

## Introduction :

L'acronyme MOOC (Massive Online Open Course) est utilisé pour désigner les plateformes d'enseignement en ligne qui proposent des cours ouverts et qui peuvent s'adresser à des centaines, des milliers, voire des dizaines de milliers d'étudiants simultanément. De nombreuses institutions (grandes écoles, universités, grandes entreprises, etc.) investissent dans la création de ce type de cours avec divers objectifs : aider au développement de la formation continue, de la formation des pays en voie de développement, accroître leur visibilité.

La finalité de ce projet n'est pas de proposer une analyse du phénomène MOOC, de ses enjeux, de ses réussites, de ses échecs ou de son avenir, mais d'offrir une vision originale d'un MOOC, au travers d'une base de sondage représentant la population des étudiants - constituée des élèves de l'IMT Atlantique, de leur profil, leur activité et de leurs motivations.

Pour cela nous allons effectuer une ACM.

L'Analyse des Correspondances Multiples (ACM ou MCA pour multiple correspondence analysis) est une extension de l'analyse factorielle des correspondances pour résumer et visualiser un tableau de données contenant plus de deux variables catégorielles. On peut aussi la considérer comme une généralisation de l'analyse en composantes principales lorsque les variables à analyser sont catégorielles plutôt que quantitatives (source STHDA, Abdi and Williams 2010).

L'objectif de cette étude est d'identifier un groupe de personnes qui partagent un profil similaire envers les cours en ligne.

Vous trouvez les données issues de notre sondage sur la même repository nommé [dataset.xlsx](#). NB : on a utilisé que les données des individus qui ont suivi les MOOCs.

On va aborder deux parties, la première sera une étude préliminaire, qui va nous permettre de choisir les individus et variables utiles pour la suite de l'étude, qui sera la deuxième partie.

## ACM préliminaire :

### Calcul :

Pour le calcul nous allons utiliser la fonction `MCA()` incluse dans le package `FactoMiner`, et pour les visualisation nous allons opter pour `factoextra`.

### Charger les librararies :

```
library(FactoMiner)
```

```
library(factoextra)
```

### Charger les données :

```
# xlsx files
library(xlsx)

my_data <- read.xlsx("dataset.xlsx",sheetIndex = 1,as.data.frame = T,encoding
= 'UTF-8')
```

### Nommer les colonnes :

```
my_col_names=c('date','a_suivi','performer_travail_projet_1','motivation_cert
ificat_1','trouver_emploi_1','mooc_inclu_etude_1','edt_chargee_1','plateforme
s_connues','aide_financiere_1','prix_1','outils_exercices_1','sciences_donnee
s_1','informatique_1','sciences_physiques_1','business_1','arts_1','langues_1
','sciences_sociales_1','performer_travail_projet','motivation_certificat','t
rouver_emploi','nbr_certificat','satisfait_gestion_temps','objectifs_atteints
','satisfaction','vacances_etude','temps_alloue_cours','mooc_inclu_etude','ed
t_chargee','plateformes_utilisees','videos_cours','transcription','diapositiv
es_cours','forum_discussion','examens_exercices','accompagnement_pedagogique
','aides_financiere','prix','outils_exercice','sciences_donnees','informatique
','sciences_physiques','business','arts','langues','sciences_sociales','sexe'
,'pays','age','est_etudiant','niveau_actuel','formation','background','mail',
'commentaire')
colnames(my_data)<-my_col_names
```

```
## j'ai nettoyé les lignes à la main donc on prend tout les lignes
my_data2<-my_data[c(1:nrow(my_data)),]
```

### Eliminer les colonnes : mail, date et commentaire

```
my_data2$date<-NULL
my_data2$mail<-NULL
my_data2$commentaire<-NULL
```

### Créer la trame de données nécessaires :

#### Librairie pour filtrer les lignes :

```
library("dplyr")
```

En remarquant le résultat de la fonction str(), on effectue le code suivant :

```
#change from factor to numeric the feature : temps_alloue_cours
my_data2$temps_alloue_cours<-as.numeric(my_data2$temps_alloue_cours)
```

### La trame de donnée voulue :

```
# Choisir que les personnes qui ont suivi des moocs et qu'ils l'ont terminé o
u pas
data_suivi<-filter(my_data2,a_suivi %in% c("Oui, je l'ai suivi mais je ne l'
ai pas terminé", "Oui, je l'ai suivi et j'ai réussi" ))
#Supprimer chaque colonne qui est totalement des NaN ( en fait ce sont celle
de la partie: jamais suivi)
data_suivi<-data_suivi[ , ! apply( data_suivi , 2 , function(x) all(is.na(x))
) ]
```

Ces données proviennent d'une enquête menée auprès des étudiants de l'IMT Atlantique à propos des MOOCs qu'ils ont suivis.

En jargon ACM :

- Individus actifs (lignes 1:115): individus qui sont utilisés dans l'ACM.
- Variables actives (toutes les colonnes sauf les colonnes numériques+ plateforme\_utilisee+ a\_suivi+sexe, est\_etudiant, formation, niveau\_actuel): variables utilisées dans l'ACM. -Variables supplémentaires (sexe, âge, satisfaction, nbr\_certificats, temps\_alloue\_cours): elles ne participent pas à l'ACM. Les coordonnées de ces variables seront prédites.

Nous commençons par extraire les individus actifs et les variables actives pour l'ACM:

```
data_suivi.active <- data_suivi[, c( "motivation_certificat", "trouver_emploi", "vacances_etude" , "mooc_inclu_etude" , "edt_chargee" , "videos_cours" , "transcription", "diapositives_cours" , "forum_discussion", "examens_exercices", "accompagnement_pedagogique", "aides_financiere", "outils_exercice" , "sciences_donnees", "informatique" , "sciences_physiques", "business", "arts", "langues" , "sciences_sociales" , "est_etudiant" , "formation" , "background" )]
```

```
str(data_suivi.active)
```

```
## 'data.frame': 115 obs. of 23 variables:
## $ motivation_certificat : Factor w/ 5 levels "D'accord","Désaccord",.
.: 1 4 1 2 5 1 2 1 2 4 ...
## $ trouver_emploi : Factor w/ 5 levels "D'accord","Désaccord",.
.: 3 4 1 5 5 2 3 1 2 2 ...
## $ vacances_etude : Factor w/ 3 levels "Les deux","Les vacances
",..: 1 1 1 3 3 3 1 3 1 1 ...
## $ mooc_inclu_etude : Factor w/ 2 levels "Non","Oui": 1 1 2 2 2 2
2 2 2 1 ...
## $ edt_chargee : Factor w/ 2 levels "Non","Oui": 1 1 2 2 2 2
2 1 1 2 ...
## $ videos_cours : Factor w/ 5 levels "Bien","Mauvais",..: 4 4
1 3 1 3 1 4 4 1 ...
## $ transcription : Factor w/ 5 levels "Bien","Mauvais",..: 1 4
1 1 1 3 3 4 1 4 ...
## $ diapositives_cours : Factor w/ 5 levels "Bien","Mauvais",..: 1 4
1 1 3 1 4 1 4 1 ...
## $ forum_discussion : Factor w/ 5 levels "Bien","Mauvais",..: 1 4
1 4 3 3 5 1 4 1 ...
## $ examens_exercices : Factor w/ 5 levels "Bien","Mauvais",..: 4 4
1 3 2 1 1 3 1 1 ...
## $ accompagnement_pedagogique: Factor w/ 5 levels "Bien","Mauvais",..: 1 4
1 3 1 2 1 1 1 3 ...
## $ aides_financiere : Factor w/ 2 levels "Non","Oui": 2 2 2 1 1 1
2 1 1 1 ...
## $ outils_exercice : Factor w/ 2 levels "Non","Oui": 2 2 2 1 2 2
2 2 1 2 ...
```

```
## $ sciences_donnees      : Factor w/ 3 levels "Inintéressant",...: 2 2
2 2 2 3 2 2 2 2 ...
## $ informatique          : Factor w/ 3 levels "Inintéressant",...: 2 2
3 2 3 2 2 2 2 2 ...
## $ sciences_physiques    : Factor w/ 3 levels "Inintéressant",...: 2 3
2 2 2 2 3 2 3 3 ...
## $ business              : Factor w/ 3 levels "Inintéressant",...: 2 3
3 3 1 2 3 1 1 1 ...
## $ arts                   : Factor w/ 3 levels "Inintéressant",...: 1 3
2 1 1 3 2 1 3 3 ...
## $ langues                 : Factor w/ 3 levels "Inintéressant",...: 1 3
2 1 1 2 1 1 3 3 ...
## $ sciences_sociales     : Factor w/ 3 levels "Inintéressant",...: 1 3
3 1 1 2 1 1 1 1 ...
## $ est_etudiant          : Factor w/ 2 levels "Non","Oui": 2 2 2 2 2 2
2 2 2 2 ...
## $ formation              : Factor w/ 6 levels "FIG","FIL","FIP",...: 1
1 1 1 1 1 1 1 1 1 ...
## $ background             : Factor w/ 4 levels "Autre école",...: 3 3 4
3 3 3 3 3 3 3 ...
```

### Résumé des données :

La fonction `summary()` peut être utilisée pour calculer la fréquence des catégories des variables.

Résumés statistiques:

```
summary(data_suivi.active)
```

```
##              motivation_certificat              trouver_emploi
## D'accord           :27          D'accord           :23
## Désaccord          :25          Désaccord          :30
## Neutre              :17          Neutre              :32
## Tout à fait d'acord :30          Tout à fait d'acord  : 6
## Tout à fait en désaccord:16      Tout à fait en désaccord:24
##
##      vacances_etude mooc_inclu_etude edt_chargee      videos_cours
## Les deux      :58      Non:25          Non:47      Bien      :52
## Les vacances: 9      Oui:90          Oui:68      Mauvais    : 3
## Vos études   :48                                Moyen      :16
##                                                    Très bien   :43
##                                                    Très mauvais: 1
##
##      transcription      diapositives_cours      forum_discussion
## Bien      :52      Bien      :56      Bien      :42
## Mauvais    : 6      Mauvais    : 4      Mauvais    : 9
## Moyen      :21      Moyen      :31      Moyen      :44
## Très bien   :35      Très bien   :23      Très bien   :14
## Très mauvais: 1      Très mauvais: 1      Très mauvais: 6
##
```

```
##      examens_exercices  accompagnement_pedagogique  aides_financiere
## Bien      :59      Bien      :41      Non:91
## Mauvais   : 2      Mauvais   :15      Oui:24
## Moyen     :31      Moyen     :36
## Très bien :22      Très bien :15
## Très mauvais: 1      Très mauvais: 8
##
##      outils_exercice      sciences_donnees      informatique
## Non:20      Inintéressant      : 8      Inintéressant      : 4
## Oui:95      Intéressant      :87      Intéressant      :91
##      Pas très intéressant:20      Pas très intéressant:20
##
##
##
##      sciences_physiques      business
## Inintéressant      :21      Inintéressant      :21
## Intéressant      :49      Intéressant      :56
## Pas très intéressant:45      Pas très intéressant:38
##
##
##
##      arts      langues
## Inintéressant      :37      Inintéressant      :35
## Intéressant      :48      Intéressant      :50
## Pas très intéressant:30      Pas très intéressant:30
##
##
##
##      sciences_sociales  est_etudiant      formation
## Inintéressant      :38      Non: 2      FIG      :97
## Intéressant      :38      Oui:113      FIL      : 4
## Pas très intéressant:39      FIP      : 3
##      Master de recherche: 1
##      Master of science : 7
##      Master spécialisé : 3
##
##      background
## Autre école: 4
## DUT      : 4
## Prépa      :95
## Université :12
##
##
```

La fonction `summary()` renvoie la taille des catégories des variables.

On va éliminer les variables suivantes car elles présentent des catégories à très faible fréquences, chose qui peut fausser l'analyse : `est_etudiant`, `formation` et `niveau_actuel`.

```
data_suivi.active <- data_suivi[, c( "motivation_certificat" , "trouver_emploi", "vacances_etude" , "mooc_inclu_etude", "edt_chargee" , "videos_cours", "transcription", "diapositives_cours" , "forum_discussion", "examens_exercices"
```

```
, "accompagnement_pedagogique", "aides_financiere", "outils_exercice", "sciences_donnees", "informatique", "sciences_physiques", "business", "arts", "langues", "sciences_sociales", "background" )]
```

```
str(data_suivi.active
)
```

```
## 'data.frame': 115 obs. of 21 variables:
## $ motivation_certificat : Factor w/ 5 levels "D'accord","Désaccord",.
.: 1 4 1 2 5 1 2 1 2 4 ...
## $ trouver_emploi : Factor w/ 5 levels "D'accord","Désaccord",.
.: 3 4 1 5 5 2 3 1 2 2 ...
## $ vacances_etude : Factor w/ 3 levels "Les deux","Les vacances
",...: 1 1 1 3 3 3 1 3 1 1 ...
## $ mooc_inclu_etude : Factor w/ 2 levels "Non","Oui": 1 1 2 2 2 2
2 2 2 1 ...
## $ edt_chargee : Factor w/ 2 levels "Non","Oui": 1 1 2 2 2 2
2 1 1 2 ...
## $ videos_cours : Factor w/ 5 levels "Bien","Mauvais",...: 4 4
1 3 1 3 1 4 4 1 ...
## $ transcription : Factor w/ 5 levels "Bien","Mauvais",...: 1 4
1 1 1 3 3 4 1 4 ...
## $ diapositives_cours : Factor w/ 5 levels "Bien","Mauvais",...: 1 4
1 1 3 1 4 1 4 1 ...
## $ forum_discussion : Factor w/ 5 levels "Bien","Mauvais",...: 1 4
1 4 3 3 5 1 4 1 ...
## $ examens_exercices : Factor w/ 5 levels "Bien","Mauvais",...: 4 4
1 3 2 1 1 3 1 1 ...
## $ accompagnement_pedagogique: Factor w/ 5 levels "Bien","Mauvais",...: 1 4
1 3 1 2 1 1 1 3 ...
## $ aides_financiere : Factor w/ 2 levels "Non","Oui": 2 2 2 1 1 1
2 1 1 1 ...
## $ outils_exercice : Factor w/ 2 levels "Non","Oui": 2 2 2 1 2 2
2 2 1 2 ...
## $ sciences_donnees : Factor w/ 3 levels "Inintéressant",...: 2 2
2 2 2 3 2 2 2 2 ...
## $ informatique : Factor w/ 3 levels "Inintéressant",...: 2 2
3 2 3 2 2 2 2 2 ...
## $ sciences_physiques : Factor w/ 3 levels "Inintéressant",...: 2 3
2 2 2 2 3 2 3 3 ...
## $ business : Factor w/ 3 levels "Inintéressant",...: 2 3
3 3 1 2 3 1 1 1 ...
## $ arts : Factor w/ 3 levels "Inintéressant",...: 1 3
2 1 1 3 2 1 3 3 ...
## $ langues : Factor w/ 3 levels "Inintéressant",...: 1 3
2 1 1 2 1 1 3 3 ...
## $ sciences_sociales : Factor w/ 3 levels "Inintéressant",...: 1 3
3 1 1 2 1 1 1 1 ...
## $ background : Factor w/ 4 levels "Autre école",...: 3 3 4
3 3 3 3 3 3 3 ...
```

## Code R pour l'ACM :

Dans le code R ci-dessous, l'ACM est effectuée uniquement sur les individus / variables actifs:

```
# On va garder 10 composantes principale pour cette étude.  
res.mca <- MCA (data_suivi.active, graph = FALSE, ncp = 10)
```

## Visualisation et interprétation :

La proportion des variances retenues par les différentes dimensions (axes) peut être extraite à l'aide de la fonction `get_eigenvalue()` [factoextra package] comme suit:

```
eig.val <- get_eigenvalue(res.mca)  
eig.val[1:15,]  
  
##      eigenvalue variance.percent cumulative.variance.percent  
## Dim.1  0.23010049         8.785655          8.785655  
## Dim.2  0.19117826         7.299534         16.085189  
## Dim.3  0.14771893         5.640177         21.725366  
## Dim.4  0.12004518         4.583543         26.308909  
## Dim.5  0.11490720         4.387366         30.696275  
## Dim.6  0.10038344         3.832822         34.529098  
## Dim.7  0.09169245         3.500985         38.030082  
## Dim.8  0.08762522         3.345690         41.375773  
## Dim.9  0.08655558         3.304850         44.680622  
## Dim.10 0.08261092         3.154235         47.834857  
## Dim.11 0.07949799         3.035378         50.870235  
## Dim.12 0.07400658         2.825706         53.695940  
## Dim.13 0.06953205         2.654860         56.350801  
## Dim.14 0.06662778         2.543970         58.894770  
## Dim.15 0.06448738         2.462245         61.357016
```

Les dimensions sont ordonnées de manière décroissante et listées en fonction de la quantité de variance expliquée. La dimension 1 explique la plus grande variance, suivie de la dimension 2 et ainsi de suite.

Le pourcentage cumulé expliqué est obtenu en ajoutant les proportions successives de variances expliquées pour obtenir le total courant. Environ 47.83% de la variance totale est expliquée par les dix premières dimensions.

Les valeurs propres peuvent être utilisées pour déterminer le nombre d'axes à retenir. Il n'y a pas de «règle générale» pour choisir le nombre de dimensions à conserver pour l'interprétation des données. Par exemple, nous optons pour 50.16% de la variance totale expliquée, soit dix Composante principale. D'ailleurs "la part de variance expliquée dans une ACM est très souvent faible, il n'est pas nécessaire d'avoir la même démarche de conservation des dimensions que pour une ACP. Ceci est purement calculatoire, à cause de la nature qualitative des variables : les contributions sont diluées entre les différentes modalités ce qui entraîne des parts d'inertie faibles." (Romain Billot, Associate Professor

Institut Mines-Télécom Atlantique Bretagne-Pays de la Loire (ex Télécom Bretagne)  
Département LUSI ).

Nous allons opter pour les dix premières composantes principales.

## Graphique des variables :

### Résultats :

Les différents composants peuvent être consultés comme suit:

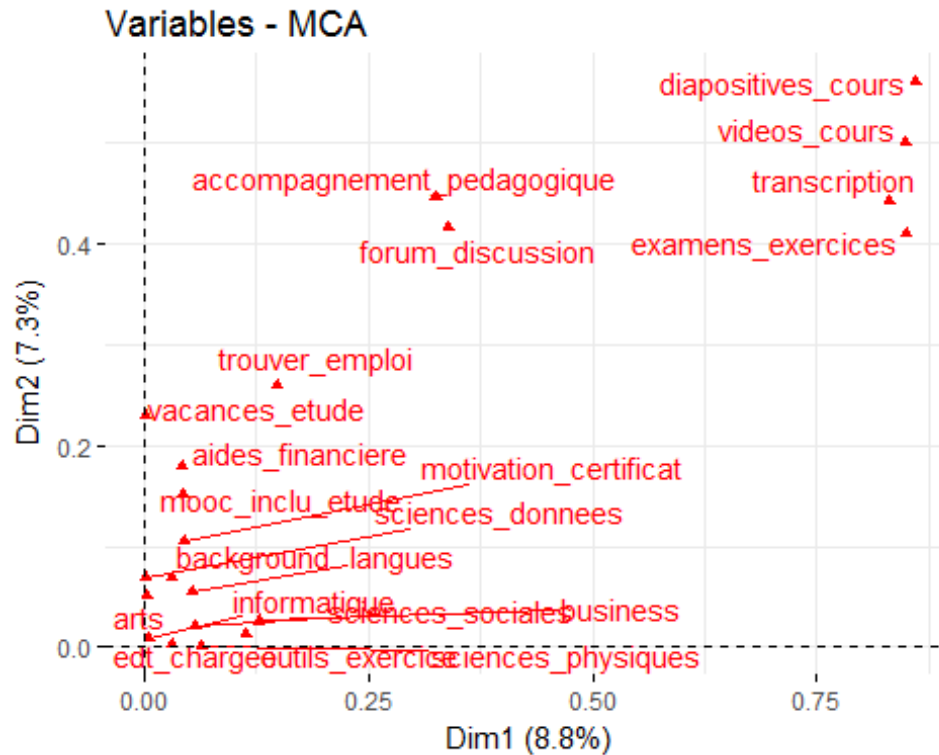
```
# Coordonnées
#print('Les coordonnées :')
#head(var$coord)
# Cos2: qualité de représentation
#print('La qualité de représentation :')
#head(var$cos2)
# Contributions aux axes
#print('La contribution des variables :')
#head(var$contrib)
```

Dans cette section, nous décrirons comment visualiser uniquement les catégories des variables. Ensuite, nous mettrons en évidence les catégories en fonction soit de leurs qualités de représentation, soit de leurs contributions aux dimensions.

*Corrélation entre les variables et les axes principaux :*

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





+ Le graphique ci-dessus permet d'identifier les variables les plus corrélées avec chaque axe. Les corrélations au carré entre les variables et les axes sont utilisées comme coordonnées.

- On constate que les variables Diapositives\_cours, videos\_cours, transcription, examens\_exercices, outils exercice sont les plus corrélées avec la dimension 1. De même, les variables accompagnement\_pedagogiques et forum\_discussion, trouver\_emploi, vacances\_etude sont les plus corrélées avec la dimension 2.

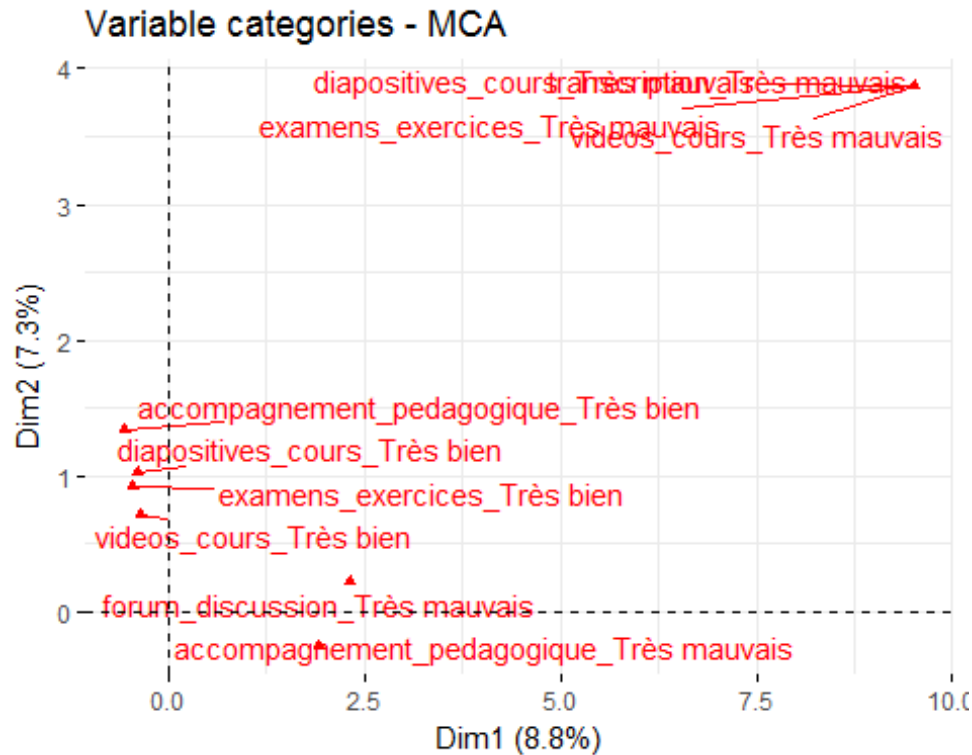
#### Coordonnées des catégories des variables :

Le code R ci-dessous affiche les coordonnées de chacune des catégories des variables dans les dimensions (1, 2, 3, 4, et 5):

```
#head(round(var$coord, 2), 4)
```

Utilisez la fonction `fviz_mca_var()` [factoextra] pour visualiser uniquement les catégories des variables: #### dimensions 1 et 2 :

```
fviz_mca_var (res.mca,
              repel = TRUE,
              axes=c(1,2),
              ggtheme = theme_minimal (),select.var = list(cos2=0.25))
```

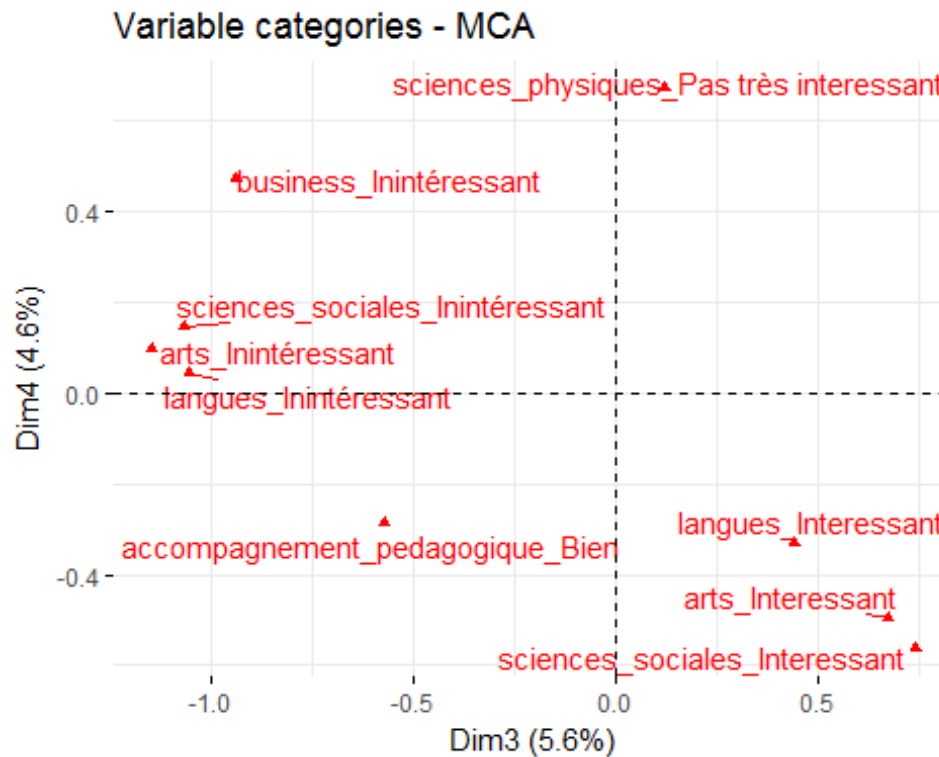


Le graphique ci-dessus montre les relations entre les catégories des variables. Il peut être interprété comme suit:

- Les catégories avec un profil similaire sont regroupées ( exemple : diapositive\_cours\_Très mauvais, videos\_cours\_Très\_mauvais).
- La distance entre les catégories et l'origine mesure la qualité des catégories. Les points qui sont loin de l'origine sont bien représentés par l'ACM( exemple : diapositive\_cours\_Très mauvais, accompagnement\_pedagogiques\_Très bien). ==> Les dimensions 1 et 2 expliquent les outils des cours ( examens\_exercices, forum\_discussion, ...). On voit très bien l'avis des sondé sur les outils utilisés lors des MOOCs ainsi que l'accompagnement pédagogiques. Ceci peut dire que les plateforme doivent faire attention à ces élément le succès des inscrits.

#### dimensions 3 et 4 :

```
fviz_mca_var (res.mca,
              repel = TRUE,
              axes=c(3,4),
              ggtheme = theme_minimal (),select.var = list(cos2=0.22))
```



==> Ces dimensions

couvrent les domaines choisis par les sondés. On voit très bien l'existence des groupes se trouvant dans des côtés opposé à l'origine, qui est indicateur qu'ils sont négativement corrélés.

#### Qualité de représentation des catégories des variables :

Les deux dimensions 1 et 2 capturent 16.08% de l'inertie totale (variation) contenue dans les données. Tous les points ne sont pas aussi bien représentés par les deux dimensions.

La qualité de représentation, appelée cosinus carré ( $\cos^2$ ), mesure le degré d'association entre les catégories des variables et les dimensions. Le  $\cos^2$  peut être extrait comme suit:

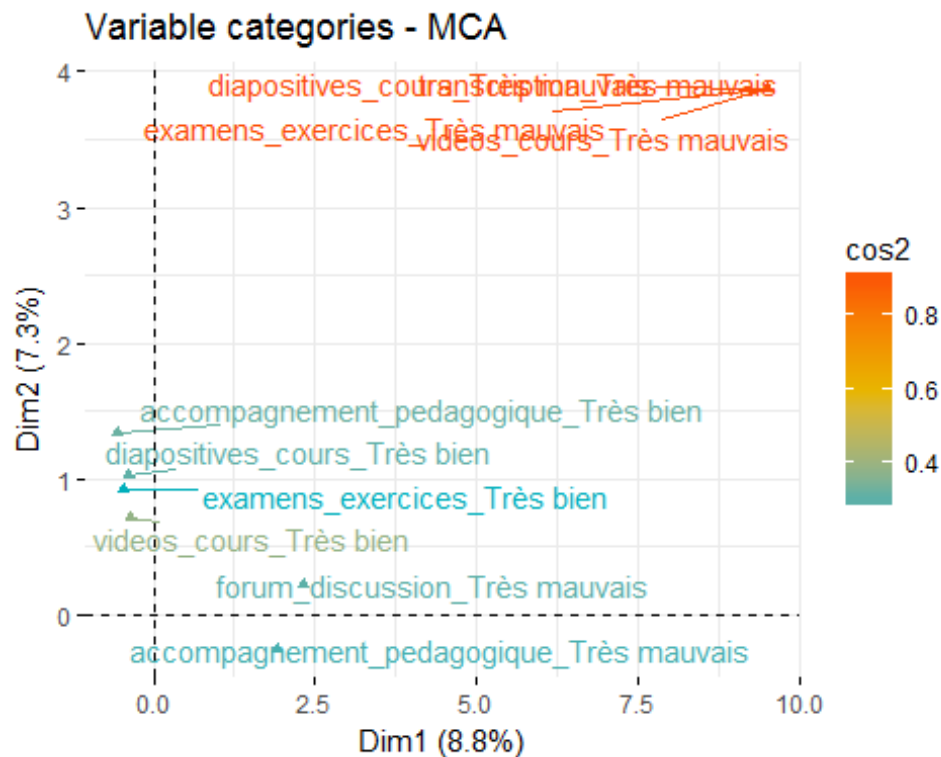
```
#head(var$cos2, 4)
```

Si une catégorie d'une variable donnée est bien représentée par deux dimensions, la somme des  $\cos^2$  est proche de 1. Pour certains éléments, plus de 2 dimensions sont nécessaires pour représenter parfaitement les données.

Il est possible de colorer les variables en fonction de la valeur de leur  $\cos^2$  à l'aide de l'argument `col.var = "cos2"`. Cela produit un gradient de couleurs. Dans ce cas, l'argument `gradient.cols` peut être utilisé pour spécifier une palette de couleur personnalisée. Par exemple, `gradient.cols = c("white", "blue", "red")` signifie que:

- les variables à faible valeur de  $\cos^2$  seront colorées en "white" (blanc)
- les variables avec des valeurs moyennes de  $\cos^2$  seront colorées en "blue" (bleu)
- les variables avec des valeurs élevées de  $\cos^2$  seront colorées en "red" (rouge)

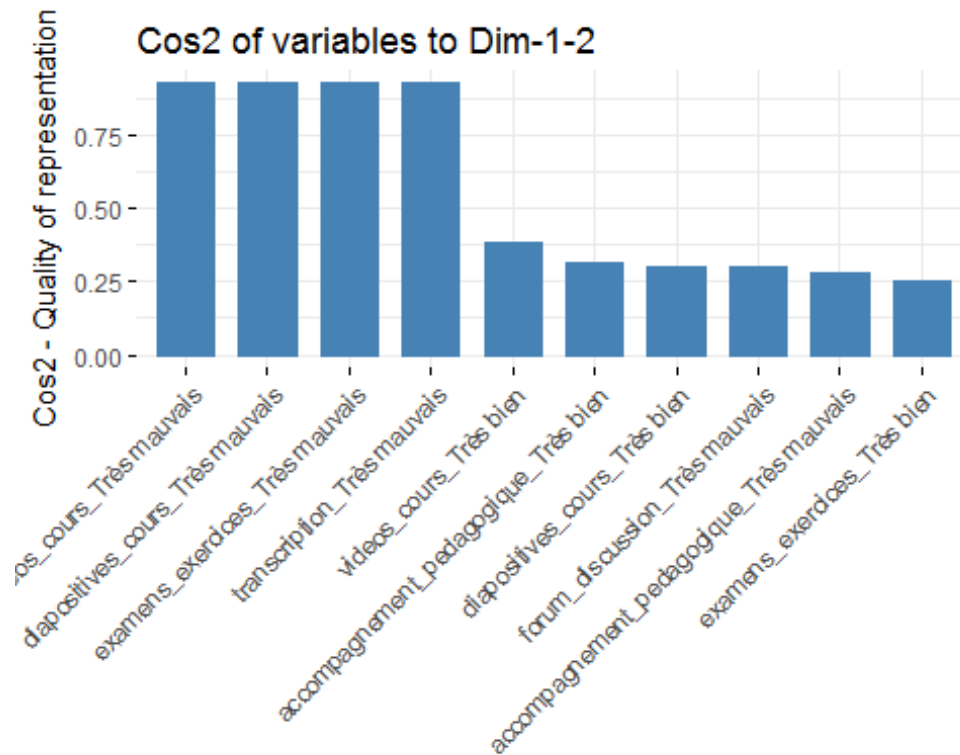
```
# Colorer en fonction du cos2
fviz_mca_var(res.mca, col.var = "cos2",
  gradient.cols = c("#00AFBB", "#E7B800", "#FC4E07"),
  repel = TRUE,
  select.var = list(cos2=0.25),
  ggtheme = theme_minimal())
```



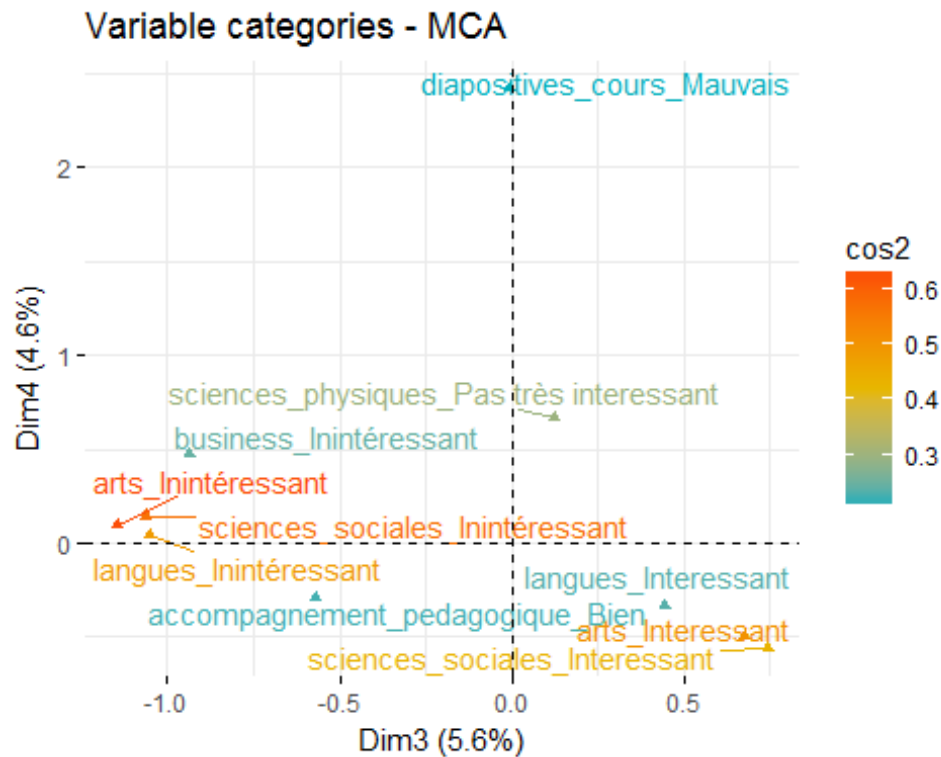
Ainsi les catégories des variables en relation avec la gestion des cours en lignes ( examens, ..) sont très bien représentées sur les dimensions 1 et 2.

Il est également possible de créer un barplot du cos2 des variables avec la fonction `fviz_cos2()` [factoextra]:

```
# Cos2 des variable sur Dim.1 et Dim.2
fviz_cos2(res.mca, choice = "var", axes = 1:2, top = 10)
```



```
# Colorer en fonction du cos2
fviz_mca_var(res.mca, col.var = "cos2",
  gradient.cols = c("#00AFBB", "#E7B800", "#FC4E07"),
  repel = TRUE,
  axes=c(3,4),
  select.var = list(cos2=0.20),
  ggtheme = theme_minimal())
```



==> ainsi les domaines des moocs sont très bien représentées dans les dimensions 3 et 4.

#### Contribution des variables aux dimensions :

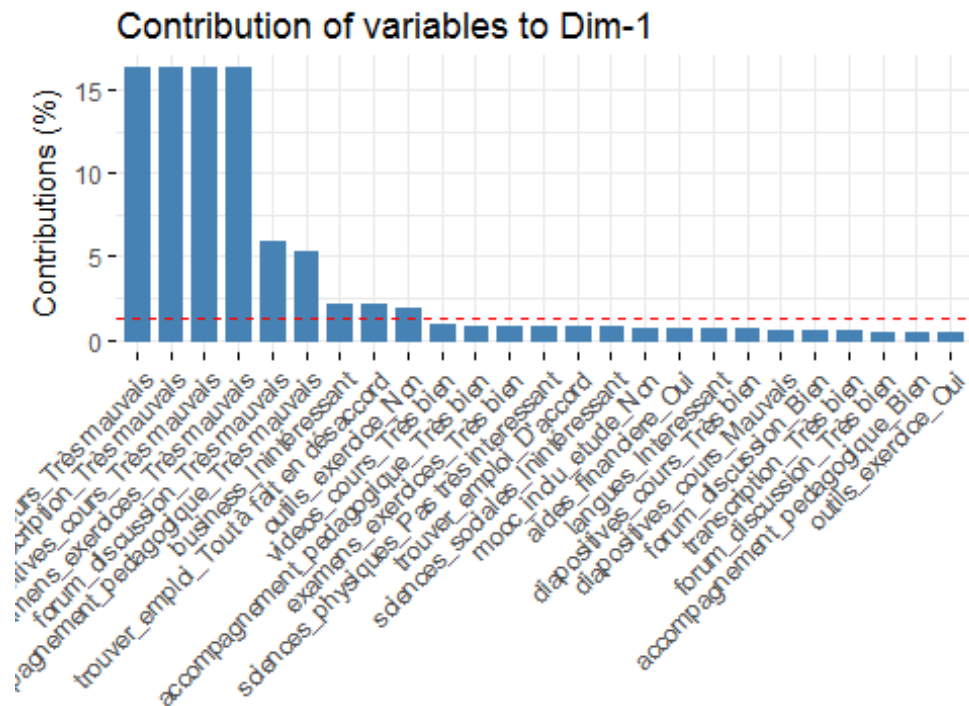
La contribution des variables (en %) à la définition des dimensions peut être extraite comme suit:

```
#head(round(var$contrib,2), 4)
```

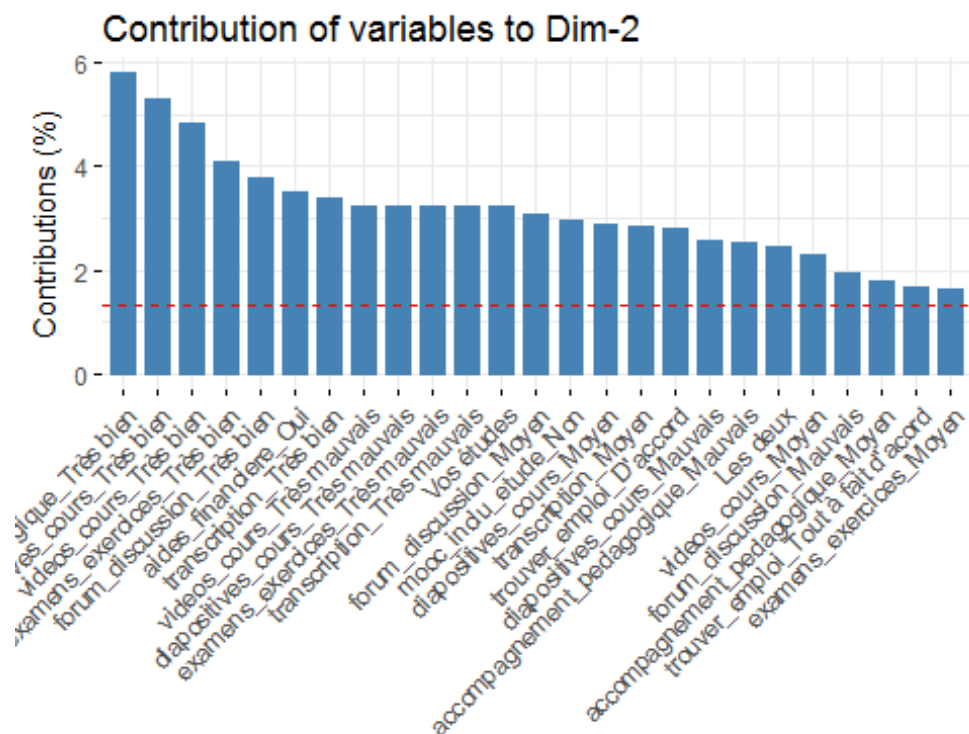
Les variables avec les plus grandes valeurs, contribuent le mieux à la définition des dimensions. Les catégories qui contribuent le plus à Dim.1 et Dim.2 sont les plus importantes pour expliquer la variabilité dans le jeu de données.

La fonction `fviz_contrib()` [factoextra] peut être utilisée pour faire un barplot de la contribution des catégories des variables. Le code R ci-dessous montre le top 25 des catégories contribuant aux dimensions:

```
# Contributions des variables à la dimension 1
fviz_contrib (res.mca, choice = "var", axes = 1, top = 25)
```



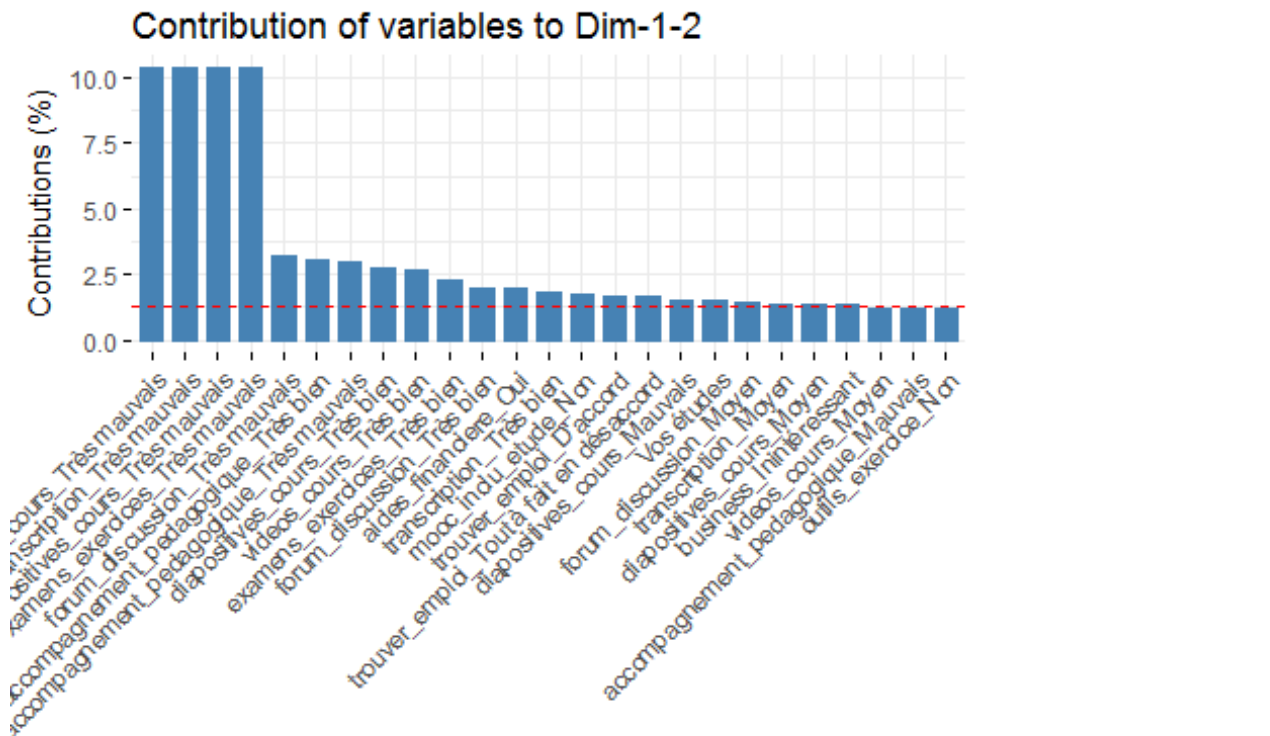
```
# Contributions des variables à la dimension 2
fviz_contrib (res.mca, choice = "var", axes = 2, top = 25)
```



Les contributions totales aux dimensions 1 et 2 sont obtenues comme suit:

```
# Contribution totale aux dimensions 1 et 2
fviz_contrib(res.mca, choice = "var", axes = 1:2, top = 25)
```

```
fviz_contrib(res.mca, choice = "var", axes = 1:2, top = 25)
```



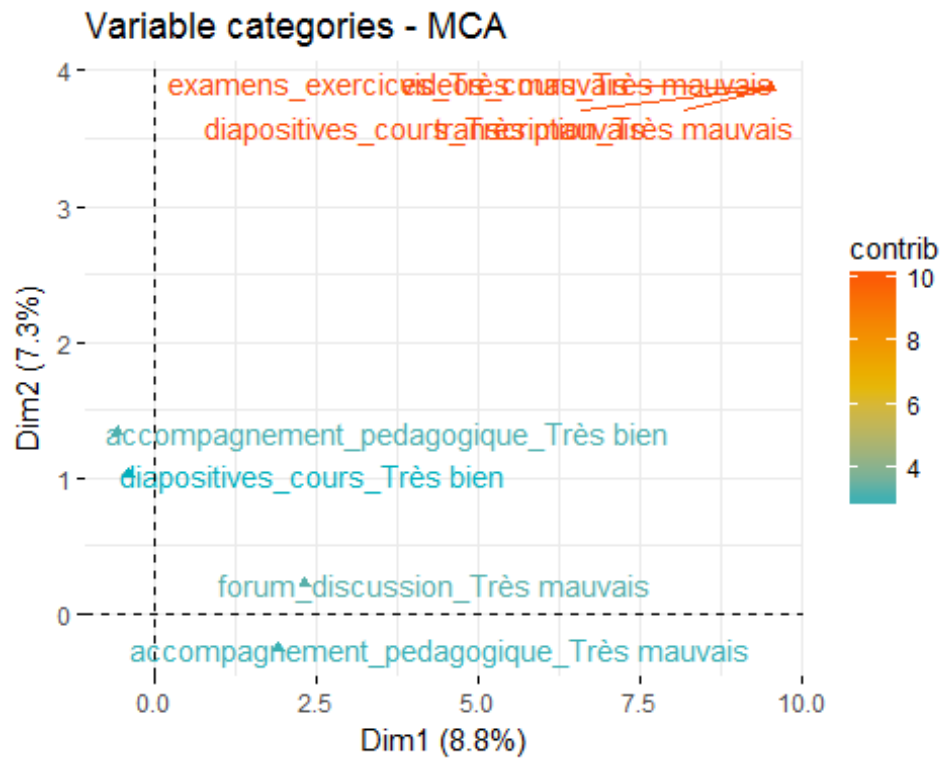
La ligne en pointillé rouge, sur le graphique ci-dessus, indique la valeur moyenne attendue sous l'hypothèse nulle.

On peut voir que:

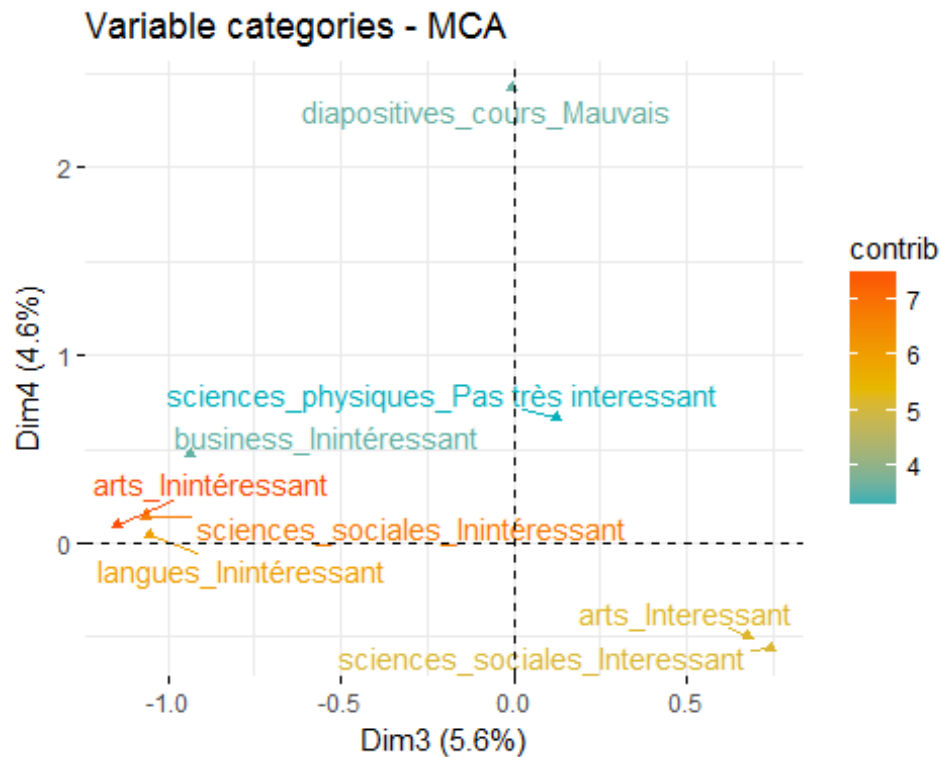
les catégories diapositive\_cours\_Très mauvais, transcription\_Très mauvais, diapositives\_cours\_très mauvais sont les plus importantes dans la définition de la première dimension. Les accompagnement\_pédagogiques Très bien, diapositives\_cours\_très Bien, vidéos\_cours\_TrèsBien contribuent le plus à la dimension 2 Les catégories les plus importantes peuvent être mises en évidence sur le graphique comme suit:

```
fviz_mca_var(res.mca, col.var = "contrib",
              gradient.cols = c("#00AFBB", "#E7B800", "#FC4E07"),
              repel = TRUE,
              select.var = list(contrib=7.5),
              ggtheme = theme_minimal()
            )
```





```
fviz_mca_var(res.mca, col.var = "contrib",
  gradient.cols = c("#00AFBB", "#E7B800", "#FC4E07"),
  repel = TRUE,
  axes=c(3,4),
  select.var = list(contrib=7.5),
  ggtheme = theme_minimal()
)
```



*Graphique des individus :*

## Résultats

La fonction `get_mca_ind()` [factoextra] sert à extraire les résultats pour les individus. Cette fonction renvoie une liste contenant les coordonnées, la cos2 et les contributions des individus:

```
ind <- get_mca_ind (res.mca)
ind

## Multiple Correspondence Analysis Results for individuals
## =====
##   Name      Description
## 1 "$coord"   "Coordinates for the individuals"
## 2 "$cos2"    "Cos2 for the individuals"
## 3 "$contrib" "contributions of the individuals"
```

Pour accéder aux différents composants, utilisez ceci:

```
print(' Coordonnées')

## [1] " Coordonnées"

#head(ind$coord)
print(' Qualité de representation')

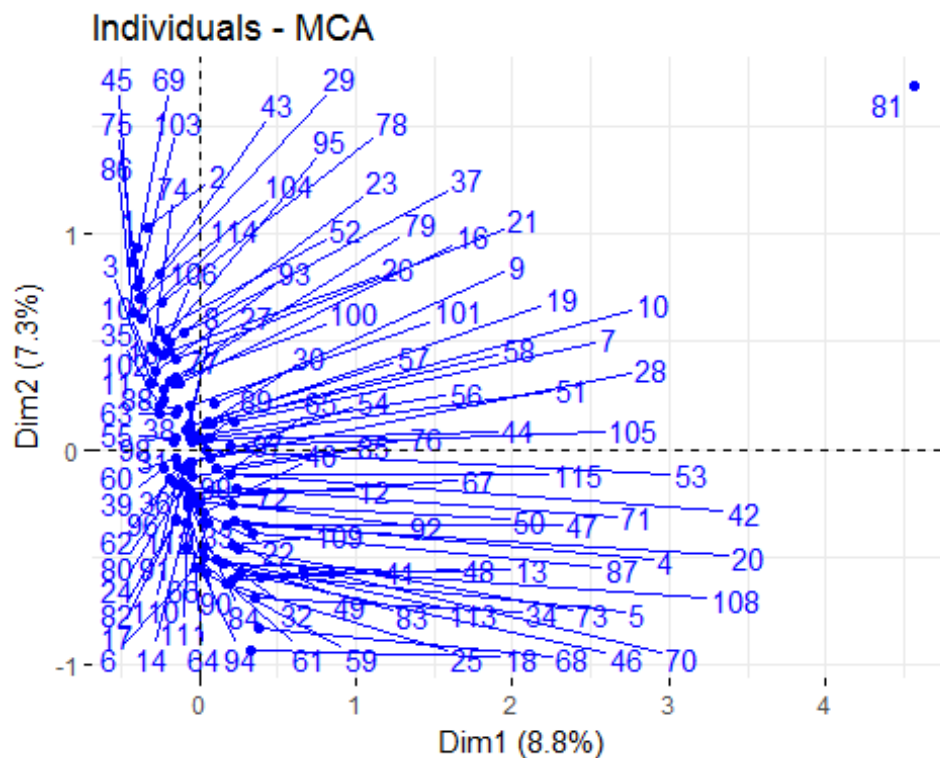
## [1] " Qualité de representation"
```

```
#head(ind$cos2)
print('Contributions')
## [1] "Contributions"
#head(ind$contrib)
```

Graphique: qualité et contribution

Utilisez la fonction `fviz_ca_row()` [dans `factoextra`] pour visualiser uniquement les points lignes: #### dimension 1 et 2 :

```
fviz_mca_ind(res.mca, repel = TRUE, axes = c(1,2))
```



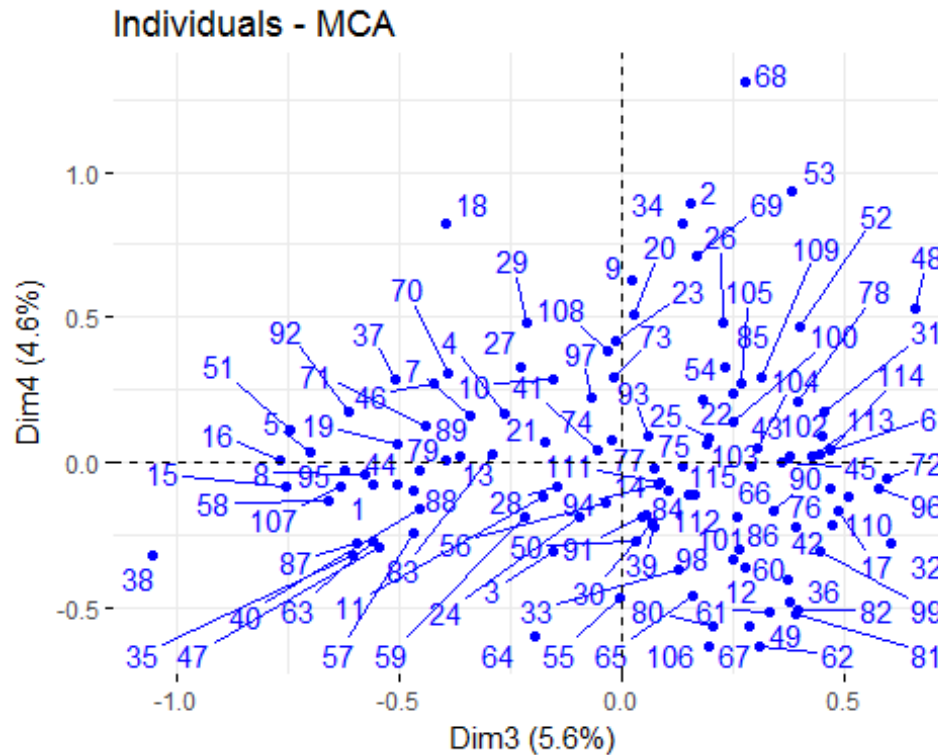
Le graphique ci-dessus montre les relations entre les points lignes:

Les lignes avec un profil similaire sont regroupées. Les lignes corrélées négativement sont positionnées sur des côtés opposés de l'origine de du graphique (quadrants opposés). La distance entre les points lignes et l'origine mesure la qualité des points lignes sur le graphique. Les points lignes qui sont loin de l'origine sont bien représentés sur le graphique.

à partir du graphe ci-dessus on peut déjà remarquer l'existence d'un individu aberrant (81). cet individu devra être utilisé comme supplémentaire car il risque de fausser l'étude.

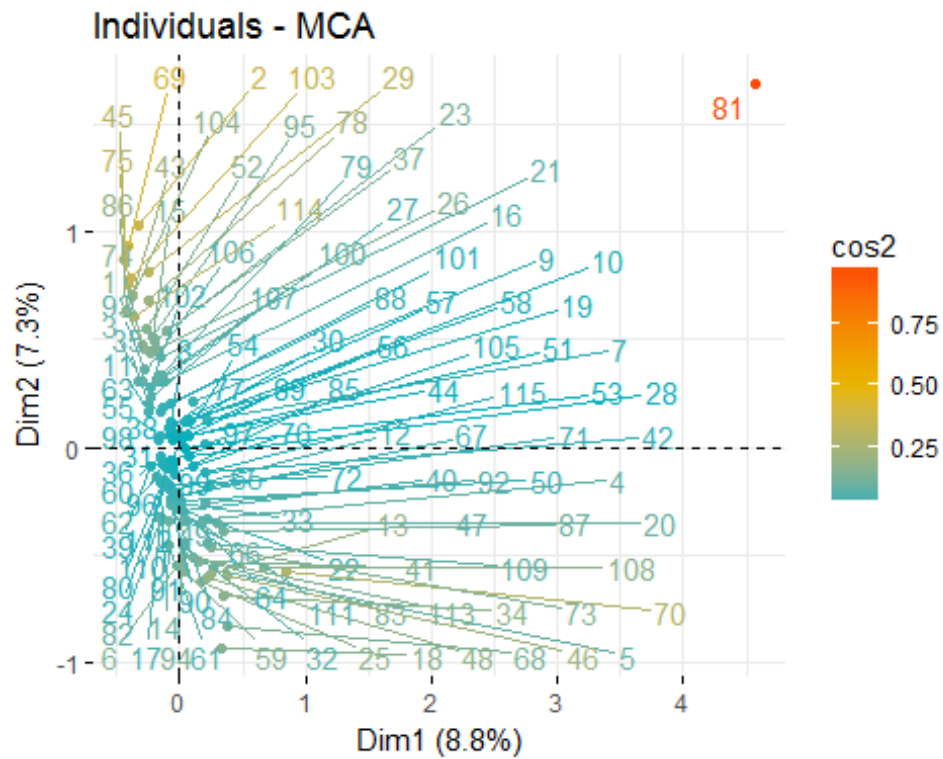
**dimension 3 et 4 :**

```
fviz_mca_ind(res.mca, repel = TRUE, axes = c(3,4))
```

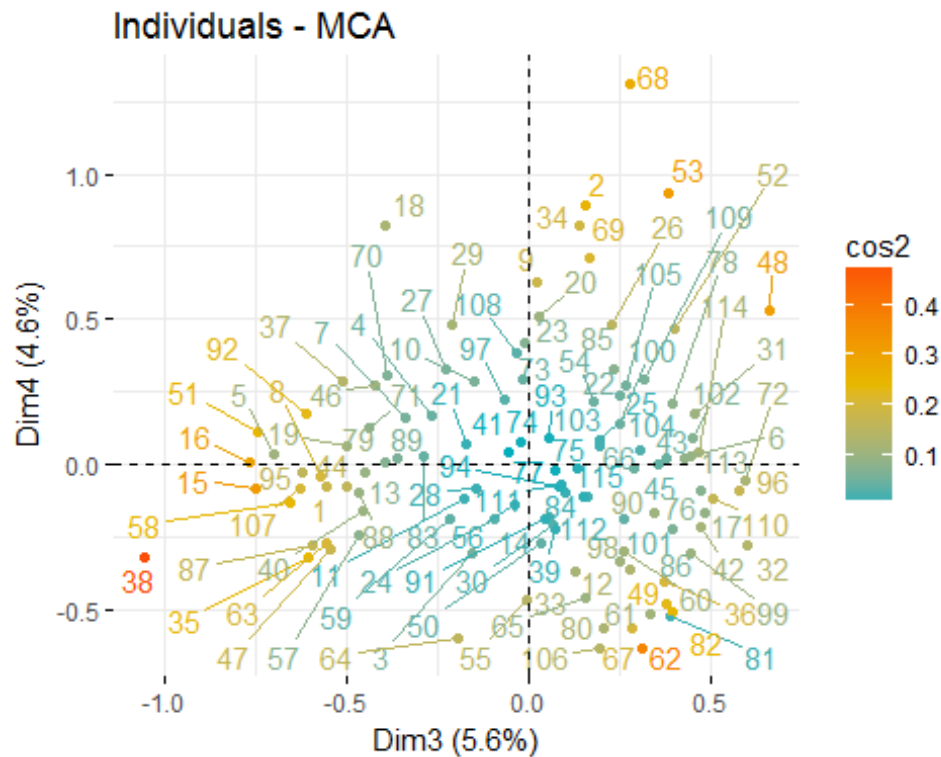


ces deux dimension  
 permet très bien de séparer les individus (sondés) . La fonction `fviz_mca_ind()` [factoextra] sert à visualiser uniquement des individus. Comme les variables, il est également possible de colorer les individus en fonction de leurs `cos2`:

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



```
fviz_mca_ind(res.mca, col.ind = "cos2",
  gradient.cols = c("#00AFBB", "#E7B800", "#FC4E07"),
  repel = TRUE,
  axes=c(3,4),
  ggtheme = theme_minimal())
```

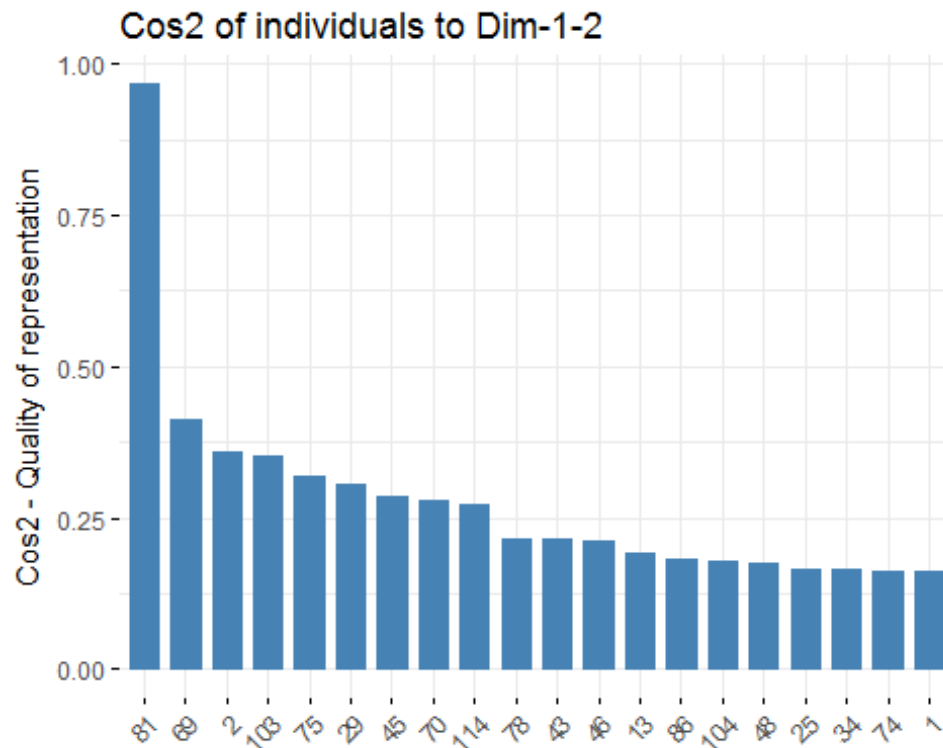


Le graphique ci-

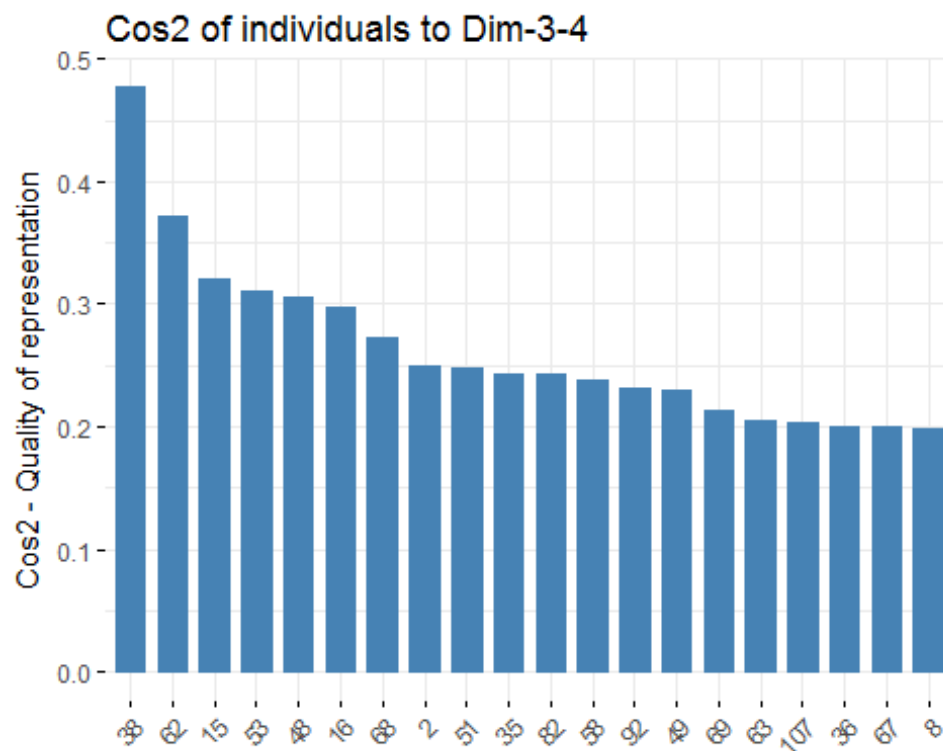
dessus montre les relations entre les points lignes:

Les lignes avec un profil similaire sont regroupées. Les lignes corrélées négativement sont positionnées sur des côtés opposés de l'origine de du graphique (quadrants opposés). La distance entre les points lignes et l'origine mesure la qualité des points lignes sur le graphique. Les points lignes qui sont loin de l'origine sont bien représentés sur le graphique. Il est également possible de créer un bar plot du  $\cos^2$  des lignes en utilisant la fonction `fviz_cos2()` [factoextra]:

```
# Cos2 des lignes sur Dim.1 et Dim.2
fviz_cos2(res.mca, choice = "ind", axes = 1:2, las=2, top = 20)
```



```
# Cos2 des lignes sur Dim.1 et Dim.2
fviz_cos2(res.mca, choice = "ind", axes = 3:4, las=2, top = 20)
```



### *Conclusion:*

Cette partie nous a permet de faire une étude préliminaire afin de trouver les informations utiles pour l'étude finale.



## ACM finale :

L'Analyse des Correspondances Multiples (ACM ou MCA pour multiple correspondence analysis) est une extension de l'analyse factorielle des correspondances pour résumer et visualiser un tableau de données contenant plus de deux variables catégorielles. On peut aussi la considérer comme une généralisation de l'analyse en composantes principales lorsque les variables à analyser sont catégorielles plutôt que quantitatives (source STHDA, Abdi and Williams 2010).

L'étude faite ci-dessous est le fruit d'une autre ACM faite précédemment. En effet, une analyse d'ACM sera pertinente si on arrive à trouver les variables et individus pertinents. Pour en savoir plus sur la première étude voir annexes.

L'objectif de cette étude est d'identifier des groupes de personnes qui partagent un profil similaire envers les cours en ligne.

les données que lesquelles on va faire l'ACM sont :

- Individus actifs (lignes 1:115 sauf 81): individus qui sont utilisés dans l'ACM.
- Variables actives (toutes les colonnes sauf les colonnes numériques+ plateforme\_utilisee+ a\_suivi+sexe, est\_etudiant, formation, niveau\_actuel): variables utilisées dans l'ACM.
- Variables supplémentaires(sexe, age, satisfaction, nbr\_certificats, temps\_alloue\_cours): elles ne participent pas à l'ACM. Les coordonnées de ces variables seront prédites.

Nous commençons par extraire les individus actifs et les variables actives pour l'ACM:

### Données :

```
final_data<- data_suivi[, c("motivation_certificat", "trouver_emploi", "vacances_etude", "mooc_inclu_etude", "edt_chargee", "videos_cours", "transcription", "diapositives_cours", "forum_discussion", "examens_exercices", "accompagnement_pedagogique", "aides_financiere", "outils_exercice", "sciences_donnees", "informatique", "sciences_physiques", "business", "arts", "langues", "sciences_sociales", "background", "sexe", "a_suivi", "age", "prix", "satisfaction", "temps_alloue_cours", "nbr_certificat")]
```

### ACM :

```
res.mca <-MCA(final_data[-c(81),],  
              quanti.sup = c(24,25,26,27,28),  
              quali.sup = c(22,23),  
              graph = FALSE,  
              ncp = 10)
```

J'ai éliminé l'individu 81 car il fausse l'étude. Ceci a été prouvé lors de la première étude. c'était un individu aberrant, chose qui a perturbé fortement les résultats. ###Résultats :  
#### Les valeurs propres : On affiche les 10 premières valeurs propres:

```
eig.val <- get_eigenvalue(res.mca)
eig.val[1:10,]

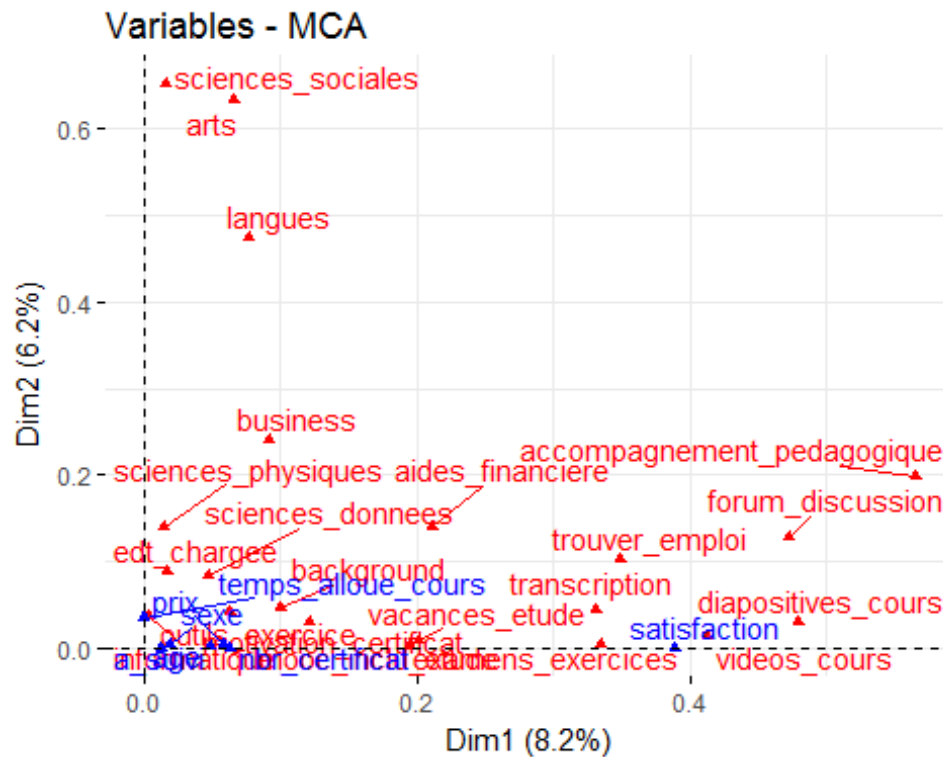
##      eigenvalue variance.percent cumulative.variance.percent
## Dim.1  0.19841592      8.170067      8.170067
## Dim.2  0.15002409      6.177463     14.347530
## Dim.3  0.12365137      5.091527     19.439057
## Dim.4  0.11590180      4.772427     24.211484
## Dim.5  0.10308801      4.244800     28.456284
## Dim.6  0.09298705      3.828879     32.285163
## Dim.7  0.08946812      3.683981     35.969145
## Dim.8  0.08745840      3.601228     39.570373
## Dim.9  0.08359114      3.441988     43.012361
## Dim.10 0.07972941      3.282976     46.295337
```

Les valeurs propres peuvent être utilisées pour déterminer le nombre d'axes à retenir. Il n'y a pas de méthode générale pour choisir le nombre de dimensions à garder pour l'interprétation des données. D'ailleurs la part de variance expliquée dans une ACM est très souvent faible. Ceci est dû au fait que les contributions des variables sont diluées entre les différentes modalités ce qui entraîne des parts d'inertie faibles. Dans notre cas c'est très probable car on a utilisé des variables à plusieurs modalités.

Nous allons étudier spécialement les deux premières composantes principales car ils présentent des résultats intéressants pour la variable satisfaction.

### Corrélation entre les variables et les axes principaux :

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

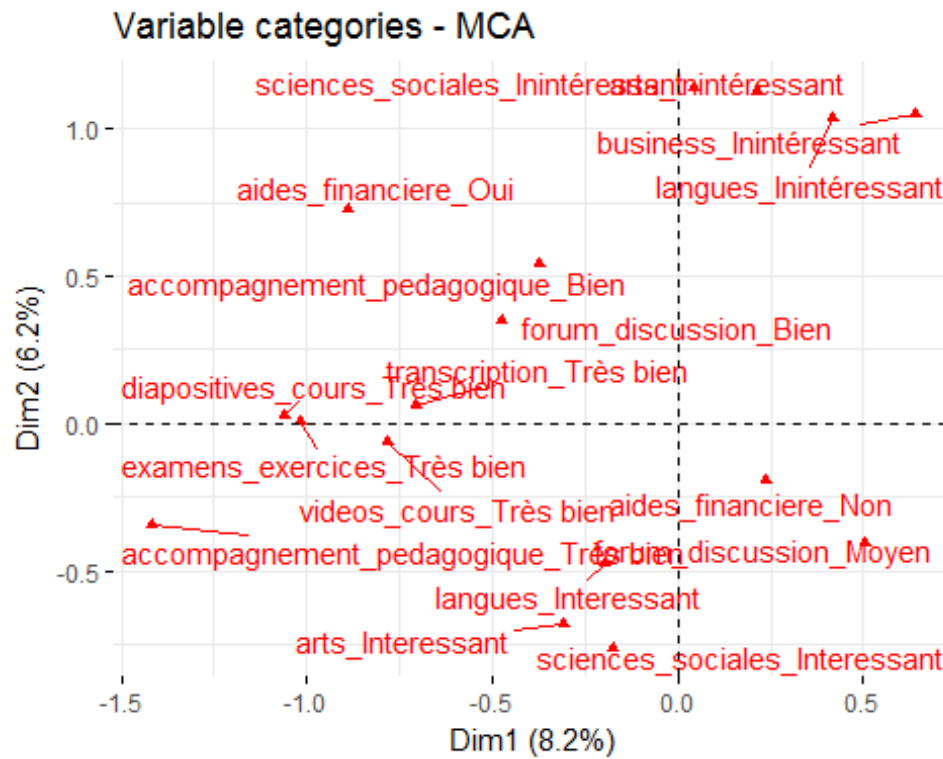


- Les outils de cours

incluant les diapositives, les examens et exercices, les forums de discussions et l'accompagnement pédagogiques sont fortement corrélés avec la dimension 1. - Tandis que les domaines relatifs aux MOOCs ( business, sciences sociales, arts, langues) sont fortement corrélés avec la dimension 2.

On reviendra sur les parties à venir pour l'interprétation des variables supplémentaire.  
 ###Graphique: catégories des variables: On va choisir que les catégories qui soient bien représentées dans les premiers axes.

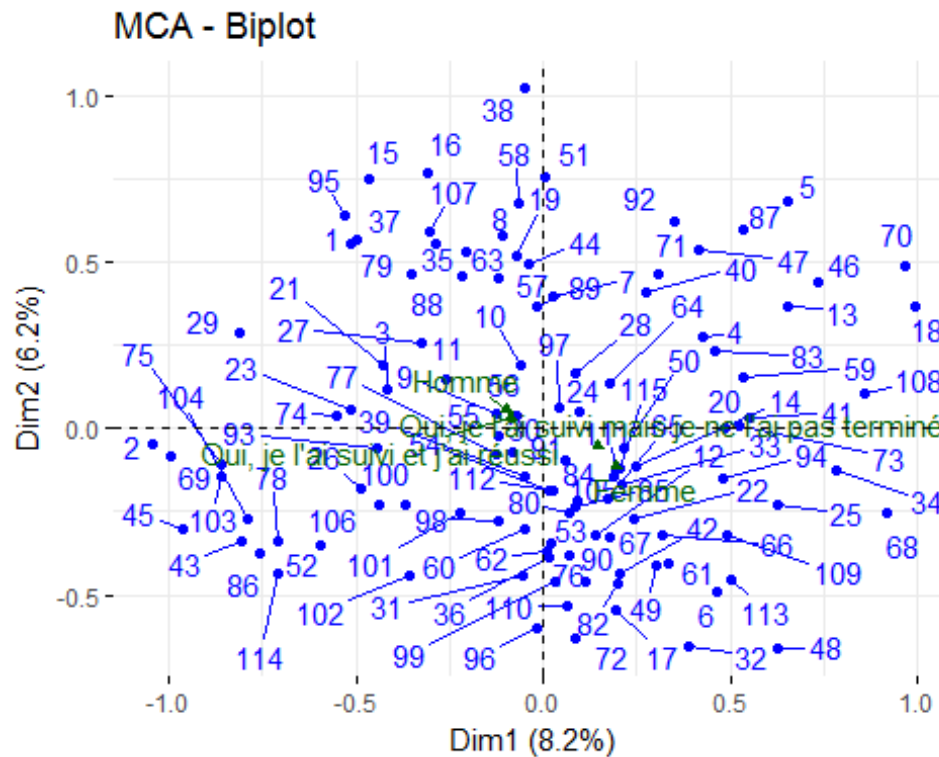
```
fviz_mca_var(res.mca, repel = TRUE,
             ggtheme= theme_minimal(),select.var = list(cos2=0.2))
```



###Graphique :

individus et variables qualitatives supplémentaires :

```
# Biplot des individus et des variables
fviz_mca_biplot (res.mca, repel = TRUE,
  ggtheme = theme_minimal (),
  axes=c(1,2),invisible = "var"
)
```

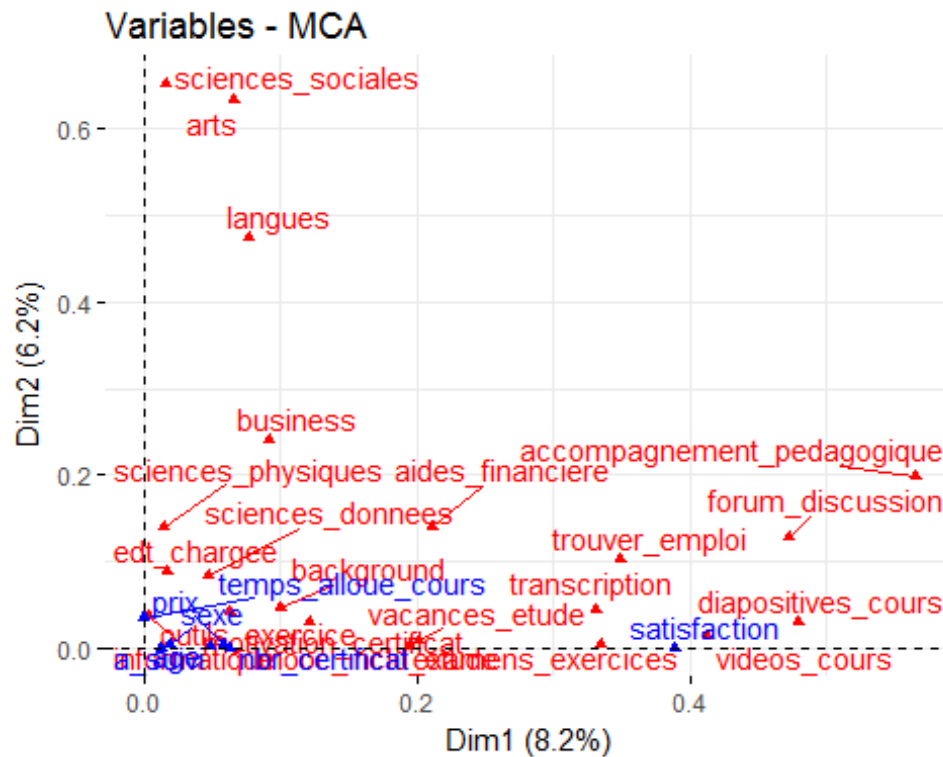


Analyse et critiques

de deux graphiques ci-dessus : - Les sondés ayant suivi et réussi leurs MOOCs sont généralement des hommes qui ont trouvé les outils de cours intéressants. - Les sondés ayant suivi mais pas réussi leurs MOOCs sont généralement des femmes qui ont trouvé les outils de cours inintéressants. .

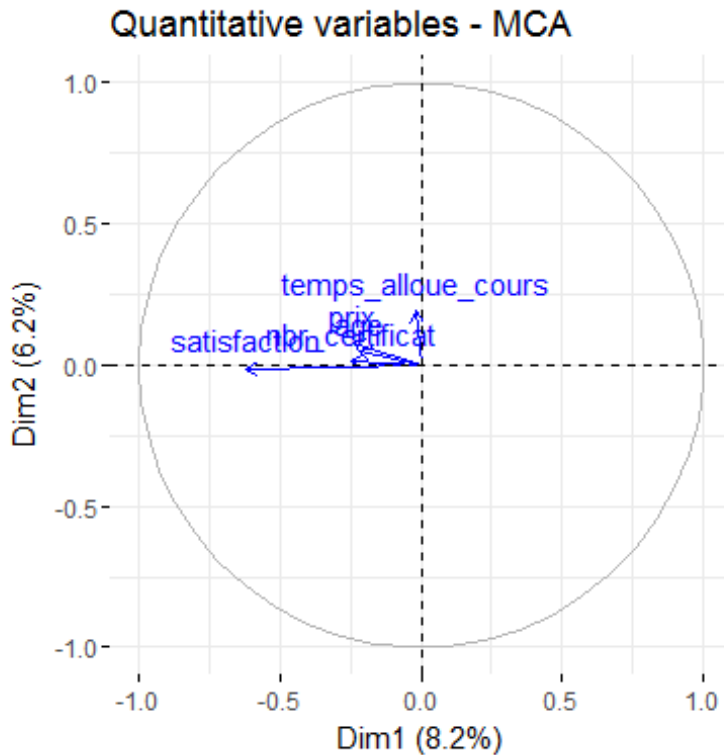
Revenons maintenant à l'interprétation des variables supplémentaires : ###Graphique: variables actives et variables supplémentaires quantitatives:

```
fviz_mca_var (res.mca, choice = "mca.cor", axes = c(1,2),
              repel = TRUE)
```



La satisfaction est très corrélée avec la dimension 1. Or la dimension 1 est liée principalement aux outils de cours. Donc on peut dire la satisfaction est lié principalement à la manière dont les cours sont faits et gérés ( les diapositifs, les examens, les exercices, les vidéos, accompagnement pédagogique etc). pour confirmer nos dires, voici le graphique suivant :

```
fviz_mca_var(res.mca, choice = "quanti.sup",
              ggtheme = theme_minimal())
```



Vous voyez très bien que la satisfaction est négativement corrélée avec la dimension 1, ce qui est très normal car la dimension 1 est négativement corrélée avec toute sorte de catégorie comme : accompagnement pédagogique bien, forum discutait bien, transcription très bien, diapositives cours, etc.

On tient à préciser qu'on a choisi de travailler seulement avec deux dimensions vu qu'elles présentent des résultats intéressants.