project

```
df = read.csv("fifa_cleaned.csv")
head(df)
```

```
##       ID              Name Age Overall Potential                Club
## 1 158023          L. Messi  31      94        94        FC Barcelona
## 2  20801 Cristiano Ronaldo  33      94        94             Juventus
## 3 190871         Neymar Jr  26      92        93 Paris Saint-Germain
## 4 193080            De Gea  27      91        93   Manchester United
## 5 192985      K. De Bruyne  27      91        92     Manchester City
## 6 183277         E. Hazard  27      91        91             Chelsea
##       Value   Wage Special Preferred.Foot International.Reputation
## 1 110500000 565000    2202           Left                        5
## 2  77000000 405000    2228          Right                        5
## 3 118500000 290000    2143          Right                        5
## 4  72000000 260000    1471          Right                        4
## 5 102000000 355000    2281          Right                        4
## 6  93000000 340000    2142          Right                        4
##   Weak.Foot Skill.Moves      Work.Rate  Body.Type Position Height Weight
## 1         4           4 Medium/ Medium      Messi       RF    170    159
## 2         4           5      High/ Low C. Ronaldo       ST    187    183
## 3         5           5   High/ Medium     Neymar       LW    175    150
## 4         3           1 Medium/ Medium       Lean       GK    193    168
## 5         5           4     High/ High     Normal      RCM    180    154
## 6         4           4   High/ Medium     Normal       LF    172    163
##   Crossing Finishing HeadingAccuracy ShortPassing Volleys Dribbling Curve
## 1       84        95              70           90      86        97    93
## 2       84        94              89           81      87        88    81
## 3       79        87              62           84      84        96    88
## 4       17        13              21           50      13        18    21
## 5       93        82              55           92      82        86    85
## 6       81        84              61           89      80        95    83
##   FKAccuracy LongPassing BallControl Acceleration SprintSpeed Agility
## 1         94          87          96           91          86      91
## 2         76          77          94           89          91      87
## 3         87          78          95           94          90      96
## 4         19          51          42           57          58      60
## 5         83          91          91           78          76      79
## 6         79          83          94           94          88      95
##   Reactions Balance ShotPower Jumping Stamina Strength LongShots
## 1        95      95        85      68      72       59        94
## 2        96      70        95      95      88       79        93
## 3        94      84        80      61      81       49        82
## 4        90      43        31      67      43       64        12
## 5        91      77        91      63      90       75        91
## 6        90      94        82      56      83       66        80
##   Aggression Interceptions Positioning Vision Penalties Composure Marking
## 1         48            22          94     94        75        96      33
## 2         63            29          95     82        85        95      28
## 3         56            36          89     87        81        94      27
## 4         38            30          12     68        40        68      15
## 5         76            61          87     94        79        88      68
```

```
## 6         54            41          87     89        86        91      34
##   StandingTackle SlidingTackle GKDiving GKHandling GKKicking GKPositioning
## 1             28            26        6         11        15             14
## 2             31            23        7         11        15             14
## 3             24            33        9          9        15             15
## 4             21            13       90         85        87             88
## 5             58            51       15         13         5             10
## 6             27            22       11         12         6              8
##   GKReflexes Release.Clause
## 1          8      226500000
## 2         11      127100000
## 3         11      228100000
## 4         94      138600000
## 5         13      196400000
## 6          8      172100000
```

Clustering

```
colnames(df)
```

```
##  [1] "ID"                       "Name"
##  [3] "Age"                      "Overall"
##  [5] "Potential"                "Club"
##  [7] "Value"                    "Wage"
##  [9] "Special"                  "Preferred.Foot"
## [11] "International.Reputation" "Weak.Foot"
## [13] "Skill.Moves"              "Work.Rate"
## [15] "Body.Type"                "Position"
## [17] "Height"                   "Weight"
## [19] "Crossing"                 "Finishing"
## [21] "HeadingAccuracy"          "ShortPassing"
## [23] "Volleys"                  "Dribbling"
## [25] "Curve"                    "FKAccuracy"
## [27] "LongPassing"              "BallControl"
## [29] "Acceleration"             "SprintSpeed"
## [31] "Agility"                  "Reactions"
## [33] "Balance"                  "ShotPower"
## [35] "Jumping"                  "Stamina"
## [37] "Strength"                 "LongShots"
## [39] "Aggression"               "Interceptions"
## [41] "Positioning"              "Vision"
## [43] "Penalties"                "Composure"
## [45] "Marking"                  "StandingTackle"
## [47] "SlidingTackle"            "GKDiving"
## [49] "GKHandling"               "GKKicking"
## [51] "GKPositioning"            "GKReflexes"
## [53] "Release.Clause"
```

```
library(dbscan)
library(deldir)
```

```
## deldir 0.1-25
```

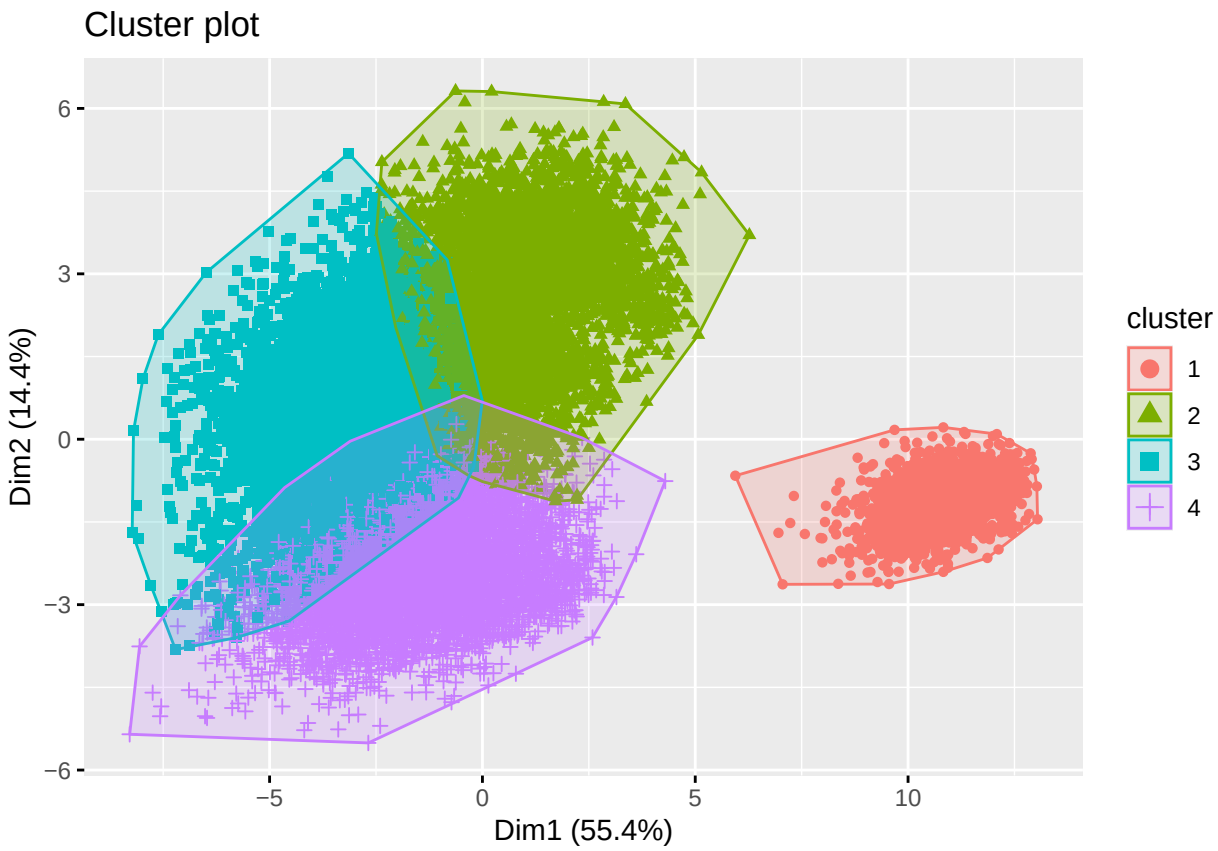
```
library(cluster)
library(RnavGraphImageData)
library(factoextra)
```

```
## Warning: package 'factoextra' was built under R version 3.6.2
## Loading required package: ggplot2
## Welcome! Want to learn more? See two factoextra-related books at https://goo.gl/ve3WBa
```

```
library(gridExtra)
```

```
set.seed(2)
dat = df[,19:52]
X=as.matrix(dat)
km.out=kmeans(X,4,nstart=15)
```

```
fviz_cluster(km.out,data = X,geom='point')
```




```
results = km.out$cluster

cluster.1 =  which(results %in% c(1))
cluster.2 =  which(results %in% c(2))
cluster.3 =  which(results %in% c(3))
cluster.4 =  which(results %in% c(4))

players1 = df[cluster.1,]
players2 = df[cluster.2,]
players3 = df[cluster.3,]
players4 = df[cluster.4,]


p1 = df$Position[cluster.1]
```

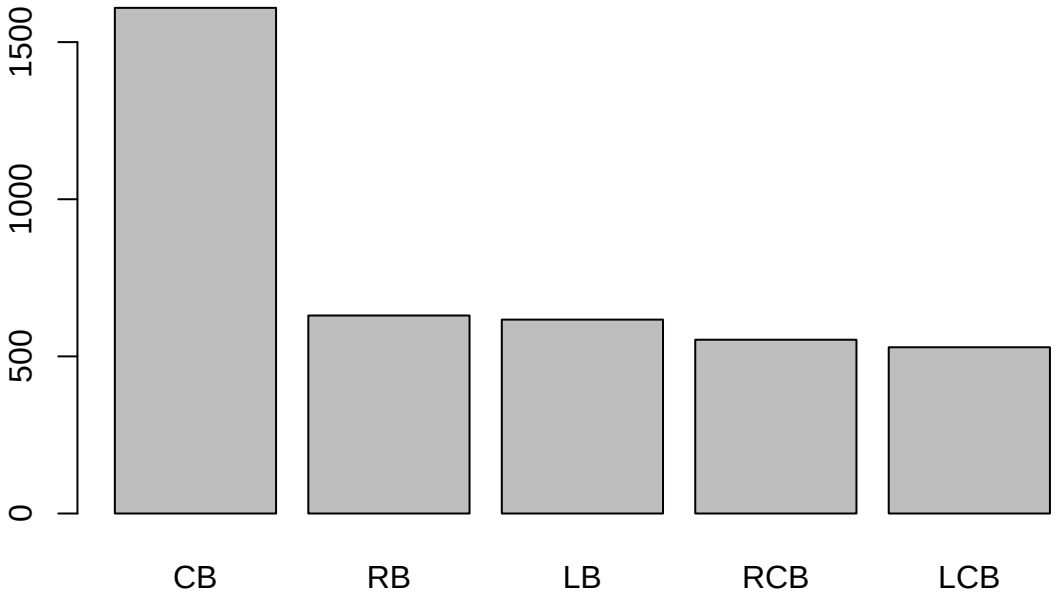
```
p2 = df$Position[cluster.2]
p3 = df$Position[cluster.3]
p4 = df$Position[cluster.4]
```

```
CBs = players2[players2['Position']=='CB',]
STs = players4[players4['Position']=='ST',]
```

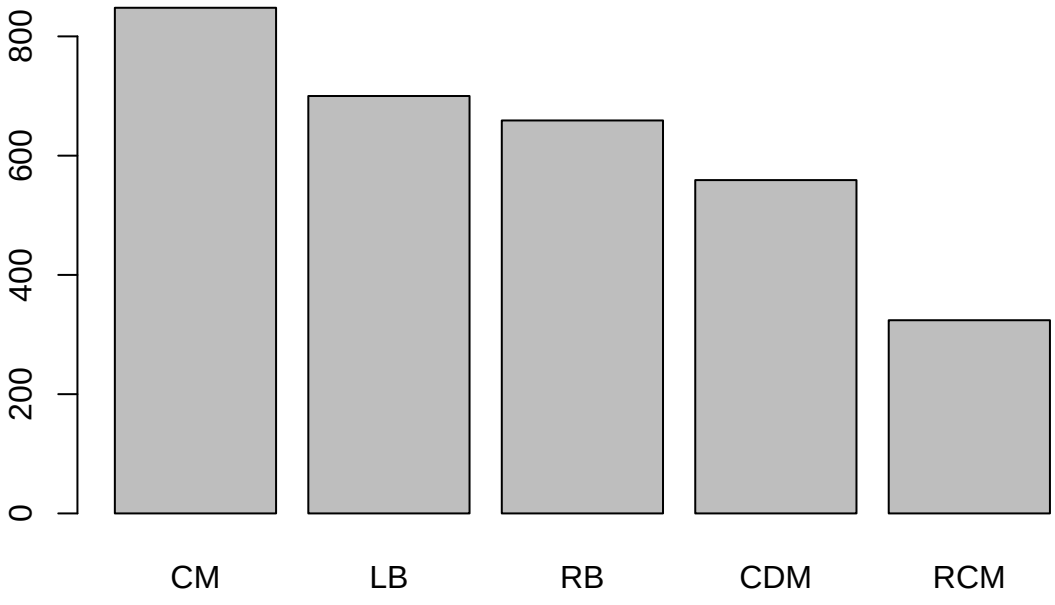
```
table(p2)
```

```
## p2
##       CAM    CB   CDM    CF    CM    GK   LAM    LB   LCB   LCM   LDM    LF    LM    LS
##     5   23  1609   387     0   341     0     0   617   529    35    46     0    35     0
##    LW   LWB   RAM    RB   RCB   RCM   RDM    RF    RM    RS    RW   RWB    ST
##     4    32     0   630   553    35    43     0    40     0     1    38     8
```

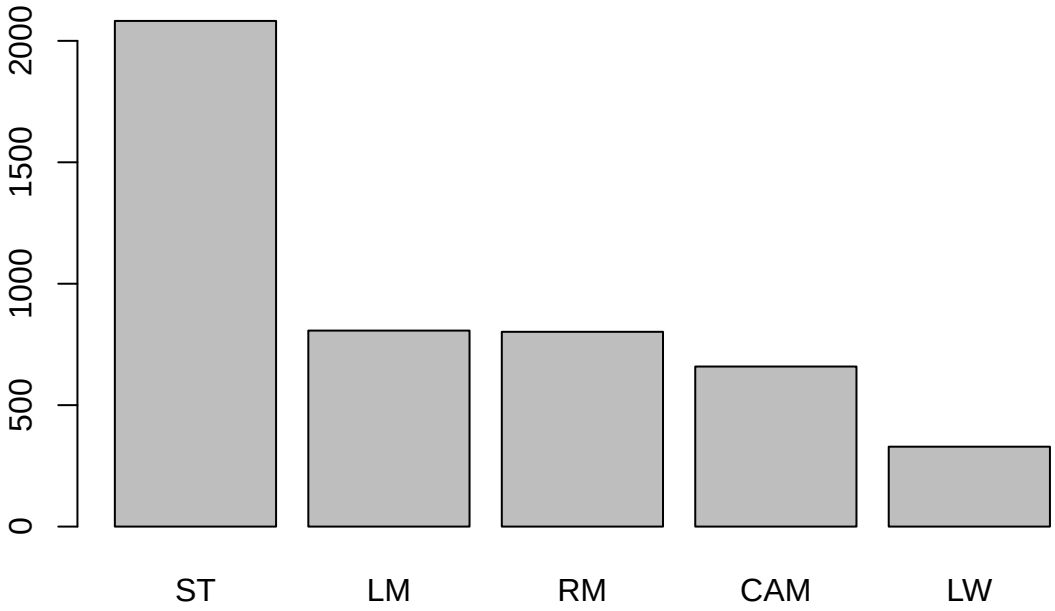
```
#barplot(table(p1))
barplot(sort(table(p2),decreasing = TRUE)[1:5])
```




```
barplot(sort(table(p3),decreasing = TRUE)[1:5])
```




```
barplot(sort(table(p4),decreasing = TRUE)[1:5])
```

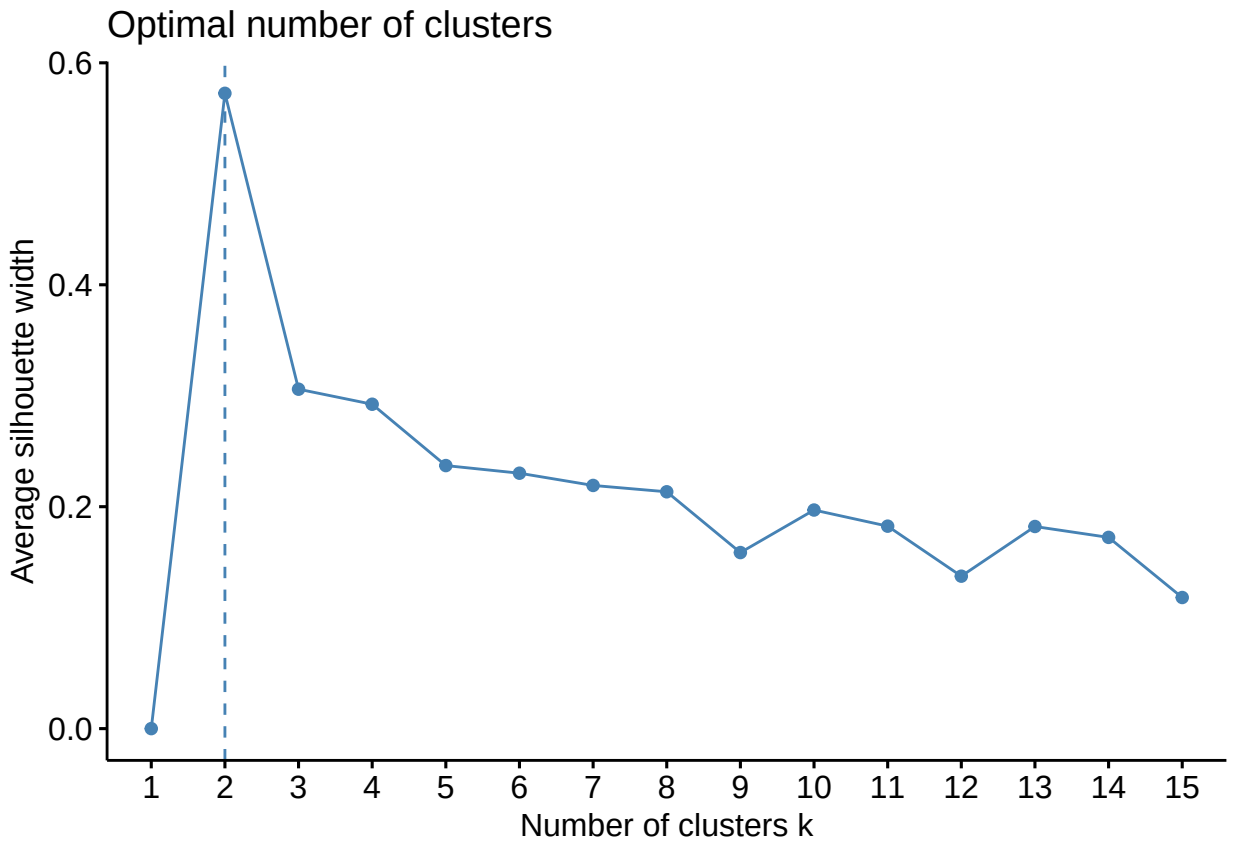



```
res_sil = vector()
for(k in 4:10){
  km.out=kmeans(X,k,nstart=15)
  ss = silhouette(km.out$cluster,dist(X))
  res_sil = c(res_sil,mean(ss[,3]))
  cat(k,mean(ss[,3]))
}
```

```
## 4 0.2923321
## Warning: Quick-TRANSfer stage steps exceeded maximum (= 907950)
## 5 0.2615736 0.2265755
## Warning: Quick-TRANSfer stage steps exceeded maximum (= 907950)
```

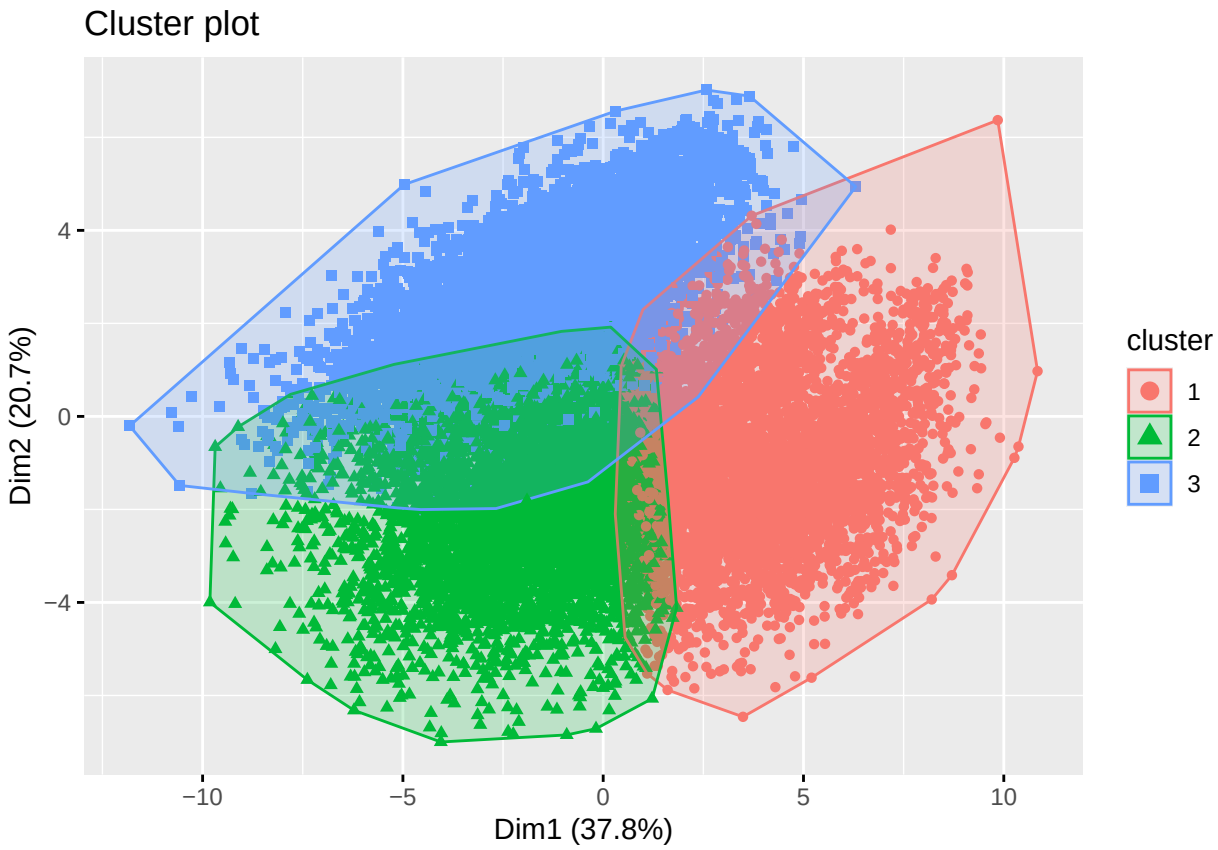
```
## 7 0.21914198 0.2036716
## Warning: did not converge in 10 iterations
## 9 0.201051
## Warning: did not converge in 10 iterations
## 10 0.1968792
```

```
fviz_nbclust(X, kmeans, method='silhouette',k.max = 15)
```



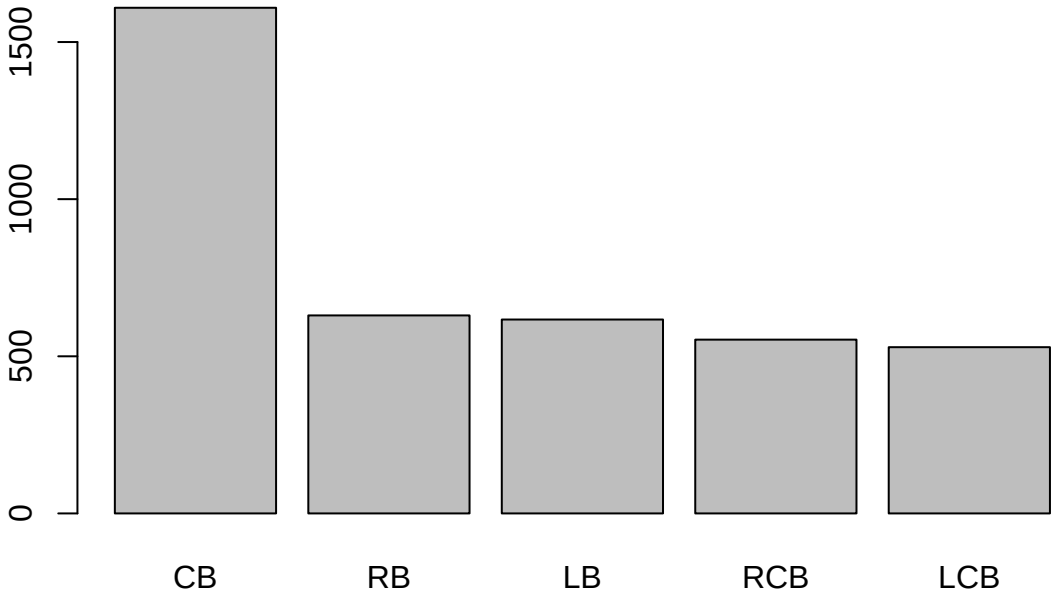
Ignore GK

```
dat = df[df$Position!="GK",]
dat = dat[,19:48]

X1=as.matrix(dat)
km.out=kmeans(X1,3,nstart=15)
fviz_cluster(km.out,data = X1,geom = 'point')
```

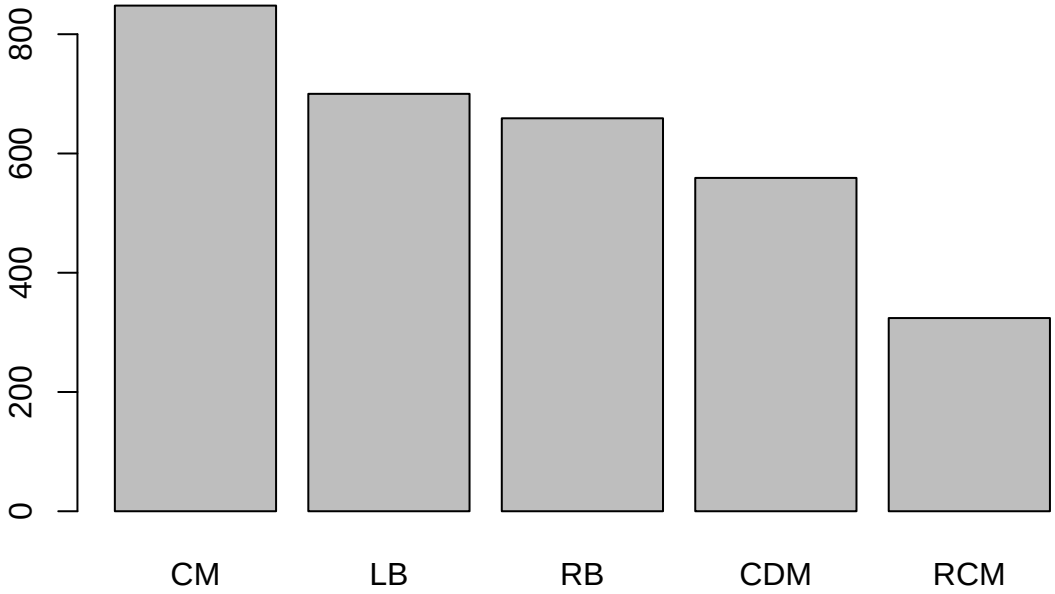



```
#fviz_nbclust(X1, kmeans, method='silhouette')
```

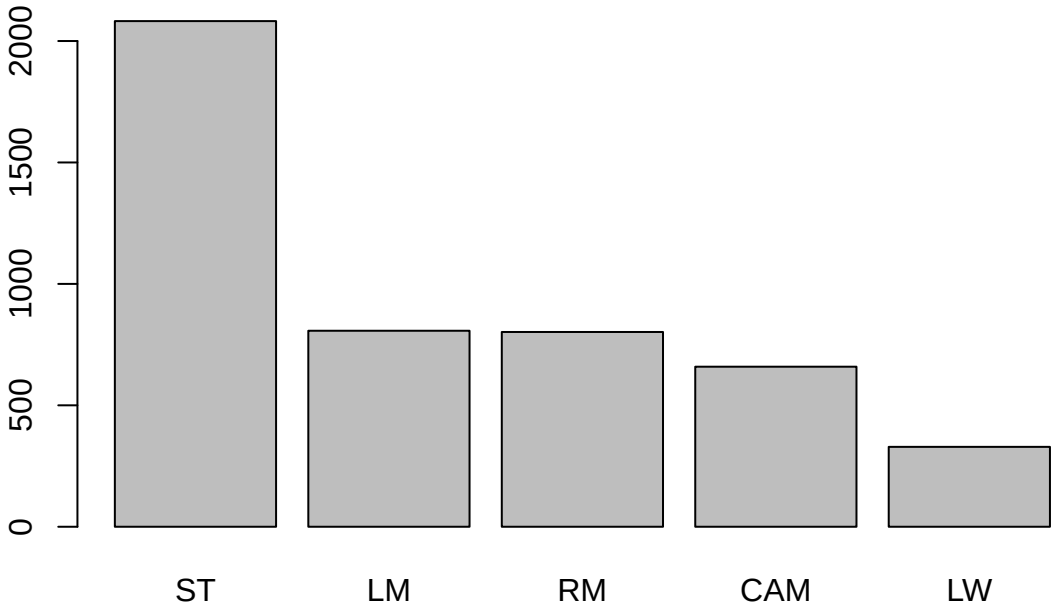
```
df.other = df[df$Position!="GK",]
results = km.out$cluster

cluster.1 =  which(results %in% c(1))
cluster.2 =  which(results %in% c(2))
cluster.3 =  which(results %in% c(3))
cluster.4 =  which(results %in% c(4))

p1 = df.other$Position[cluster.1]
p2 = df.other$Position[cluster.2]
p3 = df.other$Position[cluster.3]
#p4 = df.other$Position[cluster.4]
barplot(sort(table(p1),decreasing = TRUE)[1:5])
```




```
barplot(sort(table(p2),decreasing = TRUE)[1:5])
```




```
barplot(sort(table(p3),decreasing = TRUE)[1:5])
```




```
#table(p4)
```