

ACM : Analyse des Correspondances Multiples (MCA pour multiple correspondence analysis)

DJEBALI Wissam

October 12, 2018

```
library(ade4)
library(FactoMineR)
library(factoextra)
library(corrplot)
```

Rappel

Il existe plusieurs techniques d'analyse factorielle dont les plus courantes sont :

- l'analyse en composante principale (ACP) portant sur des variables quantitatives,
- l'analyse factorielle des correspondances (AFC) portant sur deux variables qualitatives
- l'analyse des correspondances multiples (ACM) portant sur plusieurs variables qualitatives (il s'agit d'une extension de l'AFC).

Pour combiner des variables à la fois quantitatives et qualitatives, on pourra avoir recours à l'analyse mixte de Hill et Smith.

Bien que ces techniques soient disponibles dans les extensions standards de R, il est souvent préférable d'avoir recours à deux autres extensions plus complètes, `ade4` et `FactoMineR`, chacune ayant ses avantages et des possibilités différentes.

Voici les fonctions les plus fréquentes :

Analyse	Variables	Fonction standard	Fonction <code>ade4</code>	Fonctions <code>FactoMineR</code>
ACP	plusieurs variables quantitatives	<code>princomp</code>	<code>dudi.pca</code>	PCA
AFC	deux variables qualitatives	<code>corresp</code>	<code>dudi.coa</code>	CA
ACM	plusieurs variables qualitatives	<code>mca</code>	<code>dudi.acm</code>	MCA
Analyse mixte de Hill et Smith	plusieurs variables quantitatives et/ou qualitatives	—	<code>dudi.mix</code>	—

Principe général

L'ACM est une extension de l'AFC.

L'analyse des correspondances multiples est une technique descriptive visant à résumer l'information contenu dans un grand nombre de variables afin de faciliter l'interprétation des corrélations existantes entre ces différentes variables. On cherche à savoir quelles sont les modalités corrélées entre elles. L'ensemble des individus peut être représenté dans un espace à plusieurs dimensions où chaque axe représente les différentes variables utilisées pour décrire chaque individu. Plus précisément, pour chaque variable qualitative, il y a autant d'axes que de modalités moins un. Ainsi il faut trois axes pour décrire une variable à quatre modalités.

Un tel nuage de points est aussi difficile à interpréter que de lire directement le fichier de données. On ne voit pas les corrélations qu'il peut y avoir entre modalités, par exemple qu'aller au cinéma est plus fréquent chez les personnes habitant en milieu urbain.

Afin de mieux représenter ce nuage de points, on va procéder à un changement de systèmes de coordonnées. Les individus seront dès lors projetés et représentés sur un nouveau système d'axe. Ce nouveau système d'axes est choisit de telle manière que la majorité des variations soit concentrées sur les premiers axes. Les deux-trois premiers axes permettront d'expliquer la majorité des différences observées dans l'échantillon, les autres axes n'apportant qu'une faible part additionnelle d'information. Dès lors, l'analyse pourra se concentrer sur ses premiers axes qui constitueront un bon résumé des variations observables dans l'échantillon.

Avant toute ACM, il est indispensable de réaliser une analyse préliminaire de chaque variable, afin de voir si toutes les classes sont aussi bien représentées ou s'il existe un déséquilibre. **L'ACM est sensible aux effectifs faibles**, aussi il est préférable de regrouper les classes peu représentées le cas échéant.

```
data(poison)
head(poison[, 1:7], 3)
```

```
##   Age Time   Sick Sex   Nausea Vomiting Abdominals
## 1   9   22 Sick_y  F Nausea_y  Vomit_n    Abdo_y
## 2   5    0 Sick_n  F Nausea_n  Vomit_n    Abdo_n
## 3   6   16 Sick_y  F Nausea_n  Vomit_y    Abdo_y
```

Nos données contiennent donc des:

- Individus actifs (lignes 1:55): individus qui sont utilisés dans l'ACM.
- Variables actives (colonnes 5:15): variables utilisées dans l'ACM.
- Variables supplémentaires: elles ne participent pas à l'ACM. Les coordonnées de ces variables seront prédites.
 - Variables quantitatives supplémentaires (quanti.sup): Colonnes 1 et 2 correspondant aux colonnes age et time, respectivement.
 - Variables qualitatives supplémentaires (quali.sup: Colonnes 3 et 4 correspondant aux colonnes Sick et Sex, respectivement. Ces variables seront utilisées pour colorer les individus par groupes.

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

```
poison.active <- poison[1:55, 5:15]
head(poison.active[, 1:6], 3)
```

```
##      Nausea Vomiting Abdominals   Fever   Diarrhae   Potato
## 1 Nausea_y  Vomit_n    Abdo_y Fever_y Diarrhea_y Potato_y
## 2 Nausea_n  Vomit_n    Abdo_n Fever_n Diarrhea_n Potato_y
## 3 Nausea_n  Vomit_y    Abdo_y Fever_y Diarrhea_y Potato_y
```

Résumé

La fonction `summary()` peut être utilisée pour calculer la fréquence des catégories des variables. Comme le tableau de données contient un grand nombre de variables, nous afficherons uniquement les résultats pour les 4 premières variables.

Résumés statistiques:

```
# Résumé des 4 premières variables
summary(poison.active)[, 1:4]
```

```
##      Nausea      Vomiting      Abdominals      Fever
## Nausea_n:43 Vomit_n:33 Abdo_n:18 Fever_n:20
## Nausea_y:12 Vomit_y:22 Abdo_y:37 Fever_y:35
```

code R

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

Le résultat de l'ACM est une liste:

```
print(res.mca)
```

```
## **Results of the Multiple Correspondence Analysis (MCA)**
## The analysis was performed on 55 individuals, described by 11 variables
## *The results are available in the following objects:
##
##      name                description
## 1  "$eig"                "eigenvalues"
## 2  "$var"                "results for the variables"
## 3  "$var$coord"         "coord. of the categories"
## 4  "$var$cos2"          "cos2 for the categories"
## 5  "$var$contrib"       "contributions of the categories"
## 6  "$var$v.test"        "v-test for the categories"
## 7  "$ind"                "results for the individuals"
## 8  "$ind$coord"         "coord. for the individuals"
## 9  "$ind$cos2"          "cos2 for the individuals"
## 10 "$ind$contrib"       "contributions of the individuals"
## 11 "$call"               "intermediate results"
## 12 "$call$marge.col"    "weights of columns"
## 13 "$call$marge.li"     "weights of rows"
```

Visualiation et Interprétation

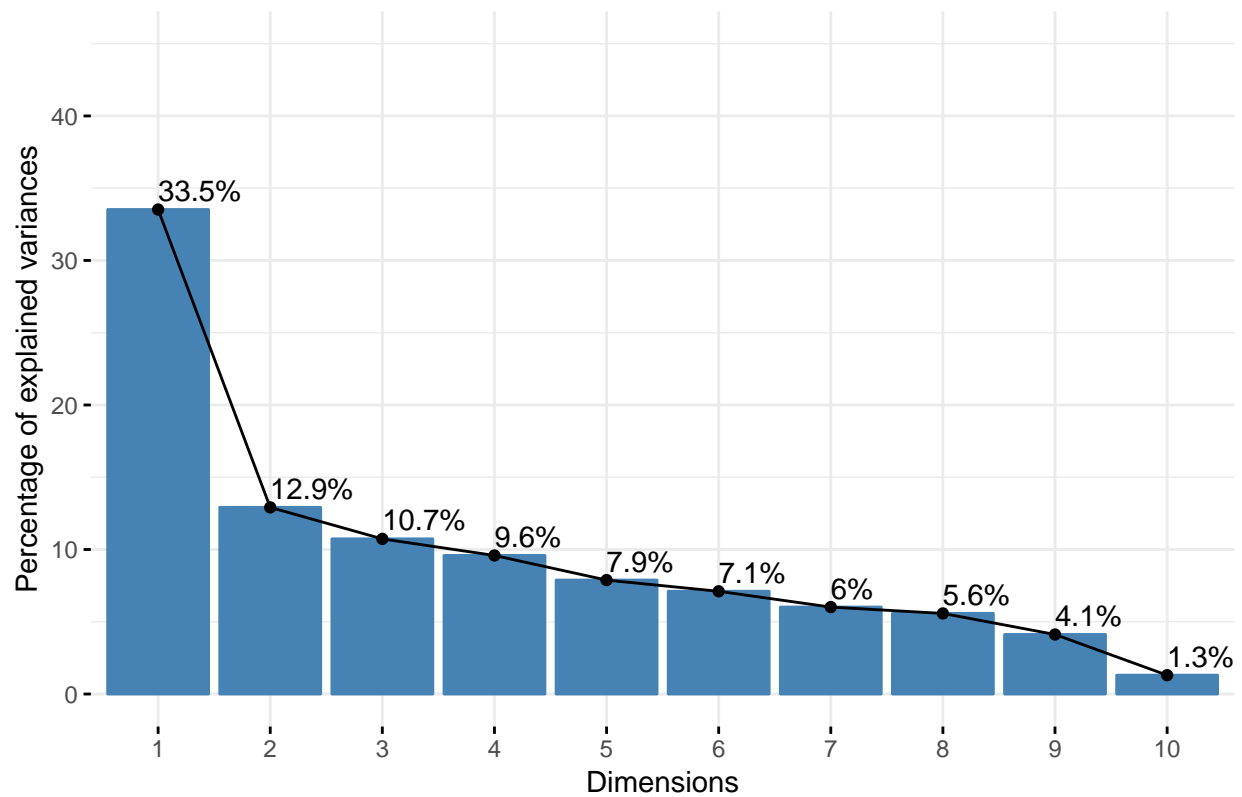
Valeurs propres / Variances

```
eig.val <- get_eigenvalue(res.mca)
```

Pour visualiser les pourcentages de variances expliquées par chaque dimension de l'ACM, utilisez la fonction *fviz_eig()* ou *fviz_screplot()*:

```
fviz_screplot (res.mca, addlabels = TRUE, ylim = c (0, 45))
```

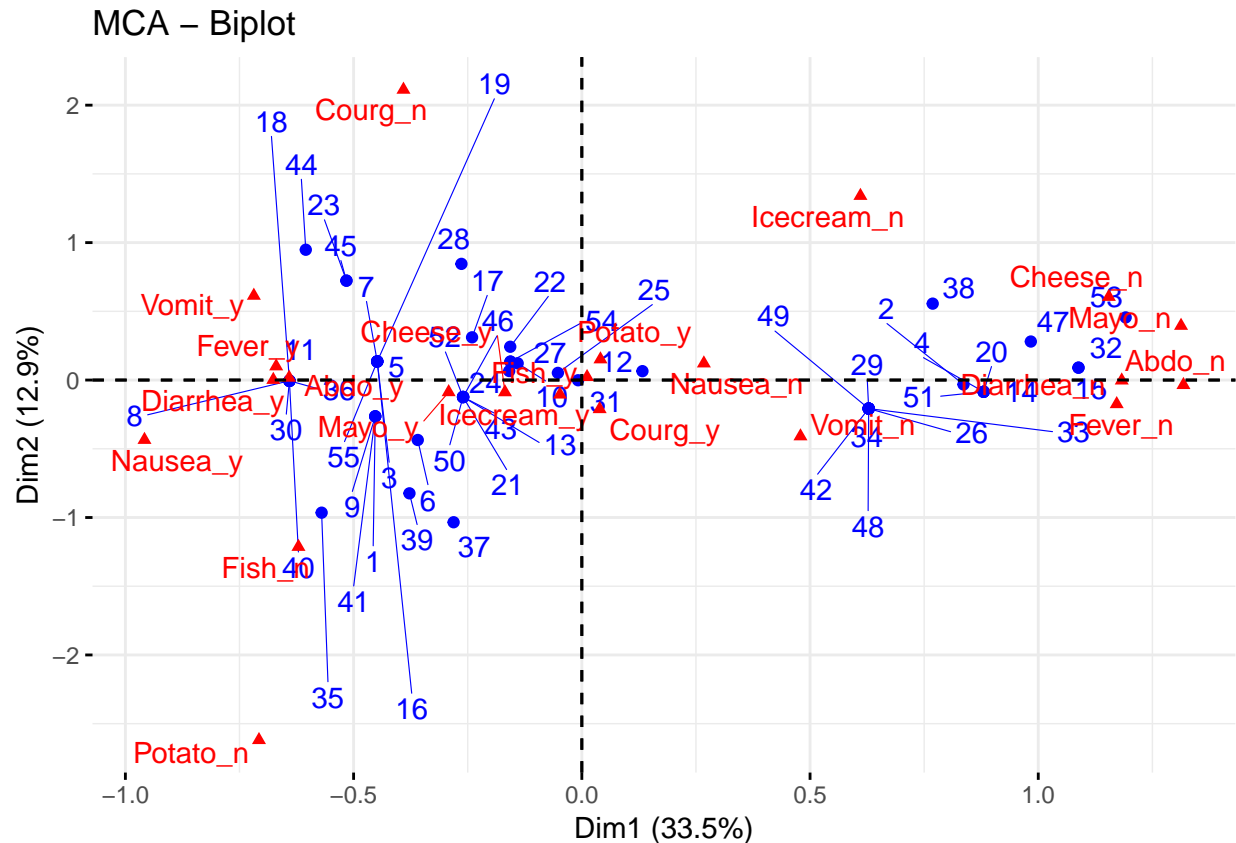
Scree plot



Biplot

La fonction `fviz_mca_biplot()` [*factoextra*] permet de visualiser le biplot des individus et des variables:

```
fviz_mca_biplot (res.mca, repel = TRUE,  
                 ggtheme = theme_minimal())
```



Dans le graphique ci-dessus, les lignes (individus) sont représentées par des points bleus et des colonnes (variables) par des triangles rouges.

La distance entre les individus donne une mesure de leur similitude (ou dissemblance). Les individus avec un profil similaire sont proches sur le graphique. Il en va de même pour les variables.

Graphiques des variables

Résultats

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

```
v <- get_mca_var(res.mca)
v
```

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

On peut consulter les différentes composantes comme suit:

```
# Coordonnées
head(v$coord)
```

```
##           Dim 1           Dim 2           Dim 3           Dim 4           Dim 5
```

```
## Nausea_n 0.2673909 0.12139029 -0.265583253 0.03376130 0.07370500
## Nausea_y -0.9581506 -0.43498187 0.951673323 -0.12097801 -0.26410958
## Vomit_n 0.4790279 -0.40919465 0.084492799 0.27361142 0.05245250
## Vomit_y -0.7185419 0.61379197 -0.126739198 -0.41041713 -0.07867876
## Abdo_n 1.3180221 -0.03574501 -0.005094243 -0.15360951 -0.06986987
## Abdo_y -0.6411999 0.01738946 0.002478280 0.07472895 0.03399075
```

```
# Cos2: qualité de représentation
head(v$cos2)
```

```
##          Dim 1          Dim 2          Dim 3          Dim 4          Dim 5
## Nausea_n 0.2562007 0.0528025759 2.527485e-01 0.004084375 0.019466197
## Nausea_y 0.2562007 0.0528025759 2.527485e-01 0.004084375 0.019466197
## Vomit_n 0.3442016 0.2511603912 1.070855e-02 0.112294813 0.004126898
## Vomit_y 0.3442016 0.2511603912 1.070855e-02 0.112294813 0.004126898
## Abdo_n 0.8451157 0.0006215864 1.262496e-05 0.011479077 0.002374929
## Abdo_y 0.8451157 0.0006215864 1.262496e-05 0.011479077 0.002374929
```

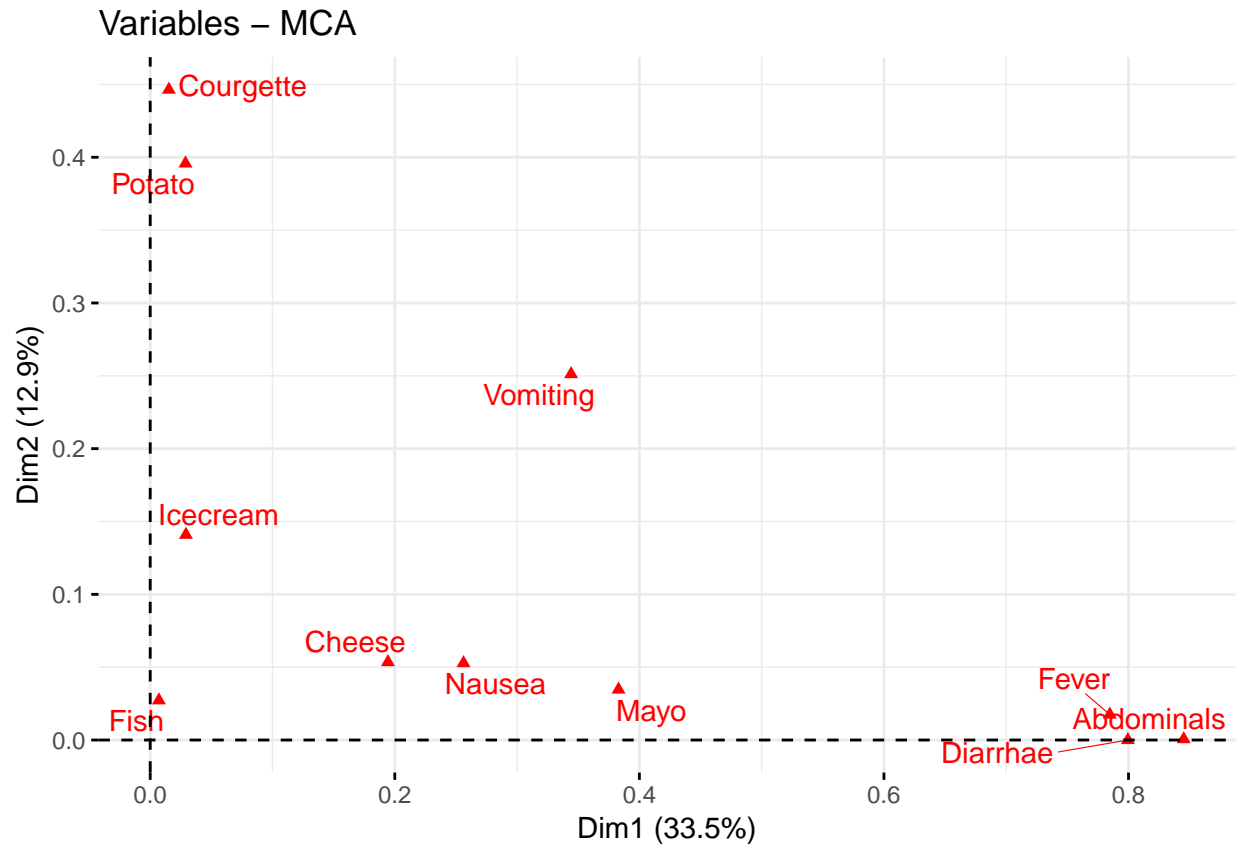
```
# Contributions aux axes
head(v$contrib)
```

```
##          Dim 1          Dim 2          Dim 3          Dim 4          Dim 5
## Nausea_n 1.515869 0.81100008 4.670018e+00 0.08449397 0.48977906
## Nausea_y 5.431862 2.90608363 1.673423e+01 0.30277007 1.75504164
## Vomit_n 3.733667 7.07226253 3.627455e-01 4.25893721 0.19036376
## Vomit_y 5.600500 10.60839380 5.441183e-01 6.38840581 0.28554563
## Abdo_n 15.417637 0.02943661 7.192511e-04 0.73219636 0.18424268
## Abdo_y 7.500472 0.01432051 3.499060e-04 0.35620363 0.08963157
```

Corrélation entre les variables et les axes principaux

Visualisation la corrélation entre les variables et les axes principaux de l'ACM :

```
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 Diarrhae, Abdominals et Fever sont les plus corrélées avec la dimension 1. De même, les variables Courgette et Potato sont les plus corrélées avec la dimension 2.

Coordonnées des catégories des variables

Affichage des coordonnées de chacune des catégories des variables dans chaque dimension(1,2 et 3):

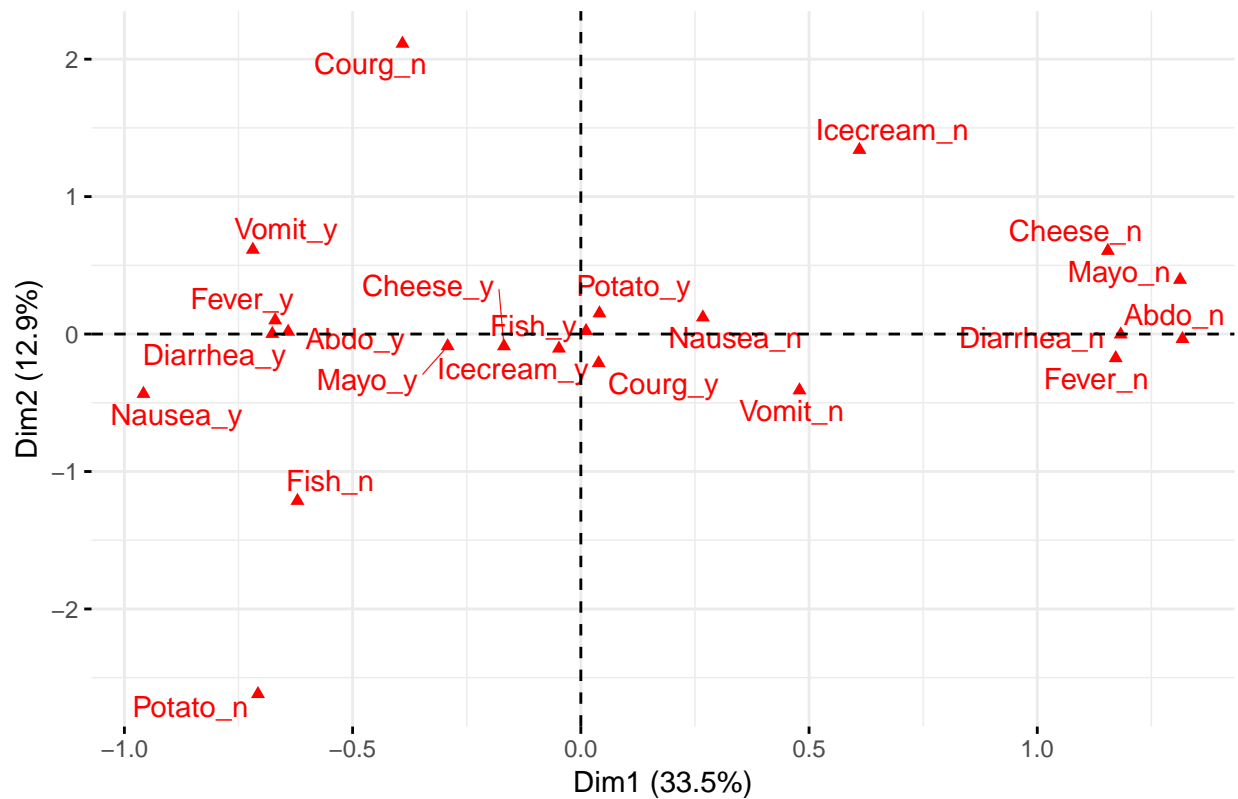
```
head(round(v$coord, 2), 4)
```

```
##          Dim 1 Dim 2 Dim 3 Dim 4 Dim 5
## Nausea_n  0.27  0.12 -0.27  0.03  0.07
## Nausea_y -0.96 -0.43  0.95 -0.12 -0.26
## Vomit_n   0.48 -0.41  0.08  0.27  0.05
## Vomit_y  -0.72  0.61 -0.13 -0.41 -0.08
```

Visualisation de la catégorie des variables:

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

Variable categories – MCA



```
#fviz_mca_var(res.mca, col.var="black", shape.var = 15, repel = TRUE)
```

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

- Les catégories avec un profil similaire sont regroupées.
- Les catégories corrélées négativement sont positionnées sur les côtés opposés de l'origine du graphique (quadrants opposés).
- 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.

Qualité de représentation des catégories des variables

Les deux dimensions 1 et 2 capturent 46% de l'inertie totale (variation) contenue dans les données. Tous les points ne sont pas aussi bien représentés par les deux dimensions.

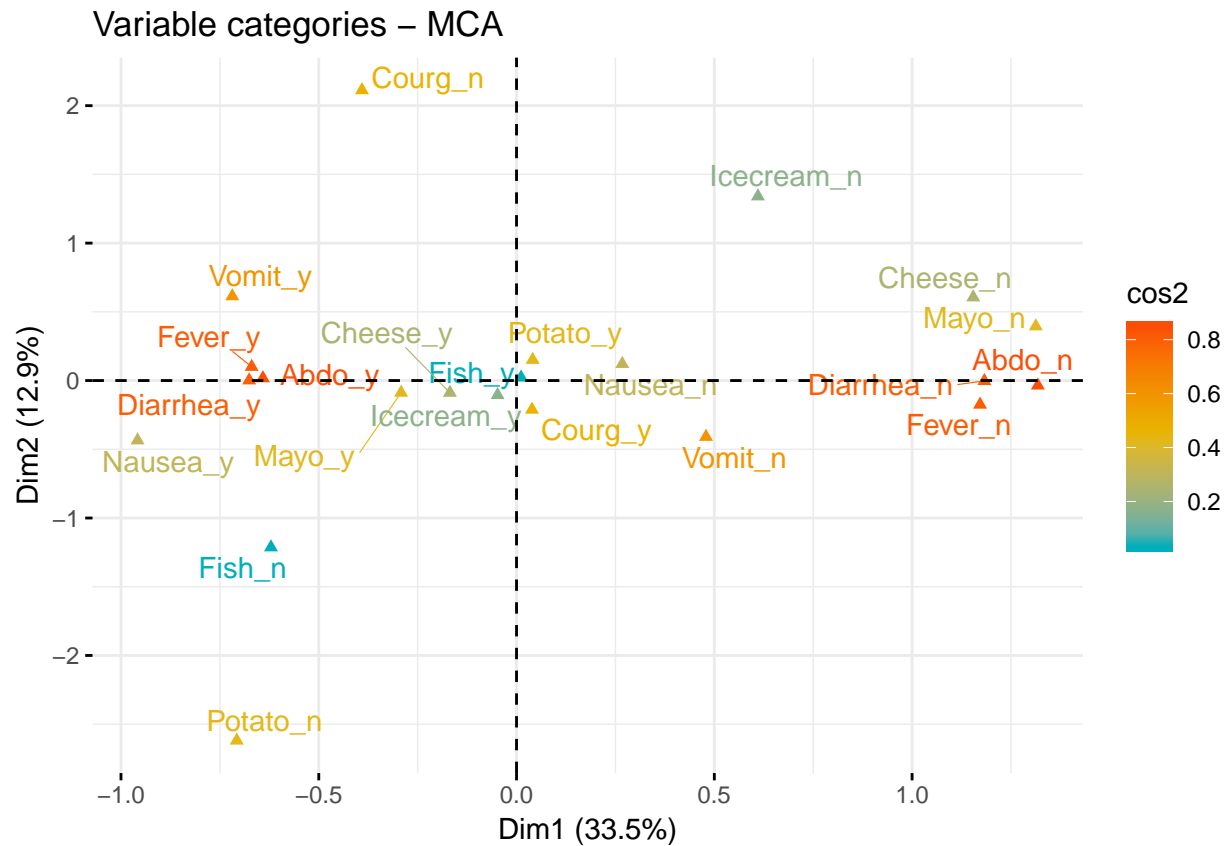
La qualité de représentation, appelée cosinus carré (\cos^2), mesure le degré d'association entre les catégories des variables et les dimensions. Le \cos^2 peut être extrait comme suit:

```
head(v$cos2, 4)
```

```
##          Dim 1      Dim 2      Dim 3      Dim 4      Dim 5
## Nausea_n 0.2562007 0.05280258 0.25274850 0.004084375 0.019466197
## Nausea_y 0.2562007 0.05280258 0.25274850 0.004084375 0.019466197
## Vomit_n  0.3442016 0.25116039 0.01070855 0.112294813 0.004126898
## Vomit_y  0.3442016 0.25116039 0.01070855 0.112294813 0.004126898
```

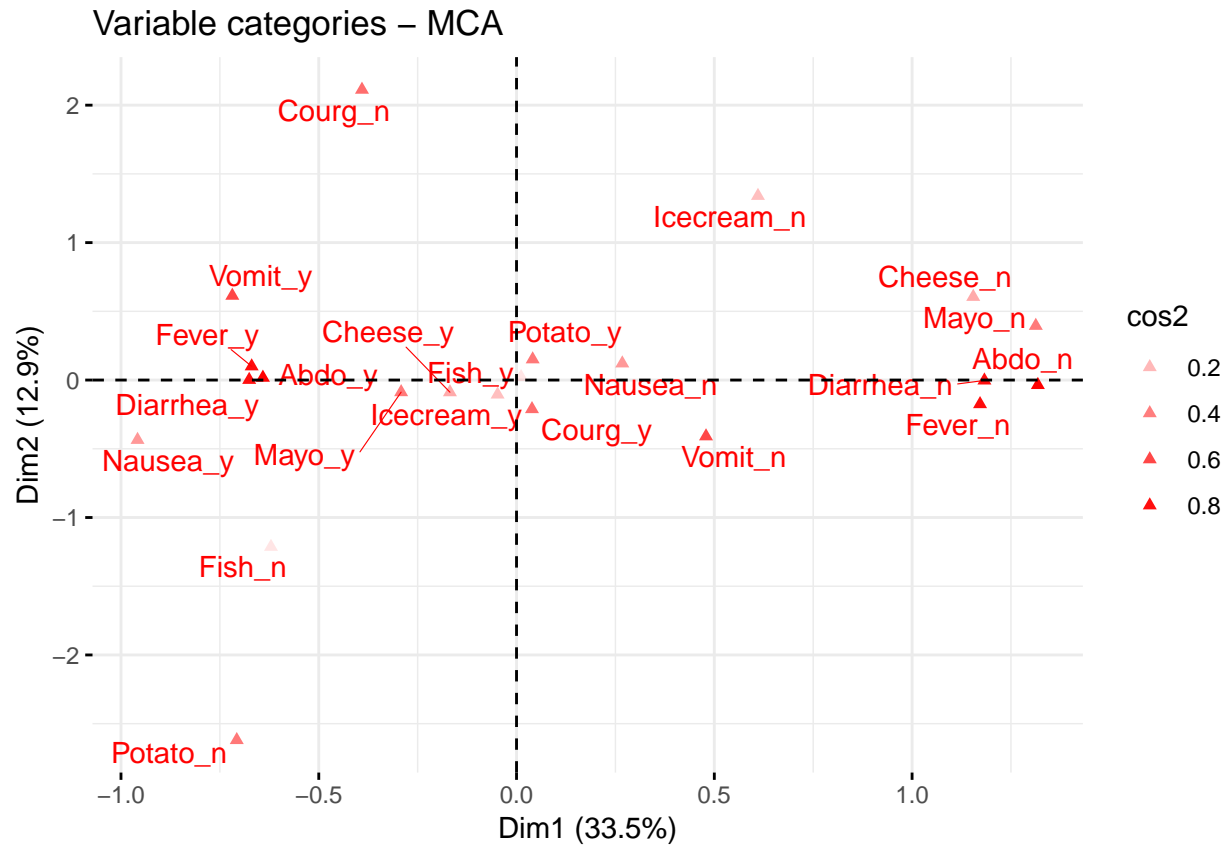
Si une catégorie d'une variable donnée est bien représentée par deux dimensions, la somme des \cos^2 est proche de 1. Pour certains éléments, plus de 2 dimensions sont nécessaires pour représenter parfaitement les données.


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



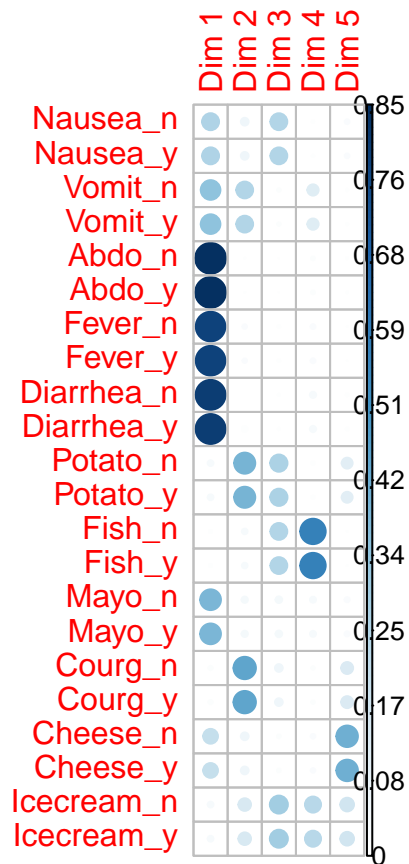
Modifier avec la transparence:

```
# Changer la transparence en fonction du cos2
fviz_mca_var(res.mca, alpha.var = "cos2",
             repel = TRUE,
             ggtheme = theme_minimal())
```



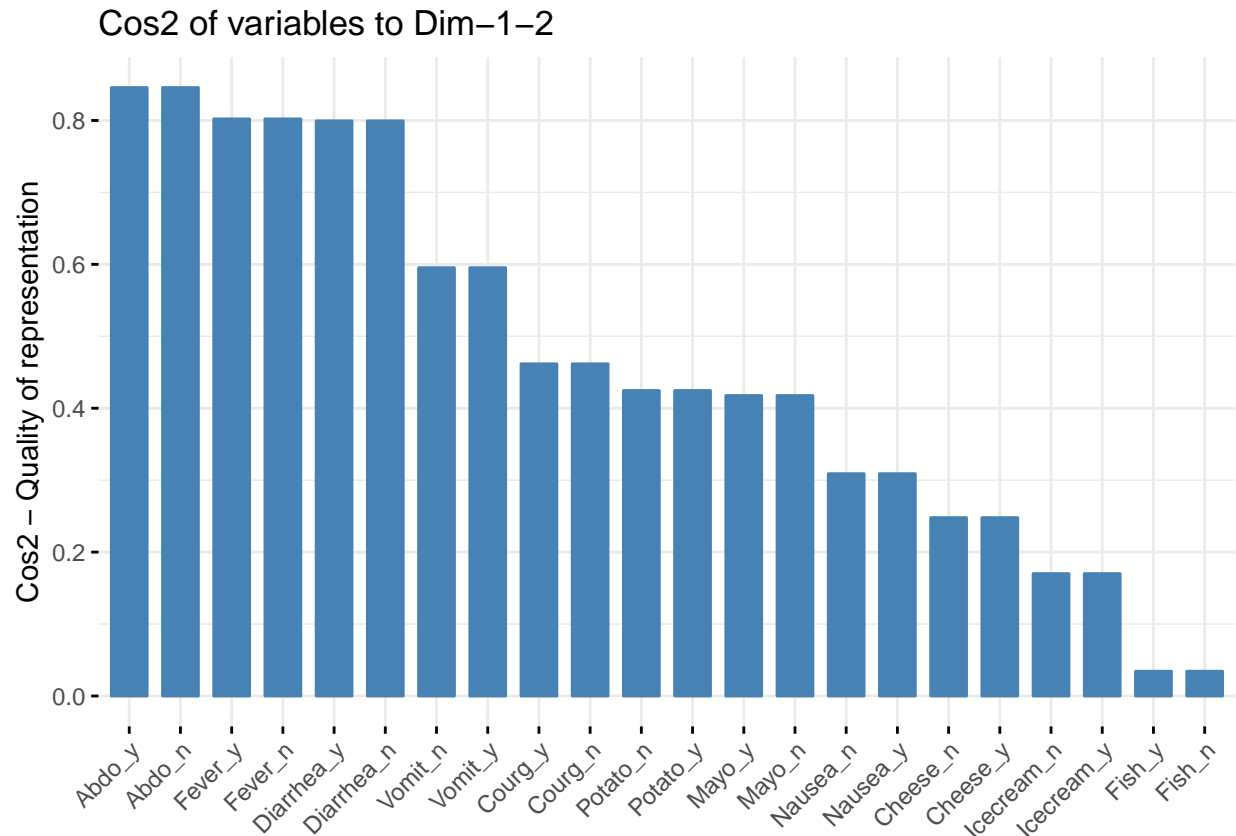
Visualisation des cos2 des catégories sur toutes les dimensions

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



Création un barplot du cos2 des variables

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



Notez que les catégories Fish_n, Fish_y, Icecream_n et Icecream_y ne sont pas très bien représentées par les deux premières dimensions. Cela implique que la position des points correspondants sur le graphique doit être interprétée avec prudence.

Contribution des variables aux dimensions

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

```
head(round(v$contrib,2), 4)
```

```
##          Dim 1 Dim 2 Dim 3 Dim 4 Dim 5
## Nausea_n  1.52  0.81  4.67  0.08  0.49
## Nausea_y  5.43  2.91 16.73  0.30  1.76
## Vomit_n   3.73  7.07  0.36  4.26  0.19
## Vomit_y   5.60 10.61  0.54  6.39  0.29
```

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

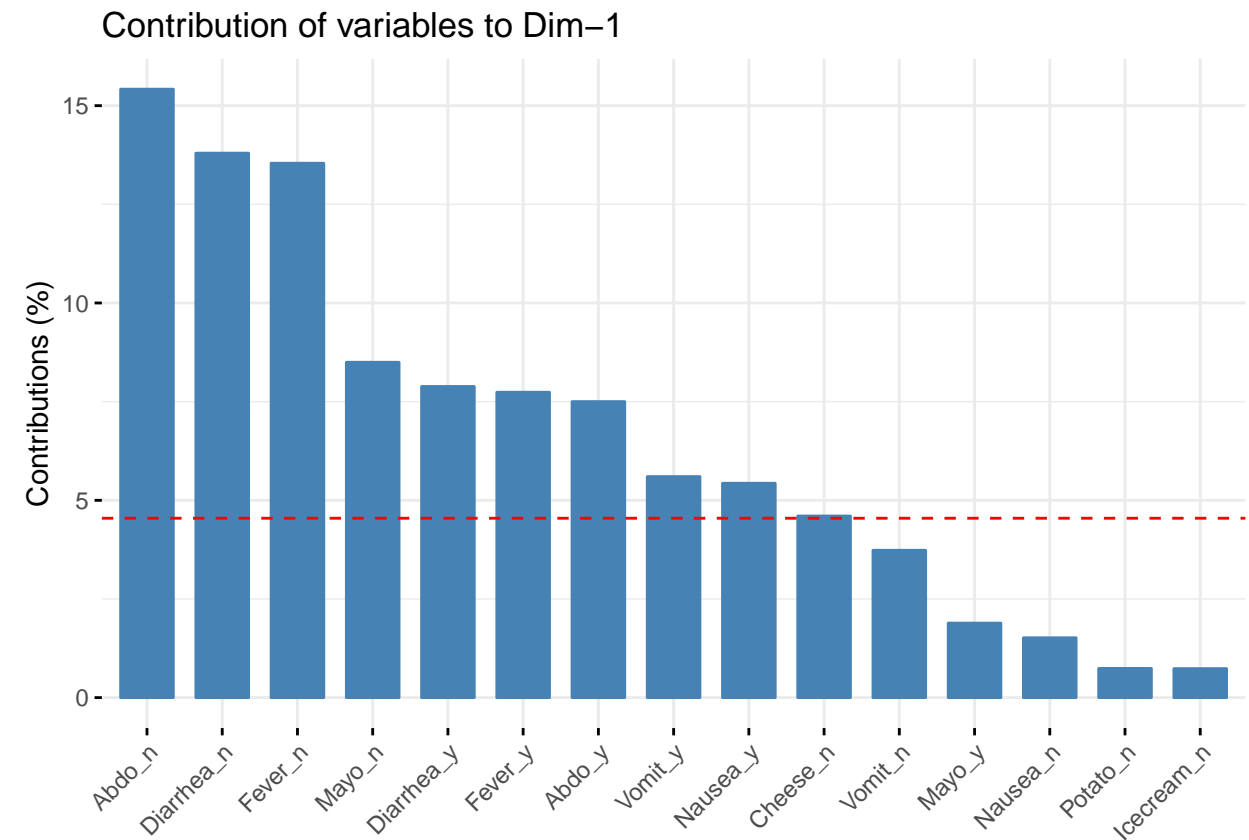
```
head(round(v$contrib,2), 4)
```

```
##          Dim 1 Dim 2 Dim 3 Dim 4 Dim 5
## Nausea_n  1.52  0.81  4.67  0.08  0.49
## Nausea_y  5.43  2.91 16.73  0.30  1.76
## Vomit_n   3.73  7.07  0.36  4.26  0.19
## Vomit_y   5.60 10.61  0.54  6.39  0.29
```

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.

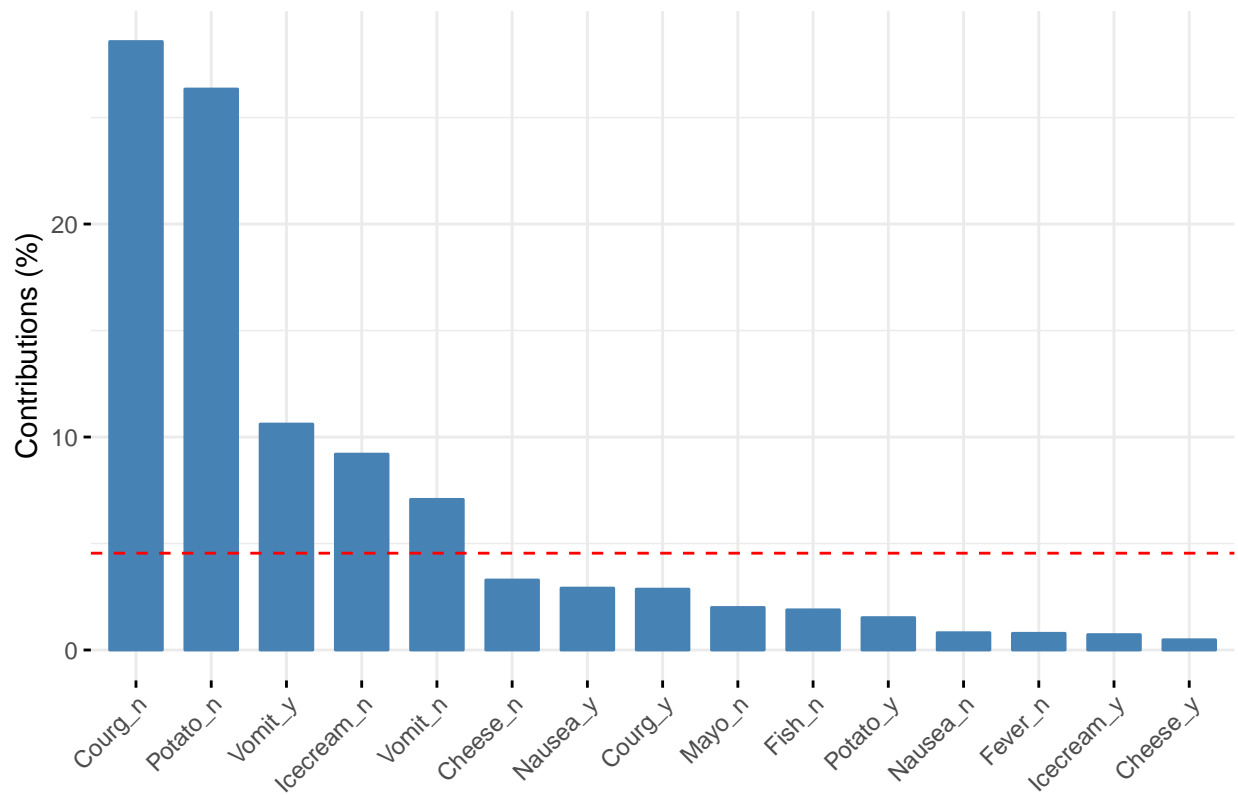
La fonction `fviz_contrib()` [*factoextra*] peut être utilisée pour faire un barplot de la contribution des catégories des variables. Le code R ci-dessous montre le top 15 des catégories contribuant aux dimensions:

```
par(mfrow=c(1,2))  
# Contributions des variables à la dimension 1  
fviz_contrib(res.mca, choice = "var", axes = 1, top = 15)
```



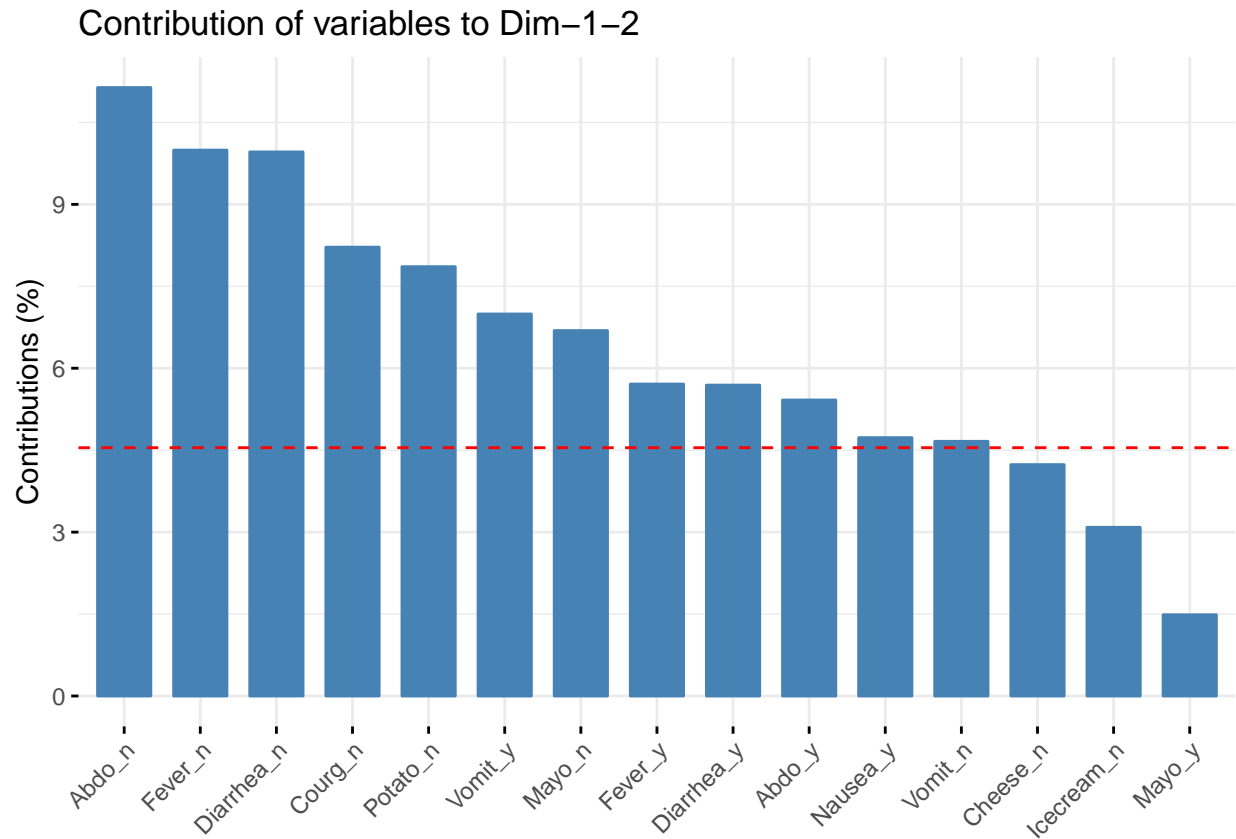
```
# Contributions des variables à la dimension 2  
fviz_contrib(res.mca, choice = "var", axes = 2, top = 15)
```

Contribution of variables to Dim-2



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 = 15)
```

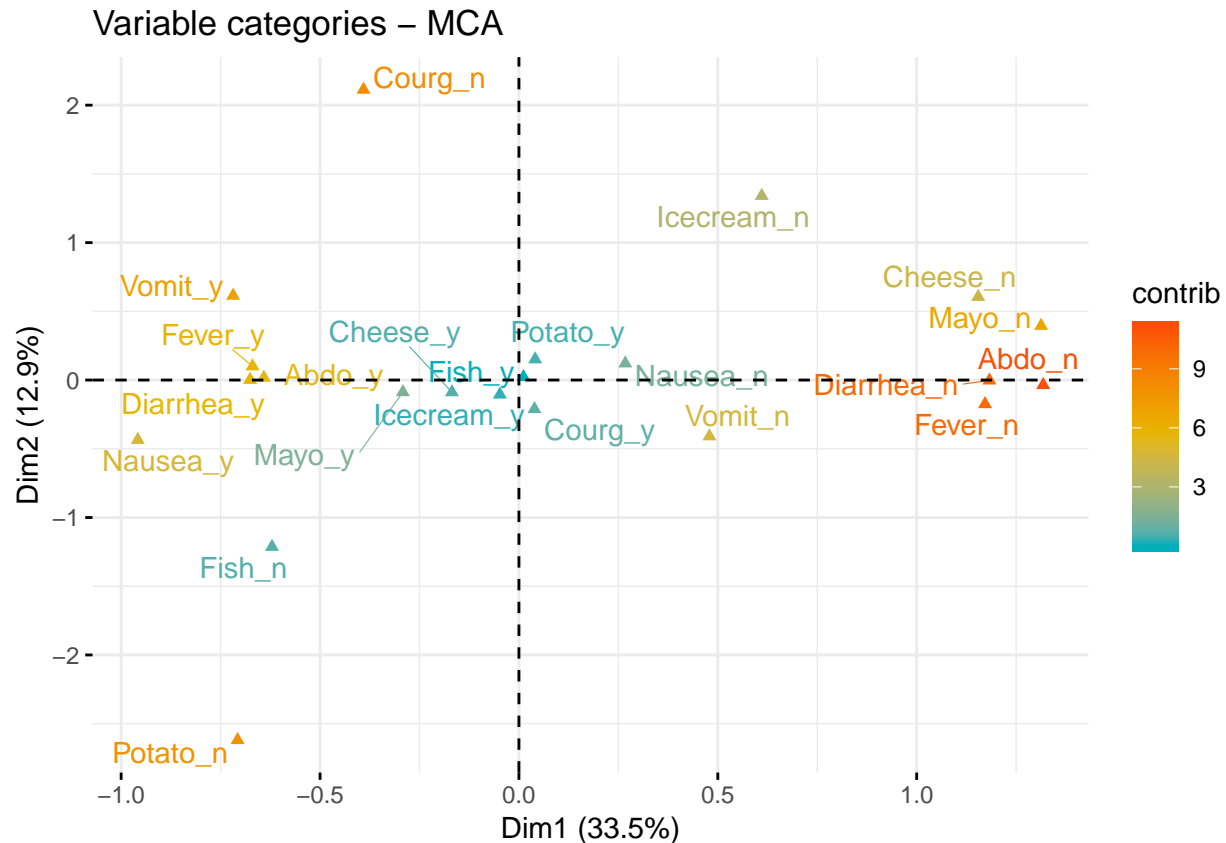


La ligne en pointillé rouge, sur le graphique ci-dessus, indique la valeur moyenne attendue sous l'hypothèse nulle.

On peut voir que:

les catégories Abdo_n, Diarrhea_n, Fever_n et Mayo_n sont les plus importantes dans la définition de la première dimension. Les catégories Courg_n, Potato_n, Vomit_y et Icecream_n 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,
              ggtheme = theme_minimal()
            )
```



Le graphique ci-dessus donne une idée du pôle des dimensions auquel les catégories contribuent réellement.

Il est évident que les catégories Abdo_n, Diarrhea_n, Fever_n et Mayo_n ont une contribution importante au pôle positif de la première dimension, tandis que les catégories Fever_y et Diarrhea_y ont une contribution majeure au pôle négatif de la première dimension; etc,

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:

```
# Coordonnées
head(ind$coord)
```



```
##          Dim 1          Dim 2          Dim 3          Dim 4          Dim 5
## 1 -0.4525811 -0.26415072  0.17151614  0.01369348 -0.11696806
## 2  0.8361700 -0.03193457 -0.07208249 -0.08550351  0.51978710
## 3 -0.4481892  0.13538726 -0.22484048 -0.14170168 -0.05004753
## 4  0.8803694 -0.08536230 -0.02052044 -0.07275873 -0.22935022
## 5 -0.4481892  0.13538726 -0.22484048 -0.14170168 -0.05004753
## 6 -0.3594324 -0.43604390 -1.20932223  1.72464616  0.04348157
```

```
# Qualité de représentation
head(ind$cos2)
```

```
##          Dim 1          Dim 2          Dim 3          Dim 4          Dim 5
## 1 0.34652591 0.1180447167 0.0497683175 0.0003172275 0.0231460846
## 2 0.55589562 0.0008108236 0.0041310808 0.0058126211 0.2148103098
## 3 0.54813888 0.0500176790 0.1379484860 0.0547920948 0.0068349171
## 4 0.74773962 0.0070299584 0.0004062504 0.0051072923 0.0507479873
## 5 0.54813888 0.0500176790 0.1379484860 0.0547920948 0.0068349171
## 6 0.02485357 0.0365775483 0.2813443706 0.5722083217 0.0003637178
```

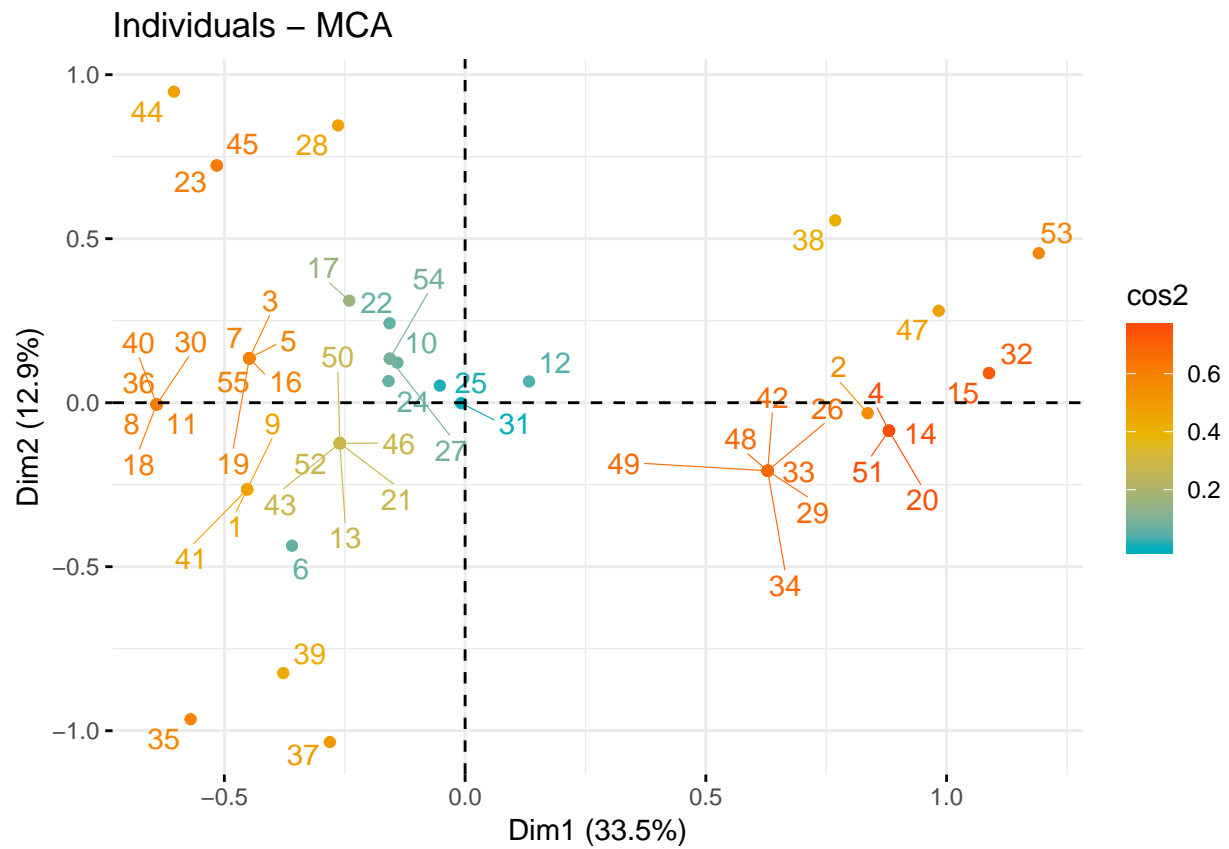
```
# Contributions
head(ind$contrib)
```

```
##          Dim 1          Dim 2          Dim 3          Dim 4          Dim 5
## 1 1.110927 0.98238297 0.498254685 0.003555817 0.31554778
## 2 3.792117 0.01435818 0.088003703 0.138637089 6.23134138
## 3 1.089470 0.25806722 0.856229950 0.380768961 0.05776914
## 4 4.203611 0.10259105 0.007132055 0.100387990 1.21319013
## 5 1.089470 0.25806722 0.856229950 0.380768961 0.05776914
## 6 0.700692 2.67693398 24.769968729 56.404214518 0.04360547
```

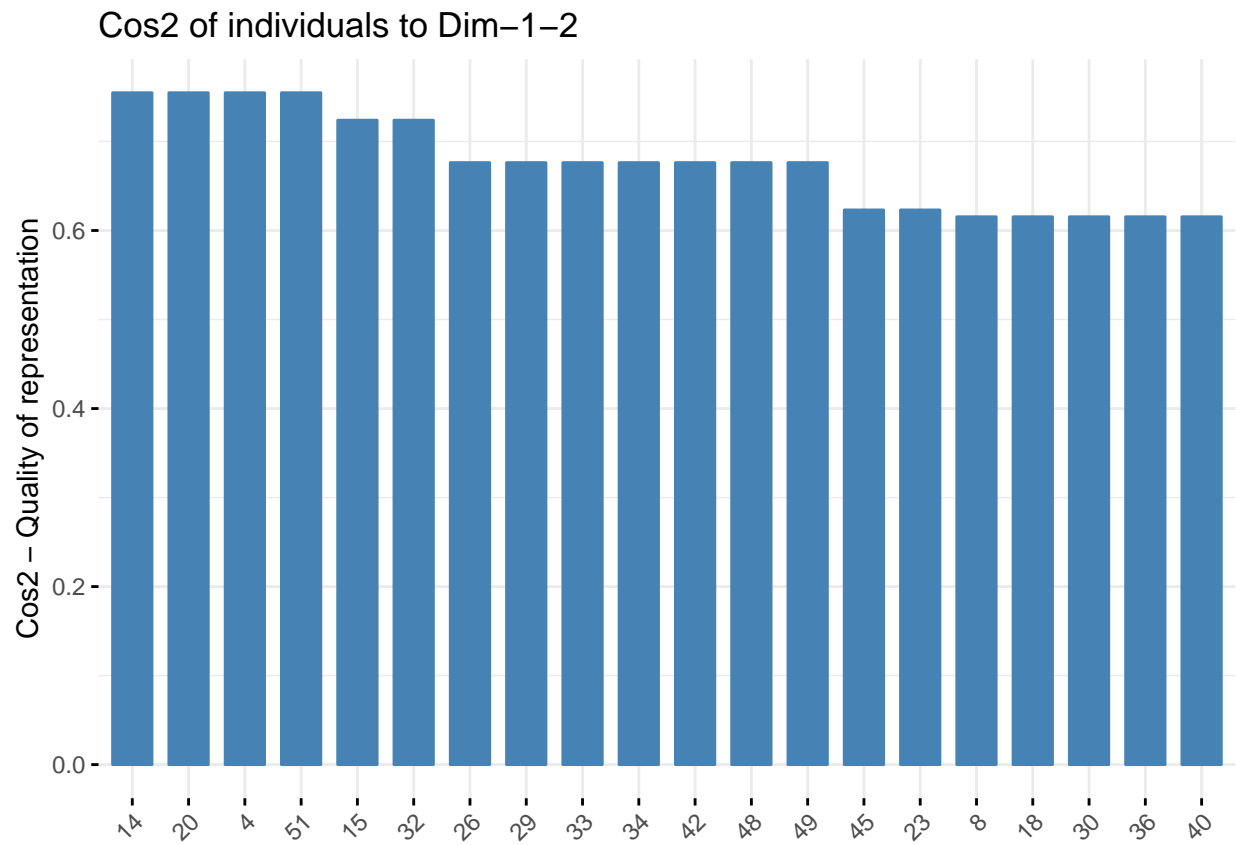
graphique: qualité et contribution

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())
```

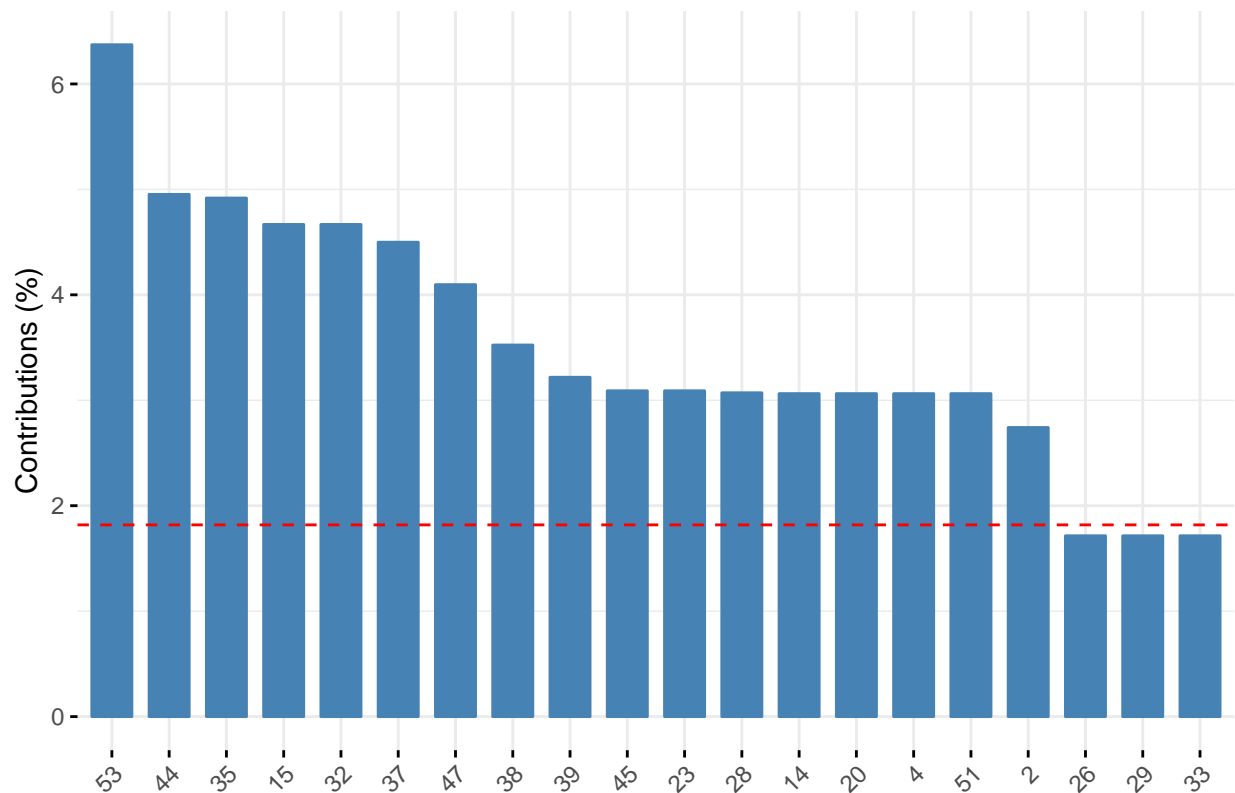


```
# Cos2 des individus
fviz_cos2(res.mca, choice = "ind", axes = 1:2, top = 20)
```



```
# Contribution des individus aux dimensions
fviz_contrib(res.mca, choice = "ind", axes = 1:2, top = 20)
```

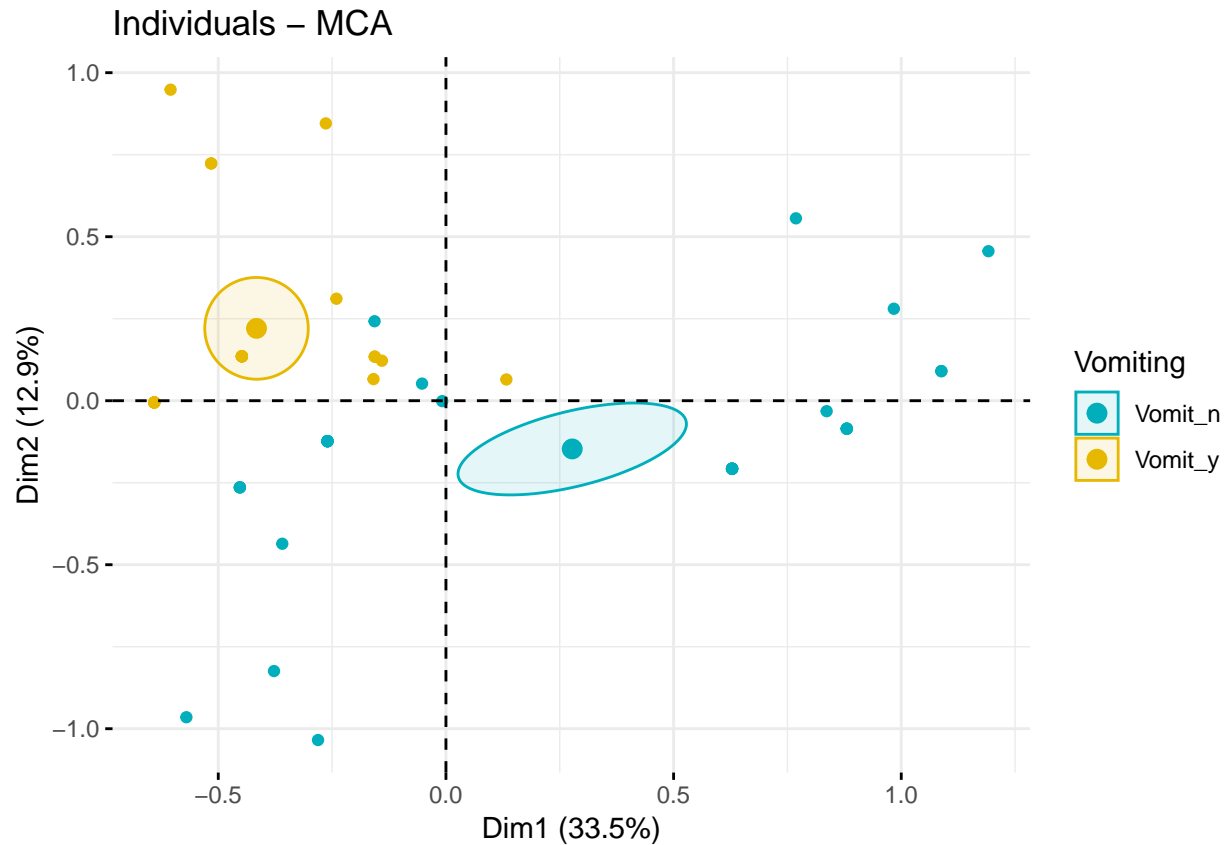
Contribution of individuals to Dim-1-2



Colorer les individus par groupes

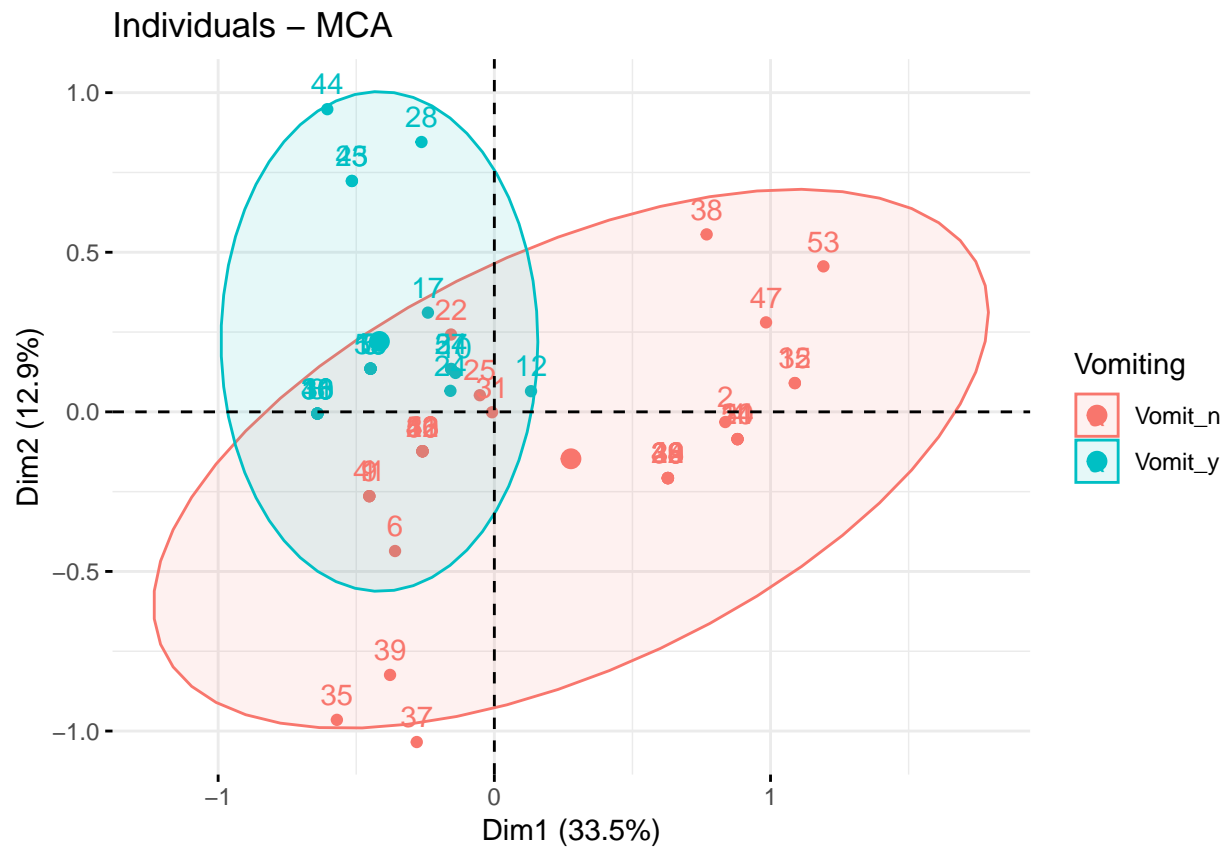
Le code R ci-dessous colore les individus par groupes en utilisant la variable Vomiting. L'argument `habillage` sert à spécifier la variable à utiliser pour colorer les individus par groupes. Une ellipse de concentration peut également être ajoutée autour de chaque groupe en utilisant l'argument `addEllipses = TRUE`. Si vous voulez une ellipse de confiance autour du point moyen (centre de gravité) des groupes, utilisez `ellipse.type = "confidence"`. L'argument `palette` permet de modifier les couleurs du groupe.

```
fviz_mca_ind (res.mca,
  label = "none", # masquer le texte des individus
  habillage = "Vomiting", # colorer par groupes
  palette = c ("#00AFBB", "#E7B800"),
  addEllipses = TRUE, ellipse.type = "confidence",
  ggtheme = theme_minimal ())
```

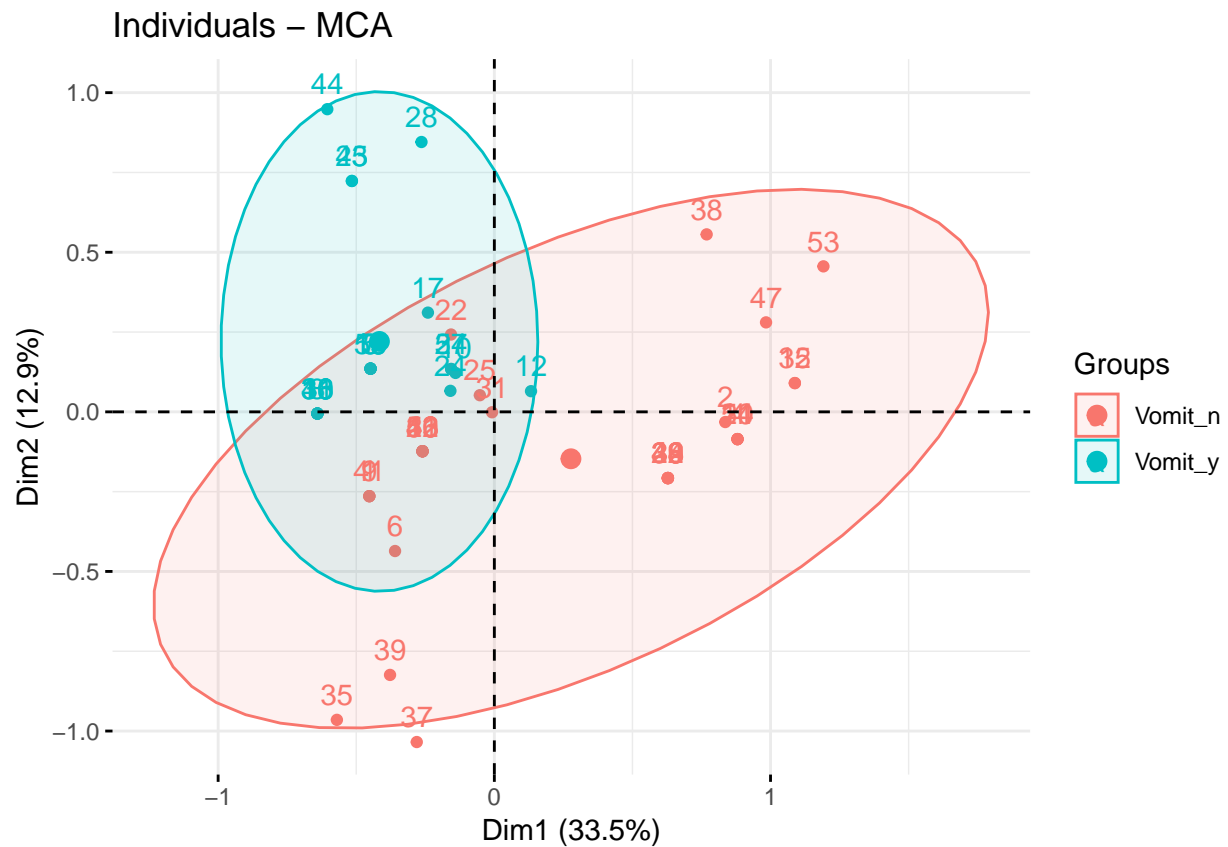


Notez que, pour spécifier l'argument **habillage**, il est également possible d'utiliser l'index de la colonne comme suit (**habillage = 2**). Vous pouvez aussi fournir une variable de regroupement externe comme suit: **habillage = poison\$Vomiting**. Par exemple:

```
# habillage = indice de la colonne
# à utiliser comme variable de regroupement
fviz_mca_ind(res.mca, habillage = 2, addEllipses = TRUE)
```

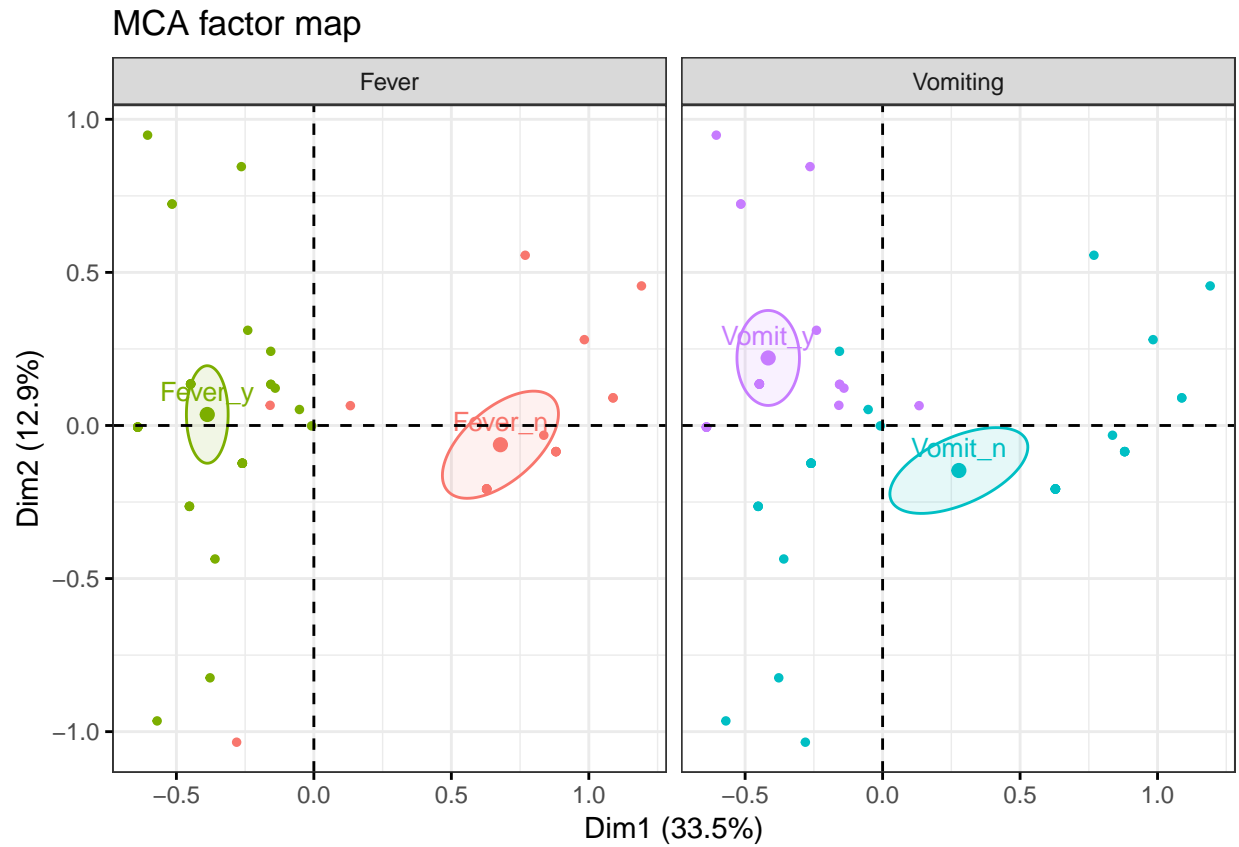


```
# habillage = variable de regroupement externe
fviz_mca_ind(res.mca, habillage = poison$Vomiting, addEllipses = TRUE)
```



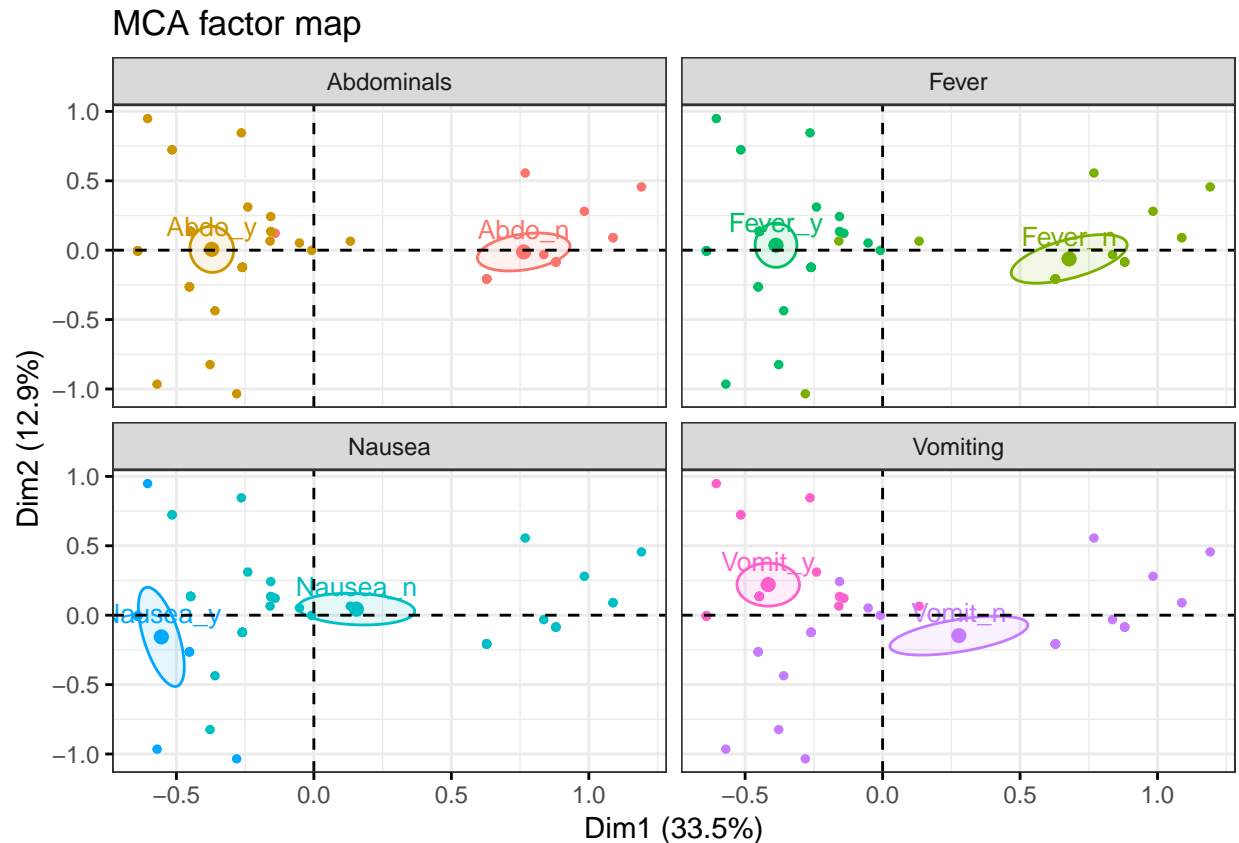
Si vous souhaitez colorer les individus à l'aide de plusieurs variables catégorielles en même temps, utilisez la fonction `fviz_ellipses()` [*factoextra*] comme suit:

```
fviz_ellipses(res.mca, c("Vomiting", "Fever"),
  geom = "point")
```



Alternativement, vous pouvez spécifier les indices des variables catégorielles:

```
fviz_ellipses (res.mca, 1:4, geom = "point")
```

Description des dimensions

La fonction `dimdesc()` [*FactoMineR*] peut être utilisée pour identifier les variables les plus corrélées avec une dimension donnée:

```
res.desc <- dimdesc (res.mca, axes = c(1,2))
# Description de la dimension 1
res.desc[[1]]
```

```
## $quali
##           R2      p.value
## Abdominals 0.8451157 4.055640e-23
## Diarrhae   0.7994680 3.910776e-20
## Fever      0.7846788 2.600566e-19
## Mayo       0.3829749 4.756234e-07
## Vomiting    0.3442016 2.510738e-06
## Nausea      0.2562007 8.062777e-05
## Cheese     0.1944181 7.534834e-04
##
## $category
##      Estimate      p.value
## Abdo_n   0.5671866 4.055640e-23
## Diarrhea_n 0.5380920 3.910776e-20
## Fever_n   0.5330918 2.600566e-19
## Mayo_n    0.4644981 4.756234e-07
## Vomit_n   0.3466915 2.510738e-06
```

```
## Nausea_n      0.3547892 8.062777e-05
## Cheese_n      0.3830043 7.534834e-04
## Cheese_y     -0.3830043 7.534834e-04
## Nausea_y     -0.3547892 8.062777e-05
## Vomit_y      -0.3466915 2.510738e-06
## Mayo_y       -0.4644981 4.756234e-07
## Fever_y      -0.5330918 2.600566e-19
## Diarrhea_y   -0.5380920 3.910776e-20
## Abdo_y       -0.5671866 4.055640e-23
# Description de la dimension 2
res.desc[[2]]
```

```
## $quali
##           R2      p.value
## Courgette 0.4464145 2.500166e-08
## Potato    0.3957543 2.690662e-07
## Vomiting  0.2511604 9.728027e-05
## Icecream  0.1409011 4.743927e-03
##
## $category
##           Estimate      p.value
## Courg_n    0.4176013 2.500166e-08
## Potato_y   0.4977523 2.690662e-07
## Vomit_y    0.1838104 9.728027e-05
## Icecream_n 0.2597197 4.743927e-03
## Icecream_y -0.2597197 4.743927e-03
## Vomit_n   -0.1838104 9.728027e-05
## Potato_n  -0.4977523 2.690662e-07
## Courg_y   -0.4176013 2.500166e-08
```

Eléments supplémentaires

Définition et types

Comme décrit ci-dessus (section @ref(mca-data-format)), le jeu de données poison contient:

- des variables continues supplémentaires (quanti.sup = 1:2, colonnes 1 et 2 correspondant aux colonnes age et time, respectivement)
- des variables qualitatives supplémentaires (quali.sup = 3:4, correspondant aux colonnes Sick et Sex, respectivement). Ces variables sont utilisées pour colorer les individus par groupes.

Les données ne contiennent pas des individus supplémentaires. Cependant, pour la démonstration, nous utiliserons les individus 53:55 en tant qu'individus supplémentaires.

Les variables et les individus supplémentaires ne sont pas utilisés pour déterminer les dimensions principales. Leurs coordonnées sont prédites en utilisant uniquement les informations fournies par l'ACM effectuée sur des variables/individus actifs.

Spécification dans l'ACM

Pour spécifier des individus et des variables supplémentaires, la fonction `MCA()` peut être utilisée comme suit:

```
res.mca <- MCA (poison, ind = 53:55,
               quanti.sup = 1:2, quali.sup = 3:4, graph = FALSE)
```

Résultats

Les résultats prédites pour les individus / variables supplémentaires peuvent être extraits comme suit:

```
# Variables qualitatives supplémentaires
```

```
res.mca$quali.sup
```

```
## $coord
##           Dim 1           Dim 2           Dim 3           Dim 4           Dim 5
## Sick_n  1.41809140  0.0020394048  0.13199139 -0.016036841 -0.08354663
## Sick_y -0.63026284 -0.0009064021 -0.05866284  0.007127485  0.03713184
## F      -0.03108147  0.1123143957  0.05033124 -0.055927173 -0.06832928
## M       0.03356798 -0.1212995474 -0.05435774  0.060401347  0.07379562
##
## $cos2
##           Dim 1           Dim 2           Dim 3           Dim 4           Dim 5
## Sick_n  0.893770319  1.848521e-06  0.007742990  0.0001143023  0.003102240
## Sick_y  0.893770319  1.848521e-06  0.007742990  0.0001143023  0.003102240
## F       0.001043342  1.362369e-02  0.002735892  0.0033780765  0.005042401
## M       0.001043342  1.362369e-02  0.002735892  0.0033780765  0.005042401
##
## $v.test
##           Dim 1           Dim 2           Dim 3           Dim 4           Dim 5
## Sick_n  6.7514655  0.009709509  0.6284047 -0.07635063 -0.3977615
## Sick_y -6.7514655 -0.009709509 -0.6284047  0.07635063  0.3977615
## F      -0.2306739  0.833551410  0.3735378 -0.41506855 -0.5071119
## M       0.2306739 -0.833551410 -0.3735378  0.41506855  0.5071119
##
## $eta2
##           Dim 1           Dim 2           Dim 3           Dim 4           Dim 5
## Sick  0.893770319  1.848521e-06  0.007742990  0.0001143023  0.003102240
## Sex   0.001043342  1.362369e-02  0.002735892  0.0033780765  0.005042401
```

```
# Variables quantitatives supplémentaires
```

```
res.mca$quanti
```

```
## $coord
##           Dim 1           Dim 2           Dim 3           Dim 4           Dim 5
## Age   0.003934896 -0.00741340 -0.26494536  0.20015501  0.02928483
## Time -0.838158507 -0.08330586 -0.08718851 -0.08421599 -0.02316931
```

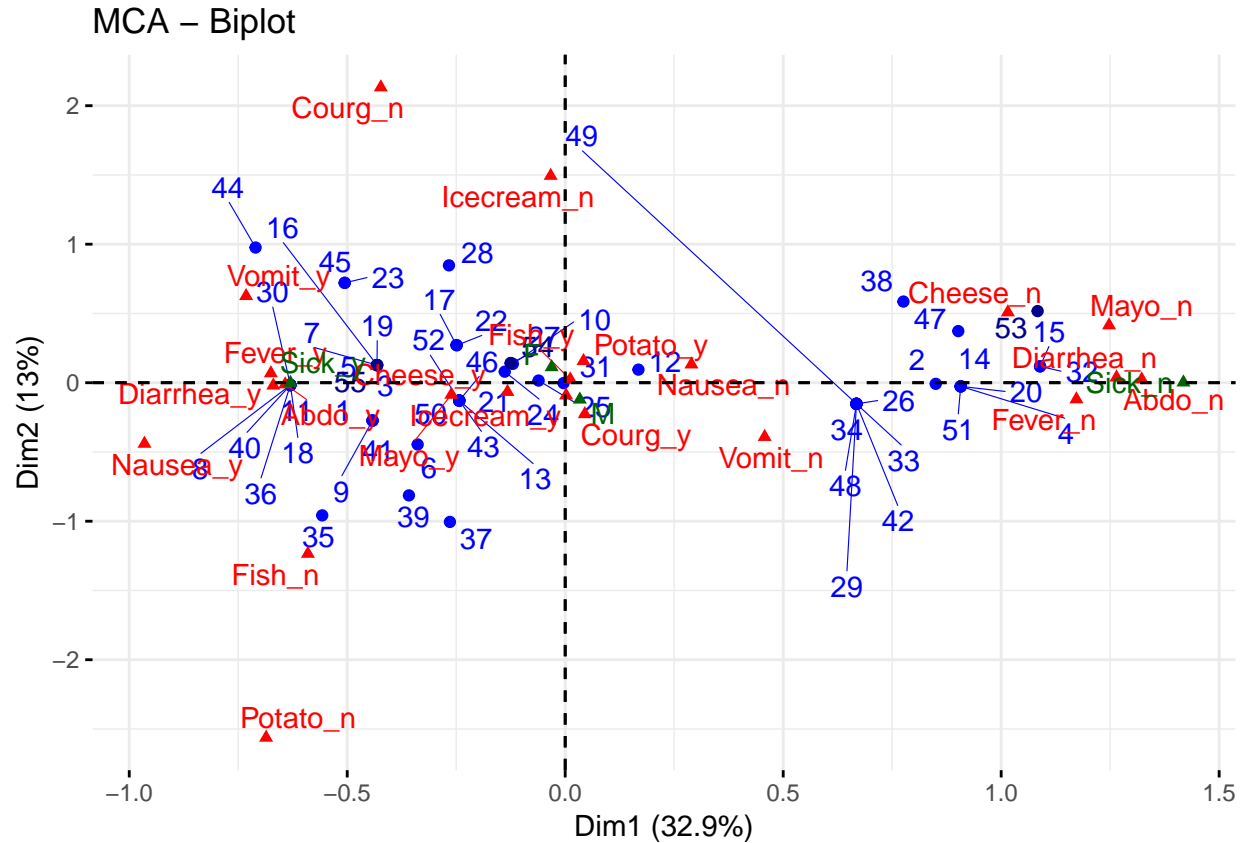
```
# Individus supplémentaires
```

```
res.mca$ind.sup
```

```
## $coord
##           Dim 1           Dim 2           Dim 3           Dim 4           Dim 5
## 53  1.0835684  0.5172478  0.5794063  0.5390903  0.4553650
## 54 -0.1249473  0.1417271 -0.1765234 -0.1526587 -0.2779565
## 55 -0.4315948  0.1270468 -0.2071580 -0.1186804 -0.1891760
##
## $cos2
##           Dim 1           Dim 2           Dim 3           Dim 4           Dim 5
## 53  0.36304957  0.08272764  0.10380536  0.08986204  0.06411692
## 54  0.03157652  0.04062716  0.06302535  0.04713607  0.15626590
## 55  0.50232519  0.04352713  0.11572730  0.03798314  0.09650827
```

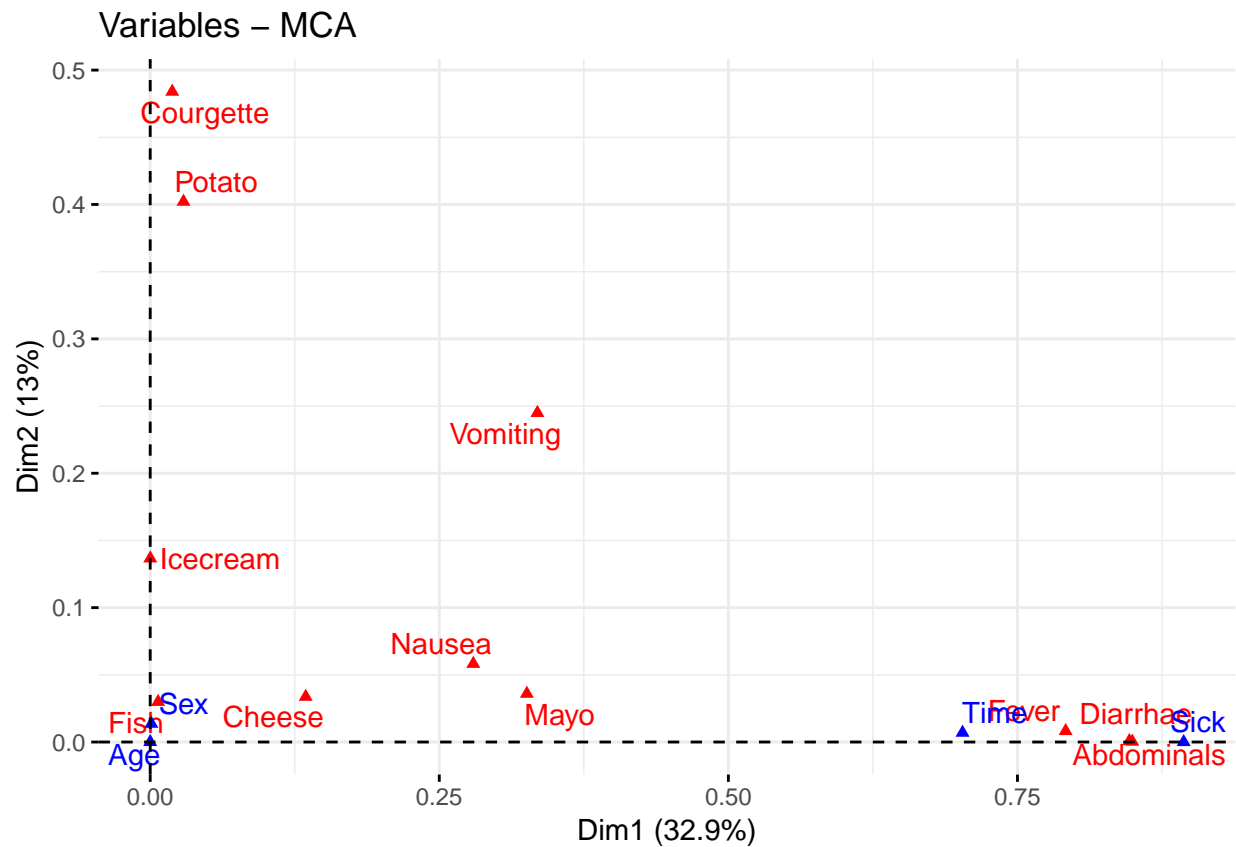
Graphique

```
# Biplot des individus et des variables
fviz_mca_biplot (res.mca, repel = TRUE,
                 ggtheme = theme_minimal ())
```



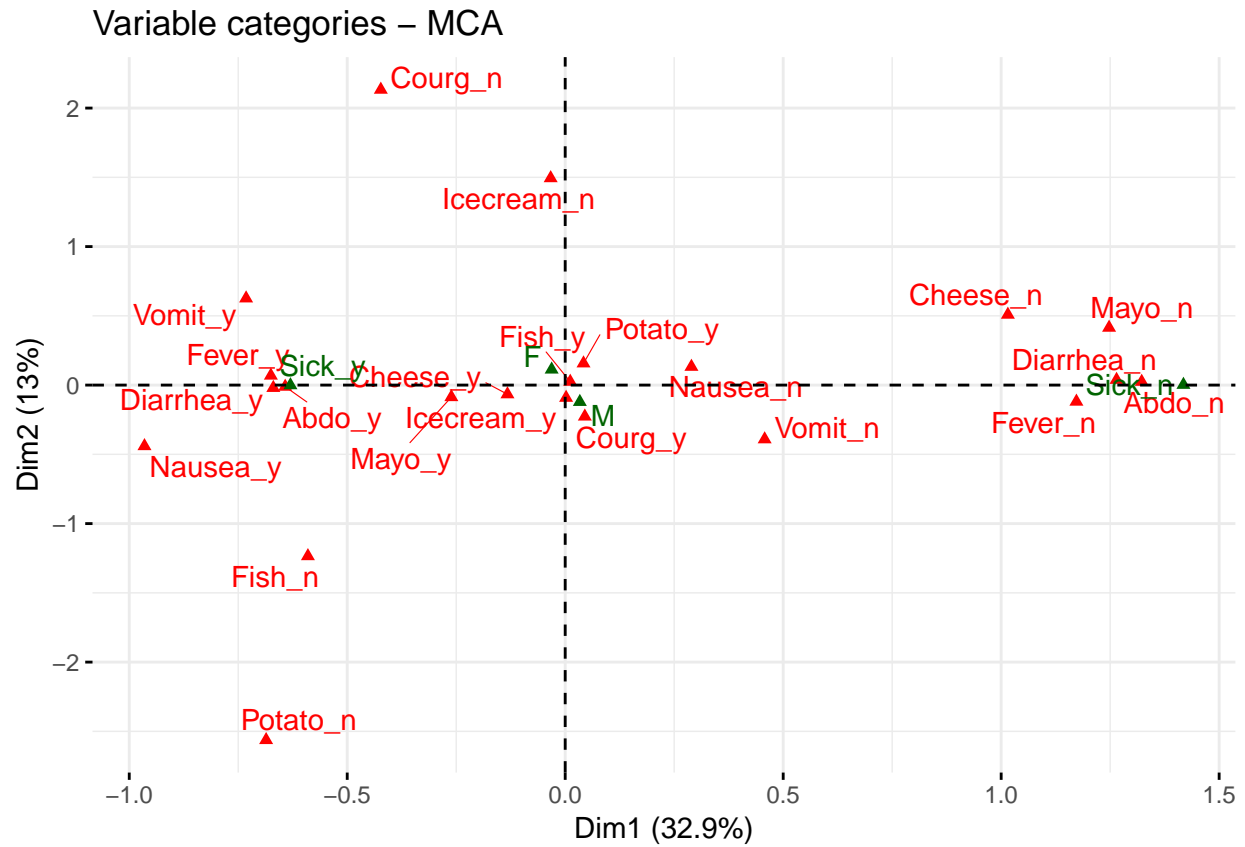
- Les individus actifs sont en bleu
- Les individus supplémentaires sont en bleu foncé
- Les catégories des variables actives sont en rouge
- Les catégories des variables supplémentaires sont en vert foncé

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



Si vous souhaitez mettre en évidence la corrélation entre les variables (actifs et supplémentaires) et les dimensions, utilisez la fonction `fviz_mca_var()` avec l'argument `choice= "mca.cor"`:

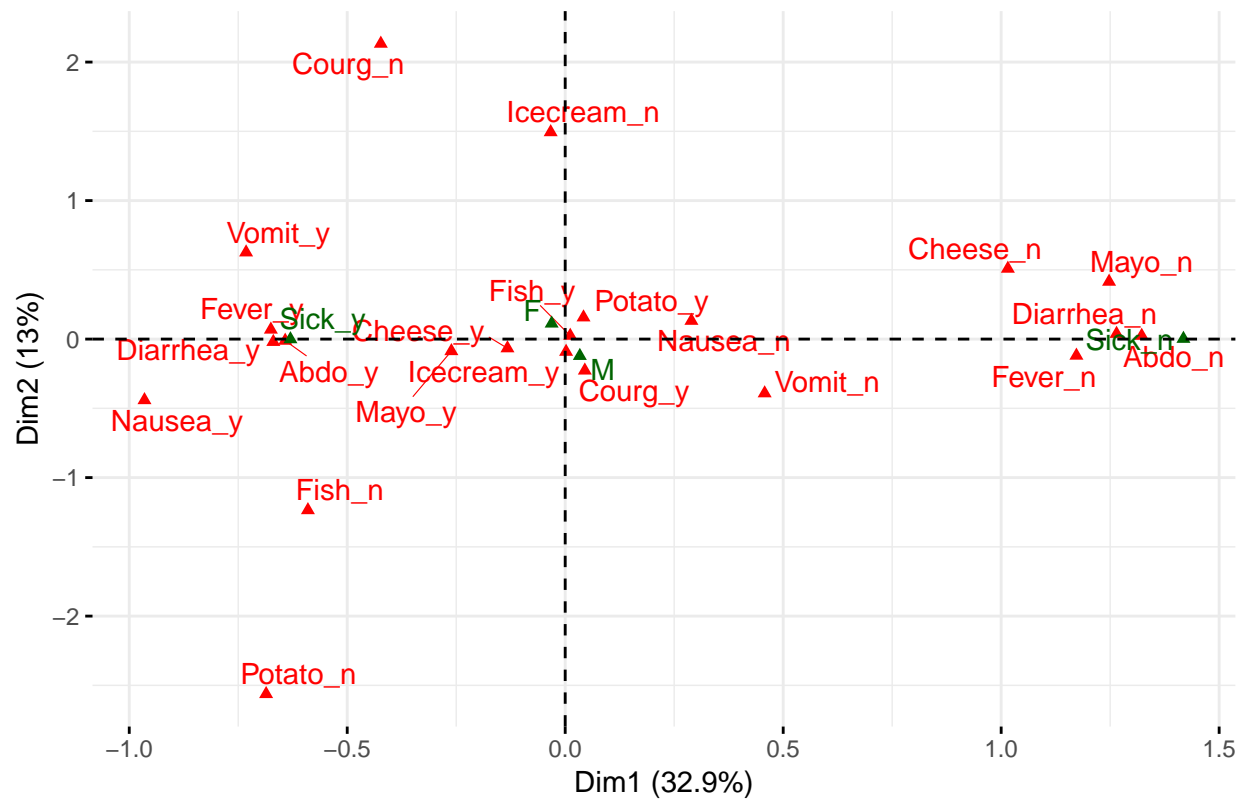
```
fviz_mca_var(res.mca, repel = TRUE,
              ggtheme= theme_minimal())
```



Le code R ci-dessous visualise les catégories des variables qualitatives (variables actives et supplémentaires):

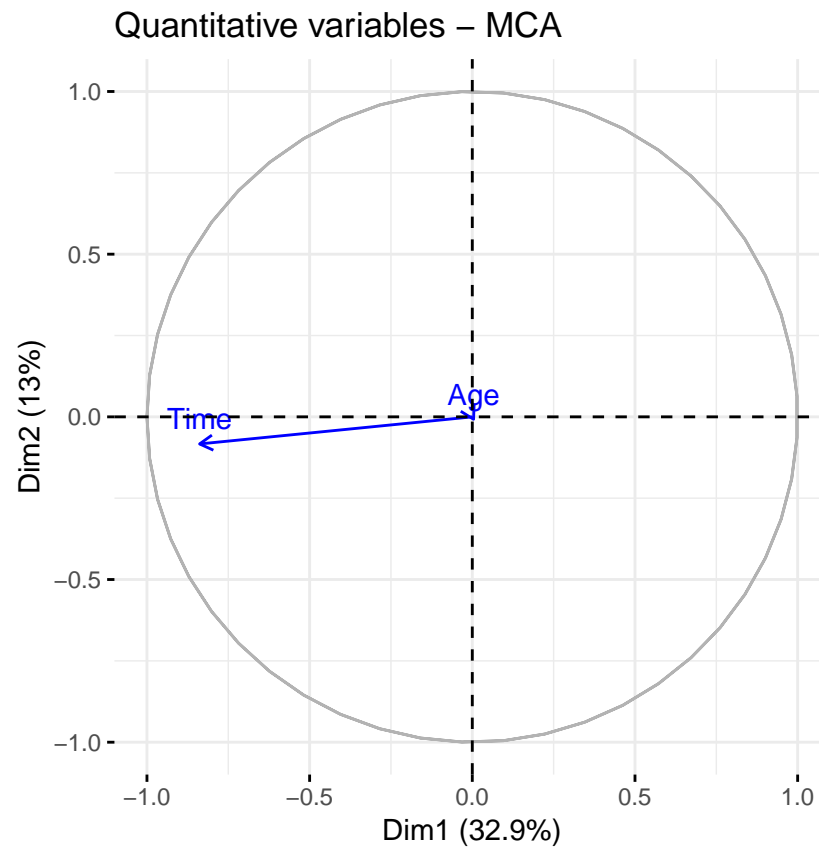
```
fviz_mca_var(res.mca, repel = TRUE,
              ggtheme= theme_minimal())
```

Variable categories – MCA



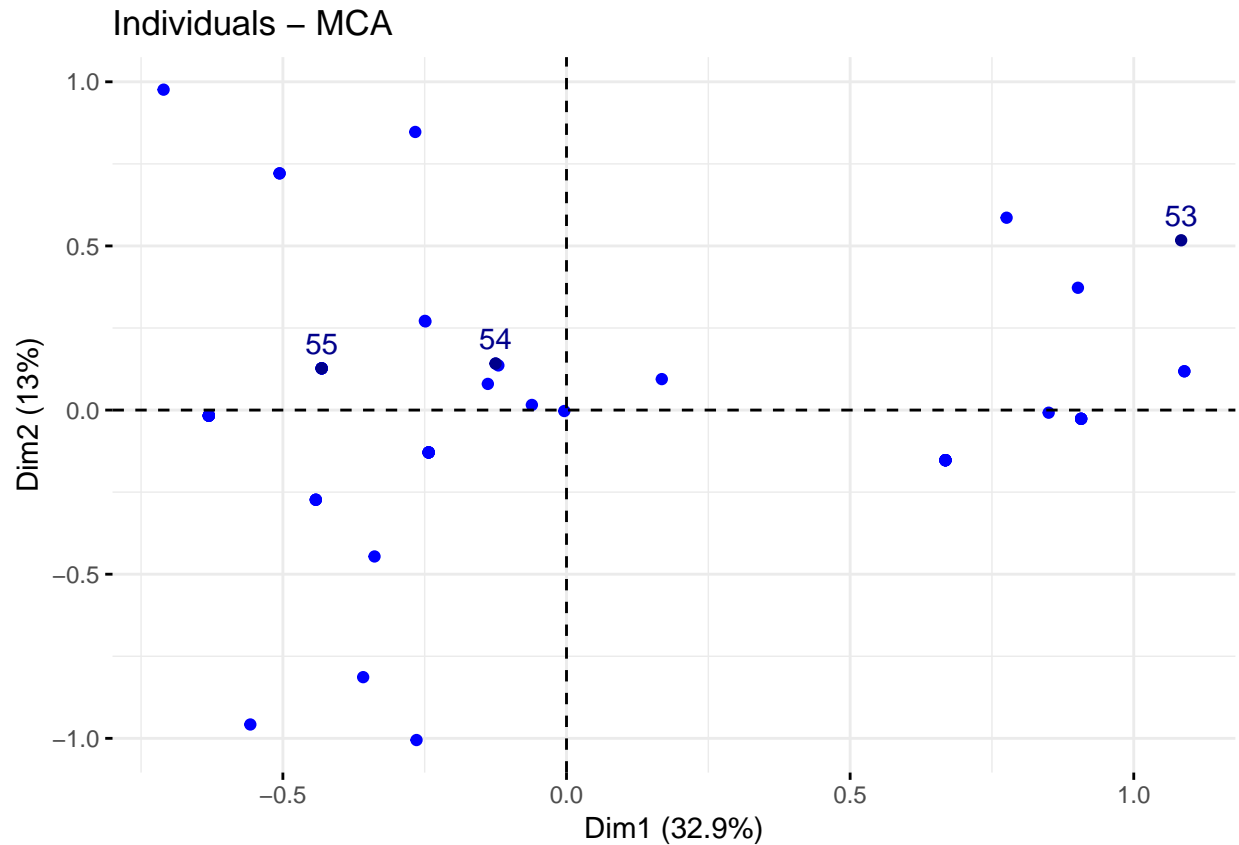
Pour les variables quantitatives supplémentaires, tapez ceci:

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



Pour visualiser les individus supplémentaires, tapez ceci:

```
fviz_mca_ind(res.mca,  
              label = "ind.sup",  
              ggtheme = theme_minimal())
```

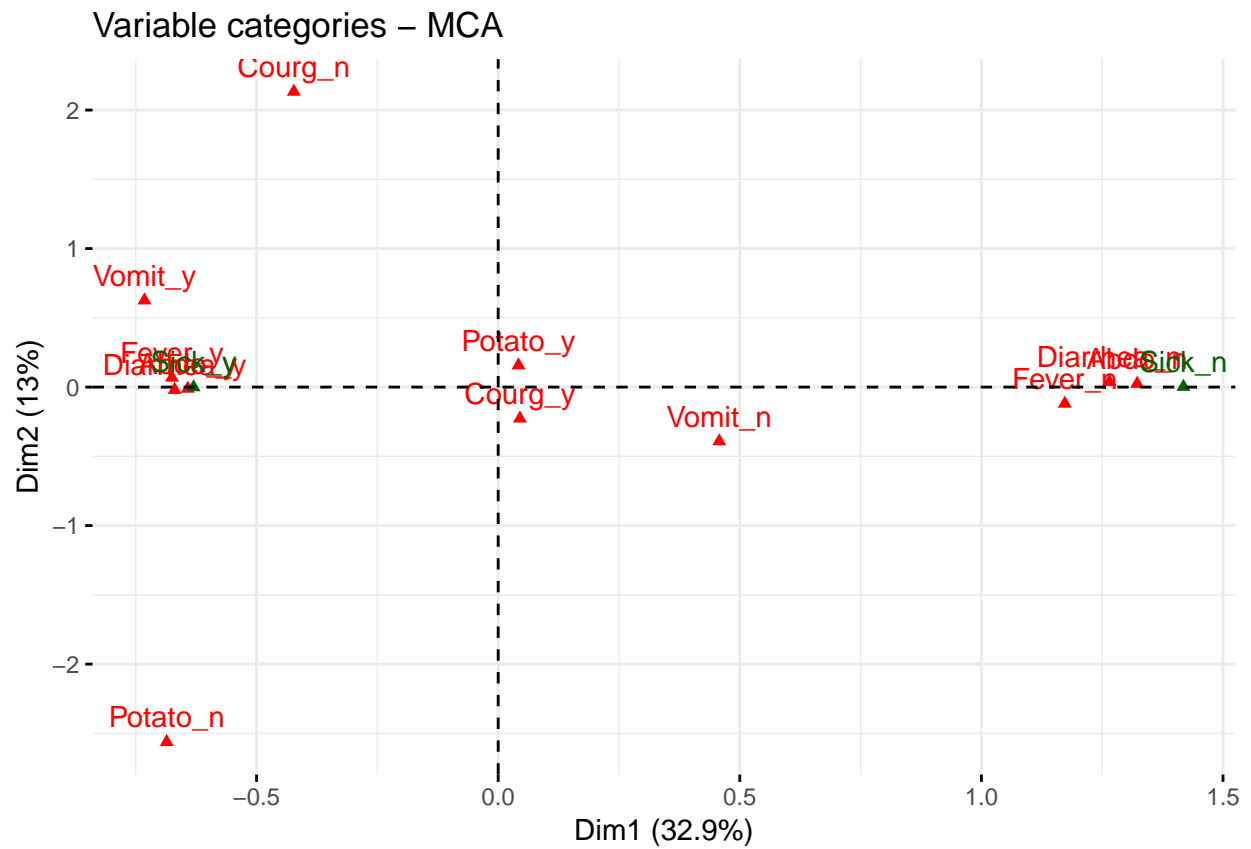
Filtrer des résultats

Si vous avez plusieurs individus / variables, il est possible de visualiser seulement certains d'entre eux en utilisant les arguments `select.ind` et `select.var`.

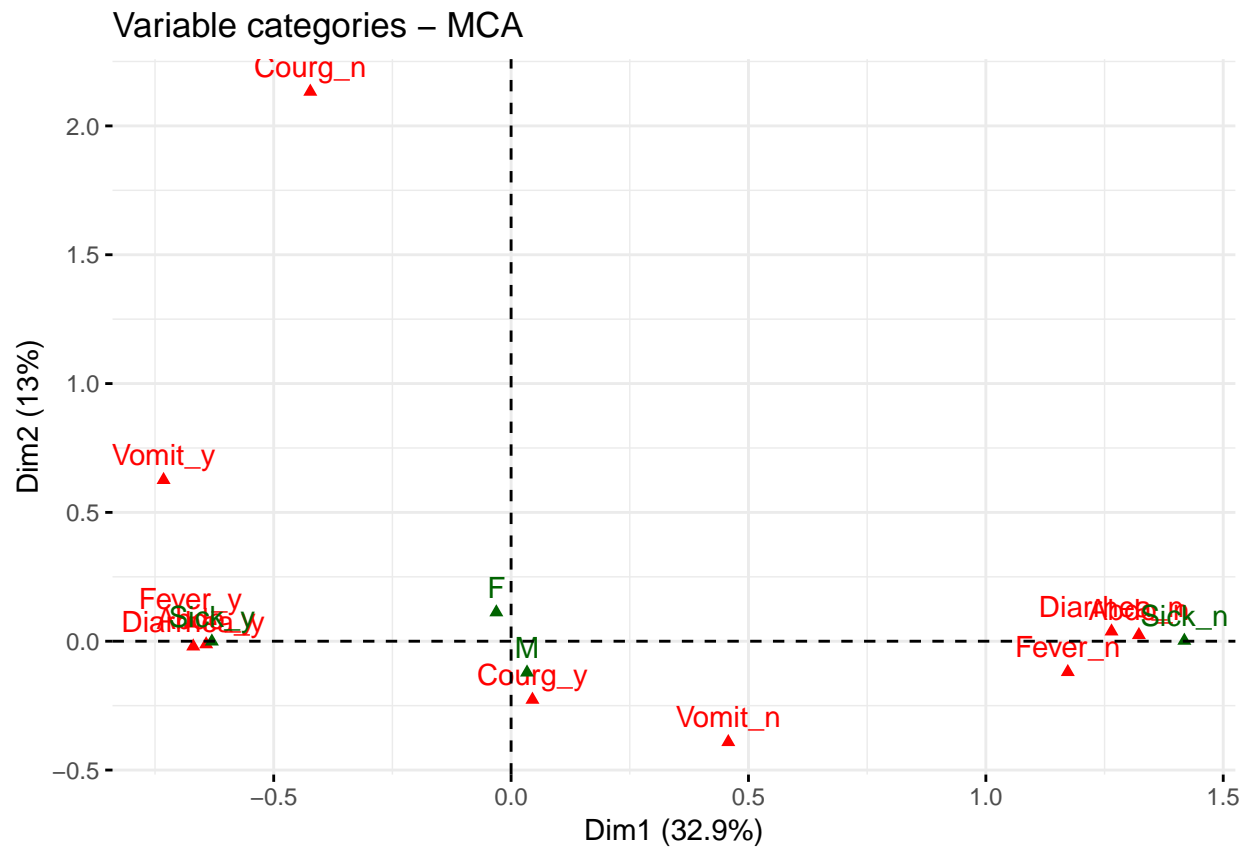
`select.ind`, `select.var`: une sélection d'individus / variables à visualiser. Les valeurs autorisées sont NULL ou une liste contenant le nom des arguments, `cos2` ou `contrib`:

`name`: est un vecteur de caractères contenant le nom des individus / variables à visualiser `cos2`: si `cos2` est dans $[0, 1]$, ex: 0.6, alors les individus / variables avec un `cos2` > 0.6 sont montrés. si `cos2` > 1, ex: 5, le top 5 des individus / variables actifs ainsi que le top 5 des individus / variables supplémentaires avec le `cos2` le plus élevé sont montrés `contrib`: si `contrib` > 1, ex: 5, alors les top 5 individus / variables avec les contributions les plus importantes sont montrés

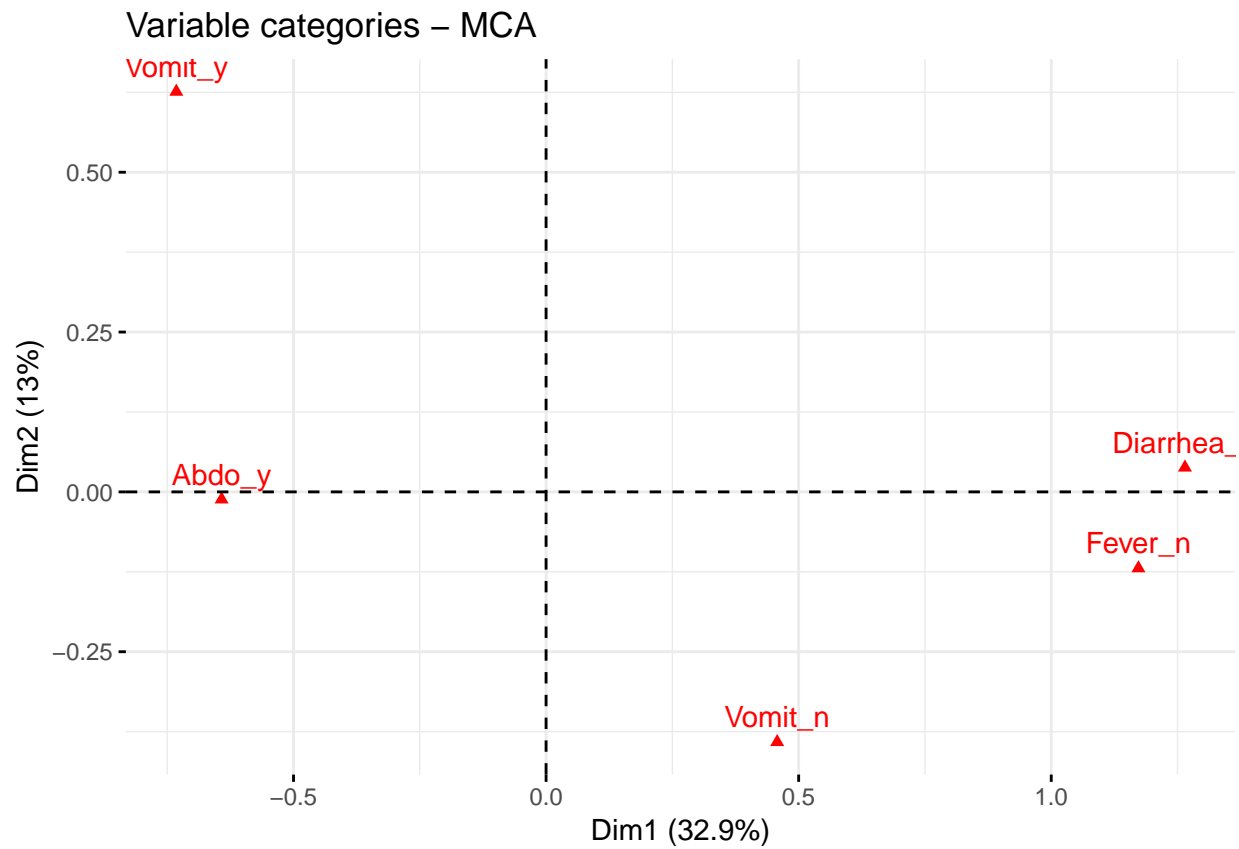
```
# Visualiser les catégories de variables avec cos2 >= 0.4
fviz_mca_var(res.mca, select.var = list(cos2 = 0.4))
```



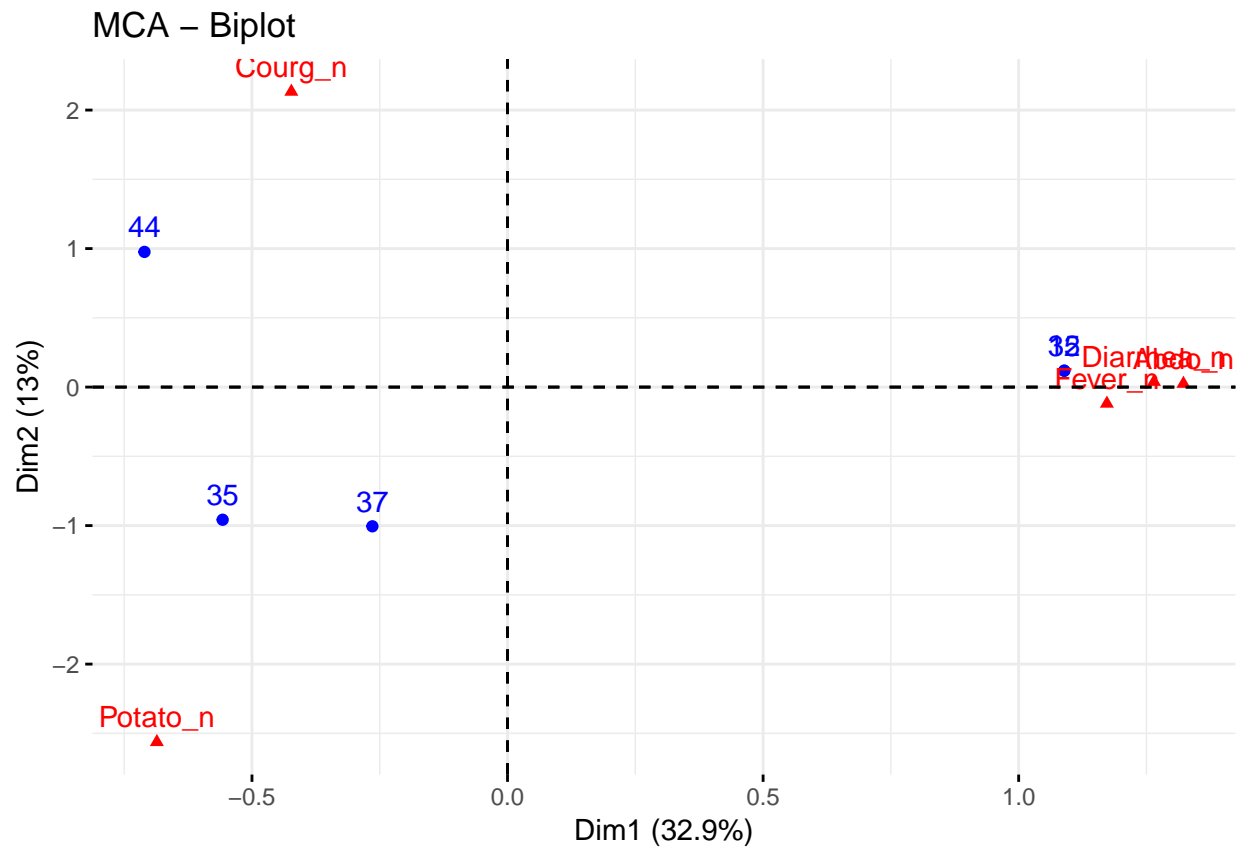
```
# Top 10 des variables actives avec le cos2 le plus élevé
fviz_mca_var (res.mca, select.var = list (cos2 = 10))
```



```
# Sélectionner par noms
name <- list(name = c("Fever_n", "Abdo_y", "Diarrhea_n",
                      "Fever_Y", "Vomit_y", "Vomit_n"))
fviz_mca_var(res.mca, select.var = name)
```



```
# Top 5 des categories de variables les plus contributifs  
fviz_mca_biplot (res.mca, select.ind = list (contrib = 5),  
  select.var = list (contrib = 5),  
  ggtheme = theme_minimal ())
```



Lorsque la sélection se fait selon les valeurs de contribution, les individus / variables supplémentaires ne sont pas représentés parce qu'ils ne contribuent pas à la construction des axes.