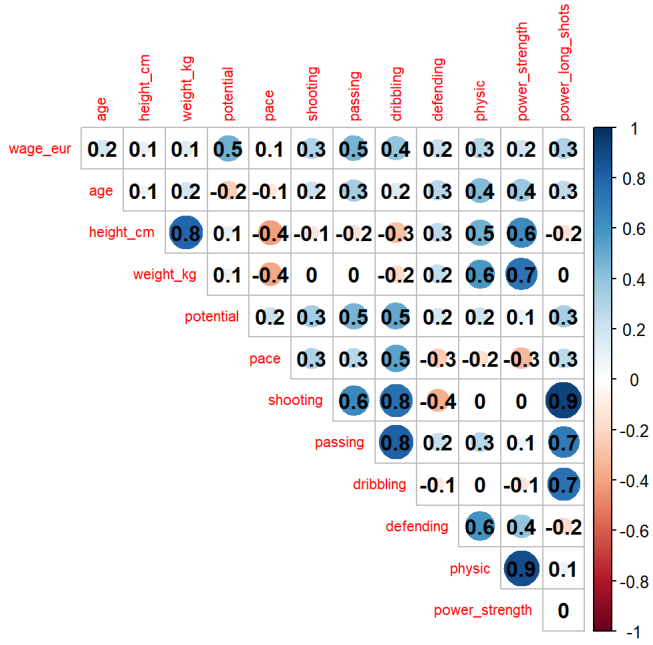
```
df_clean$"sofifa_id" <- as.character(df_clean$"sofifa_id")
df_clean$"preferred_foot" <- as.factor(df_clean$"preferred_foot")
df_clean$"high.wage.ind" <- as.factor(df_clean$"high.wage.ind")
str(df_clean)
```

```
## 'data.frame':    514 obs. of  17 variables:
##  $ sofifa_id       : chr  "155862" "218667" "189513" "146536" ...
##  $ potential       : int  89 88 85 84 88 83 83 82 82 81 ...
##  $ wage_eur        : num  300000 230000 59000 43000 75000 71000 45000 100000 65000 87000 ...
##  $ age             : int  34 25 31 34 24 29 27 31 28 33 ...
##  $ height_cm       : int  184 173 182 172 185 173 184 180 188 189 ...
##  $ weight_kg       : int  82 64 74 60 83 71 82 71 87 90 ...
##  $ club_name       : chr  "Real Madrid CF" "Manchester City" "Villarreal CF" "Sevilla FC" ...
##  $ preferred_foot  : Factor w/ 2 levels "Left","Right": 2 1 2 2 2 2 1 1 1 2 ...
##  $ pace            : int  71 81 50 87 81 73 88 64 66 45 ...
##  $ shooting        : int  70 77 83 69 78 60 76 69 66 74 ...
##  $ passing         : int  76 83 90 79 82 81 79 85 73 77 ...
##  $ dribbling       : int  73 92 81 83 87 77 83 83 72 77 ...
##  $ defending       : int  88 51 71 80 51 79 68 24 81 79 ...
##  $ physic          : int  85 67 69 58 66 72 80 55 80 78 ...
##  $ power_strength  : int  85 54 70 39 71 64 77 56 79 85 ...
##  $ power_long_shots: int  64 76 88 70 80 70 78 67 69 78 ...
##  $ high.wage.ind   : Factor w/ 2 levels "0","1": 2 2 2 2 2 2 2 2 2 2 ...
```

Now, all data types are correct.

We will try to use correlation to find the most suitable columns to be used in a regression model to fix the incorrect numbers.

```
num_col <- c("wage_eur",
             "age",
             "height_cm",
             "weight_kg",
             "potential",
             "pace",
             "shooting",
             "passing",
             "dribbling",
             "defending",
             "physic",
             "power_strength",
             "power_long_shots")

df_cor <- subset(df_clean, select=num_col)
df_cor <- df_cor[df_cor$"pace" > 0 & df_cor$"dribbling" > 0 & df_cor$"weight_kg" != 188 & df_cor$"height_cm" != 220, ]
cor_matrix <- cor(df_cor)
corrplot(cor_matrix, diag=F, type="upper", insig="p-value", number.digits=1, addCoef.col="black", tl.cex=0.7)
```



Since the `pace` value is not highly correlated with any value, the incorrect value will be replaced by the mean of the `pace` column. The column `dribbling` is highly correlated with the columns `passing` and `shooting`. Therefore, any of both values can be used to correct the negative values. The relation between `passing` and `dribbling` will be used to correct the negative values.

3. Correcting the value -66 in the `pace` column with the mean. The mean value was calculated without the -66 value. Also, correcting the value -66 in the `dribbling` column using the `passing` column since it is the most correlated to it.

```
mean_pace <- mean(df_clean[df_clean$"pace" != -66,]$"pace")
df_clean[df_clean$"pace" == -66,]["pace"] <- mean_pace
```

Checking if changes in the `pace` column were made.

```
df_clean[df_clean$"pace" < 0,]
```

0 rows | 1-10 of 17 columns

The column `pace` now does not have negative values.

Since `dribbling` is highly correlated with `passing`, I will build a regression model with `dribbling` as my dependent variable and `passing` as my independent variable to fix the -66 value.

```
model_dribbling <- lm(dribbling~passing, data=df_clean)
summary(model_dribbling)
```

```
##
## Call:
## lm(formula = dribbling ~ passing, data = df_clean)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -131.300   -3.257    0.265    3.864   20.049
##
## Coefficients:
##             Estimate Std. Error t value            Pr(>|t|)
## (Intercept) 17.55846    2.05531   8.543 <0.0000000000000002 ***
## passing      0.78265    0.03501  22.355 <0.0000000000000002 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 8.139 on 512 degrees of freedom
## Multiple R-squared:  0.494,  Adjusted R-squared:  0.493
## F-statistic: 499.8 on 1 and 512 DF,  p-value: < 0.00000000000000022
```

61 is the number in the `passing` column where the error in the `dribbling` column is located.

```
passing_value <- as.data.frame(61)
colnames(passing_value) <- c("passing")
dribbling_value <- predict(model_dribbling, newdata=passing_value)
dribbling_value
```

```
##        1
## 65.29981
```

```
df_clean[df_clean$"dribbling" == -66,]["dribbling"] <- dribbling_value
```

```
df_clean[df_clean$"dribbling" < 0,]
```

0 rows | 1-10 of 17 columns

Dribbling values below 0 have been fixed.

4. Since `weight_kg` is highly correlated with `height_cm`, I will build a regression model with `weight_kg` as my dependent variable and `height_cm` as my independent variable to fix the 188 value. 191cm is the height of the football player which weight is 188kg and that we consider a mistake.

```
# dataframe without the errors in the weight_kg column and the height_cm column
df_model_heigh_weight <- df_clean[df_clean$"weight_kg" != 188 & df_clean$"height_cm" != 220, ]
```

```
# model with weight_kg as dependent variable
model_weight_height <- lm(weight_kg~height_cm, data=df_model_heigh_weight)
summary(model_weight_height)
```

```
##
## Call:
## lm(formula = weight_kg ~ height_cm, data = df_model_heigh_weight)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -13.3609  -2.5446   0.0525   2.4765  12.0441
##
## Coefficients:
##              Estimate Std. Error t value            Pr(>|t|)
## (Intercept) -69.95846    4.89685  -14.29 <0.0000000000000002 ***
## height_cm     0.80168    0.02718   29.49 <0.0000000000000002 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 4.196 on 510 degrees of freedom
## Multiple R-squared:  0.6303, Adjusted R-squared:  0.6296
## F-statistic: 869.7 on 1 and 510 DF,  p-value: < 0.00000000000000022
```

```
height_value <- as.data.frame(191)
colnames(height_value) <- c("height_cm")
predicted_weight <- predict(model_weight_height, height_value)
predicted_weight
```

```
##        1
## 83.16263
```

```
df_clean[df_clean["weight_kg"] == 188, ]["weight_kg"] <- predicted_weight
```

The outlier on the `weight_kg` got fixed.

5. Since `height_cm` is highly correlated with `weight_kg`, I will build a regression model with `height_cm` as my dependent variable and `weight_kg` as my independent variable to fix the 220 value. 61kg is the weight of the football player which height is 220cm and that we consider a mistake.

```
model_height_weight <- lm(height_cm~weight_kg, data=df_model_heigh_weight)
summary(model_height_weight)
```

```
##
## Call:
## lm(formula = height_cm ~ weight_kg, data = df_model_heigh_weight)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -12.0890  -2.8007  -0.1655   2.7679  14.2012
##
## Coefficients:
##              Estimate Std. Error t value            Pr(>|t|)
## (Intercept) 121.54567    1.99073   61.06 <0.0000000000000002 ***
## weight_kg     0.78628    0.02666   29.49 <0.0000000000000002 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 4.156 on 510 degrees of freedom
## Multiple R-squared:  0.6303, Adjusted R-squared:  0.6296
## F-statistic: 869.7 on 1 and 510 DF,  p-value: < 0.00000000000000022
```

```
weight_value <- as.data.frame(61)
colnames(weight_value) <- c("weight_kg")
predicted_height <- predict(model_height_weight, weight_value)
predicted_height
```

```
##        1
## 169.5085
```

```
df_clean[df_clean["height_cm"] == 220, ]["height_cm"] <- predicted_height
```

The outlier on the `height_cm` got fixed.

6. Changing the duplicated id to any string that is unique in the column

```
df_clean[duplicated(df_clean$sofifa_id),]["sofifa_id"] <- "101096"
df_clean[duplicated(df_clean$sofifa_id),]
```

0 rows | 1-10 of 17 columns

The duplicated value on the `sofifa_id` was modified to 101096. Why? it is a random number I picked.

7. Changing the age value to the second max value in the column

```
df_clean[df_clean["age"]==77,]["age"] <- 38
max(df_clean$"age")
```

```
## [1] 38
```

We replaced the age of 77 by 38 because it is the max value after 77.

8. Changing the wage value to its minimum value after 4.

```
df_clean[df_clean["wage_eur"]<500,]["wage_eur"] <- 500
min(df_clean$"wage_eur")
```

```
## [1] 500
```

```
df <- data.frame(df_clean)
df
```

	sofifa_id <chr>	potential <int>	wage_eur <dbl>	a… <dbl>	height_cm <dbl>	weight_kg <dbl>	club_name <chr>
11	155862	89	300000	34	184.0000	82.00000	Real Madrid CF
36	218667	88	230000	25	173.0000	64.00000	Manchester City
61	189513	85	59000	31	182.0000	74.00000	Villarreal CF
86	146536	84	43000	34	172.0000	60.00000	Sevilla FC
111	212194	88	75000	24	185.0000	83.00000	Borussia Dortmund
136	186345	83	71000	29	173.0000	71.00000	Atlético de Madrid
161	208574	83	45000	27	184.0000	82.00000	Eintracht Frankfurt
186	176635	82	100000	31	180.0000	71.00000	Arsenal
211	210665	82	65000	28	188.0000	87.00000	RB Leipzig
236	179846	81	87000	33	189.0000	90.00000	Juventus

1-10 of 514 rows | 1-8 of 18 columns Previous **1** 2 3 4 5 6 … 52 Next

List of modifications done:

1. The row which value on `preferred_foot` was “right” was changed to “Right”. This was done because this column only admits 2 values: “Right” and “Left”.
2. `sofifa_id` column was converted to string data type. `preferred_foot` column was changed to factor data type. `high.wage.ind` column was converted to factor data type. Those changes were done because those columns are intended to be used in such way.
3. The column `pace` has 1 value below 0 which is -66. Column `pace` does not have any column which is highly correlated. Therefore, we calculated the mean of the column `pace` (without using the -66 value) and we change the -66 with the mean value. The column `dribbling` is highly correlated to `passing` (correlation of 0.8 without using the -66 value). Therefore, to fix the value, a linear regression model was built and used to predict the dribbling value.
4. `weight_kg` is highly correlated to `height_cm` (correlation of 0.8 without using the outliers). Therefore a linear regression model was used to fix the outlier of 188 on the `weight_kg` column.
5. `height_cm` is highly correlated to `weight_kg` (correlation of 0.8 without using the outliers). Therefore a linear regression model was used to fix the outlier of 220 on the `height_cm` column.
6. `sofifa_id` column had a duplicated value. An id is a unique value that identifies a player. This id could be modified to any value as long as it is not repeated. We used the value of “101096”. It is an arbitrary number but I decided to use this one as it is random and is not used in the column.
7. The second max value of a player is 38. Therefore, the player which age is 77 was changed to 38.
8. 500 is the minimum weekly wage after 4. This number was used to replace 4.

2. Exploratory Data Analysis (EDA)

2.1 EDA plan

Outline a suitable plan to explore, describe and visualize your data. (5 marks)

Univariate EDA planning:

1. For columns `potential`, `wage_eur`, `age`, `height_cm`, `weight_kg`, `pace`, `shooting`, `passing`, `dribbling`, `defending`, `physic`, `power_strength`, `power_long_shots` columns, I will explore the maximum value, the minimum value, the range value, the standard deviation and the interquartile range (IQR). The mean value and median value will also be calculated. Furthermore, outliers will be identified by using boxplots. Histograms will be graphed to check if they follow a normall distribution or if they are skewed to any side. Furthermore, qqplot graphs and Shapiro tests will be done to check for normality.
2. For columns which datatype is factor (`preferred_foot`, `high.wage.ind`), barplots and tables will be done.

Bivariate EDA planning:

Since our research questions are regarding columns `potential` and `high.wage.ind`, I will do the following:

1. Research question regarding column `potential`:
    a. I will make the column `potential` as our dependent variable and we will create scatterplots against the following columns: `wage_eur`, `age`, `height_cm`, `weight_kg`, `pace`, `shooting`, `passing`, `dribbling`, `defending`, `physic`, `power_strength`, `power_long_shots`. Furthermore, a correlation plot will be created to see insight of the correlation that exist between potential and the other columns and also look for collinearity. Then, correlation tests will be done.
    b. I will make `potential` as my dependent variable and perform ANOVA analysis with `preferred_foot` and `high.wage.ind` separately.
2. Research question regarding column `high.wage.ind`:
    a. I will create boxplots with `high.wage.ind` on the x-axis and columns `wage_eur`, `age`, `height_cm`, `weight_kg`, `pace`, `shooting`, `passing`, `dribbling`, `defending`, `physic`, `power_strength`, `power_long_shots` and `potential` in the y-axis. The I will perform t-tests to see if the is significance difference or not between their means.
    b. I will perform a chi-squared test to see if there is a relationship between `high.wage.ind` and `preferred_foot`

2.2 EDA and summary of results

Undertake and summaries the findings of your data exploration, particularly with respect to the research questions. Use appropriate summary statistics (uni- and multi-variate) and visualizations. (10 marks)

Univariate EDA

```
# This function is used to calculate max, min, range, iqr, std, mean and medians of a single numerical column
numeric_eda <- function(col){

  max_value = max(df[col])
  min_value = min(df[col])
  range_value = max_value - min_value
  iqr_value = IQR(df[, col])
  std_value = sd(df[, col])
  mean_value = mean((df[, col]))
  median_value = median(df[, col])

  max_message = paste("The max value of the column", col, "is", round(max_value,2))
  min_message = paste("The min value of the column", col, "is", round(min_value,2))
  range_message = paste("The range value of the column", col, "is", round(range_value,2))
  iqr_message = paste("The column", col, "has an interquartile range of", round(iqr_value,2))
  std_message = paste("The column", col, "has an standard deviation of", round(std_value,2))
  mean_message = paste("The mean value of the column", col, "is", round(mean_value,2))
  median_message = paste("The median value of the column", col, "is", round(median_value,2))

  print(max_message)
  print(min_message)
  print(range_message)
  print(iqr_message)
  print(std_message)
  print(mean_message)
  print(median_message)
  print("--------------------------------------------------")
}
```

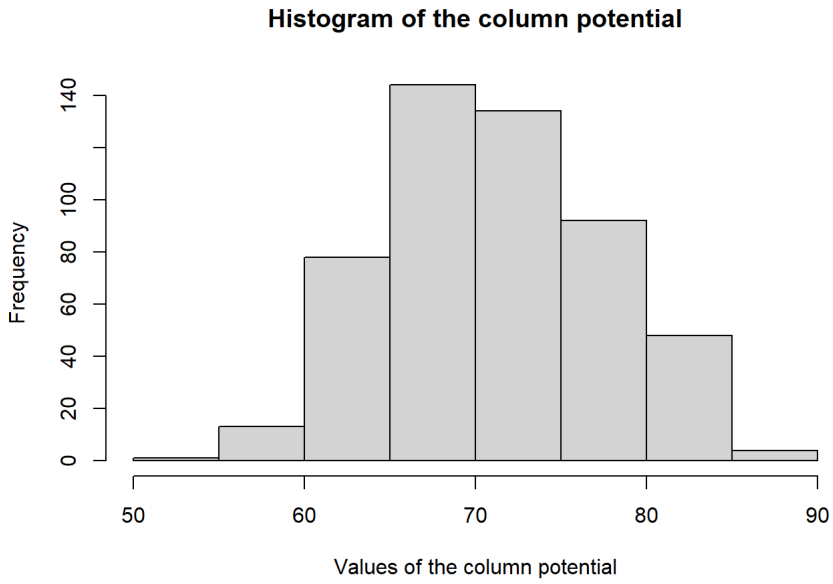
```
numeric_columns <- c("potential",
                     "wage_eur",
                     "age",
                     "height_cm",
                     "weight_kg",
                     "pace",
                     "shooting",
                     "passing",
                     "dribbling",
                     "defending",
                     "physic",
                     "power_strength",
                     "power_long_shots")

for (col in numeric_columns){
  numeric_eda(col)
}
```

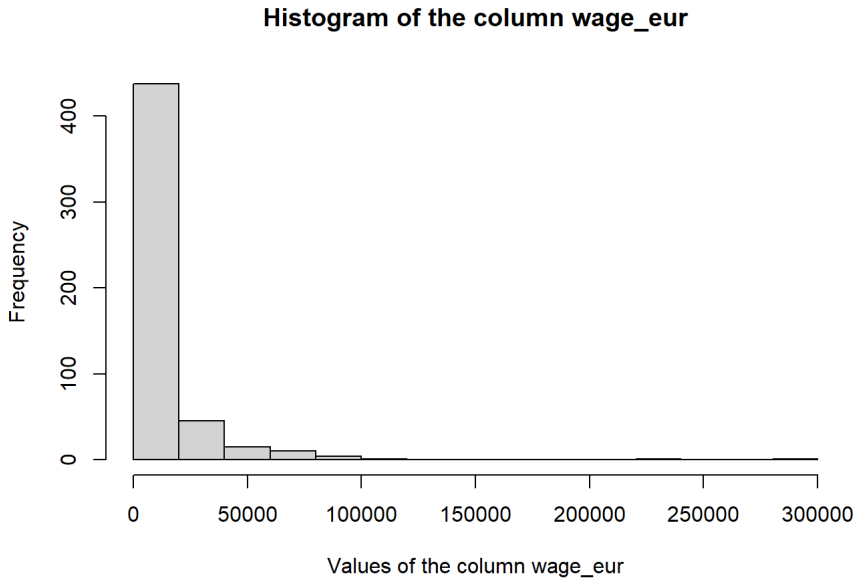
```
## [1] "The max value of the column potential is 89"
## [1] "The min value of the column potential is 54"
## [1] "The range value of the column potential is 35"
## [1] "The column potential has an interquartile range of 9"
## [1] "The column potential has an standard deviation of 6.39"
## [1] "The mean value of the column potential is 71.67"
## [1] "The median value of the column potential is 72"
## [1] "-------------------------------------------------"
## [1] "The max value of the column wage_eur is 300000"
## [1] "The min value of the column wage_eur is 500"
## [1] "The range value of the column wage_eur is 299500"
## [1] "The column wage_eur has an interquartile range of 9000"
## [1] "The column wage_eur has an standard deviation of 22823.07"
## [1] "The mean value of the column wage_eur is 10821.6"
## [1] "The median value of the column wage_eur is 3000"
## [1] "-------------------------------------------------"
## [1] "The max value of the column age is 38"
## [1] "The min value of the column age is 16"
## [1] "The range value of the column age is 22"
## [1] "The column age has an interquartile range of 6"
## [1] "The column age has an standard deviation of 4.56"
## [1] "The mean value of the column age is 25.15"
## [1] "The median value of the column age is 25"
## [1] "-------------------------------------------------"
## [1] "The max value of the column height_cm is 200"
## [1] "The min value of the column height_cm is 160"
## [1] "The range value of the column height_cm is 40"
## [1] "The column height_cm has an interquartile range of 10"
## [1] "The column height_cm has an standard deviation of 6.85"
## [1] "The mean value of the column height_cm is 180"
## [1] "The median value of the column height_cm is 180"
## [1] "-------------------------------------------------"
## [1] "The max value of the column weight_kg is 97"
## [1] "The min value of the column weight_kg is 58"
## [1] "The range value of the column weight_kg is 39"
## [1] "The column weight_kg has an interquartile range of 9"
## [1] "The column weight_kg has an standard deviation of 6.92"
## [1] "The mean value of the column weight_kg is 74.34"
## [1] "The median value of the column weight_kg is 75"
## [1] "-------------------------------------------------"
## [1] "The max value of the column pace is 94"
## [1] "The min value of the column pace is 31"
## [1] "The range value of the column pace is 63"
## [1] "The column pace has an interquartile range of 14"
## [1] "The column pace has an standard deviation of 11.15"
## [1] "The mean value of the column pace is 68.39"
## [1] "The median value of the column pace is 69"
## [1] "-------------------------------------------------"
## [1] "The max value of the column shooting is 83"
## [1] "The min value of the column shooting is 21"
## [1] "The range value of the column shooting is 62"
## [1] "The column shooting has an interquartile range of 23"
## [1] "The column shooting has an standard deviation of 14.2"
## [1] "The mean value of the column shooting is 53.04"
## [1] "The median value of the column shooting is 55"
## [1] "-------------------------------------------------"
## [1] "The max value of the column passing is 90"
## [1] "The min value of the column passing is 26"
## [1] "The range value of the column passing is 64"
## [1] "The column passing has an interquartile range of 14"
## [1] "The column passing has an standard deviation of 10.26"
## [1] "The mean value of the column passing is 57.81"
## [1] "The median value of the column passing is 58"
## [1] "-------------------------------------------------"
## [1] "The max value of the column dribbling is 92"
## [1] "The min value of the column dribbling is 32"
## [1] "The range value of the column dribbling is 60"
## [1] "The column dribbling has an interquartile range of 13"
## [1] "The column dribbling has an standard deviation of 9.91"
## [1] "The mean value of the column dribbling is 63.06"
## [1] "The median value of the column dribbling is 64"
## [1] "-------------------------------------------------"
## [1] "The max value of the column defending is 88"
## [1] "The min value of the column defending is 18"
## [1] "The range value of the column defending is 70"
## [1] "The column defending has an interquartile range of 29"
## [1] "The column defending has an standard deviation of 16.85"
## [1] "The mean value of the column defending is 51.53"
## [1] "The median value of the column defending is 57"
## [1] "-------------------------------------------------"
## [1] "The max value of the column physic is 85"
## [1] "The min value of the column physic is 35"
## [1] "The range value of the column physic is 50"
```

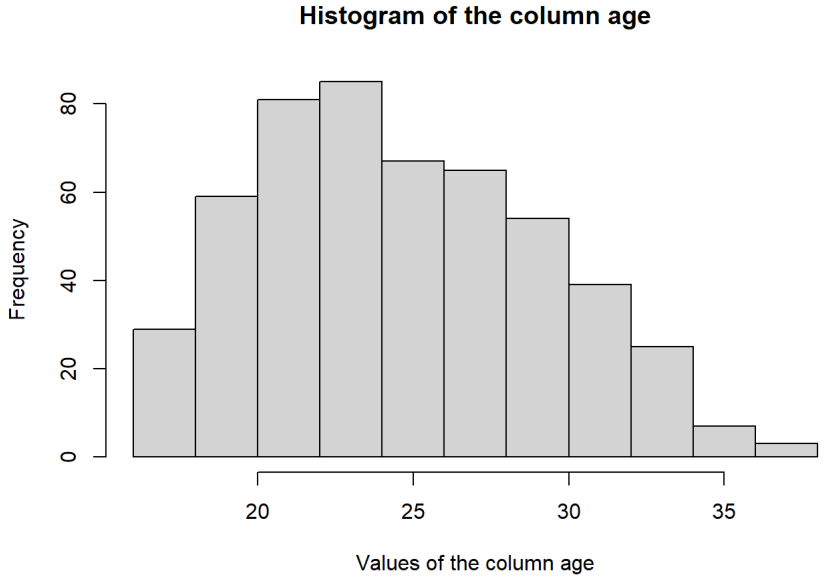
```
## [1] "The column physic has an interquartile range of 14"
## [1] "The column physic has an standard deviation of 10.03"
## [1] "The mean value of the column physic is 64.61"
## [1] "The median value of the column physic is 66"
## [1] "--------------------------------------------------"
## [1] "The max value of the column power_strength is 93"
## [1] "The min value of the column power_strength is 29"
## [1] "The range value of the column power_strength is 64"
## [1] "The column power_strength has an interquartile range of 18"
## [1] "The column power_strength has an standard deviation of 13.24"
## [1] "The mean value of the column power_strength is 65.31"
## [1] "The median value of the column power_strength is 67"
## [1] "--------------------------------------------------"
## [1] "The max value of the column power_long_shots is 88"
## [1] "The min value of the column power_long_shots is 16"
## [1] "The range value of the column power_long_shots is 72"
## [1] "The column power_long_shots has an interquartile range of 24"
## [1] "The column power_long_shots has an standard deviation of 15.65"
## [1] "The mean value of the column power_long_shots is 51.83"
## [1] "The median value of the column power_long_shots is 54"
## [1] "--------------------------------------------------"
```

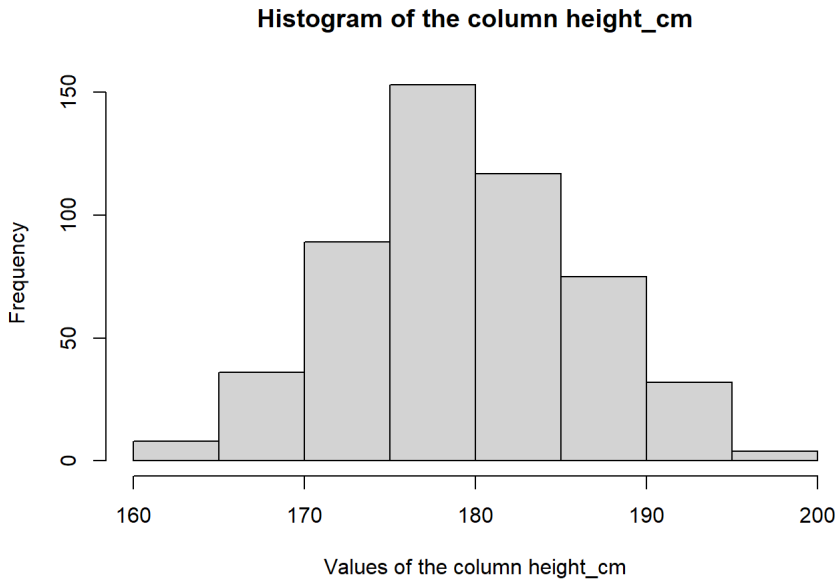
Now we will plot histograms for each column:

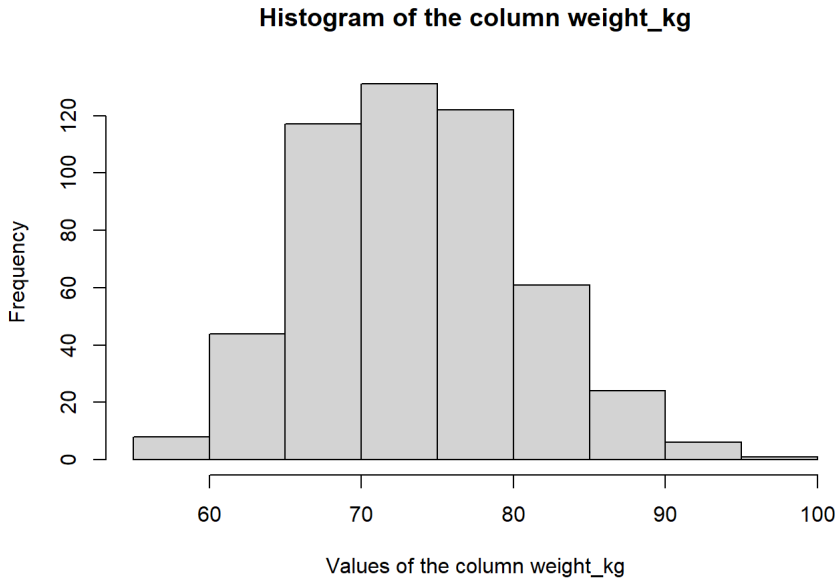
```
for (col in numeric_columns){
  title = paste("Histogram of the column", col)
  x_axis = paste("Values of the column", col)
  hist(df[,col], main=title, xlab=x_axis)
}
```

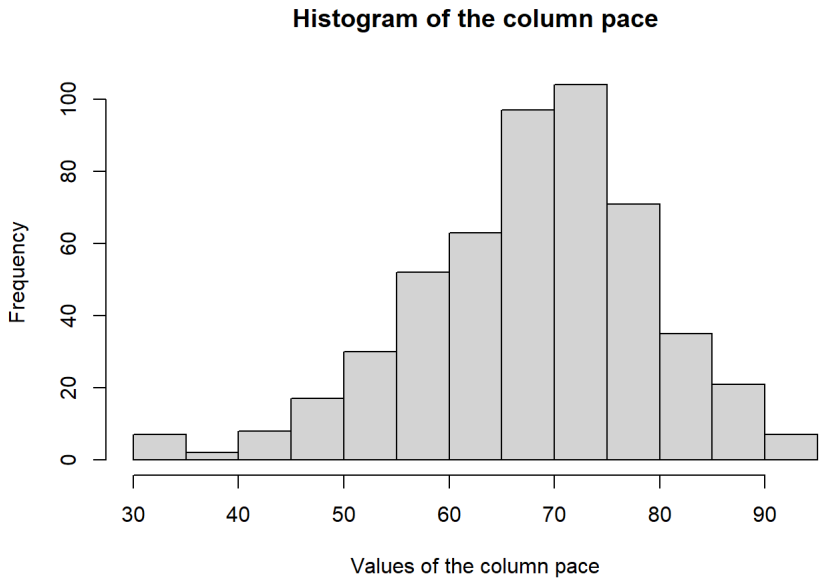

Histogram of the column potential
Frequency
0
20
40
60
80
100
140
50
60
70
80
90
Values of the column potential

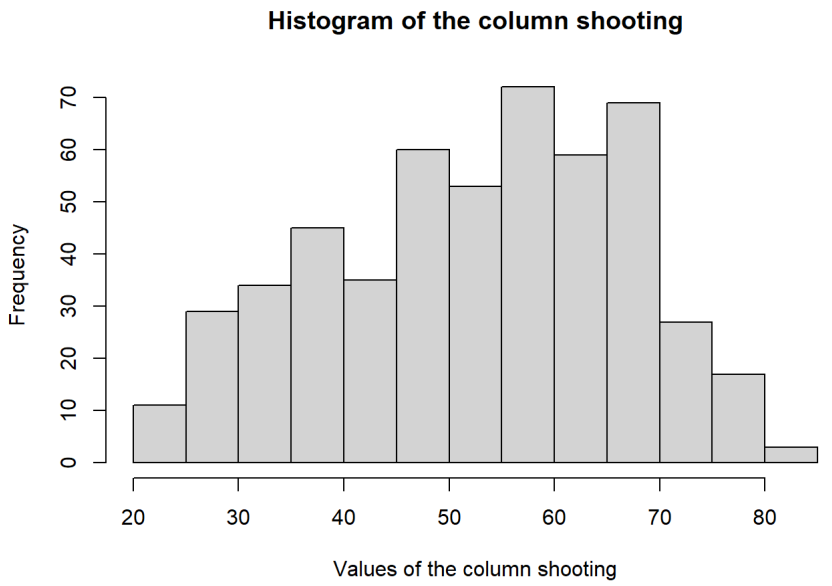

Histogram of the column wage_eur
Frequency
0
100
200
300
400
0
50000
100000
150000
200000
250000
300000
Values of the column wage_eur

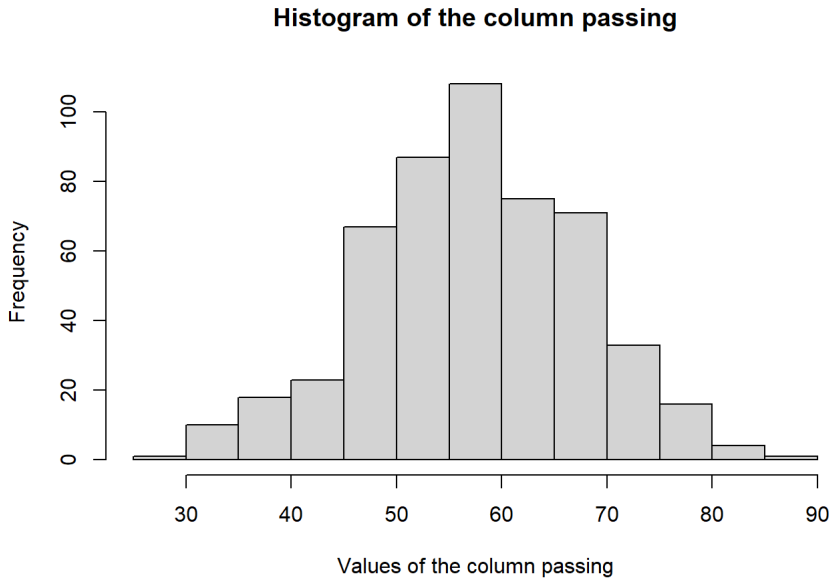

Histogram of the column age
Frequency
0
20
40
60
80
20
25
30
35
Values of the column age

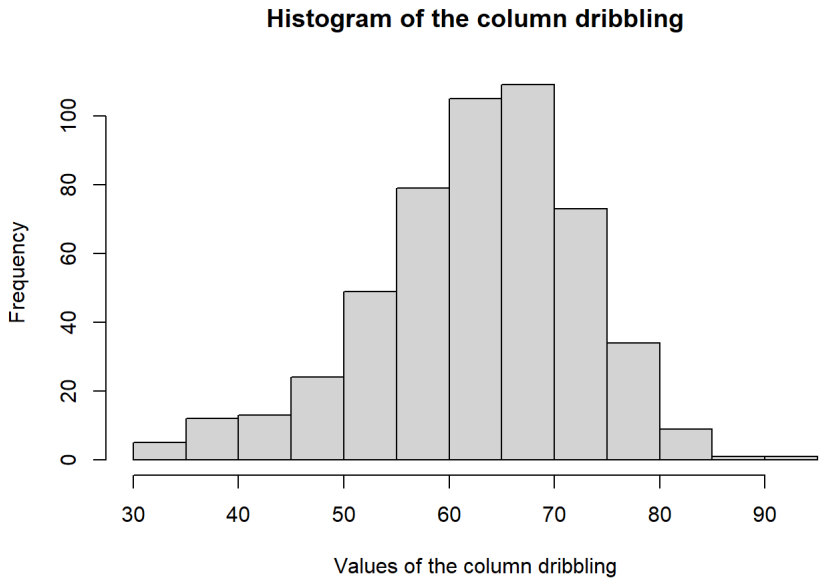

Histogram of the column height_cm
Frequency
0
50
100
150
160
170
180
190
200
Values of the column height_cm

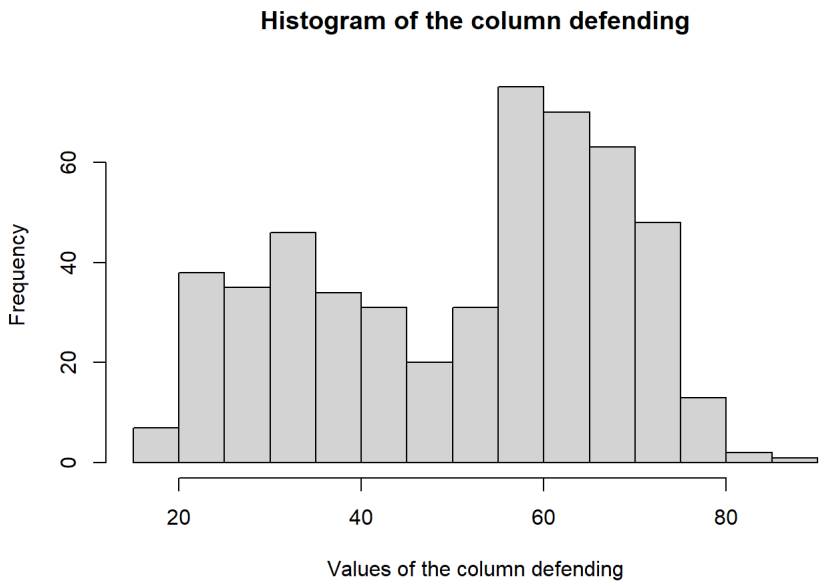

Histogram of the column weight_kg
Frequency
0
20
40
60
80
100
120
60
70
80
90
100
Values of the column weight_kg

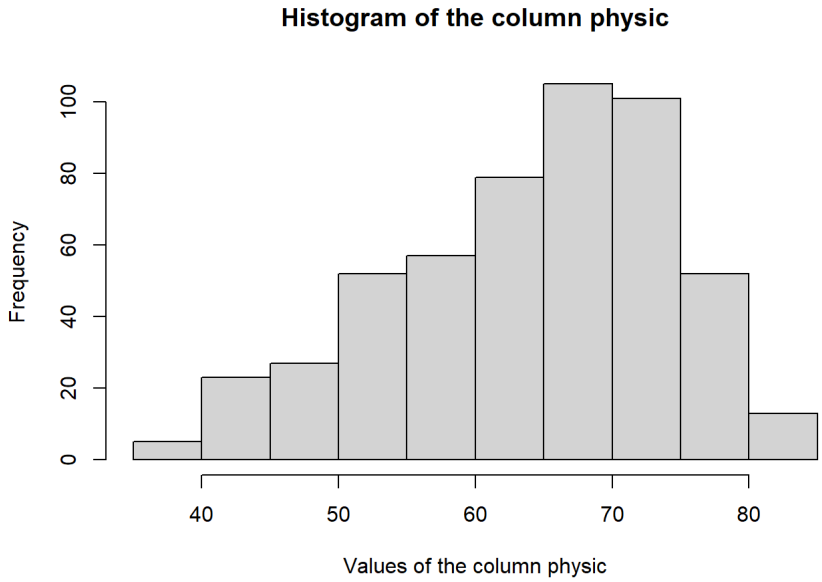

Histogram of the column pace
Frequency
0
20
40
60
80
100
30
40
50
60
70
80
90
Values of the column pace

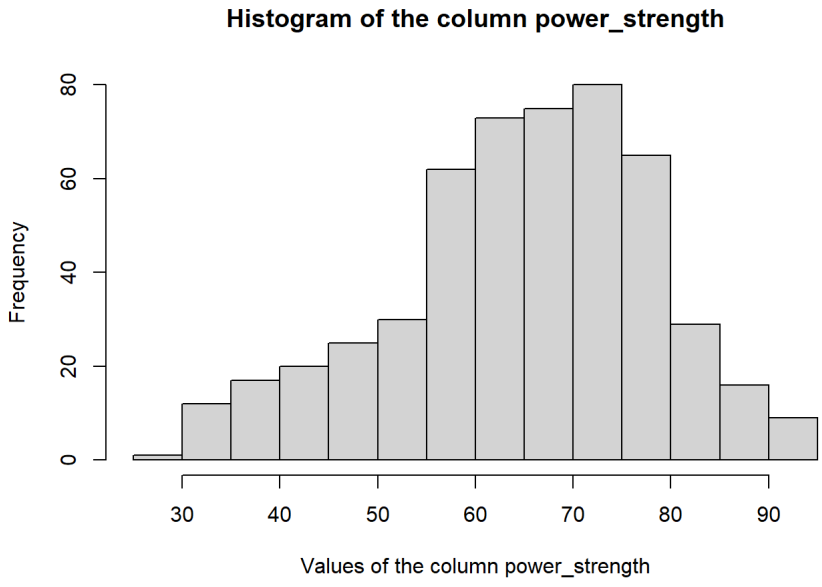

Histogram of the column shooting
Frequency
0
10
20
30
40
50
60
70
20
30
40
50
60
70
80
Values of the column shooting

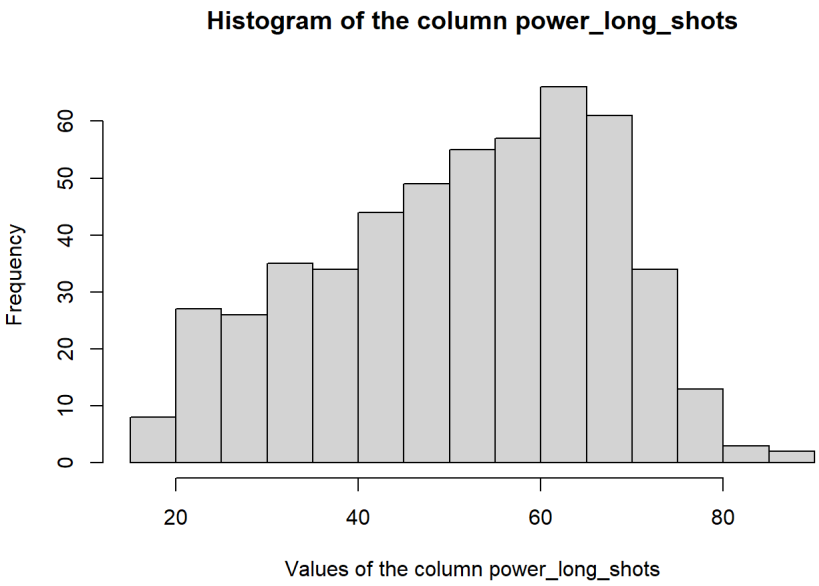

Histogram of the column passing
Frequency
0
20
40
60
80
100
30
40
50
60
70
80
90
Values of the column passing


Histogram of the column dribbling
Frequency
0
20
40
60
80
100
30
40
50
60
70
80
90
Values of the column dribbling


Histogram of the column defending
Frequency
0
20
40
60
20
40
60
80
Values of the column defending


Histogram of the column physic
Frequency
0
20
40
60
80
100
40
50
60
70
80
Values of the column physic

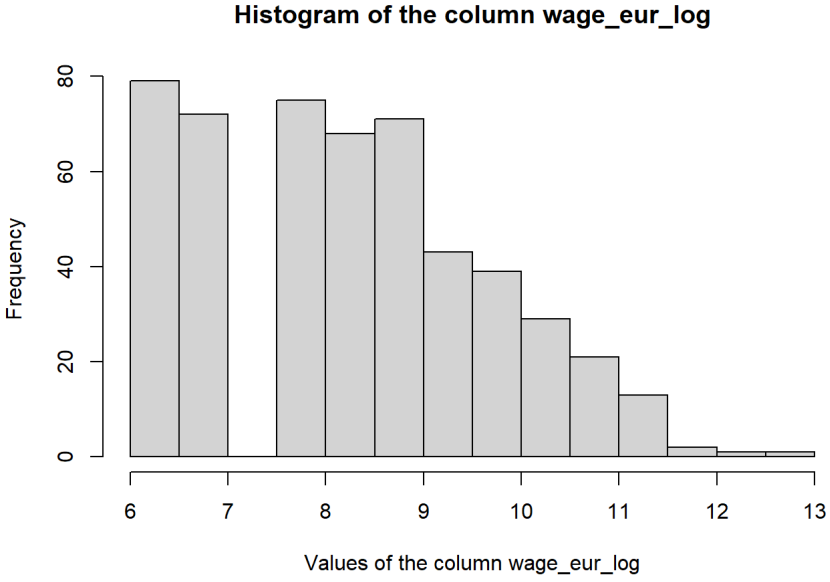

Histogram of the column power_strength
Frequency
0
20
40
60
80
30
40
50
60
70
80
90
Values of the column power_strength



- `potential` column seems to follow a normal distribution.
- `wage_eur` column seems to follow logarithmic distribution. [5]
- `age` column seems to be slightly skewed to the right.
- `height_cm` column seems to be to follow a normal distribution.
- `weight_kg` column seems to be slightly skewed to the right.
- `pace` column seems to be skewed to the left.
- `shooting` column seems to be skewed to the left.
- `passing` column seems to follow a normal distribution.
- `dribbling` column seems to be slightly skewed to the left.
- `defending` column seems to have 2 picks.
- `physic` column seems to be skewed to the left.
- `power_strength` column seems to be slightly skewed to the left.
- `power_long_shots` column seems to be skewed to the left.

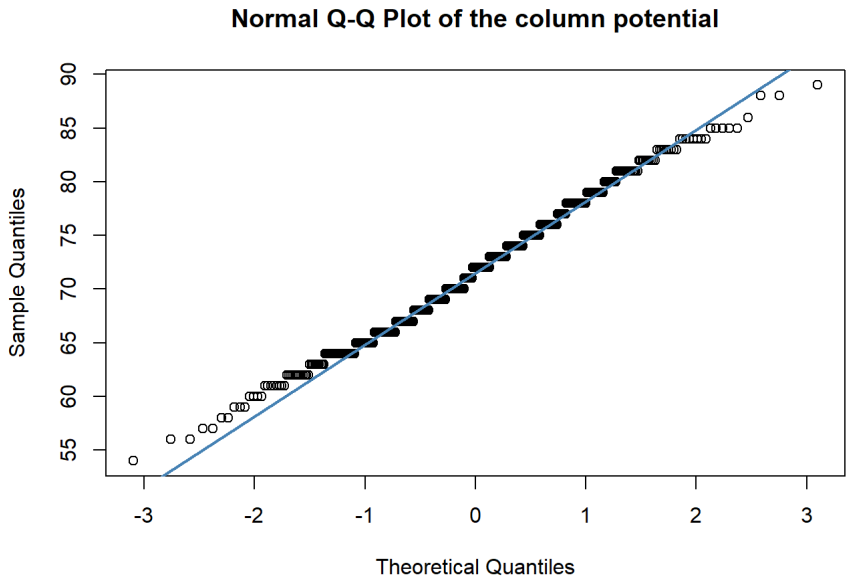
Since the column `wage_eur` seems to follow a logarithmic distribution, I will use a log function to transform values so that it tryies to follow now a normal distribution.

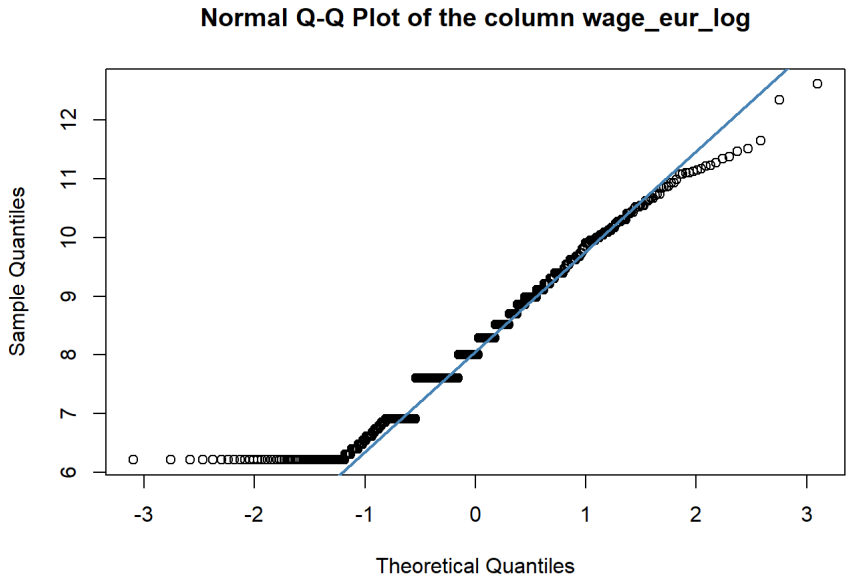
```
df$wage_eur_log <- log(df$wage_eur)
hist(df$wage_eur_log, main="Histogram of the column wage_eur_log", xlab="Values of the column wage_eur_log")
```

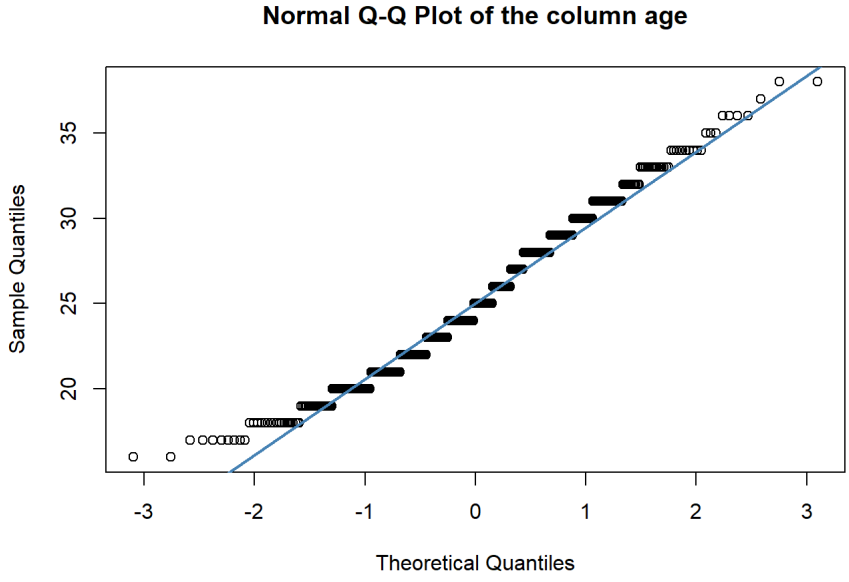


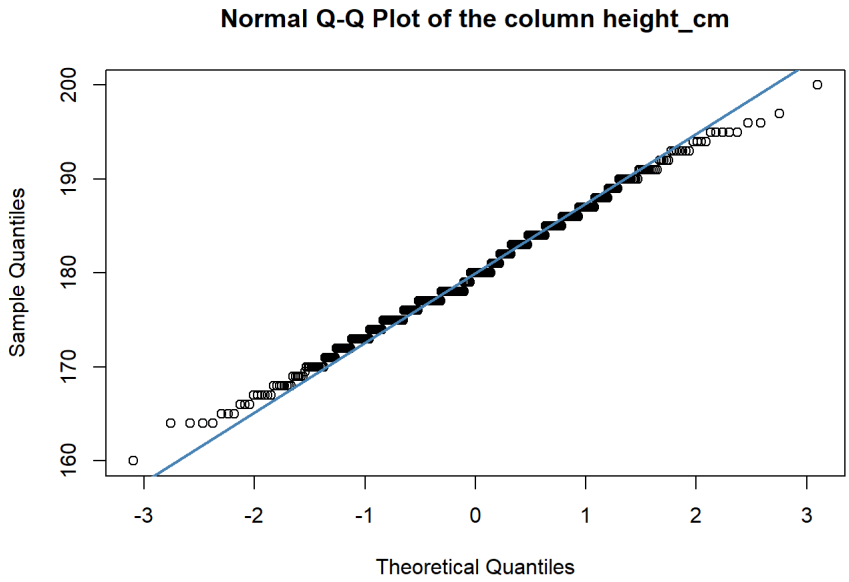
- `wage_eur_log` seems to be skewed to the right.

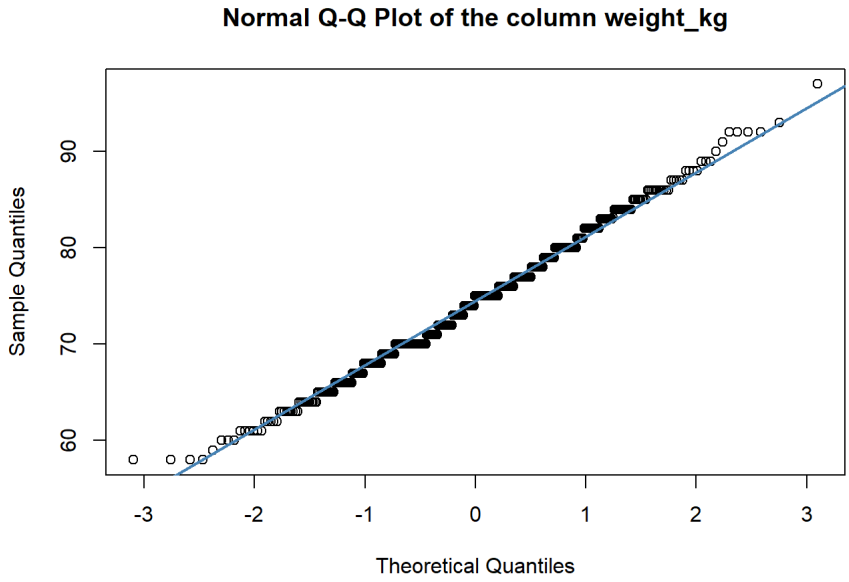
```
numeric_columns <- c("potential",
                     "wage_eur_log",
                     "age",
                     "height_cm",
                     "weight_kg",
                     "pace",
                     "shooting",
                     "passing",
                     "dribbling",
                     "defending",
                     "physic",
                     "power_strength",
                     "power_long_shots")

for (col in numeric_columns){
  title = paste("Normal Q-Q Plot of the column", col)
  qqnorm(df[,col], main=title)
  qqline(df[,col], col="steelblue", lwd=2)
}
```

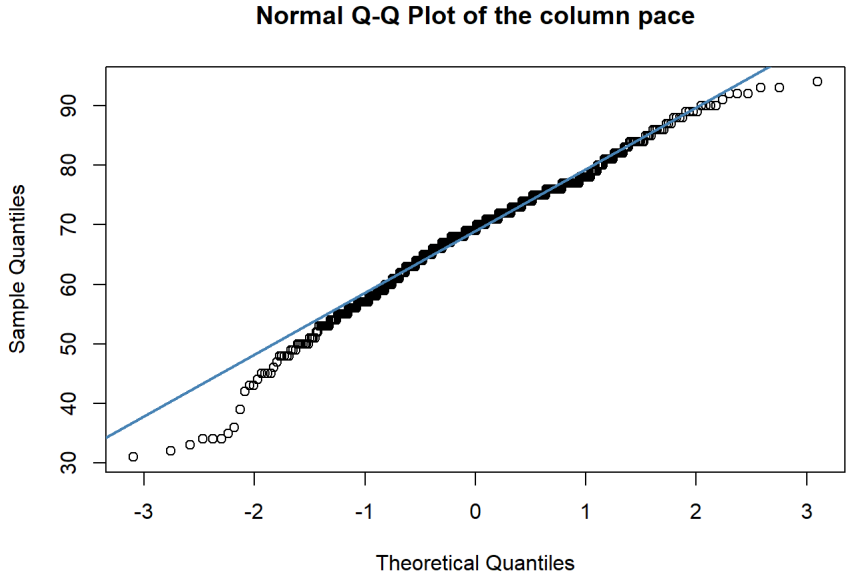

Normal Q-Q Plot of the column potential
Sample Quantiles
Theoretical Quantiles
55
60
65
70
75
80
85
90
-3
-2
-1
0
1
2
3

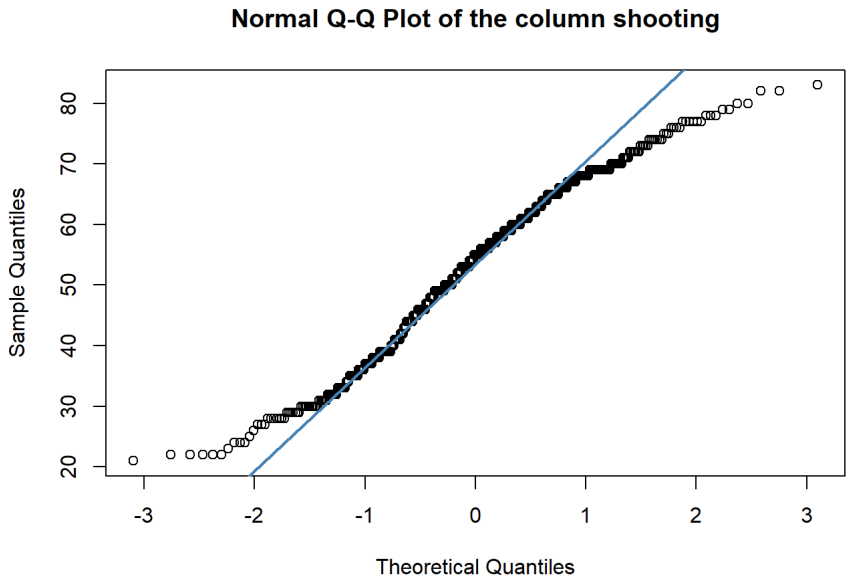

Normal Q-Q Plot of the column wage_eur_log
Sample Quantiles
Theoretical Quantiles
6
7
8
9
10
11
12
-3
-2
-1
0
1
2
3

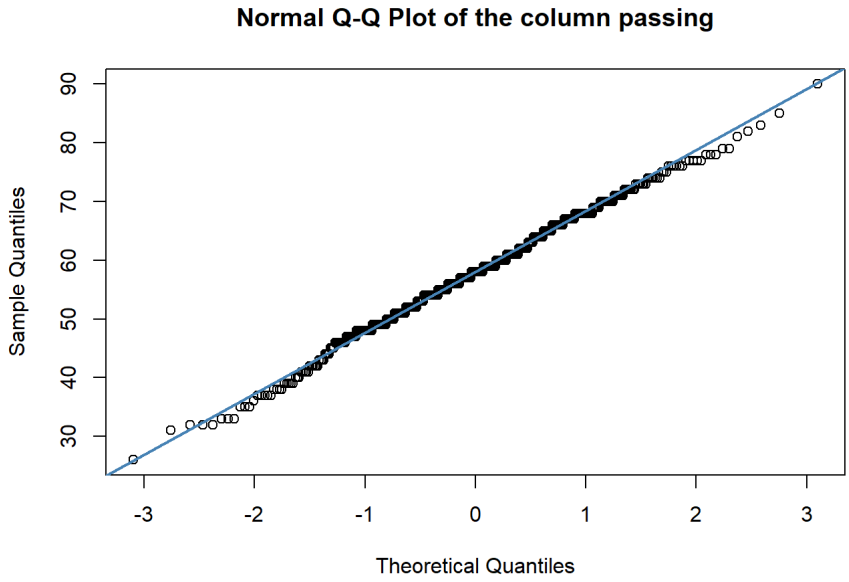

Normal Q-Q Plot of the column age
Sample Quantiles
Theoretical Quantiles
20
25
30
35
-3
-2
-1
0
1
2
3

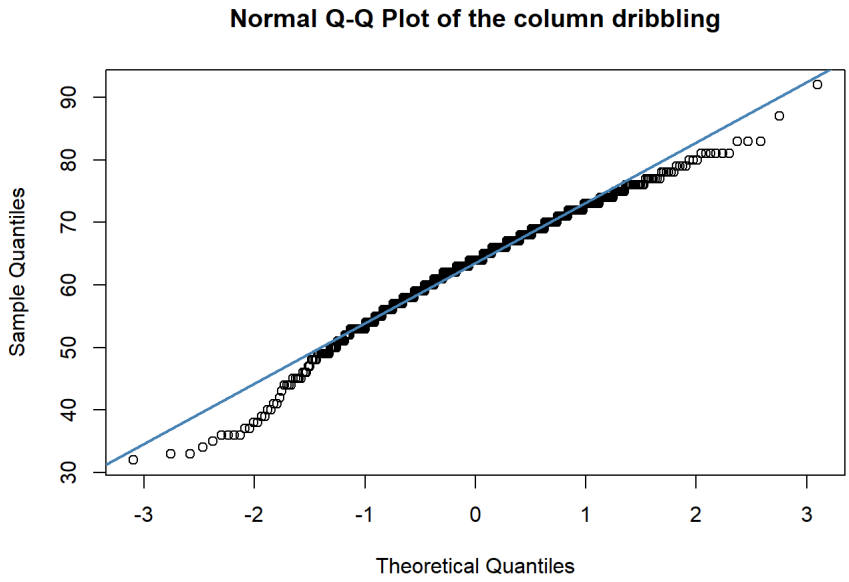

Normal Q-Q Plot of the column height_cm
Sample Quantiles
160
170
180
190
200
-3
-2
-1
0
1
2
3
Theoretical Quantiles

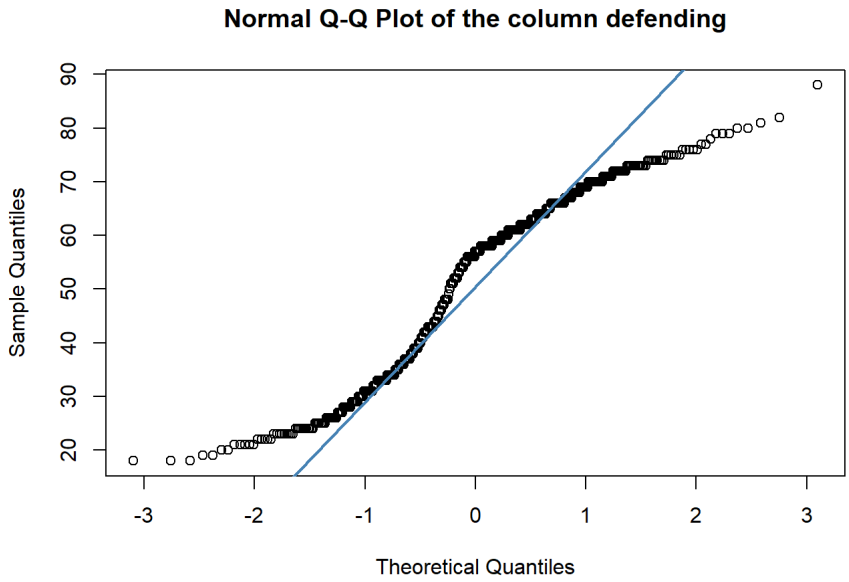

Normal Q-Q Plot of the column weight_kg
Sample Quantiles
60
70
80
90
-3
-2
-1
0
1
2
3
Theoretical Quantiles

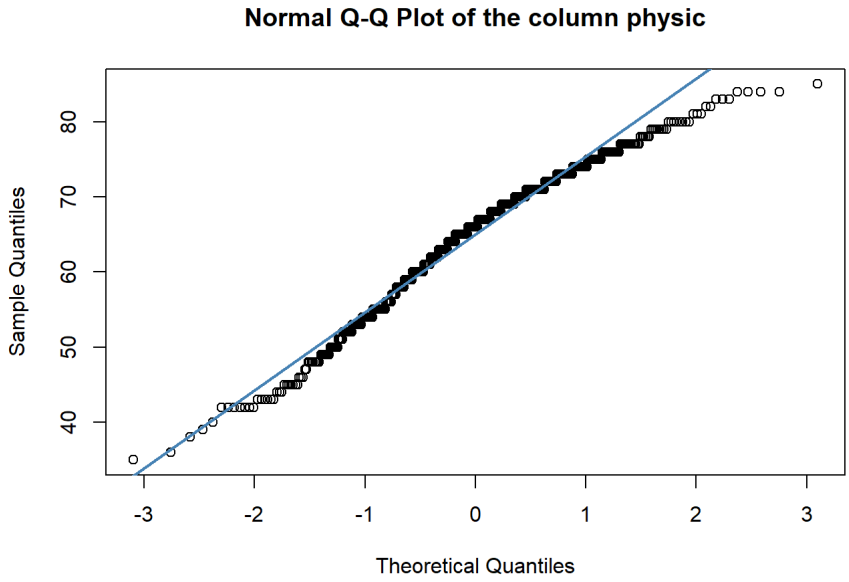

Normal Q-Q Plot of the column pace
Sample Quantiles
30
40
50
60
70
80
90
-3
-2
-1
0
1
2
3
Theoretical Quantiles

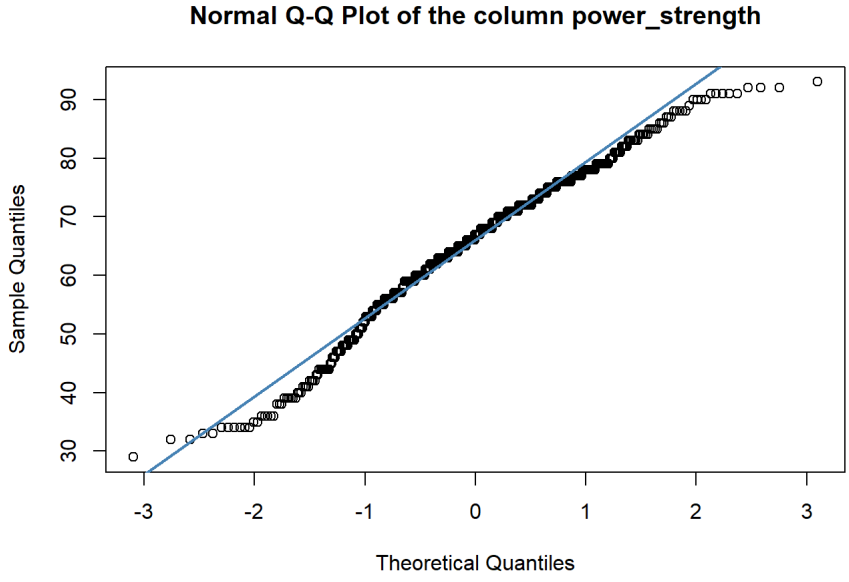

Normal Q-Q Plot of the column shooting
Sample Quantiles
Theoretical Quantiles
20
30
40
50
60
70
80
-3
-2
-1
0
1
2
3

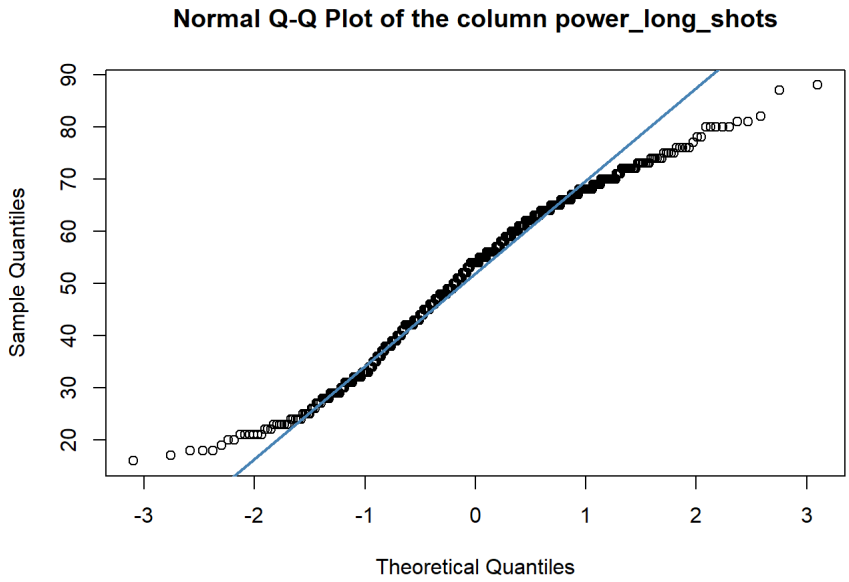

Normal Q-Q Plot of the column passing
Sample Quantiles
Theoretical Quantiles
30
40
50
60
70
80
90
-3
-2
-1
0
1
2
3


Normal Q-Q Plot of the column dribbling
Sample Quantiles
Theoretical Quantiles
30
40
50
60
70
80
90
-3
-2
-1
0
1
2
3


Normal Q-Q Plot of the column defending
Sample Quantiles
Theoretical Quantiles
20
30
40
50
60
70
80
90
-3
-2
-1
0
1
2
3


Normal Q-Q Plot of the column physic
Sample Quantiles
Theoretical Quantiles
40
50
60
70
80
-3
-2
-1
0
1
2
3


Normal Q-Q Plot of the column power_strength
Sample Quantiles
Theoretical Quantiles
30
40
50
60
70
80
90
-3
-2
-1
0
1
2
3




```
for (col in numeric_columns){
  message = paste("Shapiro-Wilk normality test for the column", col)
  print(message)
  shapiro_test <- shapiro.test(df[,col])
  print(shapiro_test[2])
  if (shapiro_test[2] < 0.05){
    shap_message <- paste("This is a very small p-value, we reject the null hypothesis that", col, "in this sample is normal
ly distributed")
    print(shap_message)
  }
  else{
    print(paste("We failed to reject the null hypothesis. Column", col, "is normally distributed"))
  }
  cat("-----------------------------------------------------------------------------------------------\n\n\n")
}
```

```
## [1] "Shapiro-Wilk normality test for the column potential"
## $p.value
## [1] 0.006241612
##
## [1] "This is a very small p-value, we reject the null hypothesis that potential in this sample is normally distributed"
## ------------------------------------------------------------------------------------------
##
##
## [1] "Shapiro-Wilk normality test for the column wage_eur_log"
## $p.value
## [1] 0.00000000002264849
##
## [1] "This is a very small p-value, we reject the null hypothesis that wage_eur_log in this sample is normally distribute
d"
## ------------------------------------------------------------------------------------------
##
##
## [1] "Shapiro-Wilk normality test for the column age"
## $p.value
## [1] 0.0000002278677
##
## [1] "This is a very small p-value, we reject the null hypothesis that age in this sample is normally distributed"
## ------------------------------------------------------------------------------------------
##
##
## [1] "Shapiro-Wilk normality test for the column height_cm"
## $p.value
## [1] 0.09240383
##
## [1] "We failed to reject the null hypothesis. Column height_cm is normally distributed"
## ------------------------------------------------------------------------------------------
##
##
## [1] "Shapiro-Wilk normality test for the column weight_kg"
## $p.value
## [1] 0.02663471
##
## [1] "This is a very small p-value, we reject the null hypothesis that weight_kg in this sample is normally distributed"
## ------------------------------------------------------------------------------------------
##
##
## [1] "Shapiro-Wilk normality test for the column pace"
## $p.value
## [1] 0.000004189181
##
## [1] "This is a very small p-value, we reject the null hypothesis that pace in this sample is normally distributed"
## ------------------------------------------------------------------------------------------
##
##
## [1] "Shapiro-Wilk normality test for the column shooting"
## $p.value
## [1] 0.0000001774451
##
## [1] "This is a very small p-value, we reject the null hypothesis that shooting in this sample is normally distributed"
## ------------------------------------------------------------------------------------------
##
##
## [1] "Shapiro-Wilk normality test for the column passing"
## $p.value
## [1] 0.3160711
##
## [1] "We failed to reject the null hypothesis. Column passing is normally distributed"
## ------------------------------------------------------------------------------------------
##
##
## [1] "Shapiro-Wilk normality test for the column dribbling"
## $p.value
## [1] 0.0000006314925
##
## [1] "This is a very small p-value, we reject the null hypothesis that dribbling in this sample is normally distributed"
## ------------------------------------------------------------------------------------------
##
##
## [1] "Shapiro-Wilk normality test for the column defending"
## $p.value
## [1] 0.0000000000001639449
##
## [1] "This is a very small p-value, we reject the null hypothesis that defending in this sample is normally distributed"
## ------------------------------------------------------------------------------------------
##
##
## [1] "Shapiro-Wilk normality test for the column physic"
## $p.value
```

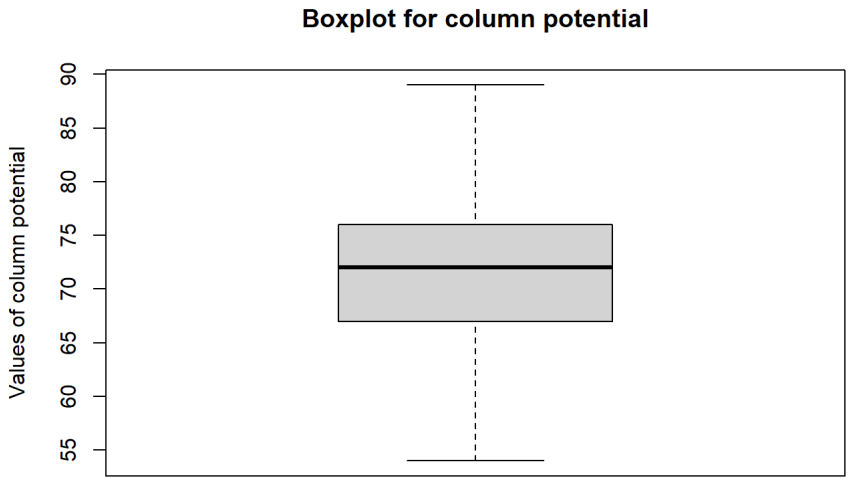
```
## [1] 0.00000003411265
##
## [1] "This is a very small p-value, we reject the null hypothesis that physic in this sample is normally distributed"
## ----------------------------------------------------------------------------------------------
##
##
## [1] "Shapiro-Wilk normality test for the column power_strength"
## $p.value
## [1] 0.0000006843208
##
## [1] "This is a very small p-value, we reject the null hypothesis that power_strength in this sample is normally distribut
ed"
## ----------------------------------------------------------------------------------------------
##
##
## [1] "Shapiro-Wilk normality test for the column power_long_shots"
## $p.value
## [1] 0.00000006352206
##
## [1] "This is a very small p-value, we reject the null hypothesis that power_long_shots in this sample is normally distrib
uted"
## ----------------------------------------------------------------------------------------------
```

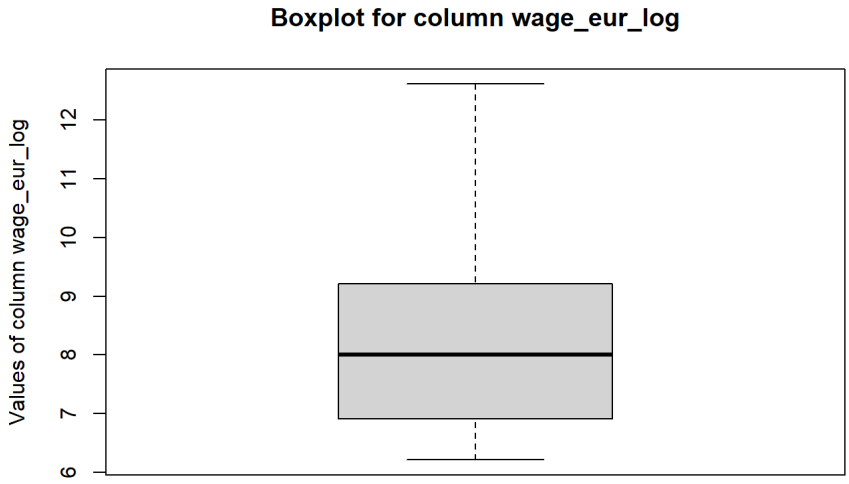
After graphing the qq-plots and peforming shapiro tests, we can conclude the following:

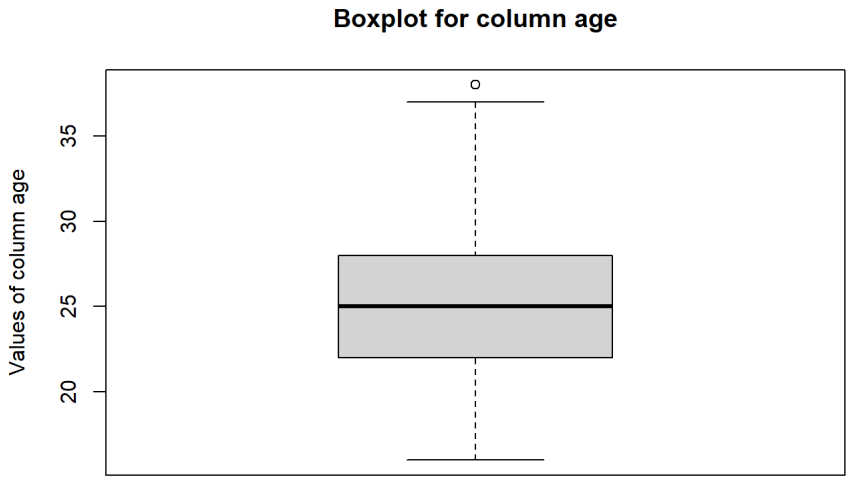
- Columns `power_long_shots`, `power_strength`, `physic`, `physic`, `defending`, `dribbling`, `shooting`, `pace`, `weight_kg`, `age`, `wage_eur_log`, `potential` are not normally distributed.
- Columns `passing`, `height_cm` are normally distributed.

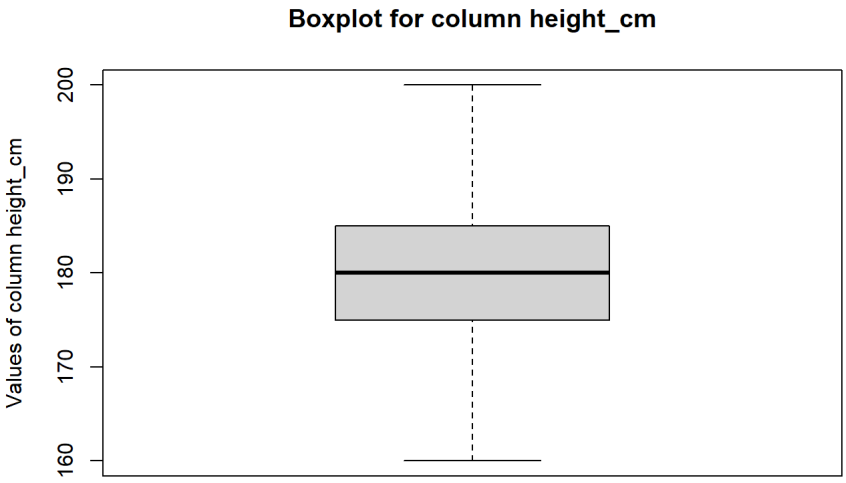
Boxplots:

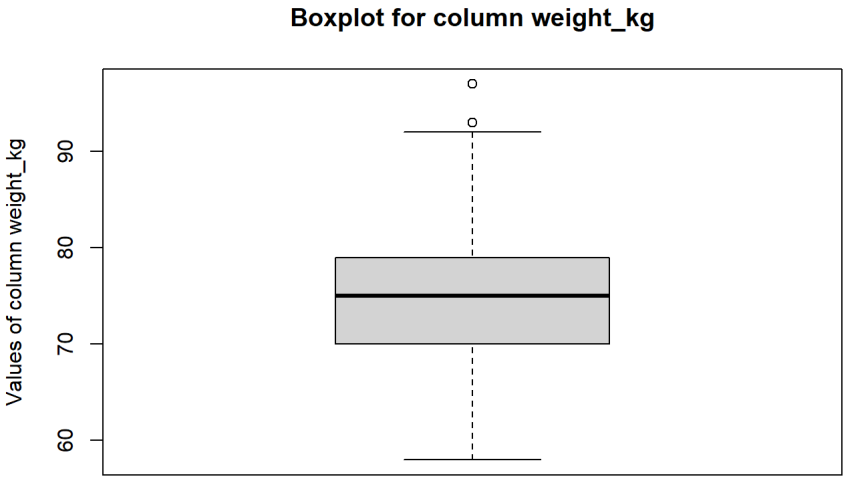
```
for (col in numeric_columns){
  title = paste("Boxplot for column", col)
  y_axis = paste("Values of column", col)
  boxplot(df[, col], main=title, ylab=y_axis)
}
```

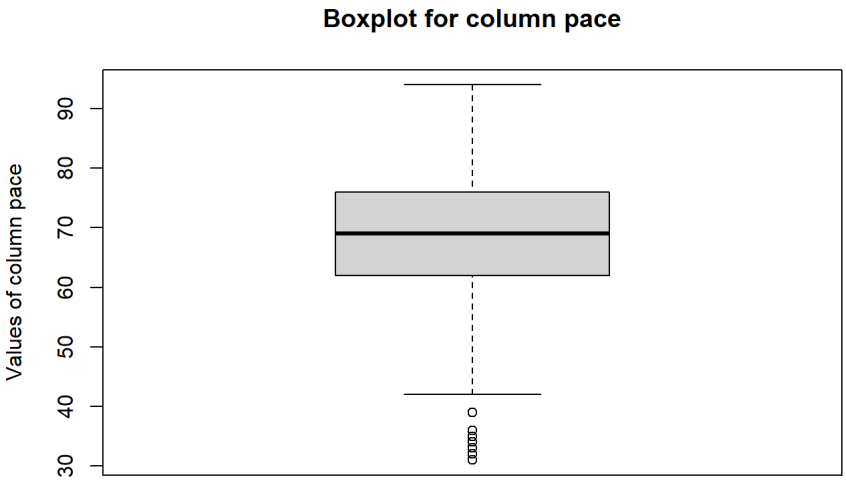

Boxplot for column potential
Values of column potential
55
60
65
70
75
80
85
90

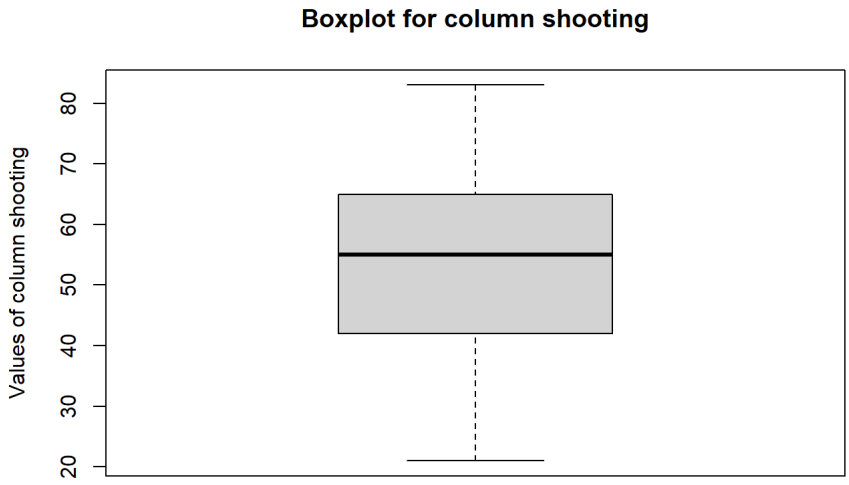

Boxplot for column wage_eur_log
Values of column wage_eur_log
6
7
8
9
10
11
12

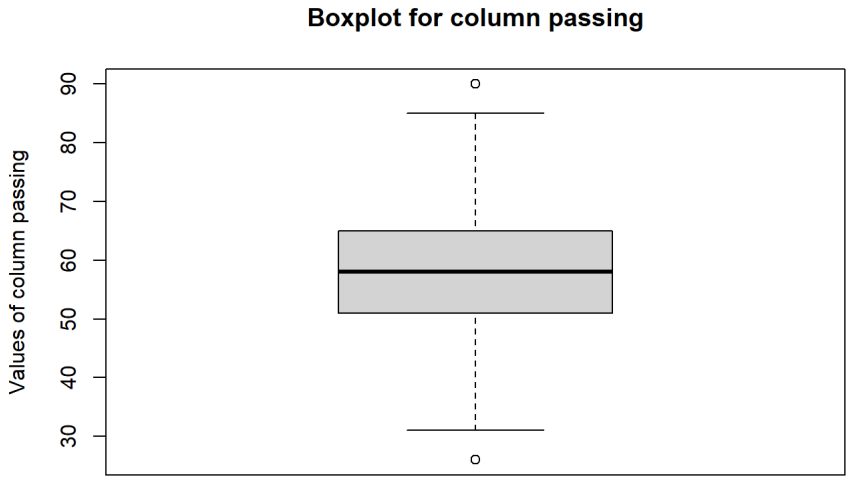

Boxplot for column age
Values of column age
20
25
30
35

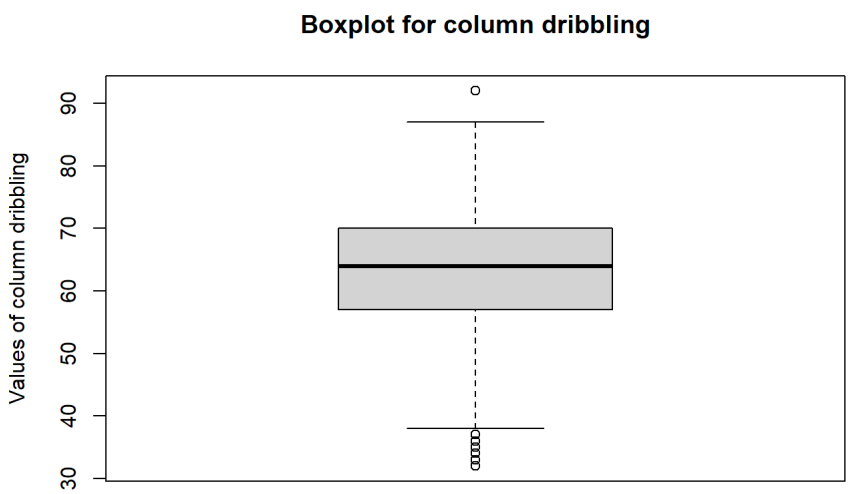

Boxplot for column height_cm
Values of column height_cm
160
170
180
190
200

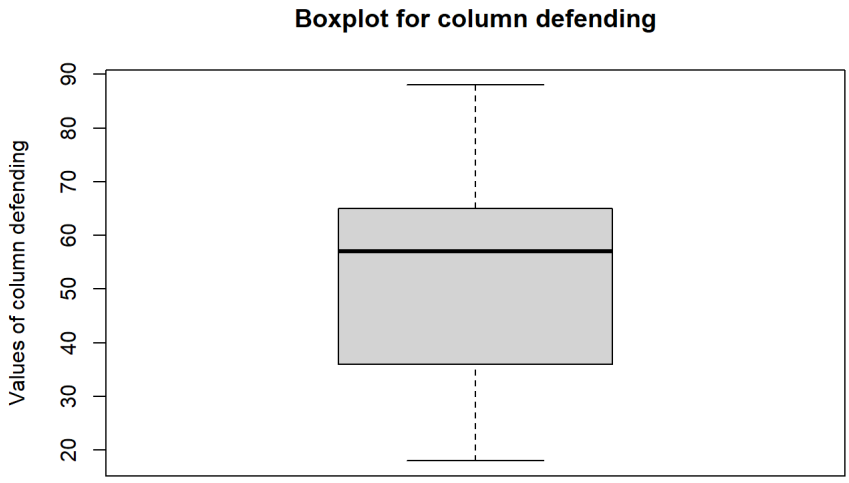

Boxplot for column weight_kg
Values of column weight_kg
60
70
80
90

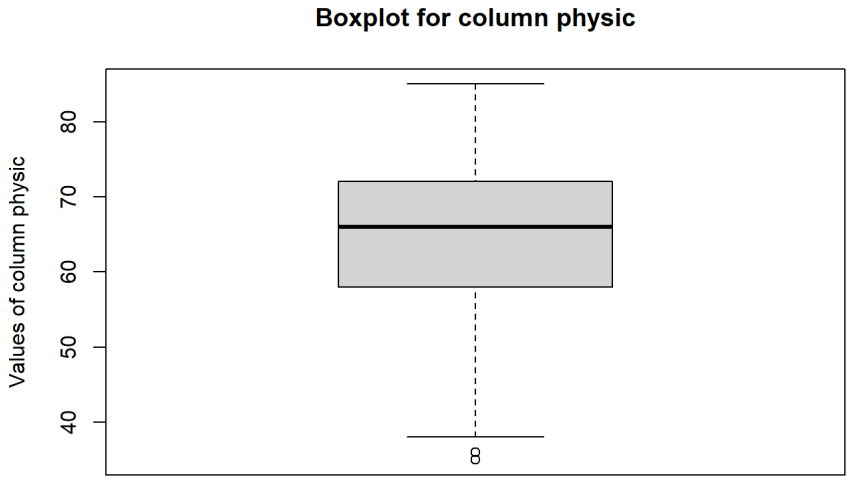

Boxplot for column pace
Values of column pace
30
40
50
60
70
80
90

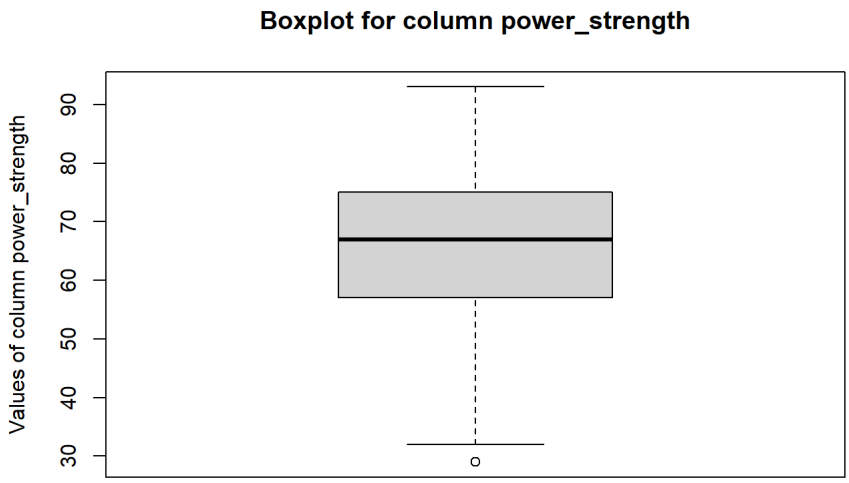

Boxplot for column shooting
Values of column shooting
20
30
40
50
60
70
80

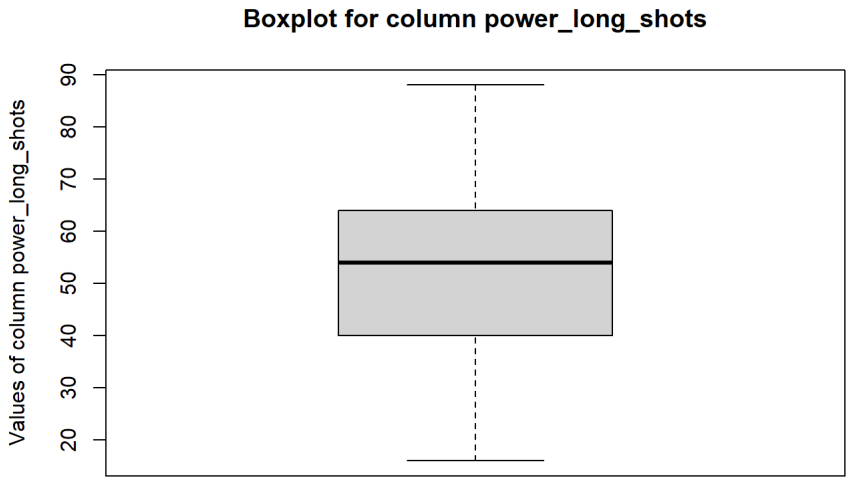

Boxplot for column passing
Values of column passing
30
40
50
60
70
80
90


Boxplot for column dribbling
Values of column dribbling
30
40
50
60
70
80
90


Boxplot for column defending
Values of column defending
20
30
40
50
60
70
80
90


Boxplot for column physic
Values of column physic
40
50
60
70
80


Boxplot for column power_strength
Values of column power_strength
30
40
50
60
70
80
90



- Even though the boxplots show that some columns do have outliers, those outliers do not represent any trouble when modeling as they seem to land on the range of normal values given by FIFA.

Now, we need to explored the categorical columns. First, wee need to create tables.

```
table_foot <- table(df$preferred_foot)
print(table_foot)
```

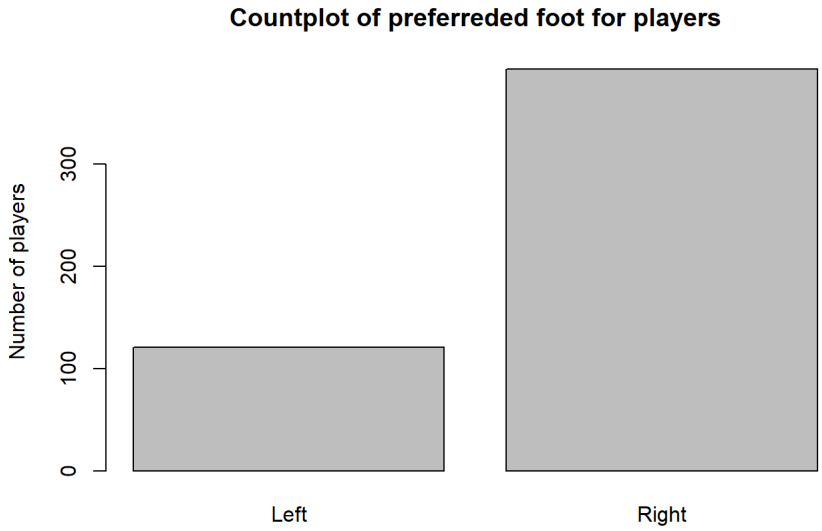
```
##
##  Left Right
##   121   393
```

```
table_wage <- table(df$high.wage.ind)
print(table_wage)
```

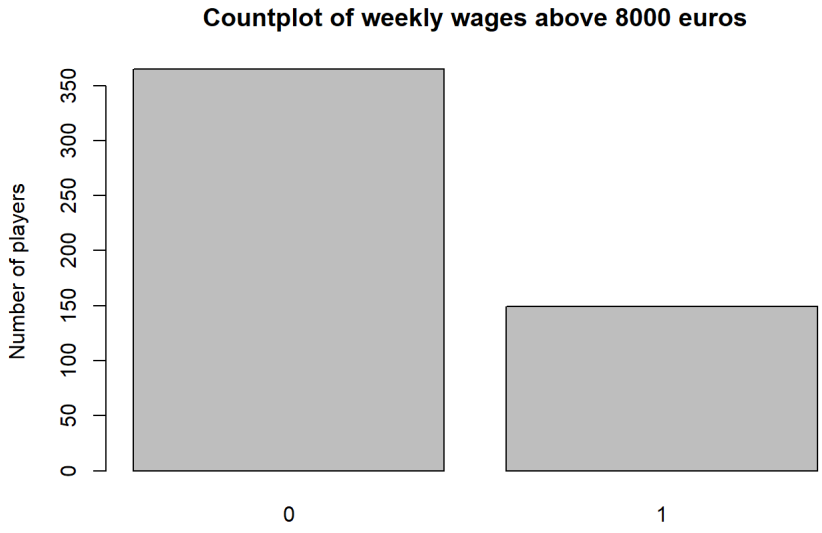
```
##
##   0   1
## 365 149
```

Now, we will create barplots

```
barplot(table_foot,
        names.arg=rownames(table_foot),
        ylab="Number of players",
        main="Countplot of preferreded foot for players")
```



```
barplot(table_wage,
        names.arg=rownames(table_wage),
        ylab="Number of players",
        main="Countplot of weekly wages above 8000 euros")
```

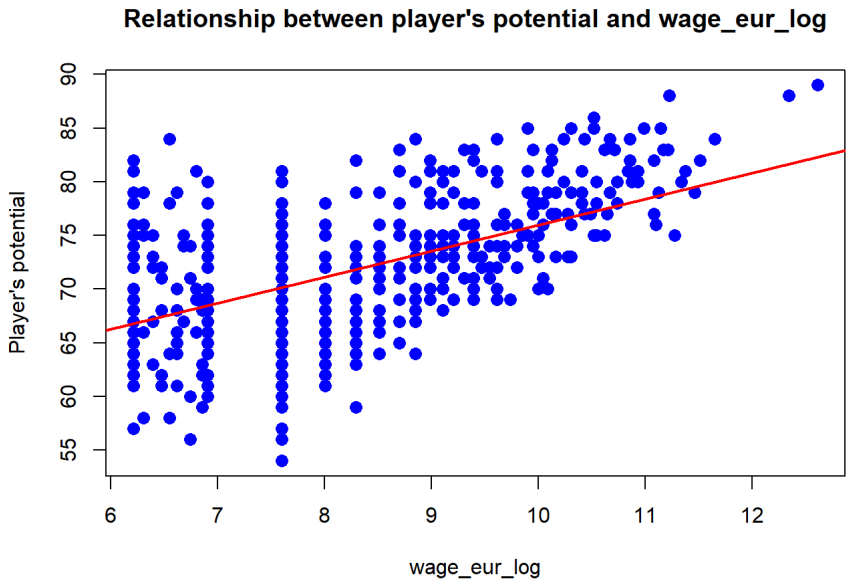


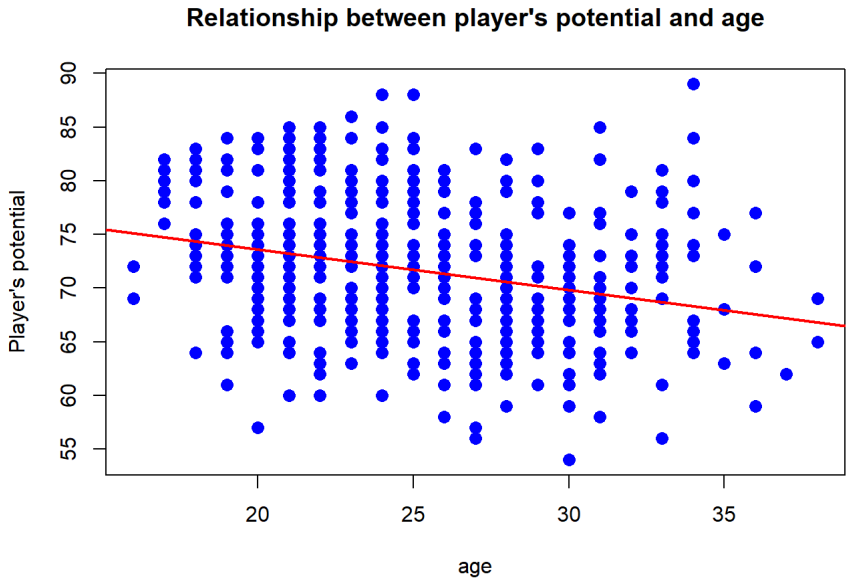
- It is clearly shown that both categorical columns are not balanced.

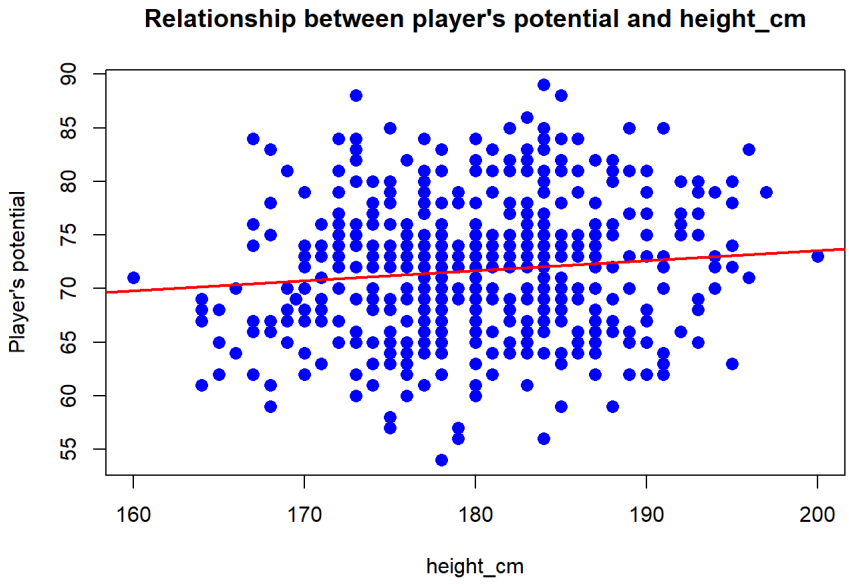
Bivariate EDA

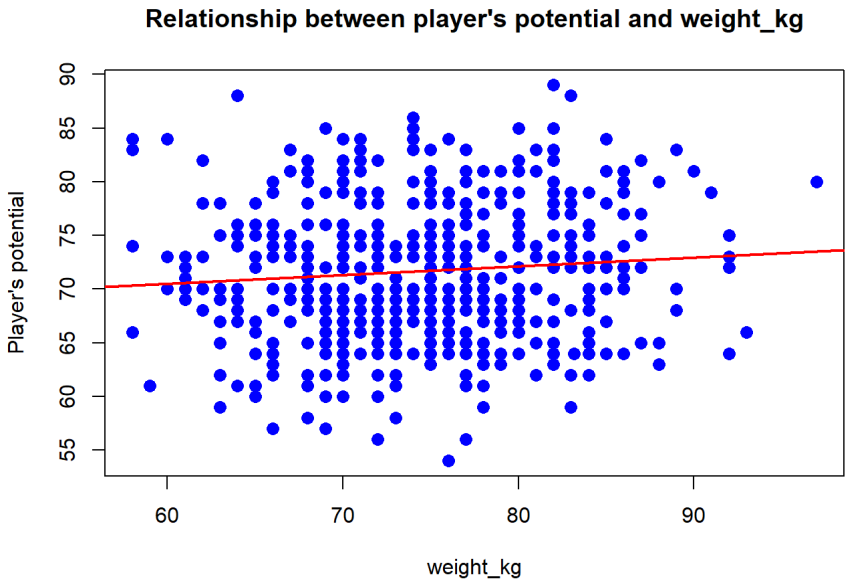
Regarding column `potential`

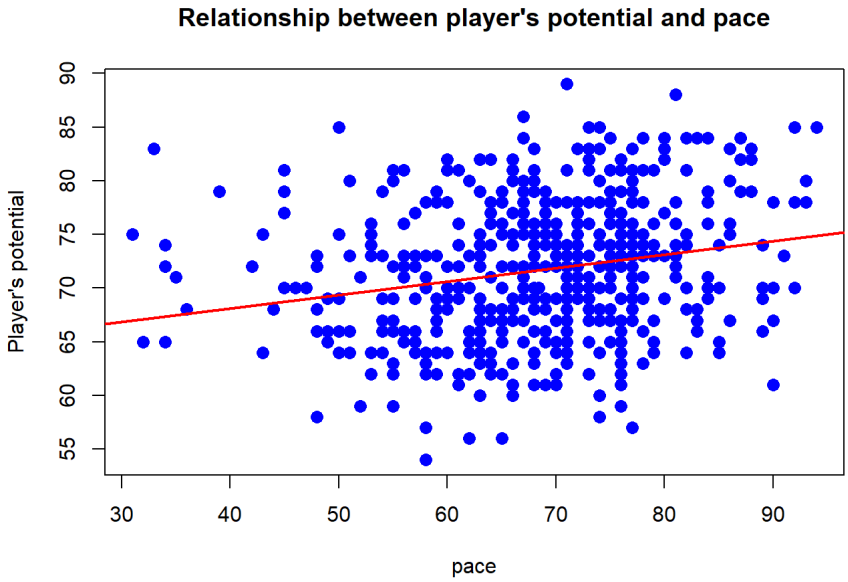
```
numeric_columns <- c("wage_eur_log",
                     "age",
                     "height_cm",
                     "weight_kg",
                     "pace",
                     "shooting",
                     "passing",
                     "dribbling",
                     "defending",
                     "physic",
                     "power_strength",
                     "power_long_shots")

for (col in numeric_columns){
  title = paste("Relationship between player's potential and", col)
  x_axis = col
  y_axis = "Player's potential"
  plot(df[,col],
       df[,"potential"],
       pch=16,
       cex=1.3,
       col="blue",
       ylab=y_axis,
       xlab=x_axis,
       main=title)

  abline(lm(df[,"potential"]~df[,col]),
         lwd=2,
         col="red")
}
```

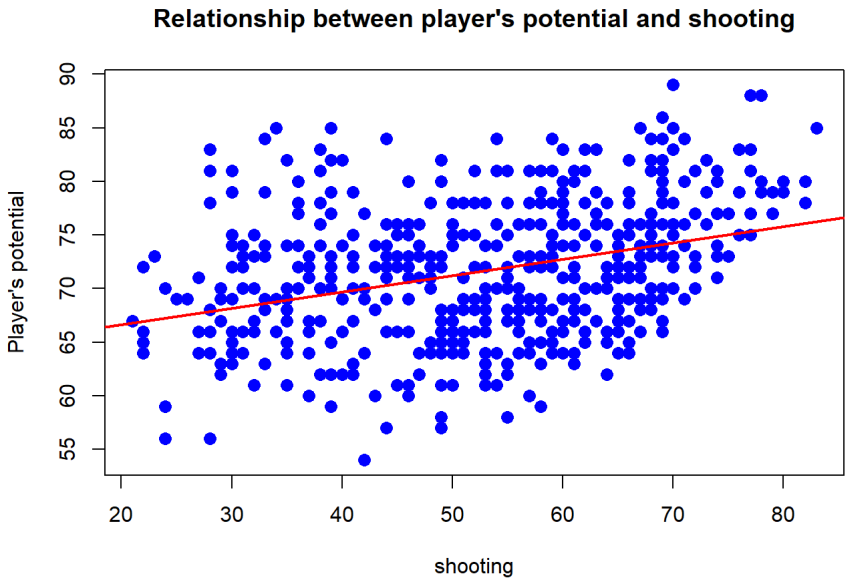

Relationship between player's potential and wage_eur_log
Player's potential
55
60
65
70
75
80
85
90
6
7
8
9
10
11
12
wage_eur_log

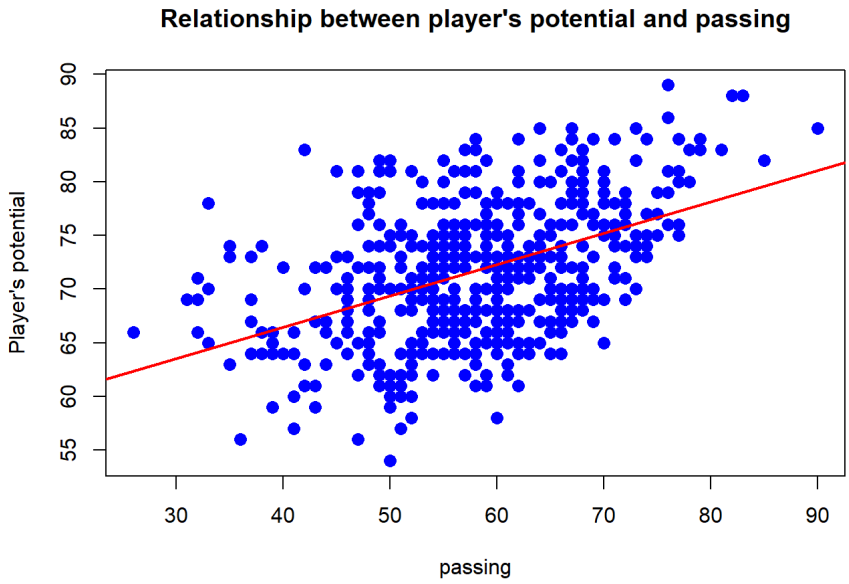

Relationship between player's potential and age
Player's potential
55
60
65
70
75
80
85
90
20
25
30
35
age

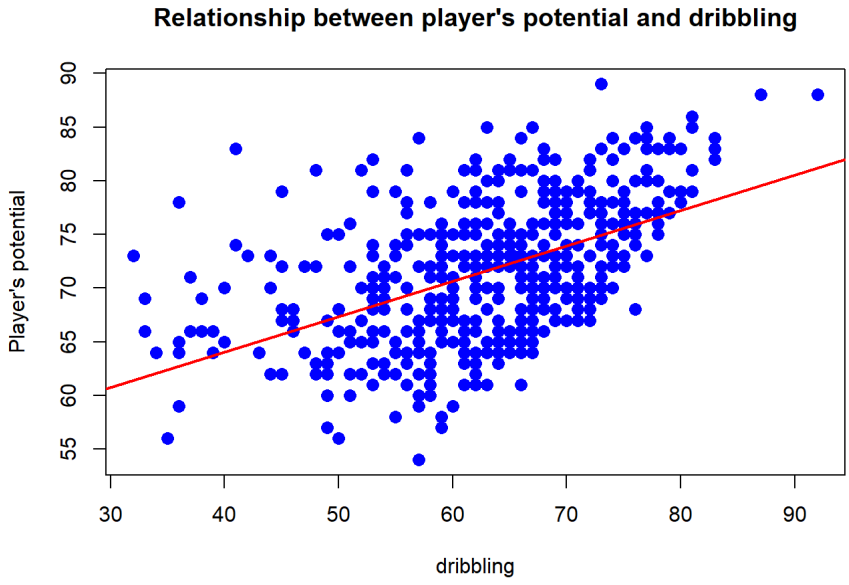

Relationship between player's potential and height_cm
Player's potential
55
60
65
70
75
80
85
90
160
170
180
190
200
height_cm

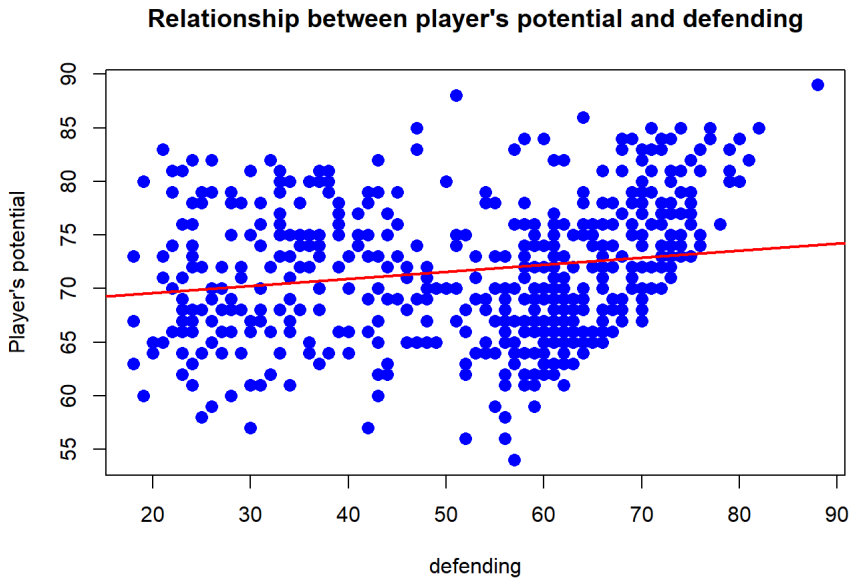

Relationship between player's potential and weight_kg
Player's potential
55
60
65
70
75
80
85
90
60
70
80
90
weight_kg

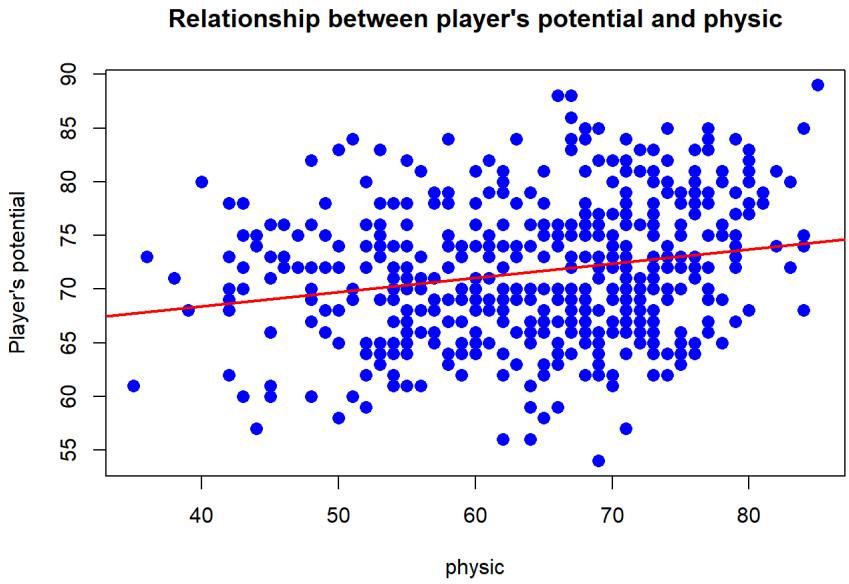

Relationship between player's potential and pace
Player's potential
55
60
65
70
75
80
85
90
30
40
50
60
70
80
90
pace

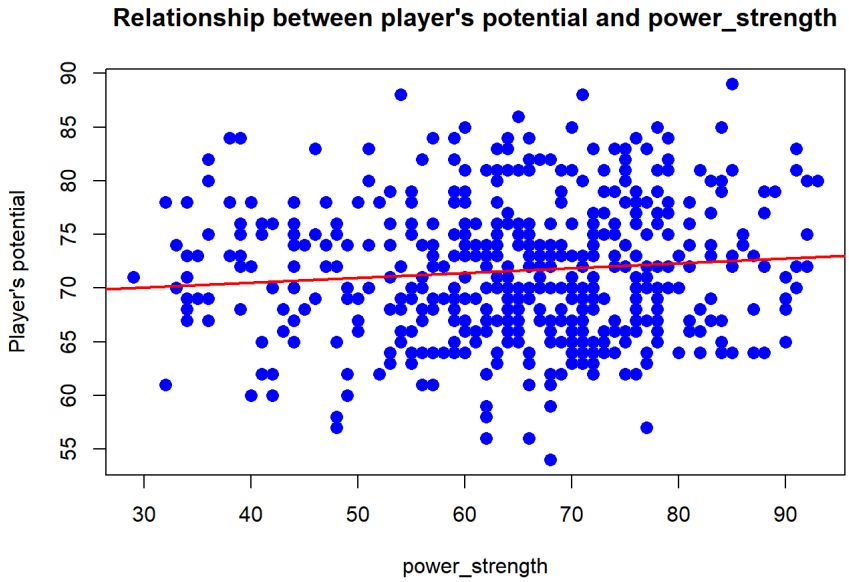

Relationship between player's potential and shooting
Player's potential
55
60
65
70
75
80
85
90
20
30
40
50
60
70
80
shooting

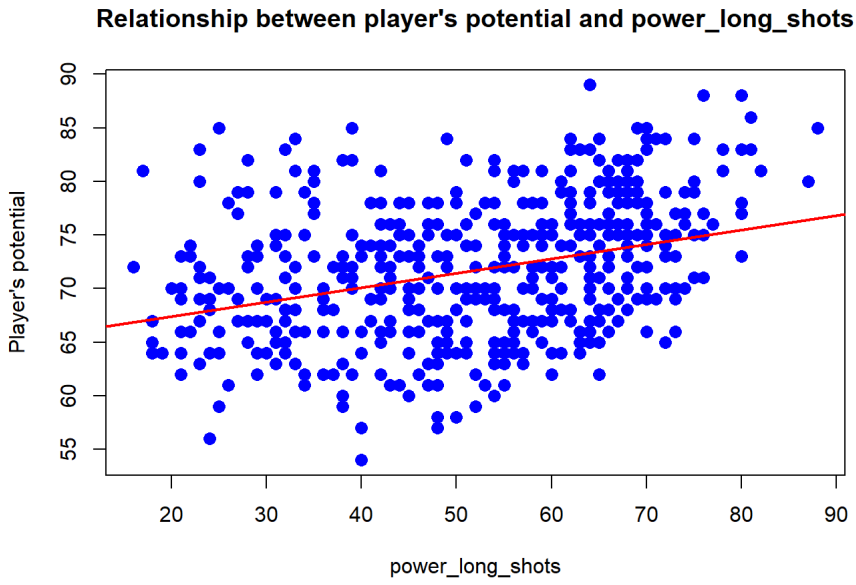

Relationship between player's potential and passing
Player's potential
passing
55
60
65
70
75
80
85
90
30
40
50
60
70
80
90


Relationship between player's potential and dribbling
Player's potential
dribbling
55
60
65
70
75
80
85
90
30
40
50
60
70
80
90


Relationship between player's potential and defending
Player's potential
defending
55
60
65
70
75
80
85
90
20
30
40
50
60
70
80
90







- `potential` seems to have a positive correlation with `wage_eur_log` .
- `potential` seems to have a quadratic relationship with `age` .
- `potential` seems to have no correlation with `height_cm` .
- `potential` seems to have no correlation with `weight_cm` .

- `potential` seems to have positive correlation with `pace` .
- `potential` seems to have positive correlation with `shooting` .
- `potential` seems to have positive correlation with `passing` .
- `potential` seems to have positive correlation with `dribbling` .
- `potential` seems to hava a quadratic relationship with `defending` .
- `potential` seems to have no correlation with `physic` .
- `potential` seems to have no correlation with `power_strength` .
- `potential` seems to have positive correlation with `power_long_shots` .

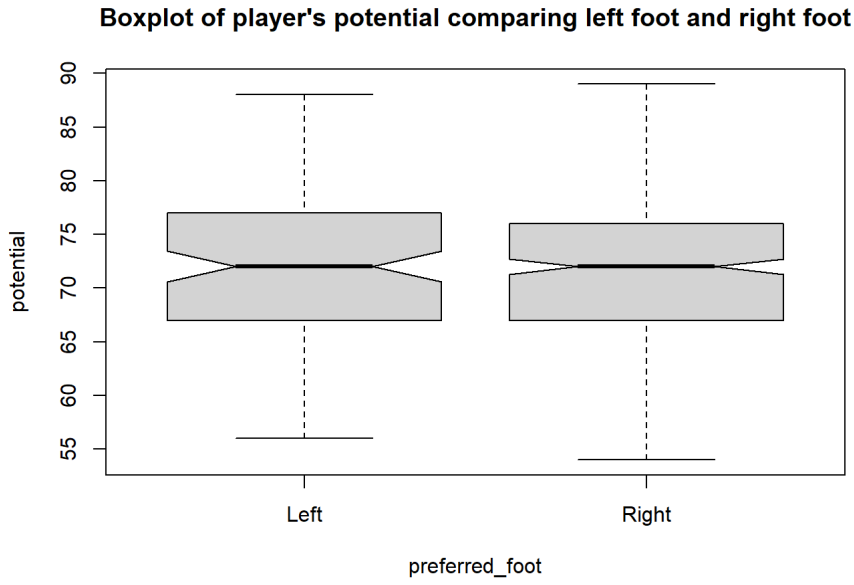
Making correlation tests:

```
for (col in numeric_columns){
  cor_test <- cor.test(df[,col], df[,"potential"], method="spearman", exact=FALSE)
  p_value <- cor_test[3]
  corr_value <- cor_test[4]
  print(p_value)
  if (p_value < 0.05){
    message <- paste("Since, we have a small p-value, we reject the null hypothesis and we say there is a relationship betwe
en player's potential and", col, ". The correlation between the 2 variables is", corr_value)
    print(message)
  } else{
    message <- paste("p-value is not small. We failed to reject the null hypothesis. There is no relationship between playe
r's potential and", col)
    print(message)
  }
  cat("-------------------------------------------------------------------------\n\n")
}
```

```
## $p.value
## [1] 0.0000000000000000000000000000000004812591
##
## [1] "Since, we have a small p-value, we reject the null hypothesis and we say there is a relationship between player's po
tential and wage_eur_log . The correlation between the 2 variables is c(rho = 0.507805698417417)"
## ------------------------------------------------------------------------
##
## $p.value
## [1] 0.000000000003466858
##
## [1] "Since, we have a small p-value, we reject the null hypothesis and we say there is a relationship between player's po
tential and age . The correlation between the 2 variables is c(rho = -0.28676537092926)"
## ------------------------------------------------------------------------
##
## $p.value
## [1] 0.0835688
##
## [1] "p-value is not small. We failed to reject the null hypothesis. There is no relationship between player's potential a
nd height_cm"
## ------------------------------------------------------------------------
##
## $p.value
## [1] 0.06292819
##
## [1] "p-value is not small. We failed to reject the null hypothesis. There is no relationship between player's potential a
nd weight_kg"
## ------------------------------------------------------------------------
##
## $p.value
## [1] 0.00000003994502
##
## [1] "Since, we have a small p-value, we reject the null hypothesis and we say there is a relationship between player's po
tential and pace . The correlation between the 2 variables is c(rho = 0.239257849172959)"
## ------------------------------------------------------------------------
##
## $p.value
## [1] 0.000000000000003338519
##
## [1] "Since, we have a small p-value, we reject the null hypothesis and we say there is a relationship between player's po
tential and shooting . The correlation between the 2 variables is c(rho = 0.338008261975809)"
## ------------------------------------------------------------------------
##
## $p.value
## [1] 0.00000000000000000000000001082434
##
## [1] "Since, we have a small p-value, we reject the null hypothesis and we say there is a relationship between player's po
tential and passing . The correlation between the 2 variables is c(rho = 0.439531262962747)"
## ------------------------------------------------------------------------
##
## $p.value
## [1] 0.0000000000000000000000000000000000006460536
##
## [1] "Since, we have a small p-value, we reject the null hypothesis and we say there is a relationship between player's po
tential and dribbling . The correlation between the 2 variables is c(rho = 0.526017679089538)"
## ------------------------------------------------------------------------
##
## $p.value
## [1] 0.000001855197
##
## [1] "Since, we have a small p-value, we reject the null hypothesis and we say there is a relationship between player's po
tential and defending . The correlation between the 2 variables is c(rho = 0.208524647149215)"
## ------------------------------------------------------------------------
##
## $p.value
## [1] 0.0000008207722
##
## [1] "Since, we have a small p-value, we reject the null hypothesis and we say there is a relationship between player's po
tential and physic . The correlation between the 2 variables is c(rho = 0.21543471838648)"
## ------------------------------------------------------------------------
##
## $p.value
## [1] 0.0206317
##
## [1] "Since, we have a small p-value, we reject the null hypothesis and we say there is a relationship between player's po
tential and power_strength . The correlation between the 2 variables is c(rho = 0.10207686183769)"
## ------------------------------------------------------------------------
##
## $p.value
## [1] 0.000000000000004010964
##
## [1] "Since, we have a small p-value, we reject the null hypothesis and we say there is a relationship between player's po
tential and power_long_shots . The correlation between the 2 variables is c(rho = 0.337081069986084)"
## ------------------------------------------------------------------------
```

Analyzing relationship between `potential` and `preferred_foot` .

```
#Anova
boxplot(potential~preferred_foot,
        data=df,
        notch=TRUE,
        main="Boxplot of player's potential comparing left foot and right foot")
```



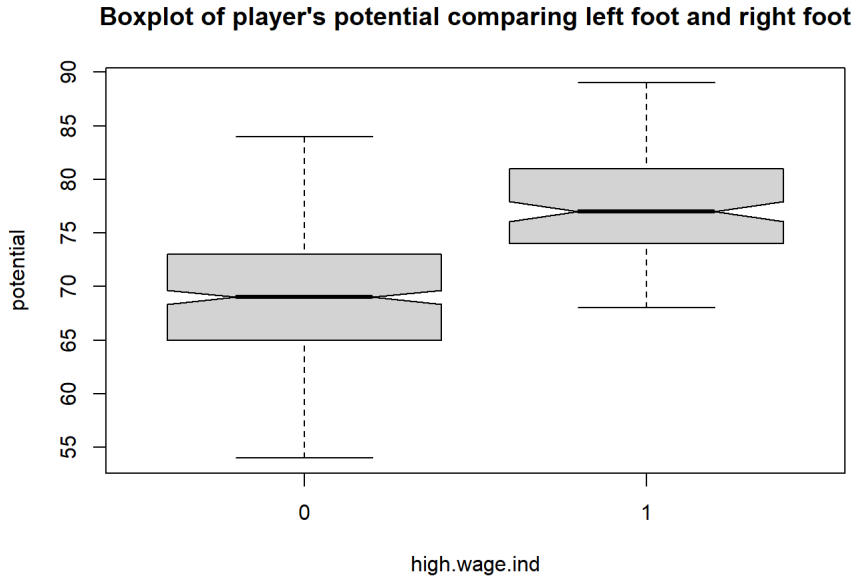

```
summary(aov(df$potential~df$preferred_foot))
```

```
##                   Df Sum Sq Mean Sq F value Pr(>F)
## df$preferred_foot   1     65   64.91   1.591  0.208
## Residuals         512  20882   40.78
```

- The boxplot shows that there is not statistical significance in the difference of medians. The ANOVA test shows that there is not statistical significance in the difference of means.

Analyzing relationship between `potential` and `high.wage.ind` .

```
#Anova
boxplot(potential~high.wage.ind,
        data=df,
        notch=TRUE,
        main="Boxplot of player's potential comparing left foot and right foot")
```



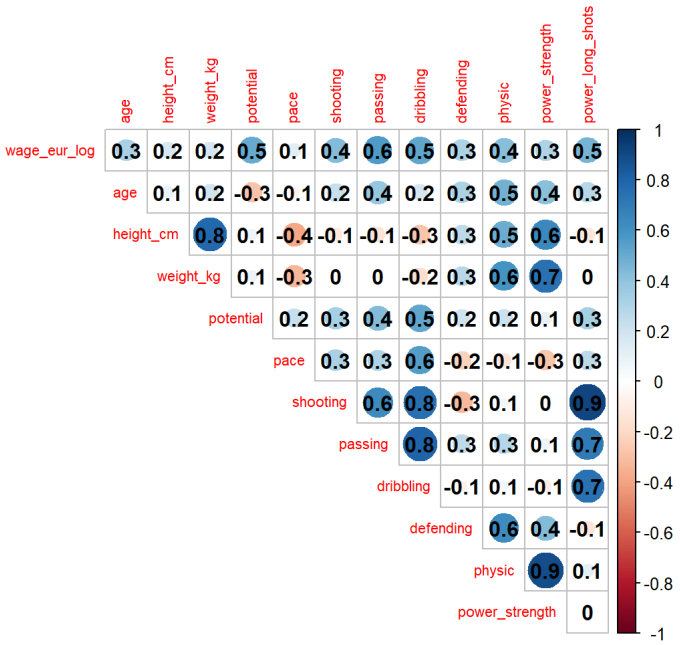

```
summary(aov(df$potential~df$high.wage.ind))
```

```
##                   Df Sum Sq Mean Sq F value             Pr(>F)
## df$high.wage.ind   1   6384    6384   224.4 <0.0000000000000002 ***
## Residuals        512  14563      28
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

The boxplot shows that there is statistical significance in the difference of medians. The ANOVA test shows that there is statistical significance in the difference of means.

Finding correlations to find issues of collinearity.

```
c_col <- c("wage_eur_log",
           "age",
           "height_cm",
           "weight_kg",
           "potential",
           "pace",
           "shooting",
           "passing",
           "dribbling",
           "defending",
           "physic",
           "power_strength",
           "power_long_shots")

df_cor <- subset(df, select=c_col)
cor_matrix <- cor(df_cor, method="spearman")
corrplot(cor_matrix, diag=F, type="upper", insig="p-value", number.digits=1, addCoef.col="black", tl.cex=0.7)
```

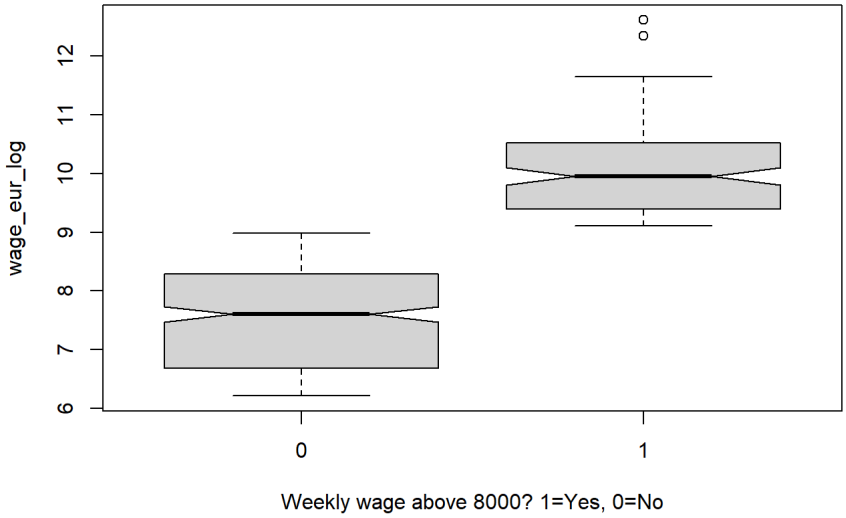


- Issues of collinearity could be found between `physic` and `power_strength`, between `dribbling` and both `shooting` and `passing`, between `power_long_shoots` and `shooting` and between `height_cm` and `weight_kg`.

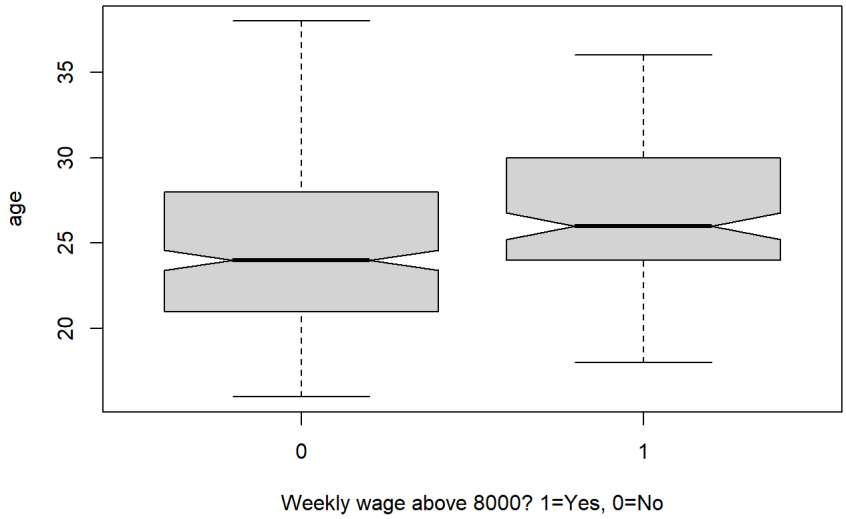
Regarding column `high.wage.ind`

```
c_col <- c("wage_eur_log",
           "age",
           "height_cm",
           "weight_kg",
           "potential",
           "pace",
           "shooting",
           "passing",
           "dribbling",
           "defending",
           "physic",
           "power_strength",
           "power_long_shots")

for (col in c_col){
  title = paste("Boxplot comparing values on", col, "between players with weekly wages below 8000 and above 8000")
  x_axis = "Weekly wage above 8000? 1=Yes, 0=No"
  y_axis = col
  boxplot(df[,col]~df[,"high.wage.ind"],
          data=df,
          notch=TRUE,
          main=title,
          xlab=x_axis,
          ylab=y_axis)
}
```

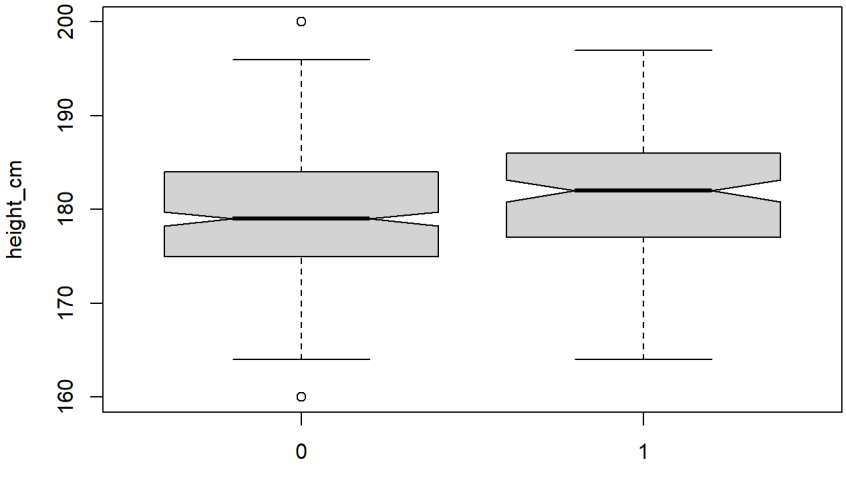
ıring values on wage_eur_log between players with weekly wages below 8(



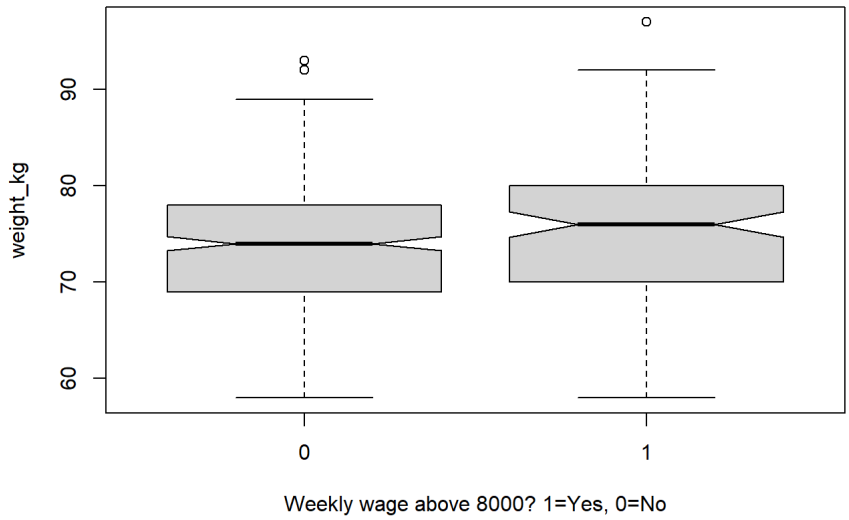
omparing values on age between players with weekly wages below 8000 ar



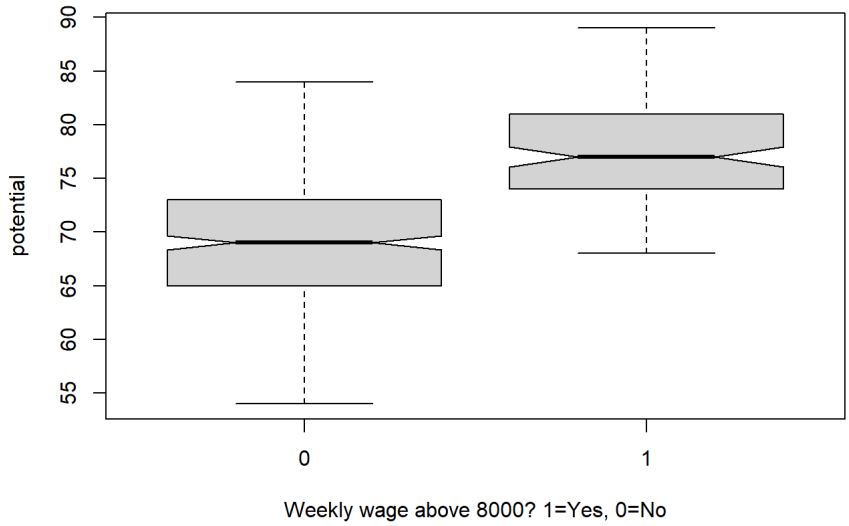
paring values on height_cm between players with weekly wages below 800



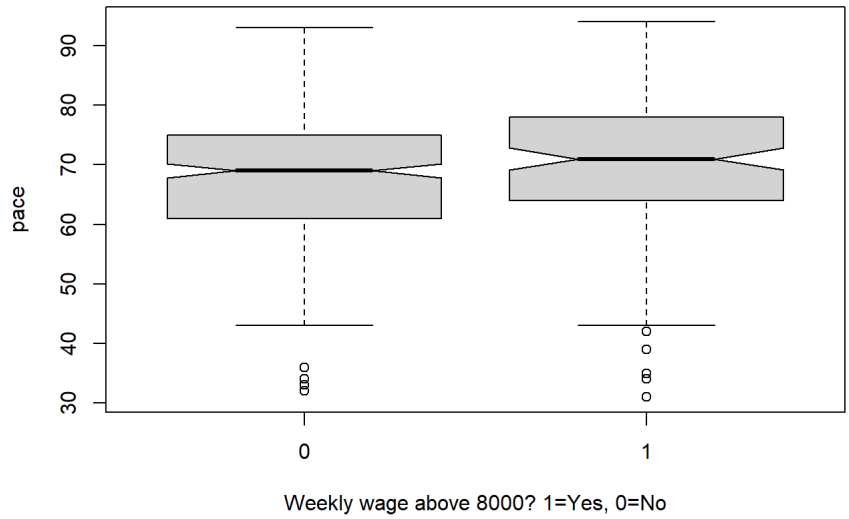
paring values on weight_kg between players with weekly wages below 800



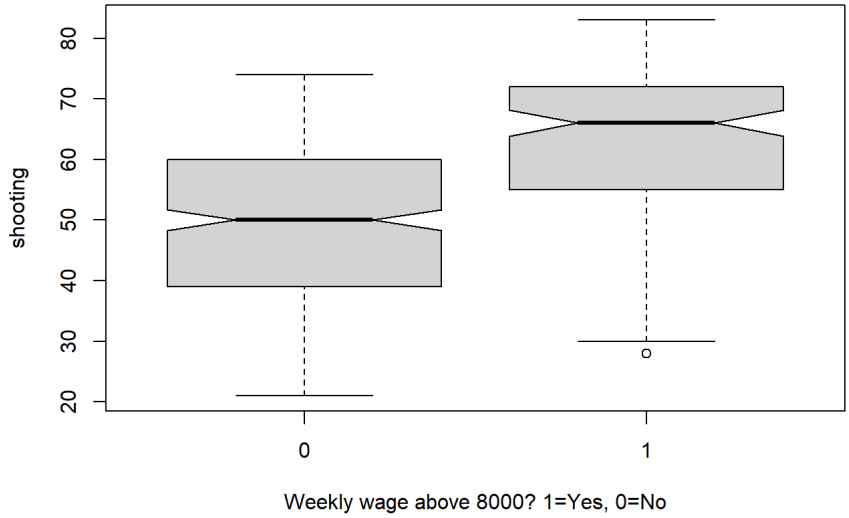
ıparing values on potential between players with weekly wages below 8000



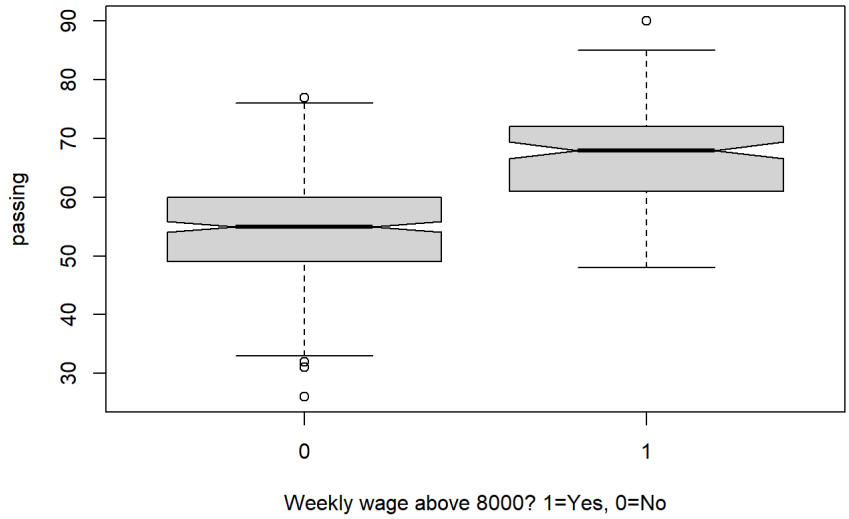
ɔmparing values on pace between players with weekly wages below 8000 a



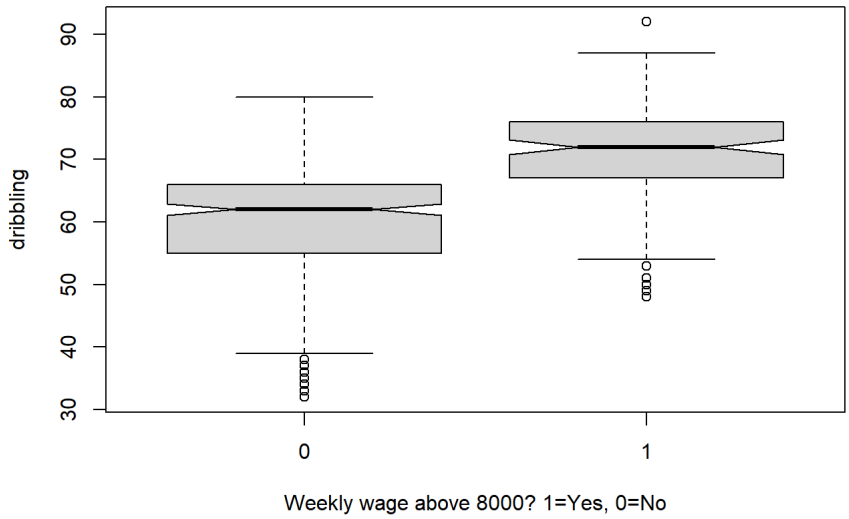
ıparing values on shooting between players with weekly wages below 8000



nparing values on passing between players with weekly wages below 8000

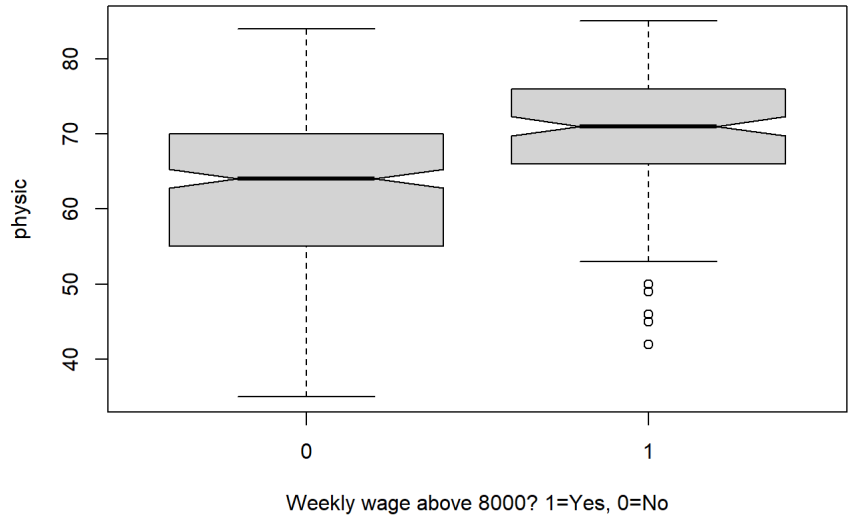


ıparing values on dribbling between players with weekly wages below 8000

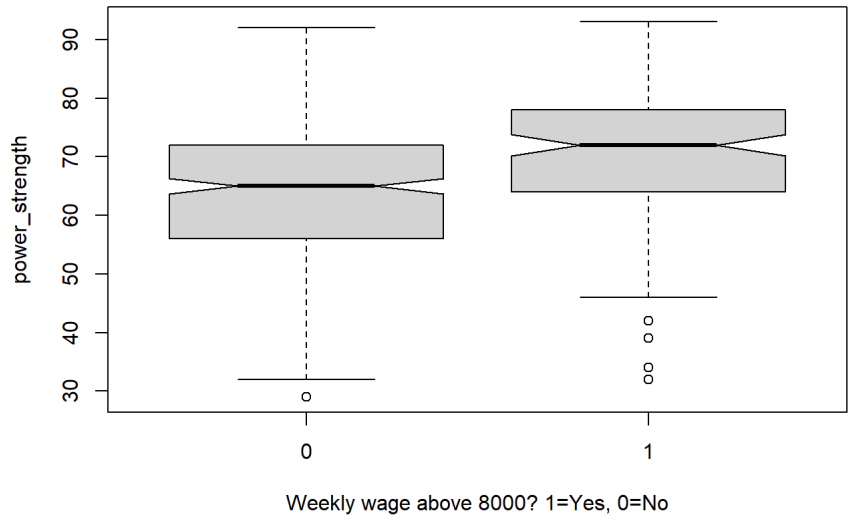


paring values on defending between players with weekly wages below 800

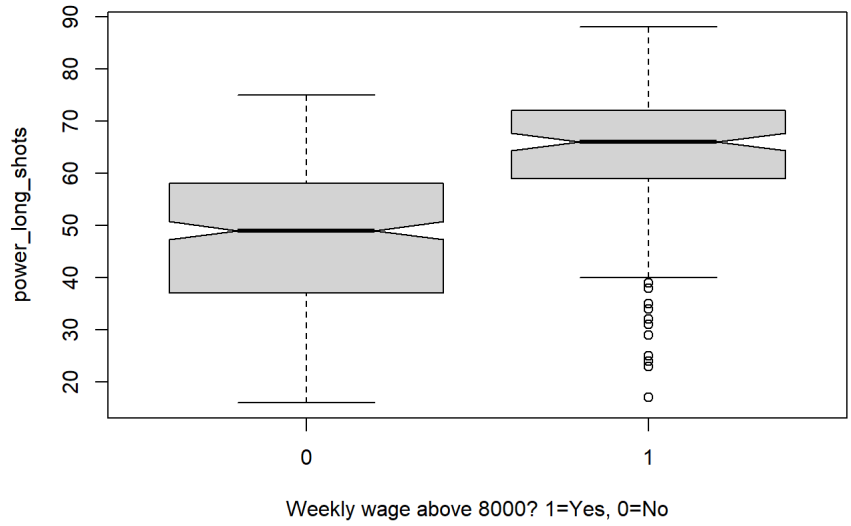
mparing values on physic between players with weekly wages below 8000 a



ring values on power_strength between players with weekly wages below 8



ng values on power_long_shots between players with weekly wages below



- Boxplots show that there is statistical significance in the difference of medians between `high.wage.ind` and the following columns: `wage_eur_log` (more than obvious since `high.wage.ind` is calculated from `wage_eur`), `age` (small), `height_cm` (barely), `potential`, `shooting`, `passing`, `dribbling`, `defending`, `physic`, `power_strength` and `power_long_shots`. Boxplots show that there is no statistical significance in the difference in medians between `high.wage.ind` and the following columns: `weight_kg` and `pace`.

Performing two sided t-test to find if there is significance difference in means between `high.wage.ind` and the other numerical columns.

```
for (col in c_col){
  message <- paste("Results for column", col)
  print(message)
  ttest_result <- t.test(df[,col]~df[,"high.wage.ind"], alternative="two.sided")
  print(ttest_result[5])
  print(ttest_result[3])
  if (ttest_result[3]<0.05){
    message_1 <- "P-value is very small, we reject null-hypothesis and we say there is statistical significance in the diffe
rence of means."
    print(message_1)
  }
  else {
    message_2 <- "P-value is big. We failed to reject the null-hypotheis. There is no evidence that there is statistical sig
nificance in the diffence of means."
    print(message_2)
  }
  cat("-------------------------------------------------------------------------------------------------------------------\n\n
\n")
}
```

```
## [1] "Results for column wage_eur_log"
## $estimate
## mean in group 0 mean in group 1
##        7.479091       10.031782
##
## $p.value
## [1] 0.0000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000021548
36
##
## [1] "P-value is very small, we reject null-hypothesis and we say there is statistical significance in the difference of m
eans."
## ----------------------------------------------------------------------------------------------------------------
##
##
## [1] "Results for column age"
## $estimate
## mean in group 0 mean in group 1
##        24.53151        26.65772
##
## $p.value
## [1] 0.0000003435215
##
## [1] "P-value is very small, we reject null-hypothesis and we say there is statistical significance in the difference of m
eans."
## ----------------------------------------------------------------------------------------------------------------
##
##
## [1] "Results for column height_cm"
## $estimate
## mean in group 0 mean in group 1
##        179.4644        181.3221
##
## $p.value
## [1] 0.004946459
##
## [1] "P-value is very small, we reject null-hypothesis and we say there is statistical significance in the difference of m
eans."
## ----------------------------------------------------------------------------------------------------------------
##
##
## [1] "Results for column weight_kg"
## $estimate
## mean in group 0 mean in group 1
##        73.72099        75.84564
##
## $p.value
## [1] 0.001767726
##
## [1] "P-value is very small, we reject null-hypothesis and we say there is statistical significance in the difference of m
eans."
## ----------------------------------------------------------------------------------------------------------------
##
##
## [1] "Results for column potential"
## $estimate
## mean in group 0 mean in group 1
##        69.41370        77.18121
##
## $p.value
## [1] 0.0000000000000000000000000000000000000000001120415
##
## [1] "P-value is very small, we reject null-hypothesis and we say there is statistical significance in the difference of m
eans."
## ----------------------------------------------------------------------------------------------------------------
##
##
## [1] "Results for column pace"
## $estimate
## mean in group 0 mean in group 1
##        67.79286        69.86577
##
## $p.value
## [1] 0.08131733
##
## [1] "P-value is big. We failed to reject the null-hypotheis. There is no evidence that there is statistical significance
in the diffence of means."
## ----------------------------------------------------------------------------------------------------------------
##
##
## [1] "Results for column shooting"
## $estimate
## mean in group 0 mean in group 1
##        49.30411        62.19463
```

```
##
## $p.value
## [1] 0.00000000000000000001383496
##
## [1] "P-value is very small, we reject null-hypothesis and we say there is statistical significance in the difference of m
eans."
## ------------------------------------------------------------------------------------------------------------------------
##
##
## [1] "Results for column passing"
## $estimate
## mean in group 0 mean in group 1
##        54.11781        66.83893
##
## $p.value
## [1] 0.0000000000000000000000000000000000000008534964
##
## [1] "P-value is very small, we reject null-hypothesis and we say there is statistical significance in the difference of m
eans."
## ------------------------------------------------------------------------------------------------------------------------
##
##
## [1] "Results for column dribbling"
## $estimate
## mean in group 0 mean in group 1
##        59.96520        70.62416
##
## $p.value
## [1] 0.00000000000000000000000000000007168102
##
## [1] "P-value is very small, we reject null-hypothesis and we say there is statistical significance in the difference of m
eans."
## ------------------------------------------------------------------------------------------------------------------------
##
##
## [1] "Results for column defending"
## $estimate
## mean in group 0 mean in group 1
##        49.00000        57.72483
##
## $p.value
## [1] 0.0000005692312
##
## [1] "P-value is very small, we reject null-hypothesis and we say there is statistical significance in the difference of m
eans."
## ------------------------------------------------------------------------------------------------------------------------
##
##
## [1] "Results for column physic"
## $estimate
## mean in group 0 mean in group 1
##        62.46027        69.86577
##
## $p.value
## [1] 0.000000000000001545619
##
## [1] "P-value is very small, we reject null-hypothesis and we say there is statistical significance in the difference of m
eans."
## ------------------------------------------------------------------------------------------------------------------------
##
##
## [1] "Results for column power_strength"
## $estimate
## mean in group 0 mean in group 1
##        63.30411        70.20805
##
## $p.value
## [1] 0.000000007551257
##
## [1] "P-value is very small, we reject null-hypothesis and we say there is statistical significance in the difference of m
eans."
## ------------------------------------------------------------------------------------------------------------------------
##
##
## [1] "Results for column power_long_shots"
## $estimate
## mean in group 0 mean in group 1
##        47.45753        62.54362
##
## $p.value
## [1] 0.00000000000000000000002527995
##
## [1] "P-value is very small, we reject null-hypothesis and we say there is statistical significance in the difference of m
```

```
eans."
## ------------------------------------------------------------------------------------------------------------------
```

Columns `wage_eur_log`, `age`, `height_cm`, `weight_cm`, `potential`, `shooting`, `passing`, `dribbling`, `defending`, `physic`, `power_strength` and `power_long_shots` have shown to have statistical significance in their difference of means between players that have weekly payments less than 8000 and above 8000. Column `pace` has shows to not have statistical significance in the difference of means.

Looking for a relationship between `preferred_foot` and `high.wage.ind`:

```
table_wage_foot <- table(df$high.wage.ind, df$preferred_foot)
table_wage_foot
```

```
##
##     Left Right
##   0   85   280
##   1   36   113
```

```
chisq.test(table_wage_foot)
```

```
##
##  Pearson's Chi-squared test with Yates' continuity correction
##
## data:  table_wage_foot
## X-squared = 0.0094454, df = 1, p-value = 0.9226
```

P-value is 0.9226, therefore failed to reject the null hypothesis and we say there is no relationship between `preferred_foot` and `high.wage.ind`.

2.3 Additional insights and issues

Highlight potential further issues or insights uncovered in 2.2. This might include follow up to findings from your initial EDA. We accept that the boundary between 2.2 and 2.3 is somewhat arbitrary so use your judgement and maximize good structure and readability. (5 marks)

No further issues where found during the EDA, all issues were found and fixed during the data cleaning.

3. Modelling

3.1 Build a model for player potential

Given the research question (i.e., player potential) outline an analysis plan that incorporates/references any findings from the data cleaning (1.3) and EDA (2.2) (5 marks). Use R to build a suitable model (10 marks).
NB Submissions where suitable models do not have good fit due to the nature of the data will not be penalized.

- From the EDA process, I have found that column `wage_eur` follows a logarithmic distribution. Therefore, a logarithmic transformation was applied to make the column more suitable for usage in the model. The transformation was stored in the `wage_eur_log` column. [6]
- From the EDA process, I have found that columns `power_long_shots`, `power_strength`, `physic`, `defending`, `dribbling`, `shooting`, `pace`, `weight_kg`, `wage_eur_log`, `age`, `potential` do not follow a normal distribution. Only columns `height_cm`, and `passing` follow a normal distribution. Therefore, when trying to find correlation between the columns, the Spearman method was used.
- In order to avoid collinearity, features that are highly correlated (>=0.8) between each other, will not be used. `height_cm` and `weight_kg` are highly correlated, for this case `height_cm` will be used because it follows a normal distribution. `dribbling` is highly correlated with `shooting` and `passing`, for this case we will use `dibbling` and remove `shooting` and `passing` in the model. In this way, we use one feature instead of two and make the model simplier, and also because dribbling has more correlation with `potential` than the other 2 features. `physic` and `power_strength` are also highly correlated, I will use `physic` and not use `power_strength` because `physic` has a higher correlation with `potential` than `power_strength` with `potential`. In summary, columns `weight_kg`, `shooting`, `passing`, `power_strenght` will not be used in the model.
- Using ANOVA and boxplots with notches, we have found that there is not statistical significance in median and mean of player's potential between players who preferred right foot and left foot. Therefore, the column `preferred_foot` will not be used in the model. On the other hand, it was found that there was a statistical signficance in median and mean of player's potential between players whose weekly wages are above 8000 euros and those below 8000 euros. However, the column `high.wage.ind` is taken from the `wage_eur` column which does not bring extra information, it only confirms that `wage_eur_log` is a good feature for our model. If `high.wage.ind` is used, an interaction problem with `wage_eur_log` will surged.
- `age` and `defending` may have a quadratic relationships on potential as seen as on the linear regressions performed on the EDA.
- First I will build a complex model and take note of the significant coefficients. Then I will build a very simple model and start adding significant coefficients found in the complex model.

model0

```
mr_model0 <- lm(potential ~ wage_eur_log*age*height_cm*pace*dribbling*defending*physic*power_long_shots +
                            I(wage_eur_log^2)+
                            I(age^2) +
                            I(height_cm^2) +
                            I(pace^2) +
                            I(dribbling^2) +
                            I(defending^2) +
                            I(physic^2) +
                            I(power_long_shots^2), data=df)
summary(mr_model0)
```

```
##
## Call:
## lm(formula = potential ~ wage_eur_log * age * height_cm * pace *
##     dribbling * defending * physic * power_long_shots + I(wage_eur_log^2) +
##     I(age^2) + I(height_cm^2) + I(pace^2) + I(dribbling^2) +
##     I(defending^2) + I(physic^2) + I(power_long_shots^2), data = df)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -6.8602 -0.9462 -0.0291  0.9572  6.1651
##
## Coefficients: (1 not defined because of singularities)
##                                                                          Estimate
## (Intercept)                                                          427405.034363127954
## wage_eur_log                                                         -40301.401010620160
## age                                                                   32839.128403441748
## height_cm                                                             -2067.053106715392
## pace                                                                   2233.809102541663
## dribbling                                                            -14884.233093490198
## defending                                                             -7296.221614994129
## physic                                                                -8818.663226317689
## power_long_shots                                                     -25832.658239883862
## I(wage_eur_log^2)                                                        0.690066755844
## I(age^2)                                                                 0.074923577237
## I(height_cm^2)                                                           0.000072019611
## I(pace^2)                                                                0.002342111916
## I(dribbling^2)                                                           0.009574794368
## I(defending^2)                                                           0.005052403570
## I(physic^2)                                                              0.002876070656
## I(power_long_shots^2)                                                    0.000385620411
## wage_eur_log:age                                                     -4643.973125183560
## wage_eur_log:height_cm                                                 164.469514853294
## age:height_cm                                                         -190.875866432536
## wage_eur_log:pace                                                     -251.577958288820
## age:pace                                                              -773.111244189751
## height_cm:pace                                                         -18.408316336233
## wage_eur_log:dribbling                                                1837.433287764628
## age:dribbling                                                         -192.306002859222
## height_cm:dribbling                                                     78.751112293509
## pace:dribbling                                                          68.925695086549
## wage_eur_log:defending                                                 905.845026878650
## age:defending                                                         -640.004713376964
## height_cm:defending                                                     38.536618968758
## pace:defending                                                         -69.832757523856
## dribbling:defending                                                    287.583337501923
## wage_eur_log:physic                                                    816.443769894335
## age:physic                                                            -399.005091641252
## height_cm:physic                                                        42.122136683417
## pace:physic                                                             -7.376064034482
## dribbling:physic                                                       252.834595842318
## defending:physic                                                       131.167522657696
## wage_eur_log:power_long_shots                                         2725.573886186652
## age:power_long_shots                                                   196.587233110556
## height_cm:power_long_shots                                             141.765502009691
## pace:power_long_shots                                                  217.153729202686
## dribbling:power_long_shots                                             535.555278311568
## defending:power_long_shots                                             418.840459733903
## physic:power_long_shots                                                435.949900754030
## wage_eur_log:age:height_cm                                              27.670895722173
## wage_eur_log:age:pace                                                   95.748419523381
## wage_eur_log:height_cm:pace                                              2.473345916236
## age:height_cm:pace                                                       4.489834544539
## wage_eur_log:age:dribbling                                              24.138134960707
## wage_eur_log:height_cm:dribbling                                        -9.391899196477
## age:height_cm:dribbling                                                  1.147638897488
## wage_eur_log:pace:dribbling                                            -11.560601876349
## age:pace:dribbling                                                       7.982890090867
## height_cm:pace:dribbling                                                -0.298437568790
## wage_eur_log:age:defending                                              82.886588920742
## wage_eur_log:height_cm:defending                                        -4.430593494107
## age:height_cm:defending                                                  3.533507513222
## wage_eur_log:pace:defending                                              3.719355543005
## age:pace:defending                                                      15.559338855313
## height_cm:pace:defending                                                 0.437914808316
## wage_eur_log:dribbling:defending                                       -37.557164060069
## age:dribbling:defending                                                  3.029865001191
## height_cm:dribbling:defending                                           -1.573908641575
## pace:dribbling:defending                                                -1.064875329101
## wage_eur_log:age:physic                                                 60.370931720731
## wage_eur_log:height_cm:physic                                           -3.373219686841
## age:height_cm:physic                                                     2.436804292044
## wage_eur_log:pace:physic                                                 1.455268190336
## age:pace:physic                                                         10.465294651960
```

```
## height_cm:pace:physic                                    0.165325966696
## wage_eur_log:dribbling:physic                          -30.424028890896
## age:dribbling:physic                                     2.104039576387
## height_cm:dribbling:physic                              -1.307304495445
## pace:dribbling:physic                                   -1.371776214836
## wage_eur_log:defending:physic                          -14.814773499251
## age:defending:physic                                     8.737357573090
## height_cm:defending:physic                              -0.664239090650
## pace:defending:physic                                    0.928964529540
## dribbling:defending:physic                              -4.490455155564
## wage_eur_log:age:power_long_shots                       -1.969179879260
## wage_eur_log:height_cm:power_long_shots                -14.636278360019
## age:height_cm:power_long_shots                          -1.044826669618
## wage_eur_log:pace:power_long_shots                     -24.259833937985
## age:pace:power_long_shots                                2.746395431753
## height_cm:pace:power_long_shots                         -1.167692774238
## wage_eur_log:dribbling:power_long_shots                -61.529809179846
## age:dribbling:power_long_shots                          -8.259402359958
## height_cm:dribbling:power_long_shots                    -2.956533479249
## pace:dribbling:power_long_shots                         -5.039526955630
## wage_eur_log:defending:power_long_shots                -46.434719022420
## age:defending:power_long_shots                          -1.480816794065
## height_cm:defending:power_long_shots                    -2.357297670236
## pace:defending:power_long_shots                         -2.925917162302
## dribbling:defending:power_long_shots                    -9.541379712589
## wage_eur_log:physic:power_long_shots                   -44.683526963194
## age:physic:power_long_shots                             -5.065746000559
## height_cm:physic:power_long_shots                       -2.361370506880
## pace:physic:power_long_shots                            -3.674735397104
## dribbling:physic:power_long_shots                       -8.691655623448
## defending:physic:power_long_shots                       -6.755120442155
## wage_eur_log:age:height_cm:pace                         -0.569396854116
## wage_eur_log:age:height_cm:dribbling                    -0.156593029568
## wage_eur_log:age:pace:dribbling                         -0.869442651129
## wage_eur_log:height_cm:pace:dribbling                    0.048709290732
## age:height_cm:pace:dribbling                            -0.046764250088
## wage_eur_log:age:height_cm:defending                    -0.471419600369
## wage_eur_log:age:pace:defending                         -1.788464253776
## wage_eur_log:height_cm:pace:defending                   -0.032253231126
## age:height_cm:pace:defending                            -0.087109872400
## wage_eur_log:age:dribbling:defending                    -0.341385760460
## wage_eur_log:height_cm:dribbling:defending              0.200097748995
## age:height_cm:dribbling:defending                       -0.016003839951
## wage_eur_log:pace:dribbling:defending                    0.235256399100
## age:pace:dribbling:defending                            -0.156033764796
## height_cm:pace:dribbling:defending                       0.005187886891
## wage_eur_log:age:height_cm:physic                       -0.374664684422
## wage_eur_log:age:pace:physic                            -1.318082428827
## wage_eur_log:height_cm:pace:physic                      -0.028363420057
## age:height_cm:pace:physic                               -0.062480467285
## wage_eur_log:age:dribbling:physic                       -0.276144740434
## wage_eur_log:height_cm:dribbling:physic                  0.152142993267
## age:height_cm:dribbling:physic                          -0.014409410323
## wage_eur_log:pace:dribbling:physic                       0.208501141767
## age:pace:dribbling:physic                               -0.109541114481
## height_cm:pace:dribbling:physic                          0.005776045847
## wage_eur_log:age:defending:physic                       -1.174821714638
## wage_eur_log:height_cm:defending:physic                  0.069235342143
## age:height_cm:defending:physic                          -0.049636991435
## wage_eur_log:pace:defending:physic                      -0.055335121516
## age:pace:defending:physic                               -0.226013261879
## height_cm:pace:defending:physic                         -0.006442295215
## wage_eur_log:dribbling:defending:physic                  0.571161208370
## age:dribbling:defending:physic                          -0.044774097852
## height_cm:dribbling:defending:physic                     0.024023080452
## pace:dribbling:defending:physic                          0.016342244766
## wage_eur_log:age:height_cm:power_long_shots             -0.005630508917
## wage_eur_log:age:pace:power_long_shots                  -0.491737407209
## wage_eur_log:height_cm:pace:power_long_shots             0.125387024943
## age:height_cm:pace:power_long_shots                     -0.016634143324
## wage_eur_log:age:dribbling:power_long_shots              0.795897661233
## wage_eur_log:height_cm:dribbling:power_long_shots        0.334970401997
## age:height_cm:dribbling:power_long_shots                 0.045851891741
## wage_eur_log:pace:dribbling:power_long_shots             0.618424114457
## age:pace:dribbling:power_long_shots                      0.022400579435
## height_cm:pace:dribbling:power_long_shots                0.027447459339
## wage_eur_log:age:defending:power_long_shots             -0.139067144832
## wage_eur_log:height_cm:defending:power_long_shots        0.256830052830
## age:height_cm:defending:power_long_shots                 0.010465545204
## wage_eur_log:pace:defending:power_long_shots             0.381311744017
## age:pace:defending:power_long_shots                     -0.090849062487
## height_cm:pace:defending:power_long_shots                0.016568453595
## wage_eur_log:dribbling:defending:power_long_shots        1.117930607660
## age:dribbling:defending:power_long_shots                 0.144128592577
```

```
## height_cm:dribbling:defending:power_long_shots                0.053523941376
## pace:dribbling:defending:power_long_shots                     0.082846642157
## wage_eur_log:age:physic:power_long_shots                      0.222215897170
## wage_eur_log:height_cm:physic:power_long_shots                0.236779689334
## age:height_cm:physic:power_long_shots                         0.025948674716
## wage_eur_log:pace:physic:power_long_shots                     0.391325119039
## age:pace:physic:power_long_shots                             -0.020182800019
## height_cm:pace:physic:power_long_shots                        0.019333396696
## wage_eur_log:dribbling:physic:power_long_shots                0.979063749558
## age:dribbling:physic:power_long_shots                         0.146741928532
## height_cm:dribbling:physic:power_long_shots                   0.047399052070
## pace:dribbling:physic:power_long_shots                        0.081595498733
## wage_eur_log:defending:physic:power_long_shots                0.721719016533
## age:defending:physic:power_long_shots                         0.044247612340
## height_cm:defending:physic:power_long_shots                   0.037412923685
## pace:defending:physic:power_long_shots                        0.044247957282
## dribbling:defending:physic:power_long_shots                   0.148354404333
## wage_eur_log:age:height_cm:pace:dribbling                     0.005331123392
## wage_eur_log:age:height_cm:pace:defending                     0.010235174316
## wage_eur_log:age:height_cm:dribbling:defending                0.002010092467
## wage_eur_log:age:pace:dribbling:defending                     0.015912542654
## wage_eur_log:height_cm:pace:dribbling:defending              -0.001132841309
## age:height_cm:pace:dribbling:defending                        0.000872429707
## wage_eur_log:age:height_cm:pace:physic                        0.008053887787
## wage_eur_log:age:height_cm:dribbling:physic                   0.002056054895
## wage_eur_log:age:pace:dribbling:physic                        0.011767038188
## wage_eur_log:height_cm:pace:dribbling:physic                 -0.000855335841
## age:height_cm:pace:dribbling:physic                           0.000667888974
## wage_eur_log:age:height_cm:defending:physic                   0.006857722152
## wage_eur_log:age:pace:defending:physic                        0.026235829030
## wage_eur_log:height_cm:pace:defending:physic                  0.000544426591
## age:height_cm:pace:defending:physic                           0.001287598198
## wage_eur_log:age:dribbling:defending:physic                   0.005221425669
## wage_eur_log:height_cm:dribbling:defending:physic            -0.002977468311
## age:height_cm:dribbling:defending:physic                      0.000258948566
## wage_eur_log:pace:dribbling:defending:physic                 -0.003470312447
## age:pace:dribbling:defending:physic                           0.002366722350
## height_cm:pace:dribbling:defending:physic                    -0.000071178432
## wage_eur_log:age:height_cm:pace:power_long_shots              0.003080992979
## wage_eur_log:age:height_cm:dribbling:power_long_shots        -0.004248840353
## wage_eur_log:age:pace:dribbling:power_long_shots             -0.002072876097
## wage_eur_log:height_cm:pace:dribbling:power_long_shots       -0.003296584231
## age:height_cm:pace:dribbling:power_long_shots                -0.000110382213
## wage_eur_log:age:height_cm:defending:power_long_shots         0.000673611437
## wage_eur_log:age:pace:defending:power_long_shots              0.012066518281
## wage_eur_log:height_cm:pace:defending:power_long_shots       -0.002079772333
## age:height_cm:pace:defending:power_long_shots                 0.000479697000
## wage_eur_log:age:dribbling:defending:power_long_shots        -0.013940982606
## wage_eur_log:height_cm:dribbling:defending:power_long_shots  -0.006194988670
## age:height_cm:dribbling:defending:power_long_shots           -0.000841499173
## wage_eur_log:pace:dribbling:defending:power_long_shots       -0.010838602315
## age:pace:dribbling:defending:power_long_shots                -0.000088096069
## height_cm:pace:dribbling:defending:power_long_shots          -0.000464535729
## wage_eur_log:age:height_cm:physic:power_long_shots           -0.000817337459
## wage_eur_log:age:pace:physic:power_long_shots                 0.005521936543
## wage_eur_log:height_cm:pace:physic:power_long_shots          -0.001975891046
## age:height_cm:pace:physic:power_long_shots                    0.000154357856
## wage_eur_log:age:dribbling:physic:power_long_shots           -0.014101702417
## wage_eur_log:height_cm:dribbling:physic:power_long_shots     -0.005266498366
## age:height_cm:dribbling:physic:power_long_shots              -0.000789704318
## wage_eur_log:pace:dribbling:physic:power_long_shots          -0.009765182913
## age:pace:dribbling:physic:power_long_shots                   -0.000561818968
## height_cm:pace:dribbling:physic:power_long_shots             -0.000436201436
## wage_eur_log:age:defending:physic:power_long_shots            0.000341945661
## wage_eur_log:height_cm:defending:physic:power_long_shots     -0.003929900706
## age:height_cm:defending:physic:power_long_shots              -0.000254878088
## wage_eur_log:pace:defending:physic:power_long_shots          -0.005451187035
## age:pace:defending:physic:power_long_shots                    0.001259932116
## height_cm:pace:defending:physic:power_long_shots             -0.000241988059
## wage_eur_log:dribbling:defending:physic:power_long_shots     -0.017042227869
## age:dribbling:defending:physic:power_long_shots              -0.002307978473
## height_cm:dribbling:defending:physic:power_long_shots        -0.000822126844
## pace:dribbling:defending:physic:power_long_shots             -0.001234261089
## wage_eur_log:age:height_cm:pace:dribbling:defending          -0.000092453578
## wage_eur_log:age:height_cm:pace:dribbling:physic             -0.000075669089
## wage_eur_log:age:height_cm:pace:defending:physic             -0.000152823531
## wage_eur_log:age:height_cm:dribbling:defending:physic        -0.000033387244
## wage_eur_log:age:pace:dribbling:defending:physic             -0.000241294735
## wage_eur_log:height_cm:pace:dribbling:defending:physic        0.000015624284
## age:height_cm:pace:dribbling:defending:physic                -0.000013579181
## wage_eur_log:age:height_cm:pace:dribbling:power_long_shots    0.000007008036
## wage_eur_log:age:height_cm:pace:defending:power_long_shots   -0.000066910323
## wage_eur_log:age:height_cm:dribbling:defending:power_long_shots  0.000079573103
## wage_eur_log:age:pace:dribbling:defending:power_long_shots    0.000013337758
```

```
## wage_eur_log:height_cm:pace:dribbling:defending:power_long_shots                0.000059515606
## age:height_cm:pace:dribbling:defending:power_long_shots                         0.000000917258
## wage_eur_log:age:height_cm:pace:physic:power_long_shots                        -0.000038378643
## wage_eur_log:age:height_cm:dribbling:physic:power_long_shots                    0.000072867433
## wage_eur_log:age:pace:dribbling:physic:power_long_shots                         0.000053245126
## wage_eur_log:height_cm:pace:dribbling:physic:power_long_shots                   0.000051135106
## age:height_cm:pace:dribbling:physic:power_long_shots                            0.000002545317
## wage_eur_log:age:height_cm:defending:physic:power_long_shots                   -0.000003131975
## wage_eur_log:age:pace:defending:physic:power_long_shots                        -0.000176859814
## wage_eur_log:height_cm:pace:defending:physic:power_long_shots                   0.000028753066
## age:height_cm:pace:defending:physic:power_long_shots                           -0.000006988649
## wage_eur_log:age:dribbling:defending:physic:power_long_shots                    0.000218769832
## wage_eur_log:height_cm:dribbling:defending:physic:power_long_shots              0.000093345658
## age:height_cm:dribbling:defending:physic:power_long_shots                       0.000013031219
## wage_eur_log:pace:dribbling:defending:physic:power_long_shots                   0.000158595081
## age:pace:dribbling:defending:physic:power_long_shots                            0.000000885272
## height_cm:pace:dribbling:defending:physic:power_long_shots                      0.000006775439
## wage_eur_log:age:height_cm:pace:dribbling:defending:physic                      0.000001442517
## wage_eur_log:age:height_cm:pace:dribbling:defending:power_long_shots            -0.000000079616
## wage_eur_log:age:height_cm:pace:dribbling:physic:power_long_shots               -0.000000185991
## wage_eur_log:age:height_cm:pace:defending:physic:power_long_shots               0.000001021421
## wage_eur_log:age:height_cm:dribbling:defending:physic:power_long_shots          -0.000001201982
## wage_eur_log:age:pace:dribbling:defending:physic:power_long_shots               -0.000000106841
## wage_eur_log:height_cm:pace:dribbling:defending:physic:power_long_shots         -0.000000854063
## age:height_cm:pace:dribbling:defending:physic:power_long_shots                  -0.000000005364
## wage_eur_log:age:height_cm:pace:dribbling:defending:physic:power_long_shots                NA
##                                                                                    Std. Error
## (Intercept)                                                                   810426.497181685059
## wage_eur_log                                                                  122813.547944598744
## age                                                                            38735.313375743157
## height_cm                                                                       4496.184263362489
## pace                                                                           11686.555341824876
## dribbling                                                                      13945.644640084638
## defending                                                                      13459.418218453275
## physic                                                                         15295.182671274888
## power_long_shots                                                               13951.856270178532
## I(wage_eur_log^2)                                                                  0.145235171449
## I(age^2)                                                                           0.010733368283
## I(height_cm^2)                                                                     0.004381462650
## I(pace^2)                                                                          0.002072281810
## I(dribbling^2)                                                                     0.005148935274
## I(defending^2)                                                                     0.001082375557
## I(physic^2)                                                                        0.003275170894
## I(power_long_shots^2)                                                              0.001315203615
## wage_eur_log:age                                                                5539.777740979263
## wage_eur_log:height_cm                                                           682.920075444882
## age:height_cm                                                                    217.621862305997
## wage_eur_log:pace                                                               1950.456345370333
## age:pace                                                                         571.705965557261
## height_cm:pace                                                                    66.479856435376
## wage_eur_log:dribbling                                                          2005.842981212068
## age:dribbling                                                                    632.918085151116
## height_cm:dribbling                                                               77.521294908730
## pace:dribbling                                                                   192.299598234302
## wage_eur_log:defending                                                          1758.858024059992
## age:defending                                                                    590.968396706115
## height_cm:defending                                                               72.684572926952
## pace:defending                                                                   177.971443259344
## dribbling:defending                                                              248.339805038806
## wage_eur_log:physic                                                             2075.629593573948
## age:physic                                                                       660.762005142207
## height_cm:physic                                                                  83.288026796875
## pace:physic                                                                      218.891982513219
## dribbling:physic                                                                 260.335989978091
## defending:physic                                                                 194.378551307029
## wage_eur_log:power_long_shots                                                   1963.687746257762
## age:power_long_shots                                                             579.858098894247
## height_cm:power_long_shots                                                        77.195051475967
## pace:power_long_shots                                                            179.189188027685
## dribbling:power_long_shots                                                       215.202741688037
## defending:power_long_shots                                                       258.629949478439
## physic:power_long_shots                                                          264.629238461172
## wage_eur_log:age:height_cm                                                        31.075209731985
## wage_eur_log:age:pace                                                             87.312472940094
## wage_eur_log:height_cm:pace                                                       10.989598525918
## age:height_cm:pace                                                                 3.272989564296
## wage_eur_log:age:dribbling                                                        87.061533591231
## wage_eur_log:height_cm:dribbling                                                  11.148392141102
## age:height_cm:dribbling                                                            3.544543921475
## wage_eur_log:pace:dribbling                                                       30.498146007141
## age:pace:dribbling                                                                 8.899333727820
## height_cm:pace:dribbling                                                           1.090049579773
## wage_eur_log:age:defending                                                        74.514440418277
## wage_eur_log:height_cm:defending                                                   9.546249739779
```

```
## age:height_cm:defending                                3.235288286923
## wage_eur_log:pace:defending                           25.685120250357
## age:pace:defending                                     8.069949577474
## height_cm:pace:defending                               0.985849145268
## wage_eur_log:dribbling:defending                      31.606458718278
## age:dribbling:defending                               10.439629593860
## height_cm:dribbling:defending                          1.358181876024
## pace:dribbling:defending                               3.312045073890
## wage_eur_log:age:physic                               88.941948241338
## wage_eur_log:height_cm:physic                         11.393973665137
## age:height_cm:physic                                   3.642855743678
## wage_eur_log:pace:physic                              31.937537108410
## age:pace:physic                                        9.695751685135
## height_cm:pace:physic                                  1.215575174577
## wage_eur_log:dribbling:physic                         34.337476044516
## age:dribbling:physic                                  10.816331141907
## height_cm:dribbling:physic                             1.421707002409
## pace:dribbling:physic                                  3.589928333759
## wage_eur_log:defending:physic                         23.491570063860
## age:defending:physic                                   7.885622765776
## height_cm:defending:physic                             1.006931384934
## pace:defending:physic                                  2.312676080519
## dribbling:defending:physic                             3.517828638066
## wage_eur_log:age:power_long_shots                     81.555731086879
## wage_eur_log:height_cm:power_long_shots               10.932890051673
## age:height_cm:power_long_shots                         3.267745375642
## wage_eur_log:pace:power_long_shots                    28.509417496257
## age:pace:power_long_shots                              7.995464778055
## height_cm:pace:power_long_shots                        1.027004618366
## wage_eur_log:dribbling:power_long_shots               28.614357409198
## age:dribbling:power_long_shots                         8.402928528966
## height_cm:dribbling:power_long_shots                   1.193165064917
## pace:dribbling:power_long_shots                        2.628342240032
## wage_eur_log:defending:power_long_shots               28.767919329372
## age:defending:power_long_shots                         9.015658198205
## height_cm:defending:power_long_shots                   1.396369412633
## pace:defending:power_long_shots                        3.237479608423
## dribbling:defending:power_long_shots                   4.241270283782
## wage_eur_log:physic:power_long_shots                  33.050638297695
## age:physic:power_long_shots                           10.394689867775
## height_cm:physic:power_long_shots                      1.437081213398
## pace:physic:power_long_shots                           3.560259794296
## dribbling:physic:power_long_shots                      4.156027851116
## defending:physic:power_long_shots                      3.928929803536
## wage_eur_log:age:height_cm:pace                        0.495064512771
## wage_eur_log:age:height_cm:dribbling                   0.486929002184
## wage_eur_log:age:pace:dribbling                        1.317773778297
## wage_eur_log:height_cm:pace:dribbling                  0.171413736602
## age:height_cm:pace:dribbling                           0.050701392856
## wage_eur_log:age:height_cm:defending                   0.409872451170
## wage_eur_log:age:pace:defending                        1.095140818957
## wage_eur_log:height_cm:pace:defending                  0.142503519257
## age:height_cm:pace:defending                           0.045255138759
## wage_eur_log:age:dribbling:defending                   1.283984735203
## wage_eur_log:height_cm:dribbling:defending             0.172999108263
## age:height_cm:dribbling:defending                      0.057392329299
## wage_eur_log:pace:dribbling:defending                  0.451132636250
## age:pace:dribbling:defending                           0.140902904494
## height_cm:pace:dribbling:defending                     0.018419560791
## wage_eur_log:age:height_cm:physic                      0.492628455912
## wage_eur_log:age:pace:physic                           1.378967302031
## wage_eur_log:height_cm:pace:physic                     0.177914057260
## age:height_cm:pace:physic                              0.054451714054
## wage_eur_log:age:dribbling:physic                      1.417352241947
## wage_eur_log:height_cm:dribbling:physic                0.188359864929
## age:height_cm:dribbling:physic                         0.059479309606
## wage_eur_log:pace:dribbling:physic                     0.506096944054
## age:pace:dribbling:physic                              0.151715147610
## height_cm:pace:dribbling:physic                        0.019900877826
## wage_eur_log:age:defending:physic                      0.946856067115
## wage_eur_log:height_cm:defending:physic                0.122910820051
## age:height_cm:defending:physic                         0.041550873731
## wage_eur_log:pace:defending:physic                     0.313478170565
## age:pace:defending:physic                              0.098909842899
## height_cm:pace:defending:physic                        0.012176743681
## wage_eur_log:dribbling:defending:physic                0.426914844337
## age:dribbling:defending:physic                         0.136550122352
## height_cm:dribbling:defending:physic                   0.018632010090
## pace:dribbling:defending:physic                        0.042625231119
## wage_eur_log:age:height_cm:power_long_shots            0.459533551961
## wage_eur_log:age:pace:power_long_shots                 1.220544535044
## wage_eur_log:height_cm:pace:power_long_shots           0.162125683317
## age:height_cm:pace:power_long_shots                    0.046609306261
## wage_eur_log:age:dribbling:power_long_shots            1.119394653110
```

```
## wage_eur_log:height_cm:dribbling:power_long_shots                    0.159456111543
## age:height_cm:dribbling:power_long_shots                             0.047297828848
## wage_eur_log:pace:dribbling:power_long_shots                         0.400264819299
## age:pace:dribbling:power_long_shots                                  0.107995787230
## height_cm:pace:dribbling:power_long_shots                            0.015075769968
## wage_eur_log:age:defending:power_long_shots                          0.971405732535
## wage_eur_log:height_cm:defending:power_long_shots                    0.155127189501
## age:height_cm:defending:power_long_shots                             0.049010084355
## wage_eur_log:pace:defending:power_long_shots                         0.361058556808
## age:pace:defending:power_long_shots                                  0.112970471569
## height_cm:pace:defending:power_long_shots                            0.017937315452
## wage_eur_log:dribbling:defending:power_long_shots                    0.456462945977
## age:dribbling:defending:power_long_shots                             0.143116343817
## height_cm:dribbling:defending:power_long_shots                       0.023085497056
## pace:dribbling:defending:power_long_shots                            0.052889070797
## wage_eur_log:age:physic:power_long_shots                             1.325598445062
## wage_eur_log:height_cm:physic:power_long_shots                       0.181468336192
## age:height_cm:physic:power_long_shots                                0.057183438902
## wage_eur_log:pace:physic:power_long_shots                            0.475070201201
## age:pace:physic:power_long_shots                                     0.146573183938
## height_cm:pace:physic:power_long_shots                               0.019755459279
## wage_eur_log:dribbling:physic:power_long_shots                       0.498608026165
## age:dribbling:physic:power_long_shots                                0.155369335433
## height_cm:dribbling:physic:power_long_shots                          0.022586321254
## pace:dribbling:physic:power_long_shots                               0.053923054024
## wage_eur_log:defending:physic:power_long_shots                       0.392803175064
## age:defending:physic:power_long_shots                                0.128093699710
## height_cm:defending:physic:power_long_shots                          0.020533824469
## pace:defending:physic:power_long_shots                               0.046583510415
## dribbling:defending:physic:power_long_shots                          0.062808876106
## wage_eur_log:age:height_cm:pace:dribbling                            0.007447327905
## wage_eur_log:age:height_cm:pace:defending                            0.006150260399
## wage_eur_log:age:height_cm:dribbling:defending                       0.007074562404
## wage_eur_log:age:pace:dribbling:defending                            0.018335336424
## wage_eur_log:height_cm:pace:dribbling:defending                      0.002505433324
## age:height_cm:pace:dribbling:defending                               0.000787836699
## wage_eur_log:age:height_cm:pace:physic                               0.007738205979
## wage_eur_log:age:height_cm:dribbling:physic                          0.007817902767
## wage_eur_log:age:pace:dribbling:physic                               0.021052544142
## wage_eur_log:height_cm:pace:dribbling:physic                         0.002810466964
## age:height_cm:pace:dribbling:physic                                  0.000847866067
## wage_eur_log:age:height_cm:defending:physic                          0.005048748438
## wage_eur_log:age:pace:defending:physic                               0.013064928607
## wage_eur_log:height_cm:pace:defending:physic                         0.001686374482
## age:height_cm:pace:defending:physic                                  0.000536316735
## wage_eur_log:age:dribbling:defending:physic                          0.016315719295
## wage_eur_log:height_cm:dribbling:defending:physic                    0.002272555235
## age:height_cm:dribbling:defending:physic                             0.000726451856
## wage_eur_log:pace:dribbling:defending:physic                         0.005598512333
## age:pace:dribbling:defending:physic                                  0.001677126692
## height_cm:pace:dribbling:defending:physic                            0.000228631408
## wage_eur_log:age:height_cm:pace:power_long_shots                     0.007014161366
## wage_eur_log:age:height_cm:dribbling:power_long_shots                0.006311165002
## wage_eur_log:age:pace:dribbling:power_long_shots                     0.016133233641
## wage_eur_log:height_cm:pace:dribbling:power_long_shots               0.002279719233
## age:height_cm:pace:dribbling:power_long_shots                        0.000632392570
## wage_eur_log:age:height_cm:defending:power_long_shots                0.005292539974
## wage_eur_log:age:pace:defending:power_long_shots                     0.012044834405
## wage_eur_log:height_cm:pace:defending:power_long_shots               0.002010985946
## age:height_cm:pace:defending:power_long_shots                        0.000640247845
## wage_eur_log:age:dribbling:defending:power_long_shots                0.014330313535
## wage_eur_log:height_cm:dribbling:defending:power_long_shots          0.002479154330
## age:height_cm:dribbling:defending:power_long_shots                   0.000780389956
## wage_eur_log:pace:dribbling:defending:power_long_shots               0.005622557787
## age:pace:dribbling:defending:power_long_shots                        0.001741546727
## height_cm:pace:dribbling:defending:power_long_shots                  0.000293900624
## wage_eur_log:age:height_cm:physic:power_long_shots                   0.007358465640
## wage_eur_log:age:pace:physic:power_long_shots                        0.019744110286
## wage_eur_log:height_cm:pace:physic:power_long_shots                  0.002659916889
## age:height_cm:pace:physic:power_long_shots                           0.000827061062
## wage_eur_log:age:dribbling:physic:power_long_shots                   0.018924046383
## wage_eur_log:height_cm:dribbling:physic:power_long_shots             0.002732418869
## age:height_cm:dribbling:physic:power_long_shots                      0.000851102862
## wage_eur_log:pace:dribbling:physic:power_long_shots                  0.006897191156
## age:pace:dribbling:physic:power_long_shots                           0.002076854978
## height_cm:pace:dribbling:physic:power_long_shots                     0.000298498302
## wage_eur_log:age:defending:physic:power_long_shots                   0.011794522819
## wage_eur_log:height_cm:defending:physic:power_long_shots             0.002029321862
## age:height_cm:defending:physic:power_long_shots                      0.000663453535
## wage_eur_log:pace:defending:physic:power_long_shots                  0.004279757673
## age:pace:defending:physic:power_long_shots                           0.001475819221
## height_cm:pace:defending:physic:power_long_shots                     0.000246737048
## wage_eur_log:dribbling:defending:physic:power_long_shots             0.006127762410
## age:dribbling:defending:physic:power_long_shots                      0.001927996631
```

```
## height_cm:dribbling:defending:physic:power_long_shots                          0.000330953228
## pace:dribbling:defending:physic:power_long_shots                               0.000725021602
## wage_eur_log:age:height_cm:pace:dribbling:defending                            0.000102592562
## wage_eur_log:age:height_cm:pace:dribbling:physic                               0.000117571672
## wage_eur_log:age:height_cm:pace:defending:physic                               0.000071979248
## wage_eur_log:age:height_cm:dribbling:defending:physic                          0.000087373376
## wage_eur_log:age:pace:dribbling:defending:physic                               0.000215912396
## wage_eur_log:height_cm:pace:dribbling:defending:physic                         0.000030326307
## age:height_cm:pace:dribbling:defending:physic                                  0.000009099105
## wage_eur_log:age:height_cm:pace:dribbling:power_long_shots                     0.000093121501
## wage_eur_log:age:height_cm:pace:defending:power_long_shots                     0.000069363086
## wage_eur_log:age:height_cm:dribbling:defending:power_long_shots                0.000078111739
## wage_eur_log:age:pace:dribbling:defending:power_long_shots                     0.000165875754
## wage_eur_log:height_cm:pace:dribbling:defending:power_long_shots               0.000031375403
## age:height_cm:pace:dribbling:defending:power_long_shots                        0.000009835699
## wage_eur_log:age:height_cm:pace:physic:power_long_shots                        0.000111784899
## wage_eur_log:age:height_cm:dribbling:physic:power_long_shots                   0.000104541874
## wage_eur_log:age:pace:dribbling:physic:power_long_shots                        0.000270231998
## wage_eur_log:height_cm:pace:dribbling:physic:power_long_shots                  0.000038512093
## age:height_cm:pace:dribbling:physic:power_long_shots                           0.000011671824
## wage_eur_log:age:height_cm:defending:physic:power_long_shots                   0.000059966062
## wage_eur_log:age:pace:defending:physic:power_long_shots                        0.000112317648
## wage_eur_log:height_cm:pace:defending:physic:power_long_shots                  0.000022535367
## age:height_cm:pace:defending:physic:power_long_shots                           0.000007904211
## wage_eur_log:age:dribbling:defending:physic:power_long_shots                   0.000157743008
## wage_eur_log:height_cm:dribbling:defending:physic:power_long_shots             0.000031855198
## age:height_cm:dribbling:defending:physic:power_long_shots                      0.000009953589
## wage_eur_log:pace:dribbling:defending:physic:power_long_shots                  0.000062502034
## age:pace:dribbling:defending:physic:power_long_shots                           0.000019937464
## height_cm:pace:dribbling:defending:physic:power_long_shots                     0.000003852540
## wage_eur_log:age:height_cm:pace:dribbling:defending:physic                     0.000001183418
## wage_eur_log:age:height_cm:pace:dribbling:defending:power_long_shots           0.000000956830
## wage_eur_log:age:height_cm:pace:dribbling:physic:power_long_shots              0.000001526263
## wage_eur_log:age:height_cm:pace:defending:physic:power_long_shots              0.000000610931
## wage_eur_log:age:height_cm:dribbling:defending:physic:power_long_shots         0.000000773746
## wage_eur_log:age:pace:dribbling:defending:physic:power_long_shots              0.000000498939
## wage_eur_log:height_cm:pace:dribbling:defending:physic:power_long_shots        0.000000327218
## age:height_cm:pace:dribbling:defending:physic:power_long_shots                 0.000000104755
## wage_eur_log:age:height_cm:pace:dribbling:defending:physic:power_long_shots                NA
##                                                                              t value
## (Intercept)                                                                    0.527
## wage_eur_log                                                                  -0.328
## age                                                                            0.848
## height_cm                                                                     -0.460
## pace                                                                           0.191
## dribbling                                                                     -1.067
## defending                                                                     -0.542
## physic                                                                        -0.577
## power_long_shots                                                              -1.852
## I(wage_eur_log^2)                                                              4.751
## I(age^2)                                                                       6.980
## I(height_cm^2)                                                                 0.016
## I(pace^2)                                                                      1.130
## I(dribbling^2)                                                                 1.860
## I(defending^2)                                                                 4.668
## I(physic^2)                                                                    0.878
## I(power_long_shots^2)                                                          0.293
## wage_eur_log:age                                                              -0.838
## wage_eur_log:height_cm                                                         0.241
## age:height_cm                                                                 -0.877
## wage_eur_log:pace                                                             -0.129
## age:pace                                                                      -1.352
## height_cm:pace                                                                -0.277
## wage_eur_log:dribbling                                                         0.916
## age:dribbling                                                                 -0.304
## height_cm:dribbling                                                            1.016
## pace:dribbling                                                                 0.358
## wage_eur_log:defending                                                         0.515
## age:defending                                                                 -1.083
## height_cm:defending                                                            0.530
## pace:defending                                                                -0.392
## dribbling:defending                                                            1.158
## wage_eur_log:physic                                                            0.393
## age:physic                                                                    -0.604
## height_cm:physic                                                               0.506
## pace:physic                                                                   -0.034
## dribbling:physic                                                               0.971
## defending:physic                                                               0.675
## wage_eur_log:power_long_shots                                                  1.388
## age:power_long_shots                                                           0.339
## height_cm:power_long_shots                                                     1.836
## pace:power_long_shots                                                          1.212
## dribbling:power_long_shots                                                     2.489
## defending:power_long_shots                                                     1.619
```

```
## physic:power_long_shots                                  1.647
## wage_eur_log:age:height_cm                               0.890
## wage_eur_log:age:pace                                    1.097
## wage_eur_log:height_cm:pace                              0.225
## age:height_cm:pace                                       1.372
## wage_eur_log:age:dribbling                               0.277
## wage_eur_log:height_cm:dribbling                        -0.842
## age:height_cm:dribbling                                  0.324
## wage_eur_log:pace:dribbling                             -0.379
## age:pace:dribbling                                       0.897
## height_cm:pace:dribbling                                -0.274
## wage_eur_log:age:defending                               1.112
## wage_eur_log:height_cm:defending                        -0.464
## age:height_cm:defending                                  1.092
## wage_eur_log:pace:defending                              0.145
## age:pace:defending                                       1.928
## height_cm:pace:defending                                 0.444
## wage_eur_log:dribbling:defending                        -1.188
## age:dribbling:defending                                  0.290
## height_cm:dribbling:defending                           -1.159
## pace:dribbling:defending                                -0.322
## wage_eur_log:age:physic                                  0.679
## wage_eur_log:height_cm:physic                           -0.296
## age:height_cm:physic                                     0.669
## wage_eur_log:pace:physic                                 0.046
## age:pace:physic                                          1.079
## height_cm:pace:physic                                    0.136
## wage_eur_log:dribbling:physic                           -0.886
## age:dribbling:physic                                     0.195
## height_cm:dribbling:physic                              -0.920
## pace:dribbling:physic                                   -0.382
## wage_eur_log:defending:physic                           -0.631
## age:defending:physic                                     1.108
## height_cm:defending:physic                              -0.660
## pace:defending:physic                                    0.402
## dribbling:defending:physic                              -1.276
## wage_eur_log:age:power_long_shots                       -0.024
## wage_eur_log:height_cm:power_long_shots                 -1.339
## age:height_cm:power_long_shots                          -0.320
## wage_eur_log:pace:power_long_shots                      -0.851
## age:pace:power_long_shots                                0.343
## height_cm:pace:power_long_shots                         -1.137
## wage_eur_log:dribbling:power_long_shots                 -2.150
## age:dribbling:power_long_shots                          -0.983
## height_cm:dribbling:power_long_shots                    -2.478
## pace:dribbling:power_long_shots                         -1.917
## wage_eur_log:defending:power_long_shots                 -1.614
## age:defending:power_long_shots                          -0.164
## height_cm:defending:power_long_shots                    -1.688
## pace:defending:power_long_shots                         -0.904
## dribbling:defending:power_long_shots                    -2.250
## wage_eur_log:physic:power_long_shots                    -1.352
## age:physic:power_long_shots                             -0.487
## height_cm:physic:power_long_shots                       -1.643
## pace:physic:power_long_shots                            -1.032
## dribbling:physic:power_long_shots                       -2.091
## defending:physic:power_long_shots                       -1.719
## wage_eur_log:age:height_cm:pace                         -1.150
## wage_eur_log:age:height_cm:dribbling                    -0.322
## wage_eur_log:age:pace:dribbling                         -0.660
## wage_eur_log:height_cm:pace:dribbling                    0.284
## age:height_cm:pace:dribbling                            -0.922
## wage_eur_log:age:height_cm:defending                    -1.150
## wage_eur_log:age:pace:defending                         -1.633
## wage_eur_log:height_cm:pace:defending                   -0.226
## age:height_cm:pace:defending                            -1.925
## wage_eur_log:age:dribbling:defending                    -0.266
## wage_eur_log:height_cm:dribbling:defending               1.157
## age:height_cm:dribbling:defending                       -0.279
## wage_eur_log:pace:dribbling:defending                    0.521
## age:pace:dribbling:defending                            -1.107
## height_cm:pace:dribbling:defending                       0.282
## wage_eur_log:age:height_cm:physic                       -0.761
## wage_eur_log:age:pace:physic                            -0.956
## wage_eur_log:height_cm:pace:physic                      -0.159
## age:height_cm:pace:physic                               -1.147
## wage_eur_log:age:dribbling:physic                       -0.195
## wage_eur_log:height_cm:dribbling:physic                  0.808
## age:height_cm:dribbling:physic                          -0.242
## wage_eur_log:pace:dribbling:physic                       0.412
## age:pace:dribbling:physic                               -0.722
## height_cm:pace:dribbling:physic                          0.290
## wage_eur_log:age:defending:physic                       -1.241
## wage_eur_log:height_cm:defending:physic                  0.563
```

```
## age:height_cm:defending:physic                                    -1.195
## wage_eur_log:pace:defending:physic                                -0.177
## age:pace:defending:physic                                         -2.285
## height_cm:pace:defending:physic                                   -0.529
## wage_eur_log:dribbling:defending:physic                            1.338
## age:dribbling:defending:physic                                    -0.328
## height_cm:dribbling:defending:physic                               1.289
## pace:dribbling:defending:physic                                    0.383
## wage_eur_log:age:height_cm:power_long_shots                       -0.012
## wage_eur_log:age:pace:power_long_shots                            -0.403
## wage_eur_log:height_cm:pace:power_long_shots                       0.773
## age:height_cm:pace:power_long_shots                               -0.357
## wage_eur_log:age:dribbling:power_long_shots                        0.711
## wage_eur_log:height_cm:dribbling:power_long_shots                  2.101
## age:height_cm:dribbling:power_long_shots                           0.969
## wage_eur_log:pace:dribbling:power_long_shots                       1.545
## age:pace:dribbling:power_long_shots                                0.207
## height_cm:pace:dribbling:power_long_shots                          1.821
## wage_eur_log:age:defending:power_long_shots                       -0.143
## wage_eur_log:height_cm:defending:power_long_shots                  1.656
## age:height_cm:defending:power_long_shots                           0.214
## wage_eur_log:pace:defending:power_long_shots                       1.056
## age:pace:defending:power_long_shots                               -0.804
## height_cm:pace:defending:power_long_shots                          0.924
## wage_eur_log:dribbling:defending:power_long_shots                  2.449
## age:dribbling:defending:power_long_shots                           1.007
## height_cm:dribbling:defending:power_long_shots                     2.319
## pace:dribbling:defending:power_long_shots                          1.566
## wage_eur_log:age:physic:power_long_shots                           0.168
## wage_eur_log:height_cm:physic:power_long_shots                     1.305
## age:height_cm:physic:power_long_shots                              0.454
## wage_eur_log:pace:physic:power_long_shots                          0.824
## age:pace:physic:power_long_shots                                  -0.138
## height_cm:pace:physic:power_long_shots                             0.979
## wage_eur_log:dribbling:physic:power_long_shots                     1.964
## age:dribbling:physic:power_long_shots                              0.944
## height_cm:dribbling:physic:power_long_shots                        2.099
## pace:dribbling:physic:power_long_shots                             1.513
## wage_eur_log:defending:physic:power_long_shots                     1.837
## age:defending:physic:power_long_shots                              0.345
## height_cm:defending:physic:power_long_shots                        1.822
## pace:defending:physic:power_long_shots                             0.950
## dribbling:defending:physic:power_long_shots                        2.362
## wage_eur_log:age:height_cm:pace:dribbling                          0.716
## wage_eur_log:age:height_cm:pace:defending                          1.664
## wage_eur_log:age:height_cm:dribbling:defending                     0.284
## wage_eur_log:age:pace:dribbling:defending                          0.868
## wage_eur_log:height_cm:pace:dribbling:defending                   -0.452
## age:height_cm:pace:dribbling:defending                             1.107
## wage_eur_log:age:height_cm:pace:physic                             1.041
## wage_eur_log:age:height_cm:dribbling:physic                        0.263
## wage_eur_log:age:pace:dribbling:physic                             0.559
## wage_eur_log:height_cm:pace:dribbling:physic                      -0.304
## age:height_cm:pace:dribbling:physic                                0.788
## wage_eur_log:age:height_cm:defending:physic                        1.358
## wage_eur_log:age:pace:defending:physic                             2.008
## wage_eur_log:height_cm:pace:defending:physic                       0.323
## age:height_cm:pace:defending:physic                                2.401
## wage_eur_log:age:dribbling:defending:physic                        0.320
## wage_eur_log:height_cm:dribbling:defending:physic                 -1.310
## age:height_cm:dribbling:defending:physic                           0.356
## wage_eur_log:pace:dribbling:defending:physic                      -0.620
## age:pace:dribbling:defending:physic                                1.411
## height_cm:pace:dribbling:defending:physic                         -0.311
## wage_eur_log:age:height_cm:pace:power_long_shots                   0.439
## wage_eur_log:age:height_cm:dribbling:power_long_shots             -0.673
## wage_eur_log:age:pace:dribbling:power_long_shots                  -0.128
## wage_eur_log:height_cm:pace:dribbling:power_long_shots            -1.446
## age:height_cm:pace:dribbling:power_long_shots                     -0.175
## wage_eur_log:age:height_cm:defending:power_long_shots              0.127
## wage_eur_log:age:pace:defending:power_long_shots                   1.002
## wage_eur_log:height_cm:pace:defending:power_long_shots            -1.034
## age:height_cm:pace:defending:power_long_shots                      0.749
## wage_eur_log:age:dribbling:defending:power_long_shots             -0.973
## wage_eur_log:height_cm:dribbling:defending:power_long_shots       -2.499
## age:height_cm:dribbling:defending:power_long_shots                -1.078
## wage_eur_log:pace:dribbling:defending:power_long_shots            -1.928
## age:pace:dribbling:defending:power_long_shots                     -0.051
## height_cm:pace:dribbling:defending:power_long_shots               -1.581
## wage_eur_log:age:height_cm:physic:power_long_shots                -0.111
## wage_eur_log:age:pace:physic:power_long_shots                      0.280
## wage_eur_log:height_cm:pace:physic:power_long_shots               -0.743
## age:height_cm:pace:physic:power_long_shots                         0.187
## wage_eur_log:age:dribbling:physic:power_long_shots                -0.745
```

```
## wage_eur_log:height_cm:dribbling:physic:power_long_shots                    -1.927
## age:height_cm:dribbling:physic:power_long_shots                             -0.928
## wage_eur_log:pace:dribbling:physic:power_long_shots                         -1.416
## age:pace:dribbling:physic:power_long_shots                                  -0.271
## height_cm:pace:dribbling:physic:power_long_shots                            -1.461
## wage_eur_log:age:defending:physic:power_long_shots                           0.029
## wage_eur_log:height_cm:defending:physic:power_long_shots                    -1.937
## age:height_cm:defending:physic:power_long_shots                             -0.384
## wage_eur_log:pace:defending:physic:power_long_shots                         -1.274
## age:pace:defending:physic:power_long_shots                                  0.854
## height_cm:pace:defending:physic:power_long_shots                            -0.981
## wage_eur_log:dribbling:defending:physic:power_long_shots                    -2.781
## age:dribbling:defending:physic:power_long_shots                             -1.197
## height_cm:dribbling:defending:physic:power_long_shots                       -2.484
## pace:dribbling:defending:physic:power_long_shots                            -1.702
## wage_eur_log:age:height_cm:pace:dribbling:defending                         -0.901
## wage_eur_log:age:height_cm:pace:dribbling:physic                            -0.644
## wage_eur_log:age:height_cm:pace:defending:physic                            -2.123
## wage_eur_log:age:height_cm:dribbling:defending:physic                       -0.382
## wage_eur_log:age:pace:dribbling:defending:physic                            -1.118
## wage_eur_log:height_cm:pace:dribbling:defending:physic                       0.515
## age:height_cm:pace:dribbling:defending:physic                               -1.492
## wage_eur_log:age:height_cm:pace:dribbling:power_long_shots                   0.075
## wage_eur_log:age:height_cm:pace:defending:power_long_shots                  -0.965
## wage_eur_log:age:height_cm:dribbling:defending:power_long_shots              1.019
## wage_eur_log:age:pace:dribbling:defending:power_long_shots                   0.080
## wage_eur_log:height_cm:pace:dribbling:defending:power_long_shots             1.897
## age:height_cm:pace:dribbling:defending:power_long_shots                      0.093
## wage_eur_log:age:height_cm:pace:physic:power_long_shots                     -0.343
## wage_eur_log:age:height_cm:dribbling:physic:power_long_shots                 0.697
## wage_eur_log:age:pace:dribbling:physic:power_long_shots                      0.197
## wage_eur_log:height_cm:pace:dribbling:physic:power_long_shots                1.328
## age:height_cm:pace:dribbling:physic:power_long_shots                         0.218
## wage_eur_log:age:height_cm:defending:physic:power_long_shots                -0.052
## wage_eur_log:age:pace:defending:physic:power_long_shots                     -1.575
## wage_eur_log:height_cm:pace:defending:physic:power_long_shots                1.276
## age:height_cm:pace:defending:physic:power_long_shots                        -0.884
## wage_eur_log:age:dribbling:defending:physic:power_long_shots                 1.387
## wage_eur_log:height_cm:dribbling:defending:physic:power_long_shots           2.930
## age:height_cm:dribbling:defending:physic:power_long_shots                    1.309
## wage_eur_log:pace:dribbling:defending:physic:power_long_shots                2.537
## age:pace:dribbling:defending:physic:power_long_shots                         0.044
## height_cm:pace:dribbling:defending:physic:power_long_shots                   1.759
## wage_eur_log:age:height_cm:pace:dribbling:defending:physic                   1.219
## wage_eur_log:age:height_cm:pace:dribbling:defending:power_long_shots         -0.083
## wage_eur_log:age:height_cm:pace:dribbling:physic:power_long_shots            -0.122
## wage_eur_log:age:height_cm:pace:defending:physic:power_long_shots             1.672
## wage_eur_log:age:height_cm:dribbling:defending:physic:power_long_shots       -1.553
## wage_eur_log:age:pace:dribbling:defending:physic:power_long_shots            -0.214
## wage_eur_log:height_cm:pace:dribbling:defending:physic:power_long_shots      -2.610
## age:height_cm:pace:dribbling:defending:physic:power_long_shots               -0.051
## wage_eur_log:age:height_cm:pace:dribbling:defending:physic:power_long_shots      NA
##                                                                          Pr(>|t|)
## (Intercept)                                                               0.59839
## wage_eur_log                                                              0.74307
## age                                                                       0.39737
## height_cm                                                                 0.64610
## pace                                                                      0.84857
## dribbling                                                                 0.28686
## defending                                                                 0.58824
## physic                                                                    0.56475
## power_long_shots                                                          0.06526
## I(wage_eur_log^2)                                                 0.0000034032091
## I(age^2)                                                          0.0000000000262
## I(height_cm^2)                                                            0.98690
## I(pace^2)                                                                 0.25947
## I(dribbling^2)                                                            0.06412
## I(defending^2)                                                    0.0000049578581
## I(physic^2)                                                               0.38071
## I(power_long_shots^2)                                                     0.76961
## wage_eur_log:age                                                          0.40266
## wage_eur_log:height_cm                                                    0.80988
## age:height_cm                                                             0.38127
## wage_eur_log:pace                                                         0.89747
## age:pace                                                                  0.17750
## height_cm:pace                                                            0.78208
## wage_eur_log:dribbling                                                    0.36052
## age:dribbling                                                             0.76150
## height_cm:dribbling                                                       0.31067
## pace:dribbling                                                            0.72032
## wage_eur_log:defending                                                    0.60699
## age:defending                                                             0.27986
## height_cm:defending                                                       0.59645
## pace:defending                                                            0.69511
```

```
## dribbling:defending                                    0.24796
## wage_eur_log:physic                                    0.69440
## age:physic                                             0.54648
## height_cm:physic                                       0.61348
## pace:physic                                            0.97315
## dribbling:physic                                       0.33239
## defending:physic                                       0.50042
## wage_eur_log:power_long_shots                          0.16637
## age:power_long_shots                                   0.73487
## height_cm:power_long_shots                             0.06747
## pace:power_long_shots                                  0.22670
## dribbling:power_long_shots                             0.01347
## defending:power_long_shots                             0.10660
## physic:power_long_shots                                0.10073
## wage_eur_log:age:height_cm                             0.37408
## wage_eur_log:age:pace                                  0.27386
## wage_eur_log:height_cm:pace                            0.82211
## age:height_cm:pace                                     0.17136
## wage_eur_log:age:dribbling                             0.78181
## wage_eur_log:height_cm:dribbling                       0.40034
## age:height_cm:dribbling                                0.74638
## wage_eur_log:pace:dribbling                            0.70496
## age:pace:dribbling                                     0.37057
## height_cm:pace:dribbling                               0.78448
## wage_eur_log:age:defending                             0.26705
## wage_eur_log:height_cm:defending                       0.64296
## age:height_cm:defending                                0.27580
## wage_eur_log:pace:defending                            0.88498
## age:pace:defending                                     0.05498
## height_cm:pace:defending                               0.65728
## wage_eur_log:dribbling:defending                       0.23585
## age:dribbling:defending                                0.77188
## height_cm:dribbling:defending                          0.24762
## pace:dribbling:defending                               0.74809
## wage_eur_log:age:physic                                0.49791
## wage_eur_log:height_cm:physic                          0.76743
## age:height_cm:physic                                   0.50416
## wage_eur_log:pace:physic                               0.96369
## age:pace:physic                                        0.28146
## height_cm:pace:physic                                  0.89193
## wage_eur_log:dribbling:physic                          0.37645
## age:dribbling:physic                                   0.84592
## height_cm:dribbling:physic                             0.35870
## pace:dribbling:physic                                  0.70270
## wage_eur_log:defending:physic                          0.52885
## age:defending:physic                                   0.26892
## height_cm:defending:physic                             0.51007
## pace:defending:physic                                  0.68826
## dribbling:defending:physic                             0.20296
## wage_eur_log:age:power_long_shots                      0.98076
## wage_eur_log:height_cm:power_long_shots                0.18187
## age:height_cm:power_long_shots                         0.74943
## wage_eur_log:pace:power_long_shots                     0.39561
## age:pace:power_long_shots                              0.73151
## height_cm:pace:power_long_shots                        0.25663
## wage_eur_log:dribbling:power_long_shots                0.03249
## age:dribbling:power_long_shots                         0.32659
## height_cm:dribbling:power_long_shots                   0.01387
## pace:dribbling:power_long_shots                        0.05632
## wage_eur_log:defending:power_long_shots                0.10776
## age:defending:power_long_shots                         0.86967
## height_cm:defending:power_long_shots                   0.09262
## pace:defending:power_long_shots                        0.36699
## dribbling:defending:power_long_shots                   0.02534
## wage_eur_log:physic:power_long_shots                   0.17760
## age:physic:power_long_shots                            0.62644
## height_cm:physic:power_long_shots                      0.10160
## pace:physic:power_long_shots                           0.30299
## dribbling:physic:power_long_shots                      0.03750
## defending:physic:power_long_shots                      0.08679
## wage_eur_log:age:height_cm:pace                        0.25118
## wage_eur_log:age:height_cm:dribbling                   0.74803
## wage_eur_log:age:pace:dribbling                        0.51000
## wage_eur_log:height_cm:pace:dribbling                  0.77652
## age:height_cm:pace:dribbling                           0.35723
## wage_eur_log:age:height_cm:defending                   0.25117
## wage_eur_log:age:pace:defending                        0.10370
## wage_eur_log:height_cm:pace:defending                  0.82113
## age:height_cm:pace:defending                           0.05538
## wage_eur_log:age:dribbling:defending                   0.79055
## wage_eur_log:height_cm:dribbling:defending             0.24852
## age:height_cm:dribbling:defending                      0.78059
## wage_eur_log:pace:dribbling:defending                  0.60249
## age:pace:dribbling:defending                           0.26919
```

```
## height_cm:pace:dribbling:defending                                    0.77844
## wage_eur_log:age:height_cm:physic                                     0.44764
## wage_eur_log:age:pace:physic                                          0.34007
## wage_eur_log:height_cm:pace:physic                                    0.87346
## age:height_cm:pace:physic                                             0.25229
## wage_eur_log:age:dribbling:physic                                     0.84568
## wage_eur_log:height_cm:dribbling:physic                               0.42001
## age:height_cm:dribbling:physic                                        0.80878
## wage_eur_log:pace:dribbling:physic                                    0.68071
## age:pace:dribbling:physic                                             0.47096
## height_cm:pace:dribbling:physic                                       0.77187
## wage_eur_log:age:defending:physic                                     0.21585
## wage_eur_log:height_cm:defending:physic                               0.57374
## age:height_cm:defending:physic                                        0.23337
## wage_eur_log:pace:defending:physic                                    0.86003
## age:pace:defending:physic                                             0.02315
## height_cm:pace:defending:physic                                       0.59723
## wage_eur_log:dribbling:defending:physic                               0.18215
## age:dribbling:defending:physic                                        0.74326
## height_cm:dribbling:defending:physic                                  0.19847
## pace:dribbling:defending:physic                                       0.70175
## wage_eur_log:age:height_cm:power_long_shots                           0.99023
## wage_eur_log:age:pace:power_long_shots                                0.68738
## wage_eur_log:height_cm:pace:power_long_shots                          0.44002
## age:height_cm:pace:power_long_shots                                   0.72148
## wage_eur_log:age:dribbling:power_long_shots                           0.47774
## wage_eur_log:height_cm:dribbling:power_long_shots                     0.03666
## age:height_cm:dribbling:power_long_shots                              0.33326
## wage_eur_log:pace:dribbling:power_long_shots                          0.12360
## age:pace:dribbling:power_long_shots                                   0.83585
## height_cm:pace:dribbling:power_long_shots                             0.06985
## wage_eur_log:age:defending:power_long_shots                           0.88628
## wage_eur_log:height_cm:defending:power_long_shots                     0.09905
## age:height_cm:defending:power_long_shots                              0.83108
## wage_eur_log:pace:defending:power_long_shots                          0.29194
## age:pace:defending:power_long_shots                                   0.42205
## height_cm:pace:defending:power_long_shots                             0.35654
## wage_eur_log:dribbling:defending:power_long_shots                     0.01501
## age:dribbling:defending:power_long_shots                              0.31487
## height_cm:dribbling:defending:power_long_shots                        0.02123
## pace:dribbling:defending:power_long_shots                             0.11851
## wage_eur_log:age:physic:power_long_shots                              0.86701
## wage_eur_log:height_cm:physic:power_long_shots                        0.19316
## age:height_cm:physic:power_long_shots                                 0.65038
## wage_eur_log:pace:physic:power_long_shots                             0.41088
## age:pace:physic:power_long_shots                                      0.89059
## height_cm:pace:physic:power_long_shots                                0.32870
## wage_eur_log:dribbling:physic:power_long_shots                        0.05068
## age:dribbling:physic:power_long_shots                                 0.34584
## height_cm:dribbling:physic:power_long_shots                           0.03685
## pace:dribbling:physic:power_long_shots                                0.13149
## wage_eur_log:defending:physic:power_long_shots                        0.06734
## age:defending:physic:power_long_shots                                 0.73006
## height_cm:defending:physic:power_long_shots                           0.06964
## pace:defending:physic:power_long_shots                                0.34310
## dribbling:defending:physic:power_long_shots                           0.01894
## wage_eur_log:age:height_cm:pace:dribbling                             0.47475
## wage_eur_log:age:height_cm:pace:defending                             0.09732
## wage_eur_log:age:height_cm:dribbling:defending                        0.77655
## wage_eur_log:age:pace:dribbling:defending                             0.38630
## wage_eur_log:height_cm:pace:dribbling:defending                       0.65155
## age:height_cm:pace:dribbling:defending                                0.26919
## wage_eur_log:age:height_cm:pace:physic                                0.29897
## wage_eur_log:age:height_cm:dribbling:physic                           0.79277
## wage_eur_log:age:pace:dribbling:physic                                0.57670
## wage_eur_log:height_cm:pace:dribbling:physic                          0.76112
## age:height_cm:pace:dribbling:physic                                   0.43160
## wage_eur_log:age:height_cm:defending:physic                           0.17559
## wage_eur_log:age:pace:defending:physic                                0.04570
## wage_eur_log:height_cm:pace:defending:physic                          0.74709
## age:height_cm:pace:defending:physic                                   0.01709
## wage_eur_log:age:dribbling:defending:physic                           0.74922
## wage_eur_log:height_cm:dribbling:defending:physic                     0.19133
## age:height_cm:dribbling:defending:physic                              0.72180
## wage_eur_log:pace:dribbling:defending:physic                          0.53591
## age:pace:dribbling:defending:physic                                   0.15943
## height_cm:pace:dribbling:defending:physic                             0.75581
## wage_eur_log:age:height_cm:pace:power_long_shots                      0.66086
## wage_eur_log:age:height_cm:dribbling:power_long_shots                 0.50142
## wage_eur_log:age:pace:dribbling:power_long_shots                      0.89787
## wage_eur_log:height_cm:pace:dribbling:power_long_shots                0.14941
## age:height_cm:pace:dribbling:power_long_shots                         0.86158
## wage_eur_log:age:height_cm:defending:power_long_shots                 0.89882
## wage_eur_log:age:pace:defending:power_long_shots                      0.31740
```

```
## wage_eur_log:height_cm:pace:defending:power_long_shots                          0.30204
## age:height_cm:pace:defending:power_long_shots                                   0.45442
## wage_eur_log:age:dribbling:defending:power_long_shots                           0.33157
## wage_eur_log:height_cm:dribbling:defending:power_long_shots                     0.01310
## age:height_cm:dribbling:defending:power_long_shots                              0.28193
## wage_eur_log:pace:dribbling:defending:power_long_shots                          0.05502
## age:pace:dribbling:defending:power_long_shots                                   0.95970
## height_cm:pace:dribbling:defending:power_long_shots                             0.11523
## wage_eur_log:age:height_cm:physic:power_long_shots                              0.91165
## wage_eur_log:age:pace:physic:power_long_shots                                   0.77996
## wage_eur_log:height_cm:pace:physic:power_long_shots                             0.45827
## age:height_cm:pace:physic:power_long_shots                                      0.85210
## wage_eur_log:age:dribbling:physic:power_long_shots                              0.45686
## wage_eur_log:height_cm:dribbling:physic:power_long_shots                        0.05506
## age:height_cm:dribbling:physic:power_long_shots                                 0.35437
## wage_eur_log:pace:dribbling:physic:power_long_shots                             0.15807
## age:pace:dribbling:physic:power_long_shots                                      0.78699
## height_cm:pace:dribbling:physic:power_long_shots                                0.14518
## wage_eur_log:age:defending:physic:power_long_shots                              0.97689
## wage_eur_log:height_cm:defending:physic:power_long_shots                        0.05392
## age:height_cm:defending:physic:power_long_shots                                 0.70118
## wage_eur_log:pace:defending:physic:power_long_shots                             0.20394
## age:pace:defending:physic:power_long_shots                                      0.39408
## height_cm:pace:defending:physic:power_long_shots                                0.32766
## wage_eur_log:dribbling:defending:physic:power_long_shots                        0.00583
## age:dribbling:defending:physic:power_long_shots                                 0.23240
## height_cm:dribbling:defending:physic:power_long_shots                           0.01364
## pace:dribbling:defending:physic:power_long_shots                                0.08992
## wage_eur_log:age:height_cm:pace:dribbling:defending                             0.36836
## wage_eur_log:age:height_cm:pace:dribbling:physic                                0.52042
## wage_eur_log:age:height_cm:pace:defending:physic                                0.03472
## wage_eur_log:age:height_cm:dribbling:defending:physic                           0.70269
## wage_eur_log:age:pace:dribbling:defending:physic                                0.26482
## wage_eur_log:height_cm:pace:dribbling:defending:physic                          0.60686
## age:height_cm:pace:dribbling:defending:physic                                   0.13686
## wage_eur_log:age:height_cm:pace:dribbling:power_long_shots                      0.94007
## wage_eur_log:age:height_cm:pace:defending:power_long_shots                      0.33565
## wage_eur_log:age:height_cm:dribbling:defending:power_long_shots                 0.30932
## wage_eur_log:age:pace:dribbling:defending:power_long_shots                      0.93598
## wage_eur_log:height_cm:pace:dribbling:defending:power_long_shots                0.05899
## age:height_cm:pace:dribbling:defending:power_long_shots                         0.92577
## wage_eur_log:age:height_cm:pace:physic:power_long_shots                         0.73164
## wage_eur_log:age:height_cm:dribbling:physic:power_long_shots                    0.48644
## wage_eur_log:age:pace:dribbling:physic:power_long_shots                         0.84396
## wage_eur_log:height_cm:pace:dribbling:physic:power_long_shots                   0.18546
## age:height_cm:pace:dribbling:physic:power_long_shots                            0.82755
## wage_eur_log:age:height_cm:defending:physic:power_long_shots                    0.95839
## wage_eur_log:age:pace:defending:physic:power_long_shots                         0.11660
## wage_eur_log:height_cm:pace:defending:physic:power_long_shots                   0.20317
## age:height_cm:pace:defending:physic:power_long_shots                            0.37745
## wage_eur_log:age:dribbling:defending:physic:power_long_shots                    0.16671
## wage_eur_log:height_cm:dribbling:defending:physic:power_long_shots              0.00370
## age:height_cm:dribbling:defending:physic:power_long_shots                       0.19166
## wage_eur_log:pace:dribbling:defending:physic:power_long_shots                   0.01177
## age:pace:dribbling:defending:physic:power_long_shots                            0.96462
## height_cm:pace:dribbling:defending:physic:power_long_shots                      0.07985
## wage_eur_log:age:height_cm:pace:dribbling:defending:physic                      0.22401
## wage_eur_log:age:height_cm:pace:dribbling:defending:power_long_shots            0.93375
## wage_eur_log:age:height_cm:pace:dribbling:physic:power_long_shots               0.90311
## wage_eur_log:age:height_cm:pace:defending:physic:power_long_shots               0.09579
## wage_eur_log:age:height_cm:dribbling:defending:physic:power_long_shots          0.12157
## wage_eur_log:age:pace:dribbling:defending:physic:power_long_shots               0.83062
## wage_eur_log:height_cm:pace:dribbling:defending:physic:power_long_shots         0.00960
## age:height_cm:pace:dribbling:defending:physic:power_long_shots                  0.95920
## wage_eur_log:age:height_cm:pace:dribbling:defending:physic:power_long_shots          NA
##
## (Intercept)
## wage_eur_log
## age
## height_cm
## pace
## dribbling
## defending
## physic
## power_long_shots                                                               .
## I(wage_eur_log^2)                                                              ***
## I(age^2)                                                                       ***
## I(height_cm^2)
## I(pace^2)
## I(dribbling^2)                                                                 .
## I(defending^2)                                                                 ***
## I(physic^2)
## I(power_long_shots^2)
## wage_eur_log:age
```

```
## wage_eur_log:height_cm
## age:height_cm
## wage_eur_log:pace
## age:pace
## height_cm:pace
## wage_eur_log:dribbling
## age:dribbling
## height_cm:dribbling
## pace:dribbling
## wage_eur_log:defending
## age:defending
## height_cm:defending
## pace:defending
## dribbling:defending
## wage_eur_log:physic
## age:physic
## height_cm:physic
## pace:physic
## dribbling:physic
## defending:physic
## wage_eur_log:power_long_shots
## age:power_long_shots
## height_cm:power_long_shots                                    .
## pace:power_long_shots
## dribbling:power_long_shots                                    *
## defending:power_long_shots
## physic:power_long_shots
## wage_eur_log:age:height_cm
## wage_eur_log:age:pace
## wage_eur_log:height_cm:pace
## age:height_cm:pace
## wage_eur_log:age:dribbling
## wage_eur_log:height_cm:dribbling
## age:height_cm:dribbling
## wage_eur_log:pace:dribbling
## age:pace:dribbling
## height_cm:pace:dribbling
## wage_eur_log:age:defending
## wage_eur_log:height_cm:defending
## age:height_cm:defending
## wage_eur_log:pace:defending
## age:pace:defending                                            .
## height_cm:pace:defending
## wage_eur_log:dribbling:defending
## age:dribbling:defending
## height_cm:dribbling:defending
## pace:dribbling:defending
## wage_eur_log:age:physic
## wage_eur_log:height_cm:physic
## age:height_cm:physic
## wage_eur_log:pace:physic
## age:pace:physic
## height_cm:pace:physic
## wage_eur_log:dribbling:physic
## age:dribbling:physic
## height_cm:dribbling:physic
## pace:dribbling:physic
## wage_eur_log:defending:physic
## age:defending:physic
## height_cm:defending:physic
## pace:defending:physic
## dribbling:defending:physic
## wage_eur_log:age:power_long_shots
## wage_eur_log:height_cm:power_long_shots
## age:height_cm:power_long_shots
## wage_eur_log:pace:power_long_shots
## age:pace:power_long_shots
## height_cm:pace:power_long_shots
## wage_eur_log:dribbling:power_long_shots                       *
## age:dribbling:power_long_shots
## height_cm:dribbling:power_long_shots                          *
## pace:dribbling:power_long_shots                               .
## wage_eur_log:defending:power_long_shots
## age:defending:power_long_shots
## height_cm:defending:power_long_shots                          .
## pace:defending:power_long_shots
## dribbling:defending:power_long_shots                          *
## wage_eur_log:physic:power_long_shots
## age:physic:power_long_shots
## height_cm:physic:power_long_shots
## pace:physic:power_long_shots
## dribbling:physic:power_long_shots                             *
## defending:physic:power_long_shots                             .
## wage_eur_log:age:height_cm:pace
```

```
## wage_eur_log:age:height_cm:dribbling
## wage_eur_log:age:pace:dribbling
## wage_eur_log:height_cm:pace:dribbling
## age:height_cm:pace:dribbling
## wage_eur_log:age:height_cm:defending
## wage_eur_log:age:pace:defending
## wage_eur_log:height_cm:pace:defending
## age:height_cm:pace:defending                                    .
## wage_eur_log:age:dribbling:defending
## wage_eur_log:height_cm:dribbling:defending
## age:height_cm:dribbling:defending
## wage_eur_log:pace:dribbling:defending
## age:pace:dribbling:defending
## height_cm:pace:dribbling:defending
## wage_eur_log:age:height_cm:physic
## wage_eur_log:age:pace:physic
## wage_eur_log:height_cm:pace:physic
## age:height_cm:pace:physic
## wage_eur_log:age:dribbling:physic
## wage_eur_log:height_cm:dribbling:physic
## age:height_cm:dribbling:physic
## wage_eur_log:pace:dribbling:physic
## age:pace:dribbling:physic
## height_cm:pace:dribbling:physic
## wage_eur_log:age:defending:physic
## wage_eur_log:height_cm:defending:physic
## age:height_cm:defending:physic
## wage_eur_log:pace:defending:physic
## age:pace:defending:physic                                       *
## height_cm:pace:defending:physic
## wage_eur_log:dribbling:defending:physic
## age:dribbling:defending:physic
## height_cm:dribbling:defending:physic
## pace:dribbling:defending:physic
## wage_eur_log:age:height_cm:power_long_shots
## wage_eur_log:age:pace:power_long_shots
## wage_eur_log:height_cm:pace:power_long_shots
## age:height_cm:pace:power_long_shots
## wage_eur_log:age:dribbling:power_long_shots
## wage_eur_log:height_cm:dribbling:power_long_shots               *
## age:height_cm:dribbling:power_long_shots
## wage_eur_log:pace:dribbling:power_long_shots
## age:pace:dribbling:power_long_shots
## height_cm:pace:dribbling:power_long_shots                       .
## wage_eur_log:age:defending:power_long_shots
## wage_eur_log:height_cm:defending:power_long_shots               .
## age:height_cm:defending:power_long_shots
## wage_eur_log:pace:defending:power_long_shots
## age:pace:defending:power_long_shots
## height_cm:pace:defending:power_long_shots
## wage_eur_log:dribbling:defending:power_long_shots               *
## age:dribbling:defending:power_long_shots
## height_cm:dribbling:defending:power_long_shots                  *
## pace:dribbling:defending:power_long_shots
## wage_eur_log:age:physic:power_long_shots
## wage_eur_log:height_cm:physic:power_long_shots
## age:height_cm:physic:power_long_shots
## wage_eur_log:pace:physic:power_long_shots
## age:pace:physic:power_long_shots
## height_cm:pace:physic:power_long_shots
## wage_eur_log:dribbling:physic:power_long_shots                  .
## age:dribbling:physic:power_long_shots
## height_cm:dribbling:physic:power_long_shots                     *
## pace:dribbling:physic:power_long_shots
## wage_eur_log:defending:physic:power_long_shots                  .
## age:defending:physic:power_long_shots
## height_cm:defending:physic:power_long_shots                     .
## pace:defending:physic:power_long_shots
## dribbling:defending:physic:power_long_shots                     *
## wage_eur_log:age:height_cm:pace:dribbling
## wage_eur_log:age:height_cm:pace:defending                       .
## wage_eur_log:age:height_cm:dribbling:defending
## wage_eur_log:age:pace:dribbling:defending
## wage_eur_log:height_cm:pace:dribbling:defending
## age:height_cm:pace:dribbling:defending
## wage_eur_log:age:height_cm:pace:physic
## wage_eur_log:age:height_cm:dribbling:physic
## wage_eur_log:age:pace:dribbling:physic
## wage_eur_log:height_cm:pace:dribbling:physic
## age:height_cm:pace:dribbling:physic
## wage_eur_log:age:height_cm:defending:physic
## wage_eur_log:age:pace:defending:physic                          *
## wage_eur_log:height_cm:pace:defending:physic
## age:height_cm:pace:defending:physic                             *
```

```
## wage_eur_log:age:dribbling:defending:physic
## wage_eur_log:height_cm:dribbling:defending:physic
## age:height_cm:dribbling:defending:physic
## wage_eur_log:pace:dribbling:defending:physic
## age:pace:dribbling:defending:physic
## height_cm:pace:dribbling:defending:physic
## wage_eur_log:age:height_cm:pace:power_long_shots
## wage_eur_log:age:height_cm:dribbling:power_long_shots
## wage_eur_log:age:pace:dribbling:power_long_shots
## wage_eur_log:height_cm:pace:dribbling:power_long_shots
## age:height_cm:pace:dribbling:power_long_shots
## wage_eur_log:age:height_cm:defending:power_long_shots
## wage_eur_log:age:pace:defending:power_long_shots
## wage_eur_log:height_cm:pace:defending:power_long_shots
## age:height_cm:pace:defending:power_long_shots
## wage_eur_log:age:dribbling:defending:power_long_shots
## wage_eur_log:height_cm:dribbling:defending:power_long_shots                  *
## age:height_cm:dribbling:defending:power_long_shots
## wage_eur_log:pace:dribbling:defending:power_long_shots                       .
## age:pace:dribbling:defending:power_long_shots
## height_cm:pace:dribbling:defending:power_long_shots
## wage_eur_log:age:height_cm:physic:power_long_shots
## wage_eur_log:age:pace:physic:power_long_shots
## wage_eur_log:height_cm:pace:physic:power_long_shots
## age:height_cm:pace:physic:power_long_shots
## wage_eur_log:age:dribbling:physic:power_long_shots
## wage_eur_log:height_cm:dribbling:physic:power_long_shots                     .
## age:height_cm:dribbling:physic:power_long_shots
## wage_eur_log:pace:dribbling:physic:power_long_shots
## age:pace:dribbling:physic:power_long_shots
## height_cm:pace:dribbling:physic:power_long_shots
## wage_eur_log:age:defending:physic:power_long_shots
## wage_eur_log:height_cm:defending:physic:power_long_shots                     .
## age:height_cm:defending:physic:power_long_shots
## wage_eur_log:pace:defending:physic:power_long_shots
## age:pace:defending:physic:power_long_shots
## height_cm:pace:defending:physic:power_long_shots
## wage_eur_log:dribbling:defending:physic:power_long_shots                     **
## age:dribbling:defending:physic:power_long_shots
## height_cm:dribbling:defending:physic:power_long_shots                        *
## pace:dribbling:defending:physic:power_long_shots                             .
## wage_eur_log:age:height_cm:pace:dribbling:defending
## wage_eur_log:age:height_cm:pace:dribbling:physic
## wage_eur_log:age:height_cm:pace:defending:physic                             *
## wage_eur_log:age:height_cm:dribbling:defending:physic
## wage_eur_log:age:pace:dribbling:defending:physic
## wage_eur_log:height_cm:pace:dribbling:defending:physic
## age:height_cm:pace:dribbling:defending:physic
## wage_eur_log:age:height_cm:pace:dribbling:power_long_shots
## wage_eur_log:age:height_cm:pace:defending:power_long_shots
## wage_eur_log:age:height_cm:dribbling:defending:power_long_shots
## wage_eur_log:age:pace:dribbling:defending:power_long_shots
## wage_eur_log:height_cm:pace:dribbling:defending:power_long_shots             .
## age:height_cm:pace:dribbling:defending:power_long_shots
## wage_eur_log:age:height_cm:pace:physic:power_long_shots
## wage_eur_log:age:height_cm:dribbling:physic:power_long_shots
## wage_eur_log:age:pace:dribbling:physic:power_long_shots
## wage_eur_log:height_cm:pace:dribbling:physic:power_long_shots
## age:height_cm:pace:dribbling:physic:power_long_shots
## wage_eur_log:age:height_cm:defending:physic:power_long_shots
## wage_eur_log:age:pace:defending:physic:power_long_shots
## wage_eur_log:height_cm:pace:defending:physic:power_long_shots
## age:height_cm:pace:defending:physic:power_long_shots
## wage_eur_log:age:dribbling:defending:physic:power_long_shots
## wage_eur_log:height_cm:dribbling:defending:physic:power_long_shots           **
## age:height_cm:dribbling:defending:physic:power_long_shots
## wage_eur_log:pace:dribbling:defending:physic:power_long_shots                *
## age:pace:dribbling:defending:physic:power_long_shots
## height_cm:pace:dribbling:defending:physic:power_long_shots                   .
## wage_eur_log:age:height_cm:pace:dribbling:defending:physic
## wage_eur_log:age:height_cm:pace:dribbling:defending:power_long_shots
## wage_eur_log:age:height_cm:pace:dribbling:physic:power_long_shots
## wage_eur_log:age:height_cm:pace:defending:physic:power_long_shots            .
## wage_eur_log:age:height_cm:dribbling:defending:physic:power_long_shots
## wage_eur_log:age:pace:dribbling:defending:physic:power_long_shots
## wage_eur_log:height_cm:pace:dribbling:defending:physic:power_long_shots      **
## age:height_cm:pace:dribbling:defending:physic:power_long_shots
## wage_eur_log:age:height_cm:pace:dribbling:defending:physic:power_long_shots
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.196 on 251 degrees of freedom
## Multiple R-squared:  0.9422, Adjusted R-squared:  0.8818
## F-statistic: 15.61 on 262 and 251 DF,  p-value: < 0.00000000000000022
```

Coefficients `I(wage_eur_log^2)`, `I(age^2)`, `I(defending^2)`, `dribbling:power_long_shots` are significant here. There are other significant coefficients as well but they will not be used because it makes the model too complex (e.g. `wage_eur_log:dribbling:power_long_shots`)

model1

Starting with a very simple model:

```
mr_model1 <- lm(potential ~ wage_eur_log +
                            age +
                            height_cm +
                            pace +
                            dribbling +
                            defending +
                            physic +
                            power_long_shots, data=df)
summary(mr_model1)
```

```
##
## Call:
## lm(formula = potential ~ wage_eur_log + age + height_cm + pace +
##     dribbling + defending + physic + power_long_shots, data = df)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -11.5916  -2.4986  -0.3558   2.2130  12.2437
##
## Coefficients:
##                  Estimate Std. Error t value             Pr(>|t|)
## (Intercept)      37.51410    6.01978   6.232       0.000000000973 ***
## wage_eur_log      1.53614    0.15548   9.880 < 0.0000000000000002 ***
## age              -0.85509    0.04340 -19.704 < 0.0000000000000002 ***
## height_cm         0.07956    0.03186   2.497               0.0128 *
## pace             -0.03197    0.01965  -1.627               0.1045
## dribbling         0.31234    0.03173   9.844 < 0.0000000000000002 ***
## defending         0.07029    0.01333   5.274       0.000000197843 ***
## physic            0.10910    0.02611   4.179       0.000034561395 ***
## power_long_shots  0.01020    0.01707   0.598               0.5503
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 3.677 on 505 degrees of freedom
## Multiple R-squared:  0.6741, Adjusted R-squared:  0.6689
## F-statistic: 130.6 on 8 and 505 DF,  p-value: < 0.00000000000000022
```

model2

`power_long_shots` has the least significant coefficient. It will be removed.

```
mr_model2 <- update(mr_model1,~.-power_long_shots)
summary(mr_model2)
```

```
##
## Call:
## lm(formula = potential ~ wage_eur_log + age + height_cm + pace +
##     dribbling + defending + physic, data = df)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -11.6786  -2.4719  -0.3568   2.2162  12.4024
##
## Coefficients:
##              Estimate Std. Error t value             Pr(>|t|)
## (Intercept)  37.44423    6.01482   6.225       0.00000000101 ***
## wage_eur_log  1.54437    0.15477   9.979 < 0.0000000000000002 ***
## age          -0.85094    0.04281 -19.876 < 0.0000000000000002 ***
## height_cm     0.07883    0.03182   2.477               0.0136 *
## pace         -0.03466    0.01912  -1.812               0.0705 .
## dribbling     0.32428    0.02465  13.158 < 0.0000000000000002 ***
## defending     0.06771    0.01260   5.375       0.00000011729 ***
## physic        0.11099    0.02590   4.285       0.00002186890 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 3.674 on 506 degrees of freedom
## Multiple R-squared:  0.6739, Adjusted R-squared:  0.6694
## F-statistic: 149.4 on 7 and 506 DF,  p-value: < 0.00000000000000022
```

model3

`pace` has the least significant coefficient. It will be removed.

```
mr_model3 <- update(mr_model2,~.-pace)
summary(mr_model3)
```

```
##
## Call:
## lm(formula = potential ~ wage_eur_log + age + height_cm + dribbling +
##     defending + physic, data = df)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -11.4664  -2.5034  -0.2909   2.2721  12.3385
##
## Coefficients:
##              Estimate Std. Error t value             Pr(>|t|)
## (Intercept)  33.00158    5.50507   5.995        0.000000000387 ***
## wage_eur_log  1.56838    0.15455  10.148 < 0.0000000000000002 ***
## age          -0.83232    0.04165 -19.982 < 0.0000000000000002 ***
## height_cm     0.09518    0.03058   3.112              0.00196 **
## dribbling     0.30461    0.02218  13.736 < 0.0000000000000002 ***
## defending     0.07092    0.01250   5.674        0.000000002347 ***
## physic        0.10384    0.02566   4.047        0.000059990173 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 3.683 on 507 degrees of freedom
## Multiple R-squared:  0.6718, Adjusted R-squared:  0.6679
## F-statistic: 172.9 on 6 and 507 DF,  p-value: < 0.00000000000000022
```

model4

`I(wage_eur_log^2)` will be added since it has been shown in the complex model that it posses a significant coefficient.

```
mr_model4 <- update(mr_model3,~.+I(wage_eur_log^2))
summary(mr_model4)
```

```
##
## Call:
## lm(formula = potential ~ wage_eur_log + age + height_cm + dribbling +
##     defending + physic + I(wage_eur_log^2), data = df)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -10.9439  -2.4533  -0.1034   2.2268  11.9656
##
## Coefficients:
##                   Estimate Std. Error t value             Pr(>|t|)
## (Intercept)       72.71138    7.52941   9.657 < 0.0000000000000002 ***
## wage_eur_log      -7.11112    1.19123  -5.970    0.000000004480176 ***
## age               -0.81786    0.03969 -20.608 < 0.0000000000000002 ***
## height_cm          0.07950    0.02918   2.724              0.00667 **
## dribbling          0.27281    0.02154  12.664 < 0.0000000000000002 ***
## defending          0.05894    0.01201   4.909    0.000001235064317 ***
## physic             0.11235    0.02444   4.596    0.000005433483767 ***
## I(wage_eur_log^2)  0.51938    0.07074   7.342    0.000000000000844 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 3.504 on 506 degrees of freedom
## Multiple R-squared:  0.7034, Adjusted R-squared:  0.6993
## F-statistic: 171.4 on 7 and 506 DF,  p-value: < 0.00000000000000022
```

model5

`I(age^2)` will be added since it has been shown in the complex model that it posses a significant coefficient.

```
mr_model5 <- update(mr_model4,~.+I(age^2))
summary(mr_model5)
```

```
##
## Call:
## lm(formula = potential ~ wage_eur_log + age + height_cm + dribbling +
##     defending + physic + I(wage_eur_log^2) + I(age^2), data = df)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -10.5296  -2.0687  -0.1365   1.9882  11.5235
##
## Coefficients:
##                    Estimate Std. Error t value             Pr(>|t|)
## (Intercept)       112.32393    7.21462  15.569 < 0.0000000000000002 ***
## wage_eur_log       -4.66319    1.05065  -4.438   0.0000111381845827 ***
## age                -4.96452    0.32246 -15.396 < 0.0000000000000002 ***
## height_cm           0.06487    0.02534   2.560               0.0108 *
## dribbling           0.30173    0.01882  16.030 < 0.0000000000000002 ***
## defending           0.06045    0.01042   5.804   0.0000000114695501 ***
## physic              0.17157    0.02169   7.908   0.0000000000000165 ***
## I(wage_eur_log^2)   0.37698    0.06235   6.046   0.0000000028856705 ***
## I(age^2)            0.07864    0.00608  12.934 < 0.0000000000000002 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 3.04 on 505 degrees of freedom
## Multiple R-squared:  0.7772, Adjusted R-squared:  0.7736
## F-statistic: 220.2 on 8 and 505 DF,  p-value: < 0.00000000000000022
```

model6

`I(defending^2)` will be added since it has been shown in the complex model that it posses a significant coefficient.

```
mr_model6 <- update(mr_model5,~.+I(defending^2))
summary(mr_model6)
```

```
##
## Call:
## lm(formula = potential ~ wage_eur_log + age + height_cm + dribbling +
##     defending + physic + I(wage_eur_log^2) + I(age^2) + I(defending^2),
##     data = df)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -10.0805  -2.0434  -0.1617   2.0537  10.3276
##
## Coefficients:
##                     Estimate  Std. Error t value             Pr(>|t|)
## (Intercept)       114.1618122   7.0543648  16.183 < 0.0000000000000002 ***
## wage_eur_log       -3.6916435   1.0437626  -3.537             0.000442 ***
## age                -4.9081759   0.3150715 -15.578 < 0.0000000000000002 ***
## height_cm           0.0636173   0.0247475   2.571             0.010436 *
## dribbling           0.3124827   0.0185030  16.888 < 0.0000000000000002 ***
## defending          -0.2199382   0.0563329  -3.904             0.000107 ***
## physic              0.1647190   0.0212273   7.760   0.0000000000000476 ***
## I(wage_eur_log^2)   0.3073120   0.0624208   4.923   0.0000011554437612 ***
## I(age^2)            0.0776776   0.0059405  13.076 < 0.0000000000000002 ***
## I(defending^2)      0.0029532   0.0005836   5.061   0.0000005867605736 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.969 on 504 degrees of freedom
## Multiple R-squared:  0.7879, Adjusted R-squared:  0.7842
## F-statistic: 208.1 on 9 and 504 DF,  p-value: < 0.00000000000000022
```

model7

`dribbling:power_long_shots` will be added since it has been shown in the complex model that it posses a significant coefficient.

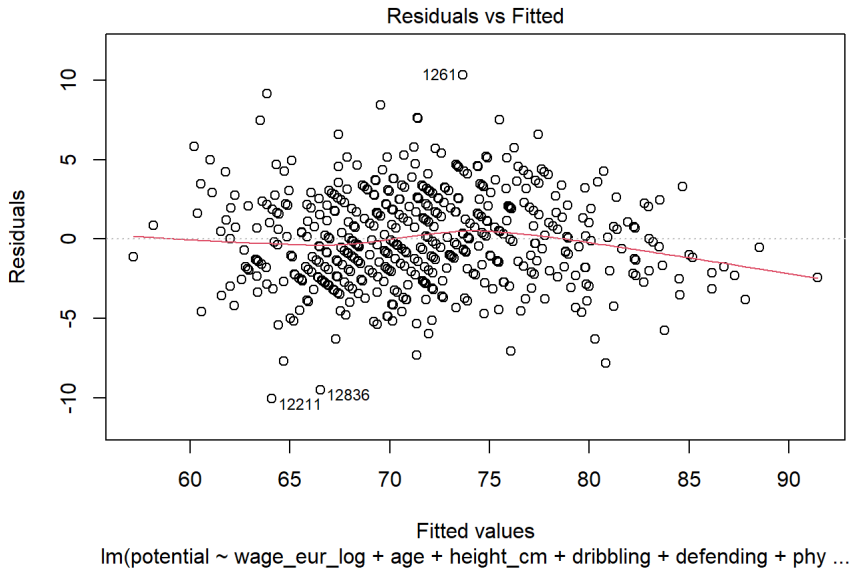
```
mr_model7 <- update(mr_model6,~. + dribbling:power_long_shots)
summary(mr_model7)
```

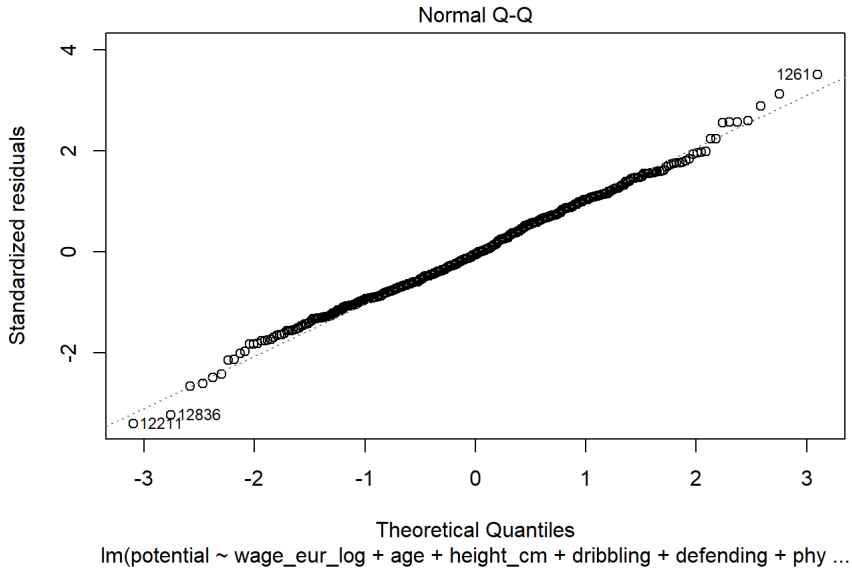
```
##
## Call:
## lm(formula = potential ~ wage_eur_log + age + height_cm + dribbling +
##     defending + physic + I(wage_eur_log^2) + I(age^2) + I(defending^2) +
##     dribbling:power_long_shots, data = df)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -9.9620 -2.0128 -0.1722  1.9967 10.1432
##
## Coefficients:
##                               Estimate  Std. Error t value             Pr(>|t|)
## (Intercept)                114.9109253   7.0808954  16.228 < 0.0000000000000002
## wage_eur_log                -3.6360710   1.0444681  -3.481             0.000542
## age                         -4.8902573   0.3153306 -15.508 < 0.0000000000000002
## height_cm                    0.0630427   0.0247434   2.548             0.011135
## dribbling                    0.2868151   0.0287175   9.987 < 0.0000000000000002
## defending                   -0.2210063   0.0563199  -3.924    0.000099154674066
## physic                       0.1623905   0.0213129   7.619    0.000000000000128
## I(wage_eur_log^2)            0.3021367   0.0625552   4.830    0.000001815737424
## I(age^2)                     0.0771825   0.0059534  12.964 < 0.0000000000000002
## I(defending^2)               0.0030008   0.0005848   5.132    0.000000411107175
## dribbling:power_long_shots   0.0002408   0.0002061   1.168             0.243190
##
## (Intercept)                ***
## wage_eur_log               ***
## age                        ***
## height_cm                  *
## dribbling                  ***
## defending                  ***
## physic                     ***
## I(wage_eur_log^2)          ***
## I(age^2)                   ***
## I(defending^2)             ***
## dribbling:power_long_shots
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.968 on 503 degrees of freedom
## Multiple R-squared:  0.7885, Adjusted R-squared:  0.7843
## F-statistic: 187.5 on 10 and 503 DF,  p-value: < 0.00000000000000022
```

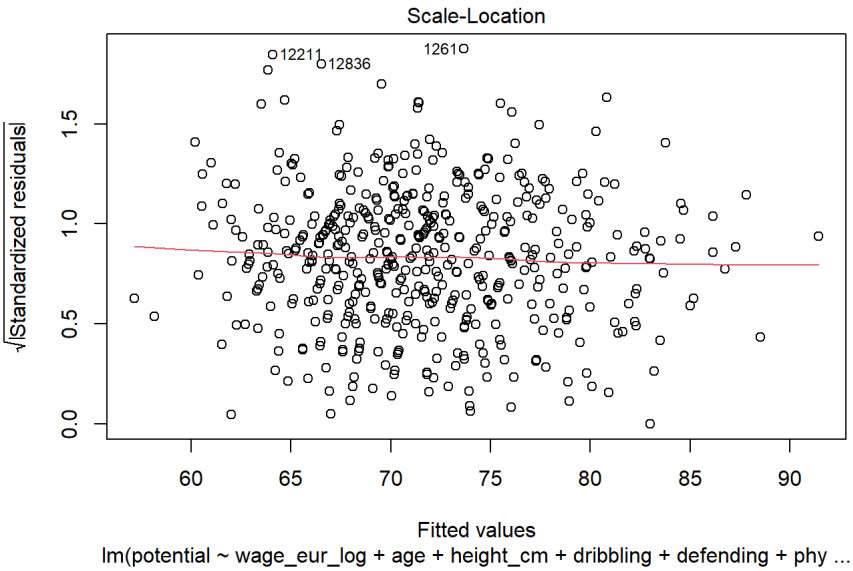
Model 7 shows that `dribbling:power_long_shots` is not significant so we will keep **model6** as our final model.

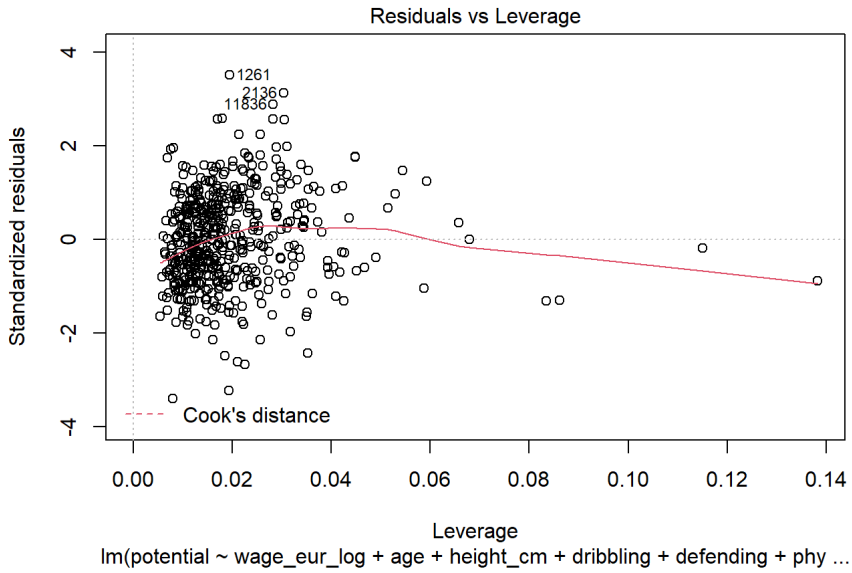
Analyzing model6:

```
plot(mr_model6)
```


Residuals vs Fitted
Residuals
Fitted values
lm(potential ~ wage_eur_log + age + height_cm + dribbling + defending + phy ...
1261
12211
12836


Normal Q-Q
Standardized residuals
Theoretical Quantiles
lm(potential ~ wage_eur_log + age + height_cm + dribbling + defending + phy ...
1261
12211
12836


Scale-Location
√|Standardized residuals|
Fitted values
lm(potential ~ wage_eur_log + age + height_cm + dribbling + defending + phy ...
1261
12211
12836



In the graph “Residuals vs Fitted”, residuals are randomly distributed which means that there are not signs of heteroscedasticity. The Normal Q-Q plot shows that the standardized residuals do follow a normal distribution which means our model is good.

3.2 Critique model using relevant diagnostics

Offer an interpretation of the model characteristics, goodness of fit and graphical diagnostics (5 marks) for the model built in 3.1. Explain any potential weaknesses (5 marks).

- First, I built a very complex model (model0) using the variables `wage_eur_log` , `age` , `height_cm` , `pace` , `dribbling` , `defending` , `physic` , `power_long_shots` , their quadratic values and their interactions between them. They were used since they have been shown to be relevant during the EDA. The output of the model was an R-squared of 0.9422 but a very low F-statistic of 15.61. However, I found that coefficients for `I(wage_eur_log^2)` , `I(age^2)` , `I(defending^2)` , `dribbling:power_long_shots` are significant. Others were too but we ignore them since adding them would make the model very complex.
- Second, I built a very simple model (model1) using the variables `wage_eur_log` , `age` , `height_cm` , `pace` , `dribbling` , `defending` , `physic` , `power_long_shots` . We found that `pace` and `power_long_shots` were the least significant. This model gave us a lower R-squared of 0.6741 but a significant higher F-statistic of 130.6 (improvement of 114.99).
- Third, I built a model (model2) by removing the `power_long_shots` column from model1 since it was no significant in model1. The R-squared slightly decreased to 0.6739 but the F-statistic increased to 149.4. Column `pace` remain to be not significant.
- Fourth, I built a model (model3) by removing the `pace` column from model2 since it was no significant in model2. The R-squared slightly increased from model2 to 0.6718 but still lower than model1 (0.6741). The F-statistic increased to 172.9.
- Fifth, I built a model (model4) by adding `I(wage_eur_log^2)` to model3. The R-squared increased to 0.7034 (the highest so far) and the F-statistic slightly decreased to 171.4.
- Sixth, I built a model (model5) by adding `I(age^2)` to model4. The R-squared increased to 0.7772 (the highest so far) and the F-statistic had a big increased to 220.2 (highest so far).
- Seventh, I built a model (model6) by adding `I(defending^2)` to model5. The R-squared slightly increased to 0.7879 (highest so far). However, the F-statistic decreased to 208.1.
- Finally, I built a model (model7) by adding `dribbling:power_long_shots` to model6. However, this column resulted to be no significant. Not only that, it led to a significant decrease in both R-squared and F-statistic.
- Therefore our final model will be model6 which had the highest R-squared and the second highest F-statistic from all my models. Furthermore, the graph “Residuals vs Fitted” shows that residuals are randomly distributed which means that there are not signs of heteroscedasticity and the Normal Q-Q plot shows that the standardized residuals do follow a normal distribution which means our model is good.
- Our final model is interpreted as follows:

$$potential = 114.1618122 - 3.6916435 \times log(wage\_eur) - 4.9081759 \times age + 0.0636173 \times height\_cm + 0.3124827 \times dribbling - 0.219938$$

- The multi-regression coefficients gives you the size of effect that the feature is having on potential and the sign gives you whether that effect is positive or negative (e.g., for every unit increase in age there is 4.9 decrease in potential).

3.3 Suggest improvements to your model

Based on the findings in 3.2 articulates possible alternative approaches to address them (5 marks).

Since Multiple Regression is an algorithm that uses distance between points for the loss function, variables which range are higher (e.g., `wage_eur_log` range is significantly smaller than the range in `height_cm`) will have a bigger impact in the model without being necessary more important than the other variables. A good way to fix this is to standardize numerical columns. In that way, your columns will follow a normal distribution and have a mean of 0 and a standard deviation of 1. In other words, they will have similar scale and will help the algorithm remove that bias.

4. Extension work

4.1 Model the likelihood of a player having a weekly wage above 8000 Euro (using the high.wage.ind variable provided).

Given this second research question (i.e., involving the binary target attribute) provide a plan of analysis based on relevant EDA for this attribute (10 marks). The model is described, explained and critiqued (10 marks). NB Submissions where suitable models do not have good fit due to the nature of the data will not be penalized.

Plan:

- Column `wage_eur_log` will not be used in the model because `high.wage.ind` is calculated from `wage_eur`. Therefore, when building the model, `wage_eur_log` will perfectly separate 1 and 0s and will lead the model to not converge. [7]
- From the EDA process, I have found that columns `power_long_shots`, `power_strength`, `physic`, `defending`, `dribbling`, `shooting`, `pace`, `weight_kg`, `wage_eur_log`, `age`, `potential` do not follow a normal distribution. Only columns `height_cm`, and `passing` follow a normal distribution.
- In order to avoid collinearity, features that are highly correlated (>=0.8) between each other, will not be used. `height_cm` and `weight_kg` are highly correlated, for this case `height_cm` will be used because it follows a normal distribution. `dribbling` is highly correlated with `shooting` and `passing`, for this case we will use `dibbling` and remove `shooting` and `passing` in the model. In this way, we use one feature instead of two and make the model simplier, and also because dribbling has more correlation with `potential` than the other 2 features. `physic` and `power_strength` are also highly correlated, I will use `physic` and not use `power_strength` because `physic` has a higher correlation with `potential` than `power_strength` with `potential`. In summary, columns `weight_kg`, `shooting`, `passing`, `power_strenght` will not be used in the model.
- `Pace` has been shown to not have statistical significance in the difference in means between `high.wage_ind` = 0 and `high.wage_ind` = 1. I will not use this column in the model.
- `weight_kg` and `pace` have been shown to not have statistical significance in the difference in medians between `high.wage_ind` = 0 and `high.wage_ind` = 1. We don't need to worry since we already said `weight_kg` will not be used to avoid collinearity and we already say we need to be careful with the `pace` column.
- First I will build a complex model and take note of the significant coefficients. Then I will build a very simple model and start adding significant coefficients found in the complex model.

model0

```
lg_model0 <- glm(high.wage.ind ~ age*height_cm*dribbling*defending*physic*power_long_shots +
                                 I(age^2) +
                                 I(height_cm^2) +
                                 I(dribbling^2) +
                                 I(defending^2) +
                                 I(physic^2) +
                                 I(power_long_shots^2) +
                                 I(potential^2), data=df, family=binomial)
```

```
## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred
```

```
summary(lg_model0)
```

```
##
## Call:
## glm(formula = high.wage.ind ~ age * height_cm * dribbling * defending *
##     physic * power_long_shots + I(age^2) + I(height_cm^2) + I(dribbling^2) +
##     I(defending^2) + I(physic^2) + I(power_long_shots^2) + I(potential^2),
##     family = binomial, data = df)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -3.5438  -0.1276  -0.0095   0.0266   2.5214
##
## Coefficients:
##                                                          Estimate
## (Intercept)                                        -104802.945687643
## age                                                   5065.505866037
## height_cm                                              550.708445183
## dribbling                                             1652.854974104
## defending                                              935.947507931
## physic                                                1878.641544470
## power_long_shots                                      1576.187413386
## I(age^2)                                                -0.063979068
## I(height_cm^2)                                           0.002322525
## I(dribbling^2)                                           0.008882192
## I(defending^2)                                           0.003725136
## I(physic^2)                                              0.002383801
## I(power_long_shots^2)                                   -0.000452889
## I(potential^2)                                           0.004064609
## age:height_cm                                          -26.867367385
## age:dribbling                                          -78.921868822
## height_cm:dribbling                                     -8.707971840
## age:defending                                          -57.902904586
## height_cm:defending                                     -4.852886988
## dribbling:defending                                    -15.785583857
## age:physic                                             -84.474018943
## height_cm:physic                                        -9.858653918
## dribbling:physic                                       -29.025406883
## defending:physic                                       -18.762725039
## age:power_long_shots                                   -79.464142280
## height_cm:power_long_shots                              -8.275082835
## dribbling:power_long_shots                             -25.329073554
## defending:power_long_shots                             -12.347570355
## physic:power_long_shots                                -28.733485006
## age:height_cm:dribbling                                  0.418933033
## age:height_cm:defending                                  0.303951247
## age:dribbling:defending                                  0.912710711
## height_cm:dribbling:defending                            0.081821250
## age:height_cm:physic                                     0.446539130
## age:dribbling:physic                                     1.305326838
## height_cm:dribbling:physic                               0.152339634
## age:defending:physic                                     0.996685213
## height_cm:defending:physic                               0.097326785
## dribbling:defending:physic                               0.299600971
## age:height_cm:power_long_shots                           0.421839868
## age:dribbling:power_long_shots                           1.250206036
## height_cm:dribbling:power_long_shots                     0.133300766
## age:defending:power_long_shots                           0.877933090
## height_cm:defending:power_long_shots                     0.063931680
## dribbling:defending:power_long_shots                     0.225766343
## age:physic:power_long_shots                              1.329560143
## height_cm:physic:power_long_shots                        0.150437571
## dribbling:physic:power_long_shots                        0.450780614
## defending:physic:power_long_shots                        0.263605106
## age:height_cm:dribbling:defending                       -0.004775410
## age:height_cm:dribbling:physic                          -0.006901944
## age:height_cm:defending:physic                          -0.005219596
## age:dribbling:defending:physic                          -0.015481253
## height_cm:dribbling:defending:physic                    -0.001552512
## age:height_cm:dribbling:power_long_shots                -0.006643097
## age:height_cm:defending:power_long_shots                -0.004628939
## age:dribbling:defending:power_long_shots                -0.014294147
## height_cm:dribbling:defending:power_long_shots          -0.001175531
## age:height_cm:physic:power_long_shots                   -0.007023534
## age:dribbling:physic:power_long_shots                   -0.020721362
## height_cm:dribbling:physic:power_long_shots             -0.002364090
## age:defending:physic:power_long_shots                   -0.015188846
## height_cm:defending:physic:power_long_shots             -0.001364800
## dribbling:defending:physic:power_long_shots             -0.004440696
## age:height_cm:dribbling:defending:physic                 0.000080878
## age:height_cm:dribbling:defending:power_long_shots       0.000075310
## age:height_cm:dribbling:physic:power_long_shots          0.000109595
## age:height_cm:defending:physic:power_long_shots          0.000079716
## age:dribbling:defending:physic:power_long_shots          0.000242649
## height_cm:dribbling:defending:physic:power_long_shots    0.000023066
```

```
## age:height_cm:dribbling:defending:physic:power_long_shots      -0.000001274
##                                                                  Std. Error
## (Intercept)                                                  135418.612237753
## age                                                            5276.956841708
## height_cm                                                       745.845280639
## dribbling                                                      1976.668166054
## defending                                                      2096.610294464
## physic                                                         2183.865143696
## power_long_shots                                               2306.738032733
## I(age^2)                                                          0.019368303
## I(height_cm^2)                                                    0.007423794
## I(dribbling^2)                                                    0.009734737
## I(defending^2)                                                    0.001456330
## I(physic^2)                                                       0.004622177
## I(power_long_shots^2)                                             0.001969161
## I(potential^2)                                                    0.000771771
## age:height_cm                                                    29.185765624
## age:dribbling                                                    77.201417306
## height_cm:dribbling                                              10.895064995
## age:defending                                                    81.724976387
## height_cm:defending                                              11.540210209
## dribbling:defending                                              30.887492019
## age:physic                                                       84.226520790
## height_cm:physic                                                 11.975421007
## dribbling:physic                                                 31.896950820
## defending:physic                                                 32.814751850
## age:power_long_shots                                             89.866121064
## height_cm:power_long_shots                                       12.693723134
## dribbling:power_long_shots                                       33.473916643
## defending:power_long_shots                                       36.050875450
## physic:power_long_shots                                          36.856371310
## age:height_cm:dribbling                                           0.427309032
## age:height_cm:defending                                           0.451460316
## age:dribbling:defending                                           1.206044413
## height_cm:dribbling:defending                                     0.170373429
## age:height_cm:physic                                              0.463373162
## age:dribbling:physic                                              1.232081610
## height_cm:dribbling:physic                                        0.175024904
## age:defending:physic                                              1.266252832
## height_cm:defending:physic                                        0.180022446
## dribbling:defending:physic                                        0.483524559
## age:height_cm:power_long_shots                                    0.496177307
## age:dribbling:power_long_shots                                    1.305359609
## height_cm:dribbling:power_long_shots                              0.184379318
## age:defending:power_long_shots                                    1.404063364
## height_cm:defending:power_long_shots                              0.198278784
## dribbling:defending:power_long_shots                              0.527728112
## age:physic:power_long_shots                                       1.419044162
## height_cm:physic:power_long_shots                                 0.202059382
## dribbling:physic:power_long_shots                                 0.535302914
## defending:physic:power_long_shots                                 0.558935651
## age:height_cm:dribbling:defending                                 0.006674756
## age:height_cm:dribbling:physic                                    0.006782747
## age:height_cm:defending:physic                                    0.006967519
## age:dribbling:defending:physic                                    0.018688359
## height_cm:dribbling:defending:physic                              0.002657522
## age:height_cm:dribbling:power_long_shots                          0.007213850
## age:height_cm:defending:power_long_shots                          0.007743780
## age:dribbling:defending:power_long_shots                          0.020565429
## height_cm:dribbling:defending:power_long_shots                    0.002908870
## age:height_cm:physic:power_long_shots                             0.007798884
## age:dribbling:physic:power_long_shots                             0.020621541
## height_cm:dribbling:physic:power_long_shots                       0.002937000
## age:defending:physic:power_long_shots                             0.021530624
## height_cm:defending:physic:power_long_shots                       0.003065135
## dribbling:defending:physic:power_long_shots                       0.008186648
## age:height_cm:dribbling:defending:physic                          0.000103005
## age:height_cm:dribbling:defending:power_long_shots                0.000113644
## age:height_cm:dribbling:physic:power_long_shots                   0.000113423
## age:height_cm:defending:physic:power_long_shots                   0.000118335
## age:dribbling:defending:physic:power_long_shots                   0.000315523
## height_cm:dribbling:defending:physic:power_long_shots             0.000044983
## age:height_cm:dribbling:defending:physic:power_long_shots         0.000001737
##                                                                  z value    Pr(>|z|)
## (Intercept)                                                       -0.774    0.438979
## age                                                                0.960    0.337091
## height_cm                                                          0.738    0.460291
## dribbling                                                          0.836    0.403052
## defending                                                          0.446    0.655301
## physic                                                             0.860    0.389658
## power_long_shots                                                   0.683    0.494419
## I(age^2)                                                          -3.303    0.000956
## I(height_cm^2)                                                     0.313    0.754396
## I(dribbling^2)                                                     0.912    0.361546
```

```
## I(defending^2)                                          2.558    0.010531
## I(physic^2)                                             0.516    0.606042
## I(power_long_shots^2)                                  -0.230    0.818099
## I(potential^2)                                          5.267 0.000000139
## age:height_cm                                          -0.921    0.357278
## age:dribbling                                          -1.022    0.306646
## height_cm:dribbling                                    -0.799    0.424141
## age:defending                                          -0.709    0.478629
## height_cm:defending                                    -0.421    0.674106
## dribbling:defending                                    -0.511    0.609304
## age:physic                                             -1.003    0.315891
## height_cm:physic                                       -0.823    0.410371
## dribbling:physic                                       -0.910    0.362836
## defending:physic                                       -0.572    0.567473
## age:power_long_shots                                   -0.884    0.376561
## height_cm:power_long_shots                             -0.652    0.514463
## dribbling:power_long_shots                             -0.757    0.449241
## defending:power_long_shots                             -0.343    0.731972
## physic:power_long_shots                                -0.780    0.435622
## age:height_cm:dribbling                                 0.980    0.326890
## age:height_cm:defending                                 0.673    0.500780
## age:dribbling:defending                                 0.757    0.449181
## height_cm:dribbling:defending                           0.480    0.631052
## age:height_cm:physic                                    0.964    0.335211
## age:dribbling:physic                                    1.059    0.289396
## height_cm:dribbling:physic                              0.870    0.384088
## age:defending:physic                                    0.787    0.431215
## height_cm:defending:physic                              0.541    0.588758
## dribbling:defending:physic                              0.620    0.535509
## age:height_cm:power_long_shots                          0.850    0.395225
## age:dribbling:power_long_shots                          0.958    0.338190
## height_cm:dribbling:power_long_shots                    0.723    0.469698
## age:defending:power_long_shots                          0.625    0.531787
## height_cm:defending:power_long_shots                    0.322    0.747124
## dribbling:defending:power_long_shots                    0.428    0.668791
## age:physic:power_long_shots                             0.937    0.348789
## height_cm:physic:power_long_shots                       0.745    0.456561
## dribbling:physic:power_long_shots                       0.842    0.399730
## defending:physic:power_long_shots                       0.472    0.637198
## age:height_cm:dribbling:defending                      -0.715    0.474335
## age:height_cm:dribbling:physic                         -1.018    0.308881
## age:height_cm:defending:physic                         -0.749    0.453777
## age:dribbling:defending:physic                         -0.828    0.407450
## height_cm:dribbling:defending:physic                   -0.584    0.559089
## age:height_cm:dribbling:power_long_shots               -0.921    0.357113
## age:height_cm:defending:power_long_shots               -0.598    0.549999
## age:dribbling:defending:power_long_shots               -0.695    0.487020
## height_cm:dribbling:defending:power_long_shots         -0.404    0.686125
## age:height_cm:physic:power_long_shots                  -0.901    0.367811
## age:dribbling:physic:power_long_shots                  -1.005    0.314974
## height_cm:dribbling:physic:power_long_shots            -0.805    0.420858
## age:defending:physic:power_long_shots                  -0.705    0.480528
## height_cm:defending:physic:power_long_shots            -0.445    0.656128
## dribbling:defending:physic:power_long_shots            -0.542    0.587521
## age:height_cm:dribbling:defending:physic                0.785    0.432347
## age:height_cm:dribbling:defending:power_long_shots      0.663    0.507535
## age:height_cm:dribbling:physic:power_long_shots         0.966    0.333921
## age:height_cm:defending:physic:power_long_shots         0.674    0.500535
## age:dribbling:defending:physic:power_long_shots         0.769    0.441871
## height_cm:dribbling:defending:physic:power_long_shots   0.513    0.608105
## age:height_cm:dribbling:defending:physic:power_long_shots -0.733  0.463524
##
## (Intercept)
## age
## height_cm
## dribbling
## defending
## physic
## power_long_shots
## I(age^2)                                              ***
## I(height_cm^2)
## I(dribbling^2)
## I(defending^2)                                        *
## I(physic^2)
## I(power_long_shots^2)
## I(potential^2)                                        ***
## age:height_cm
## age:dribbling
## height_cm:dribbling
## age:defending
## height_cm:defending
## dribbling:defending
## age:physic
## height_cm:physic
```

```
## dribbling:physic
## defending:physic
## age:power_long_shots
## height_cm:power_long_shots
## dribbling:power_long_shots
## defending:power_long_shots
## physic:power_long_shots
## age:height_cm:dribbling
## age:height_cm:defending
## age:dribbling:defending
## height_cm:dribbling:defending
## age:height_cm:physic
## age:dribbling:physic
## height_cm:dribbling:physic
## age:defending:physic
## height_cm:defending:physic
## dribbling:defending:physic
## age:height_cm:power_long_shots
## age:dribbling:power_long_shots
## height_cm:dribbling:power_long_shots
## age:defending:power_long_shots
## height_cm:defending:power_long_shots
## dribbling:defending:power_long_shots
## age:physic:power_long_shots
## height_cm:physic:power_long_shots
## dribbling:physic:power_long_shots
## defending:physic:power_long_shots
## age:height_cm:dribbling:defending
## age:height_cm:dribbling:physic
## age:height_cm:defending:physic
## age:dribbling:defending:physic
## height_cm:dribbling:defending:physic
## age:height_cm:dribbling:power_long_shots
## age:height_cm:defending:power_long_shots
## age:dribbling:defending:power_long_shots
## height_cm:dribbling:defending:power_long_shots
## age:height_cm:physic:power_long_shots
## age:dribbling:physic:power_long_shots
## height_cm:dribbling:physic:power_long_shots
## age:defending:physic:power_long_shots
## height_cm:defending:physic:power_long_shots
## dribbling:defending:physic:power_long_shots
## age:height_cm:dribbling:defending:physic
## age:height_cm:dribbling:defending:power_long_shots
## age:height_cm:dribbling:physic:power_long_shots
## age:height_cm:defending:physic:power_long_shots
## age:dribbling:defending:physic:power_long_shots
## height_cm:dribbling:defending:physic:power_long_shots
## age:height_cm:dribbling:defending:physic:power_long_shots
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 618.90  on 513  degrees of freedom
## Residual deviance: 174.95  on 443  degrees of freedom
## AIC: 316.95
##
## Number of Fisher Scoring iterations: 11
```

Columns `I(age^2)`, `I(defending^2)`, `I(potential^2)` were found to be significant.

model1

Simple model is built.

```
lg_model1 <- glm(high.wage.ind ~ age +
                                 height_cm +
                                 dribbling +
                                 defending +
                                 physic +
                                 power_long_shots +
                                 potential, data=df, family=binomial)
summary(lg_model1)
```

```
##
## Call:
## glm(formula = high.wage.ind ~ age + height_cm + dribbling + defending +
##     physic + power_long_shots + potential, family = binomial,
##     data = df)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -2.5562  -0.3949  -0.1215   0.1460   2.5659
##
## Coefficients:
##                   Estimate Std. Error z value            Pr(>|z|)
## (Intercept)      -57.201931   7.230179  -7.912 0.00000000000000254 ***
## age                0.272083   0.054216   5.018 0.00000052078334155 ***
## height_cm          0.070703   0.029736   2.378            0.017422 *
## dribbling          0.112784   0.031888   3.537            0.000405 ***
## defending          0.029214   0.011970   2.441            0.014667 *
## physic            -0.003968   0.023945  -0.166            0.868375
## power_long_shots   0.026991   0.015248   1.770            0.076711 .
## potential          0.359693   0.048786   7.373 0.00000000000016692 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 618.9  on 513  degrees of freedom
## Residual deviance: 278.6  on 506  degrees of freedom
## AIC: 294.6
##
## Number of Fisher Scoring iterations: 7
```

model2

Now I will add the significant coefficients found on model0 to model2 which are `I(age^2)`, `I(defending^2)`, and `I(potential^2)`.

```
lg_model2 <- glm(high.wage.ind ~ age +
                                 height_cm +
                                 dribbling +
                                 defending +
                                 physic +
                                 power_long_shots +
                                 potential +
                                 I(age^2) +
                                 I(defending^2) +
                                 I(potential^2), data=df, family=binomial)
summary(lg_model2)
```

```
##
## Call:
## glm(formula = high.wage.ind ~ age + height_cm + dribbling + defending +
##     physic + power_long_shots + potential + I(age^2) + I(defending^2) +
##     I(potential^2), family = binomial, data = df)
##
## Deviance Residuals:
##      Min        1Q    Median        3Q       Max
## -2.71845  -0.30933  -0.06546   0.14135   2.45539
##
## Coefficients:
##                     Estimate   Std. Error z value  Pr(>|z|)
## (Intercept)      -144.2591414   41.5384562  -3.473  0.000515 ***
## age                 2.4575823    0.6037411   4.071 0.0000469 ***
## height_cm           0.0489837    0.0319654   1.532  0.125425
## dribbling           0.0755562    0.0385326   1.961  0.049898 *
## defending          -0.2104275    0.0854109  -2.464  0.013751 *
## physic             -0.0378334    0.0274971  -1.376  0.168852
## power_long_shots    0.0385787    0.0169750   2.273  0.023046 *
## potential           2.2036432    1.1110701   1.983  0.047328 *
## I(age^2)           -0.0399280    0.0108332  -3.686  0.000228 ***
## I(defending^2)      0.0024210    0.0008875   2.728  0.006373 **
## I(potential^2)     -0.0118716    0.0074457  -1.594  0.110839
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 618.90  on 513  degrees of freedom
## Residual deviance: 240.12  on 503  degrees of freedom
## AIC: 262.12
##
## Number of Fisher Scoring iterations: 7
```

model3

`physic` column is not a significant coefficient so it will be removed from model2.

```
lg_model3 <- update(lg_model2,~.-physic)

summary(lg_model3)
```

```
##
## Call:
## glm(formula = high.wage.ind ~ age + height_cm + dribbling + defending +
##     power_long_shots + potential + I(age^2) + I(defending^2) +
##     I(potential^2), family = binomial, data = df)
##
## Deviance Residuals:
##      Min        1Q    Median        3Q       Max
## -2.80516  -0.32330  -0.06313   0.14219   2.45454
##
## Coefficients:
##                       Estimate   Std. Error z value  Pr(>|z|)
## (Intercept)      -142.9301065   40.8844611  -3.496  0.000472 ***
## age                 2.2328591    0.5733234   3.895 0.0000984 ***
## height_cm           0.0329031    0.0296654   1.109  0.267370
## dribbling           0.0860689    0.0379864   2.266  0.023465 *
## defending          -0.2188485    0.0845144  -2.589  0.009612 **
## power_long_shots    0.0352084    0.0166060   2.120  0.033987 *
## potential           2.2935743    1.0912843   2.102  0.035578 *
## I(age^2)           -0.0363405    0.0103747  -3.503  0.000460 ***
## I(defending^2)      0.0024432    0.0008817   2.771  0.005586 **
## I(potential^2)     -0.0126722    0.0073044  -1.735  0.082763 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 618.90  on 513  degrees of freedom
## Residual deviance: 242.06  on 504  degrees of freedom
## AIC: 262.06
##
## Number of Fisher Scoring iterations: 7
```

model4

`height_cm` column is not a significant coefficient so it will be removed from model3.

```
lg_model4 <- update(lg_model3,~.-height_cm)

summary(lg_model4)
```

```
##
## Call:
## glm(formula = high.wage.ind ~ age + dribbling + defending + power_long_shots +
##     potential + I(age^2) + I(defending^2) + I(potential^2), family = binomial,
##     data = df)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -2.8620  -0.3202  -0.0668   0.1361   2.5007
##
## Coefficients:
##                       Estimate   Std. Error z value  Pr(>|z|)
## (Intercept)      -138.8253817   40.4575875  -3.431  0.000601 ***
## age                 2.3847270    0.5613062   4.249 0.0000215 ***
## dribbling           0.0656418    0.0329937   1.990  0.046643 *
## defending          -0.2202835    0.0837244  -2.631  0.008512 **
## power_long_shots    0.0379794    0.0163750   2.319  0.020376 *
## potential           2.2971500    1.0862434   2.115  0.034450 *
## I(age^2)           -0.0389509    0.0101871  -3.824  0.000132 ***
## I(defending^2)      0.0024510    0.0008724   2.810  0.004961 **
## I(potential^2)     -0.0125484    0.0072689  -1.726  0.084290 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 618.9  on 513  degrees of freedom
## Residual deviance: 243.3  on 505  degrees of freedom
## AIC: 261.3
##
## Number of Fisher Scoring iterations: 7
```

model5

`I(potential^2)` has been shown to have low significance. Therefore it will be removed from model 4.

```
lg_model5 <- update(lg_model4,~.-I(potential^2))

summary(lg_model5)
```

```
##
## Call:
## glm(formula = high.wage.ind ~ age + dribbling + defending + power_long_shots +
##     potential + I(age^2) + I(defending^2), family = binomial,
##     data = df)
##
## Deviance Residuals:
##      Min        1Q    Median        3Q       Max
## -2.86148  -0.32449  -0.08563   0.11175   2.45009
##
## Coefficients:
##                     Estimate  Std. Error z value           Pr(>|z|)
## (Intercept)      -73.0617277  10.3315670  -7.072 0.0000000000015305 ***
## age                2.6121957   0.5614537   4.653 0.0000032784273046 ***
## dribbling          0.0615119   0.0324929   1.893            0.05835 .
## defending         -0.2239394   0.0830827  -2.695            0.00703 **
## power_long_shots   0.0393219   0.0164313   2.393            0.01671 *
## potential          0.4380596   0.0567646   7.717 0.0000000000000119 ***
## I(age^2)          -0.0429493   0.0101425  -4.235 0.0000228961024058 ***
## I(defending^2)     0.0024820   0.0008672   2.862            0.00421 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 618.90  on 513  degrees of freedom
## Residual deviance: 246.36  on 506  degrees of freedom
## AIC: 262.36
##
## Number of Fisher Scoring iterations: 7
```

Description, explanation and critique of the model:

- First, I built a complex model with the features that were found important for `high.wage.ind` during the EDA with their quadratic relationship and interactions between them. `I(age^2)`, `I(defending^2)`, `I(potential^2)` were found to be significant. The AIC score of model0 is 316.95
- Second, I built a very simple model (model1). The column `physic` was no significant and column `power_long_shots` had very small significance. The AIC score decrease to 294.6 (lowest so far).
- Third, I added the significant coefficients found in model0 which were `I(age^2)`, `I(defending^2)`, and `I(potential^2)` to model1 to build model2. The AIC score decrease to 262.12 (lowest so far).
- Fourth, for model3, I removed column `physic` from model2 because it was not significant. The AIC score slightly decreased to 262.06 (lowest so far).
- Fifth, for model4, I removed column `height_cm` from model3 because it was not significant. The AIC score slightly decreased to 261.3 (lowest so far).
- Finally, for model5, I removed `I(potential^2)` from model4 because it had very low significance. However, the AIC score slightly increased to 262.36.
- Since model4 is the model which has the smallest AIC score, then this will be our model I will use.
- The model can be interpreted as follows:

$$log(\frac{p}{1-p}) = -138.8253817 + 2.3847270 \times age + 0.0656418 \times dribbling - 0.2202835 \times defending + 0.0379794 \times power\_long\_shots + 2.$$

- The logistic regression coefficients give the change in the log odds of the outome for a one unit increase in the explanatory variable (e.g., for every unit change for age the log odds of making having a weekly wage above 8000 increases by 2.38).
- To improve the model, same concept for multi-regression applies. Since Logistic Regression is an algorithm that uses distance between points for the loss function, variables which range are higher (e.g., `wage_eur_log` range is significantly smaller than the range in `height_cm`) will have a bigger impact in the model without being necessary more important than the other variables. A good way to fix this is to standardize numerical columns. In that way, your columns will follow a normal distribution and have a mean of 0 and a standard deviation of 1. In other words, they will have similar scale and will help the algorithm remove that bias.
- Another issue found in the model is that the variable `high.wage_ind` is imbalanced which means that the model will have bias towards learning more about values 1 than from values 0. A way to fix this is to apply a “rare events correction to the intercept” [8]

References

Add any references here. NB You can either do this manually or automatically with a `.bib` file (which then must be submitted along with your `.Rmd` file). See the RMarkdown documentation (https://bookdown.org/yihui/rmarkdown-cookbook/bibliography.html) for guidance.

[1] J. Press, “The Japan News,” 13 August 2021. [Online]. Available: https://the-japan-news.com/news/article/0007672187 (https://the-japan-news.com/news/article/0007672187). [Accessed 05 January 2022].

[2] FIFA, “FIFA,” 25 February 2021. [Online]. Available: https://www.fifa.com/news/kazuyoshi-miura-i-don-t-think-i-ll-ever-leave-football (https://www.fifa.com/news/kazuyoshi-miura-i-don-t-think-i-ll-ever-leave-football). [Accessed 06 January 2022].

[3] “Transfer Markt,” [Online]. Available: https://www.transfermarkt.com/kyle-hudlin/profil/spieler/829766 (https://www.transfermarkt.com/kyle-hudlin/profil/spieler/829766). [Accessed 06 January 2022].

[4] “Colgadosporelfutbol.com,” [Online]. Available: https://colgadosporelfutbol.com/el-peso-ideal-de-un-futbolista-profesional-y-su-importancia/# (https://colgadosporelfutbol.com/el-peso-ideal-de-un-futbolista-profesional-y-su-importancia/#):~:text=El%20margen%20estimado%20es%20de,por%20cada%202.5%20cent%C3%ADmetros%20adicionales. [Accessed 06 January 2022].

[5] S. Glen, “Statistics How to,” [Online]. Available: https://www.statisticshowto.com/logarithmic-distribution/ (https://www.statisticshowto.com/logarithmic-distribution/). [Accessed 06 January 2022].

[6] K. S. Htoon, “Medium,” 29 February 2020. [Online]. Available: https://medium.com/@kyawsawhtoon/log-transformation-purpose-and-interpretation-9444b4b049c9 (https://medium.com/@kyawsawhtoon/log-transformation-purpose-and-interpretation-9444b4b049c9). [Accessed 06 January 2022].

[7] “Statology,” 01 October 2021. [Online]. Available: https://www.statology.org/glm-fit-algorithm-did-not-converge/ (https://www.statology.org/glm-fit-algorithm-did-not-converge/). [Accessed 06 January 2022].

[8] G. King y L. Zeng, «Logistic Regression in Rare,» Harvard University, Cambridge, MA, , 2001.