

Multiple Correspondence Analysis (MCA) with FactoMineR (tea dataset)

Magalie Houée-Bigot & François Husson

Import data

```
setwd("C:/Users/houee/Downloads") # select the working directory
tea <- read.table("data_MCA_Tea.csv", header=TRUE, sep=";")
```

header=TRUE : indicates that the file contains the names of the variables

sep=";" : indicates the fields separator (usually “;” or “,” for csv files)

It is important to check that the import is well done

```
summary(tea)
```

```
##          breakfast          tea.time          evening          lunch
## breakfast   :144  Not.tea time:131  evening   :103  lunch     : 44
## Not.breakfast:156  tea time    :169  Not.evening:197  Not.lunch:256
##
##
##
##
##          dinner          always          home          work
## dinner     : 21  always   :103  home     :291  Not.work:213
## Not.dinner:279  Not.always:197  Not.home:  9  work    : 87
##
##
##
##          tearoom          friends          resto          pub
## Not.tearoom:242  friends   :196  Not.resto:221  Not.pub:237
## tearoom     : 58  Not.friends:104  resto     : 79  pub      : 63
##
##
##
##          Tea          How          sugar          how
## black   : 74  alone:195  No.sugar:155  tea bag      :170
## Earl Grey:193  lemon: 33  sugar   :145  tea bag+unpacked: 94
## green   : 33  milk : 63          unpackaged   : 36
##
##          other: 9
##
##
```

```

##           where           price           age           sex
## chain store      :192  p_branded      : 95  Min.    :15.00  F:178
## chain store+tea shop: 78  p_cheap      : 7  1st Qu.:23.00  M:122
## tea shop        : 30  p_private label: 21  Median  :32.00
##                :      p_unknown      : 12  Mean    :37.05
##                :      p_upscale      : 53  3rd Qu.:48.00
##                :      p_variable     :112  Max.    :90.00
##
##           SPC           Sport           age_Q           frequency
## employee      :59  Not.sportsman:121  +60 :38  +2/day    :127
## middle        :40  sportsman     :179  15-24:92  1 to 2/week: 44
## non-worker    :64                25-34:69  1/day     : 95
## other worker :20                35-44:40  3 to 6/week: 34
## senior        :35                45-59:61
## student       :70
## workman       :12
##           escape.exoticism           spirituality           healthy
## escape-exoticism :142  Not.spirituality:206  healthy :210
## Not.escape-exoticism:158  spirituality : 94  Not.healthy: 90
##
##
##
##
##           diuretic           friendliness           iron.absorption
## diuretic      :174  friendliness :242  iron absorption : 31
## Not.diuretic:126  Not.friendliness: 58  Not.iron absorption:269
##
##
##
##
##           feminine           sophisticated           slimming
## feminine     :129  Not.sophisticated: 85  No.slimming:255
## Not.feminine:171  sophisticated :215  slimming : 45
##
##
##
##           exciting           relaxing           effect.on.health
## exciting     :116  No.relaxing:113  effect on health : 66
## No.exciting:184  relaxing :187  No.effect on health:234
##
##
##
##

```

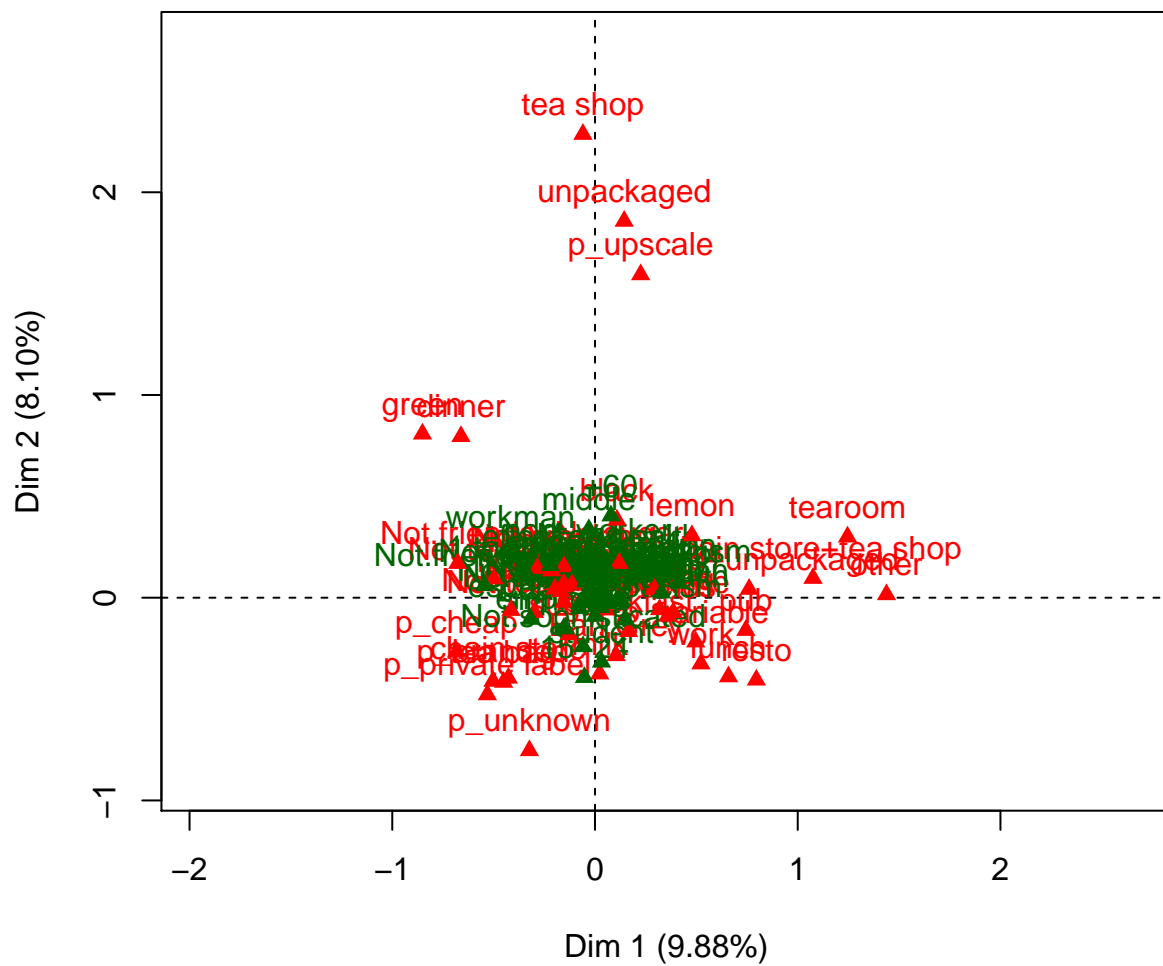
Loading FactoMineR

```
library(FactoMineR)
```

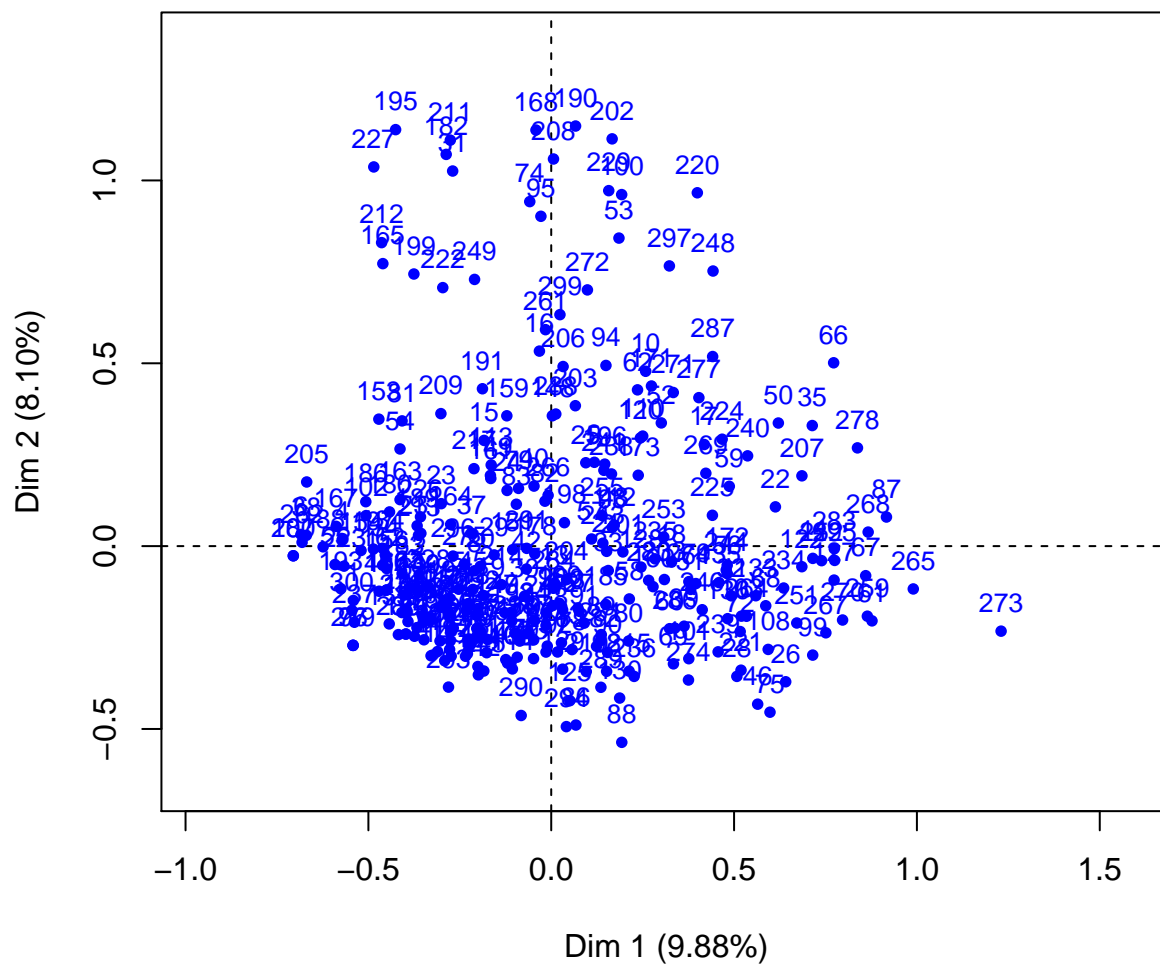
MCA

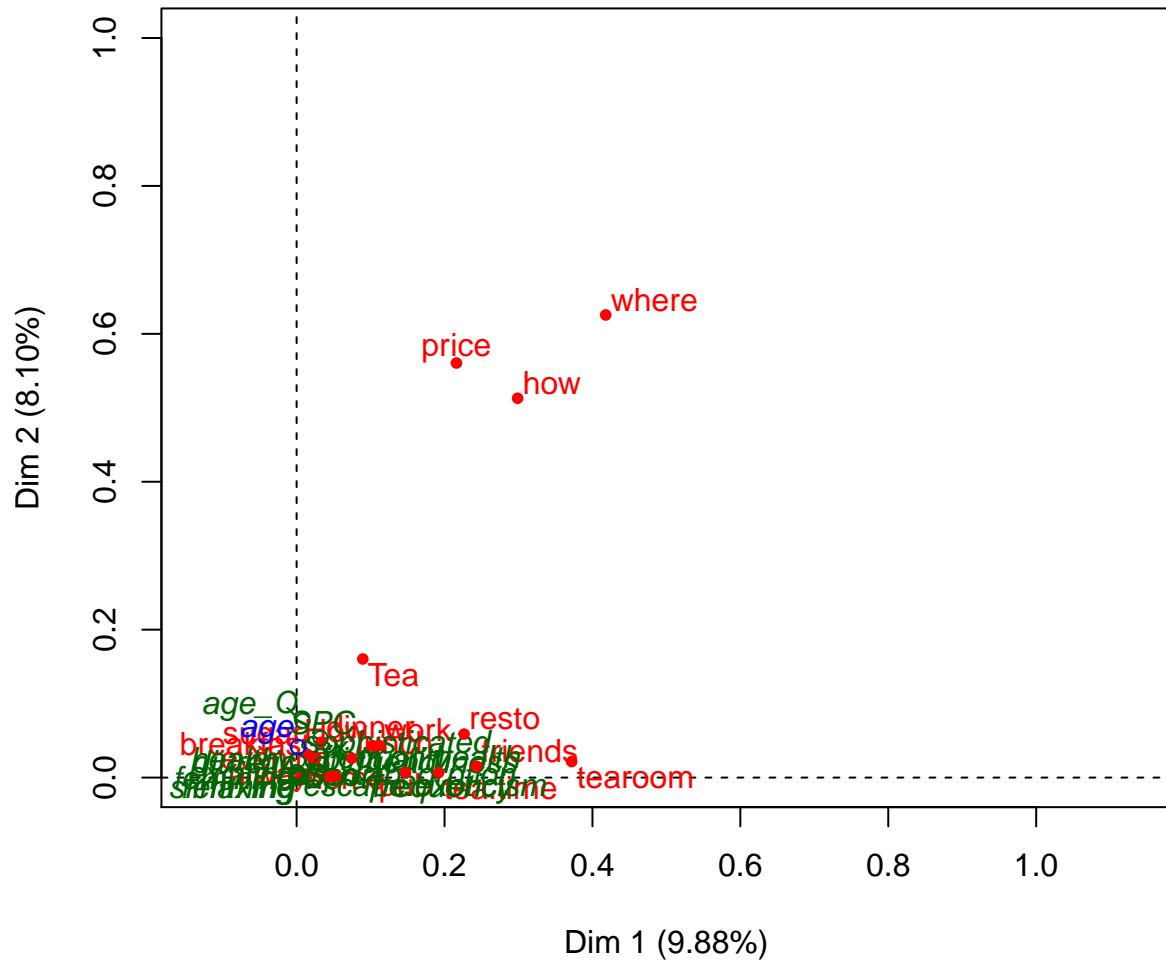
```
res <- MCA(tea, quanti.sup=19, quali.sup=c(20:36))
```

MCA factor map

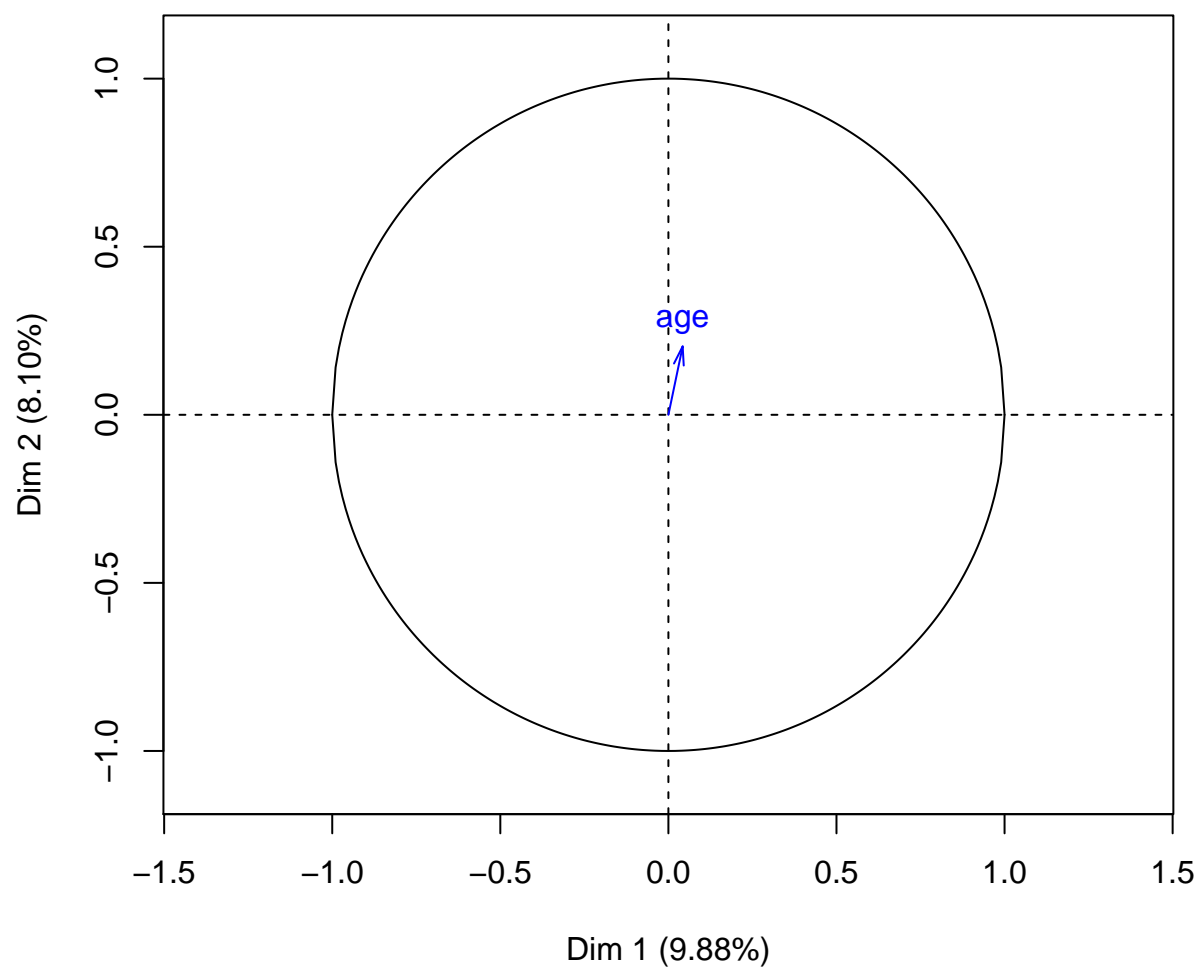


MCA factor map





Supplementary variables on the MCA factor map



Outputs can be summarized with the function `summary`.

```
summary(res, ncp=3, nbelements=Inf) ## fonction summary.MCA
```

Outputs are given for the first 2 dimensions (by default 3 dimensions are given) and for the 5 first elements.

```
summary(res, ncp=2, nbelements=5)
```

```
##  
## Call:  
## MCA(X = tea, quanti.sup = 19, quali.sup = c(20:36))  
##  
##  
## 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 5 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 |
##
## Categories (the 5 first)
##                   Dim.1  ctr  cos2 v.test  Dim.2  ctr  cos2 v.test
## breakfast |  0.166  0.495  0.025  2.756 | -0.166  0.607  0.026 -2.764
## Not.breakfast | -0.153  0.457  0.025 -2.756 |  0.154  0.560  0.026  2.764
## Not.tea time | -0.498  4.053  0.192 -7.578 |  0.093  0.174  0.007  1.423
## tea time |  0.386  3.142  0.192  7.578 | -0.072  0.135  0.007 -1.423
## evening |  0.319  1.307  0.053  3.985 | -0.058  0.053  0.002 -0.728
##
## breakfast |
## Not.breakfast |
## Not.tea time |
## tea time |
## evening |
##
## Categorical variables (eta2)
##                   Dim.1  Dim.2
## breakfast | 0.025 0.026 |
## tea.time | 0.192 0.007 |
## evening | 0.053 0.002 |
## lunch | 0.075 0.026 |
## dinner | 0.033 0.048 |
##
## Supplementary categories (the 5 first)
##                   Dim.1  cos2 v.test  Dim.2  cos2 v.test
## F |  0.151  0.033  3.158 | -0.109  0.017 -2.278 |
## M | -0.221  0.033 -3.158 |  0.159  0.017  2.278 |
## employee | -0.153  0.006 -1.313 | -0.151  0.006 -1.289 |

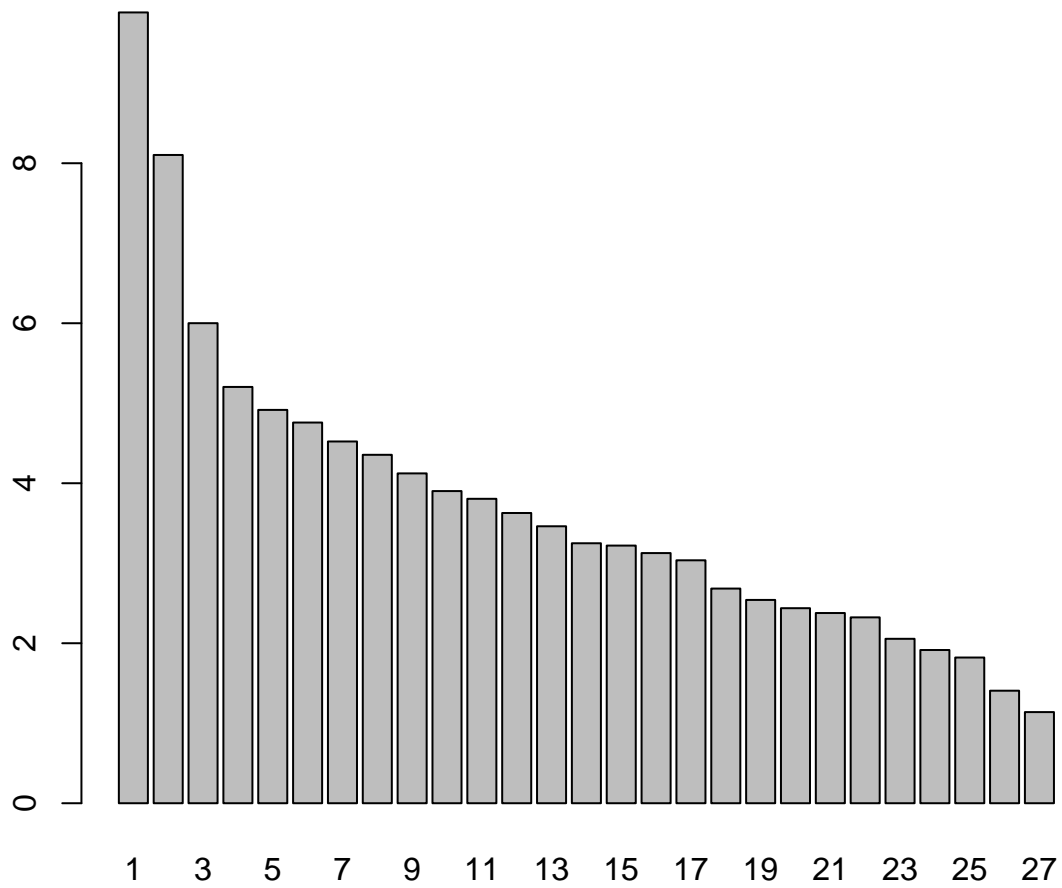
```

```
## middle      | -0.030  0.000 -0.205 |  0.336  0.017  2.281 |
## non-worker  | -0.036  0.000 -0.324 |  0.185  0.009  1.666 |
##
## Supplementary categorical variables (eta2)
##           Dim.1 Dim.2
## sex       | 0.033  0.017 |
## SPC       | 0.032  0.053 |
## Sport     | 0.001  0.000 |
## age_Q     | 0.008  0.077 |
## frequency | 0.094  0.006 |
##
## Supplementary continuous variable
##           Dim.1  Dim.2
## age       | 0.042 | 0.204 |
```

Graph of the eigenvalues

```
barplot(res$eig[,2],main="Eigenvalues", names.arg=1:nrow(res$eig))
```

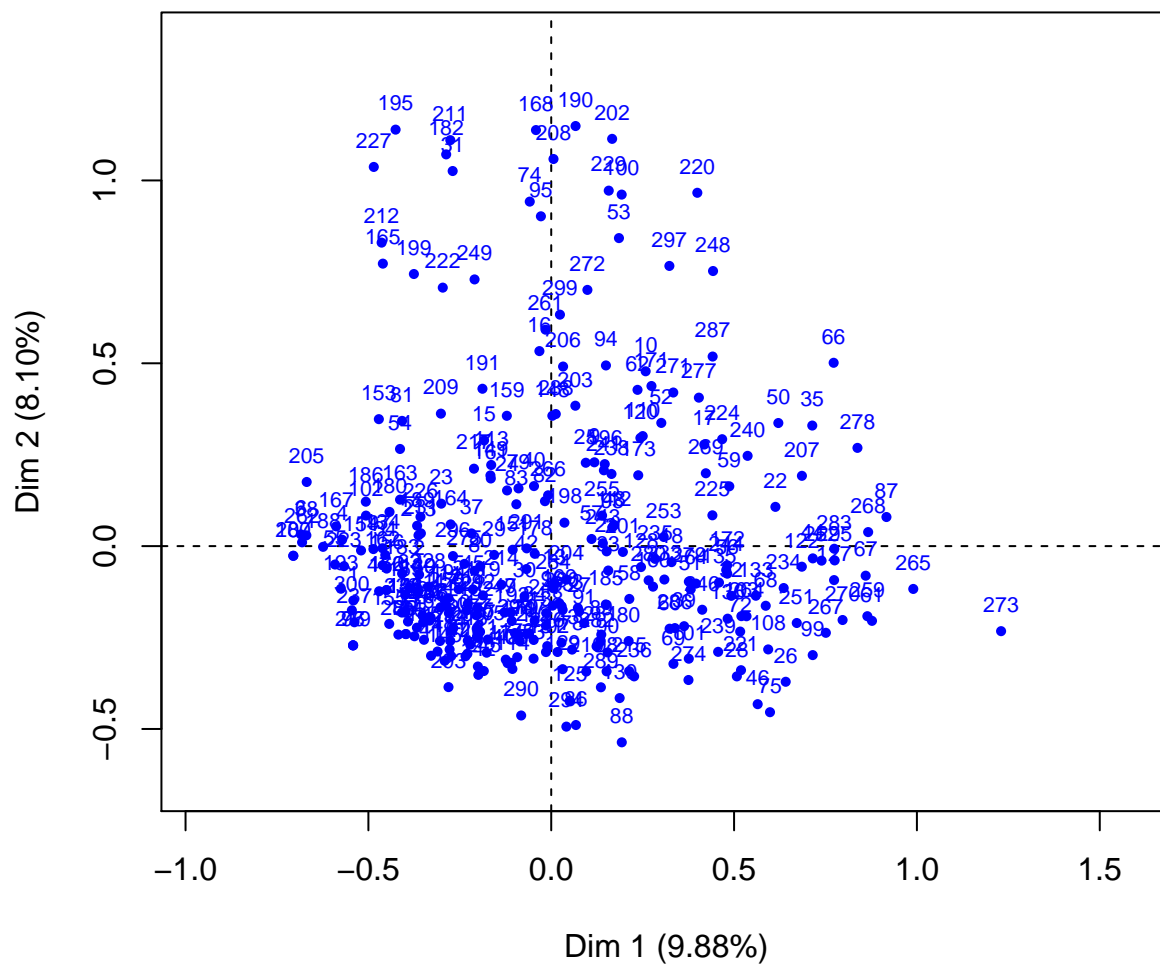

Eigenvalues



Graph with some labels

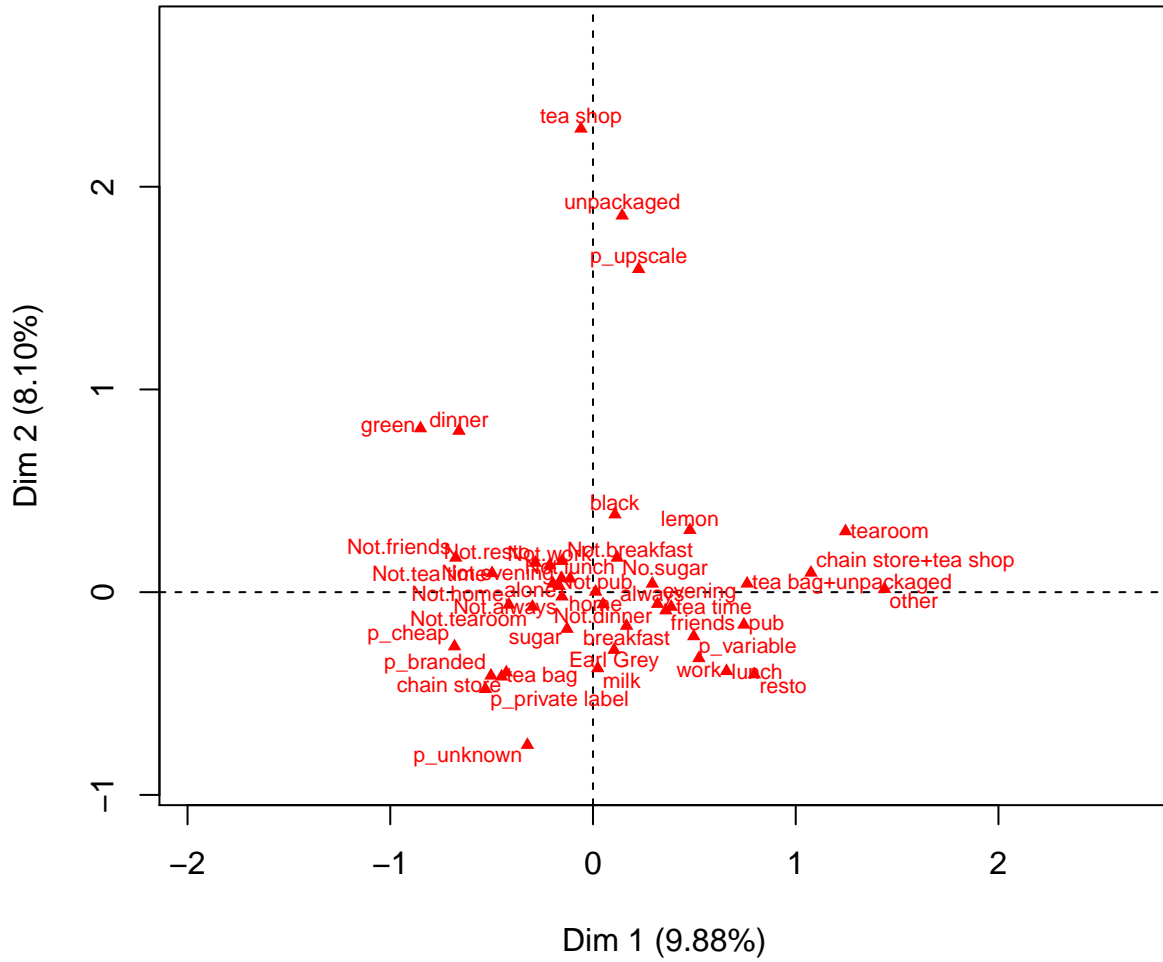
```
plot(res, label=c("var", "quali.sup"), cex=0.7)
```


MCA factor map



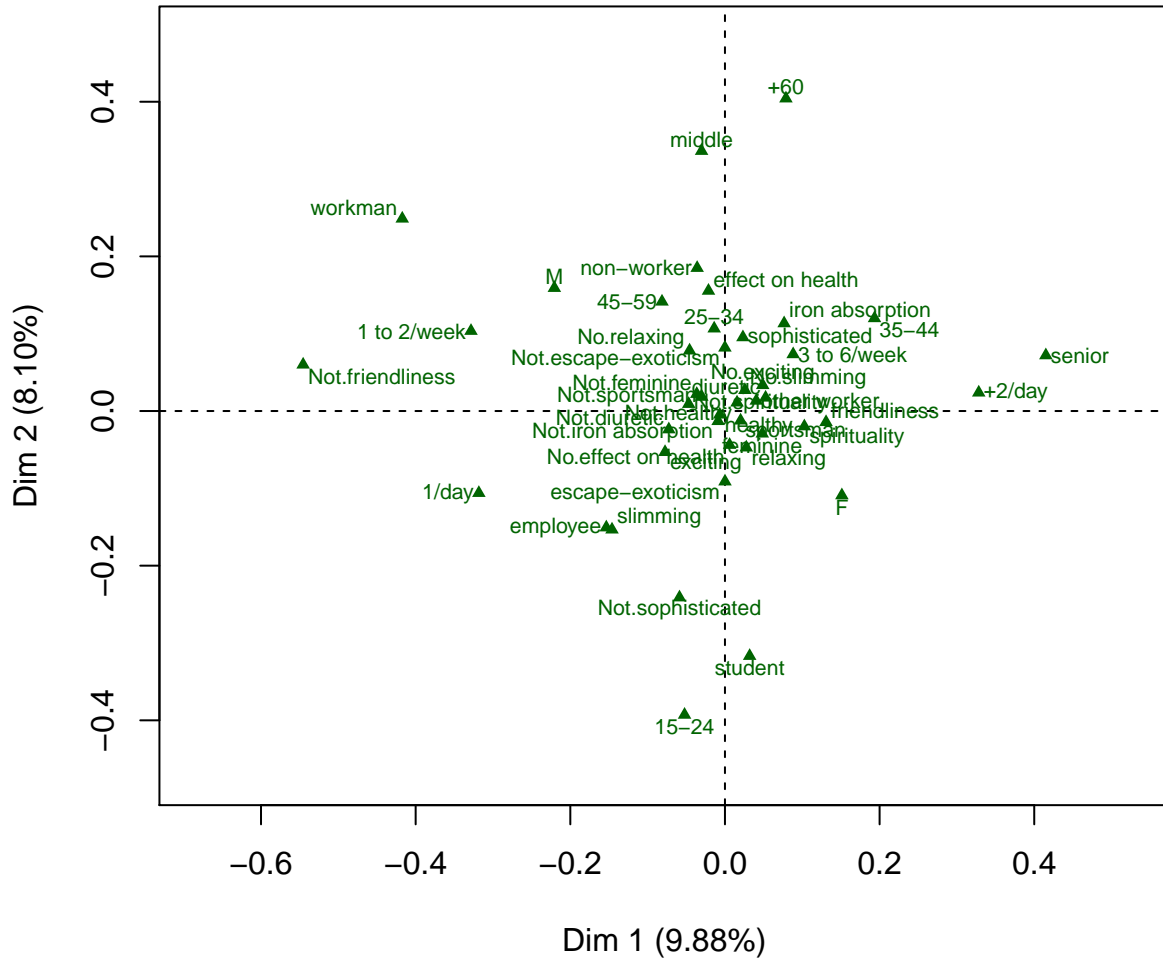
```
plot(res, invisible=c("ind","quali.sup"),autoLab="y",cex=0.7,  
      title="Active categories")
```

Active categories



```
plot(res, invisible=c("ind", "var"), autoLab="y", cex=0.7,  
      title="Supplementary categories")
```

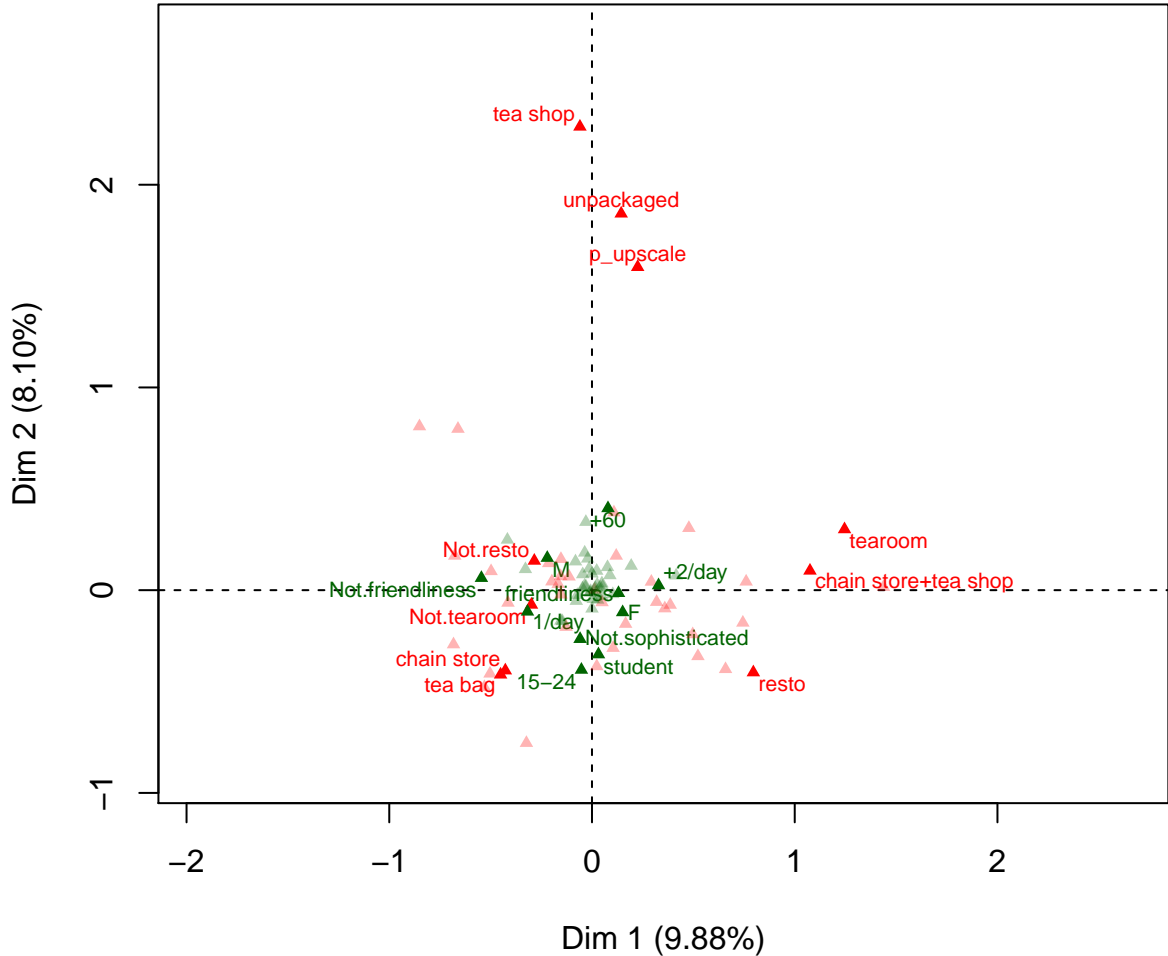
Supplementary categories



Selection of some categories

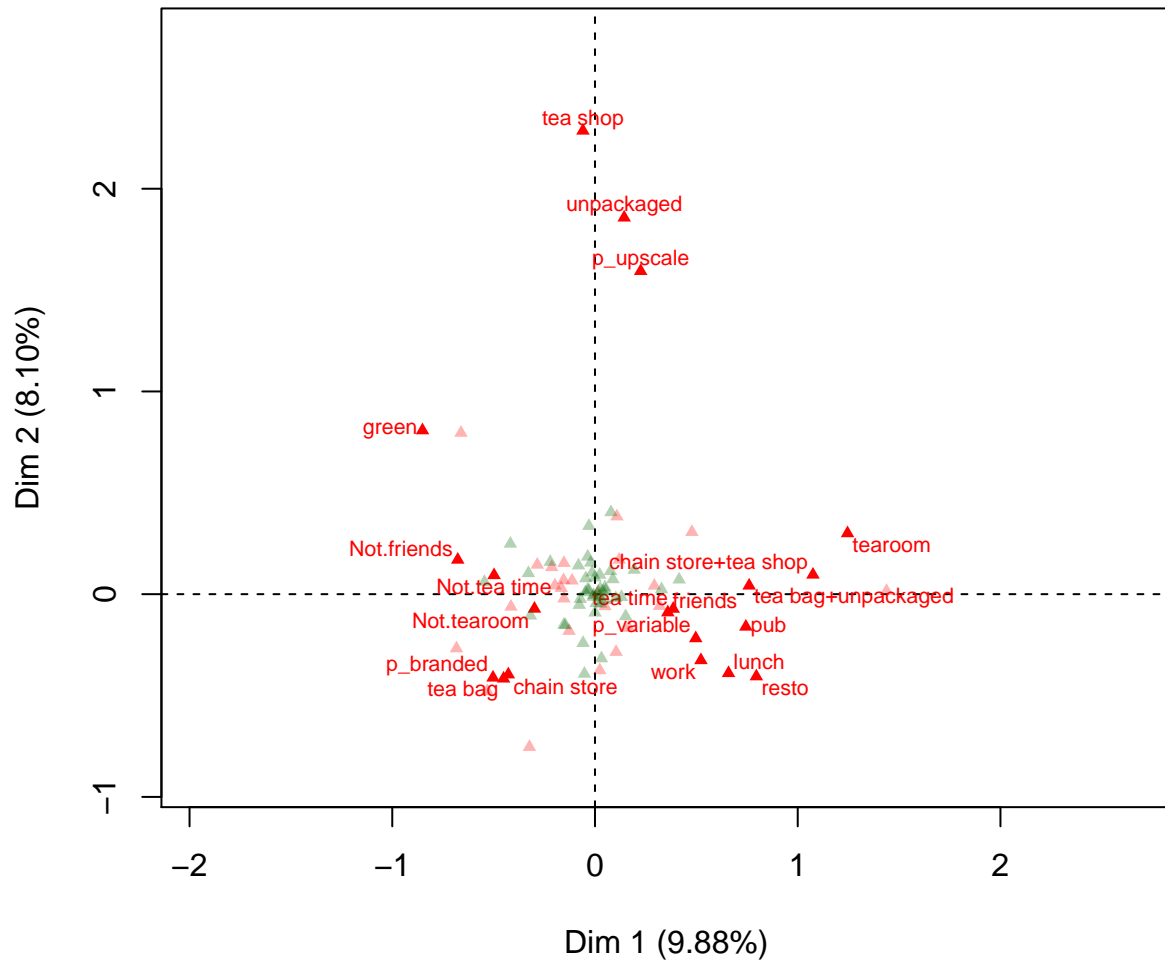
```
plot(res, invisible="ind", autoLab="y", cex=0.7, selectMod="cos2 10")
```

MCA factor map



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

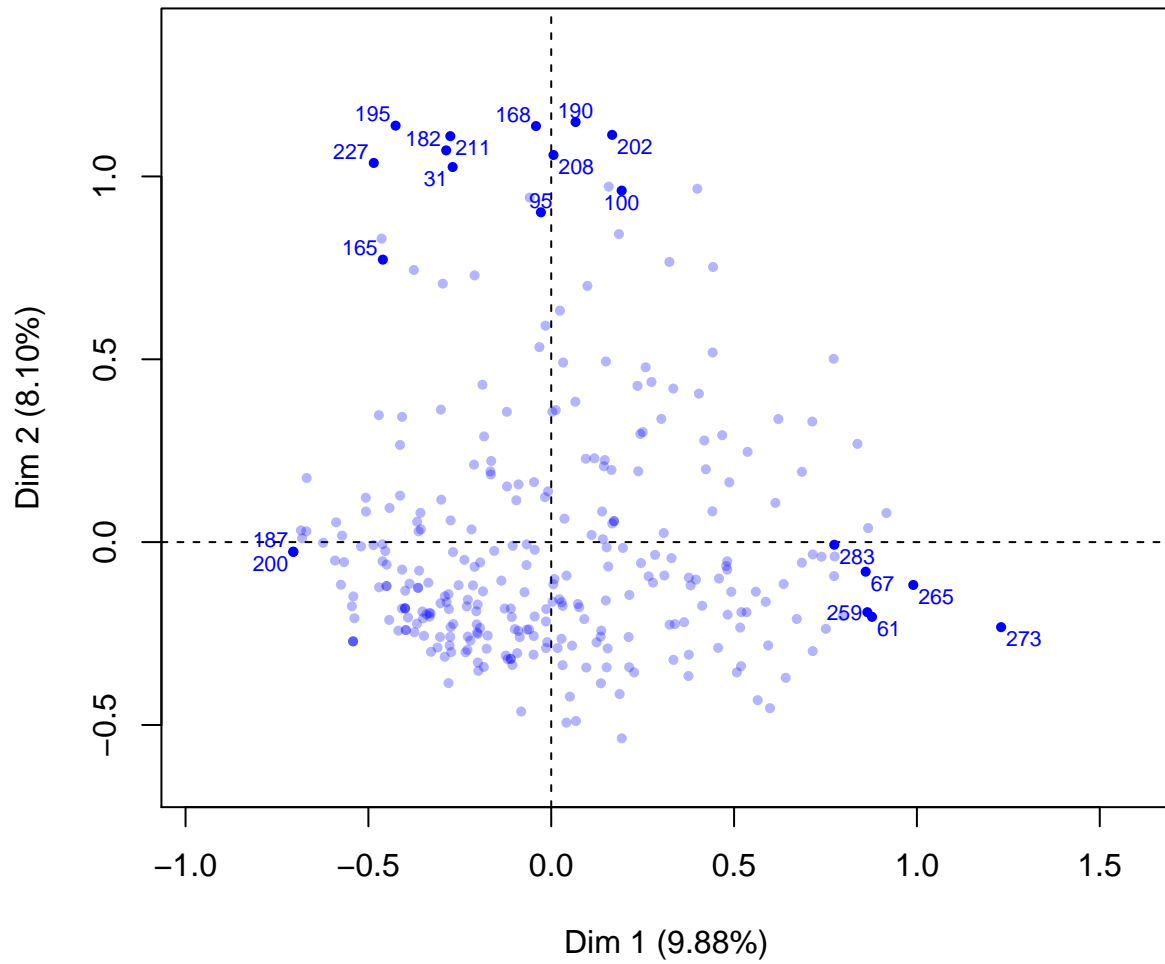
MCA factor map



Selection of some individuals

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

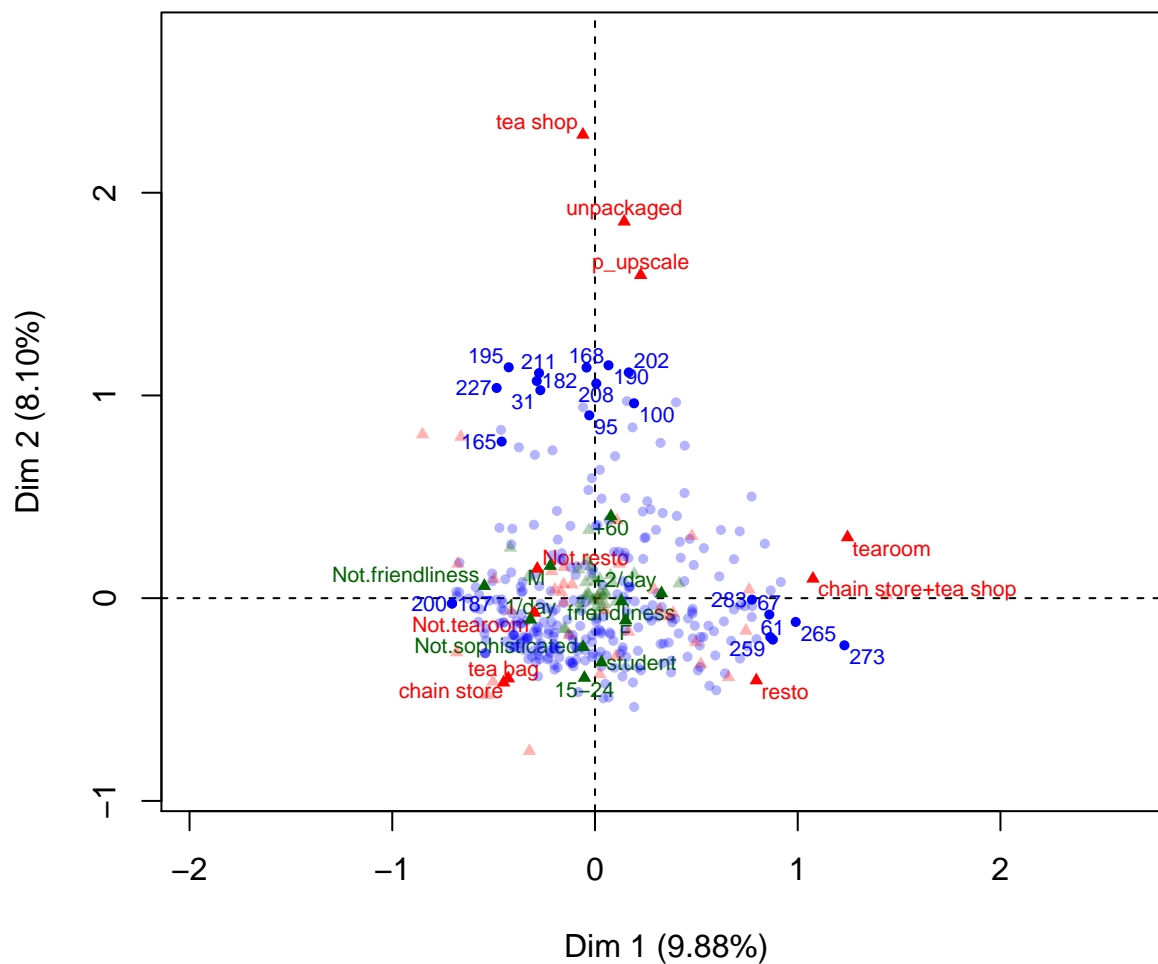
MCA factor map



Selection of some categories and some individuals

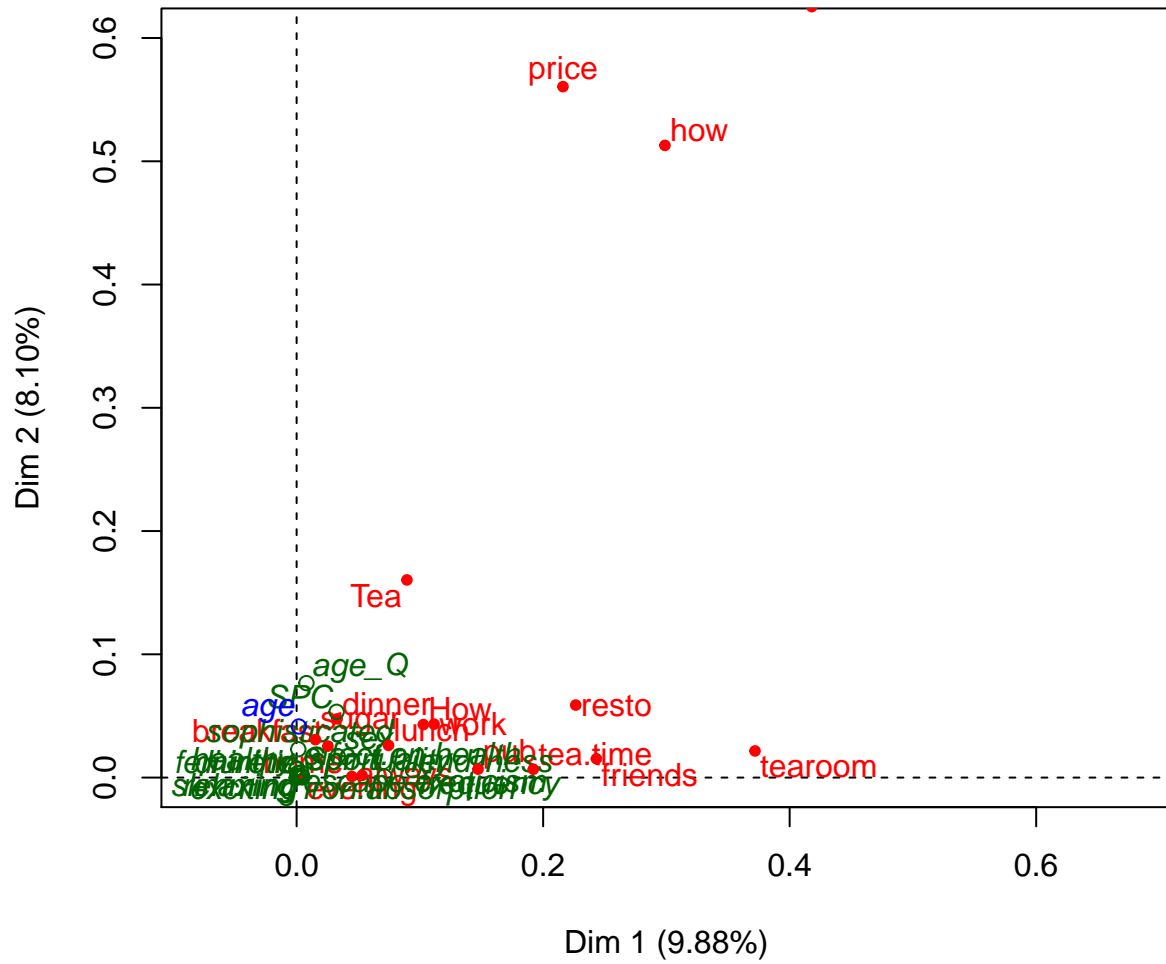
```
plot(res, autoLab="y", cex=0.7, select="cos2 20", selectMod="cos2 10")
```


MCA factor map

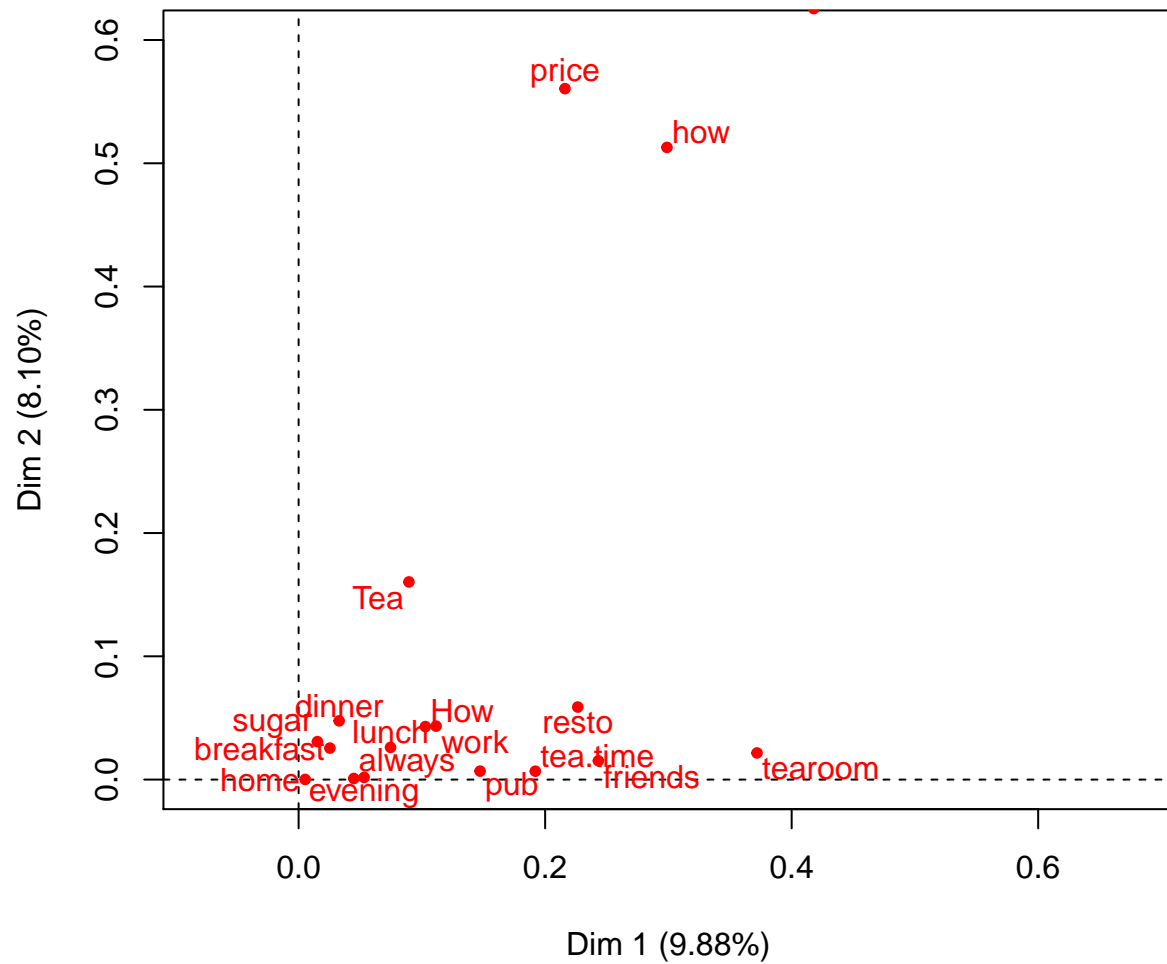


Graphs of the variables

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



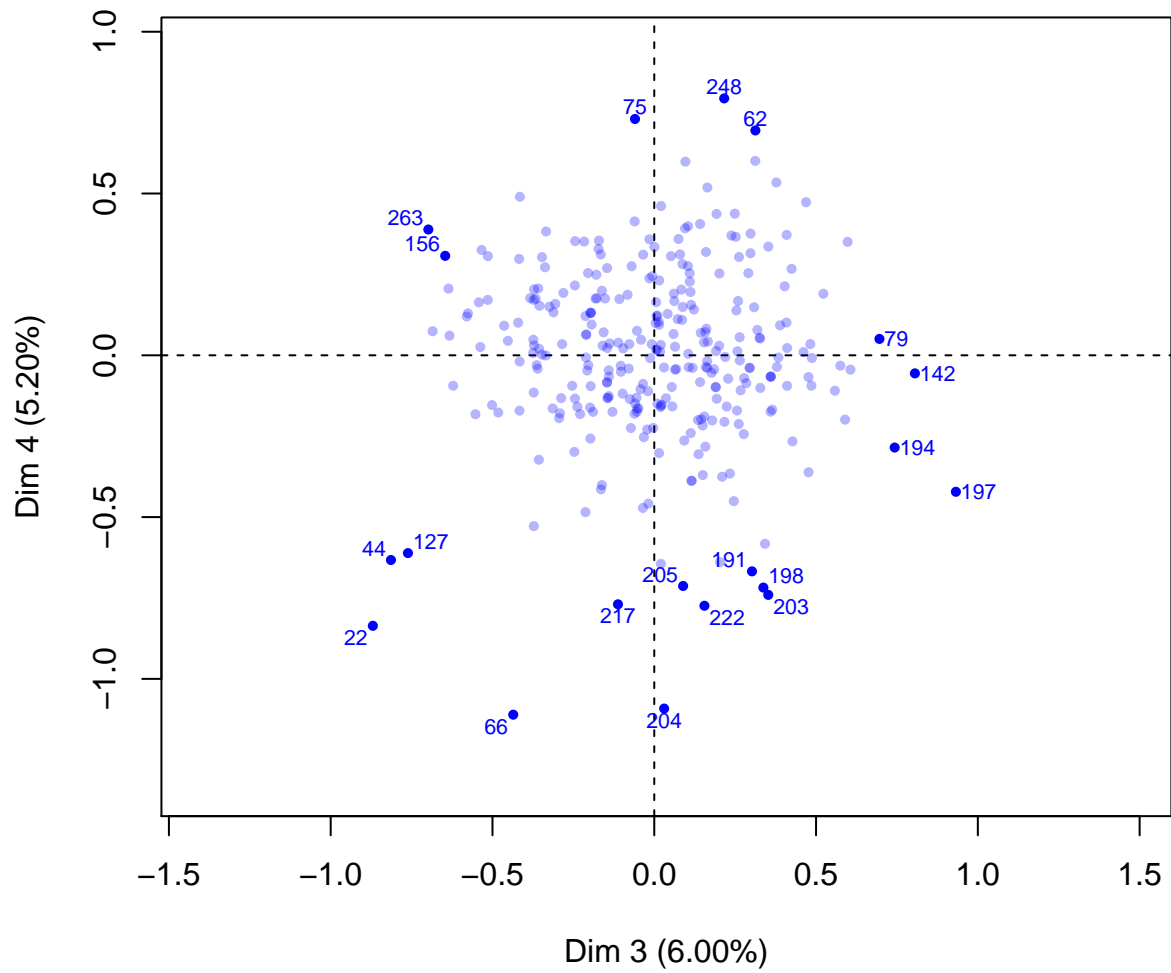
```
plot(res, choix="var",xlim=c(0,0.6),ylim=c(0,0.6),invisible=c("quali.sup","quanti.sup"))
```



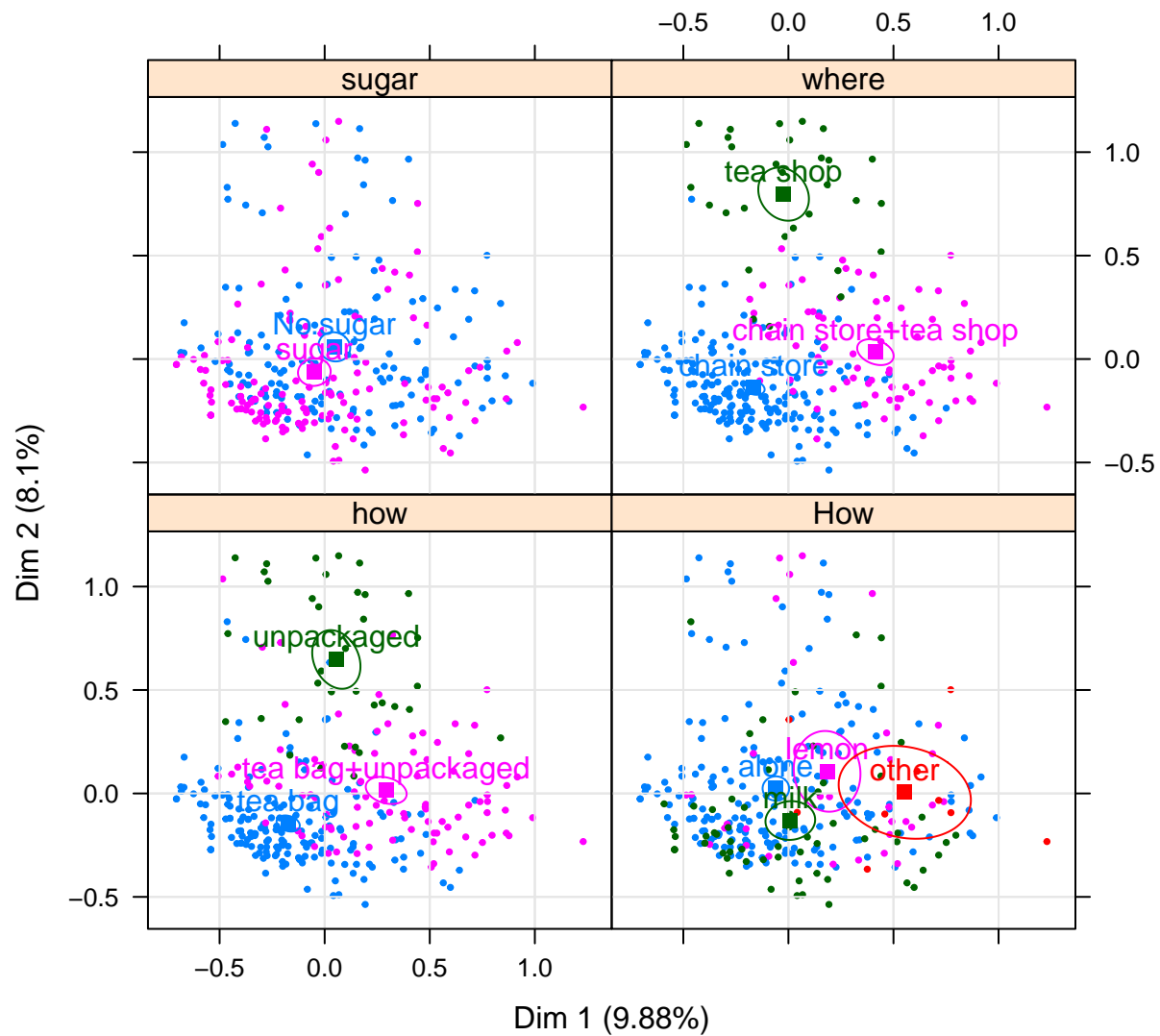
Graphs on dimensions 3 and 4

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

MCA factor map



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

Description of the dimensions

```
dimdesc(res)
```

```
## $`Dim 1`
## $`Dim 1`$quali
##           R2      p.value
## where    0.41793014 1.255462e-35
## tearoom  0.37189109 6.082138e-32
## how      0.29882863 1.273180e-23
## friends  0.24319952 8.616289e-20
## resto    0.22646759 2.319804e-18
## tea.time 0.19203800 1.652462e-15
## price    0.21609382 4.050469e-14
```

```

## pub          0.14722360 5.846592e-12
## work         0.11153590 3.000872e-09
## How          0.10285191 4.796010e-07
## Tea         0.08950330 8.970954e-07
## lunch       0.07458227 1.570629e-06
## frequency   0.09438792 1.849071e-06
## friendliness 0.07132511 2.706357e-06
## evening     0.05311759 5.586801e-05
## always      0.04479873 2.219503e-04
## sex         0.03335969 1.487620e-03
## dinner      0.03289362 1.608077e-03
## breakfast   0.02539639 5.667604e-03
## sugar       0.01527654 3.234986e-02
##
## $`Dim 1`$category
##              Estimate      p.value
## chain store+tea shop 0.33853776 1.344557e-35
## tearoom             0.29731072 6.082138e-32
## tea bag+unpackaged  0.23457030 1.361423e-21
## friends             0.19950832 8.616289e-20
## resto               0.20802605 2.319804e-18
## tea time            0.17011357 1.652462e-15
## pub                 0.18137133 5.846592e-12
## p_variable          0.27595067 5.956230e-12
## work                0.14170406 3.000872e-09
## +2/day              0.14855615 7.380937e-07
## lunch               0.14862636 1.570629e-06
## friendliness        0.13020388 2.706357e-06
## other                0.38192443 9.244911e-06
## evening             0.09345270 5.586801e-05
## always              0.08582336 2.219503e-04
## F                   0.07158886 1.487620e-03
## Not.dinner          0.13685745 1.608077e-03
## lemon               0.01223478 3.515252e-03
## breakfast           0.06141392 5.667604e-03
## senior              0.16802844 8.886876e-03
## Earl Grey           0.12203299 1.547110e-02
## No.sugar            0.04761975 3.234986e-02
## sugar               -0.04761975 3.234986e-02
## 1 to 2/week         -0.10430402 1.829818e-02
## p_private label     -0.11979263 1.156245e-02
## Not.breakfast       -0.06141392 5.667604e-03
## dinner              -0.13685745 1.608077e-03
## M                   -0.07158886 1.487620e-03
## alone               -0.23140428 2.326233e-04
## Not.always          -0.08582336 2.219503e-04
## 1/day               -0.10038745 1.556219e-04
## Not.evening         -0.09345270 5.586801e-05
## Not.friendliness    -0.13020388 2.706357e-06
## Not.lunch           -0.14862636 1.570629e-06
## green               -0.24569103 1.281162e-07
## Not.work            -0.14170406 3.000872e-09
## p_branded           -0.10910793 1.116908e-09
## Not.pub             -0.18137133 5.846592e-12

```

```

## Not.tea time          -0.17011357 1.652462e-15
## Not.resto             -0.20802605 2.319804e-18
## Not.friends           -0.19950832 8.616289e-20
## tea bag               -0.23182447 8.877561e-22
## chain store           -0.24012436 3.008256e-27
## Not.tearoom           -0.29731072 6.082138e-32
##
##
## $`Dim 2`
## $`Dim 2`$quanti
##      correlation      p.value
## age  0.2035108 0.0003890693
##
## $`Dim 2`$quali
##              R2      p.value
## where      0.62550194 4.542155e-64
## price      0.56056797 1.837909e-50
## how        0.51288621 4.103156e-47
## Tea        0.16034278 5.359827e-12
## resto      0.05883014 2.165287e-05
## age_Q      0.07663110 9.613084e-05
## dinner     0.04764166 1.385133e-04
## work       0.04334283 2.825934e-04
## sugar      0.03078909 2.286813e-03
## How        0.04300447 4.565763e-03
## lunch      0.02609615 5.035226e-03
## breakfast  0.02554407 5.527765e-03
## sophisticated 0.02298649 8.531637e-03
## tearoom    0.02159669 1.081515e-02
## SPC        0.05335498 1.284774e-02
## sex        0.01734823 2.250375e-02
## friends    0.01527530 3.235693e-02
##
## $`Dim 2`$category
##              Estimate      p.value
## tea shop      0.56623933 3.435386e-58
## p_upscale     0.58675674 6.819842e-53
## unpackaged    0.47523577 4.876111e-43
## green         0.17636551 5.660702e-07
## Not.resto     0.09599600 2.165287e-05
## black         0.02833892 1.280976e-04
## dinner        0.14912299 1.385133e-04
## Not.work      0.07997829 2.825934e-04
## No.sugar      0.06120846 2.286813e-03
## Not.lunch     0.07959849 5.035226e-03
## Not.breakfast 0.05576534 5.527765e-03
## +60           0.11445238 7.567898e-03
## sophisticated 0.05865034 8.531637e-03
## tearoom       0.06486865 1.081515e-02
## middle        0.09793868 2.228820e-02
## M             0.04674127 2.250375e-02
## Not.friends   0.04527027 3.235693e-02
## friends       -0.04527027 3.235693e-02
## p_private label -0.13563786 2.301393e-02

```



```

## F -0.04674127 2.250375e-02
## Not.tearoom -0.06486865 1.081515e-02
## Not.sophisticated -0.05865034 8.531637e-03
## p_unknown -0.23171262 7.581675e-03
## breakfast -0.05576534 5.527765e-03
## lunch -0.07959849 5.035226e-03
## p_variable -0.04476924 3.517223e-03
## student -0.12977090 2.381679e-03
## sugar -0.06120846 2.286813e-03
## milk -0.13222339 7.496148e-04
## work -0.07997829 2.825934e-04
## Not.dinner -0.14912299 1.385133e-04
## resto -0.09599600 2.165287e-05
## 15-24 -0.16342547 4.501578e-06
## p_branded -0.11247568 8.001515e-07
## Earl Grey -0.20470443 6.499738e-12
## tea bag -0.31755648 2.123707e-18
## chain store -0.36891861 5.804094e-23
##
##
## $`Dim 3`
## $`Dim 3`$quanti
## correlation p.value
## age -0.3397736 1.530157e-09
##
## $`Dim 3`$quali
## R2 p.value
## Tea 0.33178708 9.996725e-27
## breakfast 0.21498315 2.138445e-17
## sugar 0.21026537 5.278399e-17
## How 0.19600156 5.867378e-14
## home 0.13440219 5.642852e-11
## age_Q 0.14565571 1.852397e-09
## evening 0.10637896 7.269113e-09
## friends 0.10294629 1.307065e-08
## always 0.10133291 1.721112e-08
## tea.time 0.05424196 4.636352e-05
## frequency 0.06361133 2.167789e-04
## pub 0.04311129 2.936636e-04
## SPC 0.07625129 6.739742e-04
## where 0.04166084 1.801385e-03
## exciting 0.02425859 6.872690e-03
## dinner 0.02152288 1.095256e-02
## spirituality 0.01606168 2.817777e-02
## lunch 0.01561039 3.050262e-02
## diuretic 0.01302499 4.827448e-02
##
## $`Dim 3`$category
## Estimate p.value
## Earl Grey 0.18122366 5.474630e-22
## Not.breakfast 0.13921762 2.138445e-17
## sugar 0.13764789 5.278399e-17
## Not.home 0.32238212 5.642852e-11
## evening 0.10304138 7.269113e-09

```

## friends	0.10113377	1.307065e-08
## always	0.10056783	1.721112e-08
## lemon	0.29113799	3.532927e-05
## Not.tea time	0.07044079	4.636352e-05
## 1 to 2/week	0.13367067	4.787775e-05
## 25-34	0.15458550	9.102105e-05
## student	0.11762964	1.467086e-04
## pub	0.07646920	2.936636e-04
## 15-24	0.12278166	4.286711e-04
## alone	0.13200504	7.535155e-04
## tea shop	0.12875140	1.905017e-03
## exciting	0.04797670	6.872690e-03
## dinner	0.08625296	1.095256e-02
## spirituality	0.04098581	2.817777e-02
## lunch	0.05297807	3.050262e-02
## unpackaged	0.07670099	3.765383e-02
## Not.diuretic	0.03468681	4.827448e-02
## diuretic	-0.03468681	4.827448e-02
## +2/day	-0.07763057	3.323605e-02
## Not.lunch	-0.05297807	3.050262e-02
## Not.spirituality	-0.04098581	2.817777e-02
## chain store+tea shop	-0.09887763	2.197901e-02
## 35-44	-0.07139235	1.961314e-02
## Not.dinner	-0.08625296	1.095256e-02
## non-worker	-0.08793735	8.568111e-03
## No.exciting	-0.04797670	6.872690e-03
## 45-59	-0.07791412	1.331486e-03
## +60	-0.12806070	4.108178e-04
## Not.pub	-0.07646920	2.936636e-04
## tea time	-0.07044079	4.636352e-05
## other	-0.33576960	1.254567e-05
## milk	-0.08737342	7.989362e-08
## Not.always	-0.10056783	1.721112e-08
## Not.friends	-0.10113377	1.307065e-08
## Not.evening	-0.10304138	7.269113e-09
## home	-0.32238212	5.642852e-11
## No.sugar	-0.13764789	5.278399e-17
## breakfast	-0.13921762	2.138445e-17
## black	-0.22772627	4.313358e-26