

Pourquoi faire du sport ?

Skander Kolsi

2024-03-22

Introduction

Ce projet a pour but d'étudier les différents avis des tunisiens sur les bienfaits généraux de l'activité physique. Les réponses ont été enregistrés au cours de la periode du 05/03/2024 jusqu'à 31/03/2024. Ce projet est réalisé par Kolsi Skander et supervisé par Mr.Ghazi BelMufti.

Chargement des paquets

```
library(readxl)
```

```
## Warning: package 'readxl' was built under R version 4.3.3
```

```
library(dplyr)
```

```
##
```

```
## Attaching package: 'dplyr'
```

```
## The following objects are masked from 'package:stats':
```

```
##
```

```
##      filter, lag
```

```
## The following objects are masked from 'package:base':
```

```
##
```

```
##      intersect, setdiff, setequal, union
```

```
library(FactoMineR)
```

```
library(ggplot2)
```

```
library(factoextra)
```

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

```
library(corrplot)
```

```
## corrplot 0.92 loaded
```

```
library(questionr)
```

```
## Warning: package 'questionr' was built under R version 4.3.3
```

Importation de la base

```
df <- read_excel("C:\\Users\\skouz\\OneDrive\\Desktop\\ACP\\Projet.xlsx")
#Pour faciliter la lecture
colnames(df)=c('sexe','age','categorie','mental','stress','confiance','mdv','sante','physique','poids',
```

```
head(df)
```

```
## # A tibble: 6 x 31
##   sexe age      categorie mental stress confiance mdv sante physique poids
##   <chr> <chr>      <chr>      <dbl> <dbl>      <dbl> <dbl> <dbl>      <dbl> <dbl>
## 1 Femme 20 - 22 ans Etudiant~ 4      4      4      4      4      1      5
## 2 Femme 20 - 22 ans Etudiant~ 4      4      3      4      4      4      5
## 3 Homme 20 - 22 ans Etudiant~ 5      2      4      5      5      5      4
## 4 Femme 20 - 22 ans Etudiant~ 3      4      1      5      4      2      1
## 5 Femme 20 - 22 ans Etudiant~ 3      3      2      3      4      1      1
## 6 Femme 20 - 22 ans Etudiant~ 4      4      4      5      2      3      5
## # i 21 more variables: competition <dbl>, connaissances <dbl>, plaisir <dbl>,
## #   nature <dbl>, musique <dbl>, smarphone <dbl>, amis <dbl>, manger <dbl>,
## #   doucement <dbl>, sechauffer <dbl>, epuisement <dbl>, dormir <dbl>,
## #   eau <dbl>, etirement <dbl>, regulierement <chr>, coach <chr>,
## #   important <chr>, programme <chr>, blessure <chr>, objectif <chr>,
## #   continuer <chr>
```

La base de données est construite à partir d'un questionnaire destiné à des individus ayant des différentes catégories socio-professionnelles et plusieurs classes d'âge. Les données contiennent principalement les bienfaits de l'activités sportives dans un premier lieu et les habitudes des tunisiens lors d'une séance de sport.

I/ Statistique descriptive

1.Repartition selon les pratiquants et les non pratiquants :

```
rp <- df$regulierement
OUI = sum(df$rp == "Oui")
```

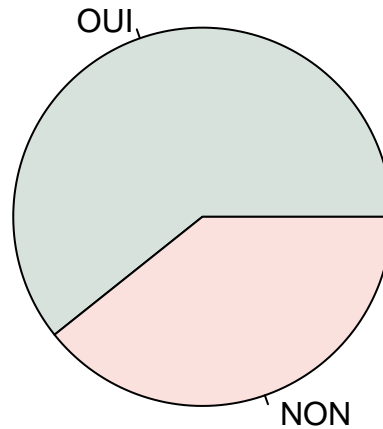
```
## Warning: Unknown or uninitialised column: 'rp'.
```

```
NON = sum(df$rp == "Non")
```

```
## Warning: Unknown or uninitialised column: 'rp'.
```

```
lbls2 <- c("OUI", "NON")
df1 <- data.frame(rp) %>% group_by(rp) %>% count(rp)
pie(table(rp), labels = lbls2, col=c("#d8e2dc", "#fae1dd"), main="Répartition selon les pratiquants et les
```

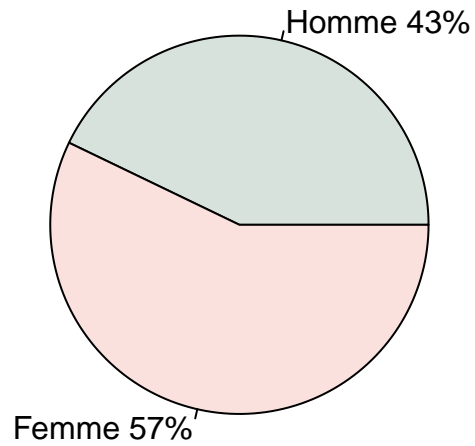
Répartition selon les pratiquants et les non pratiquants



2. Répartition selon le sexe :

```
sexe <- df$sexe
freq1=table(sexe)
pct1 <- round(freq1/sum(freq1)*100)
lbls1 <- c("Homme", "Femme")
lbls1 <- paste(lbls1, pct1)
lbls1 <- paste(lbls1, "%", sep="")
pie(freq1, labels = lbls1, col=c("#d8e2dc", "#fae1dd"), main="Répartition selon le Sexe")
```

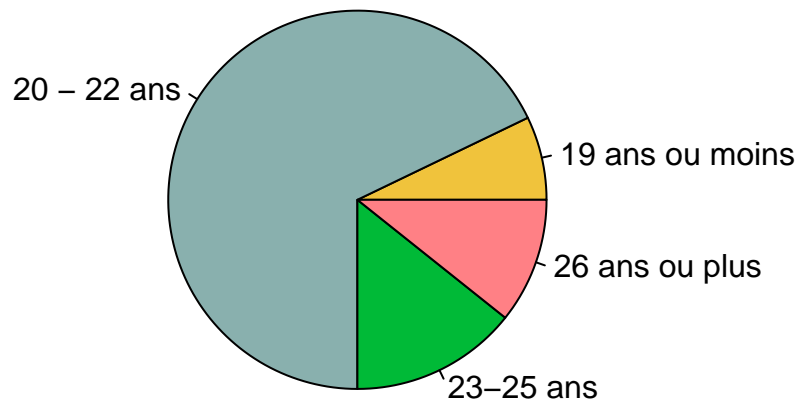
Répartition selon le Sexe



3. Répartition selon l'âge :

```
lbls2 <- c("19 ans ou moins", "20 - 22 ans", "23-25 ans", "26 ans ou plus")
df1 <- data.frame(df$age) %>%
  group_by(df$age) %>%
  count(df$age)
pie(table(df$age), labels = lbls2, col=c("#f0c33c", "#89b0ae", "#00ba37", "#ff8085"), main="Répartition selon l'âge")
```

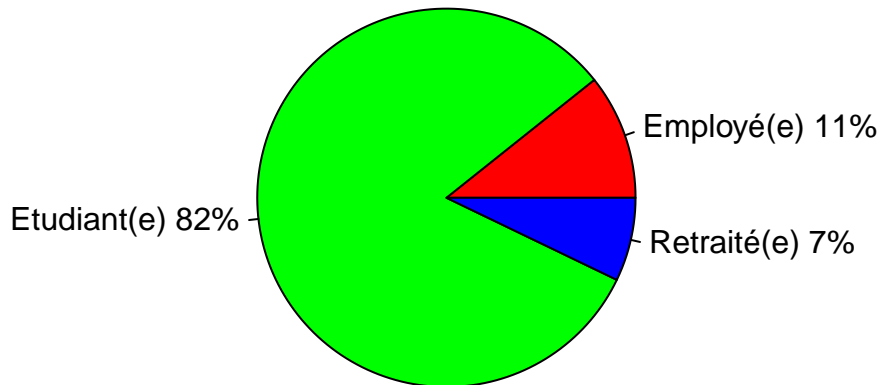
Répartition selon l'age



4. Répartition selon La catégorie socio-professionnelle :

```
occ<- df$catégorie
freq4<-table(occ)
pct4 <- round(freq4/sum(freq1)*100)
lbls4 <- names(freq4)
lbls4 <- paste(lbls4, pct4)
lbls4 <- paste(lbls4,"%",sep="")
pie(freq4,labels = lbls4, col=rainbow(length(lbls4))
    ,main="Répartition selon la catégorie socio-professionnelle")
```

Répartition selon la catégorie socio-professionnelle



II/ But de l'ACP : On veut déterminer les causes principales qui poussent les Tunisiens à faire du sport

```
sport <- select(df,c(4:13))
```

```
res.pca1 <- PCA (sport,graph=FALSE)
```

1.les axes à conserver :

On commence par le critère du taux d'inertie cumulé et de Kaiser.

```
eig.val1 <- get_eigenvalue(res.pca1)  
eig.val1
```

##	eigenvalue	variance.percent	cumulative.variance.percent
## Dim.1	3.4798184	34.798184	34.79818
## Dim.2	1.6715328	16.715328	51.51351
## Dim.3	1.2909823	12.909823	64.42334
## Dim.4	1.1478859	11.478859	75.90219
## Dim.5	0.7791403	7.791403	83.69360

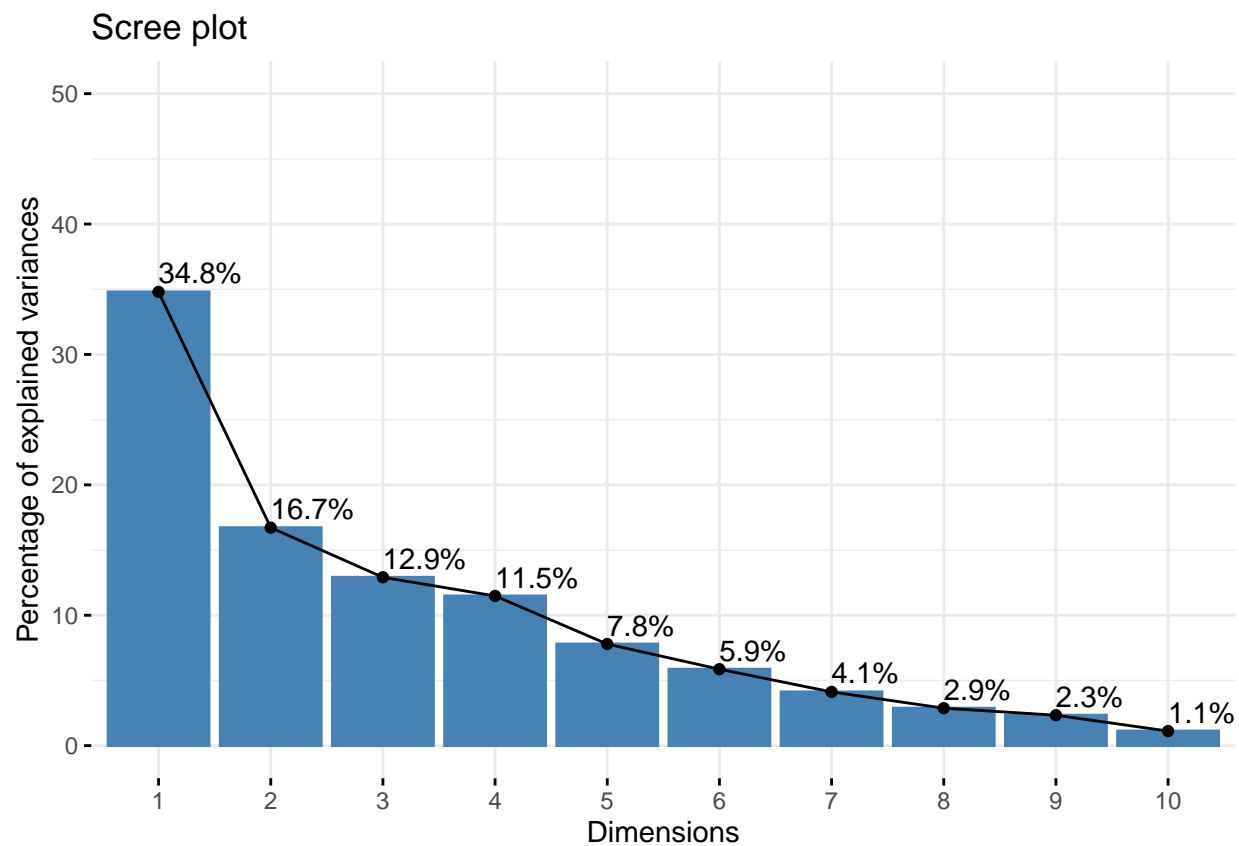
## Dim.6	0.5862218	5.862218	89.55582
## Dim.7	0.4120603	4.120603	93.67642
## Dim.8	0.2871667	2.871667	96.54809
## Dim.9	0.2333241	2.333241	98.88133
## Dim.10	0.1118673	1.118673	100.00000

Critère de Kaiser : On a 4 valeurs propres supérieures à 1.

Critère d'inertie cumulée : On pourrait retenir les 4 premières composantes principales qui expliquent 75.9% de l'inertie totale.

On passe maintenant au critère du coude

```
fviz_eig(res.pca1, addlabels = TRUE, ylim = c(0, 50))
```



D'après le critère du Coude, on observe un décrochement (coude) suivi d'une décroissance . On sélectionne les axes avant le coude. Selon le tracé de l'éboulis présenté ci-dessus, le nombre d'axe à conserver est de 2.

D'après ces trois critères le nombre d'axe à conserver est de 2 axes.

Après avoir fixé les axes à conserver, on commence l'étude afin d'aboutir aux cartes des variables et des individus et les interprète.

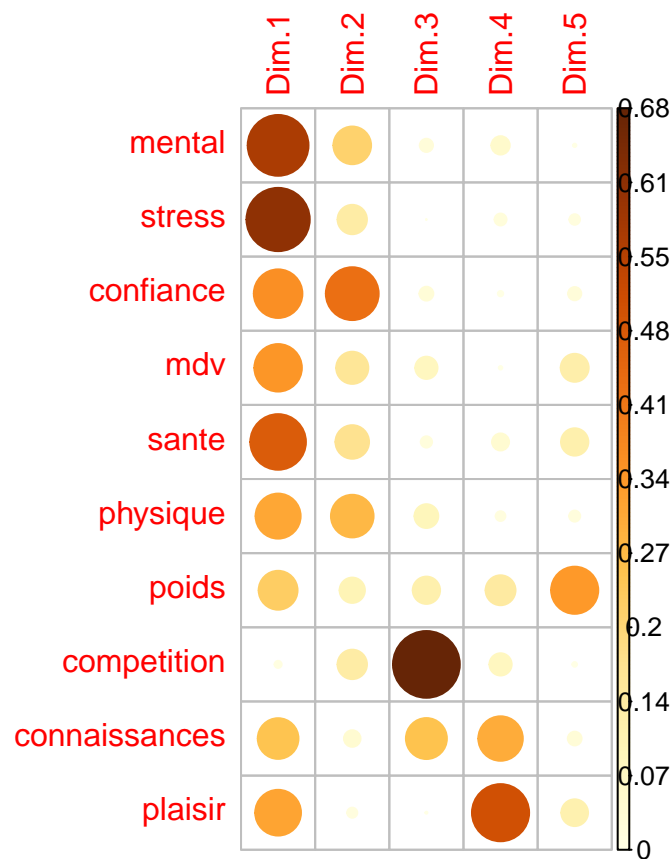
2.Interprétation de la carte des variables:

```
var1 <- get_pca_var(res.pca1)
var1
```

```
## Principal Component Analysis Results for variables
## =====
##   Name      Description
## 1 "$coord"   "Coordinates for the variables"
## 2 "$cor"     "Correlations between variables and dimensions"
## 3 "$cos2"    "Cos2 for the variables"
## 4 "$contrib" "contributions of the variables"
```

Qualité de représentation des variables

```
corrplot(var1$cos2, is.corr=FALSE)
```



Plus le cos2 est élevé plus la variable est mieux représentée. par exemple, la variable “competition” est bien représentée sur le 3ème axe alors qu’elle est mal représentée sur le 1er et 2ème axe.

Carte des variables

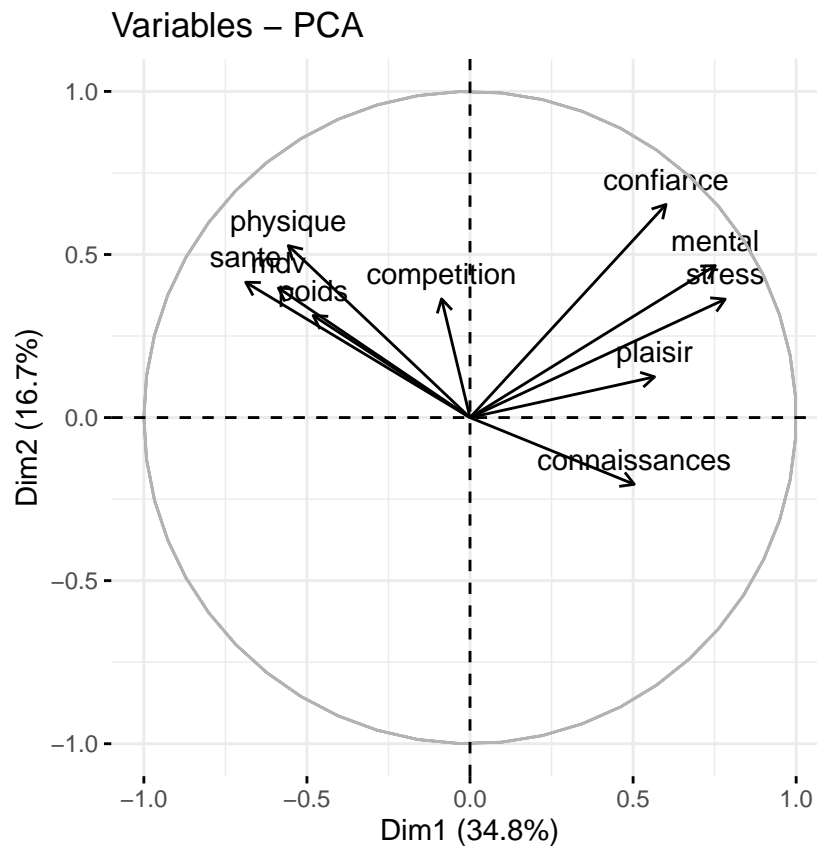
```
var1$coord
```

```
##           Dim.1      Dim.2      Dim.3      Dim.4      Dim.5
## mental      0.75211937  0.4658457 -0.163198549 -2.266441e-01  0.04162801
```



```
## stress      0.78251161  0.3630188 -0.009543775 -1.476052e-01 -0.13050232
## confiance   0.60093012  0.6533635 -0.171471571  2.922047e-05  0.16298025
## mdv         -0.58746198  0.3975135 -0.275036622  4.321190e-02 -0.34601405
## sante       -0.68764496  0.4149671  0.139156246  2.100741e-01 -0.33785390
## physique    -0.55704320  0.5271283  0.294119955  1.202618e-01  0.13620808
## poids       -0.48094388  0.3128427 -0.337754873  3.698681e-01  0.58308282
## competition -0.08718056  0.3639500  0.825790756 -2.748330e-01  0.05783874
## connaissances 0.50344499 -0.2045909  0.506646307  5.499711e-01  0.16994338
## plaisir     0.56574270  0.1248499 -0.025289587  7.067291e-01 -0.33042202
```

```
fviz_pca_var(res.pca1)
```



Interprétation

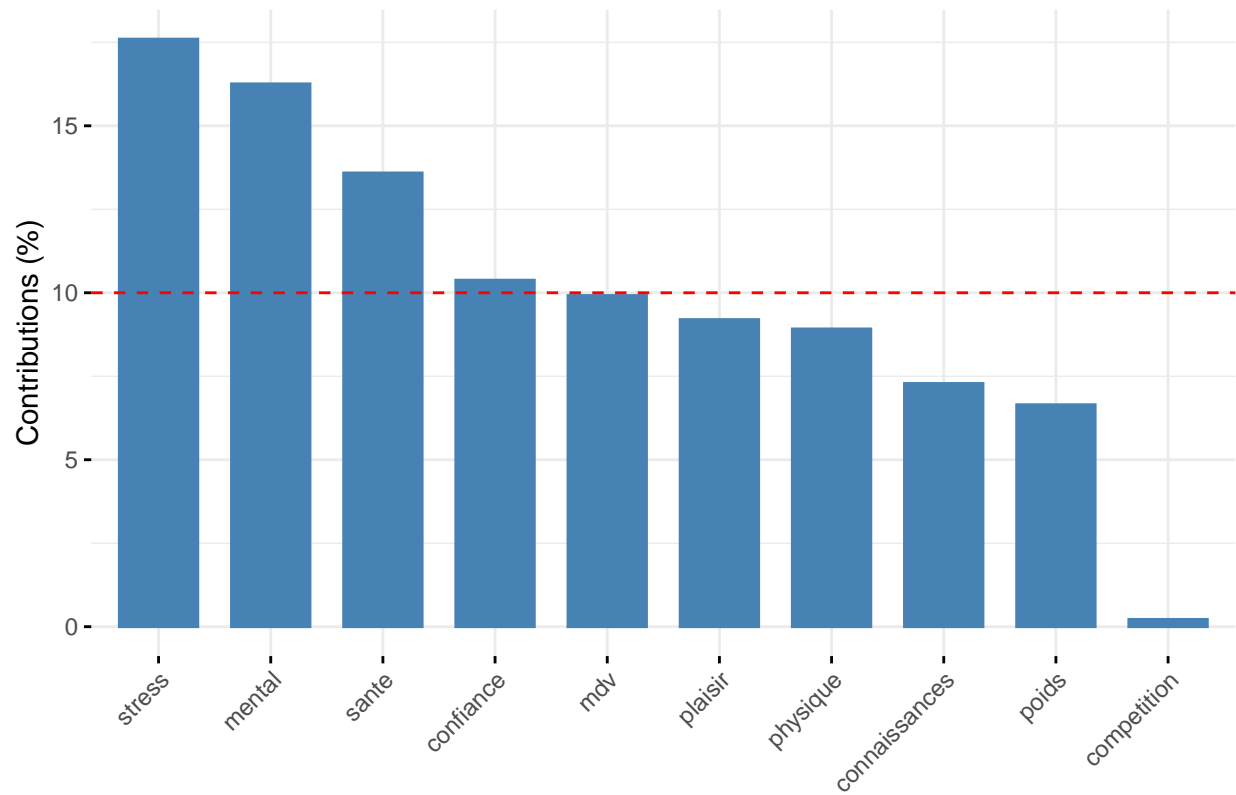
Les deux premières dimensions contiennent 51.5% de l'inertie totale. Les variables "mental", "stress" et "santé" interviennent dans la contribution et la construction du premier axe. ce premier axe oppose les personnes qui font du sport pour le plaisir et le dévouement à ceux qui font du sport pour améliorer leurs conditions physiques et pour perdre du poids.

Le deuxième axe est déterminé par les variables "confiance" et "physique".

Pour vérifier la contribution des axes :

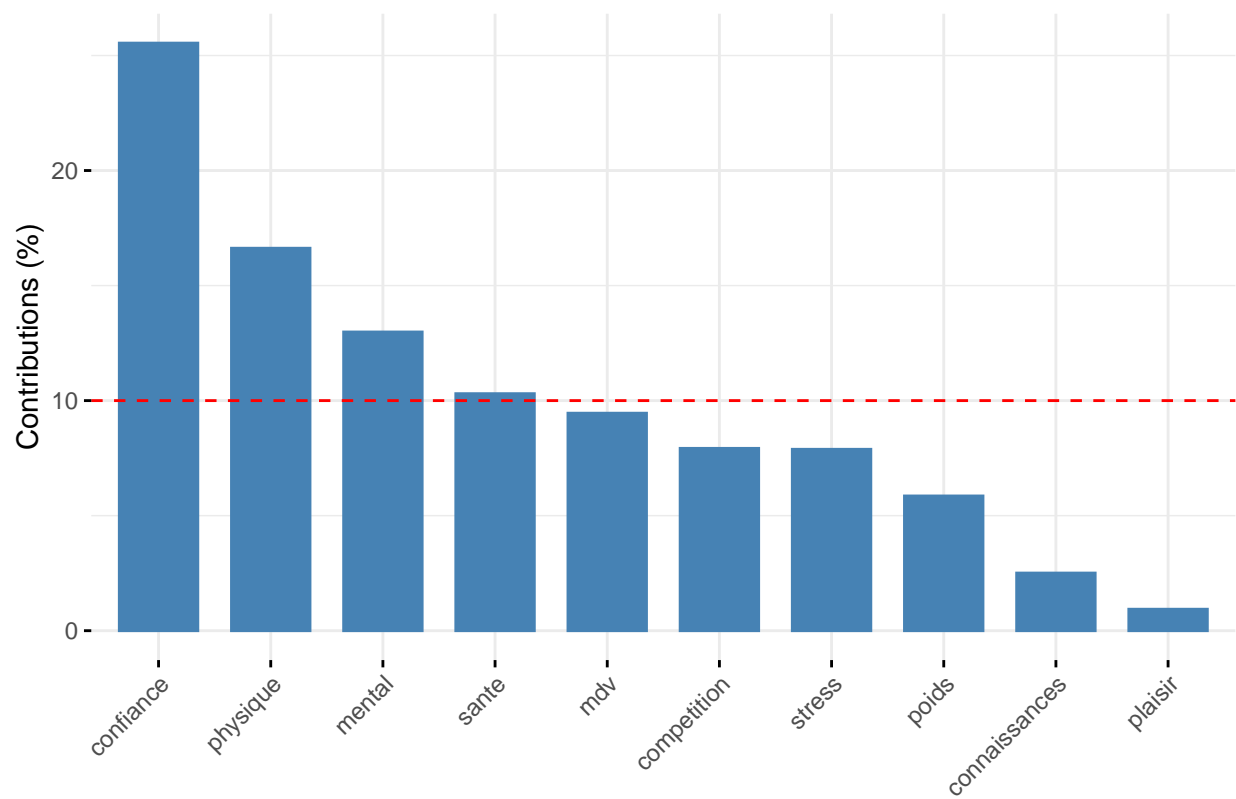
```
fviz_contrib(res.pca1, choice = "var", axes = 1, top = 10)
```

Contribution of variables to Dim-1

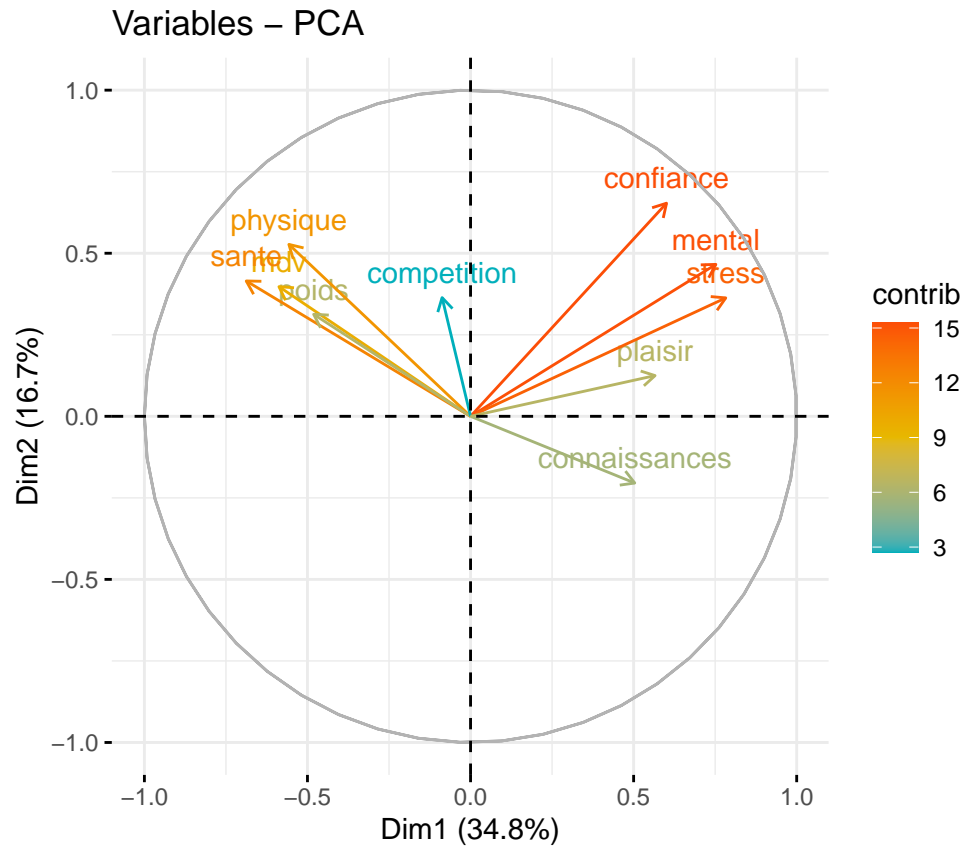


```
fviz_contrib(res.pca1, choice = "var", axes = 2, top = 10)
```

Contribution of variables to Dim-2

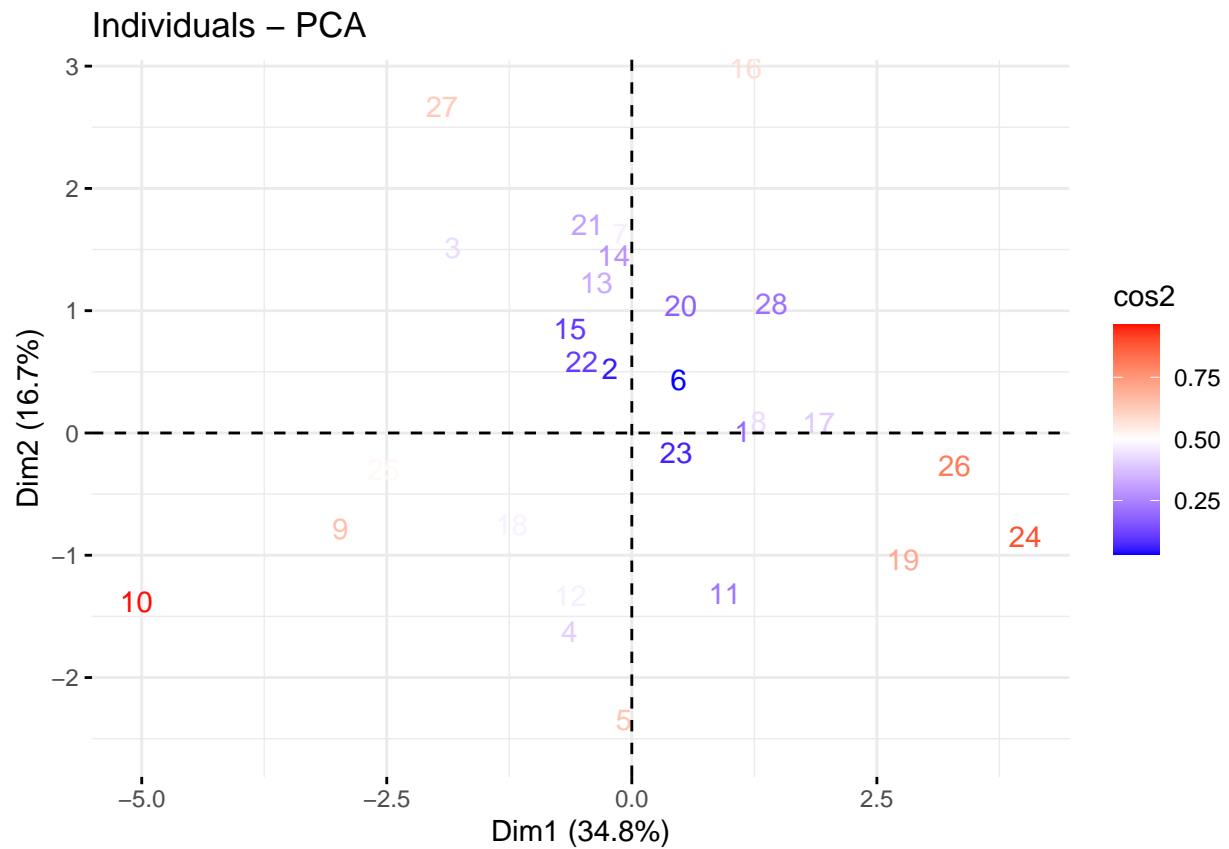


```
fviz_pca_var(res.pca1, col.var = "contrib",  
             gradient.cols = c("#00AFBB", "#E7B800", "#FC4E07")  
             )
```



2.La carte des individus :

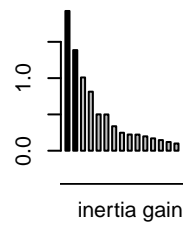
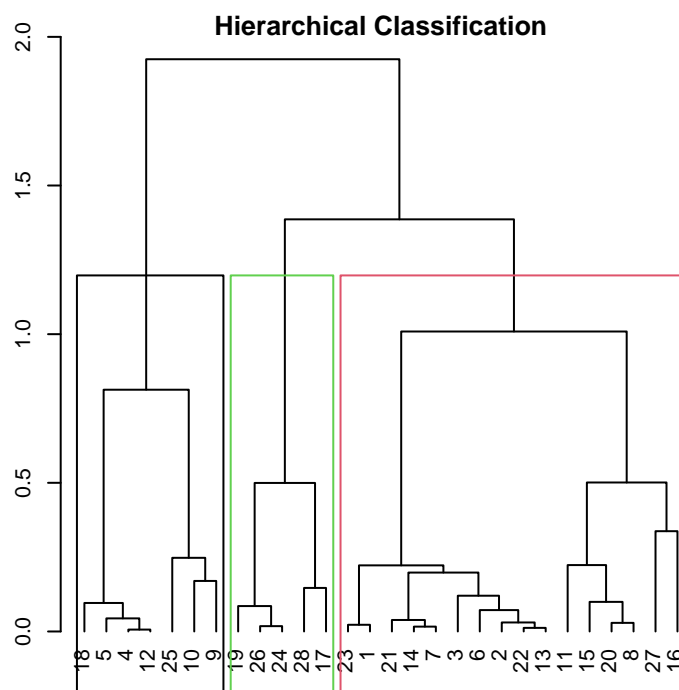
```
fviz_pca_ind(res.pca1,geom ="text",col.ind="cos2")+
scale_color_gradient2(low="blue", mid="white", high="red", midpoint=0.5)
```



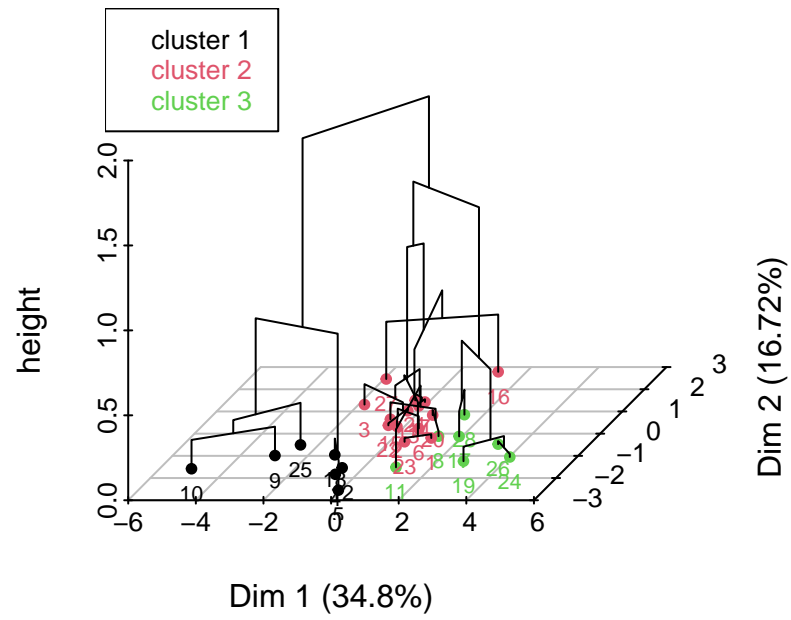
III/ Classification

```
res.pca <- PCA(sport, ncp = 3, graph = FALSE)
res.hcpc <- HCPC(res.pca, nb.clust = 3, graph = FALSE)
HCPC (res.pca1, nb.clust = 3, graph = TRUE)
```

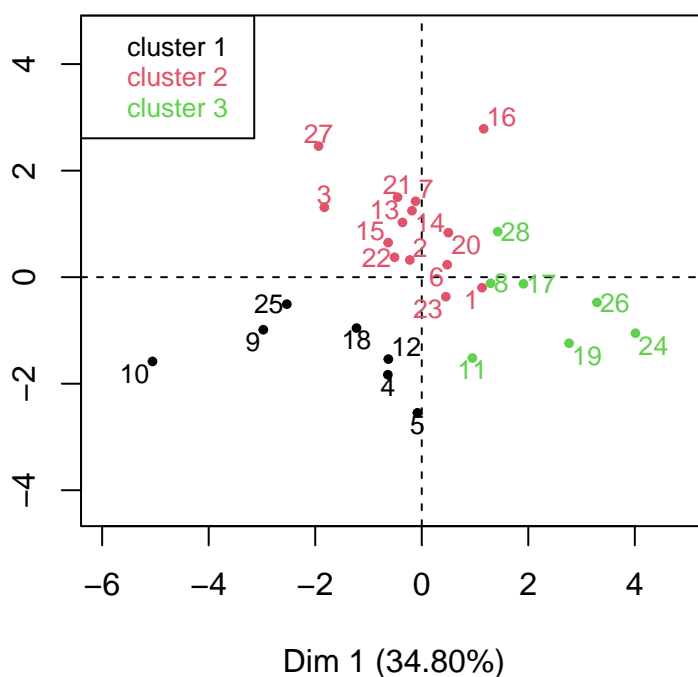
Hierarchical Clustering



Hierarchical clustering on the factor map



Factor map

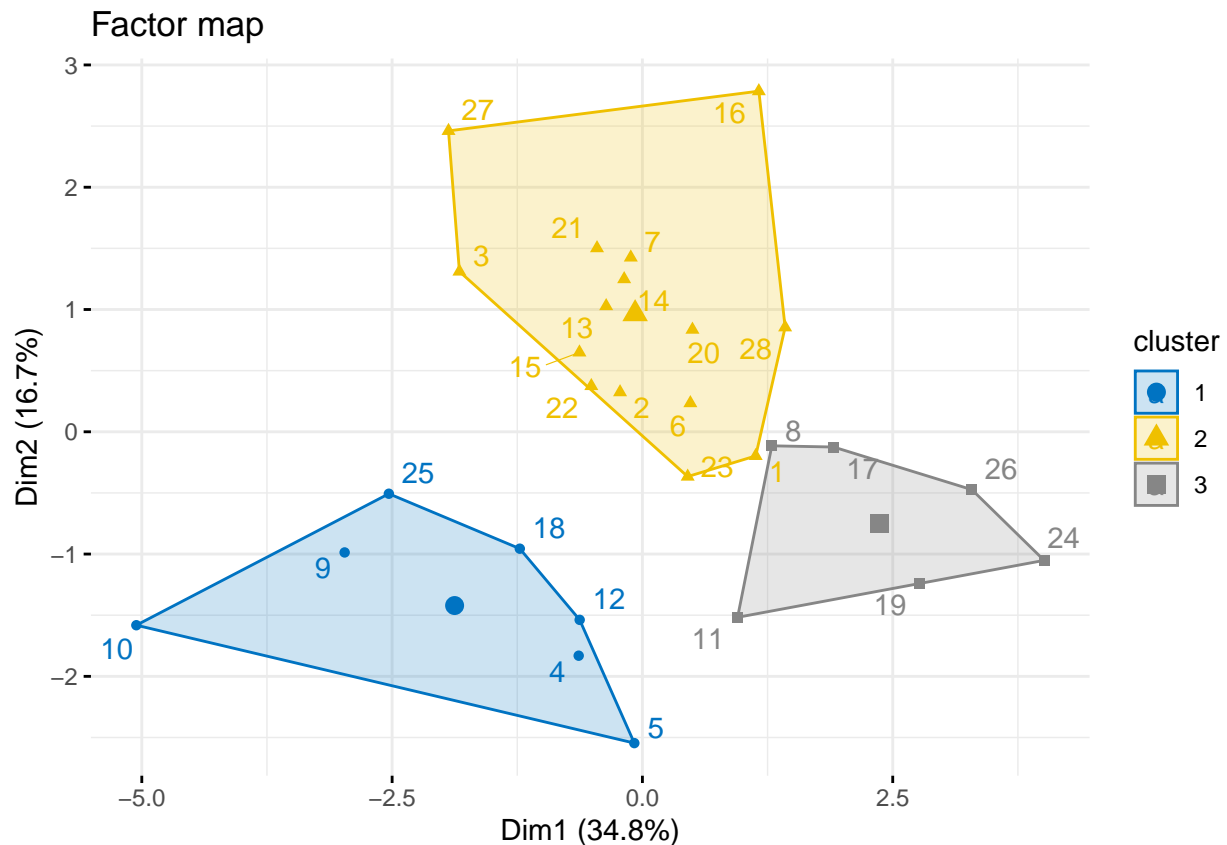


```
## **Results for the Hierarchical Clustering on Principal Components**
##   name
## 1  "$data.clust"
## 2  "$desc.var"
## 3  "$desc.var$quanti.var"
## 4  "$desc.var$quanti"
## 5  "$desc.axes"
## 6  "$desc.axes$quanti.var"
## 7  "$desc.axes$quanti"
## 8  "$desc.ind"
## 9  "$desc.ind$para"
## 10 "$desc.ind$dist"
## 11 "$call"
## 12 "$call$t"
##   description
## 1  "dataset with the cluster of the individuals"
## 2  "description of the clusters by the variables"
## 3  "description of the cluster var. by the continuous var."
## 4  "description of the clusters by the continuous var."
## 5  "description of the clusters by the dimensions"
## 6  "description of the cluster var. by the axes"
## 7  "description of the clusters by the axes"
## 8  "description of the clusters by the individuals"
## 9  "parangons of each clusters"
## 10 "specific individuals"
## 11 "summary statistics"
```



```
## 12 "description of the tree"
```

```
fviz_cluster(res.hcpc,  
repel = TRUE,  
show.clust.cent = TRUE,  
palette = "jco",  
ggtheme = theme_minimal(),  
main = "Factor map"  
)
```



Les classes observées

Cluster 1 : la catégorie des personnes qui font du sport pour être en bonne forme physique.

Cluster 2 : la catégorie des personnes qui font du sport pour se débarrasser du stress et avoir confiance en soi.

Cluster 3 : la catégorie des personnes qui font du sport pour le plaisir et pour faire des nouvelles connaissances.

IV/ But de l'ACM: voir le comportement de ceux qui pratiquent une activité sportive

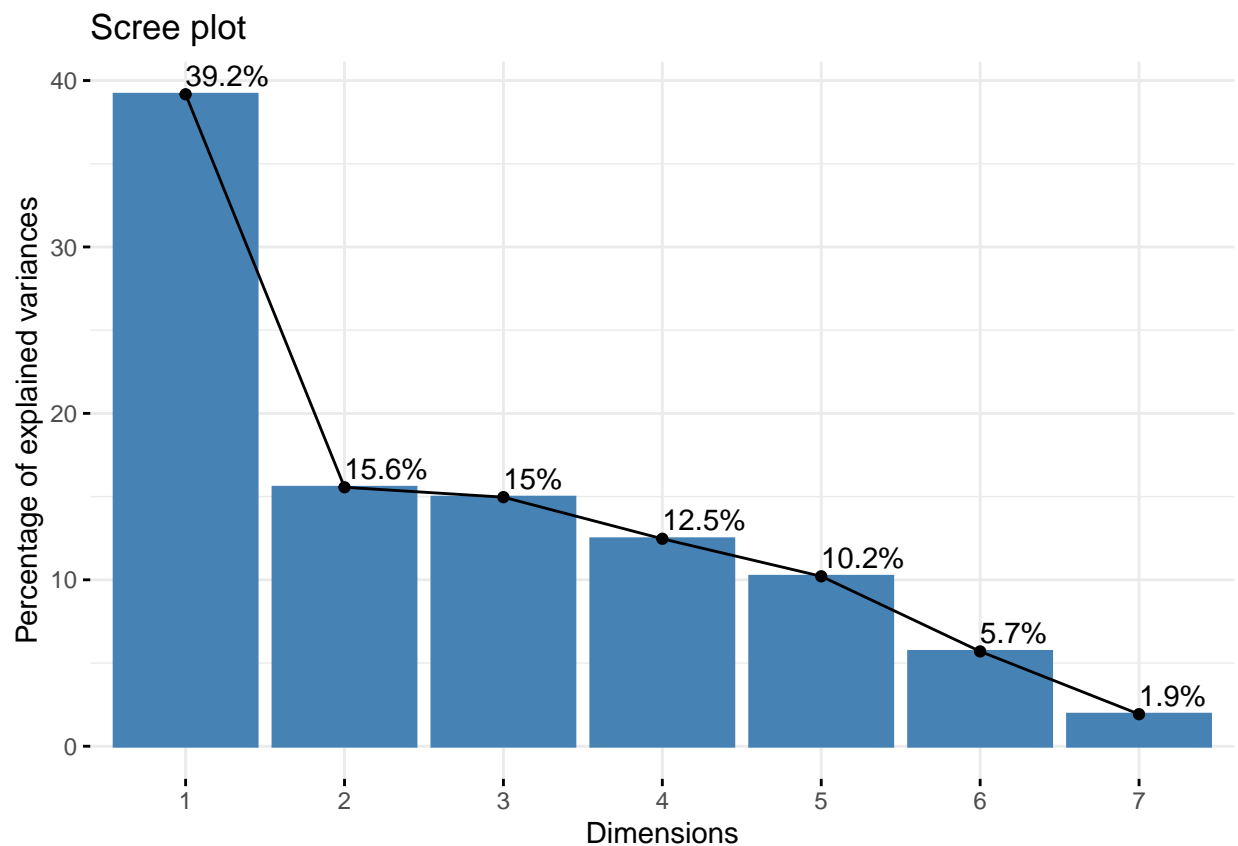
```
df_2=select(df,c(25,26,27,28,29,30,31))  
def2 <- MCA(df_2,graph = FALSE)
```

1.les axes à conserver :

```
def2$eig
```

```
##      eigenvalue percentage of variance cumulative percentage of variance
## dim 1 0.39173309          39.173309          39.17331
## dim 2 0.15559897          15.559897          54.73321
## dim 3 0.14966110          14.966110          69.69932
## dim 4 0.12464864          12.464864          82.16418
## dim 5 0.10211323          10.211323          92.37550
## dim 6 0.05698011           5.698011          98.07351
## dim 7 0.01926485           1.926485         100.00000
```

```
fviz_eig(def2,addlabels= TRUE)
```



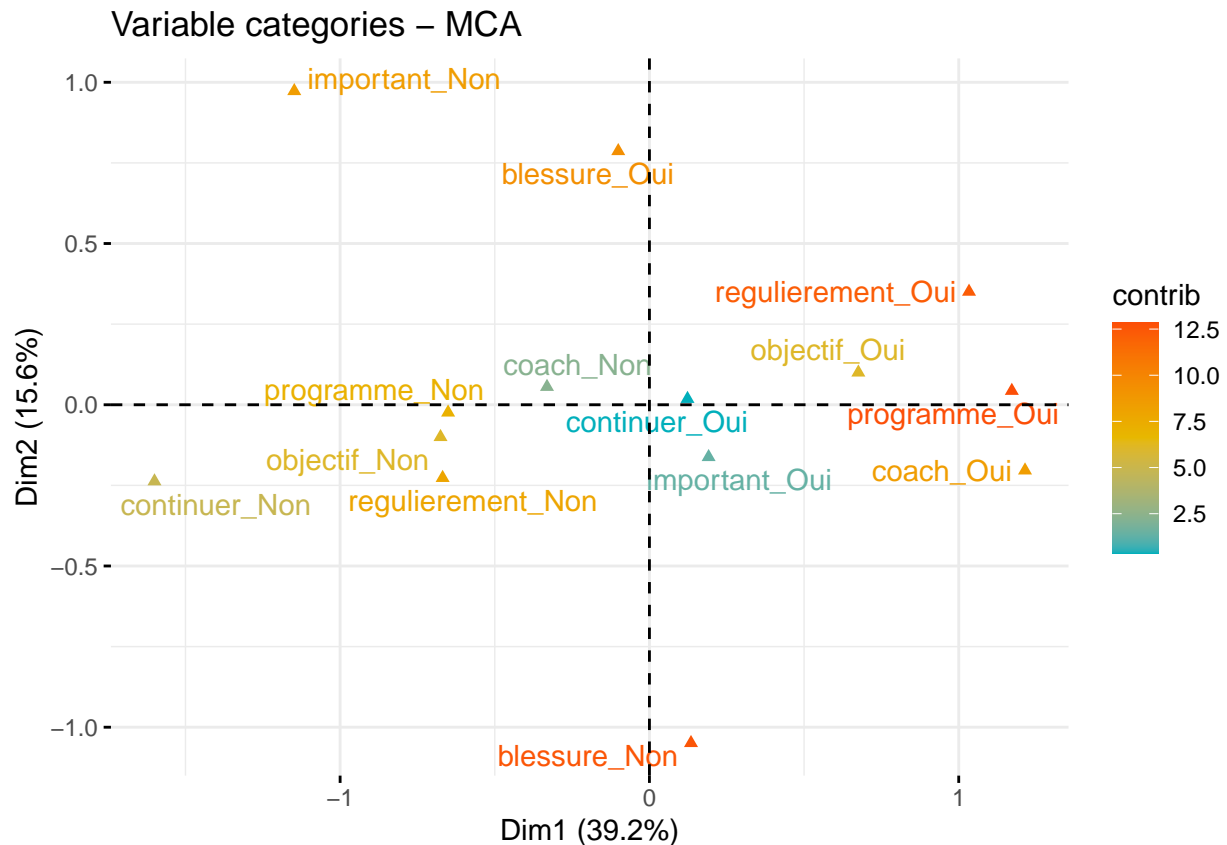
Interprétation

Critère du taux d'inertie cumulée : les deux premières dimensions cumulent une proportion de variance d'environ 54.7%, ce qui suggère que ces deux dimensions sont importantes pour expliquer les données.

Critère du coude : On peut voir que le coude se situe après la deuxième composante principale, ce qui indique que l'on peut se contenter de retenir les deux premières composantes principales.

2. La carte des modalités :

```
fviz_mca_var(def2,  
  col.var = "contrib", # Couleur en fonction de la contribution  
  gradient.cols = c("#00AFBB", "#E7B800", "#FC4E07"), # Choix de couleurs  
  repel = TRUE, # Évite le chevauchement des labels  
  ggtheme = theme_minimal() # Style du graphique  
)
```



Interprétation

Le premier axe semble être un axe qui indique la persévérance sportive

Le deuxième axe semble être un axe qui indique la blessure sportive

3. La carte des individus :

```
fviz_mca_ind (def2, select.ind = list(cos2 = 0.4),  
  repel = TRUE,  
  ggtheme = theme_minimal ())
```

