

# TP: Analyse des Correspondances Multiples - Correction

```
rm(list=ls())
library("tidyverse")

## -- Attaching packages ----- tidyverse 1.3.0 --

## v ggplot2 3.3.3      v purrr 0.3.4
## v tibble 3.0.6       v dplyr 1.0.4
## v tidyr 1.1.2        v stringr 1.4.0
## v readr 1.4.0        v forcats 0.5.1

## -- Conflicts ----- tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()     masks stats::lag()

library("FactoMineR") #pour effectuer l'ACP
library("factoextra") #pour extraire et visualiser les résultats issus de FactoMineR

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

## 1 Loisirs

### 1.1 Importation du jeu de données

```
Loisirs=read.table("AnaDo_JeuDonnees_Loisirs.csv",header=TRUE,sep=";",na.strings = "NA")
# On enlève les variables Sexe, Age, Situation matrimoniale, Profession
Loisirs <- Loisirs %>% dplyr::select(-Age,-Sexe,-Situationmatrimoniale,-Profession,-NbActivites)
# On transforme la variable TV en facteur
Loisirs$TV <- as.factor(Loisirs$TV)
str(Loisirs)

## 'data.frame': 8403 obs. of 18 variables:
## $ Lecture : chr "0" "0" "0" "0" ...
## $ EcouterMusique : chr "0" "N" "0" "N" ...
## $ Cinema : chr "0" "N" "N" "N" ...
## $ Spectacle : chr "0" "N" "N" "N" ...
## $ Exposition : chr "0" "0" "0" "0" ...
## $ Ordinateur : chr "N" "N" "N" "N" ...
## $ Sport : chr "0" "0" "N" "0" ...
## $ Marche : chr "0" "0" "N" "N" ...
## $ Voyage : chr "0" "N" "0" "0" ...
## $ JouerMusique : chr "N" "N" "N" "N" ...
```

```
## $ Collection      : chr  "N" "O" "N" "N" ...
## $ ActiviteBenevole: chr  "O" "O" "N" "N" ...
## $ Bricolage       : chr  "O" "O" "N" "N" ...
## $ Jardinage       : chr  "N" "O" "N" "N" ...
## $ Tricot          : chr  "N" "N" "N" "N" ...
## $ Cuisine         : chr  "N" "N" "N" "N" ...
## $ Peche           : chr  "N" "N" "N" "N" ...
## $ TV              : Factor w/ 5 levels "0","1","2","3",...: 3 5 5 2 4 4 4 1 2 2 ...
```

```
dim(Loisirs)
```

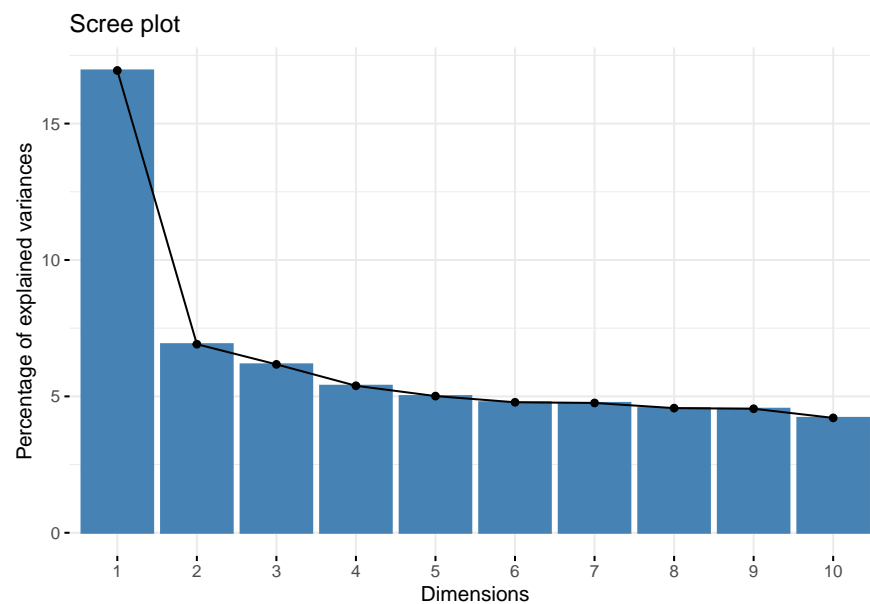
```
## [1] 8403  18
```

## 1.2 ACM sur toutes les variables sauf Nbactivités

```
res.mca <- MCA(Loisirs, graph=FALSE)
```

## 1.3 Graphe des valeurs propres

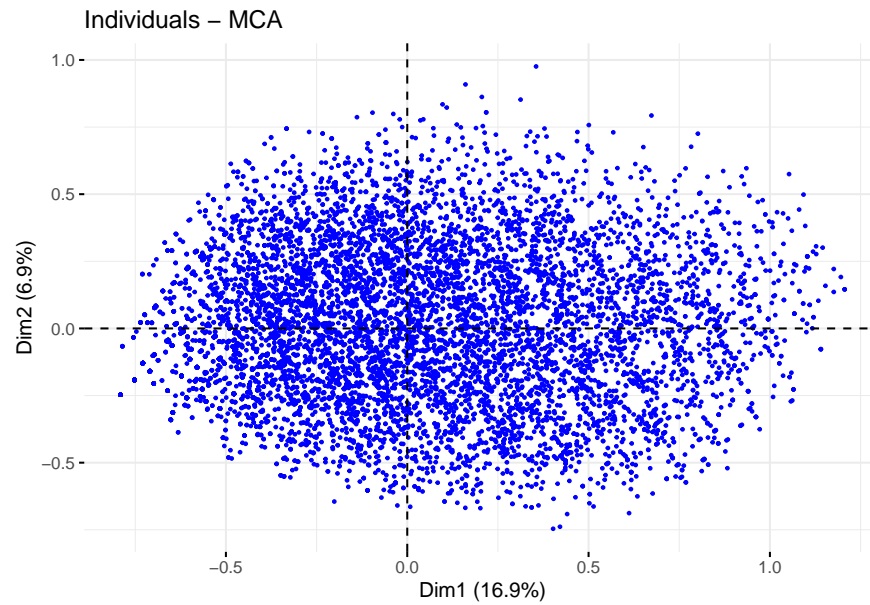
```
fviz_eig(res.mca)
```



Le pourcentage d'inertie portée par le plan principal est de 24% ## Analyse

## 1.4 Les individus

```
fviz_mca_ind(res.mca,repel=TRUE,label="none",pointsize = 0.5)
```

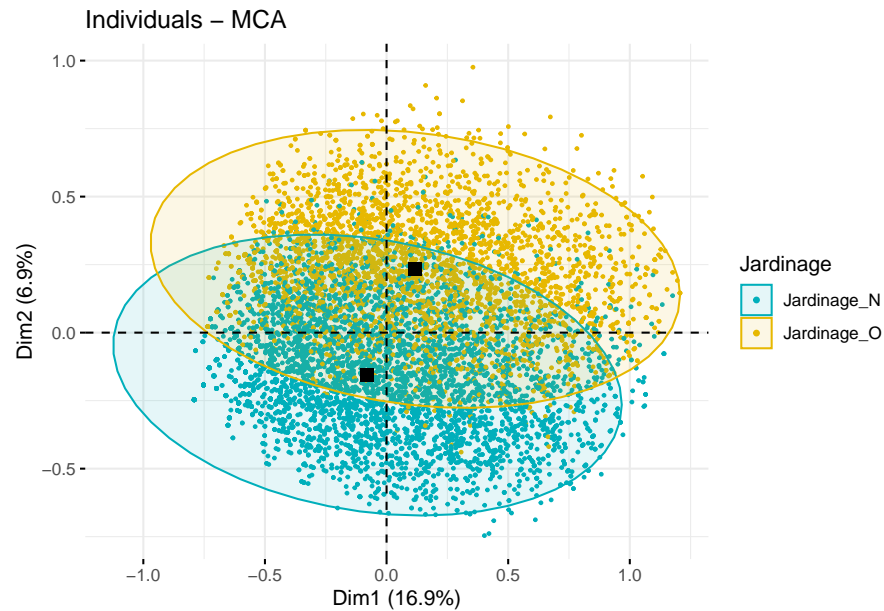


## 1.5 Les individus selon le ‘Jardinage’

*# Calcul des coordonnées du barycentre des individus par modalité de jardinage  
# (une modalité est au barycentre des individus qui la possède)*

```
coord_in_12 <- res.mca$ind$coord[,1:2]
JN <- colMeans(coord_in_12[Loisirs$Jardinage=="N",])
JO <- colMeans(coord_in_12[Loisirs$Jardinage=="O",])

p <- fviz_mca_ind(res.mca,
  label = "none", # masquer le texte des individus
  habillage = "Jardinage", # colorer par groupes
  palette = c("#00AFBB", "#E7B800"),
  addEllipses = TRUE,
  ggtheme = theme_minimal(), pointsize = 0.5)
p+
  geom_point(aes(x=JO[1],y=JO[2]),size=3, shape=15)+
  geom_point(aes(x=JN[1],y=JN[2]),size=3, shape=15)
```



```
table(Loisirs$Jardinage)
```

```
##
##      N      O
## 5047 3356
```

L'axe 2 semble séparer les individus selon qu'ils jardinent ou non. Le barycentre des individus qui ne jardinent pas est plus proche de l'origine que celui des individus qui jardinent. On en déduit qu'il y a plus d'individus qui ne jardinent pas.

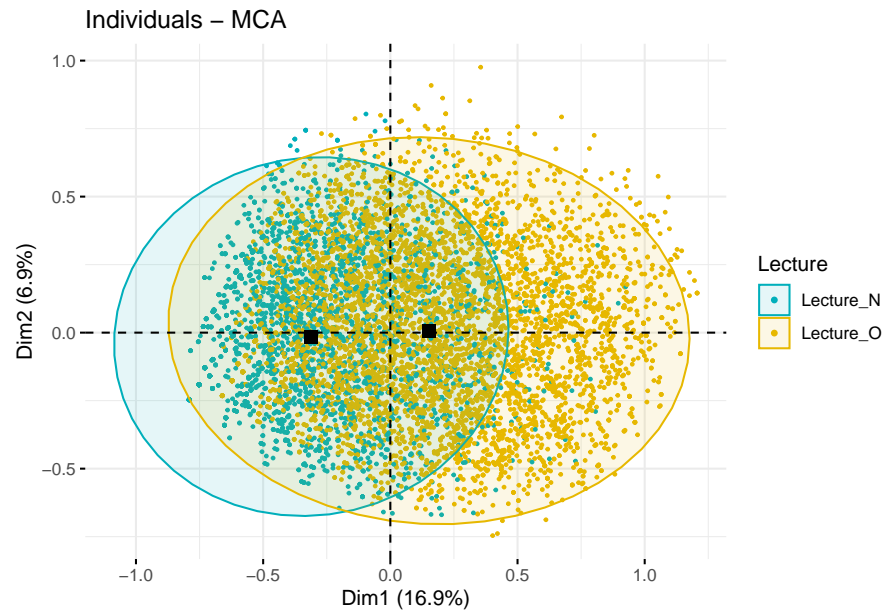
## 1.6 Les individus selon le 'Lecture'

```
# Calcul des coordonnées du barycentre des individus par modalité de Lecture
# (une modalité est au barycentre des individus qui la possède)

LN <- colMeans(coord_in_12[Loisirs$Lecture=="N",])
LO <- colMeans(coord_in_12[Loisirs$Lecture=="O",])

p <- fviz_mca_ind(res.mca,
  label = "none", # masquer le texte des individus
  habillage = "Lecture", # colorer par groupes
  palette = c("#00AFBB", "#E7B800"),
  addEllipses = TRUE,
  ggtheme = theme_minimal(), pointsize = 0.5)

p+
  geom_point(aes(x=LO[1],y=LO[2]),size=3, shape=15)+
  geom_point(aes(x=LN[1],y=LN[2]),size=3, shape=15)
```



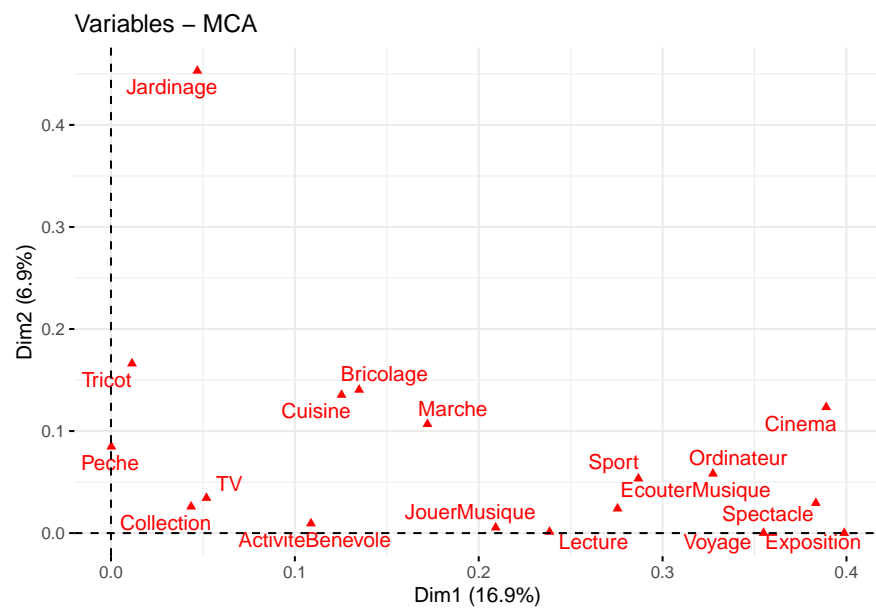
```
table(Loisirs$Lecture)
```

```
##
##      N      0
## 2757 5646
```

L'axe 1 semble séparer les individus qui lisent des autres. Il y a plus d'individus qui lisent comparé à ceux qui ne lisent pas.

## 1.7 Les variables

```
fviz_mca_var(res.mca, choice = "mca.cor", repel = TRUE)
```

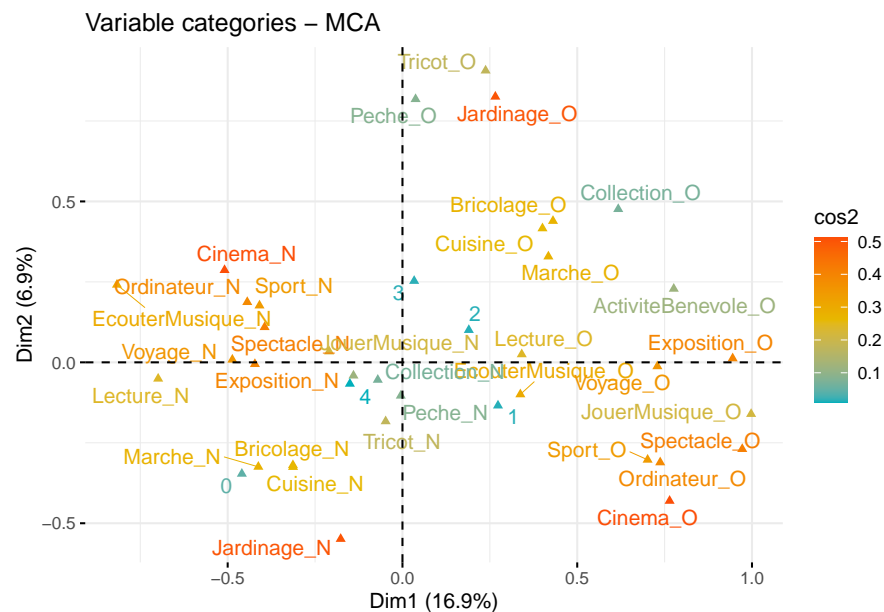


L'axe 1 est très corrélée aux activités culturelles et l'axe 2 au jardinage et dans une moindre mesure aux activités de Bricolage et Tricot.

## 1.8 Les modalités

```
fviz_mca_var(res.mca, col.var = "cos2",
  gradient.cols = c("#00AFBB", "#E7B800", "#FC4E07"),
  repel = TRUE,
  ggtheme = theme_minimal())
```

```
## Warning: ggrepel: 1 unlabeled data points (too many overlaps). Consider
## increasing max.overlaps
```



## 1.9 Contribution des modalités aux deux premiers axes

```
res.mca$var$contrib[,1:2]
```

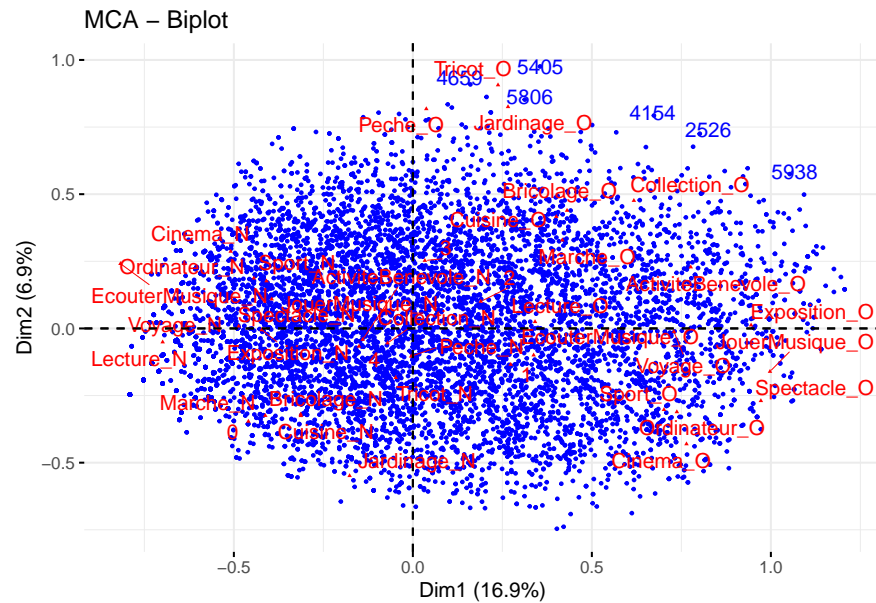
##	Dim 1	Dim 2
## Lecture_N	4.5032219225	0.058354237
## Lecture_O	2.1989696848	0.028494975
## EcouterMusique_N	5.4784495896	1.170355021
## EcouterMusique_O	2.2624974259	0.483334779
## Cinema_N	4.3693964253	3.398013220
## Cinema_O	6.5612490531	5.102583710
## Spectacle_N	3.1086402651	0.585807244
## Spectacle_O	7.6632789710	1.444105445
## Exposition_N	3.4605242601	0.001417903
## Exposition_O	7.7451733730	0.003173479
## Ordinateur_N	3.4573805209	1.506765703

## Ordinateur_0	5.7422295225	2.502528851
## Sport_N	2.9686665720	1.355928936
## Sport_0	5.0913351097	2.325450983
## Marche_N	2.4029937376	3.653801047
## Marche_0	2.4334988078	3.700184629
## Voyage_N	3.9912696968	0.002797257
## Voyage_0	5.9815638632	0.004192143
## JouerMusique_N	1.0214897558	0.064877132
## JouerMusique_0	4.8576735441	0.308521867
## Collection_N	0.1255703527	0.183124605
## Collection_0	1.0985220762	1.602021630
## ActiviteBenevole_N	0.4673868258	0.099302145
## ActiviteBenevole_0	2.5889956621	0.550064333
## Bricolage_N	1.5968707659	4.073308690
## Bricolage_0	2.1947384587	5.598353622
## Jardinage_N	0.5270330951	12.464019778
## Jardinage_0	0.7925911893	18.744311031
## Tricot_N	0.0540307420	1.924809068
## Tricot_0	0.2672858361	9.521879255
## Cuisine_N	1.5455816599	4.091989564
## Cuisine_0	1.9778916684	5.236547687
## Peche_N	0.0005641581	0.656555533
## Peche_0	0.0044523716	5.181577954
## 0	0.7180192943	1.000293241
## 1	0.3054837556	0.180849640
## 2	0.2599259171	0.178151342
## 3	0.0066553120	0.930772183
## 4	0.1688987584	0.081380139

## 1.10 Représentation simultanée

```
fviz_mca_biplot(res.mca,repel=TRUE,pointsize = 0.5)
```

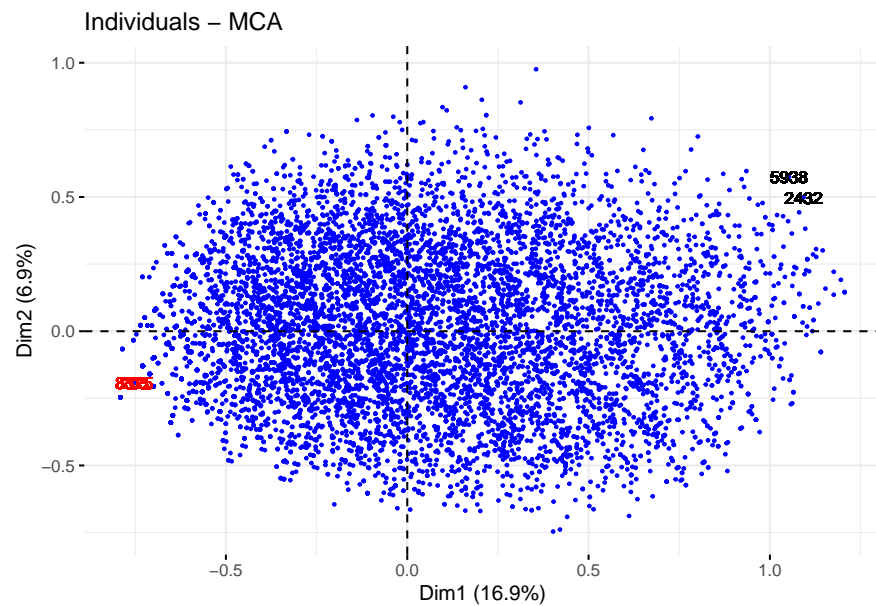
```
## Warning: ggrepel: 8397 unlabeled data points (too many overlaps). Consider
## increasing max.overlaps
```



### 1.11 Quelques individus particulier

```
p <- fviz_mca_ind(res.mca,repel=TRUE,label="none",pointsize = 0.5)

p+geom_text(aes(x=coord_in_12[2432,1],y=coord_in_12[2432,2]),size=3, label="2432")+
  geom_text(aes(x=coord_in_12[5938,1],y=coord_in_12[5938,2]),label="5938",size=3)+
  geom_text(aes(x=coord_in_12[203,1],y=coord_in_12[203,2]),size=3, color="red", label="203")+
  geom_text(aes(x=coord_in_12[8325,1],y=coord_in_12[8325,2]),size=3, color="red", label="8325")
```



```
Loisirs[c(2432,5938,203,8325),]
```

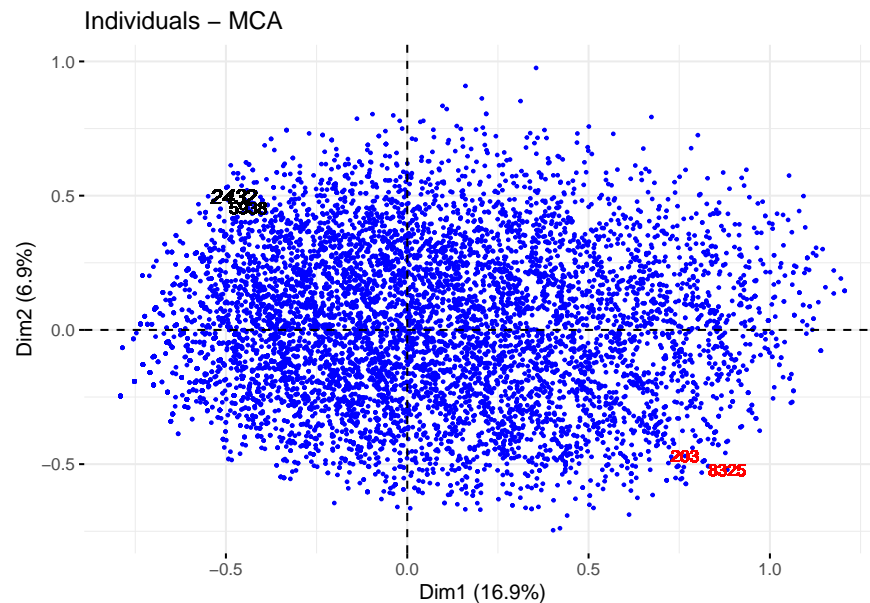
```
##      Lecture EcouterMusique Cinema Spectacle Exposition Ordinateur Sport Marche
```



##	2432	0	0	0	0	0	0	N	0
##	5938	0	0	N	0	0	0	0	0
##	203	N	N	N	N	N	N	N	N
##	8325	N	N	N	N	N	N	N	N
##		Voyage	JouerMusique	Collection	ActiviteBenevole	Bricolage	Jardinage	Tricot	
##	2432	0	0	0	0	0	0	0	
##	5938	0	0	0	0	0	0	0	
##	203	N	N	N	N	N	N	N	
##	8325	N	N	N	N	N	N	N	
##		Cuisine	Pêche	TV					
##	2432	0	N	2					
##	5938	0	N	3					
##	203	N	N	4					
##	8325	N	N	4					

Les individus 2432 et 5938 sont du côté des réponses ‘Oui’, ils ont donc beaucoup d’activités loisirs. A contrario, les individus 203 et 8325 sont du côté des ‘Non’, ils ont donc peu d’activités.

```
p <- fviz_mca_ind(res.mca,repel=TRUE,label="none",pointsize = 0.5)
p+
  geom_text(aes(x=coord_in_12[5676,1],y=coord_in_12[2432,2]),fontface=3, label="2432")+
  geom_text(aes(x=coord_in_12[1143,1],y=coord_in_12[1143,2]),size=3, label="5938")+
  geom_text(aes(x=coord_in_12[6766,1],y=coord_in_12[6766,2]),size=3, color="red", label="203")+
  geom_text(aes(x=coord_in_12[255,1],y=coord_in_12[255,2]),size=3, color="red", label="8325")
```



```
Loisirs[c(5676,1143,6766,255),]
```

##		Lecture	EcouterMusique	Cinema	Spectacle	Exposition	Ordinateur	Sport	Marche
##	5676	N	N	N	N	N	N	N	N
##	1143	0	N	N	N	N	N	N	N
##	6766	0	0	0	0	0	0	0	0
##	255	0	0	0	0	0	0	0	0

```
##      Voyage JouerMusique Collection ActiviteBenevole Bricolage Jardinage Tricot
## 5676      N              N          N              N          0          0          0
## 1143      N              N          N              N          0          0          0
## 6766      0              0          N              N          N          N          N
## 255       0              0          N              0          N          N          N
##      Cuisine Peche TV
## 5676      0          N  4
## 1143      N          N  4
## 6766      0          N  0
## 255       N          N  1
```

Les individus 5676 et 1143 sont du côté des Tricot, Jardinage ‘Oui’ et Cinema, Ordinateur ‘Non’, ils ont des activités plutôt ‘tranquilles’ et pas d’activités ‘jeunes’ . A contrario, les individus 6766 et 255 sont du côté des Sport, Cinema, Ordinateur ‘Oui’ et Jardin, Cuisine ‘Non’, ils ont des loisirs plutôt ‘jeunes’ et pas d’activités ‘tranquilles’. Et ce pour un même nombre d’activités (environ mêmes coordonnées sur la première bissectrice).

### 1.12 Interprétation des deux bissectrices:

- 1ère bissectrice: intensité de l’activité (oppose beaucoup d’activités à très peu d’activités)
- 2ème bissectrice : type d’activités (oppose les activités plutôt ‘jeunes’ aux activités plutôt ‘sénior’). On peut noter que cette dimension sépare les activités des individus à nombre d’activités constant (orthogonal à la 1ère bissectrice)

#Chiens

### 1.13 Importation du jeu de données

```
load("dogs.rda")
Chiens <- data
head(Chiens)
```

```
##      Size Weight Velocity Intelligence Affectivity Aggressivness
## Beauceron      S++      W+      V++          I+          Af+          Ag+
## BassetHound     S-      W-      V-          I-          Af-          Ag+
## GermanShepherd S++      W+      V++          I++          Af+          Ag+
## Boxer           S+      W+      V+          I+          Af+          Ag+
## Bulldog         S-      W-      V-          I+          Af+          Ag-
## BullMastiff     S++      W++      V-          I++          Af-          Ag+
```

```
dim(Chiens)
```

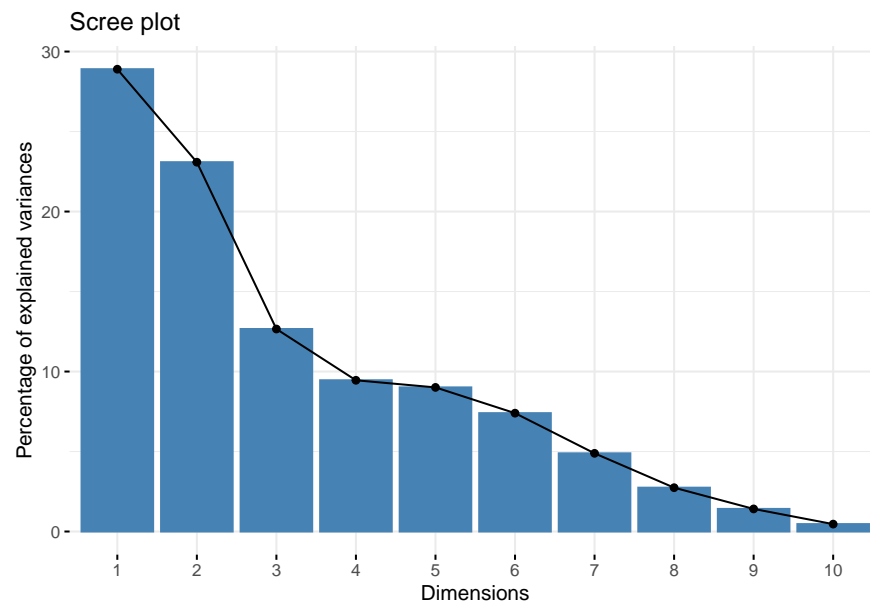
```
## [1] 27  6
```

### 1.14 ACM sur toutes les variables sauf Nbactivités

```
res.mca <- MCA(Chiens, graph=FALSE)
```

## 1.15 Graphe des valeurs propres

```
fviz_eig(res.mca)
```



```
res.mca$eig
```

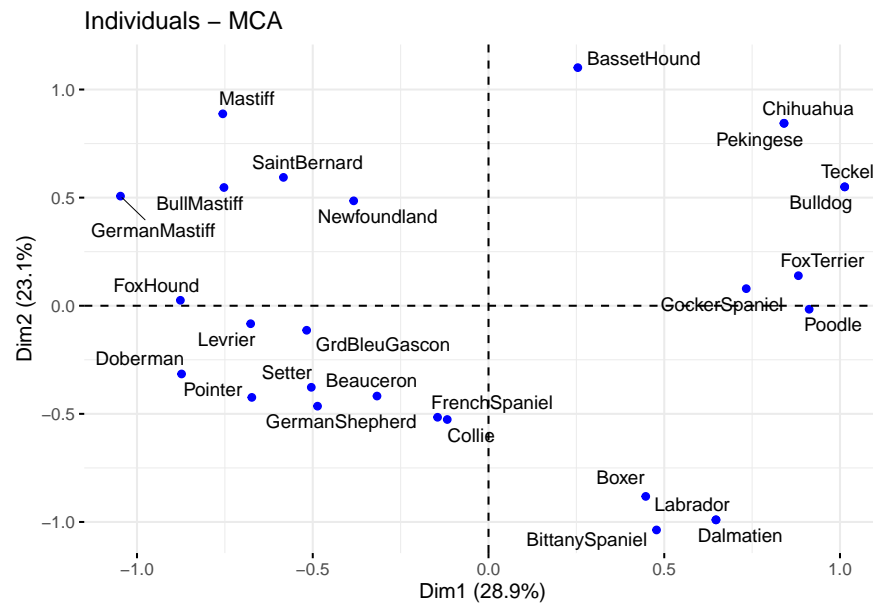
	eigenvalue	percentage of variance	cumulative percentage of variance
## dim 1	0.481606165	28.896370	28.89637
## dim 2	0.384737288	23.084237	51.98061
## dim 3	0.210954049	12.657243	64.63785
## dim 4	0.157554025	9.453242	74.09109
## dim 5	0.150132670	9.007960	83.09905
## dim 6	0.123295308	7.397718	90.49677
## dim 7	0.081462460	4.887748	95.38452
## dim 8	0.045669757	2.740185	98.12470
## dim 9	0.023541911	1.412515	99.53722
## dim 10	0.007713034	0.462782	100.00000

Le pourcentage d'inertie portée par le plan principal est de 52%

## 1.16 Analyse

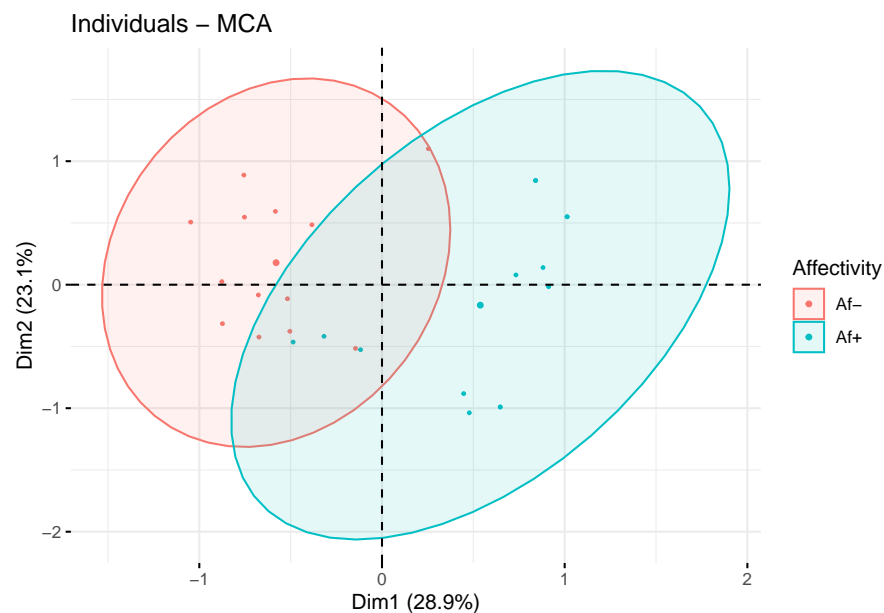
## 1.17 Les individus

```
fviz_mca_ind(res.mca,repel=TRUE,labels=0.4)
```



## 1.18 Les chiens selon l' 'Affectivity'

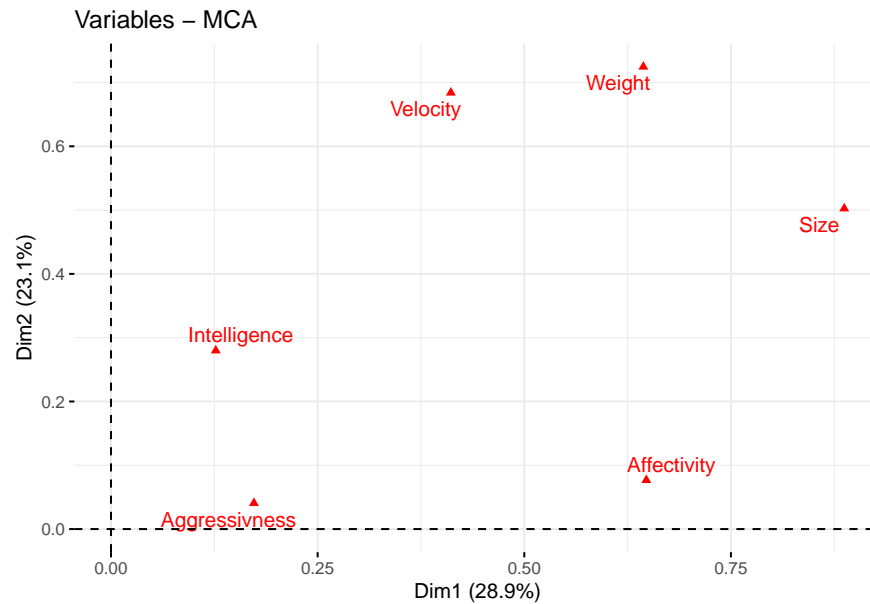
```
fviz_mca_ind(res.mca,
  label = "none", # masquer le texte des individus
  habillage = "Affectivity", # colorer par groupes
  addEllipses = TRUE,
  ggtheme = theme_minimal (),pointsize = 0.5)
```



L'axe 1 semble séparer les chiens affectives de ceux qui ne le sont pas.

## 1.19 Les variables

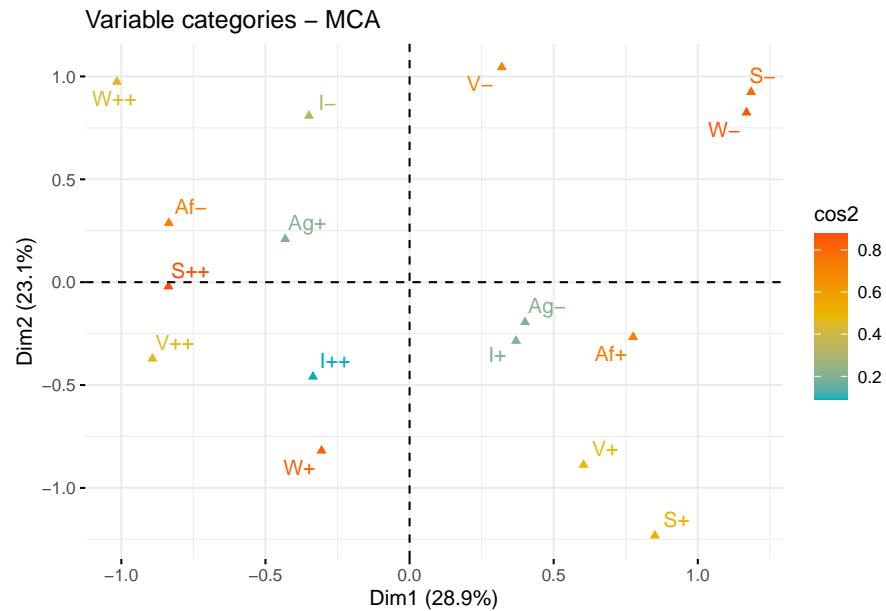
```
fviz_mca_var(res.mca, choice = "mca.cor", repel = TRUE)
```



L'axe 1 est très corrélée aux variables 'Affectivity', 'Size' et 'Weight' et l'axe 2 aux variables 'Velocity', 'Weight' et 'Size'.

## 1.20 Les modalités

```
fviz_mca_var(res.mca, col.var = "cos2",  
  gradient.cols = c("#00AFBB", "#E7B800", "#FC4E07"),  
  repel = TRUE,  
  ggtheme = theme_minimal())
```



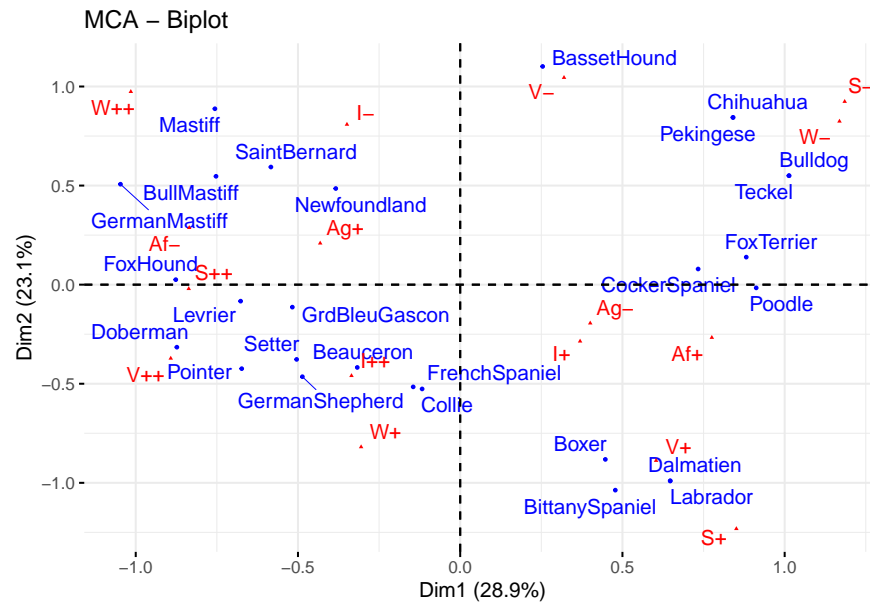
## 1.21 Contribution des modalités aux deux premiers axes

```
res.mca$var$contrib[,1:2]
```

```
##          Dim 1          Dim 2
## S-   12.5978150   9.58661729
## S+    4.6420727  12.17067028
## S++  13.4585463   0.01019149
## W-   14.0104164   8.72224556
## W+    1.6736860  15.06207234
## W++   6.6040417   7.60886705
## V-    1.3119931  17.51742290
## V+    3.7368537  10.11705778
## V++   9.1804174   1.99644722
## I-    1.2492400   8.39130827
## I+    2.2742083   1.70014447
## I++   0.8633836   2.03240794
## Af-  11.6215827   1.72364624
## Af+  10.7914697   1.60052866
## Ag-    2.8813167   0.84758676
## Ag+    3.1029565   0.91278575
```

## 1.22 Représentation simultanée et interprétation

```
fviz_mca_biplot(res.mca,repel=TRUE,pointsize = 0.5)
```



- L'axe 1 oppose des petits et moyens chiens affectifs et peu agressifs à de très gros chiens rapides et agressifs (et non affectifs)
- L'axe 2 est un axe de taille, de vitesse et d'intelligence. Il oppose les tous petits et gros chiens lents et peu intelligents à des chiens de corpulence moyenne, rapide et intelligents.