

Analyse des Correspondances Multiples (ACM) avec FactoMineR

François Husson

Script et sorties R de la vidéo du didacticiel sur les données de consommation du thé. Le jeu de données `the.csv` est disponible sur le site Web : <http://factominer.free.fr/livre>.

Importation du jeu de données depuis le site

```
the <- read.table("http://factominer.free.fr/livre/the.csv",
                 header=TRUE, sep=";", fileEncoding="latin1")
```

`header=TRUE` : précise que le nom des variables est présent

`sep=";"` : précise que le séparateur de colonnes est le point-virgule (fréquent dans les fichiers csv, pour une tabulation il faudrait écrire `sep="\t"`)

```
##          ptt.dèj          goûter          soirée          après.déjeuner
## Pas.ptt dèj:156  goûter    :169  Pas.soirée:197  après dèj    : 44
## ptt dèj      :144  Pas.goûter:131  soirée      :103  Pas.après dèj:256
##
##
##
##          après.dîner          tt.moment          maison
## après dîner    : 21  Pas.tt moment:197  maison      :291
## Pas.après dîner:279  tt moment   :103  Pas.maison: 9
##
##
##
##          travail          salon.de.t          amis          resto
## Pas.travail:213  Pas.salon de t:242  amis        :196  Pas.resto:221
## travail      : 87  salon de t    : 58  Pas.amis:104  resto      : 79
##
##
##
##          bar          variété          comment          sucre
## bar      : 63  noir    : 74  autre : 9  Pas.sucré:155
## Pas.bar:237  parfumé:193  citron: 33  sucré   :145
##          vert    : 33  lait   : 63
##          pur     :195
##
##
```

```

##
##      forme          lieuachat          type      sexe
## sachet      :170   GMS              :192   t_bas_de_gamme : 7   F:178
## sachet+vrac: 94   GMS+mag.spé.      : 78   t_haut_de_gamme: 53  H:122
## vrac        : 36   magasin spécialisé: 30   t_inconnu      : 12
##
##                                t_marque_connue: 95
##                                t_MDD           : 21
##                                t_variable       :112
##
##      CSP          Sport          âge          âge_Q
## autre actif:20   Pas.sportif:121   Min.      :15.00   15-24 :92
## cadre moyen:40   sportif          :179   1st Qu.:23.00   25-34 :69
## cadre sup   :35                                Median   :32.00   35-44 :40
## employé    :59                                Mean     :37.05   45-59 :61
## étudiant   :70                                3rd Qu.:48.00   60 et +:38
## non actif  :64                                Max.     :90.00
## ouvrier    :12
##
##      fréquence          évasion.exotisme          spiritualité
## + de 2/jour:127   évasion-exotisme      :142   Pas.spiritualité:206
## 1 à 2/sem   : 44   Pas.évasion-exotisme:158   spiritualité    : 94
## 1/jour      : 95
## 3 à 6/sem   : 34
##
##
##
##      bon.pr.la.santé          diurétique          convivialité
## bon pr la santé      :210   diurétique      :174   convivialité    :242
## Pas.bon pr la santé: 90   Pas.diurétique:126   Pas.convivialité: 58
##
##
##
##
##      absorption.fer          féminin          raffiné
## absorption fer      : 31   féminin        :129   Pas.raffiné: 85
## Pas.absorption fer:269   Pas.féminin:171   raffiné       :215
##
##
##
##
##      amaigrissant          excitant          relaxant
## amaigrissant      : 45   excitant       :116   Pas.relaxant:113
## Pas.amaigrissant:255   Pas.excitant:184   relaxant      :187
##
##
##
##
##      ss.effet.santé
## Pas.ss effet santé:234
## ss effet santé    : 66
##
##

```

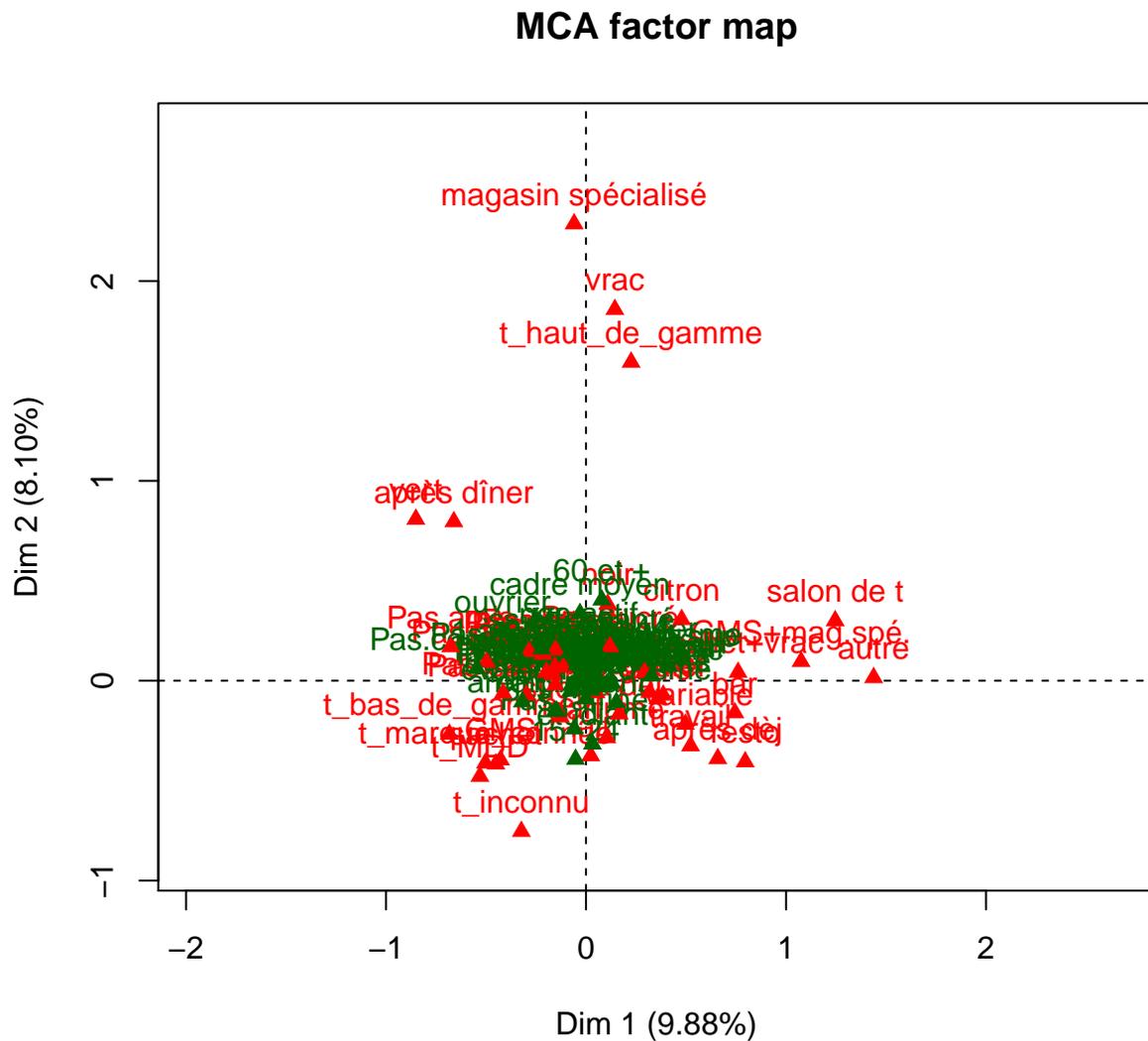
```
##  
##  
##
```

Chargement de FactoMineR

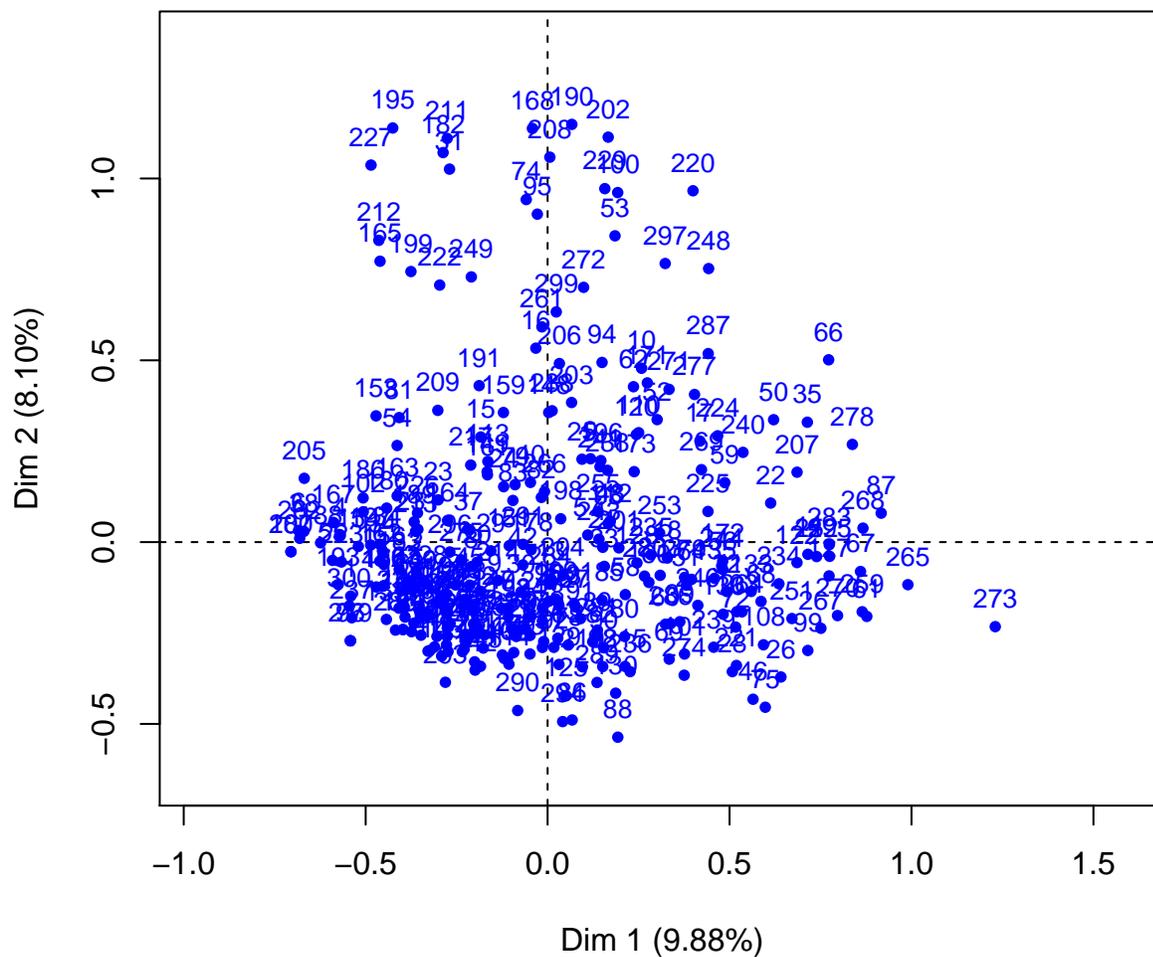
```
library(FactoMineR)
```

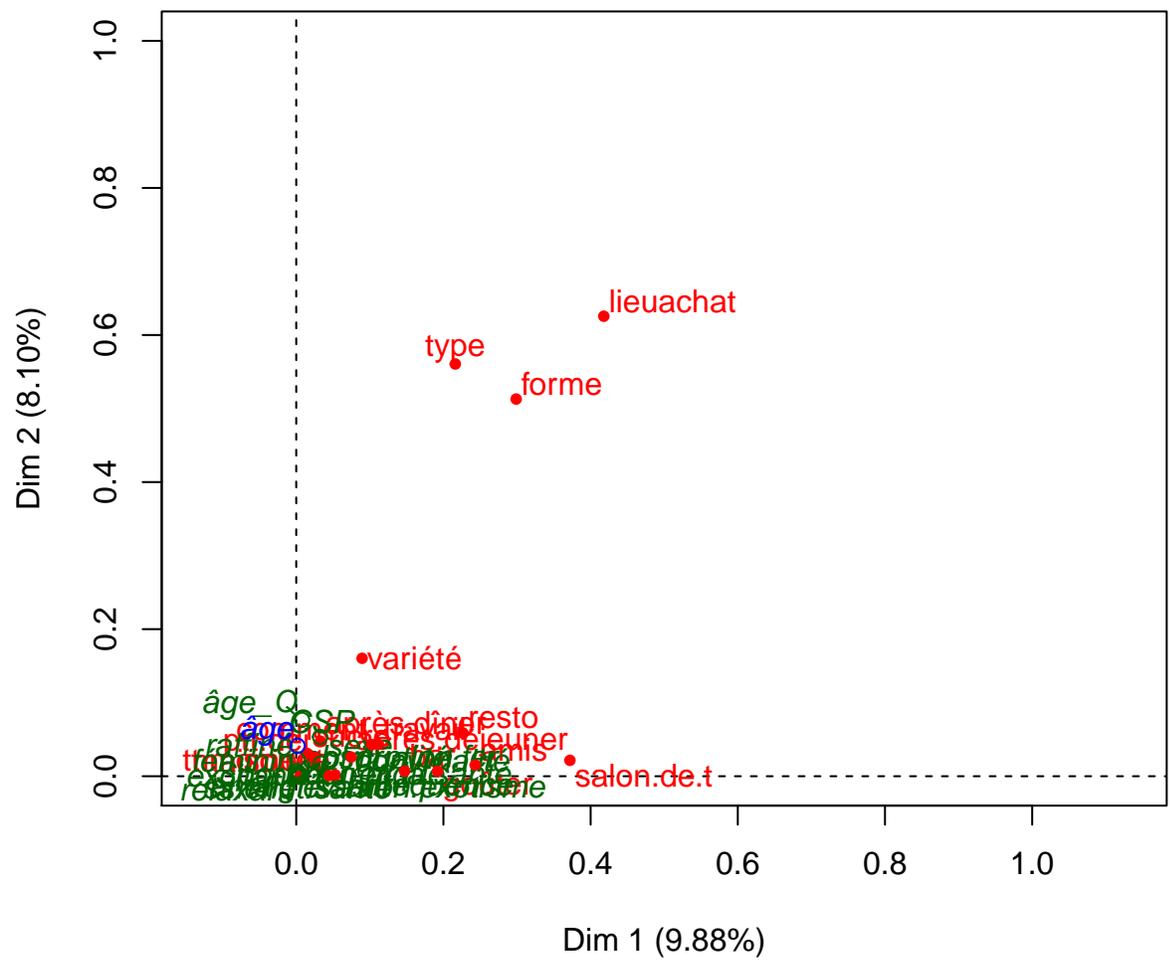
L'ACM et les graphes par défaut

```
res <- MCA(the, quanti.sup=22, quali.sup=c(19:21,23:36))
```

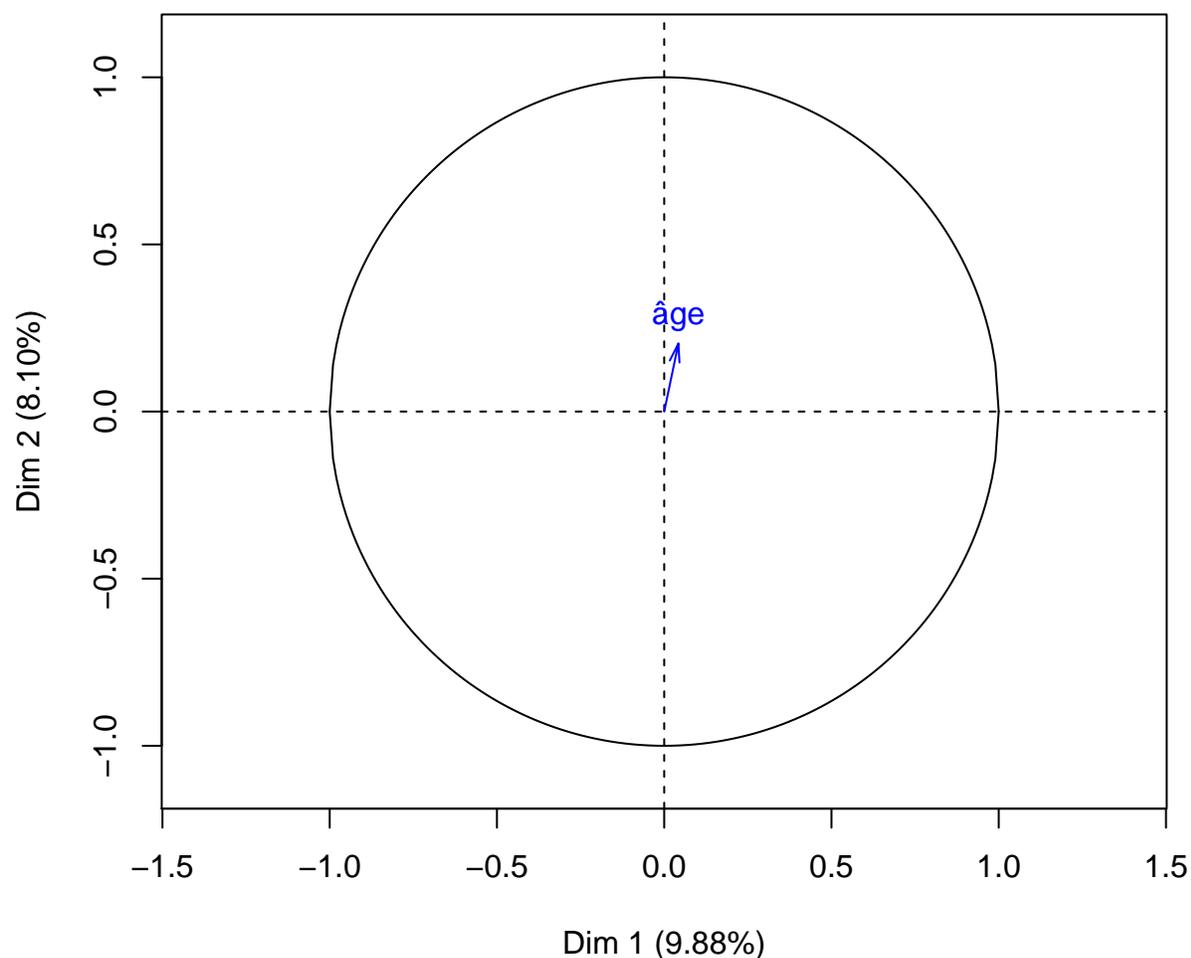


MCA factor map





Supplementary variables on the MCA factor map



Résumé des principaux résultats

Pour éviter de surcharger les sorties, on a précisé dans la fonction `summary.MCA`, appelée par `summary`, que l'on veut les résultats sur les 8 (`nbelements=8`) premiers éléments (individus et modalités), et uniquement sur les 2 premières dimensions (`ncp=2`)

```
summary(res, ncp=2, nbelements=8) ## fonction summary.MCA
```

```
##  
## Call:  
## MCA(X = the, quanti.sup = 22, quali.sup = c(19:21, 23:36))  
##  
##  
## Eigenvalues  
##
```

	Dim.1	Dim.2	Dim.3	Dim.4	Dim.5	Dim.6
--	-------	-------	-------	-------	-------	-------

```

## Variance          0.148  0.122  0.090  0.078  0.074  0.071
## % of var.        9.885  8.103  6.001  5.204  4.917  4.759
## Cumulative % of var. 9.885 17.988 23.989 29.192 34.109 38.868
##                   Dim.7  Dim.8  Dim.9  Dim.10  Dim.11  Dim.12
## Variance          0.068  0.065  0.062  0.059  0.057  0.054
## % of var.         4.522  4.355  4.123  3.902  3.805  3.628
## Cumulative % of var. 43.390 47.745 51.867 55.769 59.574 63.202
##                   Dim.13  Dim.14  Dim.15  Dim.16  Dim.17  Dim.18
## Variance          0.052  0.049  0.048  0.047  0.046  0.040
## % of var.         3.462  3.250  3.221  3.127  3.037  2.683
## Cumulative % of var. 66.664 69.914 73.135 76.262 79.298 81.982
##                   Dim.19  Dim.20  Dim.21  Dim.22  Dim.23  Dim.24
## Variance          0.038  0.037  0.036  0.035  0.031  0.029
## % of var.         2.541  2.438  2.378  2.323  2.055  1.915
## Cumulative % of var. 84.523 86.961 89.339 91.662 93.717 95.633
##                   Dim.25  Dim.26  Dim.27
## Variance          0.027  0.021  0.017
## % of var.         1.821  1.407  1.139
## Cumulative % of var. 97.454 98.861 100.000
##
## Individuals (the 8 first)
##                   Dim.1  ctr  cos2  Dim.2  ctr  cos2
## 1 | -0.541  0.658  0.143 | -0.149  0.061  0.011 |
## 2 | -0.361  0.293  0.133 | -0.078  0.017  0.006 |
## 3 |  0.073  0.012  0.003 | -0.169  0.079  0.018 |
## 4 | -0.572  0.735  0.235 |  0.018  0.001  0.000 |
## 5 | -0.253  0.144  0.079 | -0.118  0.038  0.017 |
## 6 | -0.684  1.053  0.231 |  0.032  0.003  0.001 |
## 7 | -0.111  0.027  0.022 | -0.182  0.090  0.059 |
## 8 | -0.210  0.099  0.043 | -0.068  0.013  0.004 |
##
## Categories (the 8 first)
##                   Dim.1  ctr  cos2  v.test  Dim.2  ctr  cos2  v.test
## Pas.ptt dèj | -0.153  0.457  0.025 -2.756 |  0.154  0.560  0.026  2.764
## ptt dèj |  0.166  0.495  0.025  2.756 | -0.166  0.607  0.026 -2.764
## goûter |  0.386  3.142  0.192  7.578 | -0.072  0.135  0.007 -1.423
## Pas.goûter | -0.498  4.053  0.192 -7.578 |  0.093  0.174  0.007  1.423
## Pas.soirée | -0.167  0.683  0.053 -3.985 |  0.030  0.028  0.002  0.728
## soirée |  0.319  1.307  0.053  3.985 | -0.058  0.053  0.002 -0.728
## après dèj |  0.659  2.385  0.075  4.722 | -0.390  1.018  0.026 -2.793
## Pas.après dèj | -0.113  0.410  0.075 -4.722 |  0.067  0.175  0.026  2.793
##
## Pas.ptt dèj |
## ptt dèj |
## goûter |
## Pas.goûter |
## Pas.soirée |
## soirée |
## après dèj |
## Pas.après dèj |
##
## Categorical variables (eta2)
##                   Dim.1  Dim.2
## ptt.dèj | 0.025  0.026 |

```

```

## goûter          | 0.192 0.007 |
## soirée          | 0.053 0.002 |
## après.déjeuner | 0.075 0.026 |
## après.dîner    | 0.033 0.048 |
## tt.moment      | 0.045 0.001 |
## maison         | 0.005 0.000 |
## travail        | 0.112 0.043 |
##
## Supplementary categories (the 8 first)
##          Dim.1  cos2 v.test  Dim.2  cos2 v.test
## F        | 0.151 0.033 3.158 | -0.109 0.017 -2.278 |
## H        | -0.221 0.033 -3.158 | 0.159 0.017 2.278 |
## autre actif | 0.040 0.000 0.187 | 0.013 0.000 0.061 |
## cadre moyen | -0.030 0.000 -0.205 | 0.336 0.017 2.281 |
## cadre sup   | 0.415 0.023 2.608 | 0.072 0.001 0.452 |
## employé    | -0.153 0.006 -1.313 | -0.151 0.006 -1.289 |
## étudiant   | 0.032 0.000 0.305 | -0.317 0.031 -3.022 |
## non actif  | -0.036 0.000 -0.324 | 0.185 0.009 1.666 |
##
## Supplementary categorical variables (eta2)
##          Dim.1 Dim.2
## sexe     | 0.033 0.017 |
## CSP      | 0.032 0.053 |
## Sport    | 0.001 0.000 |
## âge_Q    | 0.008 0.077 |
## fréquence | 0.094 0.006 |
## évasion.exotisme | 0.000 0.007 |
## spiritualité | 0.005 0.000 |
## bon.pr.la.santé | 0.000 0.000 |
##
## Supplementary continuous variable
##          Dim.1  Dim.2
## âge      | 0.042 | 0.204 |

```

La fonction `summary.MCA` permet de précéder le nombre de composantes avec `ncp` pour lequel les résultats sont fournis, et l'argument `nbelements=Inf` permet de préciser qu'on veut les résultats sur tous les individus.

```
summary(res, ncp=3, nbelements=Inf)
```

Description des dimensions

On peut décrire les dimensions de l'ACM avec la commande suivante. Par défaut, les 3 premières dimensions sont décrites.

```
dimdesc(res)
```

Pour éviter d'avoir trop de modalités, on ne décrit que les axes 1 et 2 et on conserve uniquement les modalités qui ont une probabilité critique inférieure à 10^{-5} au lieu de 0.05 par défaut.

```
dimdesc(res, axes=c(1,2), proba=1e-05)
```

```
## $`Dim 1`  
## $`Dim 1`$quali  
##           R2      p.value  
## lieuachat    0.41793014 1.255462e-35  
## salon.de.t   0.37189109 6.082138e-32  
## forme        0.29882863 1.273180e-23  
## amis         0.24319952 8.616289e-20  
## resto        0.22646759 2.319804e-18  
## goûter       0.19203800 1.652462e-15  
## type         0.21609382 4.050469e-14  
## bar          0.14722360 5.846592e-12  
## travail      0.11153590 3.000872e-09  
## comment      0.10285191 4.796010e-07  
## variété      0.08950330 8.970954e-07  
## après.déjeuner 0.07458227 1.570629e-06  
## fréquence    0.09438792 1.849071e-06  
## convivialité 0.07132511 2.706357e-06  
##  
## $`Dim 1`$category  
##           Estimate      p.value  
## GMS+mag.spé.    0.3385378 1.344557e-35  
## salon de t      0.2973107 6.082138e-32  
## sachet+vrac     0.2345703 1.361423e-21  
## amis            0.1995083 8.616289e-20  
## resto           0.2080260 2.319804e-18  
## goûter          0.1701136 1.652462e-15  
## bar             0.1813713 5.846592e-12  
## t_variable      0.2759507 5.956230e-12  
## travail         0.1417041 3.000872e-09  
## + de 2/jour     0.1485561 7.380937e-07  
## après dèj       0.1486264 1.570629e-06  
## convivialité    0.1302039 2.706357e-06  
## autre           0.3819244 9.244911e-06  
## Pas.convivialité -0.1302039 2.706357e-06  
## Pas.après dèj   -0.1486264 1.570629e-06  
## vert           -0.2456910 1.281162e-07  
## Pas.travail     -0.1417041 3.000872e-09  
## t_marque_connue -0.1091079 1.116908e-09  
## Pas.bar         -0.1813713 5.846592e-12  
## Pas.goûter      -0.1701136 1.652462e-15  
## Pas.resto       -0.2080260 2.319804e-18
```

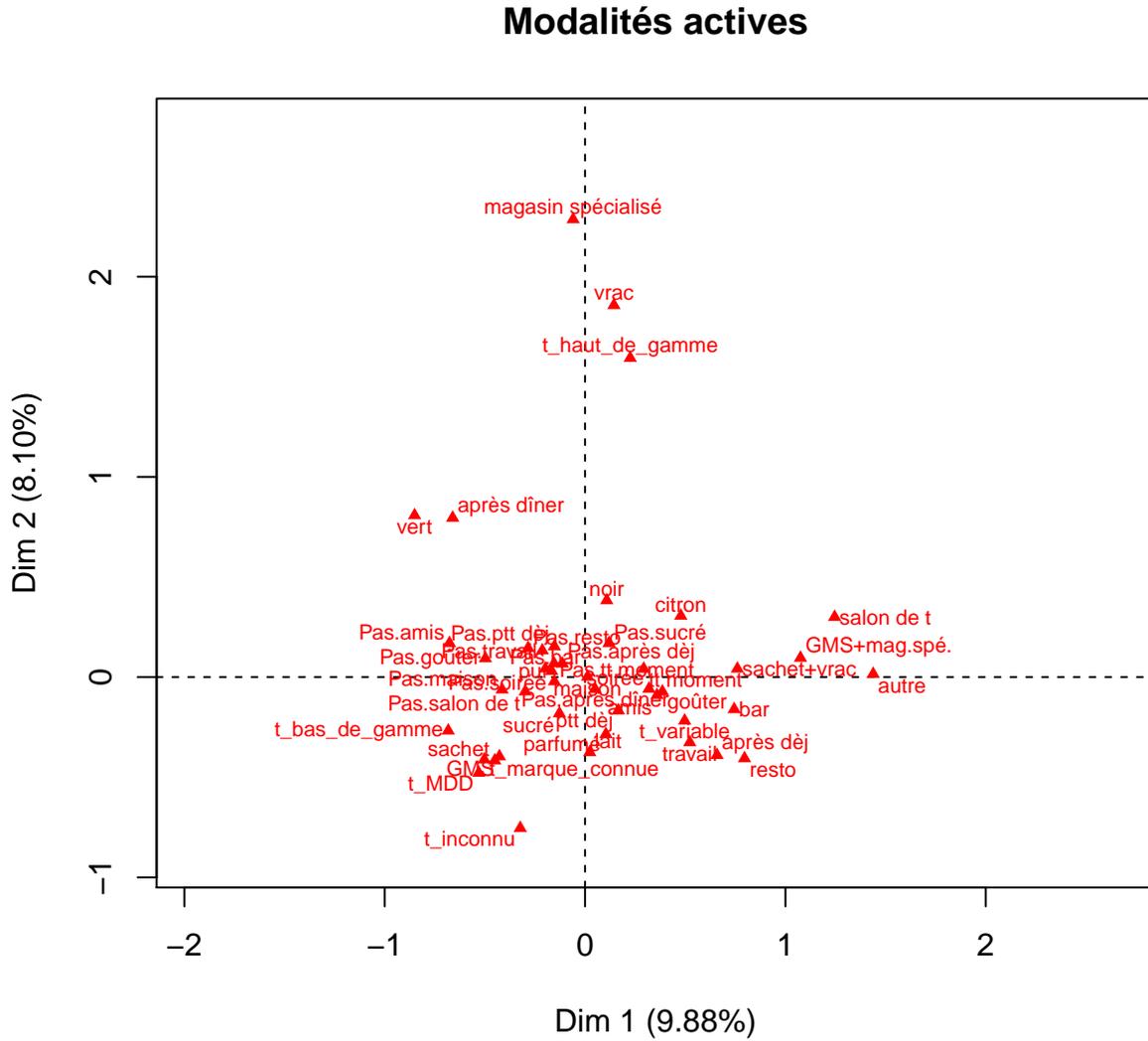
```

## Pas.amis          -0.1995083 8.616289e-20
## sachet            -0.2318245 8.877561e-22
## GMS               -0.2401244 3.008256e-27
## Pas.salon de t   -0.2973107 6.082138e-32
##
##
## $`Dim 2`
## $`Dim 2`$quali
##                R2          p.value
## lieuachat 0.6255019 4.542155e-64
## type      0.5605680 1.837909e-50
## forme     0.5128862 4.103156e-47
## variété   0.1603428 5.359827e-12
##
## $`Dim 2`$category
##                Estimate      p.value
## magasin spécialisé 0.5662393 3.435386e-58
## t_haut_de_gamme   0.5867567 6.819842e-53
## vrac               0.4752358 4.876111e-43
## vert              0.1763655 5.660702e-07
## 15-24             -0.1634255 4.501578e-06
## t_marque_connue   -0.1124757 8.001515e-07
## parfumé           -0.2047044 6.499738e-12
## sachet            -0.3175565 2.123707e-18
## GMS               -0.3689186 5.804094e-23

```


Graphe avec les individus et les modalités supplémentaires invisibles, et donc avec uniquement les modalités actives :

```
plot(res, invisible=c("ind", "quali.sup"), autoLab="y", cex=0.7,
      title="Modalités actives")
```



Sélection de certaines modalités

Les modalités peuvent être sélectionnées avec l'argument `selectMod`.

`selectMod="cos2 0.7"` : sélectionne les modalités actives et supplémentaires qui ont, sur le plan tracé, une qualité de projection supérieure à 0.7

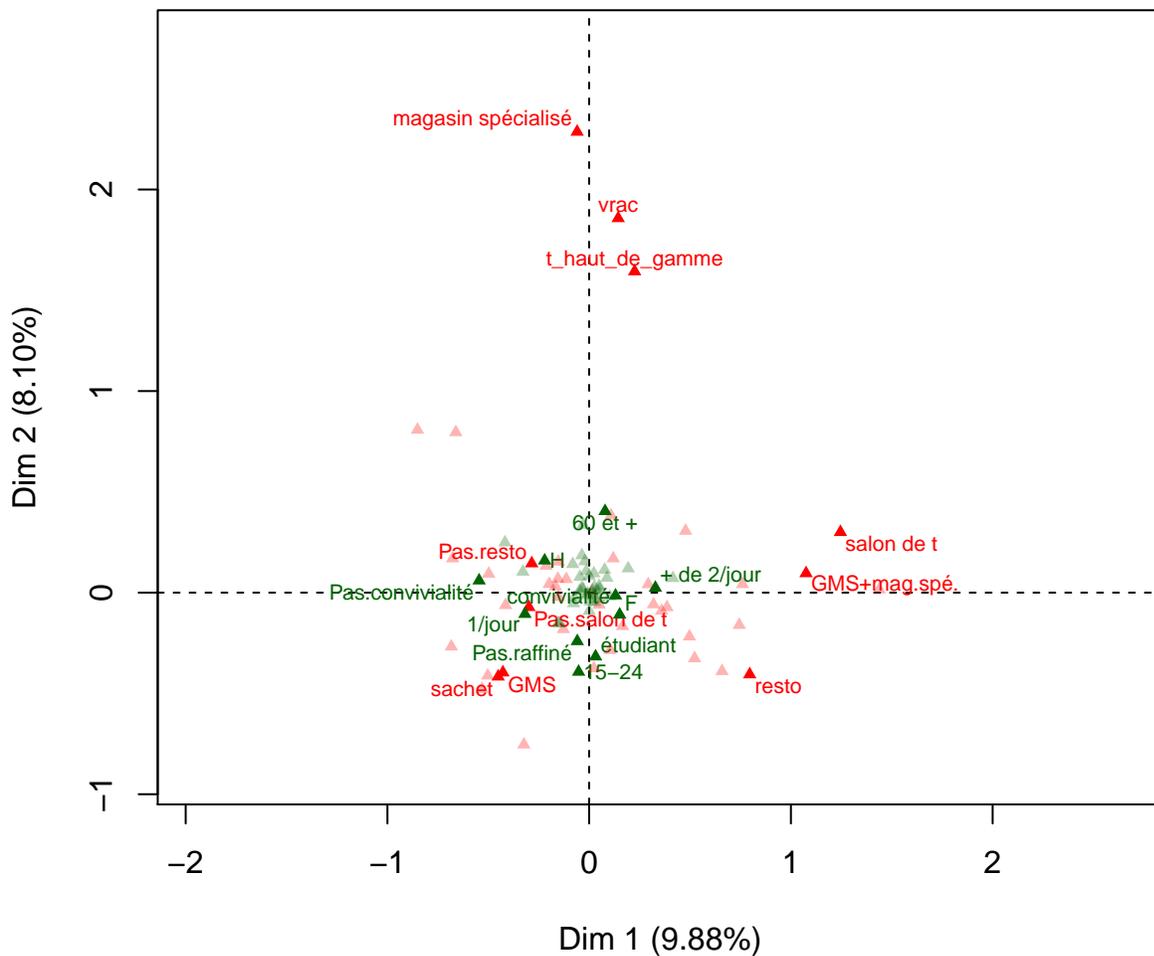
`selectMod="cos2 5"` : sélectionne les 5 modalités, 5 actives et 5 supplémentaires, qui ont la meilleure qualité de projection sur le plan tracé

`selectMod="contrib 5"` : sélectionne les 5 modalités (nécessairement actives) qui ont le plus contribué à la construction du plan tracé

`selectMod="dist 5"` : sélectionne les 5 modalités, 5 actives et (supplémentaires) les plus loin de l'origine des axes

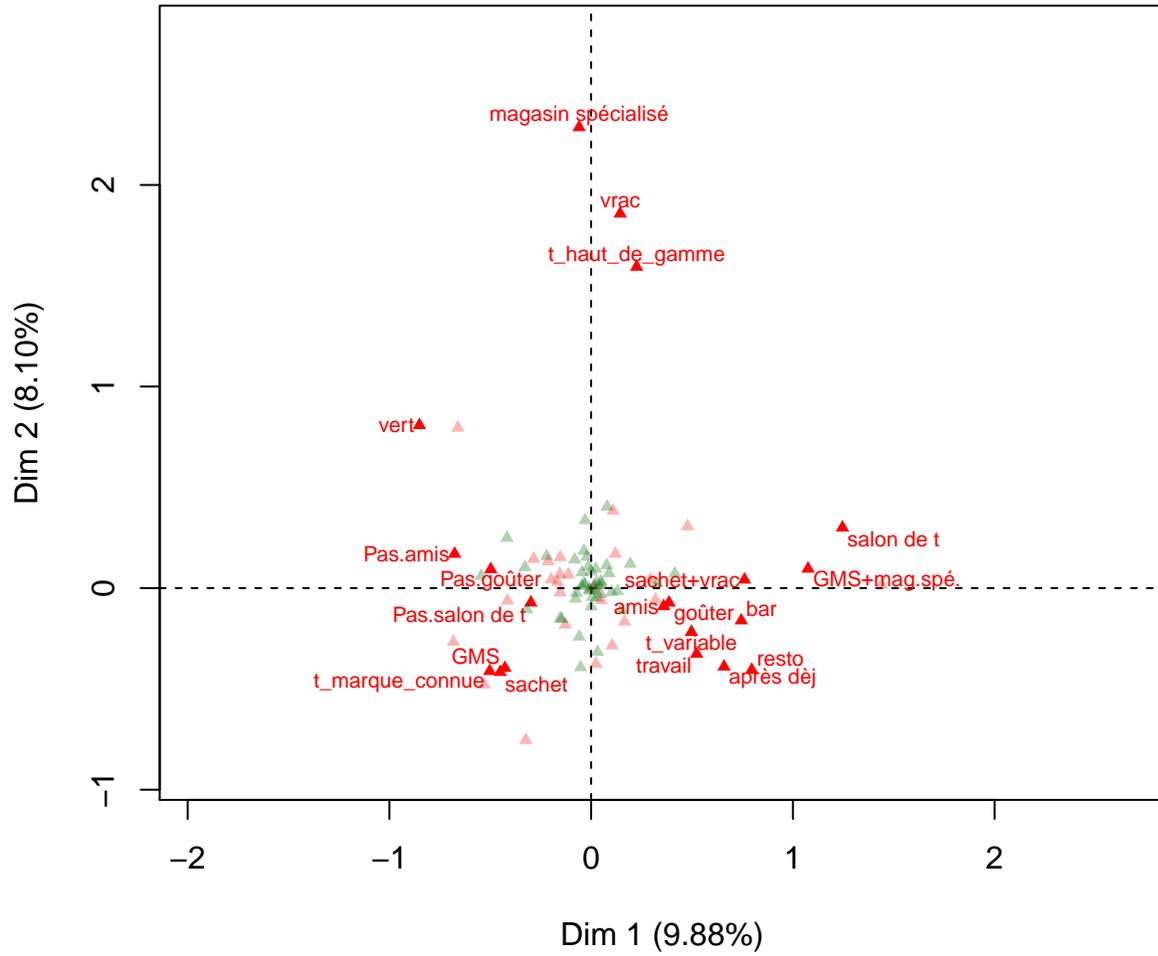
```
plot(res, invisible="ind", autoLab="y", cex=0.7, selectMod="cos2 10")
```

MCA factor map



```
plot(res, invisible="ind", autoLab="y", cex=0.7, selectMod="contrib 20")
```

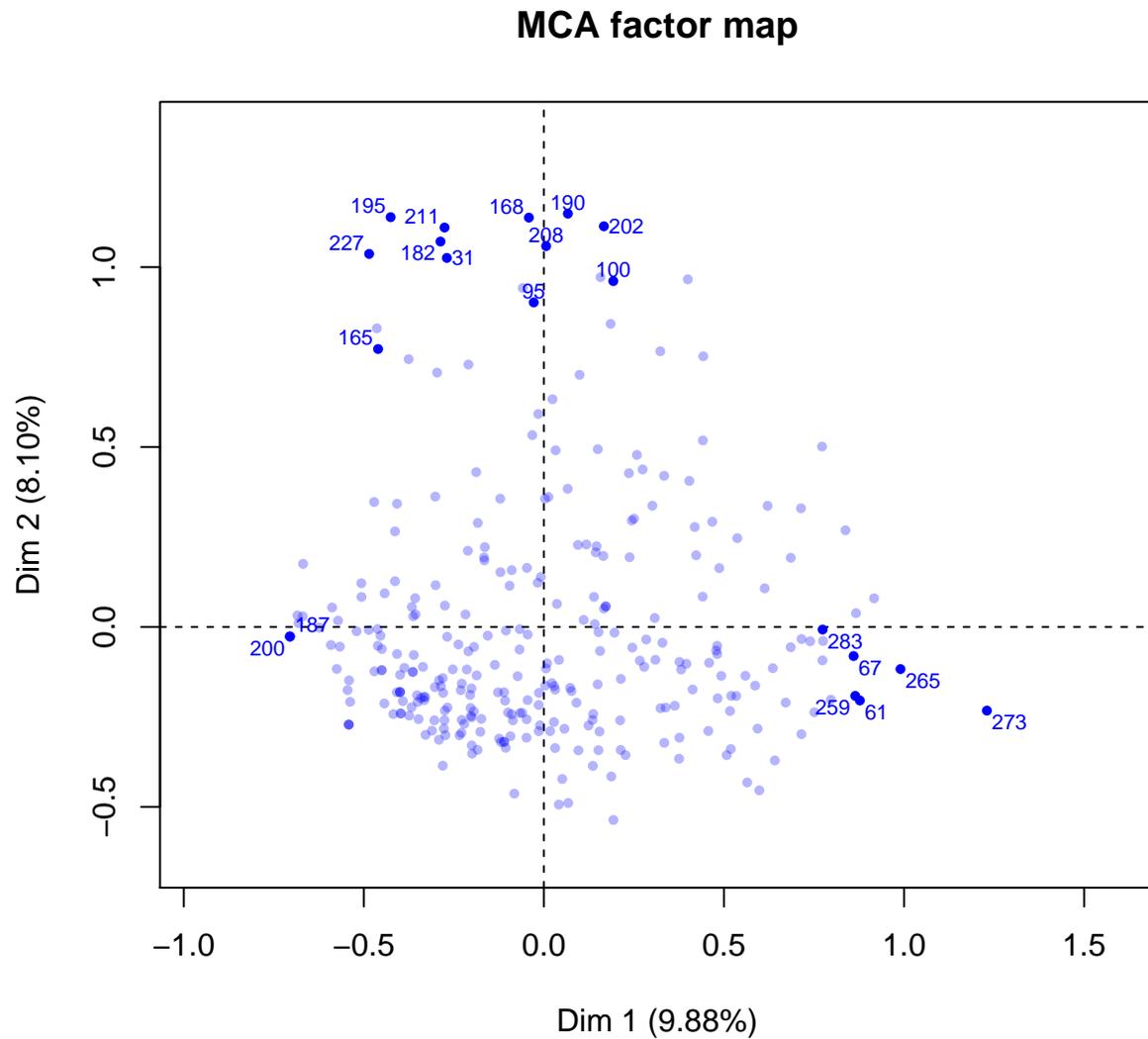
MCA factor map



Sélection des individus

Pour les individus, on utilise l'argument `select` au lieu de `selectMod` exactement de la même façon que pour les modalités.

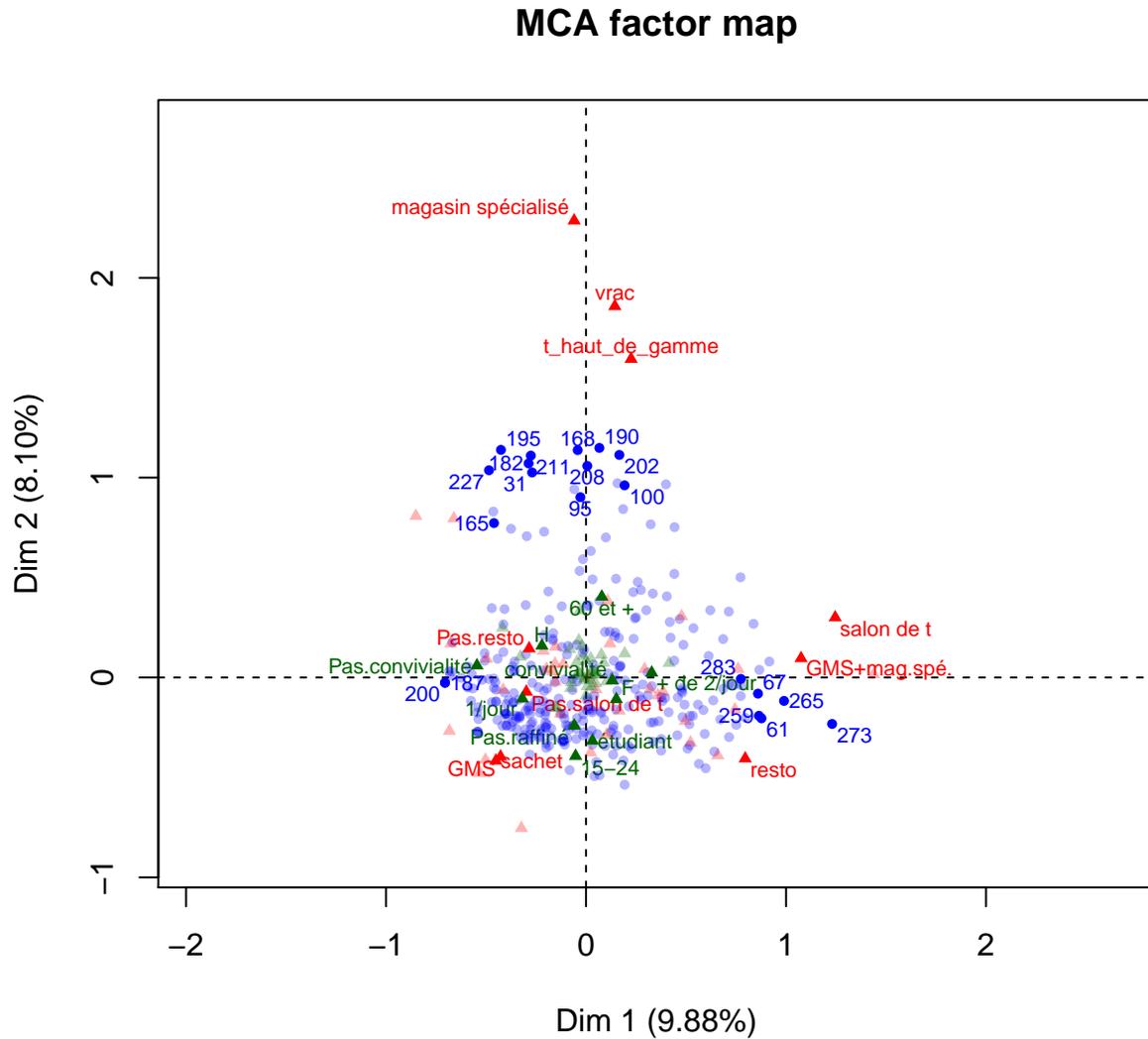
```
plot(res, invisible=c("var", "quali.sup"), autoLab="y", cex=0.7, select="cos2 20")
```



Sélection des individus et des modalités

On peut construire un graphe simultanée et sélectionner à la fois des individus avec `select` et des modalités avec `selectMod`.

```
plot(res, autoLab="y", cex=0.7, select="cos2 20", selectMod="cos2 10")
```

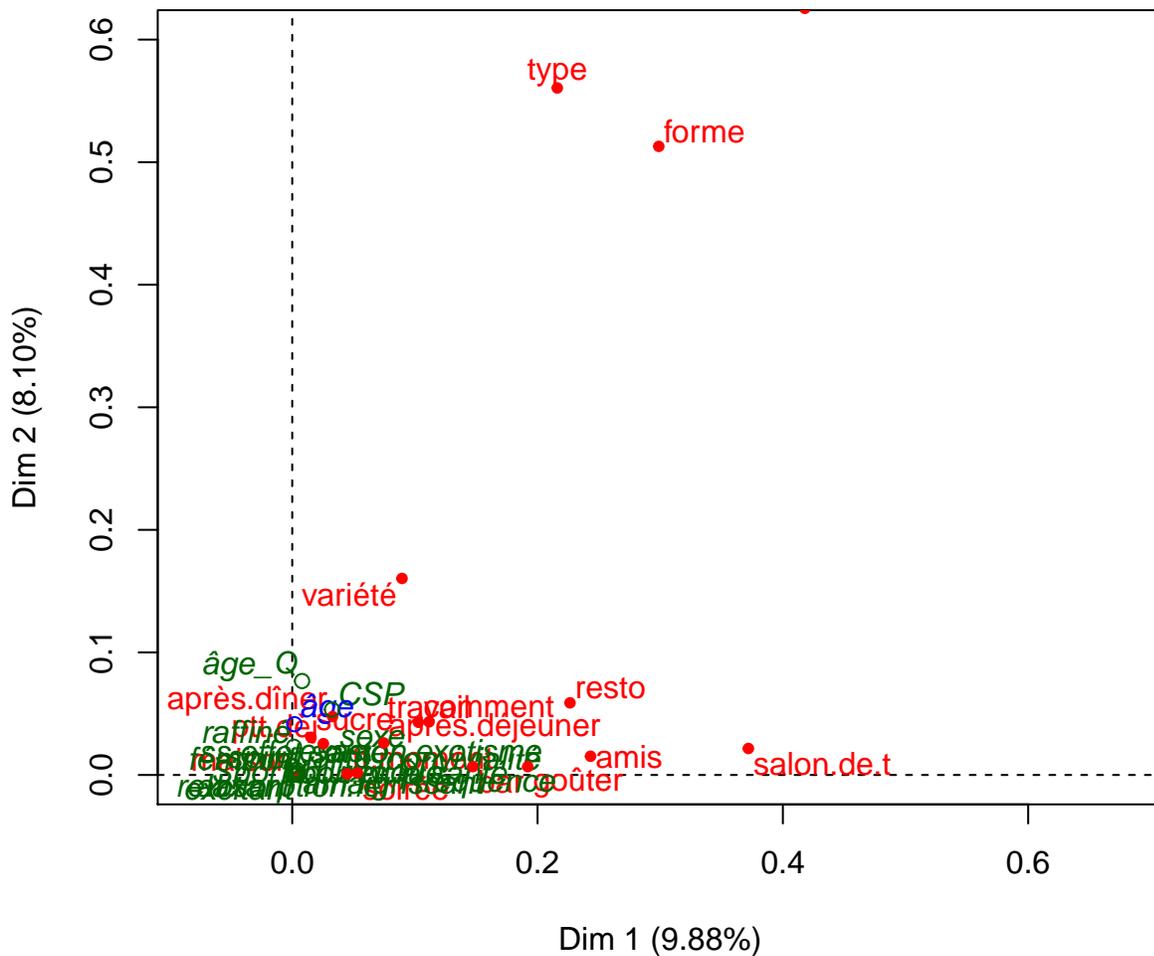


Graphes des variables

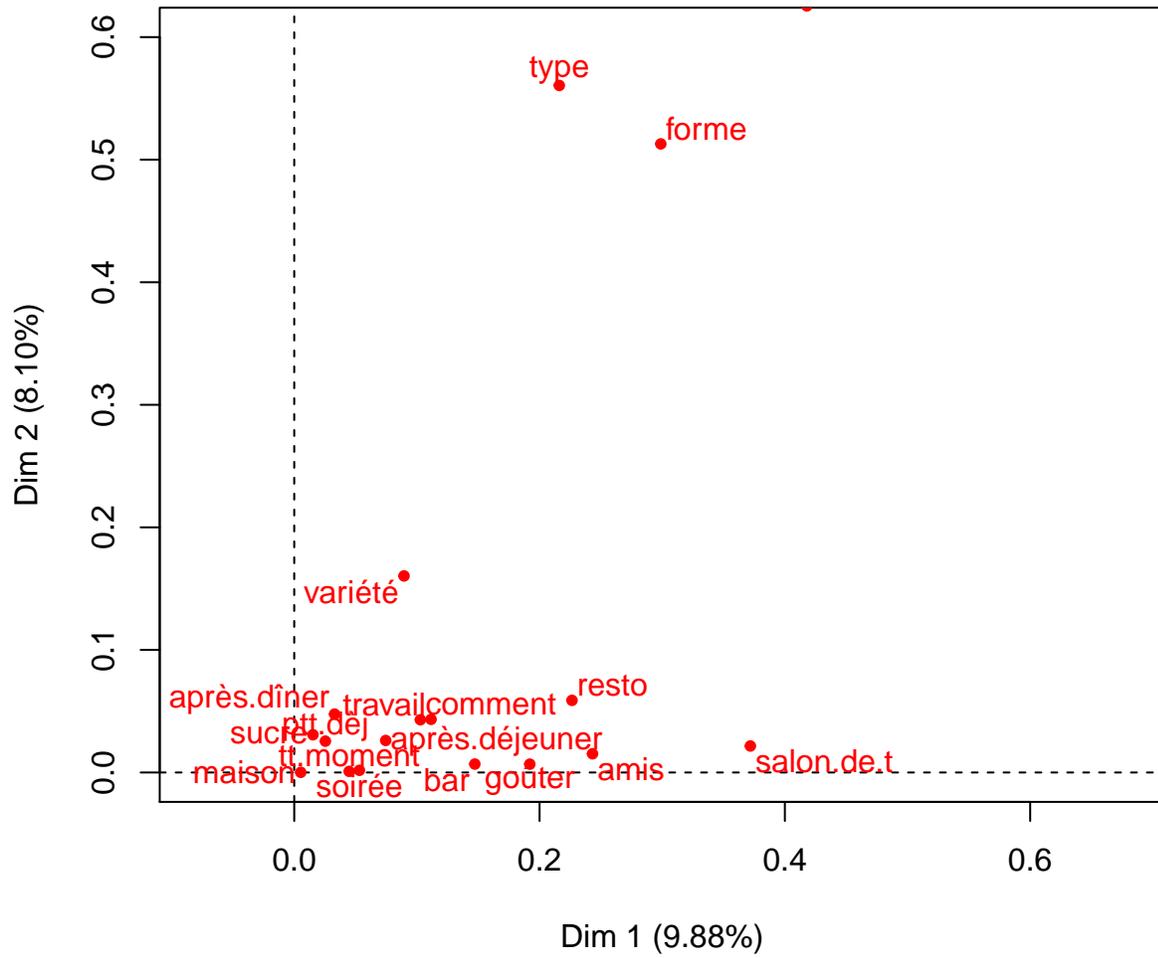
On peut construire le graphe des liaisons. Pour les variables qualitatives, la corrodnnée sur un axe est le rapport de corrélation au carré entre la variable qualitative et la dimension; pour les variables quantitatives supplémentaire, on calcule le coeeficient de corrélation au carré entre la dimension et la variable quantitative.

On zoome ici sur les graphes en précisant les intervalles sur x et y avec xlim et ylim.

```
plot(res, choix="var",xlim=c(0,0.6),ylim=c(0,0.6))
```

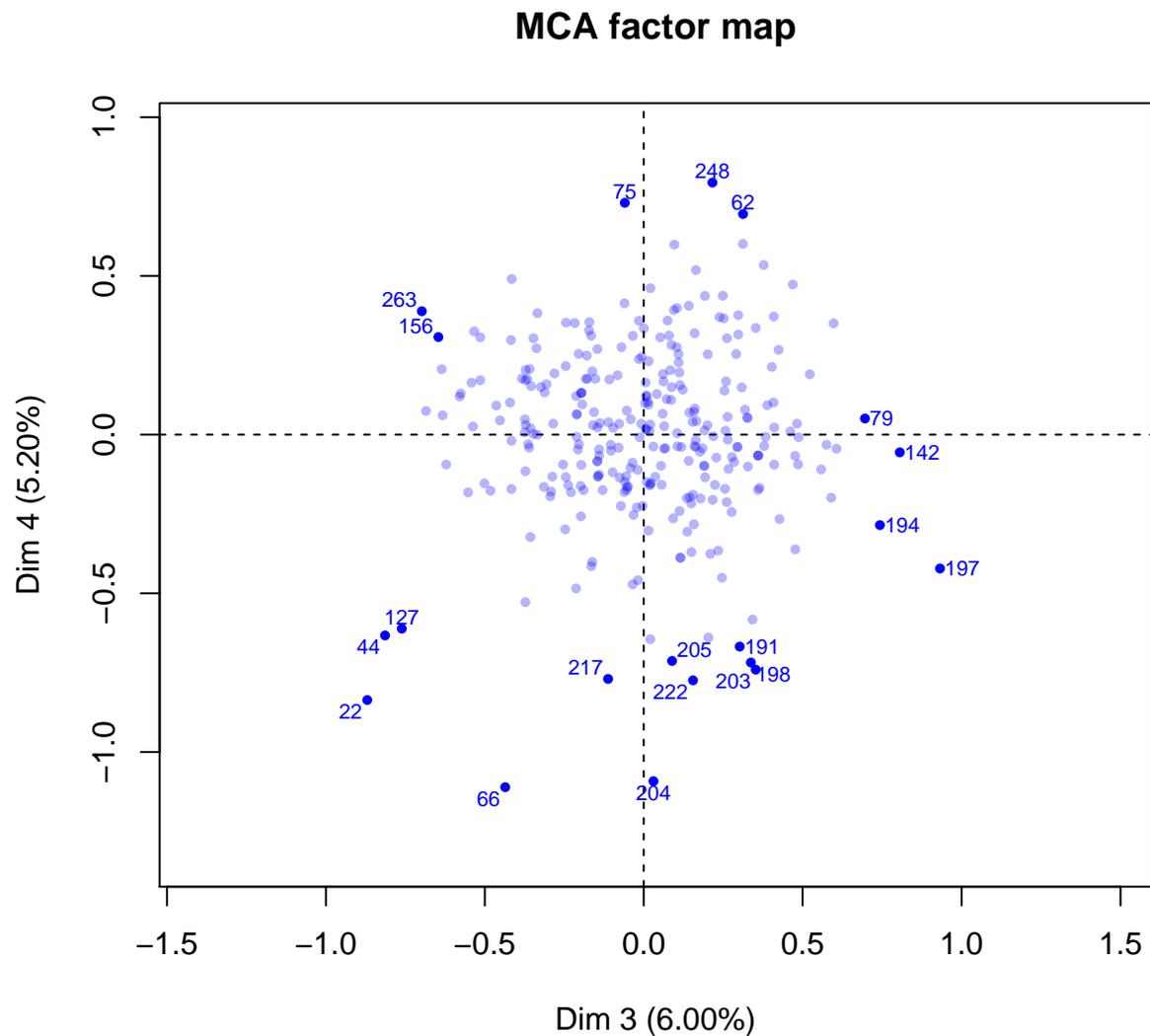


```
plot(res, choix="var",xlim=c(0,0.6),ylim=c(0,0.6),invisible=c("quali.sup","quanti.sup"))
```



Graphes sur les axes 3 et 4

```
plot(res, invisible=c("var", "quali.sup"), cex=0.7, select="contrib 20", axes=3:4)
```



```
plot(res, invisible="ind", autoLab="y", cex=0.7, selectMod="cos2 20", axes=3:4)
```


Ellipses de confiance autour de modalités

On peut construire des ellipses de confiance autour des modalités de chaque variable. Par défaut un graphe est construit par variable. On précise ici avec `keepvar` les variables pour lesquelles nous voulons un graphe.

```
plotellipses(res,keepvar=c(14:17))
```

