Fifa 19_Overview Analysis

Dylan Brussier

2020-11-23

Contents

1 Introduction

I use an existing database available on kaggle to train myself and play with the different functions and arguments to see what they are doing.

We will see in this report:

- an exploratory Data Analysis
- an overall rating vs age, overall rating vs position
- a teams analysis
- a k-means clustering to try to find "bargains"; ie if there is someone with the same skills/potential, can they be found for a bargain?
- a regression or tree analysis by position to find a pattern with player > 80 overall rating (coming soon)

Data :

Variables	Description	Type of Data	Level of Measurement
X	row number		
ID	unique id for every player	Categorical	Nominal
Name	name	Categorical	
Age	age	Numerical Discrete	Ratio
Photo	url to the player's photo		
Nationality	nationality	Categorical	Nominal
Flag	url to players's country flag		
Overall	overall rating	Numerical Continuous	Ratio
Potential	potential rating	Numerical Continuous	Ratio
Club	current club	Categorical	Nominal
Club.Logo	url to club logo		
Value	current market value	Numerical Continuous	Ratio
Wage	current wage	Numerical Continuous	Ratio
Special			
Preferred Foot	left/right	Categorical	Nominal
International Reputation	rating on scale of 5	Categorical	Ordinal
Weak.Foot	rating on scale of 5	Categorical	Ordinal
Skill.Moves	rating on scale of 5	Categorical	Ordinal
Work.Rate	attack work rate (High or Low) / defence work rate (High or Low)	Categorical	Nominal
Body.Type	body type of player	Categorical	Nominal
Real.Face	Yes / No	Categorical	Nominal
Position	position on the pitch *	Categorical	Nominal
Jersey.Number	jersey number	Numerical Discrete	Ratio
Joined	joined date	Categorical	Nominal
Loaned From	club name if applicable	Categorical	Nominal
Contract.Valid.Until	contract end date	Categorical	Nominal
Height	height of the player	Numerical Continuous	Ratio
Weight	weight of the player	Numerical Continuous	Ratio
LS	rating on scale of 100	Numerical Discrete	Ratio
ST	rating on scale of 100	Numerical Discrete	Ratio
RS	rating on scale of 100	Numerical Discrete	Ratio
LW	rating on scale of 100	Numerical Discrete	Ratio
LF	rating on scale of 100	Numerical Discrete	Ratio
CF	rating on scale of 100	Numerical Discrete	Ratio
RF	rating on scale of 100	Numerical Discrete	Ratio
RW	rating on scale of 100	Numerical Discrete	Ratio
LAM	rating on scale of 100	Numerical Discrete	Ratio
CAM	rating on scale of 100	Numerical Discrete	Ratio

Variables	Description	Type of Data	Level of Measurement
RAM	rating on scale of 100	Numerical Discrete	Ratio
LM	rating on scale of 100	Numerical Discrete	Ratio
LCM	rating on scale of 100	Numerical Discrete	Ratio
CM	rating on scale of 100	Numerical Discrete	Ratio
RCM	rating on scale of 100	Numerical Discrete	Ratio
RM	rating on scale of 100	Numerical Discrete	Ratio
LWB	rating on scale of 100	Numerical Discrete	Ratio
LDM	rating on scale of 100	Numerical Discrete	Ratio
CDM	rating on scale of 100	Numerical Discrete	Ratio
RDM	rating on scale of 100	Numerical Discrete	Ratio
RWB	rating on scale of 100	Numerical Discrete	Ratio
LB	rating on scale of 100	Numerical Discrete	Ratio
LCB	rating on scale of 100	Numerical Discrete	Ratio
CB	rating on scale of 100	Numerical Discrete	Ratio
RCB	rating on scale of 100	Numerical Discrete	Ratio
RB	rating on scale of 100	Numerical Discrete	Ratio
Crossing	rating on scale of 100	Numerical Discrete	Ratio
Finishing	rating on scale of 100	Numerical Discrete	Ratio
HeadingAccuracy	rating on scale of 100	Numerical Discrete	Ratio
ShortPassing	rating on scale of 100	Numerical Discrete	Ratio
Volleys	rating on scale of 100	Numerical Discrete	Ratio
Dribbling	rating on scale of 100	Numerical Discrete	Ratio
Curve	rating on scale of 100	Numerical Discrete	Ratio
FKAccuracy	rating on scale of 100	Numerical Discrete	Ratio
LongPassing	rating on scale of 100	Numerical Discrete	Ratio
BallControl	rating on scale of 100	Numerical Discrete	Ratio
Acceleration	rating on scale of 100	Numerical Discrete	Ratio
SprintSpeed	rating on scale of 100	Numerical Discrete	Ratio
Agility	rating on scale of 100	Numerical Discrete	Ratio
Reactions	rating on scale of 100	Numerical Discrete	Ratio
Balance	rating on scale of 100	Numerical Discrete	Ratio
ShotPower	rating on scale of 100	Numerical Discrete	Ratio
Jumping	rating on scale of 100	Numerical Discrete	Ratio
Stamina	rating on scale of 100	Numerical Discrete	Ratio
Strength	rating on scale of 100	Numerical Discrete	Ratio
LongShots	rating on scale of 100	Numerical Discrete	Ratio
Aggression	rating on scale of 100	Numerical Discrete	Ratio
Interceptions	rating on scale of 100	Numerical Discrete	Ratio
Positioning	rating on scale of 100	Numerical Discrete	Ratio
Vision	rating on scale of 100	Numerical Discrete	Ratio
Penalties	rating on scale of 100	Numerical Discrete	Ratio
Composure	rating on scale of 100	Numerical Discrete	Ratio
Marking	rating on scale of 100	Numerical Discrete	Ratio
StandingTackle	rating on scale of 100	Numerical Discrete	Ratio
SlidingTackle	rating on scale of 100	Numerical Discrete	Ratio
GKDiving	rating on scale of 100	Numerical Discrete	Ratio
GKHandling	rating on scale of 100	Numerical Discrete	Ratio
GKKicking	rating on scale of 100	Numerical Discrete	Ratio
GKPositioning	rating on scale of 100	Numerical Discrete	Ratio
GKReflexes	rating on scale of 100	Numerical Discrete	Ratio
Release.Clause	release clause value	Numerical Continuous	Ratio

- more details about positions : SOCCER POSITIONS & ABBREVIATIONS

2 Data Wrangling

```
## [1] "Structure of our data"

## 'data.frame':    18207 obs. of  89 variables:
##  $ X                     : int  0 1 2 3 4 5 6 7 8 9 ...
##  $ ID                    : int  158023 20801 190871 193080 192985 183277 177003 176580 155862 20038
##  $ Name                  : chr  "L. Messi" "Cristiano Ronaldo" "Neymar Jr" "De Gea" ...
##  $ Age                   : int  31 33 26 27 27 27 32 31 32 25 ...
##  $ Photo                 : chr  "https://cdn.sofifa.org/players/4/19/158023.png" "https://cdn.sofi
##  $ Nationality           : chr  "Argentina" "Portugal" "Brazil" "Spain" ...
##  $ Flag                  : chr  "https://cdn.sofifa.org/flags/52.png" "https://cdn.sofifa.org/flag
##  $ Overall               : int  94 94 92 91 91 91 91 91 91 90 ...
##  $ Potential             : int  94 94 93 93 92 91 91 91 91 93 ...
##  $ Club                  : chr  "FC Barcelona" "Juventus" "Paris Saint-Germain" "Manchester United
##  $ Club.Logo             : chr  "https://cdn.sofifa.org/teams/2/light/241.png" "https://cdn.sofifa
##  $ Value                 : chr  "€110.5M" "€77M" "€118.5M" "€72M" ...
##  $ Wage                  : chr  "€565K" "€405K" "€290K" "€260K" ...
##  $ Special               : int  2202 2228 2143 1471 2281 2142 2280 2346 2201 1331 ...
##  $ Preferred.Foot        : chr  "Left" "Right" "Right" "Right" ...
##  $ International.Reputation: int  5 5 5 4 4 4 4 5 5 3 ...
##  $ Weak.Foot             : int  4 4 5 3 5 4 4 4 3 3 ...
##  $ Skill.Moves           : int  4 5 5 1 4 4 4 3 3 1 ...
##  $ Work.Rate             : chr  "Medium/ Medium" "High/ Low" "High/ Medium" "Medium/ Medium" ...
##  $ Body.Type             : chr  "Messi" "C. Ronaldo" "Neymar" "Lean" ...
##  $ Real.Face             : chr  "Yes" "Yes" "Yes" "Yes" ...
##  $ Position              : chr  "RF" "ST" "LW" "GK" ...
##  $ Jersey.Number         : int  10 7 10 1 7 10 10 9 15 1 ...
##  $ Joined                : chr  "Jul 1, 2004" "Jul 10, 2018" "Aug 3, 2017" "Jul 1, 2011" ...
##  $ Loaned.From           : chr  "" "" "" "" ...
##  $ Contract.Valid.Until  : chr  "2021" "2022" "2022" "2020" ...
##  $ Height                : chr  "5'7" "6'2" "5'9" "6'4" ...
##  $ Weight                : chr  "159lbs" "183lbs" "150lbs" "168lbs" ...
##  $ LS                    : chr  "88+2" "91+3" "84+3" "" ...
##  $ ST                    : chr  "88+2" "91+3" "84+3" "" ...
##  $ RS                    : chr  "88+2" "91+3" "84+3" "" ...
##  $ LW                    : chr  "92+2" "89+3" "89+3" "" ...
##  $ LF                    : chr  "93+2" "90+3" "89+3" "" ...
##  $ CF                    : chr  "93+2" "90+3" "89+3" "" ...
##  $ RF                    : chr  "93+2" "90+3" "89+3" "" ...
##  $ RW                    : chr  "92+2" "89+3" "89+3" "" ...
##  $ LAM                   : chr  "93+2" "88+3" "89+3" "" ...
##  $ CAM                   : chr  "93+2" "88+3" "89+3" "" ...
##  $ RAM                   : chr  "93+2" "88+3" "89+3" "" ...
##  $ LM                    : chr  "91+2" "88+3" "88+3" "" ...
##  $ LCM                   : chr  "84+2" "81+3" "81+3" "" ...
##  $ CM                    : chr  "84+2" "81+3" "81+3" "" ...
##  $ RCM                   : chr  "84+2" "81+3" "81+3" "" ...
##  $ RM                    : chr  "91+2" "88+3" "88+3" "" ...
##  $ LWB                   : chr  "64+2" "65+3" "65+3" "" ...
##  $ LDM                   : chr  "61+2" "61+3" "60+3" "" ...
##  $ CDM                   : chr  "61+2" "61+3" "60+3" "" ...
```

```
##  $ RDM                      : chr  "61+2" "61+3" "60+3" "" ...
##  $ RWB                      : chr  "64+2" "65+3" "65+3" "" ...
##  $ LB                       : chr  "59+2" "61+3" "60+3" "" ...
##  $ LCB                      : chr  "47+2" "53+3" "47+3" "" ...
##  $ CB                       : chr  "47+2" "53+3" "47+3" "" ...
##  $ RCB                      : chr  "47+2" "53+3" "47+3" "" ...
##  $ RB                       : chr  "59+2" "61+3" "60+3" "" ...
##  $ Crossing                 : int  84 84 79 17 93 81 86 77 66 13 ...
##  $ Finishing                : int  95 94 87 13 82 84 72 93 60 11 ...
##  $ HeadingAccuracy          : int  70 89 62 21 55 61 55 77 91 15 ...
##  $ ShortPassing             : int  90 81 84 50 92 89 93 82 78 29 ...
##  $ Volleys                  : int  86 87 84 13 82 80 76 88 66 13 ...
##  $ Dribbling                : int  97 88 96 18 86 95 90 87 63 12 ...
##  $ Curve                    : int  93 81 88 21 85 83 85 86 74 13 ...
##  $ FKAccuracy               : int  94 76 87 19 83 79 78 84 72 14 ...
##  $ LongPassing              : int  87 77 78 51 91 83 88 64 77 26 ...
##  $ BallControl              : int  96 94 95 42 91 94 93 90 84 16 ...
##  $ Acceleration             : int  91 89 94 57 78 94 80 86 76 43 ...
##  $ SprintSpeed              : int  86 91 90 58 76 88 72 75 75 60 ...
##  $ Agility                  : int  91 87 96 60 79 95 93 82 78 67 ...
##  $ Reactions                : int  95 96 94 90 91 90 90 92 85 86 ...
##  $ Balance                  : int  95 70 84 43 77 94 94 83 66 49 ...
##  $ ShotPower                : int  85 95 80 31 91 82 79 86 79 22 ...
##  $ Jumping                  : int  68 95 61 67 63 56 68 69 93 76 ...
##  $ Stamina                  : int  72 88 81 43 90 83 89 90 84 41 ...
##  $ Strength                 : int  59 79 49 64 75 66 58 83 83 78 ...
##  $ LongShots                : int  94 93 82 12 91 80 82 85 59 12 ...
##  $ Aggression               : int  48 63 56 38 76 54 62 87 88 34 ...
##  $ Interceptions            : int  22 29 36 30 61 41 83 41 90 19 ...
##  $ Positioning              : int  94 95 89 12 87 87 79 92 60 11 ...
##  $ Vision                   : int  94 82 87 68 94 89 92 84 63 70 ...
##  $ Penalties                : int  75 85 81 40 79 86 82 85 75 11 ...
##  $ Composure                : int  96 95 94 68 88 91 84 85 82 70 ...
##  $ Marking                  : int  33 28 27 15 68 34 60 62 87 27 ...
##  $ StandingTackle           : int  28 31 24 21 58 27 76 45 92 12 ...
##  $ SlidingTackle            : int  26 23 33 13 51 22 73 38 91 18 ...
##  $ GKDiving                 : int  6 7 9 90 15 11 13 27 11 86 ...
##  $ GKHandling               : int  11 11 9 85 13 12 9 25 8 92 ...
##  $ GKKicking                : int  15 15 15 87 5 6 7 31 9 78 ...
##  $ GKPositioning            : int  14 14 15 88 10 8 14 33 7 88 ...
##  $ GKReflexes               : int  8 11 11 94 13 8 9 37 11 89 ...
##  $ Release.Clause           : chr  "€226.5M" "€127.1M" "€228.1M" "€138.6M" ...
## [1] "Summary of our data"

##        X               ID              Name                Age
##  Min.   :    0   Min.   :    16   Length:18207       Min.   :16.00
##  1st Qu.: 4552   1st Qu.:200316   Class :character   1st Qu.:21.00
##  Median : 9103   Median :221759   Mode  :character   Median :25.00
##  Mean   : 9103   Mean   :214298                      Mean   :25.12
##  3rd Qu.:13654   3rd Qu.:236530                      3rd Qu.:28.00
##  Max.   :18206   Max.   :246620                      Max.   :45.00
##
##      Photo            Nationality           Flag               Overall
##  Length:18207       Length:18207        Length:18207        Min.   :46.00
```

```
##    Class :character   Class :character   Class :character   1st Qu.:62.00
##    Mode  :character   Mode  :character   Mode  :character   Median :66.00
##                                                             Mean   :66.24
##                                                             3rd Qu.:71.00
##                                                             Max.   :94.00
##
##       Potential         Club              Club.Logo             Value
##    Min.   :48.00   Length:18207       Length:18207        Length:18207
##    1st Qu.:67.00   Class :character   Class :character    Class :character
##    Median :71.00   Mode  :character   Mode  :character    Mode  :character
##    Mean   :71.31
##    3rd Qu.:75.00
##    Max.   :95.00
##
##        Wage             Special       Preferred.Foot    International.Reputation
##    Length:18207     Min.   : 731    Length:18207        Min.   :1.000
##    Class :character 1st Qu.:1457    Class :character    1st Qu.:1.000
##    Mode  :character Median :1635    Mode  :character    Median :1.000
##                     Mean   :1598                        Mean   :1.113
##                     3rd Qu.:1787                        3rd Qu.:1.000
##                     Max.   :2346                        Max.   :5.000
##                                                         NA's   :48
##      Weak.Foot       Skill.Moves      Work.Rate         Body.Type
##    Min.   :1.000   Min.   :1.000   Length:18207       Length:18207
##    1st Qu.:3.000   1st Qu.:2.000   Class :character   Class :character
##    Median :3.000   Median :2.000   Mode  :character   Mode  :character
##    Mean   :2.947   Mean   :2.361
##    3rd Qu.:3.000   3rd Qu.:3.000
##    Max.   :5.000   Max.   :5.000
##    NA's   :48      NA's   :48
##     Real.Face          Position          Jersey.Number      Joined
##    Length:18207     Length:18207      Min.   : 1.00     Length:18207
##    Class :character Class :character  1st Qu.: 8.00     Class :character
##    Mode  :character Mode  :character  Median :17.00     Mode  :character
##                                       Mean   :19.55
##                                       3rd Qu.:26.00
##                                       Max.   :99.00
##                                       NA's   :60
##    Loaned.From      Contract.Valid.Until    Height             Weight
##    Length:18207     Length:18207         Length:18207       Length:18207
##    Class :character Class :character      Class :character   Class :character
##    Mode  :character Mode  :character      Mode  :character   Mode  :character
##
##
##
##
##         LS               ST                RS                LW
##    Length:18207     Length:18207      Length:18207       Length:18207
##    Class :character Class :character  Class :character   Class :character
##    Mode  :character Mode  :character  Mode  :character   Mode  :character
##
##
##
##
```

```
##       LF                  CF                  RF                  RW
## Length:18207        Length:18207        Length:18207        Length:18207
## Class :character    Class :character    Class :character    Class :character
## Mode  :character    Mode  :character    Mode  :character    Mode  :character
##
##
##
##
##       LAM                 CAM                 RAM                 LM
## Length:18207        Length:18207        Length:18207        Length:18207
## Class :character    Class :character    Class :character    Class :character
## Mode  :character    Mode  :character    Mode  :character    Mode  :character
##
##
##
##
##       LCM                 CM                  RCM                 RM
## Length:18207        Length:18207        Length:18207        Length:18207
## Class :character    Class :character    Class :character    Class :character
## Mode  :character    Mode  :character    Mode  :character    Mode  :character
##
##
##
##
##       LWB                 LDM                 CDM                 RDM
## Length:18207        Length:18207        Length:18207        Length:18207
## Class :character    Class :character    Class :character    Class :character
## Mode  :character    Mode  :character    Mode  :character    Mode  :character
##
##
##
##
##       RWB                 LB                  LCB                 CB
## Length:18207        Length:18207        Length:18207        Length:18207
## Class :character    Class :character    Class :character    Class :character
## Mode  :character    Mode  :character    Mode  :character    Mode  :character
##
##
##
##
##       RCB                 RB                  Crossing         Finishing
## Length:18207        Length:18207        Min.   : 5.00    Min.   : 2.00
## Class :character    Class :character    1st Qu.:38.00    1st Qu.:30.00
## Mode  :character    Mode  :character    Median :54.00    Median :49.00
##                                         Mean   :49.73    Mean   :45.55
##                                         3rd Qu.:64.00    3rd Qu.:62.00
##                                         Max.   :93.00    Max.   :95.00
##                                         NA's   :48       NA's   :48
##  HeadingAccuracy  ShortPassing      Volleys         Dribbling
##  Min.   : 4.0    Min.   : 7.00   Min.   : 4.00   Min.   : 4.00
##  1st Qu.:44.0    1st Qu.:54.00   1st Qu.:30.00   1st Qu.:49.00
##  Median :56.0    Median :62.00   Median :44.00   Median :61.00
##  Mean   :52.3    Mean   :58.69   Mean   :42.91   Mean   :55.37
##  3rd Qu.:64.0    3rd Qu.:68.00   3rd Qu.:57.00   3rd Qu.:68.00
```

```
##    Max.   :94.0    Max.   :93.00    Max.   :90.00    Max.    :97.00
##    NA's   :48       NA's   :48       NA's   :48       NA's    :48
##       Curve            FKAccuracy       LongPassing       BallControl
##    Min.   : 6.00    Min.   : 3.00    Min.   : 9.00    Min.   : 5.00
##    1st Qu.:34.00    1st Qu.:31.00    1st Qu.:43.00    1st Qu.:54.00
##    Median :48.00    Median :41.00    Median :56.00    Median :63.00
##    Mean   :47.17    Mean   :42.86    Mean   :52.71    Mean   :58.37
##    3rd Qu.:62.00    3rd Qu.:57.00    3rd Qu.:64.00    3rd Qu.:69.00
##    Max.   :94.00    Max.   :94.00    Max.   :93.00    Max.   :96.00
##    NA's   :48       NA's   :48       NA's   :48       NA's   :48
##     Acceleration     SprintSpeed       Agility          Reactions        Balance
##    Min.   :12.00    Min.   :12.00    Min.   :14.0     Min.   :21.00    Min.   :16.00
##    1st Qu.:57.00    1st Qu.:57.00    1st Qu.:55.0     1st Qu.:56.00    1st Qu.:56.00
##    Median :67.00    Median :67.00    Median :66.0     Median :62.00    Median :66.00
##    Mean   :64.61    Mean   :64.73    Mean   :63.5     Mean   :61.84    Mean   :63.97
##    3rd Qu.:75.00    3rd Qu.:75.00    3rd Qu.:74.0     3rd Qu.:68.00    3rd Qu.:74.00
##    Max.   :97.00    Max.   :96.00    Max.   :96.0     Max.   :96.00    Max.   :96.00
##    NA's   :48       NA's   :48       NA's   :48       NA's   :48       NA's   :48
##      ShotPower         Jumping          Stamina          Strength
##    Min.   : 2.00    Min.   :15.00    Min.   :12.00    Min.   :17.00
##    1st Qu.:45.00    1st Qu.:58.00    1st Qu.:56.00    1st Qu.:58.00
##    Median :59.00    Median :66.00    Median :66.00    Median :67.00
##    Mean   :55.46    Mean   :65.09    Mean   :63.22    Mean   :65.31
##    3rd Qu.:68.00    3rd Qu.:73.00    3rd Qu.:74.00    3rd Qu.:74.00
##    Max.   :95.00    Max.   :95.00    Max.   :96.00    Max.   :97.00
##    NA's   :48       NA's   :48       NA's   :48       NA's   :48
##      LongShots        Aggression      Interceptions     Positioning       Vision
##    Min.   : 3.00    Min.   :11.00    Min.   : 3.0     Min.   : 2.00    Min.   :10.0
##    1st Qu.:33.00    1st Qu.:44.00    1st Qu.:26.0     1st Qu.:38.00    1st Qu.:44.0
##    Median :51.00    Median :59.00    Median :52.0     Median :55.00    Median :55.0
##    Mean   :47.11    Mean   :55.87    Mean   :46.7     Mean   :49.96    Mean   :53.4
##    3rd Qu.:62.00    3rd Qu.:69.00    3rd Qu.:64.0     3rd Qu.:64.00    3rd Qu.:64.0
##    Max.   :94.00    Max.   :95.00    Max.   :92.0     Max.   :95.00    Max.   :94.0
##    NA's   :48       NA's   :48       NA's   :48       NA's   :48       NA's   :48
##      Penalties        Composure         Marking       StandingTackle SlidingTackle
##    Min.   : 5.00    Min.   : 3.00    Min.   : 3.00    Min.   : 2.0     Min.   : 3.00
##    1st Qu.:39.00    1st Qu.:51.00    1st Qu.:30.00    1st Qu.:27.0     1st Qu.:24.00
##    Median :49.00    Median :60.00    Median :53.00    Median :55.0     Median :52.00
##    Mean   :48.55    Mean   :58.65    Mean   :47.28    Mean   :47.7     Mean   :45.66
##    3rd Qu.:60.00    3rd Qu.:67.00    3rd Qu.:64.00    3rd Qu.:66.0     3rd Qu.:64.00
##    Max.   :92.00    Max.   :96.00    Max.   :94.00    Max.   :93.0     Max.   :91.00
##    NA's   :48       NA's   :48       NA's   :48       NA's   :48       NA's   :48
##      GKDiving         GKHandling        GKKicking       GKPositioning
##    Min.   : 1.00    Min.   : 1.00    Min.   : 1.00    Min.   : 1.00
##    1st Qu.: 8.00    1st Qu.: 8.00    1st Qu.: 8.00    1st Qu.: 8.00
##    Median :11.00    Median :11.00    Median :11.00    Median :11.00
##    Mean   :16.62    Mean   :16.39    Mean   :16.23    Mean   :16.39
##    3rd Qu.:14.00    3rd Qu.:14.00    3rd Qu.:14.00    3rd Qu.:14.00
##    Max.   :90.00    Max.   :92.00    Max.   :91.00    Max.   :90.00
##    NA's   :48       NA's   :48       NA's   :48       NA's   :48
##      GKReflexes       Release.Clause
##    Min.   : 1.00    Length:18207
##    1st Qu.: 8.00    Class :character
##    Median :11.00    Mode  :character
```

```
##  Mean   :16.71
##  3rd Qu.:14.00
##  Max.   :94.00
##  NA's   :48
```

Some modification are needed :

1. **Change the Player Valuation**: The player valuations feature is a character string, with a designation at the end specifying whether the value is thousands or millions. It's essential the type is as numrical value.

2. **Change Player Wage**: See Point 1 above. Same transformation has occurred for player Wage

3. **Regroup Player Positions into 5 big groups**: There are 28 different positions in FIFA2019. To make analysis less granular,it is a common thing to group the position into this 5 big groups: GK, DEF, MID, FWD and Unknown.

4. **Regroup Player Positions into 6 other groups**: See Point 3. There is an another way to group the positions: GK, DEFWB (left/right wing/full back), DEFC (center defenders), DEFMID (defensive Mid), WingFor (left/right side winger/mid), MIDCA (centr/attack mid), SC (striker/ center forward) and Unknown. *Warning*: This groups are not a subdivision of the groups of point 3. I choose to regroup left/right midfielders with left/right forwards. Also, I choose to separate the defensive midfielders and the center midfielder, because I think this positions are different and require differents quality.

5. **Group Player age**: group players into age buckets, in 5 year increments other than for players 20 years and younger, and players 35 years and over.

3 General Overview Analysis

3.1 Correlation

The first thing I would like to check is the correlation between the variables and thus analyze that should give good insights. To do this, I will evaluate the Spearman Correlation and the Pearson Correlation. The Pearson model evaluates the linearity between two variables whereas the Spaerman evaluates if two variables have a tendency to change together.

```
## [1] "Pearson Matrix"
```

[1] "Spearman Matrix"

The correlation matrix (Spearman coefficients) for player attributes:

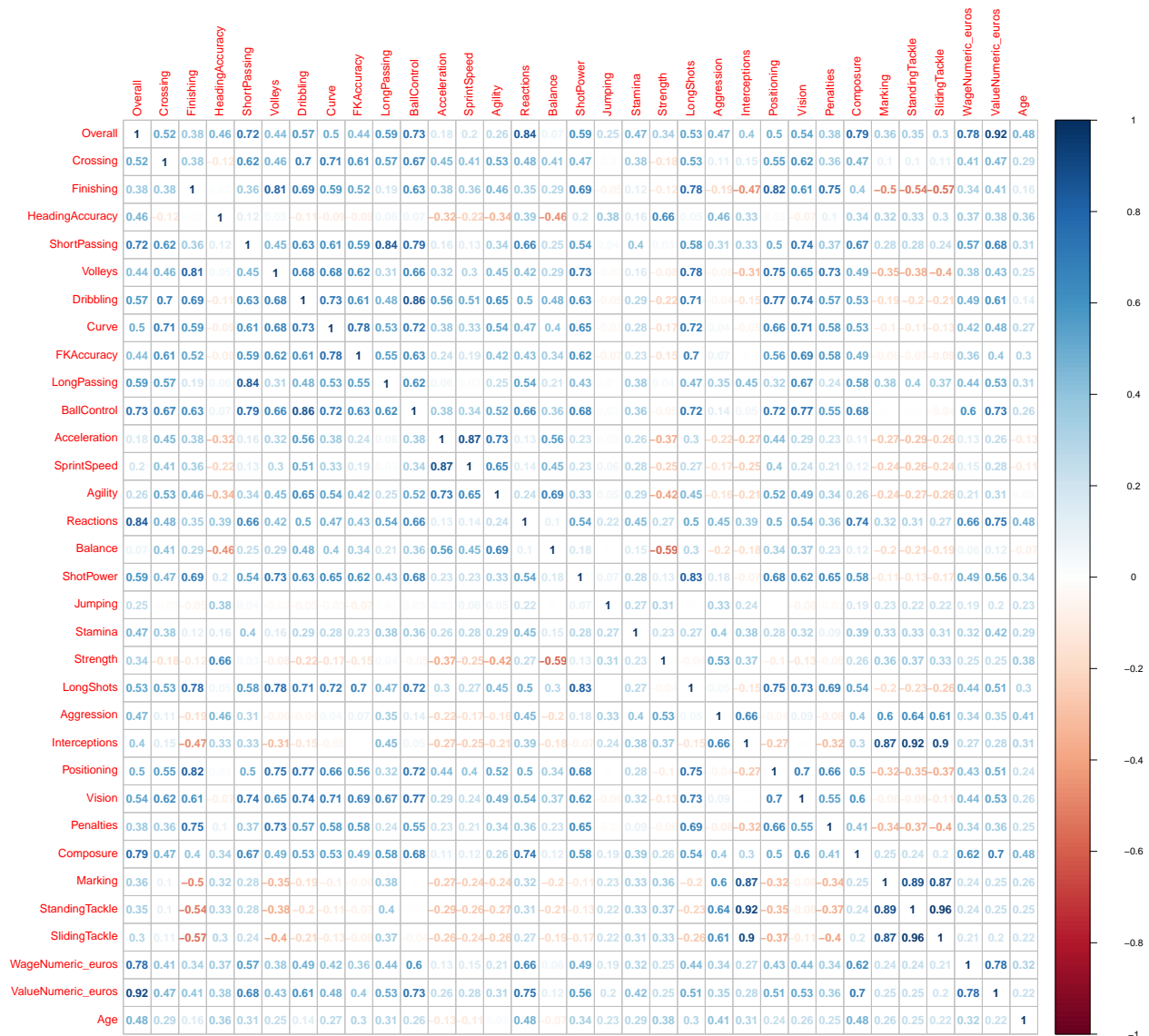|	Overall	Crossing	Finishing	HeadingAccuracy	ShortPassing	Volleys	Dribbling	Curve	FKAccuracy	LongPassing	BallControl	Acceleration	SprintSpeed	Agility	Reactions	Balance	ShotPower	Jumping	Stamina	Strength	LongShots	Aggression	Interceptions	Positioning	Vision	Penalties	Composure	Marking	StandingTackle	SlidingTackle	WageNumeric_euros	ValueNumeric_euros	Age
Overall	1	0.52	0.38	0.46	0.72	0.44	0.57	0.5	0.44	0.59	0.73	0.18	0.2	0.26	0.84		0.59	0.25	0.47	0.34	0.53	0.47	0.4	0.5	0.54	0.38	0.79	0.36	0.35	0.3	0.78	0.92	0.48
Crossing	0.52	1	0.38	-0.12	0.62	0.46	0.7	0.71	0.61	0.57	0.67	0.45	0.41	0.53	0.48	0.41	0.47		0.38	-0.18	0.53	0.11	0.15	0.55	0.62	0.36	0.47	0.1	0.1	0.11	0.41	0.47	0.29
Finishing	0.38	0.38	1		0.36	0.81	0.69	0.59	0.52	0.19	0.63	0.38	0.36	0.46	0.35	0.29	0.69		0.12	-0.12	0.78	-0.19	-0.47	0.82	0.61	0.75	0.4	-0.5	-0.54	-0.57	0.34	0.41	0.16
HeadingAccuracy	0.46	-0.12		1	0.12		-0.1	-0.09			0.07	-0.32	-0.22	-0.34	0.39	-0.46	0.2	0.38	0.16	0.66		0.46	0.33			0.1	0.34	0.32	0.33	0.3	0.37	0.38	0.36
ShortPassing	0.72	0.62	0.36	0.12	1	0.45	0.63	0.61	0.59	0.84	0.79	0.16	0.13	0.34	0.66	0.25	0.54		0.4		0.58	0.31	0.33	0.5	0.74	0.37	0.67	0.28	0.28	0.24	0.57	0.68	0.31
Volleys	0.44	0.46	0.81		0.45	1	0.68	0.68	0.62	0.31	0.66	0.32	0.3	0.45	0.42	0.29	0.73		0.16		0.78		-0.31	0.75	0.65	0.73	0.49	-0.35	-0.38	-0.4	0.38	0.43	0.25
Dribbling	0.57	0.7	0.69		0.63	0.68	1	0.73	0.61	0.48	0.86	0.56	0.51	0.65	0.5	0.48	0.63		0.29	-0.22	0.71		-0.15	0.77	0.74	0.57	0.53	-0.19	-0.2	-0.21	0.49	0.61	0.14
Curve	0.5	0.71	0.59		0.61	0.68	0.73	1	0.78	0.53	0.72	0.38	0.33	0.54	0.47	0.4	0.65		0.28	-0.17	0.72			0.66	0.71	0.58	0.53	-0.1	-0.11	-0.13	0.42	0.48	0.27
FKAccuracy	0.44	0.61	0.52		0.59	0.62	0.61	0.78	1	0.55	0.63	0.24	0.19	0.42	0.43	0.34	0.62		0.23	-0.15	0.7			0.56	0.69	0.58	0.49			-0.09	0.36	0.4	0.3
LongPassing	0.59	0.57	0.19		0.84	0.31	0.48	0.53	0.55	1	0.62			0.25	0.54	0.21	0.43		0.38		0.47	0.35	0.45	0.32	0.67	0.24	0.58	0.38	0.4	0.37	0.44	0.53	0.31
BallControl	0.73	0.67	0.63		0.79	0.66	0.86	0.72	0.63	0.62	1	0.38	0.34	0.52	0.66	0.36	0.68		0.36		0.72	0.14		0.72	0.77	0.55	0.68				0.6	0.73	0.26
Acceleration	0.18	0.45	0.38	-0.32	0.16	0.32	0.56	0.38	0.24		0.38	1	0.87	0.73	0.13	0.56	0.23		0.26	-0.37	0.3	-0.22	-0.27	0.44	0.29	0.23	0.11	-0.27	-0.29	-0.26	0.13	0.26	-0.13
SprintSpeed	0.2	0.41	0.36	-0.22	0.13	0.3	0.51	0.33	0.19		0.34	0.87	1	0.65	0.14	0.45	0.23		0.28	-0.25	0.27	-0.17	-0.25	0.4	0.24	0.21	0.12	-0.24	-0.26	-0.24	0.15	0.28	-0.11
Agility	0.26	0.53	0.46	-0.34	0.34	0.45	0.65	0.54	0.42	0.25	0.52	0.73	0.65	1	0.24	0.69	0.33		0.29	-0.42	0.45	-0.16	-0.21	0.52	0.49	0.34	0.26	-0.24	-0.27	-0.26	0.21	0.31	
Reactions	0.84	0.48	0.35	0.39	0.66	0.42	0.5	0.47	0.43	0.54	0.66	0.13	0.14	0.24	1	0.1	0.54	0.22	0.45	0.27	0.5	0.45	0.39	0.5	0.54	0.36	0.74	0.32	0.31	0.27	0.66	0.75	0.48
Balance	0.07	0.41	0.29	-0.46	0.25	0.29	0.48	0.4	0.34	0.21	0.36	0.56	0.45	0.69	0.1	1	0.18		0.15	-0.59	0.3	-0.2	-0.18	0.34	0.37	0.23		-0.2	-0.21	-0.19	0.06	0.12	
ShotPower	0.59	0.47	0.69	0.2	0.54	0.73	0.63	0.65	0.62	0.43	0.68	0.23	0.23	0.33	0.54	0.18	1		0.28	0.13	0.83	0.18		0.68	0.62	0.65	0.58	-0.11	-0.13	-0.17	0.49	0.56	0.34
Jumping	0.25			0.38											0.22		0.07	1	0.27	0.31		0.33	0.24				0.19	0.23	0.22	0.22	0.19	0.2	0.23
Stamina	0.47	0.38	0.12	0.16	0.4	0.16	0.29	0.28	0.23	0.38	0.36	0.26	0.28	0.29	0.45	0.15	0.28	0.27	1	0.23	0.27	0.4	0.38	0.28	0.32		0.39	0.33	0.33	0.31	0.32	0.42	0.29
Strength	0.34	-0.18	-0.12	0.66			-0.22	-0.17	-0.15			-0.37	-0.25	-0.42	0.27	-0.59	0.13	0.31	0.23	1		0.53	0.37		-0.13		0.26	0.36	0.37	0.33	0.25	0.25	0.38
LongShots	0.53	0.53	0.78		0.58	0.78	0.71	0.72	0.7	0.47	0.72	0.3	0.27	0.45	0.5	0.3	0.83		0.27		1		-0.15	0.75	0.73	0.69	0.54	-0.2	-0.23	-0.26	0.44	0.51	0.3
Aggression	0.47	0.11	-0.19	0.46	0.31			0.34	0.07	0.35	0.14	-0.22	-0.17	-0.16	0.45	-0.2	0.18	0.33	0.4	0.53		1	0.66		0.09		0.4	0.6	0.64	0.61	0.34	0.35	0.41
Interceptions	0.4	0.15	-0.47	0.33	0.33	-0.31	-0.15			0.45		-0.27	-0.25	-0.21	0.39	-0.18			0.24	0.38	0.37	-0.15	0.66	1	-0.27		-0.32	0.3	0.87	0.92	0.9	0.27	0.28	0.31
Positioning	0.5	0.55	0.82		0.5	0.75	0.77	0.66	0.56	0.32	0.72	0.44	0.4	0.52	0.5	0.34	0.68		0.28	-0.1	0.75		-0.27	1	0.7	0.66	0.5	-0.32	-0.35	-0.37	0.43	0.51	0.24
Vision	0.54	0.62	0.61		0.74	0.65	0.74	0.71	0.69	0.67	0.77	0.29	0.24	0.49	0.54	0.37	0.62		0.32	-0.13	0.73	-0.09		0.7	1	0.55	0.6		-0.18	-0.19	0.44	0.53	0.26
Penalties	0.38	0.36	0.75	0.1	0.37	0.73	0.57	0.58	0.58	0.24	0.55	0.23	0.21	0.34	0.36	0.23	0.65		0.09		0.69		-0.32	0.66	0.55	1	0.41	-0.34	-0.37	-0.4	0.34	0.36	0.25
Composure	0.79	0.47	0.4	0.34	0.67	0.49	0.53	0.53	0.49	0.58	0.68	0.11	0.12	0.26	0.74	0.12	0.58	0.19	0.39	0.26	0.54	0.4	0.3	0.5	0.6	0.41	1	0.25	0.24	0.2	0.62	0.7	0.48
Marking	0.36	0.1	-0.5	0.32	0.28	-0.35	-0.19	-0.1			0.38		-0.27	-0.24	-0.24	0.32	-0.2		0.23	0.33	0.36	-0.2	0.6	0.87	-0.32		-0.34	0.25	1	0.89	0.87	0.24	0.25	0.26
StandingTackle	0.35	0.1	-0.54	0.33	0.28	-0.38	-0.2	-0.11			0.4		-0.29	-0.26	-0.27	0.31	-0.21	-0.13	0.22	0.33	0.37	-0.23	0.64	0.92	-0.35		-0.37	0.24	0.89	1	0.96	0.24	0.25	0.25
SlidingTackle	0.3	0.11	-0.57	0.3	0.24	-0.4	-0.21	-0.13	-0.09		0.37		-0.26	-0.24	-0.26	0.27	-0.19	-0.17	0.22	0.31	0.33	-0.26	0.61	0.9	-0.37	-0.11	-0.4	0.87	0.96	1	0.21	0.2	0.22
WageNumeric_euros	0.78	0.41	0.34	0.37	0.57	0.38	0.49	0.42	0.36	0.44	0.6	0.13	0.15	0.21	0.66		0.49	0.19	0.32	0.25	0.44	0.34	0.27	0.43	0.44	0.34	0.62	0.24	0.24	0.21	1	0.78	0.32
ValueNumeric_euros	0.92	0.47	0.41	0.38	0.68	0.43	0.61	0.48	0.4	0.53	0.73	0.26	0.28	0.31	0.75	0.12	0.56	0.2	0.42	0.25	0.51	0.35	0.28	0.51	0.53	0.36	0.7	0.25	0.25	0.2	0.78	1	0.22
Age	0.48	0.29	0.16	0.36	0.31	0.25	0.14	0.27	0.3	0.31	0.26	-0.13	-0.11		0.48		0.34	0.23	0.29	0.38	0.3	0.41	0.31	0.24	0.26	0.25	0.48	0.26	0.25	0.22	0.32	0.22	1

[1] "Top 10"

The Spearman coefficients show a high correlation between Overall and Value, Reactions, Composure and Wage. But Value and Wage don't appear in the top 3 of the Pearson coefficients because both they probably have anomaly (like extrems values) in their values.

3.2 Overall Ratings

What define a player is his overall rating value. As we can suppose from looking the correlation table that this statistique define the value and the wage for a player. After plotting the value, player ratings appear normally distributed in FIFA19, with a mean of 66.24 and standard deviation of 6.91.

Distribution of player ratings
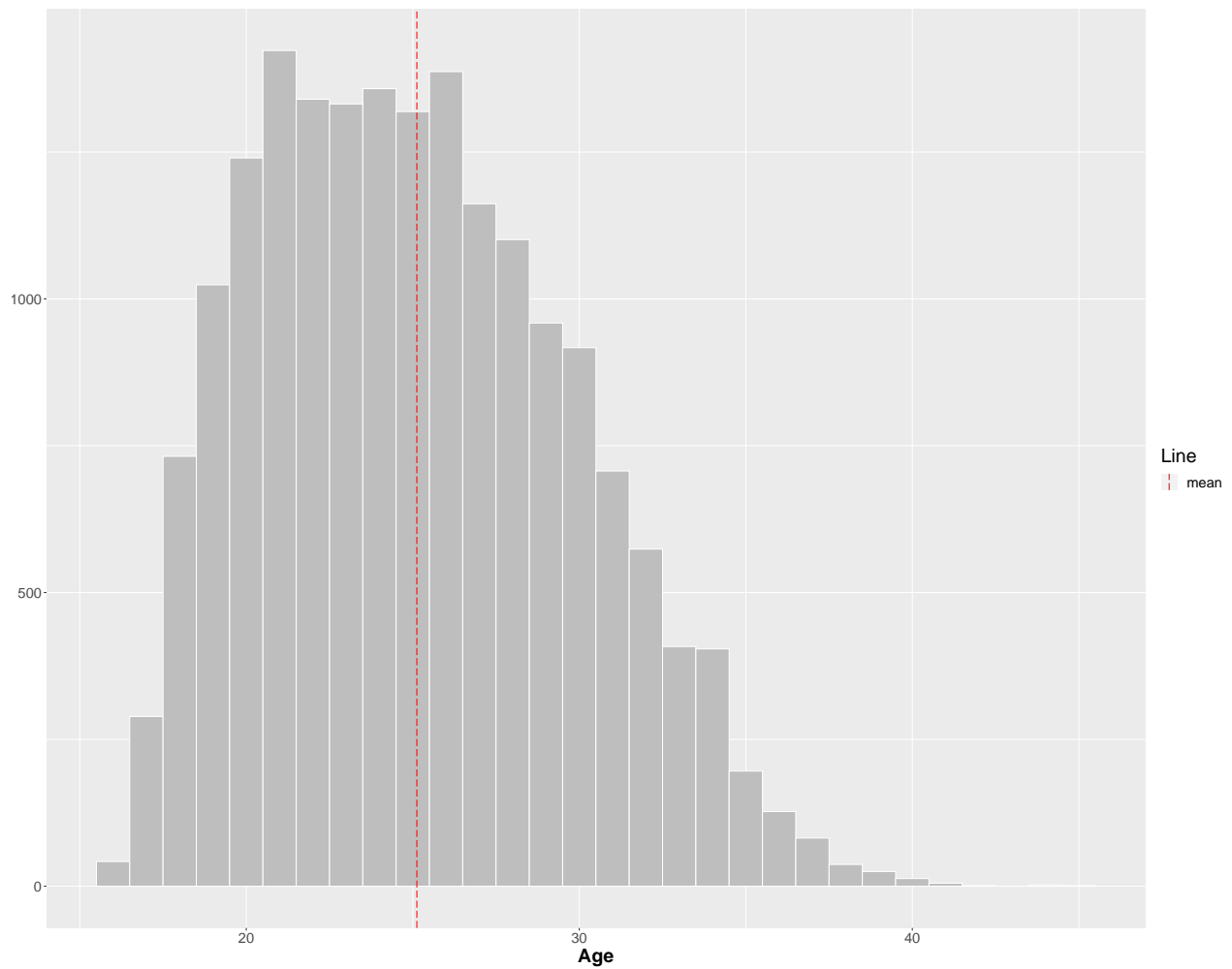*Mean: 66.24 / Standard deviation: 6.91*
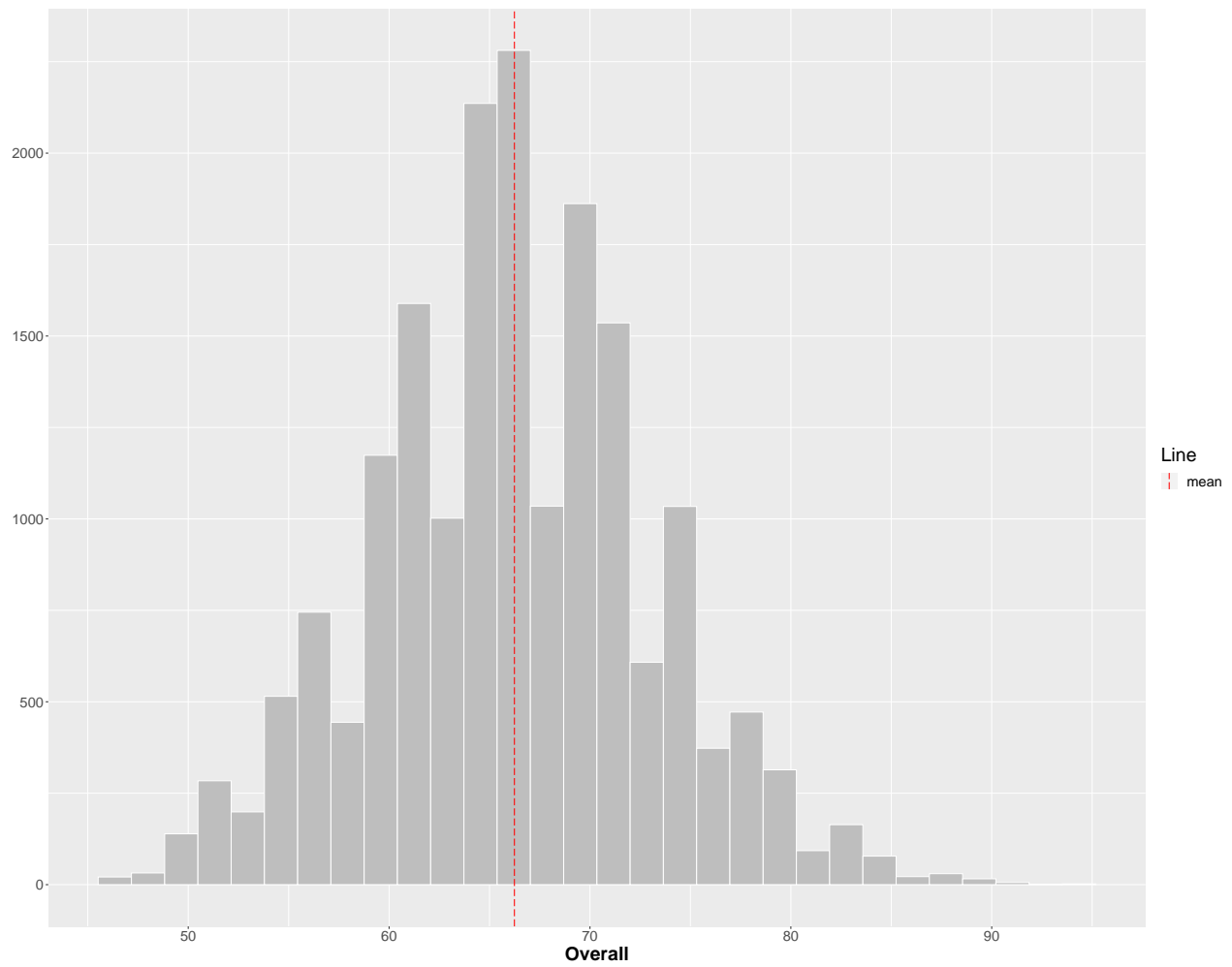
3.3 Age Distribution

An another good statistic to analyse is the age. The age average is around 25 yo, the younger player has 16 yo and the older 45 yo. The distribution follows a normal distibution.

Distribution of player ages
*Mean: 25.12 / Standard deviation: 4.67 / min: 16 / max: 45*

3.4 Ages vs Overall Rating

Now we analyse each statistic alone, it will be good to see how the overall rating and the age are related. It's a common thinking which the older the player are getting, the more they gains experience as a result an increasing of the oevrall score.

The graph shows that the overall ratings increase until 30 yo and begin declining after.

Age vs Overall Rating Curve
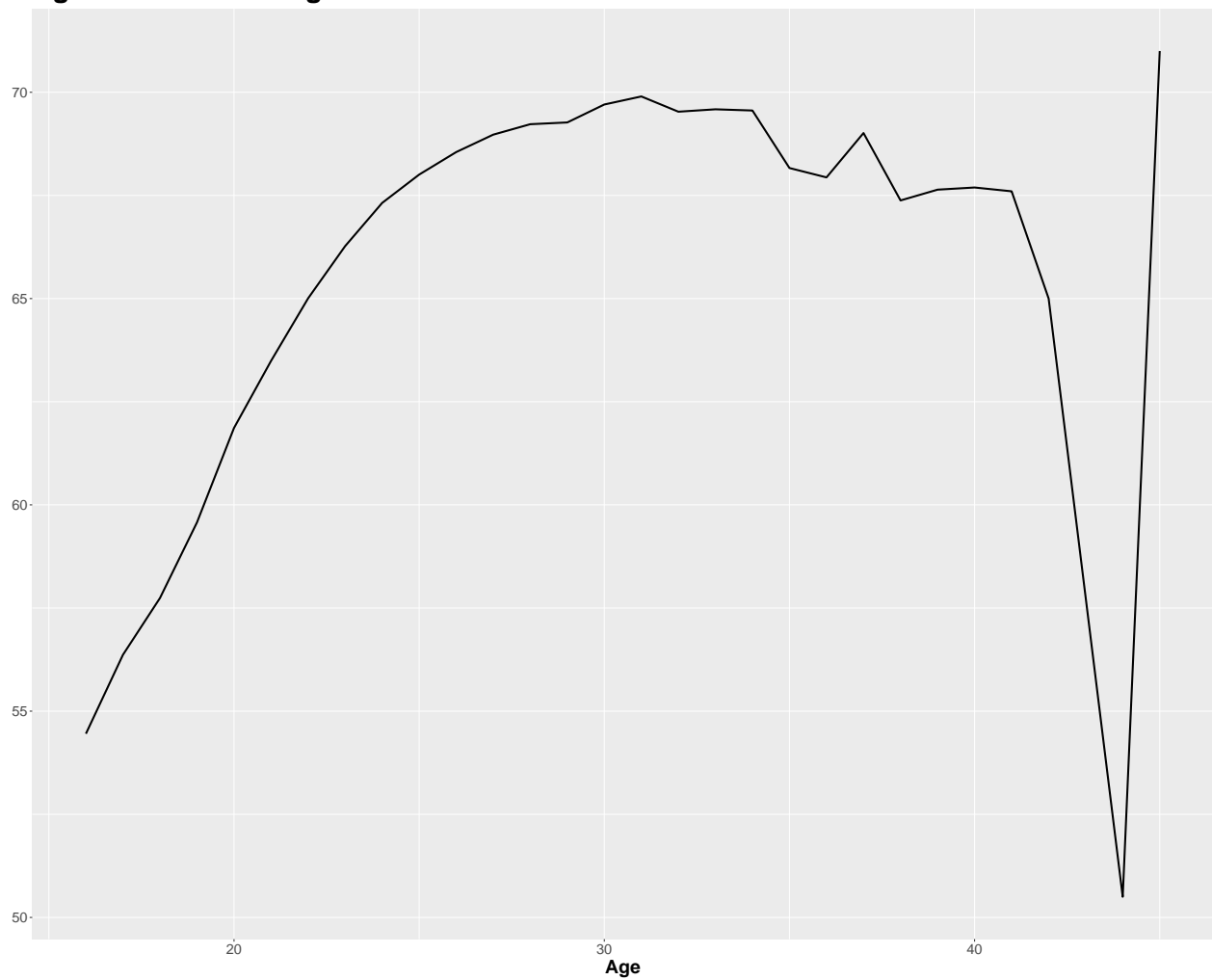
I want to investigate the players above 40 years with a rating above 70.

So, only 9 players are this criteria, and it appaers that they are almost all goalkeeper. In fact this position allows playing more longer.

Thus to have a more detail overview, an analysis per position could have a good income.

4 Potential Analysis

Before analysing by position, another statistic is related to the overall score, the potential. It define if a player (young or older) can continue to grow or not. So it's a prediction of a future overall score.

4.1 Young Player Analysis

So when you are a club, prepare the future is essential, so an analysis of the potiential of the young player will be a good outcome.

Young player Potential Overview

Player	Overall	RoomToGrow (total)
A. Dabo — LM, ...650K, 17	62	86
R. Griffiths — ST, ...575K, 18	61	84
J. von Moos — ST, ...280K, 17	58	84
B. Gilmour — CM, ...550K, 17	61	84
G. Azzinnari — ST, ...350K, 17	59	83
Y. Lenze — GK, ...190K, 17	57	82
J. Romero — ST, ...280K, 17	58	81
A. Appiah — RW, ...280K, 17	58	81
Y. Toure — ST, ...230K, 17	57	80
I. Sauter — CB, ...180K, 17	56	80
B. Mumba — CM, ...190K, 16	55	80
B. Kavanagh — LM, ...230K, 17	57	80
A. Wilson — LM, ...230K, 18	57	80
S. Giménez — ST, ...200K, 17	56	79
O. Aga — ST, ...200K, 17	55	78
K. Askildsen — CM, ...110K, 17	52	77
D. Campbell — CM, ...60K, 17	50	76
K. Lara — RB, ...60K, 16	50	74
E. Destanoglu — GK, ...50K, 17	50	74
N. Ayéva — ST, ...70K, 16	48	72

Rate

Criteria: RoomToGrow, Overall

All these players are talented, but *J. von Moos*, a striker (17 yo and costs €280K), *Y. Lenze*, a goalkeeper (17 yo and €190K) are a really good potential and a less investment related to the others. Also, the left-middle *A.Dabo* (17 yo and €650K) could be a great opportunity, he is a little bit expensive related to the other, but with this potential, he coulb become a extraordinary player.

4.2   When Potential end?

It is really interesting to know if the players reach their potential at one point, or they continues to grow their skills throughout their careers. My first impression is the player reach their maximum level near 30 yo and then start to decline, especially physically. And after analyzed the data, the potential and the overall rating converge averaging near 30 yo.

Overall vs Potential
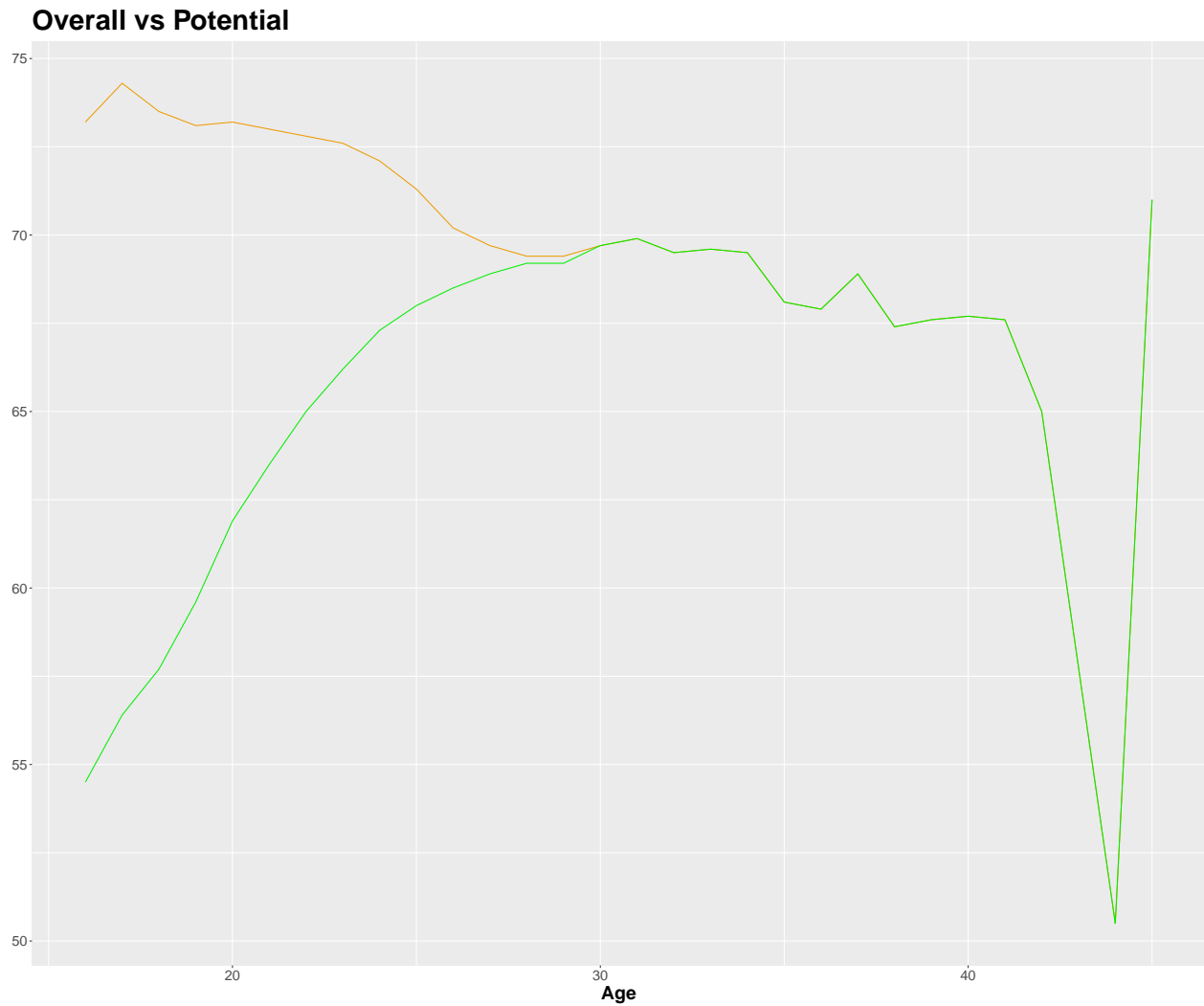
5   Position Analysis

But all position should have their own characteristic, their own specifity.

5.1   Position Bar Plot

It coulb be interesting to see the repartition of player's positions and see which position contains the most players. When the transfer period comes, it could be valuable to know where put the most effort to fill a need.

Bar Plot of player Position
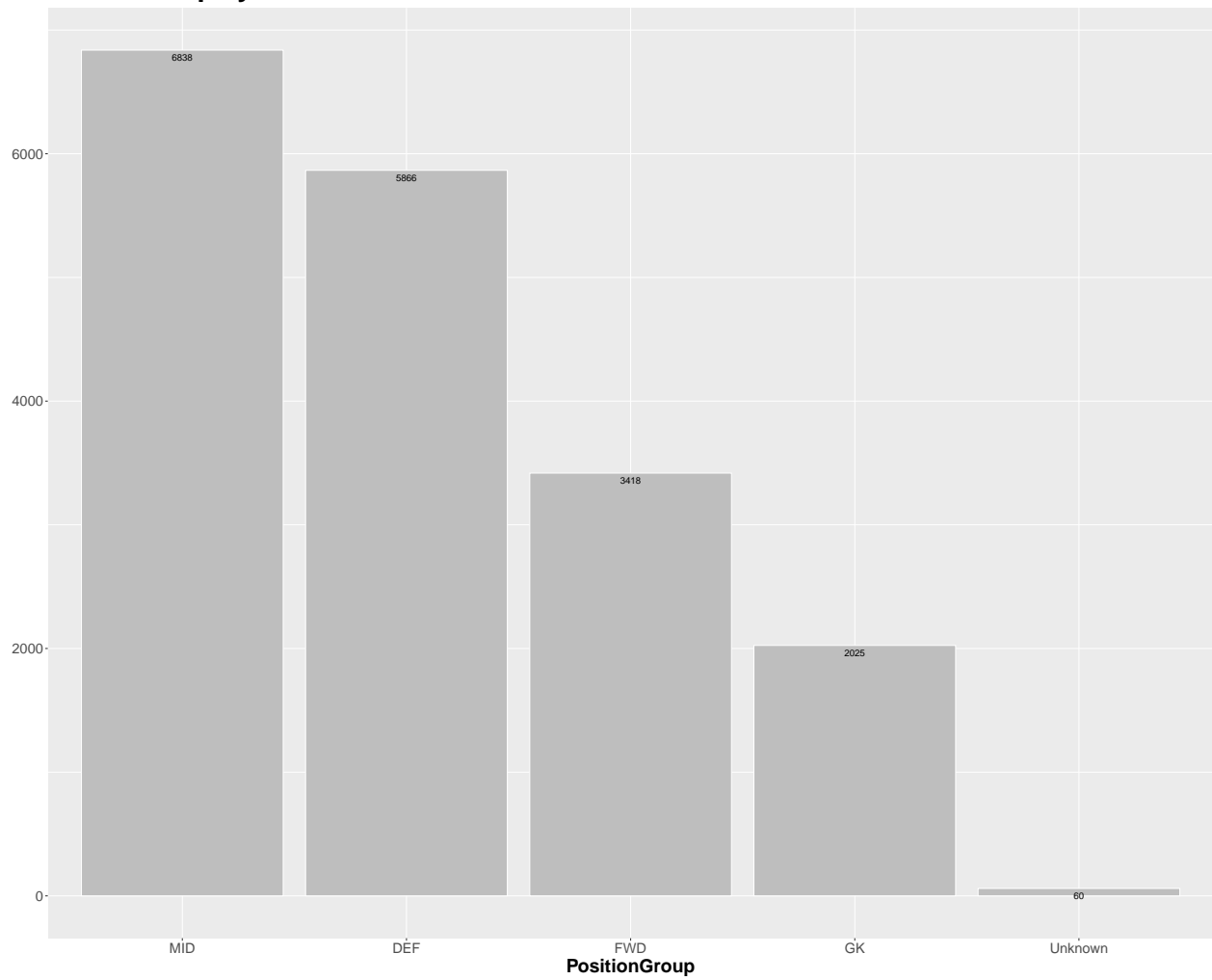
It seems there are less players who play forward. To validate that, poltting a new graph with the detailed positions will give us a good income

Bar Plot of player Position
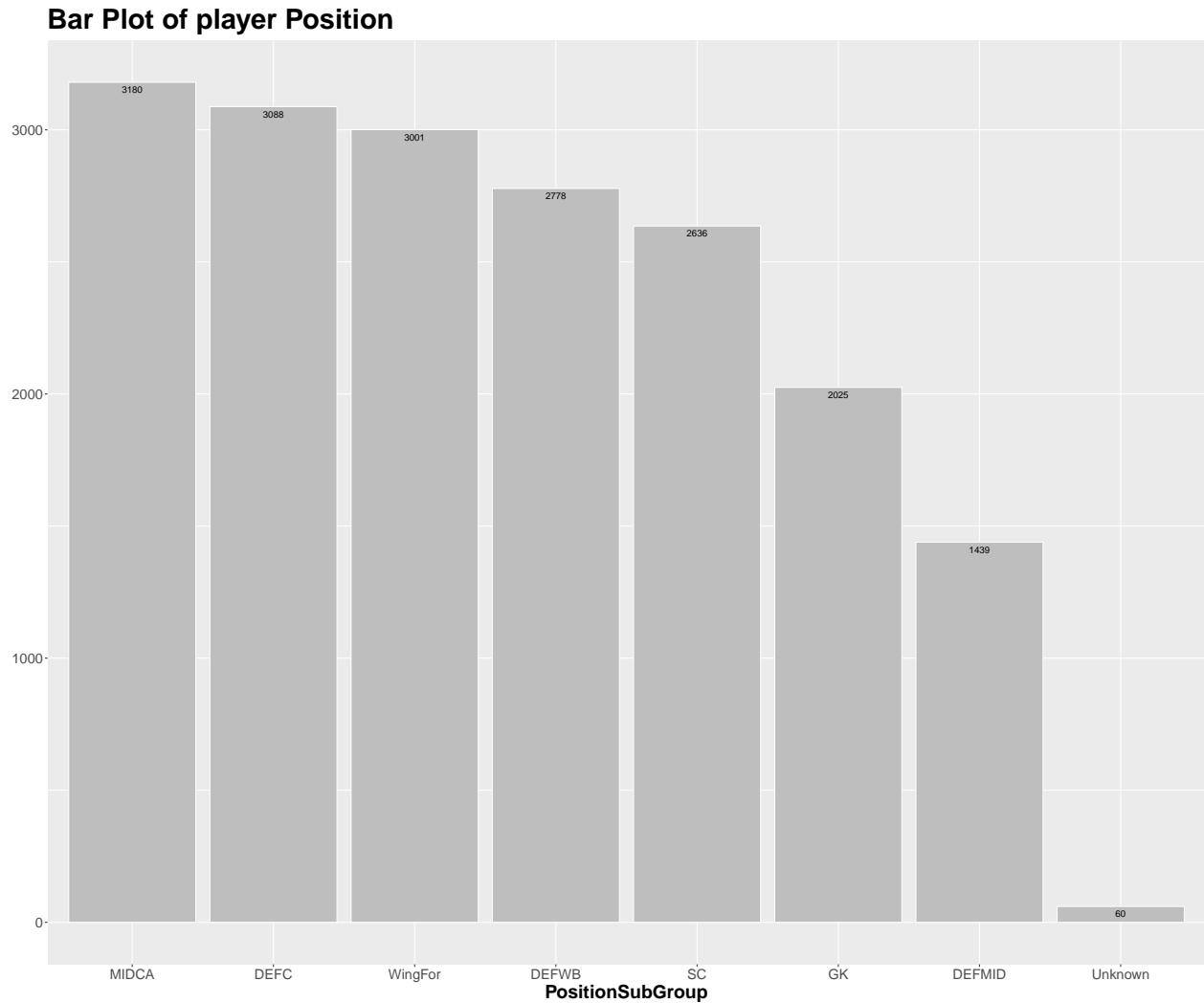
With the "subgroup", the defensive midfield players are less compare to the other groups because it's a specific position.

5.2 Overall Distribution for each group of position

After plotting the histogram for the overall score for each position, their distributions seem to be normal with a mean around 66 for each group. So all groups appear to be similar except one, the defensive midfield , which has the fewer player but with the best average. Again, this specific players should be difficult finding this players, likely increasing the price. But they are most valuable player for each money spent.

Overall Rating by Position
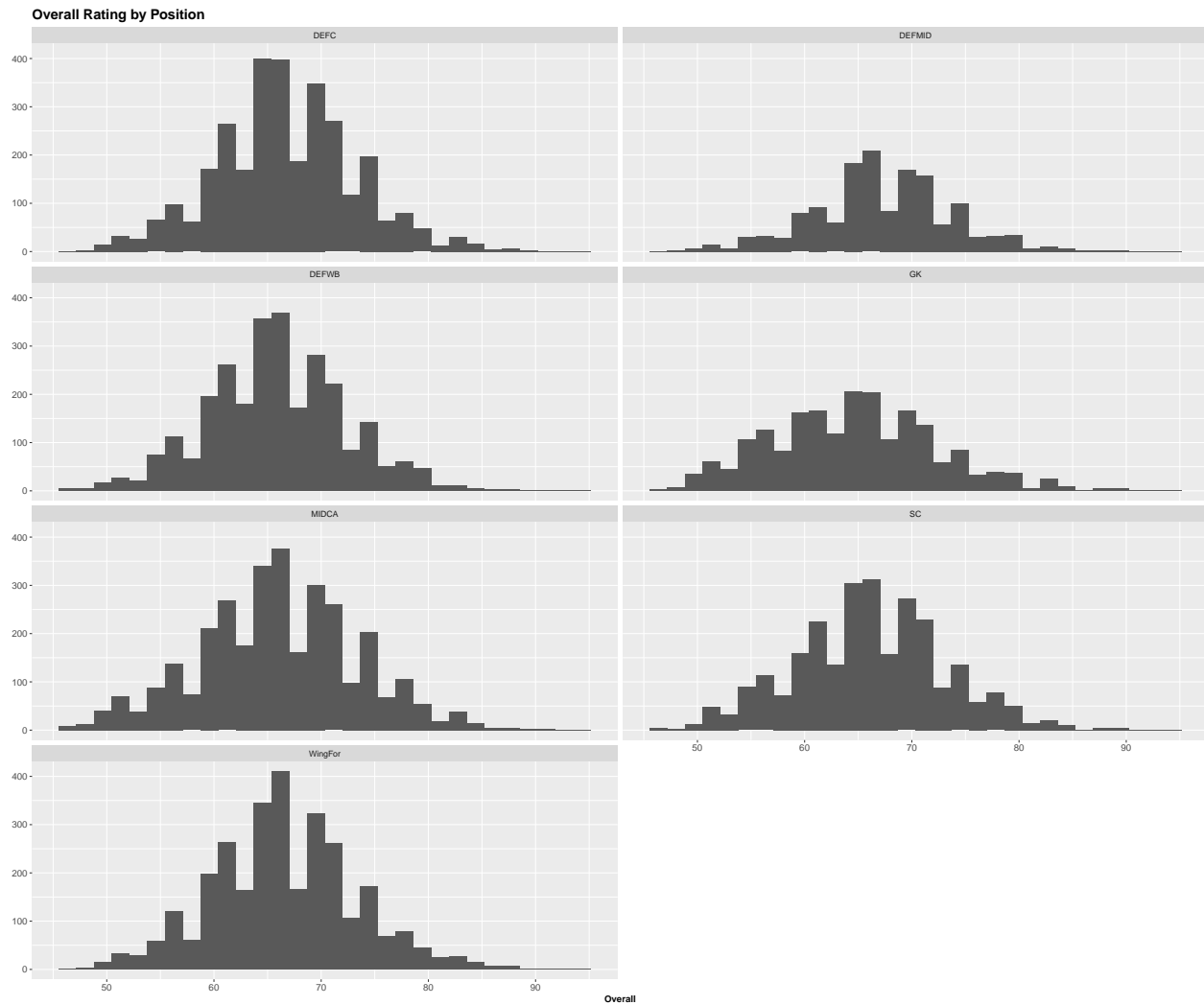
5.3 Age distribution for each group of position

I did the same analysis but for the age value and the conclusions are quite similar, all groups appear to be similar except for one, the goalkeepers where the average is a little bit higher. In fact, the goalkeepers have a longer career than the others. The average is around 25 years old. (Maybe I will doing later a statistical test to verify the hypothesis that all groups are similar)

Age by Position
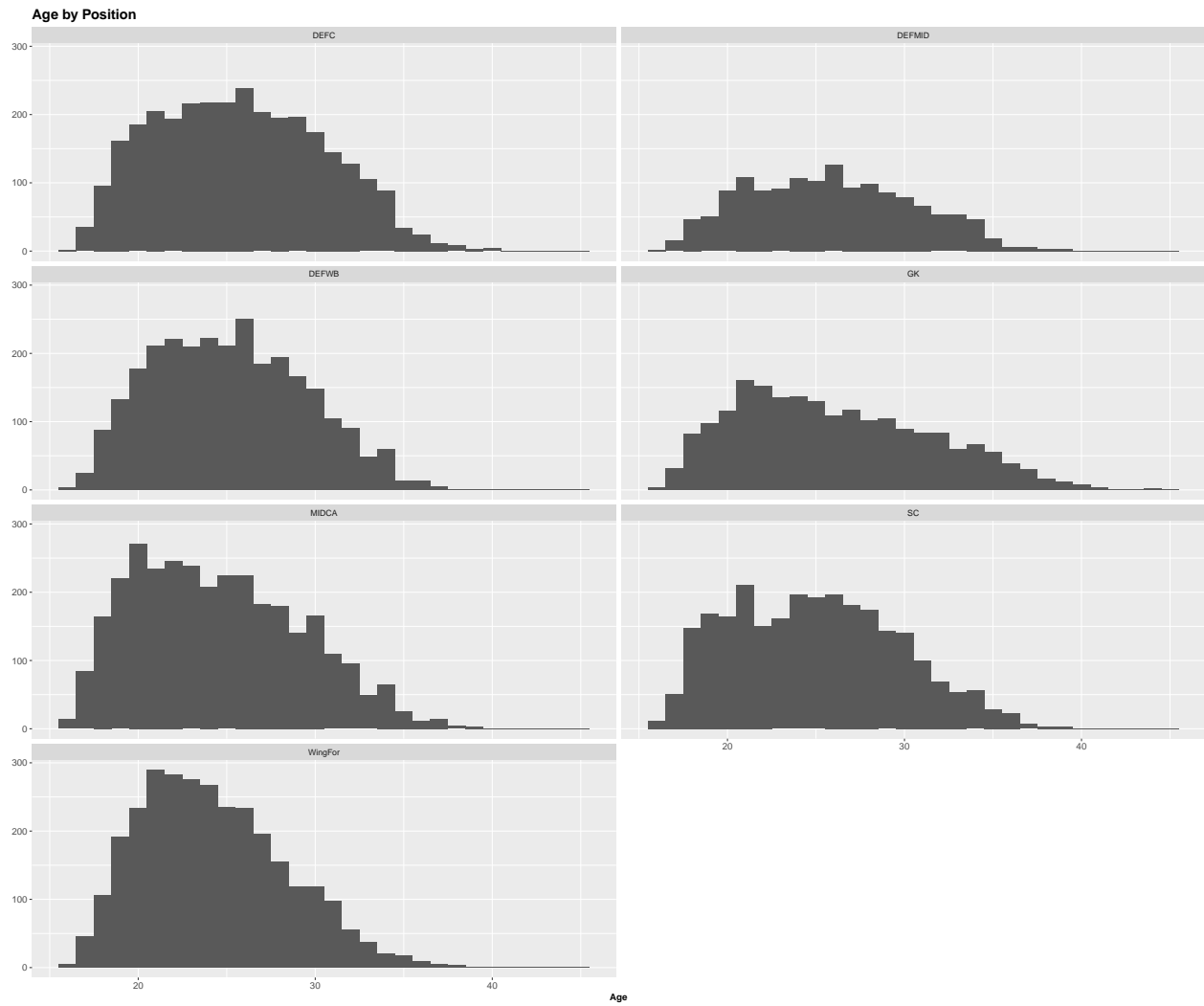
5.4 Age vs Overall for each group of positions

Also I want to compare the age vs the overall score for each position. The data shows that all score increase until 30 yo and stabilizing after to decrease until the end of their career. So, during the transfert period, a players around 30 yo are a safety but more expensive choice.

Age vs Overall Rating Curve
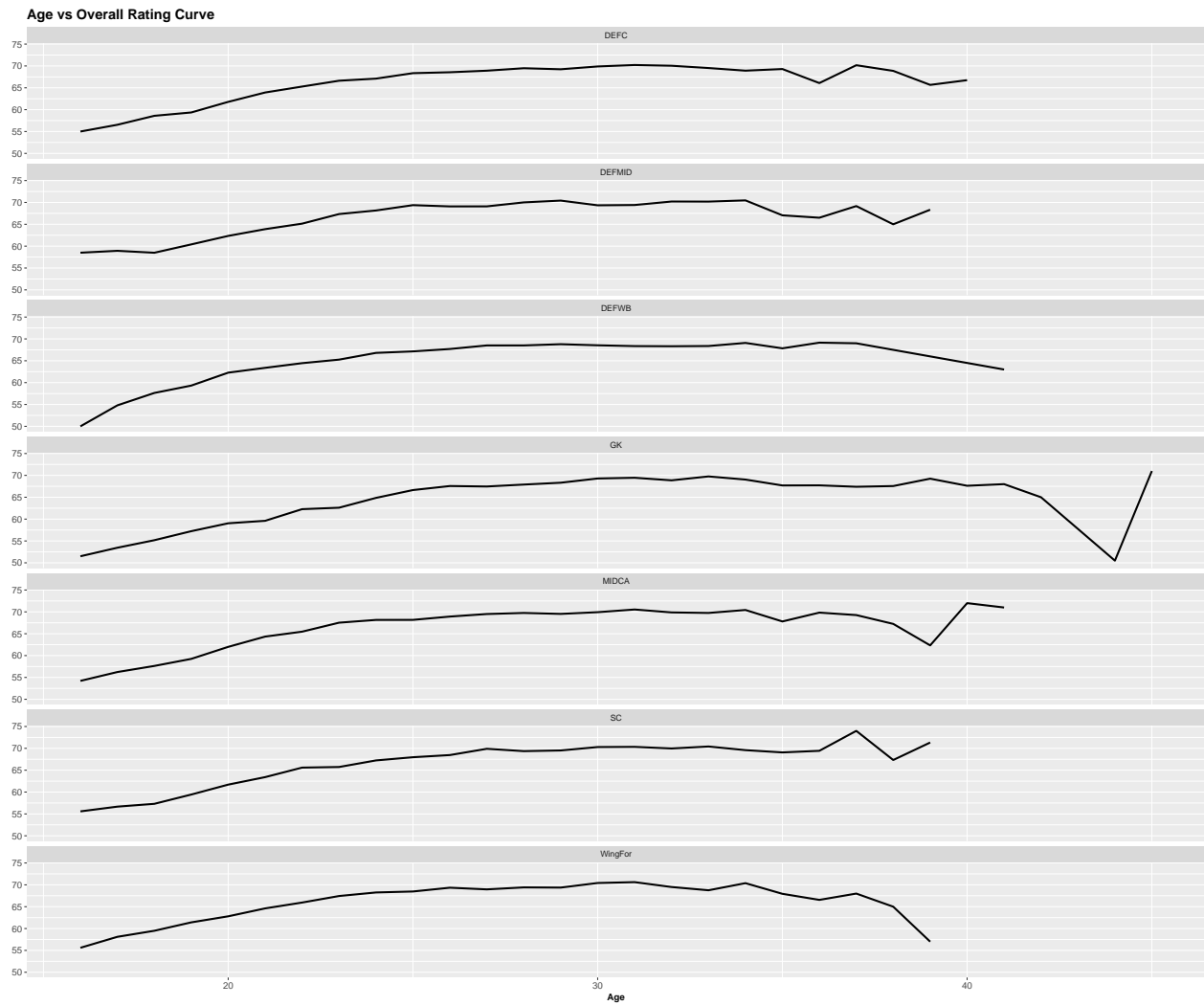
5.5 Positions vs skills

To conclude the overview per position, I think a coach wants to know the common skills of players in the same position to compare players helping to take decision.

Skills Overview per Positions

DEFC

	Acceleration	Aggression	Agility	Balance	BallControl	Composure	Crossing	Curve	Dribbling	Finishing	FKAccuracy	HeadingAccuracy	Interceptions	Jumping	LongPassing	LongShots	Marking	Penalties	Positioning	Reactions	ShortPassing	ShotPower	SlidingTackle	SprintSpeed	Stamina	StandingTackle	Strength	Vision	Volleys
RCB	59	76	56	56	61	67	44	40	52	34	36	74	73	74	62	38	73	44	38	69	65	56	72	63	69	74	80	48	35
LCB	60	75	57	57	62	68	46	39	53	35	37	73	73	76	62	40	72	44	38	69	66	59	72	63	70	74	80	49	35
CB	58	75	55	55	62	68	45	39	51	34	35	74	72	73	61	38	72	44	37	69	66	58	72	61	67	74	79	47	33

DEFMID

	Acceleration	Aggression	Agility	Balance	BallControl	Composure	Crossing	Curve	Dribbling	Finishing	FKAccuracy	HeadingAccuracy	Interceptions	Jumping	LongPassing	LongShots	Marking	Penalties	Positioning	Reactions	ShortPassing	ShotPower	SlidingTackle	SprintSpeed	Stamina	StandingTackle	Strength	Vision	Volleys
RDM	66	76	67	68	72	70	60	60	68	55	55	63	72	70	71	65	70	58	62	72	74	71	69	66	80	72	74	68	54
LDM	67	73	67	66	73	71	63	65	68	58	62	62	71	67	72	67	70	59	63	71	74	72	66	65	77	70	72	71	53
CDM	64	76	66	66	72	71	60	59	66	52	55	63	73	70	72	64	70	55	59	70	74	70	70	63	77	73	73	68	51

DEFWB

	Acceleration	Aggression	Agility	Balance	BallControl	Composure	Crossing	Curve	Dribbling	Finishing	FKAccuracy	HeadingAccuracy	Interceptions	Jumping	LongPassing	LongShots	Marking	Penalties	Positioning	Reactions	ShortPassing	ShotPower	SlidingTackle	SprintSpeed	Stamina	StandingTackle	Strength	Vision	Volleys
RWB	77.5	71.5	72	69.5	74.5	67.5	73.5	61	72	49.5	43.5	63	72	70.5	68	53	71	52.5	67.5	73	73	66	73.5	78	78	72.5	69.5	65	49
RB	76	72	73	71	71	67	72	63	70	47	47	63	71	72	65	56	70	50	63	71	70	66	73	77	78	73	68	60	47
LWB	77	72	74	72	74	70	73	67	73	51	62	59	72	71	67	61	69	59	68	72	72	70	74	77	79	74	67	65	47
LB	76	72	73	72	71	68	73	67	71	49	55	62.5	71	72	65.5	59	70	51	63	71	71	68	72	76	78	73	68	61	49

MIDCA

	Acceleration	Aggression	Agility	Balance	BallControl	Composure	Crossing	Curve	Dribbling	Finishing	FKAccuracy	HeadingAccuracy	Interceptions	Jumping	LongPassing	LongShots	Marking	Penalties	Positioning	Reactions	ShortPassing	ShotPower	SlidingTackle	SprintSpeed	Stamina	StandingTackle	Strength	Vision	Volleys
RCM	67	72	72	71	74	72	66	68	71	63	62	59	69	67	73	71	66	61	68	73	76	74	65	67	77	69	69	73	60
RAM	78	46	79	76	74	71	73	71	74	64	73	49	47	55	63	66	30	69	72	73	74	63	36	76	66	47	55	74	66
LCM	69	71.5	73	74	74.5	72.5	67	68	72	62	65	57	69.5	67	72	70	66	60	68	72	75.5	73	64	68	77	70	68	72	59.5
LAM	73	48	77	74	75	72	74	70	75	61	64	56	42	59	68	65	26	66	71	70	73	69	27	72	64	34	62	75	63
CM	68.5	70.5	72	72	75	72	67	67	72.5	62	65	59	69	67	74	70	66	61	68	72	76	73.5	64	67	76	69	69	73	61
CAM	73	56	77	75	75	71.5	69	71	75	68	69	54	43.5	63	70	71	44	66	70	71	74	72	40	71	71	45	61	74	65

SC

	Acceleration	Aggression	Agility	Balance	BallControl	Composure	Crossing	Curve	Dribbling	Finishing	FKAccuracy	HeadingAccuracy	Interceptions	Jumping	LongPassing	LongShots	Marking	Penalties	Positioning	Reactions	ShortPassing	ShotPower	SlidingTackle	SprintSpeed	Stamina	StandingTackle	Strength	Vision	Volleys
ST	72	62	70	66	72	70	58	62	70	74	55	72	30	73	53	67	31	69	75	71	67	75	23	73	69	27	76	64	69
RS	74	65	73	68	73	71	58	63	71	75	54	71	33	73	57	67	36	69	75	71	67	75	24	76	71	30	76	65	68
LS	73	64	73	67	74	72	61	66	73	76	56	70	33	71	56	70	35	70	76	71	68	76	26	76	70	32	74	65	70
CF	77	60.5	79	78	77	72	66	69.5	77.5	72	65	63	30.5	68	63	70	40	69	72.5	73	73.5	74.5	31	75	71.5	36.5	64	71.5	70

WingFor

|	Acceleration	Aggression	Agility	Balance	BallControl	Composure	Crossing	Curve	Dribbling	Finishing	FKAccuracy	HeadingAccuracy	Interceptions	Jumping	LongPassing	LongShots	Marking	Penalties	Positioning	Reactions	ShortPassing	ShotPower	SlidingTackle	SprintSpeed	Stamina	StandingTackle	Strength	Vision	Volleys
RW	82	56	82	77	75	70	71	71	77	68	65	53	36	61	65	68	35	63	71	70	72	73	30	80	72	32	60	70	64
RM	79	58	79	76	74	69	71	69	76	67	62	54	43	66	66	67.5	43	62	70	70	72	71	38	79	73	41	63	70	62
RF	76	53	75	77	78	74	72	76	77	73	75	66	36	64	70	71	39	74	75	74	73	72	35	75	72	40	62	73	67
LW	82	57	81	77	75	71	69	70	77	69	64	55	36	65	63	70	40	66	70	70	70	73	31	80	71	36	62	69	66
LM	80	59	80	77	74	69	70	70	76	68	64	55	40	66	65	69	42	63	70	70	71	73	35	79	73	39	63	70	62
LF	76	56	84	79	80	74	75	80	78	70	70	58	41	64	75	74	40	71	73	78	76	79	42	76	66	40	65	74	71

Sop, some skill seems to be specific to the defensive player, other to the attack players, etc... Some position require some polyvalence, like the defensive midfield and central player It's logical, but the data confirms this.

6 Valuation and skills Analysis

After analyzing the technical part, a more financial analysis is necessary to have a complete understanding.

6.1 Player Valuations

So, the first thing I want to analyse is the distribution of the valuation.

Player Valuations Repartition
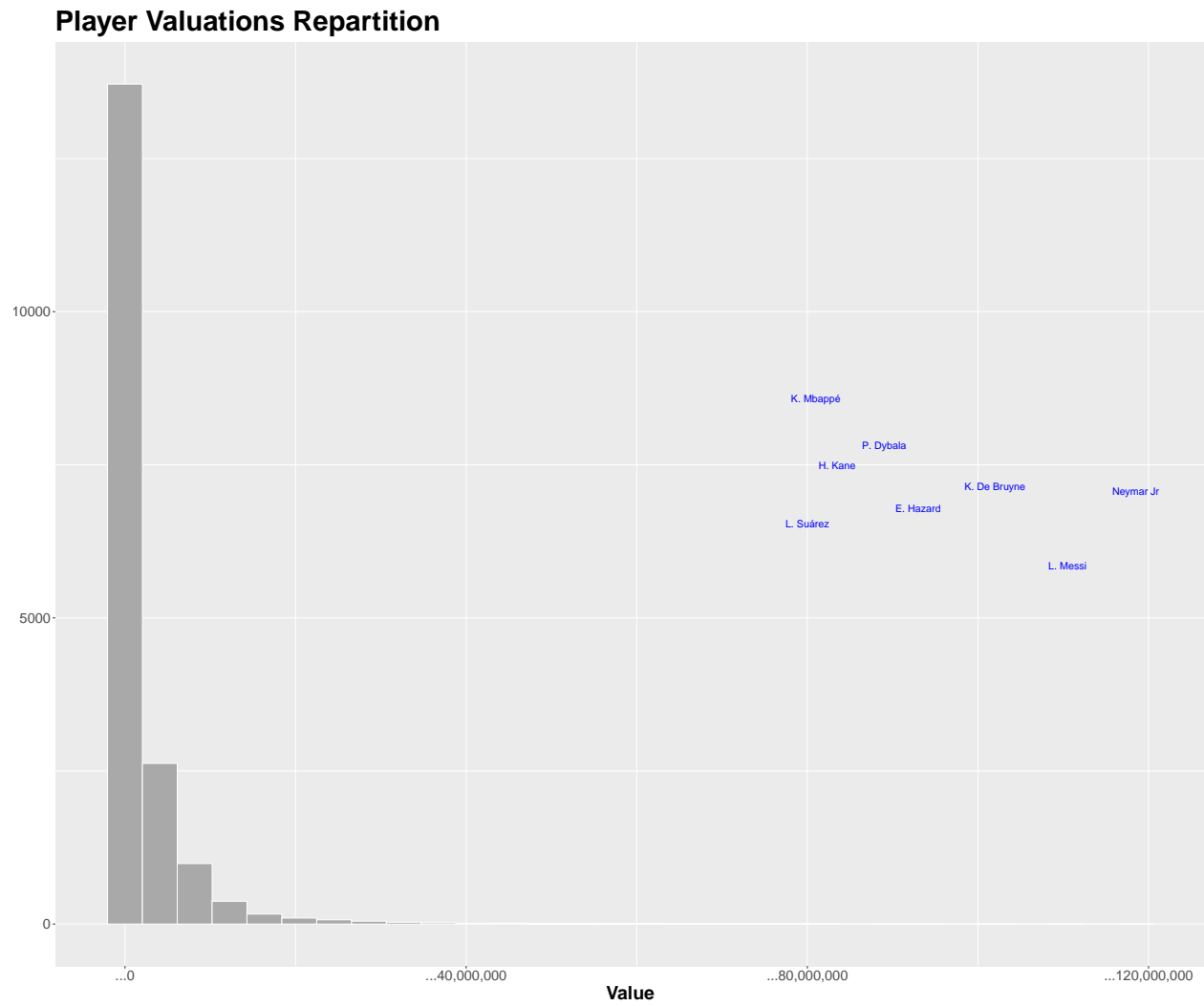
There are 8 players with a value $>= 80{,}000{,}000$ € who create a heavily positive skew. So I remove this extreme value for the rest of this study to not influence the results.
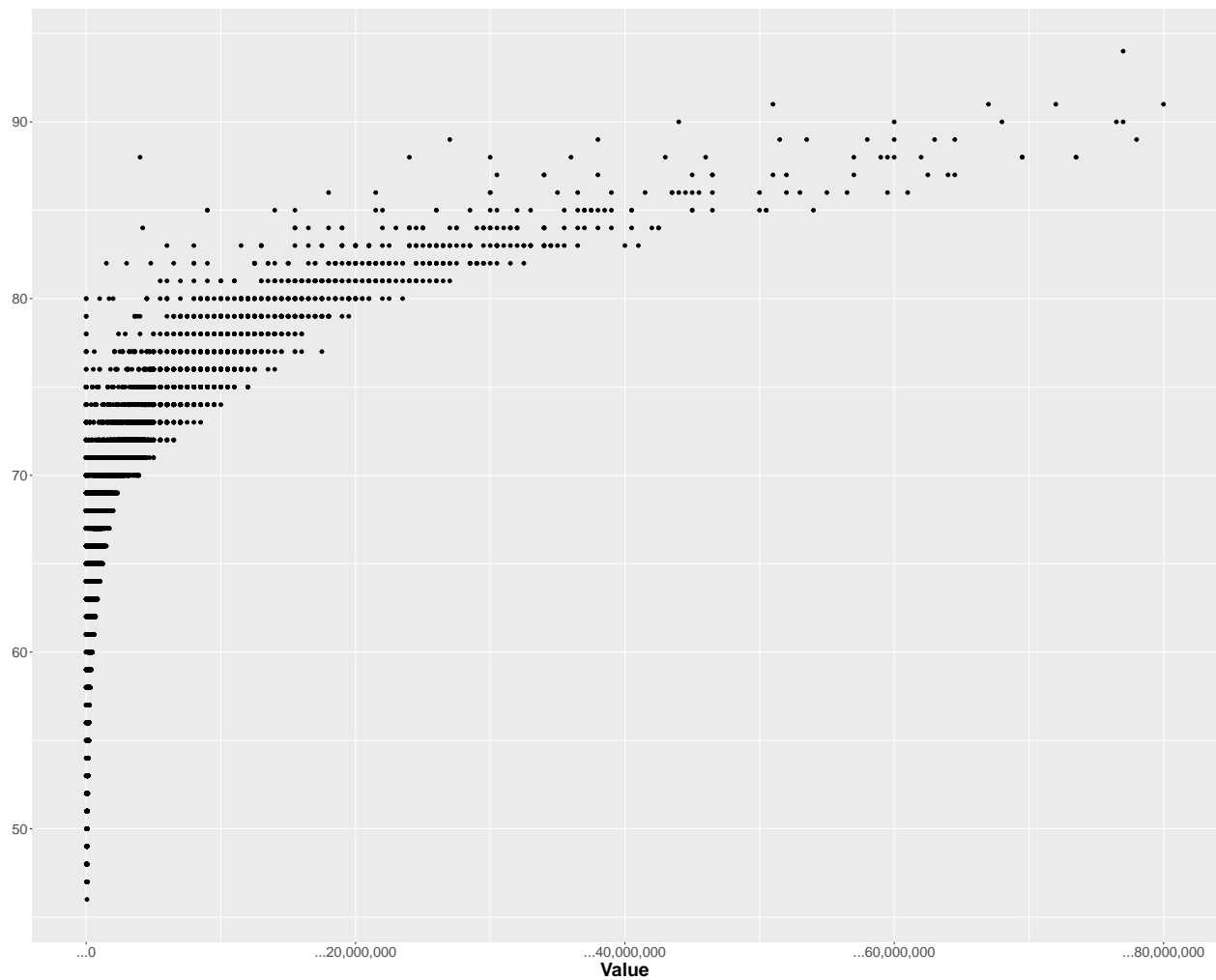
Player Valuations Repartition
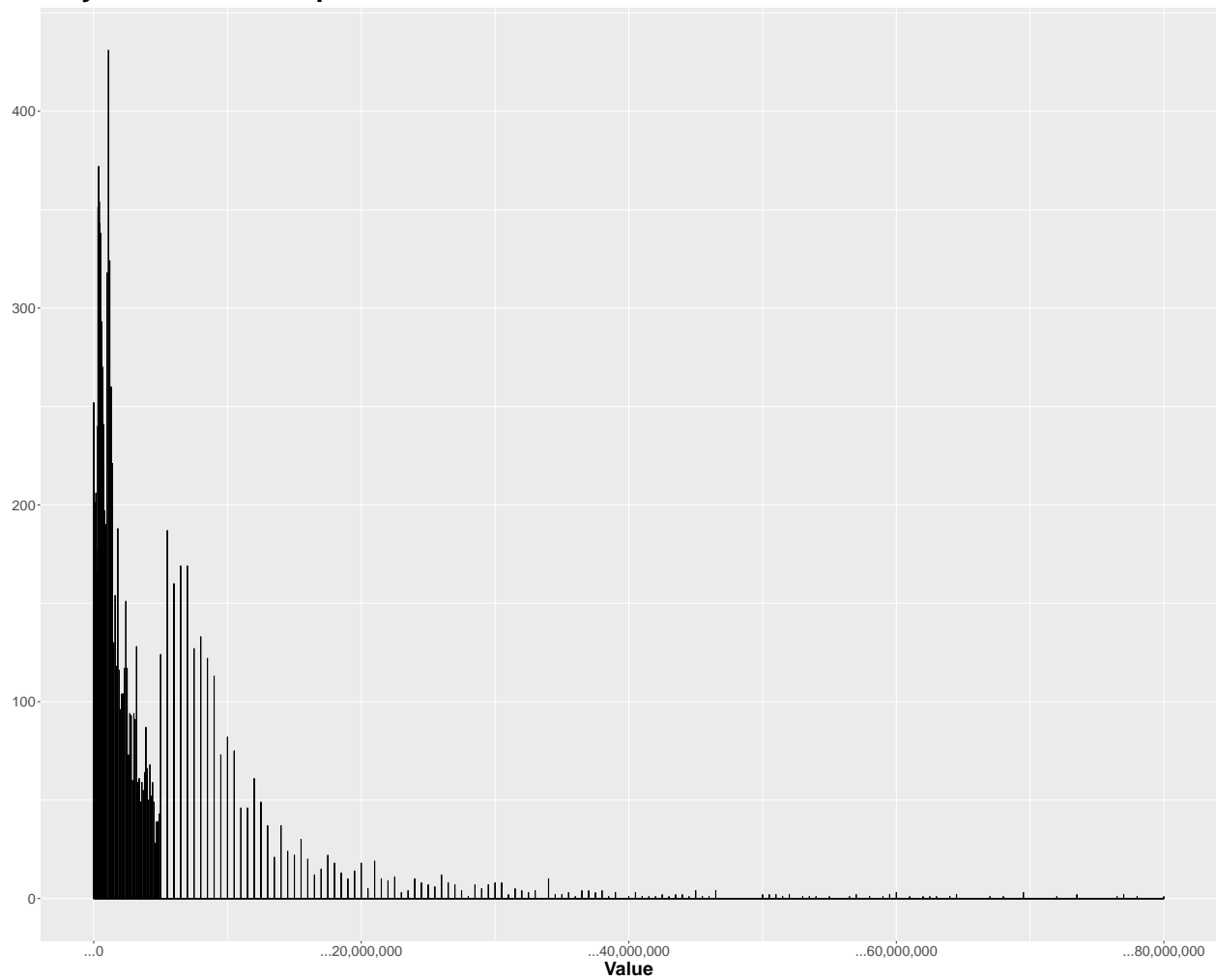
Most players are a valuation below 20,000,000 Euros.

6.2   Overall Valuations

Excluding the 8 more valuable players, I want to know how much a overall point costs. As we can see below, the valuation is related to the overall score, there is a strong non-linear correlation between this two variables. But around an overall score between 80 and 90, there is a lot disparity, so an another variable should be influencing the valuation.

Overall score vs Valuations

6.3 Age vs Valuations

And my fisrt intuition is to think with age and experience the player valuation would increase and so reflect a relationship, but it's not as relevant as I thought. On the graph we can see an increasing of the value over the year until approximately 25 yo and start decreasing after but there are some disparity on the valuation between players.

Age vs Valuations
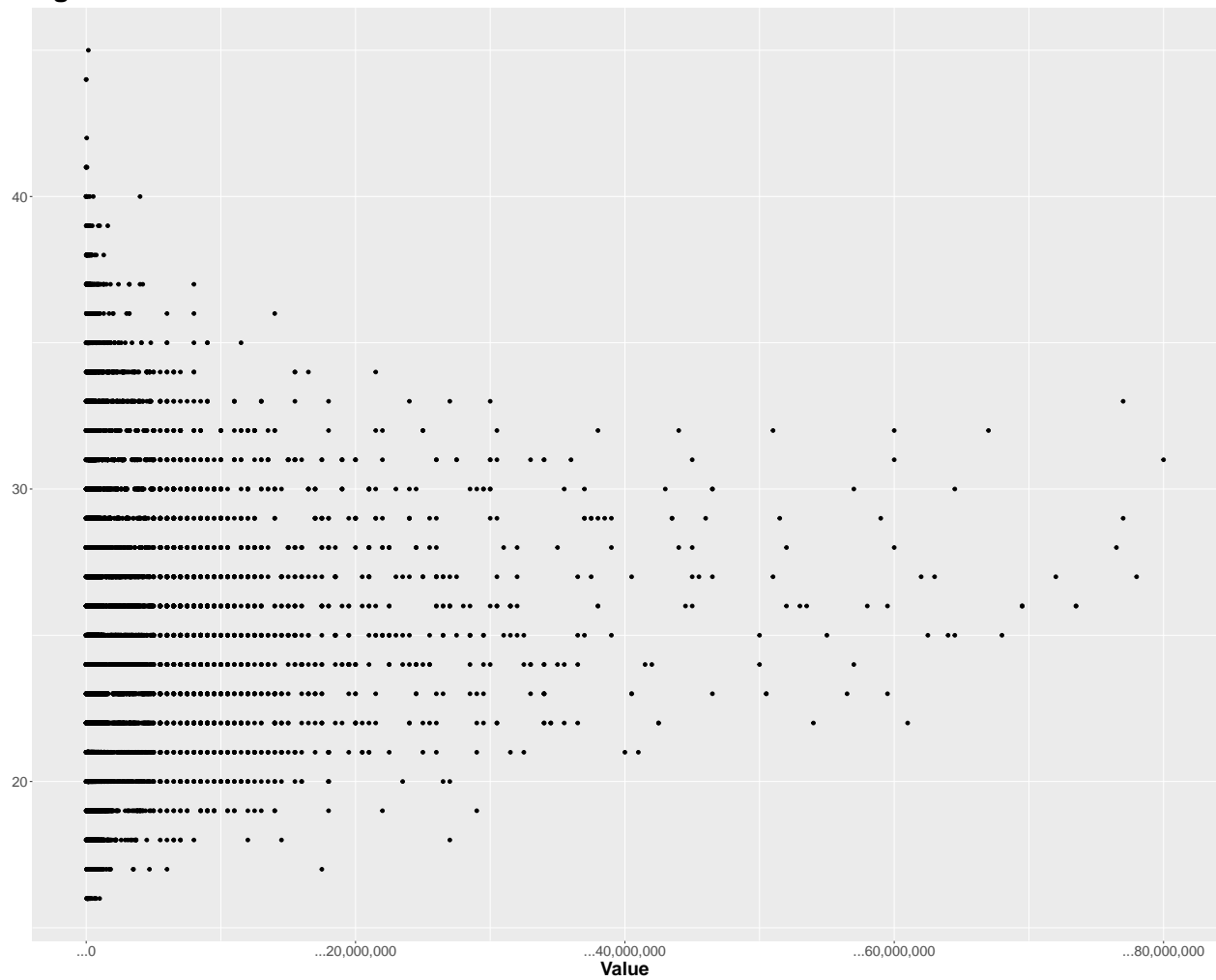
The trend is more visible with the boxplot but we can say that the value is define by others variables. (we will show them with the regression study).

Age Group vs Valuations
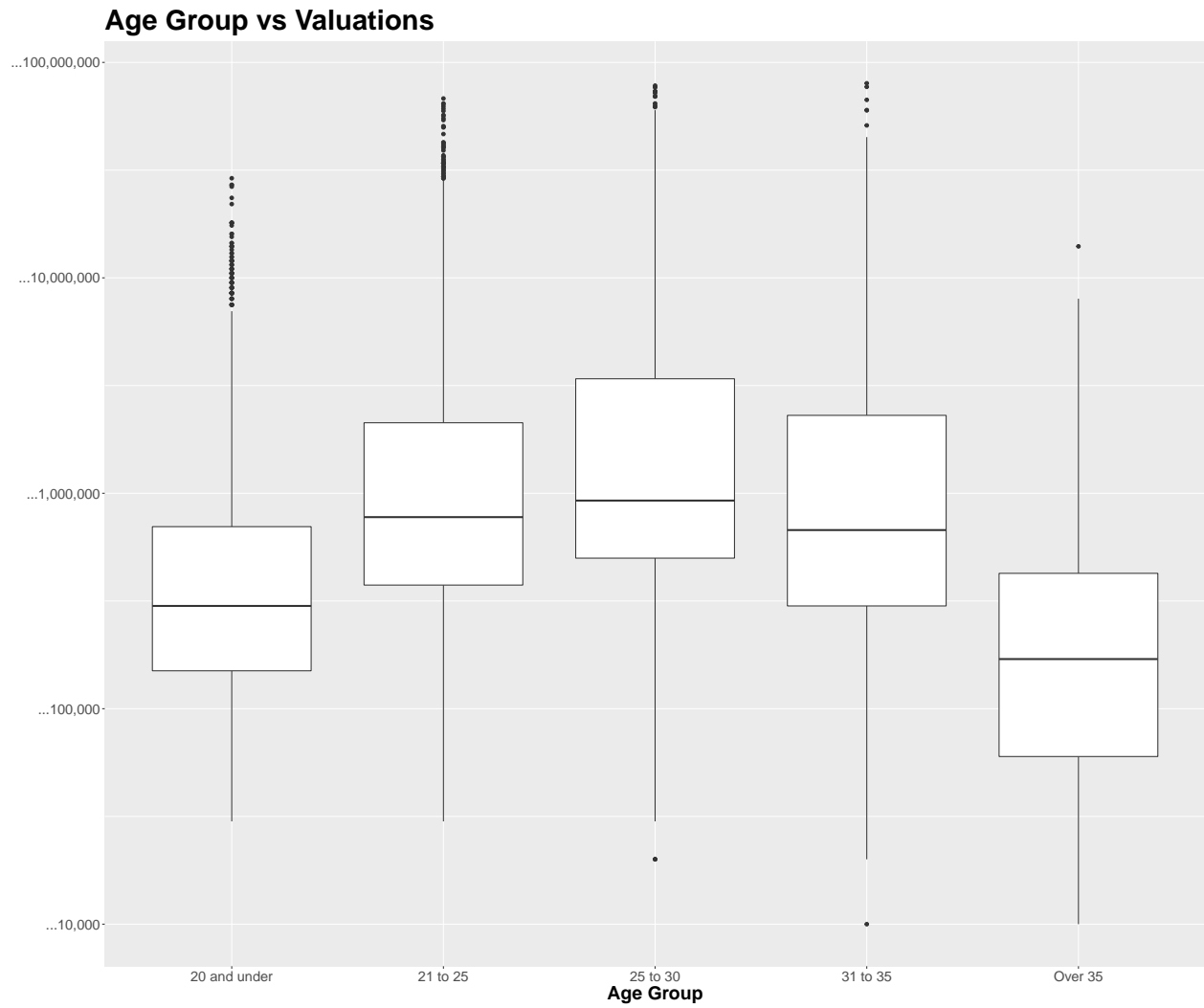
6.4 Position vs Valuations

Also, all players are not the same valuation according to their positions. With the main group, it is diffucult to see a difference between them but it seems offensive players are a greater valuation.

Position vs Valuations
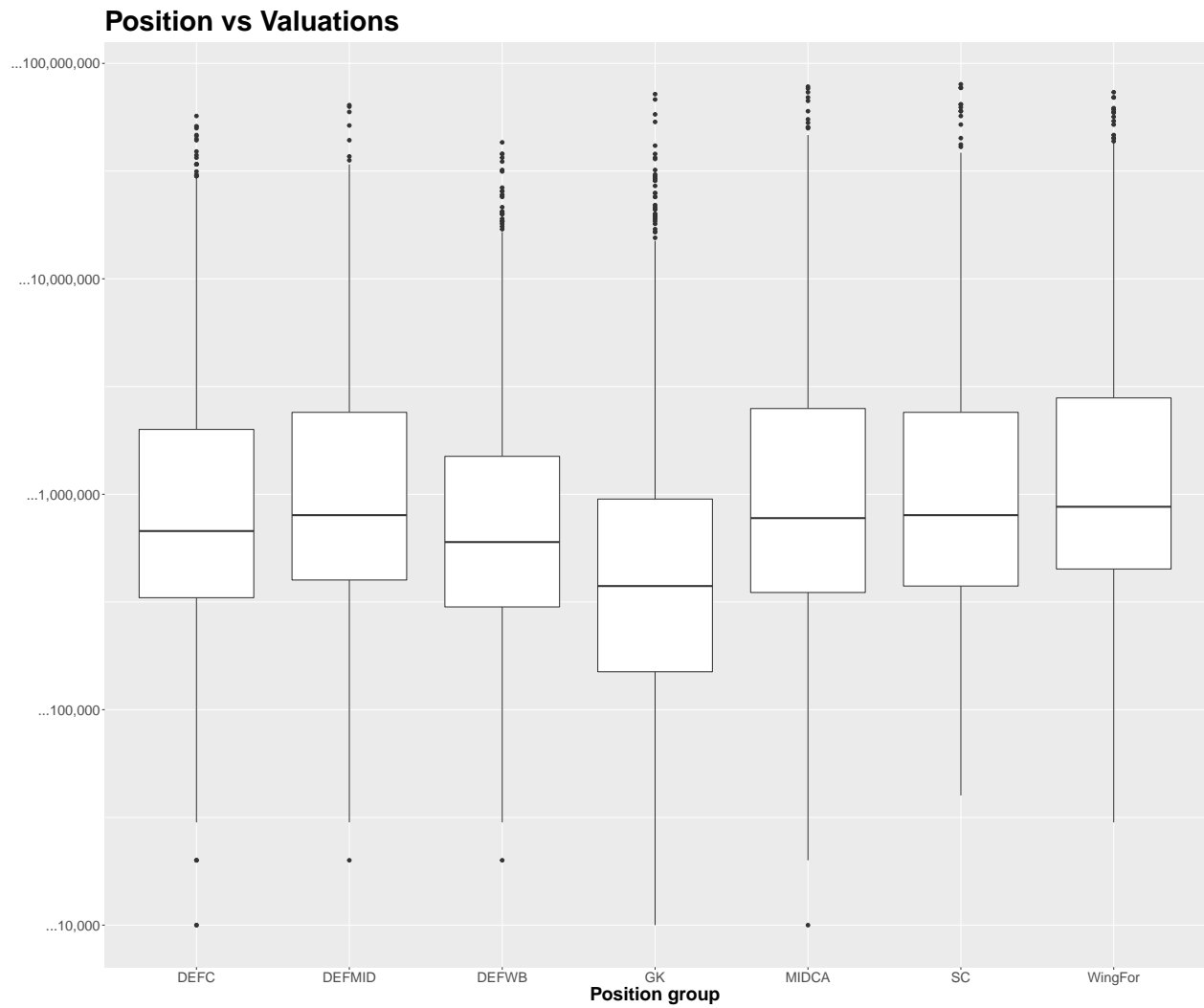
It is more distinct when the position are not grouped together. Specially, the wing-forward players (left or right) cost more than others.
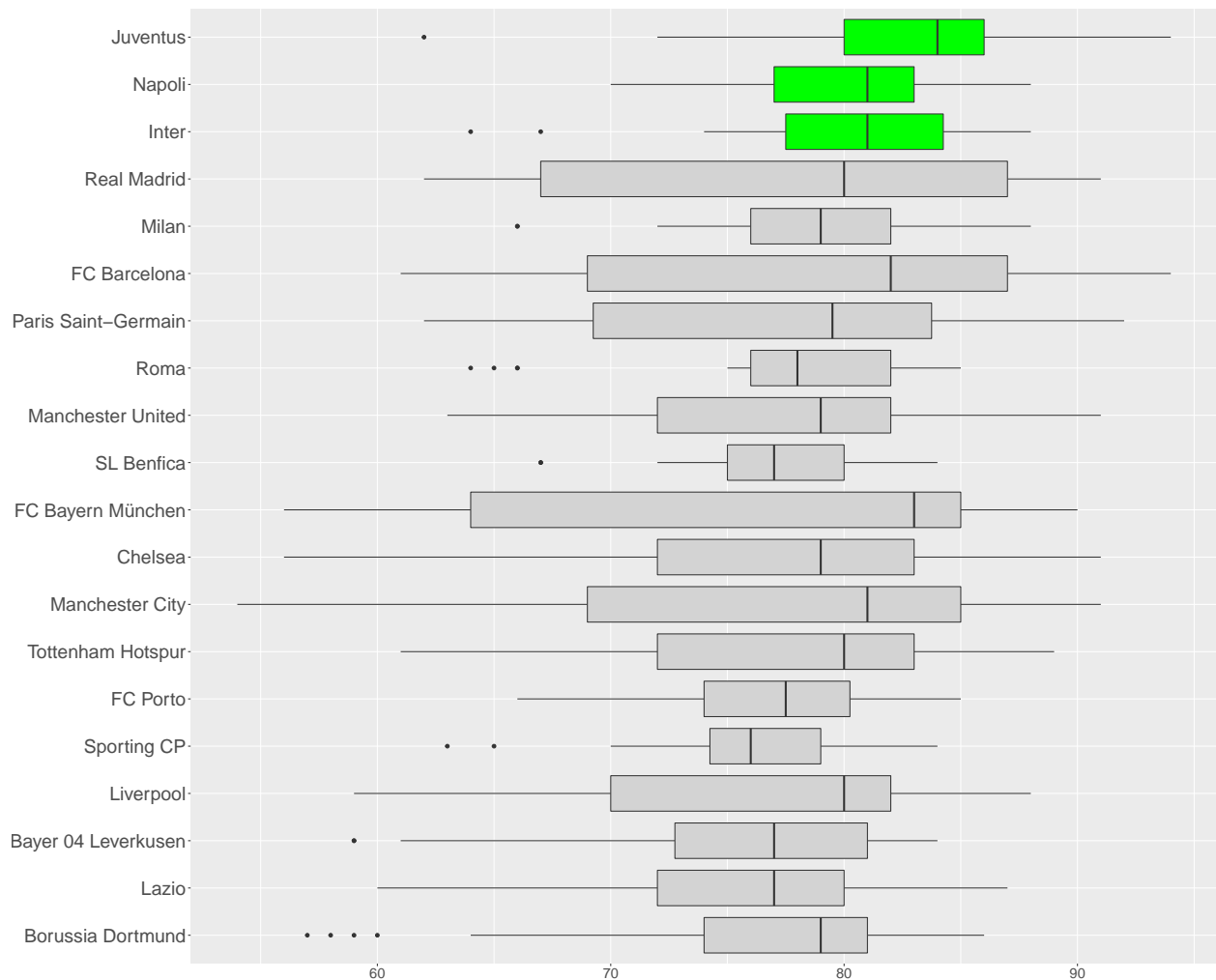
Position group vs Valuations
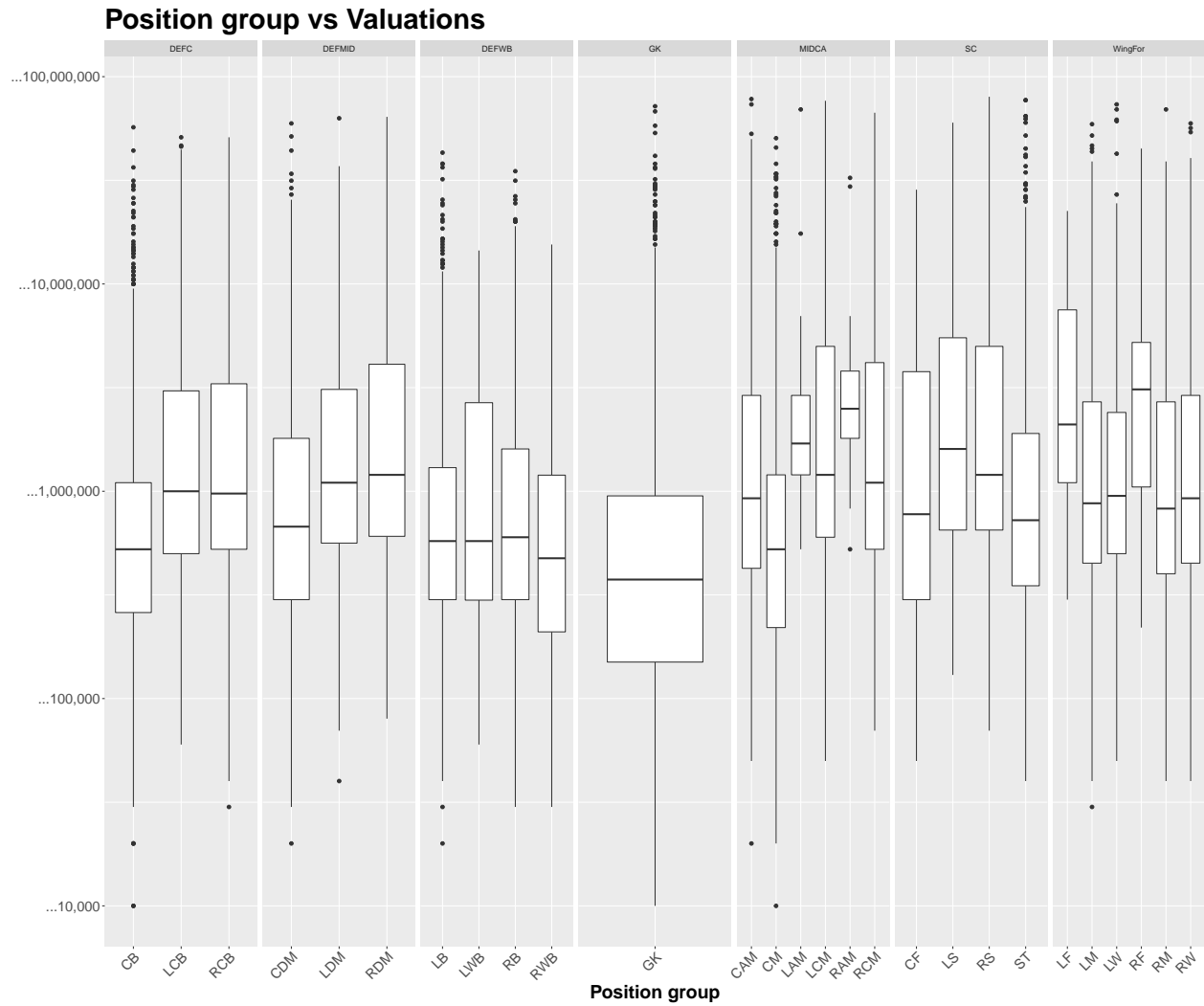
7 Team Analysis

After doing an individual analysis, I would like to compare the teams with each other, and see their difference in management.

7.1 Team Overall Talent

And the first statistic I want to compare between the team is the overall score.

Teams Overall Talent

The top 3 is compose with only italian club but this result need to be nuanced (or qualify) because they have less players in their team. If we take the 25 best players in each team (because each has around 25 main players), so we will probably see a different graph. The last players are generally the young players who play with the reserve team and not with the first team.

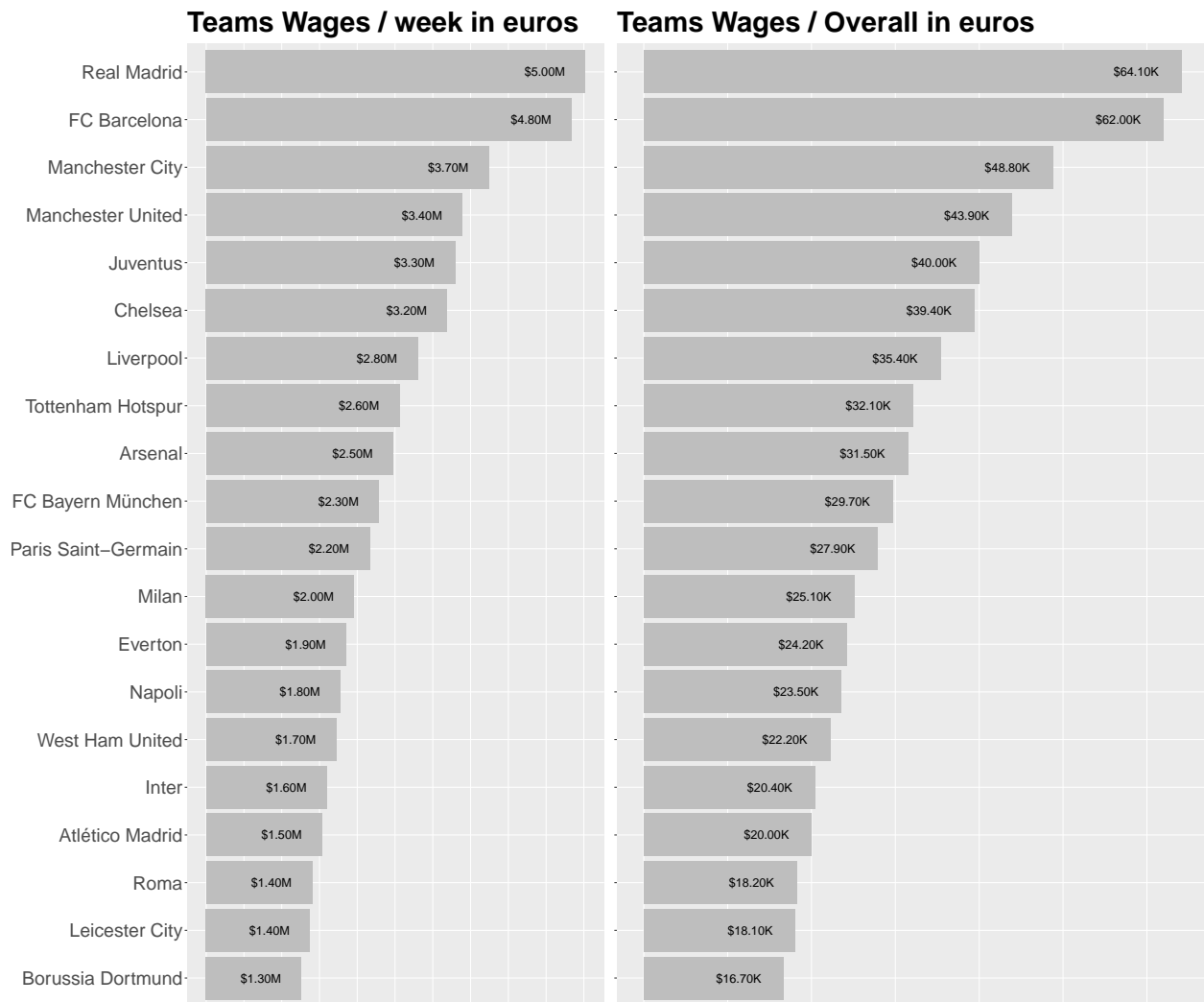**Teams Overall Talent with the 25 best player**
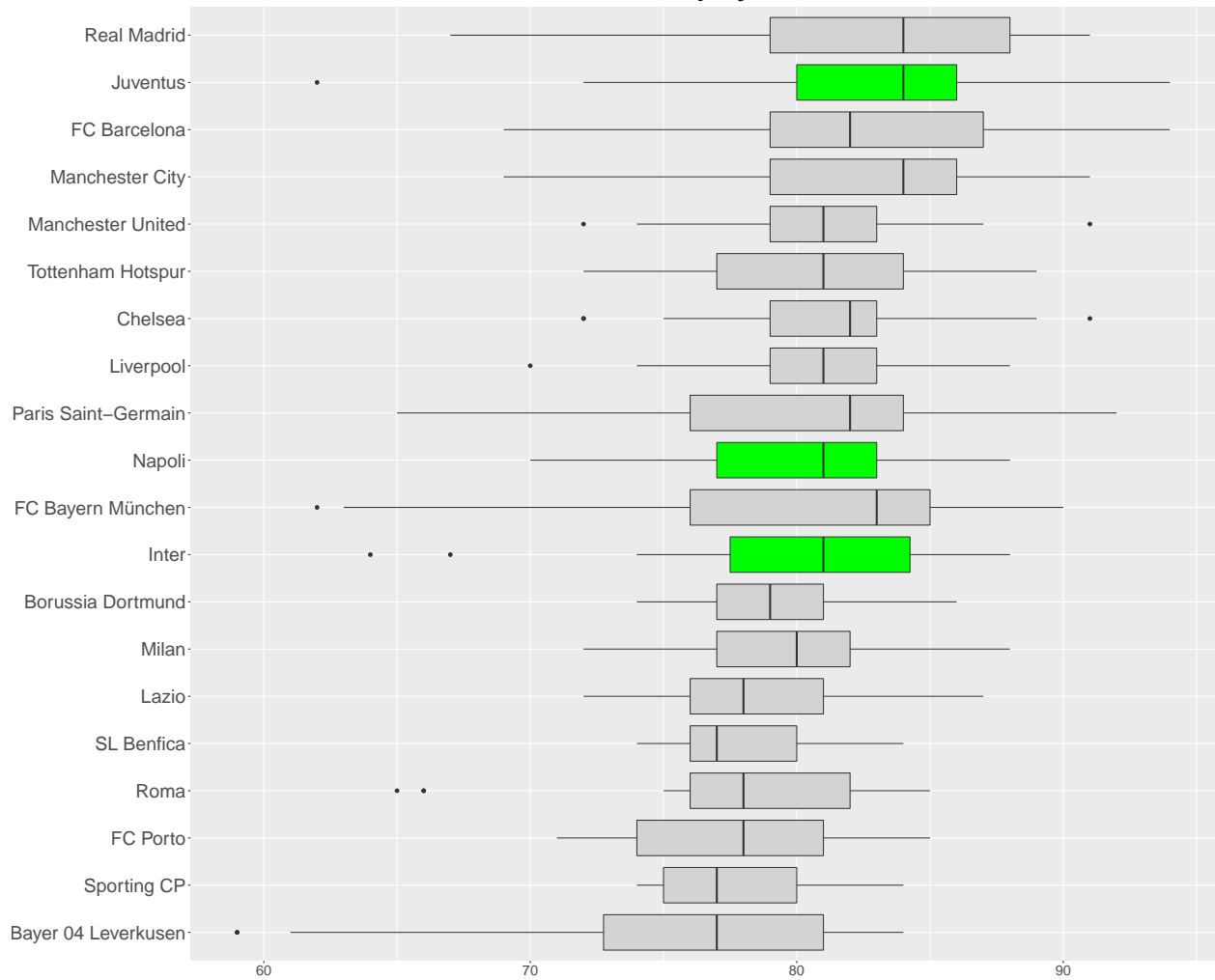
Now the top 3 is composed with Real Madrid, Juventus and FC Barcelona followed closely by Manchester City. This rating reflects more accurately the reality.

7.2   Team Wage Bills and Value

But having the best players should have a big wage so the team with the most best players should have a big wage per week.

The two spanish team have a bigger amount of wage per week so each overall point costs more than the other teams. If we look the 4 teams with the most overall talent, this 2 club seem to over pay their players.

Teams Wages / week in euros   **Teams Wages / Overall in euros**

Team	Wages / week	Wages / Overall
Real Madrid	$5.00M	$64.10K
FC Barcelona	$4.80M	$62.00K
Manchester City	$3.70M	$48.80K
Manchester United	$3.40M	$43.90K
Juventus	$3.30M	$40.00K
Chelsea	$3.20M	$39.40K
Liverpool	$2.80M	$35.40K
Tottenham Hotspur	$2.60M	$32.10K
Arsenal	$2.50M	$31.50K
FC Bayern München	$2.30M	$29.70K
Paris Saint–Germain	$2.20M	$27.90K
Milan	$2.00M	$25.10K
Everton	$1.90M	$24.20K
Napoli	$1.80M	$23.50K
West Ham United	$1.70M	$22.20K
Inter	$1.60M	$20.40K
Atlético Madrid	$1.50M	$20.00K
Roma	$1.40M	$18.20K
Leicester City	$1.40M	$18.10K
Borussia Dortmund	$1.30M	$16.70K

7.3 Teams Age

Let starting by analyzing the age average and see the 20 younger and older team. After looking at this graph, a first observation is that the majority of the older team come from South America and all younger team from Europa.
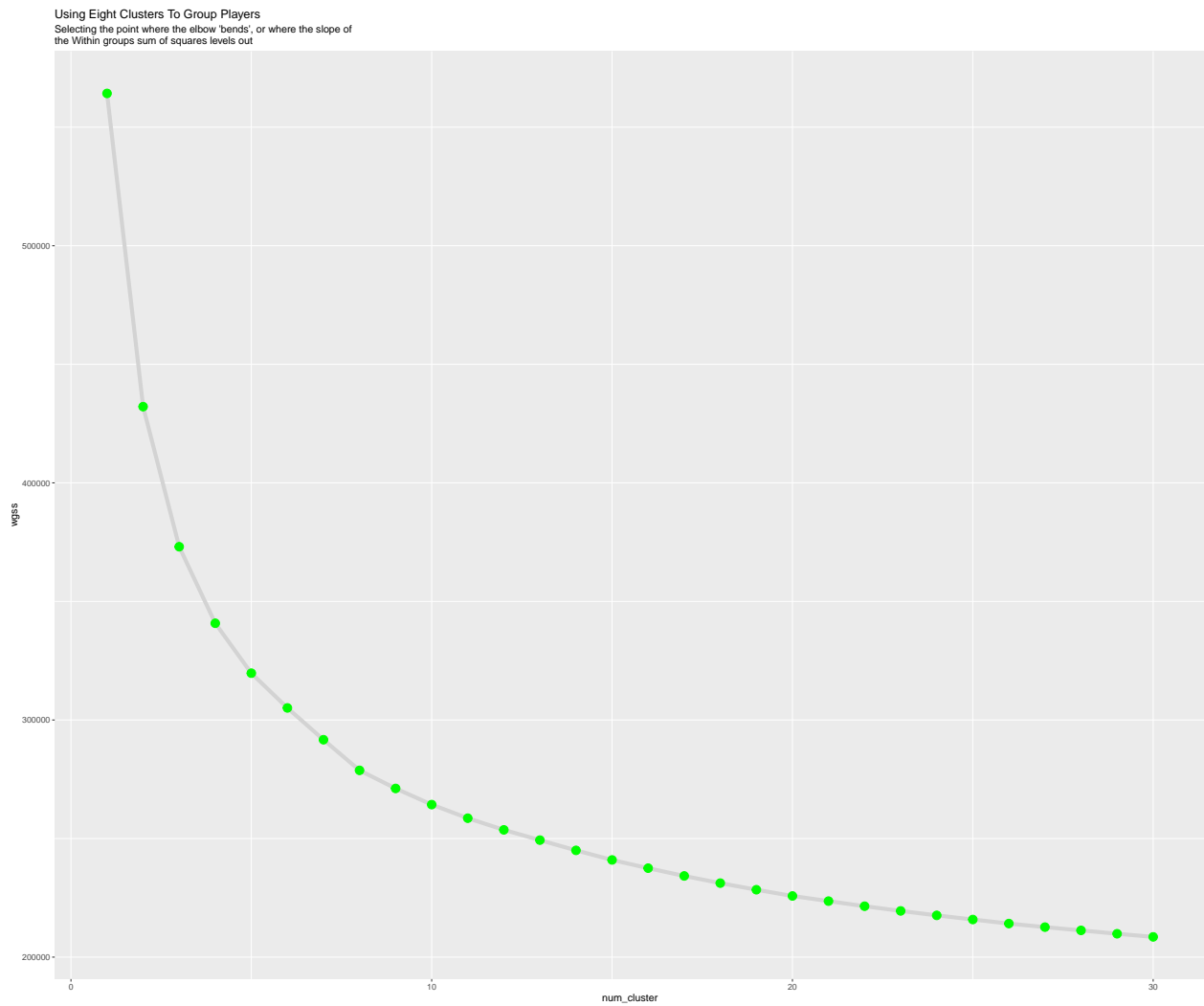
Teams Average Age
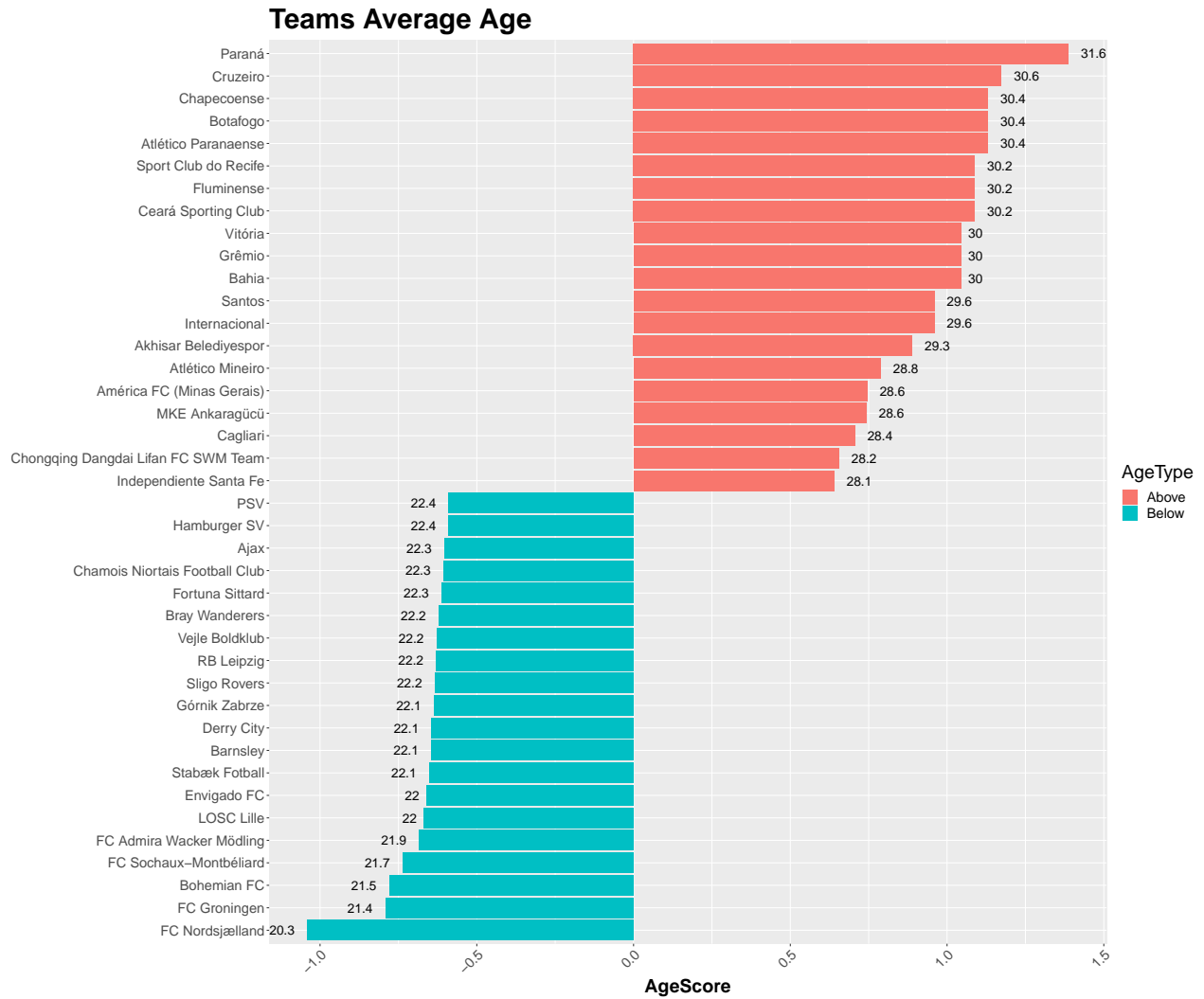
8 K-means clustering

The overview analysis is completed, the last thing I want to do is to try to group the players with the same characteristic.

8.1 Prepare the data

The first step is to determine the optimal number of cluster.

Using Eight Clusters To Group Players
Selecting the point where the elbow 'bends', or where the slope of
the Within groups sum of squares levels out

8.2 Train the model

Here, eight clusters seem to be a good choice because it represents the "elbow bends".

The K-means analysis has split the data into 8 groups that we can see above. Let's see some statistic for each cluster.

The cluster 6 has a overall average higher

Overall vs Value for the top 20 players in each cluster
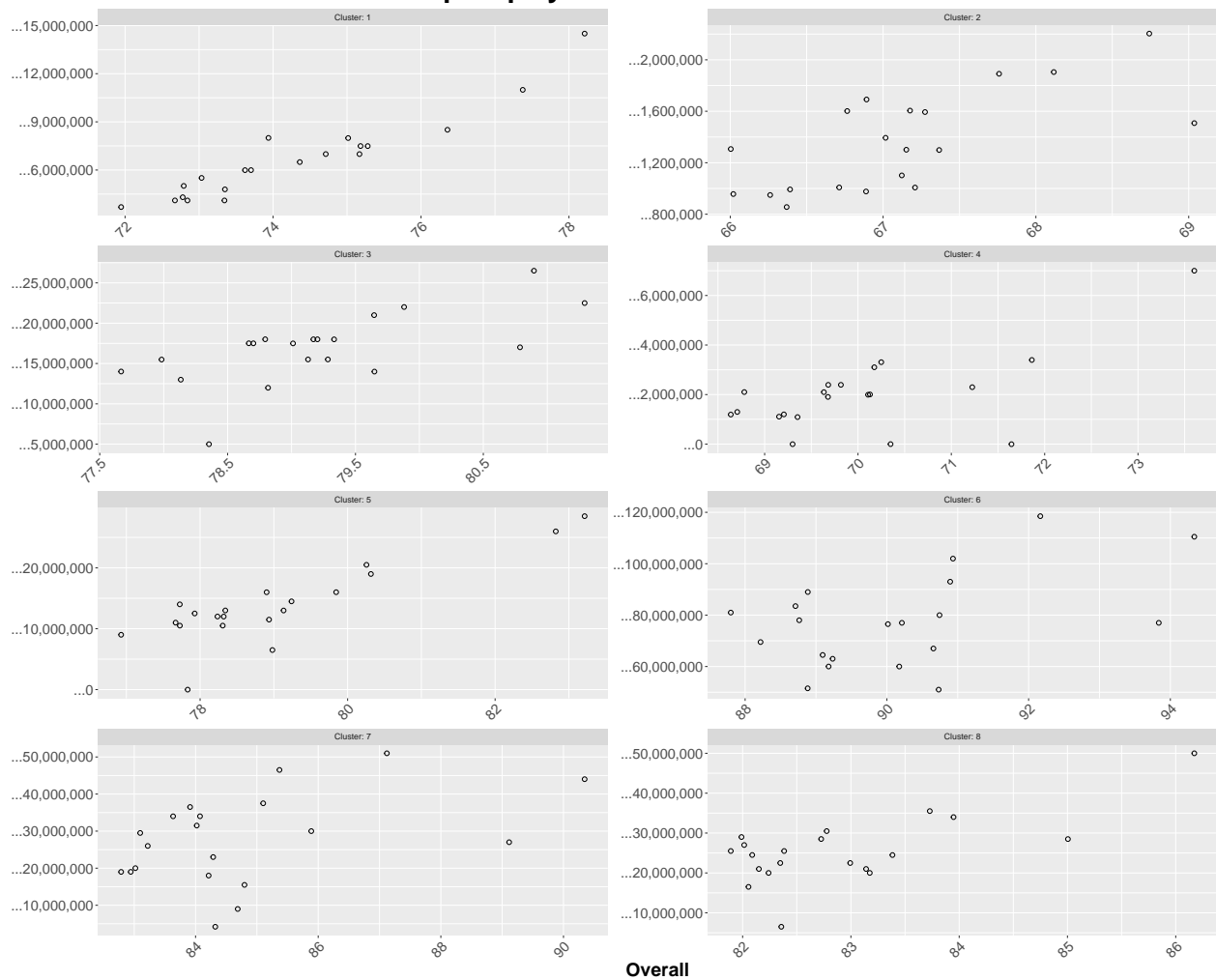
Just for curiosity, for each cluster this is the top 3 players.

Name	Club	Age	PositionSubGroup	Overall	Cluster
J. Gomez	Liverpool	21	DEFC	78	1
A. Barreca	AS Monaco	23	DEFWB	75	1
A. Tuanzebe	Aston Villa	20	DEFC	72	1
Abel Ruiz	FC Barcelona	18	SC	67	2
D. Cole	Burton Albion	23	SC	66	2
F. Feuillassier	Real Madrid	20	WingFor	65	2
C. Wilson	Bournemouth	26	SC	79	3
A. Iwobi	Arsenal	22	WingFor	78	3
D. Ings	Southampton	25	SC	76	3
F. Ceccherini	Fiorentina	26	DEFC	72	4
C. Carter-Vickers	Swansea City	20	DEFC	69	4
K. Mavropanos	Arsenal	20	DEFC	68	4
Javi Sánchez	Real Madrid	21	DEFC	67	4
De Tomás	Rayo Vallecano	23	SC	76	5
Borja Mayoral	Levante UD	21	SC	75	5
O. Niasse	Everton	28	SC	75	5
L. Messi	FC Barcelona	31	WingFor	94	6

35

Name	Club	Age	PositionSubGroup	Overall	Cluster
L. Modrić	Real Madrid	32	MIDCA	91	6
L. Suárez	FC Barcelona	31	SC	91	6
G. Chiellini	Juventus	33	DEFC	89	7
M. Benatia	Juventus	31	DEFC	86	7
Sokratis	Arsenal	30	DEFC	84	7
R. Varane	Real Madrid	25	DEFC	86	8
N. Otamendi	Manchester City	30	DEFC	85	8
Nacho Fernández	Real Madrid	28	DEFC	83	8