

# Projet AD par Seif Eddine Mejri

*Mejri\_Seif\_eddine*

*20/06/2020*

## Chargement des packages à utiliser

```
library(readr)
library(FactoMineR)
library(factoextra)

## Loading required package: ggplot2

## Welcome! Want to learn more? See two factoextra-related books at https://goo.gl/ve3WBa
library(corrplot)

## corrplot 0.84 loaded

library(RColorBrewer)
library(questionr)
```

## Importation de la base et préparation des données

```
dataset<-read_delim("base.csv", ";", escape_double = FALSE,
                    trim_ws = TRUE)

## Parsed with column specification:
## cols(
##   .default = col_double(),
##   sexe = col_character(),
##   age = col_character(),
##   `situation professionnelle` = col_character(),
##   Francais = col_character(),
##   Anglais = col_character(),
##   Espagnol = col_character(),
##   Allemand = col_character(),
##   Italien = col_character(),
##   `equipe-tun` = col_character(),
##   nation = col_character(),
##   Joueurs = col_character(),
##   transmis_passion = col_character()
## )

## See spec(...) for full column specifications.

team<-dataset[dataset$LFC!=5 & dataset$MUFC!=5 & dataset$AFC!=5 &
              dataset$CFC!=5 & dataset$MCFC!=5 & dataset$FCB!=5 &
              dataset$RM!=5 & dataset$JFC!=5 & dataset$ACM!=5 &
              dataset$IM!=5 & dataset$BM!=5 & dataset$BVB!=5 &
              dataset$PSG!=5,
              c("sexe", "age", "situation professionnelle", "Anglais", "Espagnol",
                "Allemand", "Italien", "fan_since", "LFC", "MUFC", "AFC", "CFC", "MCFC", "FCB", "RM", "JFC", "ACM", "IM", "BM", "BVB", "PSG",
                "equipe-tun", "nation", "Joueurs", "attachement",
```

```
"enfance","joueur_prefere","gagne","histoire","plus_talentueuse","principe","spectacle",
"satisfaction resultat",
"MoneyBall","Gegenpressing","Tiki_Taka","Catinaccio","Grinta","Pragmatisme",
"transmis_passion"]]
```

```
reasons<-team[c("enfance","joueur_prefere","gagne","histoire",
               "plus_talentueuse","principe","spectacle")]
```

## Intérêt de l'ACP :

L'intérêt cette ACP est principalement d'investiguer les corrélations entre les différentes raisons pour supporter une équipe étrangère chez la Tunisie afin de produire des nouvelles variables qui peuvent caractériser les supporters d'une équipe. Dans cette partie je vais me limiter aux supporters de Liverpool FC, qui sont exactement 38 personnes.

## 2) Pertinence de l'ACP:

On va mesurer la pertinence de l'ACP via le corrélogramme et la matrice des corrélations.

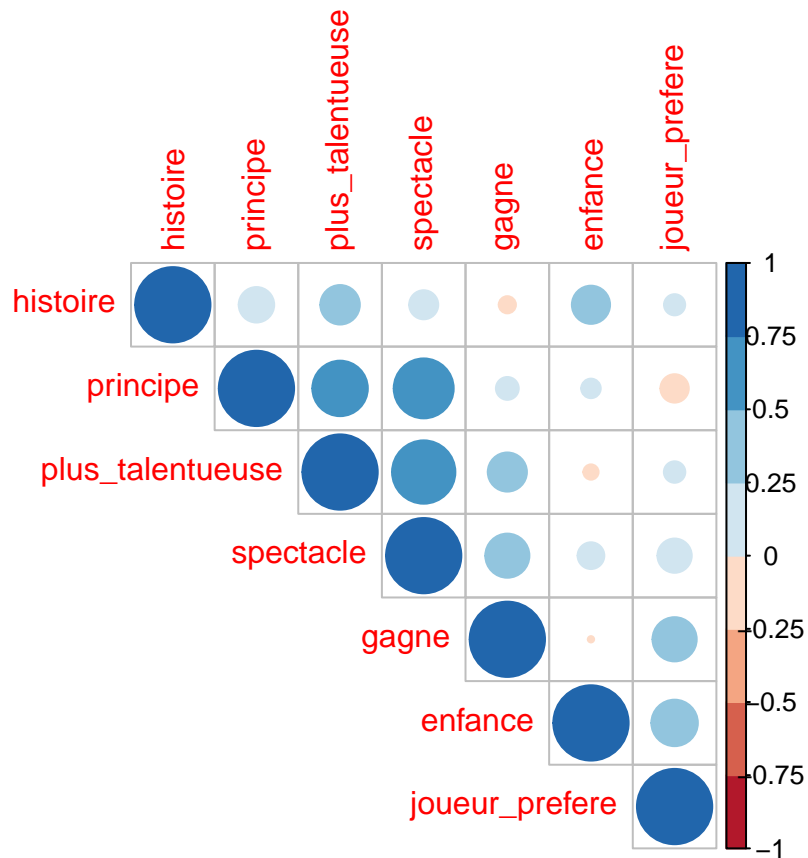
```
reasons_matrix<-as.matrix(team[c("enfance","joueur_prefere","gagne",
                                "histoire","plus_talentueuse","principe","spectacle")])
```

```
M2<-round(cor(reasons_matrix),3)
```

```
M2
```

```
##          enfance joueur_prefere  gagne histoire plus_talentueuse
## enfance          1.000          0.381 -0.007   0.257          -0.041
## joueur_prefere    0.381          1.000  0.345   0.079           0.081
## gagne            -0.007          0.345  1.000  -0.052           0.270
## histoire          0.257          0.079 -0.052  1.000           0.275
## plus_talentueuse -0.041          0.081  0.270  0.275           1.000
## principe          0.067         -0.142  0.092  0.222           0.543
## spectacle         0.126          0.207  0.343  0.149           0.707
##
##          principe spectacle
## enfance          0.067    0.126
## joueur_prefere   -0.142    0.207
## gagne            0.092    0.343
## histoire         0.222    0.149
## plus_talentueuse 0.543    0.707
## principe         1.000    0.634
## spectacle        0.634    1.000
```

```
corrplot(M2, type="upper", order="hclust",
         col=brewer.pal(n=8, name="RdBu"))
```



On peut observer l'existence de fortes corrélations entre les variables : « la plus talentueuse », « principe » et « spectacle ». ce qui signifie que cette ACP devrait être pertinente.

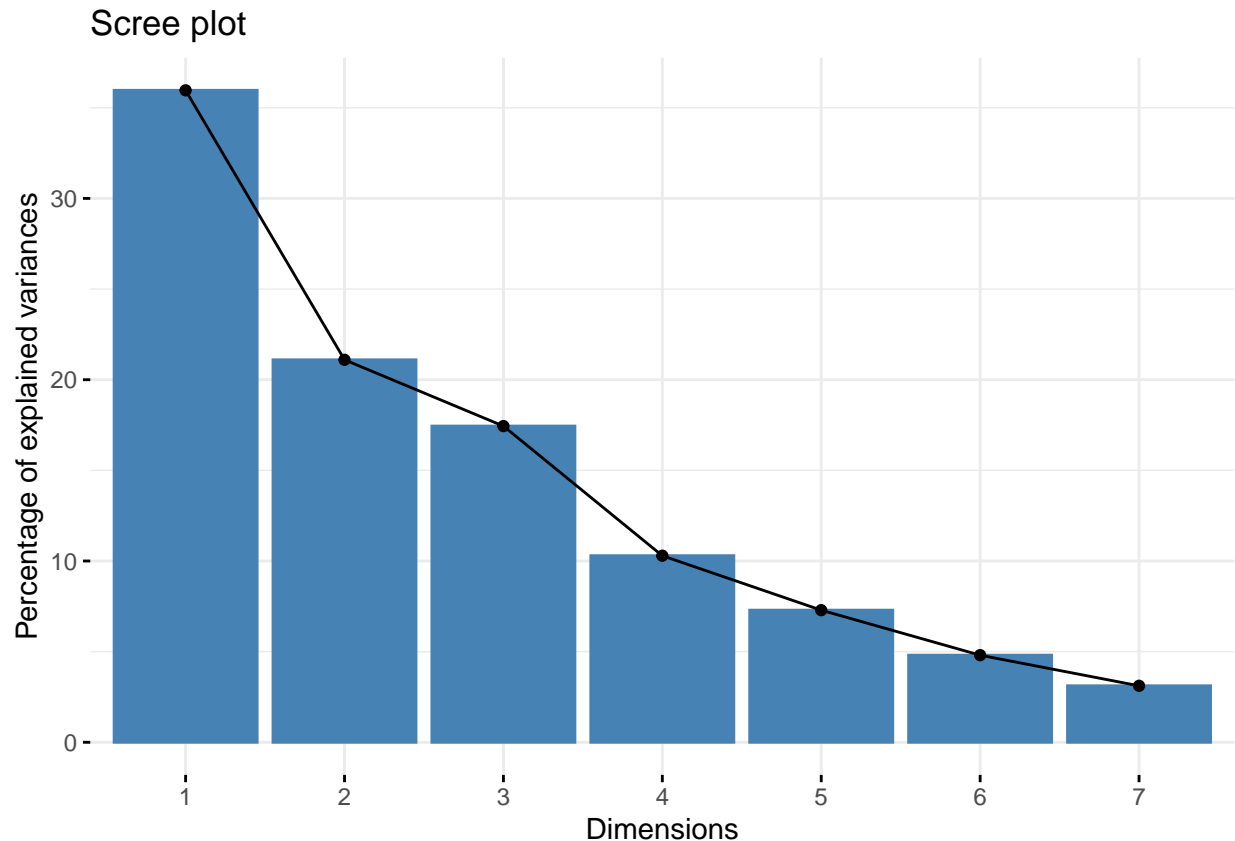
##3) Choix du nombre d'axes à retenir :

```
reasons_pca<-PCA(reasons,graph = F)
```

```
head(round(reasons_pca$eig,3))
```

```
##      eigenvalue percentage of variance cumulative percentage of variance
## comp 1      2.517                35.961                35.961
## comp 2      1.477                21.097                57.058
## comp 3      1.221                17.442                74.500
## comp 4      0.720                10.291                84.791
## comp 5      0.510                 7.287                92.079
## comp 6      0.336                 4.805                96.883
```

```
fviz_screplot(reasons_pca , ncp=10)
```



a) Critère de Kaiser: On remarque qu'il y a 3 axes dont les valeurs propres sont supérieures à 1 => On retient 3 axes d'après ce critère.

b) Critère du taux d'inertie cumulée: On remarque que le taux d'inertie cumulé des 2 premiers axes est de 57.06% qui est un taux important si on prend en considération le nombre des variables (7). => On va donc, d'après ce critère, retenir les 2 premiers axes.

c) Critère du coude: On remarque que le coude se trouve au niveau du deuxième axe, d'après ce critère. => On devrait retenir les 2 premiers axes.

==> Sur les 3 critères, 2 d'entre eux ont permis de retenir 2 axes, on va donc retenir 2 axes.

###4) Interprétation des axes :

```
round(reasons_pca$var$coord,3)
```

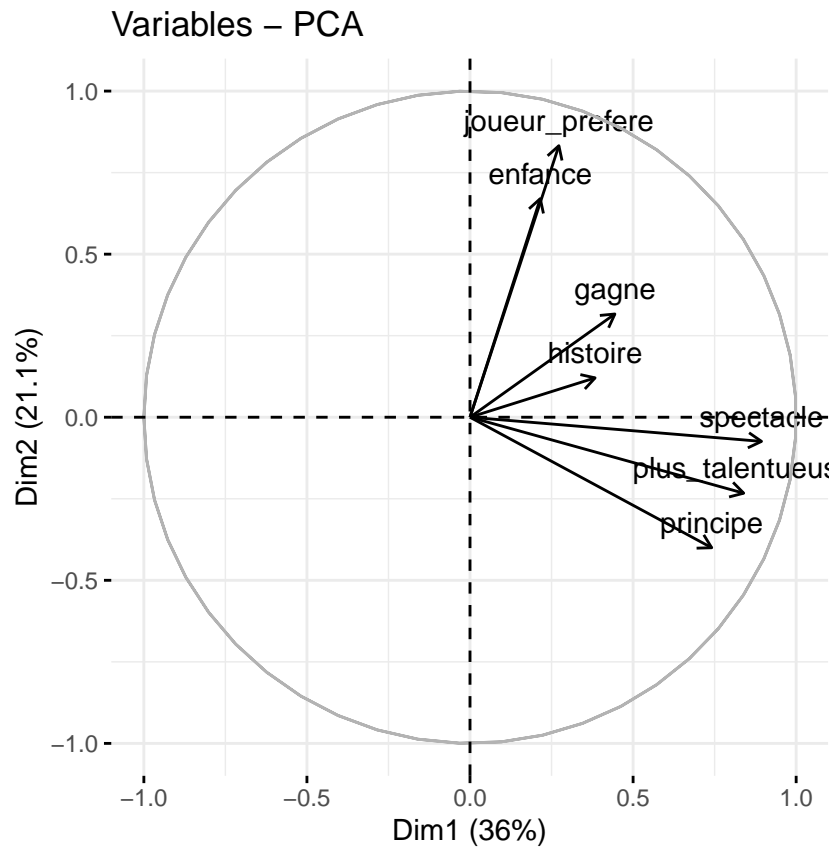
```
##          Dim.1 Dim.2 Dim.3 Dim.4 Dim.5
## enfance      0.215  0.670  0.490 -0.405  0.215
## joueur_prefere 0.272  0.833 -0.185  0.032 -0.352
## gagne        0.444  0.317 -0.655  0.251  0.456
## histoire     0.384  0.121  0.682  0.594  0.091
## plus_talentueuse 0.839 -0.233 -0.060  0.146 -0.266
## principe     0.741 -0.400  0.181 -0.282  0.195
## spectacle    0.894 -0.074 -0.123 -0.195 -0.122
```

```
round(reasons_pca$var$cos2,3)
```

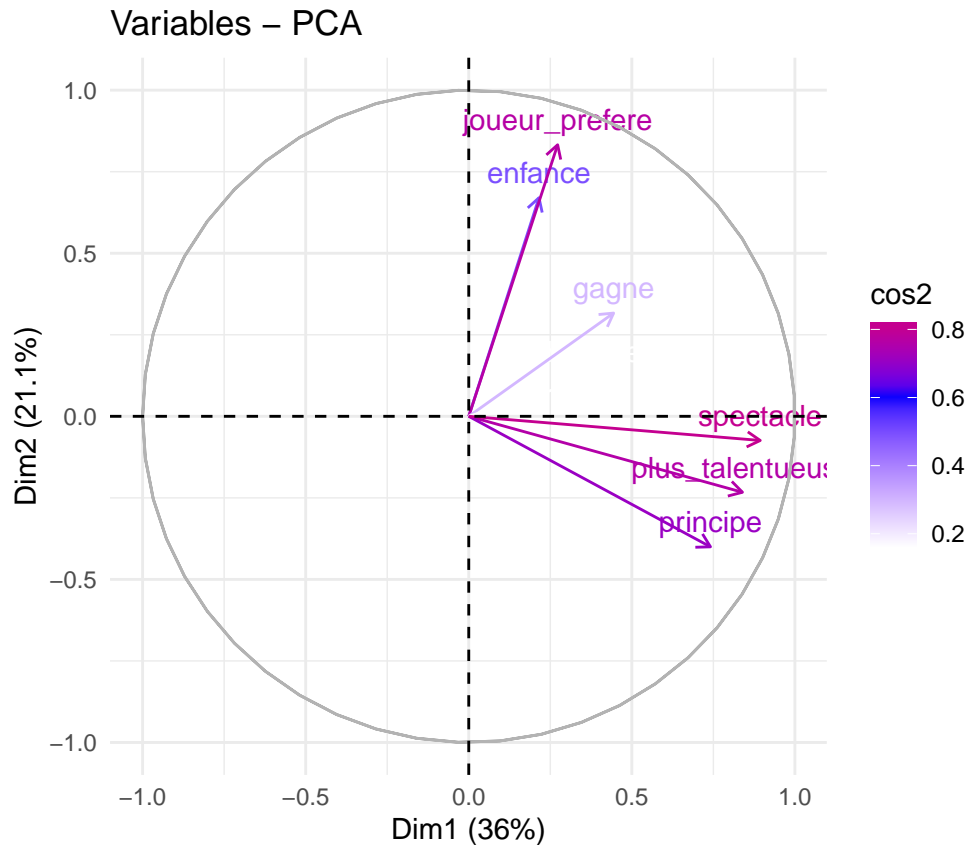
```
##          Dim.1 Dim.2 Dim.3 Dim.4 Dim.5
## enfance      0.046  0.449  0.240  0.164  0.046
## joueur_prefere 0.074  0.693  0.034  0.001  0.124
```

```
## gagne          0.197 0.100 0.429 0.063 0.208
## histoire       0.147 0.015 0.466 0.353 0.008
## plus_talentueuse 0.704 0.054 0.004 0.021 0.071
## principe       0.549 0.160 0.033 0.080 0.038
## spectacle      0.799 0.005 0.015 0.038 0.015
```

```
fviz_pca_var(reasons_pca)
```



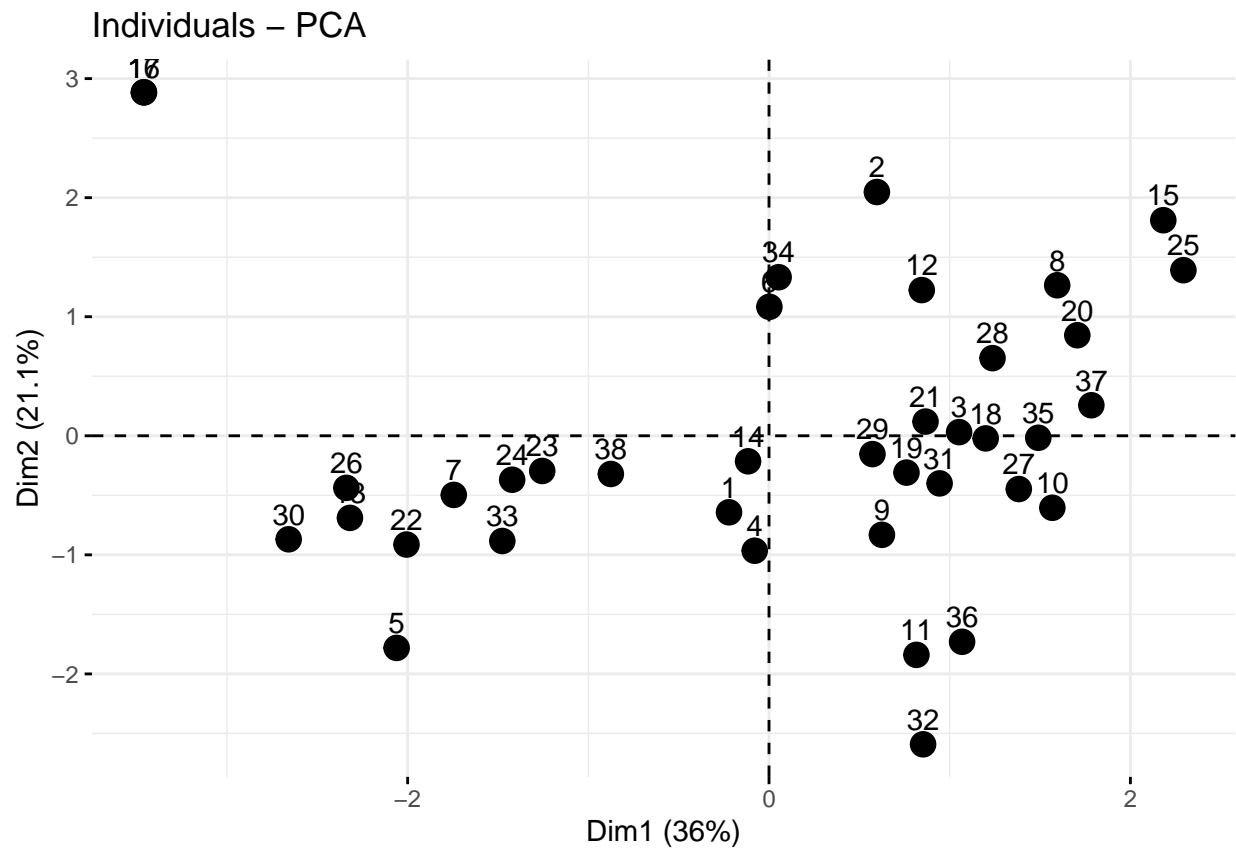
```
fviz_pca_var(reasons_pca, col.var="cos2") +
  scale_color_gradient2(low="white", mid="blue",
                        high="red", midpoint=0.6) +
  theme_minimal()
```



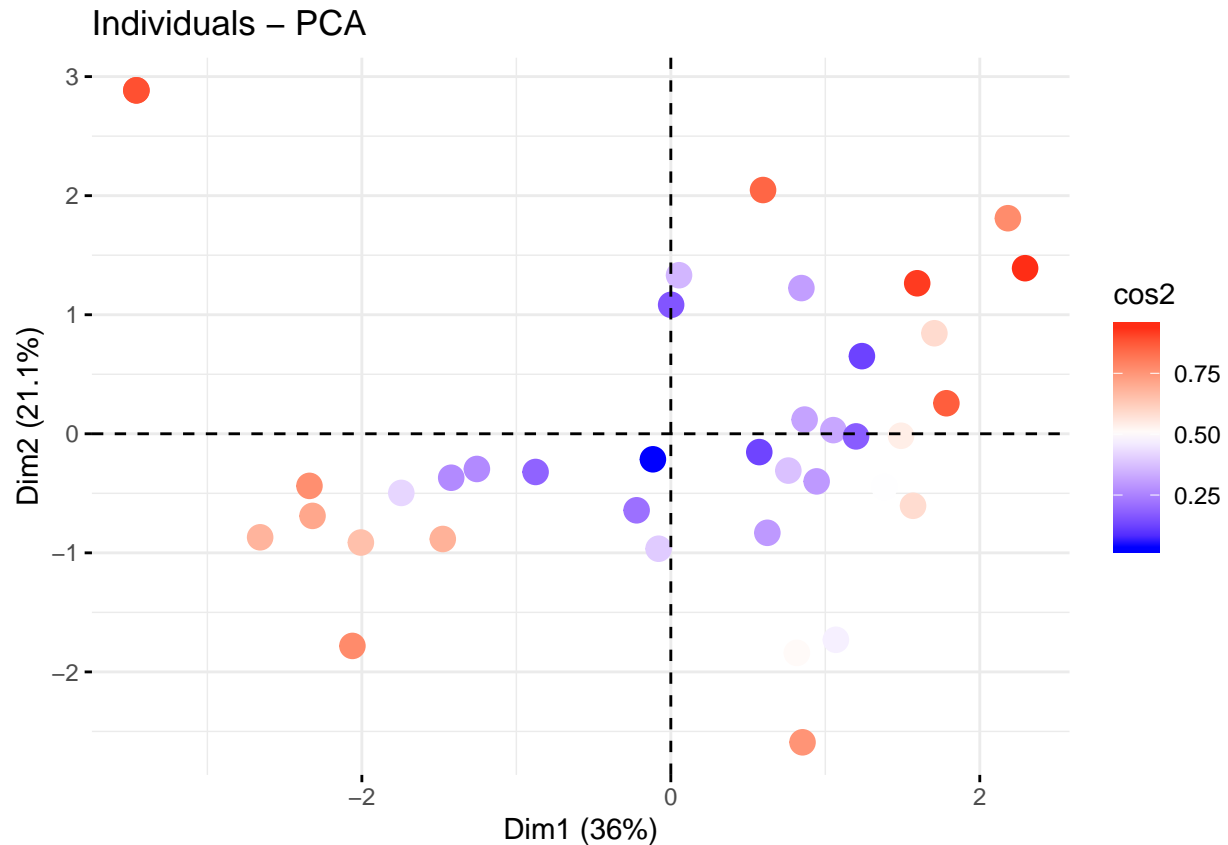
#Axe1 : Cet axe est fortement corrélé aux variables « l'équipe la plus talentueuse », « l'équipe qui présente mes principes footballistiques » et « l'équipe qui propose le meilleur spectacle ». Cet axe regroupe les raisons liées au style de jeux de l'équipe (c'est un choix rationnel pris dans un âge d'adolescence généralement). Soit l'axe du Goût footballistique ou des raisons rationnelles. #Axe 2 : Cet axe est fortement corrélé aux variables « l'équipe de l'enfance » et « l'équipe du joueur préféré ». Cet axe regroupe les raisons liées à la personnification de l'équipe en un Joueur. La raison qu'on trouve « équipe de l'enfance » c'est que le choix de l'équipe préférée est décidé par le club où le joueur préféré évolue. (C'est un choix en quelque sorte irrationnel pris dans l'âge de la préadolescence). D'où on peut le qualifier comme l'axe du Choix d'enfance ou des Raisons irrationnelles.

###5)Interprétation de la carte des individus :

```
fviz_pca_ind(reasons_pca, pointsize = 4 , col.ind.sup = 'blue')
```



```
fviz_pca_ind(reasons_pca,geom = "point",pointsize = 4 , col.ind="cos2")+
  scale_color_gradient2(low="blue", mid="white",
    high="red", midpoint=0.5)
```



On peut identifier 2 lots de supporters tunisiens du Liverpool: #Le 1er lot se distingue à gauche du 1er axe en bleu, regroupant les supporters les plus anciens (ont commencé de supporter Liverpool en moyenne depuis 2001,7) et les plus fidèle de l'équipe qui ont survécu les moments difficiles de l'équipe. C'est un lot qui n'accorde pas d'importance au spectacle footballistique présenté par le club(ils supporteront l'équipe quoi qu'il arrive). Ils sont plus ou moins indépendant du 2ème axe. #Le 2ème lot se distingue à droite du 1er axe en jaune, regroupant les nouveaux supporters et les enthousiastes du club, ce lot est largement influencé par style de jeux très impressionnant proposé par Liverpool dernièrement. (Ils ont commencé de supporter Liverpool en moyenne depuis 2005.2)

### Quelque chose de spécial -carte des individus 3d-

```
library(pca3d)

pca <- prcomp(reasons, scale.=TRUE)
pca3d(pca)

## [1] 0.06827592 0.05693111 0.07028234
## Creating new device
```

AFC sur le lien entre le joueur préféré et l'équipe nationale préférée:

### Préparation des données

```
tab3<-table(dataset$nation,dataset$Joueurs)

lprop(tab3)
```



```
##
##      autre Cristiano Ronaldo Eden Hazard Kylian Mbappe Lionel Messi
##  Allemagne 22.3 10.6          0.0          0.0          13.8
##  Angleterre 32.5 10.0          10.0          0.0          12.5
##  Argentine  0.0  0.0          0.0          0.0          98.6
##  Autre      29.0 12.9          9.7          0.0          19.4
##  Bresil     11.5 23.1          0.0          0.0          42.3
##  Espagne    9.6 11.5          1.9          0.0          71.2
##  France     13.3 6.7          0.0          26.7          33.3
##  Italie     48.4 4.7          1.6          0.0          20.3
##  Pays-bas   21.4 14.3          0.0          0.0          32.1
##  Ensemble   21.5 9.1          2.1          1.0          39.9
##
##      Luis Suarez Luka Modric Mohamed Salah Mourad el Hedhli Neymar
##  Allemagne  0.0          1.1          1.1          1.1          0.0
##  Angleterre  2.5          0.0          0.0          0.0          2.5
##  Argentine   0.0          0.0          0.0          0.0          0.0
##  Autre       0.0          0.0          3.2          3.2          0.0
##  Bresil      0.0          0.0          0.0          0.0          15.4
##  Espagne     0.0          1.9          1.9          0.0          0.0
##  France      0.0          0.0          0.0          0.0          6.7
##  Italie      0.0          4.7          1.6          1.6          4.7
##  Pays-bas    3.6          0.0          3.6          0.0          0.0
##  Ensemble    0.5          1.2          1.2          0.7          2.1
##
##      Paul Pogba Robert Lewandowski Sadio Mane Sergio Ramos
##  Allemagne  0.0          40.4          5.3          3.2
##  Angleterre  2.5          0.0          17.5          0.0
##  Argentine   0.0          0.0          0.0          0.0
##  Autre       0.0          9.7          3.2          3.2
##  Bresil      0.0          0.0          0.0          3.8
##  Espagne     0.0          0.0          0.0          1.9
##  France      0.0          0.0          6.7          0.0
##  Italie      0.0          0.0          4.7          3.1
##  Pays-bas    0.0          7.1          3.6          7.1
##  Ensemble    0.2          10.3          4.3          2.4
##
##      Virgil Van Dijk Total
##  Allemagne  1.1          100.0
##  Angleterre 10.0          100.0
##  Argentine   1.4          100.0
##  Autre       6.5          100.0
##  Bresil      3.8          100.0
##  Espagne     0.0          100.0
##  France      6.7          100.0
##  Italie      4.7          100.0
##  Pays-bas    7.1          100.0
##  Ensemble    3.6          100.0
```

```
cprop(tab3)
```

```
##
##      autre Cristiano Ronaldo Eden Hazard Kylian Mbappe Lionel Messi
##  Allemagne 23.3 26.3          0.0          0.0          7.8
##  Angleterre 14.4 10.5          44.4          0.0          3.0
```

```
##      Argentine      0.0      0.0      0.0      0.0      40.7
##      Autre      10.0     10.5     33.3      0.0      3.6
##      Bresil      3.3     15.8      0.0      0.0      6.6
##      Espagne      5.6     15.8     11.1      0.0     22.2
##      France      2.2      2.6      0.0     100.0      3.0
##      Italie     34.4      7.9     11.1      0.0      7.8
##      Pays-bas      6.7     10.5      0.0      0.0      5.4
##      Total     100.0    100.0     100.0     100.0     100.0
##
##      Luis Suarez Luka Modric Mohamed Salah Mourad el Hedhli Neymar
##      Allemagne      0.0      20.0      20.0      33.3      0.0
##      Angleterre    50.0      0.0      0.0      0.0     11.1
##      Argentine      0.0      0.0      0.0      0.0      0.0
##      Autre      0.0      0.0     20.0     33.3      0.0
##      Bresil      0.0      0.0      0.0      0.0     44.4
##      Espagne      0.0     20.0     20.0      0.0      0.0
##      France      0.0      0.0      0.0      0.0     11.1
##      Italie      0.0     60.0     20.0     33.3     33.3
##      Pays-bas     50.0      0.0     20.0      0.0      0.0
##      Total     100.0     100.0     100.0     100.0     100.0
##
##      Paul Pogba Robert Lewandowski Sadio Mane Sergio Ramos
##      Allemagne      0.0     88.4      27.8     30.0
##      Angleterre    100.0      0.0     38.9      0.0
##      Argentine      0.0      0.0      0.0      0.0
##      Autre      0.0      7.0      5.6     10.0
##      Bresil      0.0      0.0      0.0     10.0
##      Espagne      0.0      0.0      0.0     10.0
##      France      0.0      0.0      5.6      0.0
##      Italie      0.0      0.0     16.7     20.0
##      Pays-bas      0.0      4.7      5.6     20.0
##      Total     100.0     100.0     100.0     100.0
##
##      Virgil Van Dijk Ensemble
##      Allemagne      6.7      22.4
##      Angleterre     26.7      9.5
##      Argentine      6.7     16.5
##      Autre     13.3      7.4
##      Bresil      6.7      6.2
##      Espagne      0.0     12.4
##      France      6.7      3.6
##      Italie     20.0     15.3
##      Pays-bas     13.3      6.7
##      Total     100.0     100.0
```

###1)Intérêt de l'AFC : L'AFC dans cette partie va croiser les deux variables qualitatives : "Joueurs Préféré" et "équipe nationale préférée" possédant respectivement 15 modalités et 9 modalités. NB : l'AFC va être réalisé sur la totalité de l'échantillon(419 individus) et pas comme l'ACP réalisée seulement sur les supporters de Liverpool FC.

###2)Pertinence de l'AFC (test de chi-2):

```
test<-chisq.test(tab3)
```

```
## Warning in chisq.test(tab3): Chi-squared approximation may be incorrect
```

```
test
```

```
##  
## Pearson's Chi-squared test  
##  
## data:  tab3  
## X-squared = 519.77, df = 112, p-value < 2.2e-16
```

- La valeur du chi-2 est de 519.77
- Le degré de liberté du chi-2 est de 112
- La p-value (= p-value < 2.2e-16) étant presque nulle (elle est de l'ordre de 10 puissance -16), on peut donc rejeter l'hypothèse d'indépendance.

==>l'AFC est donc pertinente.

## Exécution de l'AFC

```
afc_gen<-CA(tab3,graph=F)
```

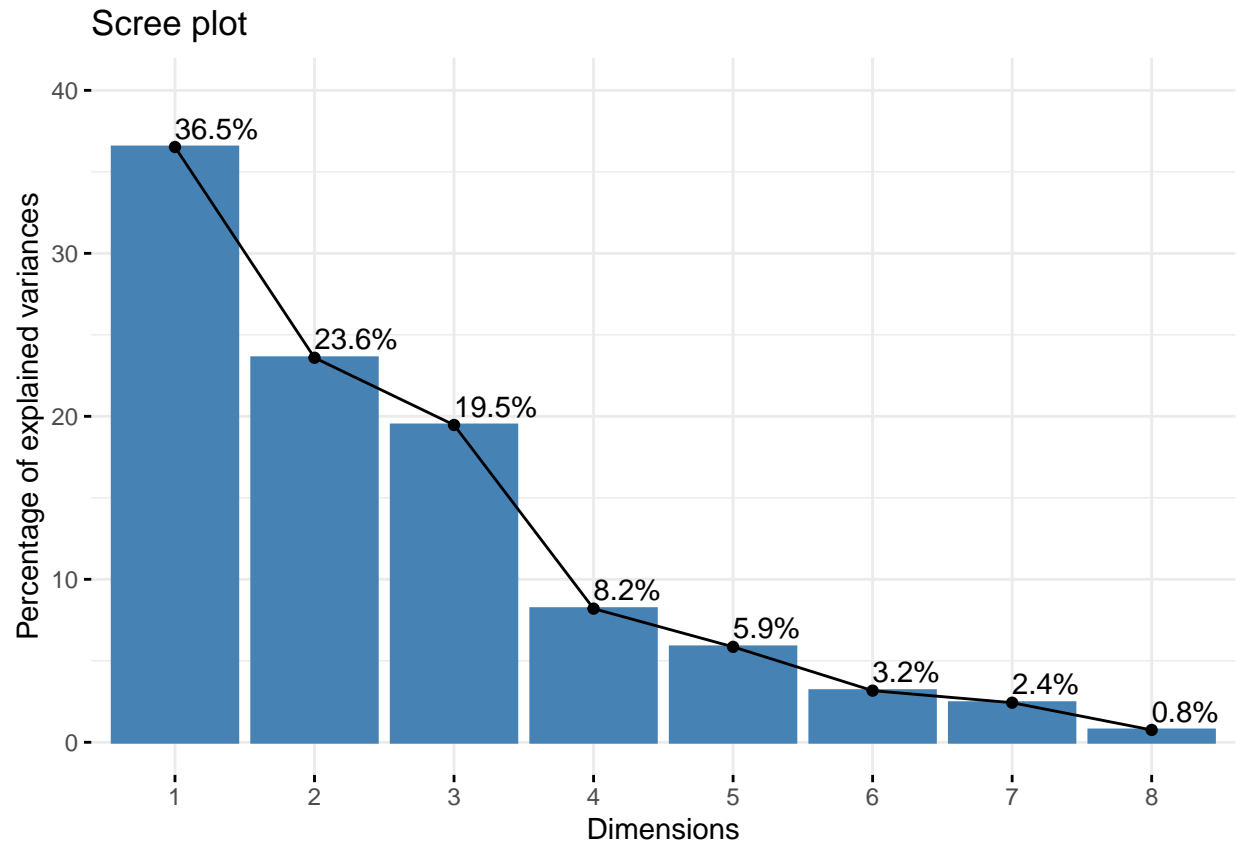
```
###3)Choix du nombre d'axe:
```

```
eig_val_gen <- afc_gen$eig
```

```
round(afc_gen$eig,3)
```

```
##      eigenvalue percentage of variance cumulative percentage of variance  
## dim 1      0.453             36.523             36.523  
## dim 2      0.293             23.593             60.116  
## dim 3      0.242             19.470             79.586  
## dim 4      0.102              8.201             87.787  
## dim 5      0.073              5.856             93.642  
## dim 6      0.039              3.170             96.812  
## dim 7      0.030              2.430             99.242  
## dim 8      0.009              0.758            100.000
```

```
fviz_screplot (afc_gen, addlabels = TRUE, ylim = c(0, 40))
```



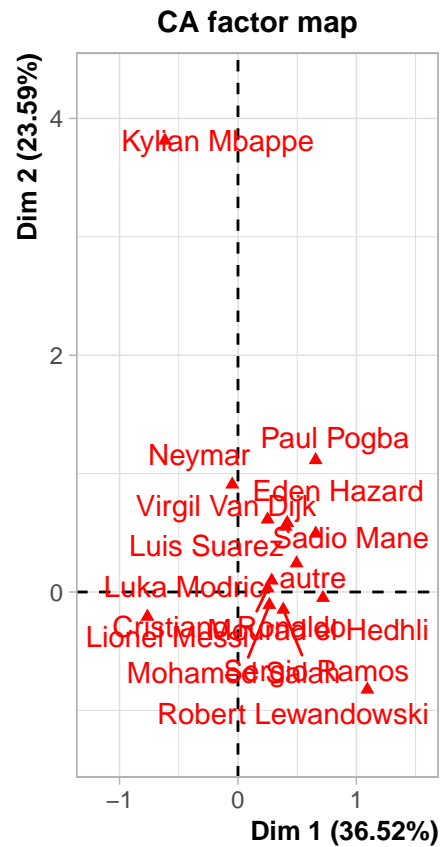
Le taux d'inertie cumulée du premier plan étant de 60.12%. il est possible de se contenter des 2 premiers axes.

###4) Représentation et interprétation de la carte des profils-colonnes (les joueurs):

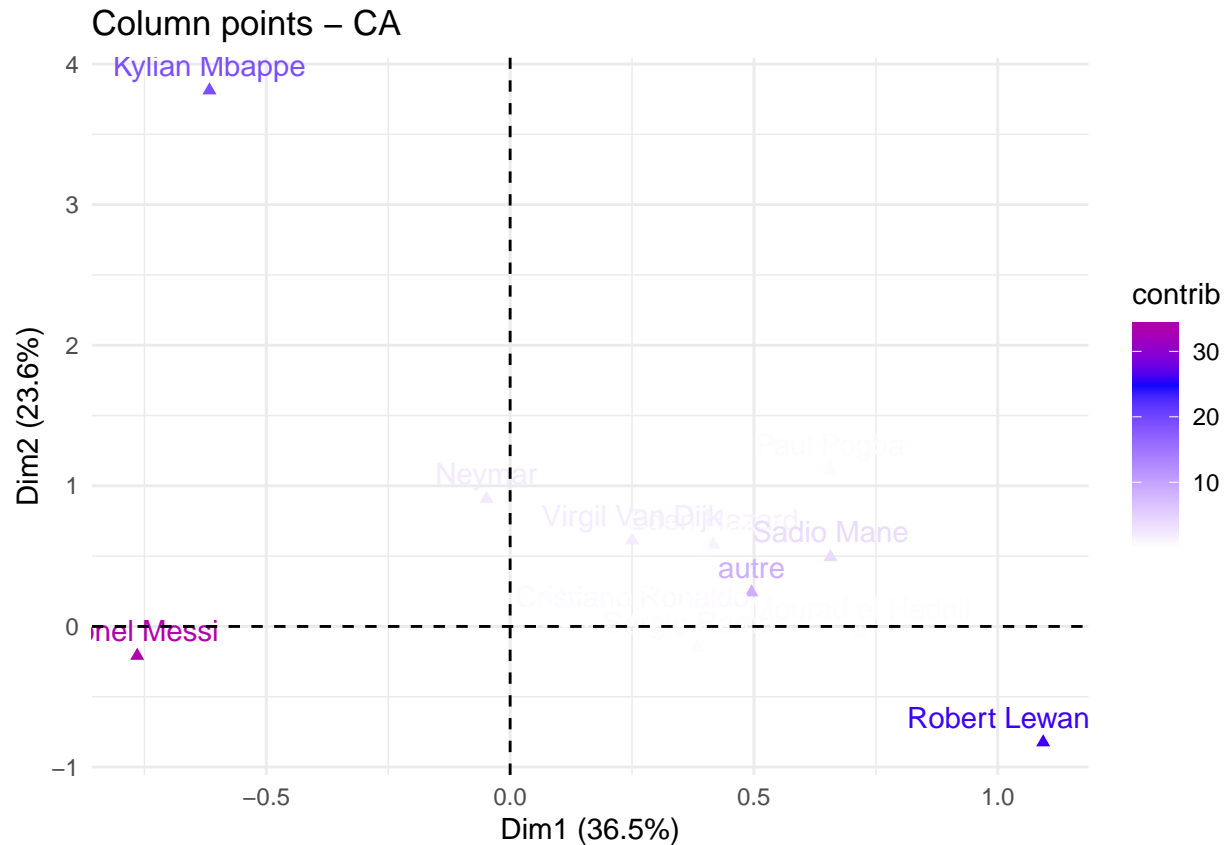
```
round(cbind(afc_gen$col$coord[,1:2], afc_gen$col$contrib[,1:2], afc_gen$col$cos2[,1:2]), 2)
```

	Dim 1	Dim 2	Dim 1	Dim 2	Dim 1	Dim 2
## autre	0.50	0.24	11.62	4.31	0.49	0.12
## Cristiano Ronaldo	0.25	0.03	1.26	0.03	0.16	0.00
## Eden Hazard	0.42	0.58	0.82	2.50	0.06	0.12
## Kylian Mbappe	-0.62	3.81	0.80	47.43	0.01	0.54
## Lionel Messi	-0.76	-0.21	51.45	5.92	0.92	0.07
## Luis Suarez	0.40	0.55	0.17	0.50	0.03	0.06
## Luka Modric	0.29	0.10	0.21	0.04	0.04	0.01
## Mohamed Salah	0.27	-0.11	0.19	0.05	0.08	0.01
## Mourad el Hedhli	0.72	-0.05	0.81	0.01	0.30	0.00
## Neymar	-0.05	0.91	0.01	6.05	0.00	0.24
## Paul Pogba	0.66	1.11	0.23	1.01	0.05	0.13
## Robert Lewandowski	1.09	-0.82	27.07	23.81	0.46	0.26
## Sadio Mane	0.66	0.49	4.09	3.57	0.34	0.19
## Sergio Ramos	0.38	-0.15	0.77	0.18	0.23	0.03
## Virgil Van Dijk	0.25	0.61	0.49	4.59	0.08	0.50

```
plot(afc_gen, invisible="row")
```



```
fviz_ca_col(afc_gen, col.col = "contrib")+
  scale_color_gradient2(low = "white", mid = "blue",
                        high = "red", midpoint = 25) +
  theme_minimal()
```



#Le premier axe oppose les joueurs : Lionel Messi (contrib=51.45%) et Robert Lewandowski (contrib=27.07%).

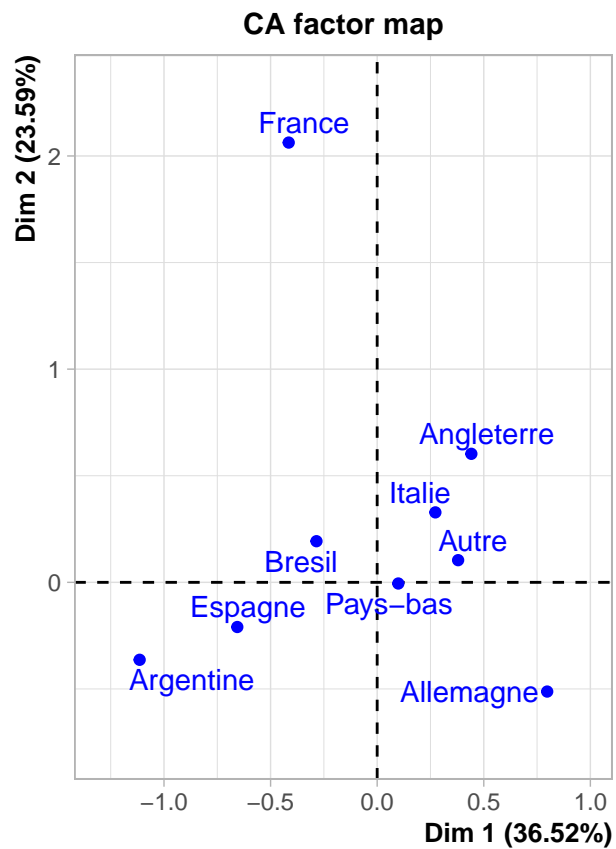
#Le deuxième axe est essentiellement expliqué par Kyllian Mbappé (contrib=47.43%) et dans une moindre mesure par Robert Lewandowski (contrib=23.81%).

###5) Représentation et interprétation de la carte des profils-lignes (les équipes nationales):

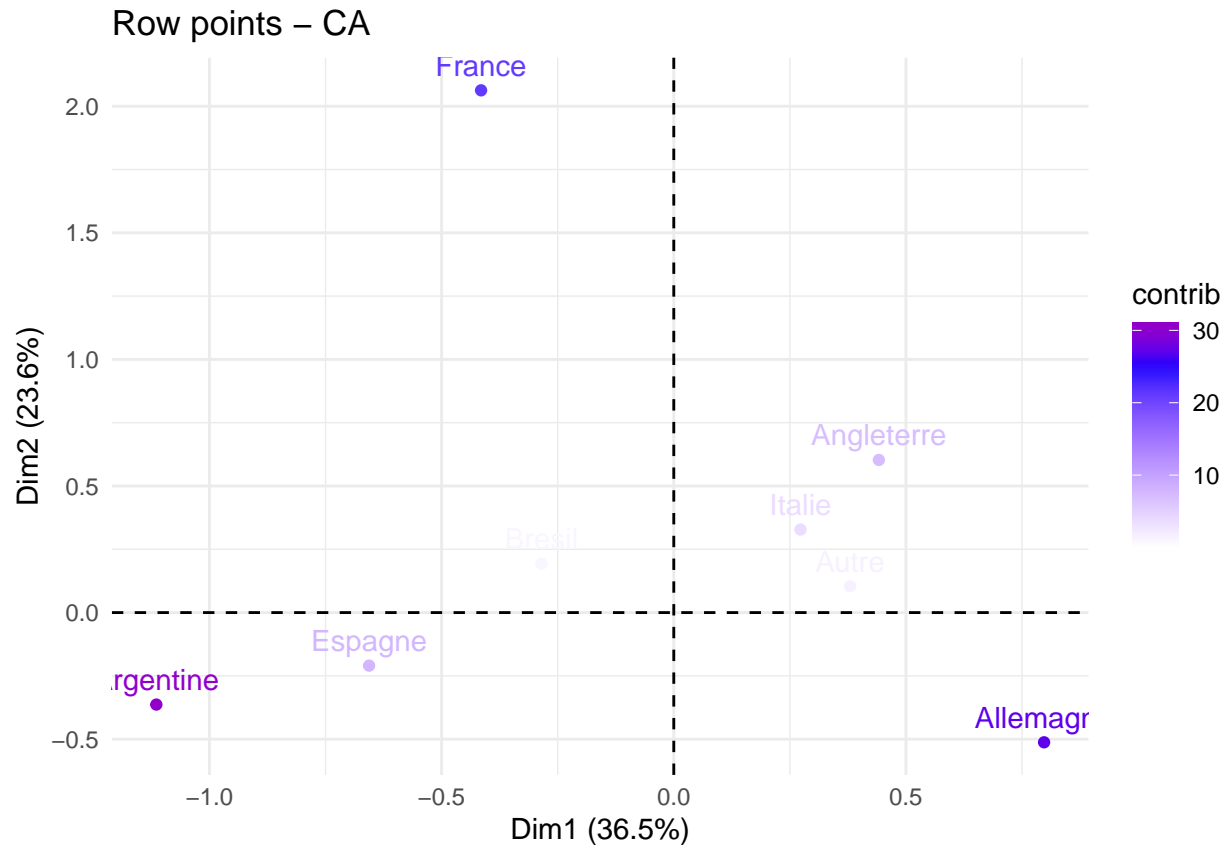
```
round(cbind(afc_gen$row$coord[,1:2], afc_gen$row$contrib[,1:2], afc_gen$row$cos2[,1:2]),2)
```

```
##          Dim 1 Dim 2 Dim 1 Dim 2 Dim 1 Dim 2
## Allemagne  0.80 -0.51 31.49 20.12  0.56  0.23
## Angleterre  0.44  0.60  4.11 11.85  0.13  0.24
## Argentine -1.11 -0.36 45.13  7.43  0.86  0.09
## Autre      0.38  0.10  2.35  0.27  0.23  0.02
## Bresil     -0.28  0.19  1.11  0.79  0.06  0.03
## Espagne    -0.66 -0.21 11.79  1.86  0.78  0.08
## France     -0.41  2.06  1.36 52.05  0.02  0.58
## Italie      0.27  0.33  2.51  5.62  0.10  0.15
## Pays-bas   0.10 -0.01  0.14  0.00  0.02  0.00
```

```
plot(afc_gen, invisible="col")
```



```
fviz_ca_row(afc_gen, col.row = "contrib") +  
  scale_color_gradient2(low = "white", mid = "blue",  
    high = "red", midpoint = 25) +  
  theme_minimal()
```

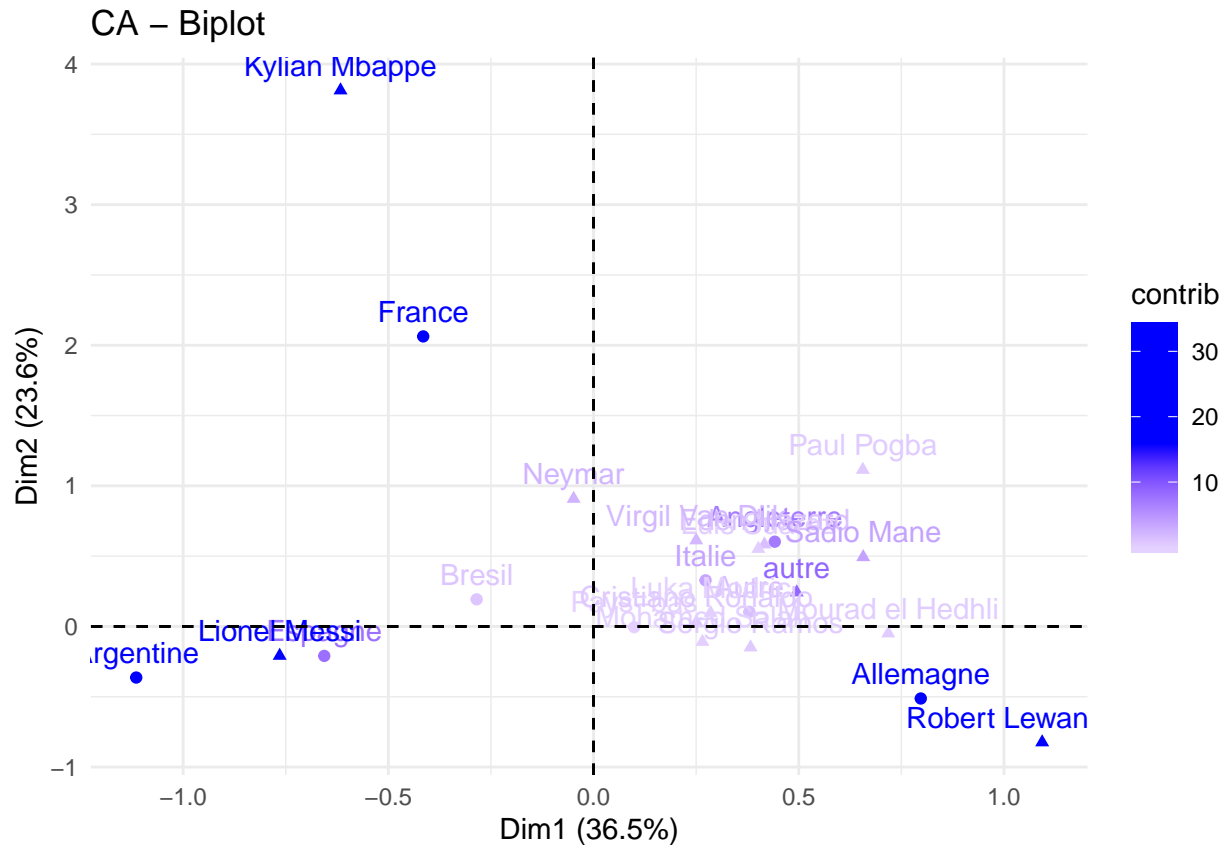


#Le premier axe: oppose aux équipes nationales: Argentine (contrib=45.13%), Allemagne (contrib=31.49%) et faiblement Espagne (contrib=11.79%) #Le deuxième axe: est expliqué par essentiellement par la France (contrib=52.05%) avec des faibles contributions par l'Allemagne (contrib=20.12%) et L'Angleterre (contrib=11.85%)

###6) Carte de représentation simultanée:

```
fviz_ca_biplot(afc_gen,col.col = "contrib" ,col.row = "contrib") + scale_color_gradient2(low = "white",
high = "blue", midpoint = 15) +
theme_minimal()
```





La carte de représentation simultanée montre des correspondances entre: Lionel Messi et Argentine, Kylian Mbappé et la France, Robert Lewandowski et l'Allemagne.

\*\*\*Lionel Messi et Argentine : c'est la correspondance la plus prévisible comme le montre les profils-colonnes où 98,6% des supporters de l'Argentine trouve Lionel Messi comme joueur préféré.

\*\*\*Kylian Mbappé et la France contrairement à la première correspondance cela était un résultat imprévisible à cause de la présence d'un autre joueur français dans le questionnaire : Paul Pogba. Avec 26.7% des supporters de la France préfèrent Kylian Mbappé.

\*\*\*Robert Lewandowski et L'Allemagne avec 40.4% des supporters de «die Mannschaft» préfère Robert Lewandowski cela peut être expliqué par le nombre important des réponses reçues de la part des supporters du Bayern München qui préfèrent leur Butteur Polonais.