Philologie computationnelle - TP Chrétien

JBC & FC

2023-2024

```r
source("functions.R")
```

```
## Le chargement a nécessité le package : ggplot2
```

Data

Load texts

```r
data = t(as.matrix(read.csv("chrestien_fw_abs.csv", row.names = 1, header = TRUE)[, c(-1,-2)]))
auts = read.csv("chrestien_fw_abs.csv", row.names = 1, header = TRUE)[, 1]
data = data[rowSums(data) > 0,]
colnames(data) = gsub(pattern = "000", replacement = "k", colnames(data))
data_abs_save = data
```

Features extraction, culling and selection

Now, we will use survey method to keep only reliable features,

```r
# select = selection(data, z = 1.645)
# select = select[,4]
# # Get relative frequencies
# data_abs_save = data
# data = relativeFreqs(data)
# # keep only reliable features
# data = data[select,]
```

Dimensionality reduction

Principal component analysis (PCA)

Let's use `FactoMineR`.

```r
# Load FactoMineR and factoextra for better visuals
library('FactoMineR')
library('factoextra')
```

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

```r
# And compute ACP. As the function expects individuals as lines and variables as columns, I'm transposi
myCA = CA(t(data)[1:79,], graph = FALSE)
```

```
## Warning in CA(t(data)[1:79, ], graph = FALSE): The columns quelque, quiconque,
## enjusque, cil, chasque, enteimes, endementre, eluec, dementre, entre_tant,
## neporuec, seviaus, decoste, hu, entrementre sum at 0. They were suppressed from
```

the analysis

I can now look at the significance of my axes.

```
myCA$eig
```

```
##           eigenvalue percentage of variance cumulative percentage of variance
## dim 1  0.0216368184               8.5698642                          8.569864
## dim 2  0.0166046192               6.5767216                         15.146586
## dim 3  0.0139160724               5.5118478                         20.658434
## dim 4  0.0102660639               4.0661604                         24.724594
## dim 5  0.0077787306               3.0809828                         27.805577
## dim 6  0.0072173837               2.8586457                         30.664223
## dim 7  0.0063110215               2.4996557                         33.163878
## dim 8  0.0062240192               2.4651961                         35.629074
## dim 9  0.0057194726               2.2653563                         37.894431
## dim 10 0.0053613494               2.1235117                         40.017942
## dim 11 0.0050022579               1.9812835                         41.999226
## dim 12 0.0048186474               1.9085595                         43.907785
## dim 13 0.0046388866               1.8373602                         45.745146
## dim 14 0.0045959744               1.8203636                         47.565509
## dim 15 0.0044201211               1.7507120                         49.316221
## dim 16 0.0043146564               1.7089398                         51.025161
## dim 17 0.0041727533               1.6527351                         52.677896
## dim 18 0.0041265838               1.6344484                         54.312344
## dim 19 0.0039568622               1.5672254                         55.879570
## dim 20 0.0038616606               1.5295182                         57.409088
## dim 21 0.0037799199               1.4971425                         58.906230
## dim 22 0.0035865143               1.4205388                         60.326769
## dim 23 0.0034955987               1.3845292                         61.711298
## dim 24 0.0034220312               1.3553907                         63.066689
## dim 25 0.0033203676               1.3151240                         64.381813
## dim 26 0.0031768907               1.2582960                         65.640109
## dim 27 0.0031200367               1.2357774                         66.875887
## dim 28 0.0030636935               1.2134611                         68.089348
## dim 29 0.0030045652               1.1900417                         69.279389
## dim 30 0.0028204293               1.1171095                         70.396499
## dim 31 0.0027476833               1.0882965                         71.484795
## dim 32 0.0027017449               1.0701013                         72.554897
## dim 33 0.0026058388               1.0321150                         73.587012
## dim 34 0.0025502981               1.0101166                         74.597128
## dim 35 0.0024935364               0.9876345                         75.584763
## dim 36 0.0024647323               0.9762258                         76.560989
## dim 37 0.0024161439               0.9569810                         77.517970
## dim 38 0.0023130170               0.9161348                         78.434104
## dim 39 0.0022966039               0.9096339                         79.343738
## dim 40 0.0022325085               0.8842471                         80.227985
## dim 41 0.0021622005               0.8563997                         81.084385
## dim 42 0.0020958342               0.8301135                         81.914499
## dim 43 0.0020694669               0.8196700                         82.734169
## dim 44 0.0020357537               0.8063169                         83.540485
## dim 45 0.0019885422               0.7876175                         84.328103
## dim 46 0.0019436013               0.7698174                         85.097920
## dim 47 0.0018858736               0.7469527                         85.844873
## dim 48 0.0018164916               0.7194721                         86.564345
```

```
## dim 49 0.0017722433                0.7019463                      87.266291
## dim 50 0.0016960891                0.6717833                      87.938075
## dim 51 0.0016720488                0.6622615                      88.600336
## dim 52 0.0016212456                0.6421395                      89.242476
## dim 53 0.0015641643                0.6195308                      89.862007
## dim 54 0.0015069426                0.5968665                      90.458873
## dim 55 0.0015011086                0.5945558                      91.053429
## dim 56 0.0014477376                0.5734168                      91.626846
## dim 57 0.0014329349                0.5675537                      92.194399
## dim 58 0.0014063535                0.5570255                      92.751425
## dim 59 0.0013136170                0.5202946                      93.271720
## dim 60 0.0012777317                0.5060812                      93.777801
## dim 61 0.0012066674                0.4779342                      94.255735
## dim 62 0.0011884395                0.4707145                      94.726449
## dim 63 0.0011608382                0.4597823                      95.186232
## dim 64 0.0010564026                0.4184176                      95.604649
## dim 65 0.0010466251                0.4145450                      96.019194
## dim 66 0.0009796248                0.3880077                      96.407202
## dim 67 0.0009677434                0.3833017                      96.790504
## dim 68 0.0009120235                0.3612323                      97.151736
## dim 69 0.0008676085                0.3436405                      97.495377
## dim 70 0.0008408166                0.3330288                      97.828405
## dim 71 0.0008080430                0.3200479                      98.148453
## dim 72 0.0007803217                0.3090681                      98.457521
## dim 73 0.0007632162                0.3022930                      98.759815
## dim 74 0.0007047781                0.2791470                      99.038962
## dim 75 0.0006710316                0.2657807                      99.304742
## dim 76 0.0006142720                0.2432995                      99.548042
## dim 77 0.0005860199                0.2321095                      99.780151
## dim 78 0.0005550645                0.2198488                     100.000000
```

or plot it

```
barplot(myCA$eig[,2], main="%age of var", names.arg=1:nrow(myCA$eig))
```

%age of var
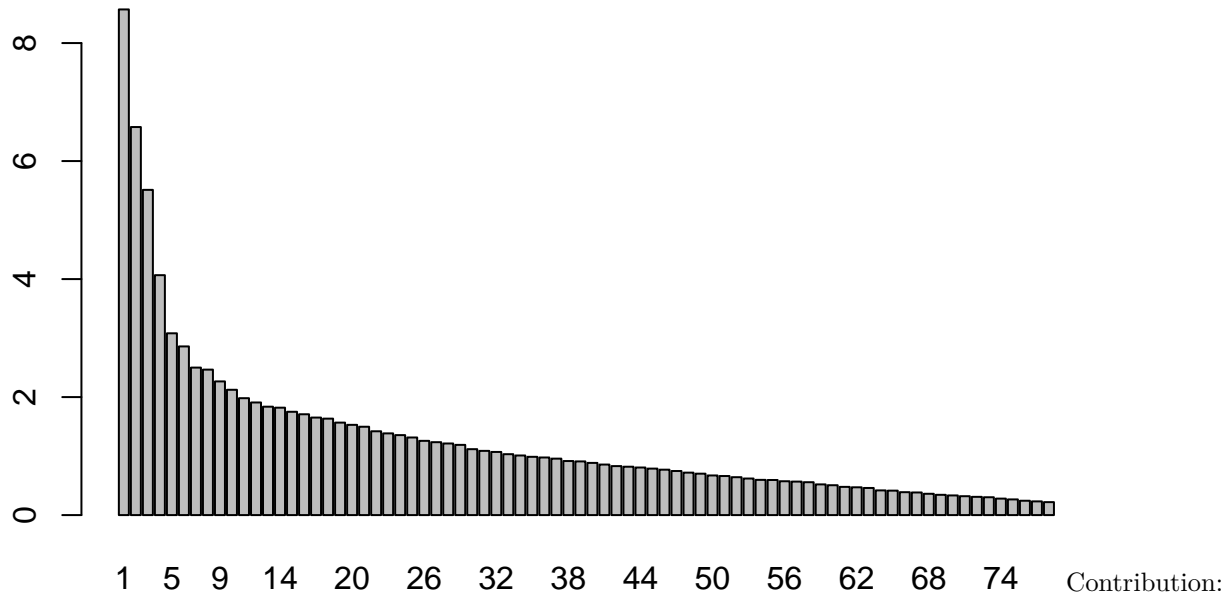

```
sort(myCA$col$contrib[,1], decreasing = TRUE)
```

```
##           je           le          vos          mon           ne
## 1.615921e+01 8.933089e+00 7.395664e+00 6.834900e+00 4.738901e+00
##           de          lor           ja           se       vostre
## 3.913434e+00 3.560978e+00 2.377325e+00 2.302046e+00 1.935862e+00
##           tu          qui           et          cel            a
## 1.816474e+00 1.623276e+00 1.560787e+00 1.479400e+00 1.301531e+00
##           mi          que       dusque          sor       chascun
## 1.254799e+00 1.252237e+00 1.235490e+00 1.138650e+00 1.119984e+00
##        maint         donc         rien         mais          par
## 9.851958e-01 9.689506e-01 9.567072e-01 9.302864e-01 8.026662e-01
##           il           un          ton        joste          son
## 7.519637e-01 7.157719e-01 7.092466e-01 6.761997e-01 5.961186e-01
##         mout           ci        nient         trop          las
## 5.931190e-01 5.575250e-01 5.108279e-01 5.080809e-01 5.038535e-01
##         tres          coi         quel       devers       encore
## 4.891818e-01 4.703406e-01 4.484381e-01 4.476847e-01 4.282005e-01
##        autel       defors        apres         mien          car
## 3.984280e-01 3.641907e-01 3.618329e-01 3.077652e-01 3.043391e-01
##       certes           or       espoir          mar      daerrain
## 2.904417e-01 2.882406e-01 2.778044e-01 2.752867e-01 2.741498e-01
##          ier         tant      ensement         cest          hai
## 2.587600e-01 2.480688e-01 2.406760e-01 2.355719e-01 2.150210e-01
##       derrier        parmi           on       devant       dedenz
## 2.135527e-01 2.132003e-01 2.123756e-01 2.040077e-01 2.007404e-01
##         main         fors          enz          nul        outre
## 1.941636e-01 1.931429e-01 1.869405e-01 1.844943e-01 1.809046e-01
##        encui       endroit          des      trestot          ore
## 1.799439e-01 1.736429e-01 1.692578e-01 1.633041e-01 1.527915e-01
##     premerain        entre         piec    autrement          pis
## 1.519508e-01 1.507673e-01 1.483817e-01 1.481913e-01 1.421892e-01
```

```
##         adonc            tel             ai       encontre         laienz
## 1.416629e-01   1.409767e-01   1.346311e-01   1.333408e-01   1.201727e-01
##            ha      quelconque        igaument          jehui          desor
## 1.114095e-01   1.109927e-01   1.109927e-01   1.098724e-01   1.098704e-01
##         entor            jus            icel         tempre           ades
## 1.084205e-01   1.084094e-01   1.077506e-01   1.031056e-01   1.020627e-01
##        aucant            non           ensus            sis           lues
## 1.017408e-01   1.008067e-01   1.004634e-01   9.892273e-02   9.623419e-02
##           soi         petitet         ersoir        aillors           itel
## 9.588019e-02   9.443397e-02   9.423743e-02   9.396588e-02   9.359493e-02
##         iluec            iqui            hui           puis            cui
## 9.343788e-02   9.237605e-02   9.000822e-02   8.832314e-02   8.716598e-02
##          ainc          desque         autrui          enson         partot
## 8.136292e-02   8.084136e-02   7.882788e-02   7.877872e-02   7.872248e-02
##           tot          sempre         contre       aussiment         meisme
## 7.766460e-02   7.738736e-02   7.295257e-02   7.227160e-02   7.184550e-02
##     volentiers          avant          selonc          anuit           lors
## 7.170335e-02   7.146479e-02   6.945654e-02   6.945508e-02   6.912737e-02
##        huimais         cantque          ariere            mie           cant
## 6.624932e-02   6.563338e-02   6.449133e-02   6.443298e-02   6.319187e-02
##         nenil           onque             ce           aussi       endemain
## 6.257585e-02   5.915212e-02   5.858678e-02   5.716722e-02   5.447173e-02
##      lendemain         orainz         comfait            pas           ainz
## 5.419922e-02   5.384295e-02   5.369788e-02   5.237669e-02   5.134639e-02
##       autretel       nequedent           avoi          nesun            poi
## 5.070434e-02   5.055355e-02   5.055262e-02   4.847610e-02   4.736491e-02
##            ca             en           tuen           gote            ice
## 4.719826e-02   4.677691e-02   4.408766e-02   4.365315e-02   4.357791e-02
##         autre          desus        entresait          dalez           sien
## 4.266348e-02   4.258702e-02   4.236731e-02   4.180675e-02   3.994931e-02
##         adens            oil             ui           issi            oan
## 3.991601e-02   3.907249e-02   3.848370e-02   3.721413e-02   3.680499e-02
##         gaire         ambedos         que_que           enmi           plus
## 3.331434e-02   3.310283e-02   3.282577e-02   3.266008e-02   3.097170e-02
##           por           dont          antan          amont       rechief_
## 2.989003e-02   2.977892e-02   2.946492e-02   2.824780e-02   2.631388e-02
##        jusque          caienz           neis             ez       enquenuit
## 2.580722e-02   2.542121e-02   2.525714e-02   2.517116e-02   2.492902e-02
##            la            lez           suen           hors    autresiment
## 2.262099e-02   2.197590e-02   2.069397e-02   2.034352e-02   1.917891e-02
##            el            res           dela            soz        environ
## 1.847392e-02   1.821562e-02   1.800351e-02   1.706124e-02   1.660171e-02
##         aucun           sore        combien              e           voir
## 1.640159e-02   1.510177e-02   1.480478e-02   1.422524e-02   1.404322e-02
##     maintenant           come            ome  endementieres           aie
## 1.349991e-02   1.339200e-02   1.319197e-02   1.309524e-02   1.287517e-02
##        sovent          ainsi           deci           lonc           sauf
## 1.268628e-02   1.245580e-02   1.206210e-02   1.125358e-02   1.064239e-02
##        coment              i           vers          moins          icest
## 1.025577e-02   1.010684e-02   1.000712e-02   9.924996e-03   9.381093e-03
##      autretant       orendroit           aval      contremont     meismement
## 9.062476e-03   8.862121e-03   8.248175e-03   8.219273e-03   8.178342e-03
##         matin             fi            nos       demanois             ou
## 7.680408e-03   7.003644e-03   6.946239e-03   6.453542e-03   6.110733e-03
```

```
##        nostre             es          mieus            bas              o
## 5.287943e-03   5.162209e-03   5.130595e-03   5.029301e-03   4.900117e-03
##        demain         autresi         envers           prim            sus
## 4.747042e-03   4.679129e-03   4.664639e-03   4.474251e-03   4.292912e-03
##      neporcant         encoste          autant        aincois           pres
## 4.069478e-03   3.857984e-03   3.790271e-03   3.710471e-03   3.702487e-03
##         point          avuec           lasus           atot           tost
## 3.528190e-03   3.492847e-03   3.205524e-03   2.926073e-03   2.571534e-03
##     entretant          itant          empres            ici        tantost
## 2.349249e-03   2.132577e-03   2.114943e-03   1.989343e-03   1.713705e-03
##          buer         manois           chies           foiz           sans
## 1.712492e-03   1.625663e-03   1.548670e-03   1.152168e-03   1.081585e-03
##      ensemble        dejoste           atant          ambes          jadis
## 1.072412e-03   9.218338e-04   8.715644e-04   4.850633e-04   4.257764e-04
##        travers          estre           aluec     dementieres           deca
## 3.493798e-04   2.467580e-04   2.359028e-04   2.218690e-04   2.091529e-04
##          naje         auques          tresque             si     franchement
## 8.524891e-05   7.996916e-05   7.412979e-05   7.269963e-05   6.479806e-05
##         desoz          assez           voire       contreval         plusor
## 6.070934e-05   5.406519e-05   5.251069e-05   8.326010e-06   2.958636e-06
##    sodainement
## 1.784145e-06
```

```r
#fviz_ca_row(myCA, geom=c("text"), labelsize=2)
fviz_ca_row(myCA, col.row = sub("\\_.*", "", rownames(myCA$row$coord)), title="", geom=c("point"), label
```
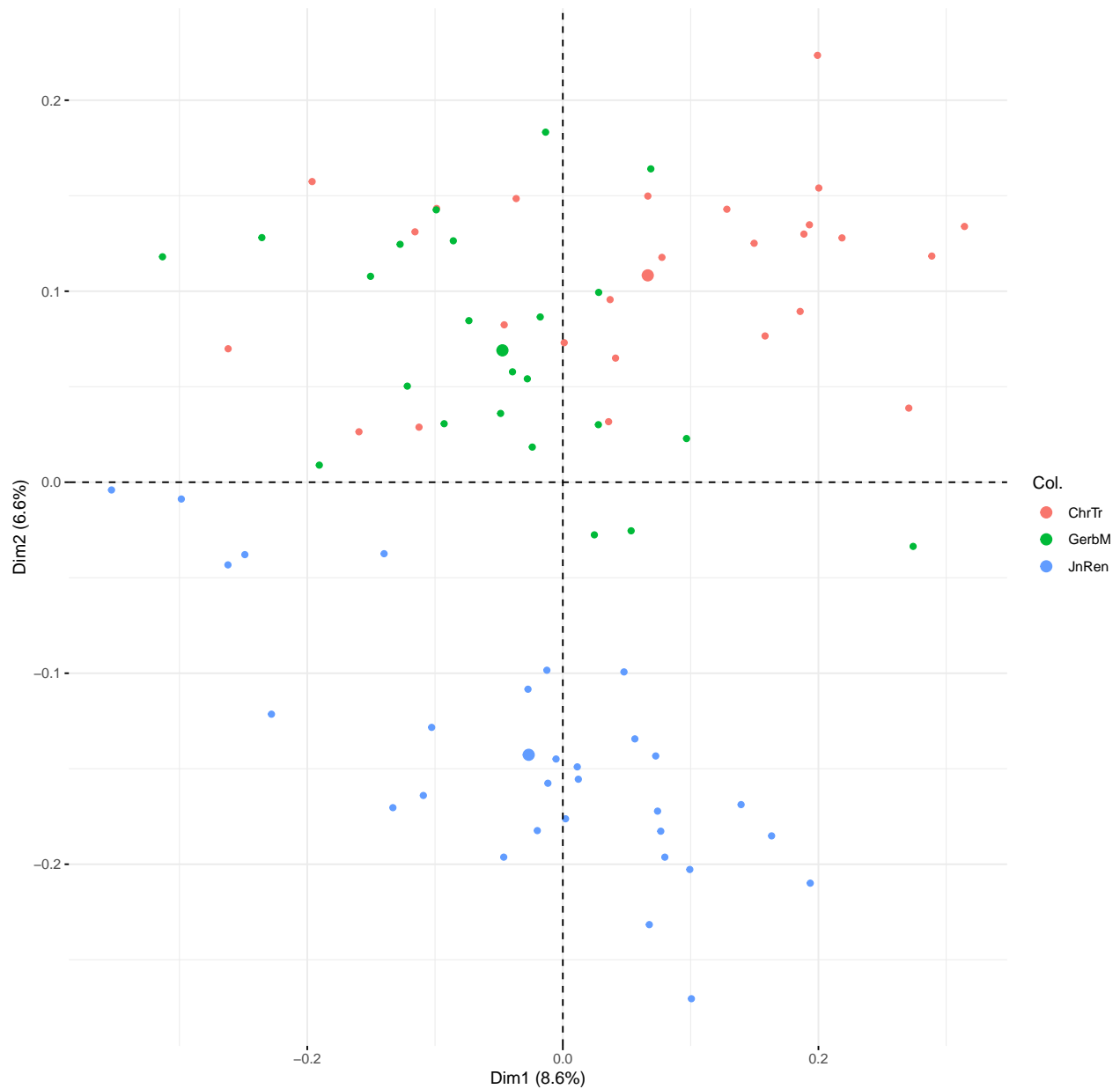
```
#fviz_ca_col(myCA, col.col = sub("\\_.*", "", rownames(myCA$col$coord)), title="", geom=c("point", "tex
```

Clustering

Selecting features

Weighting

In oppositions to analyses like Correspondance analysis, that perform their own kind of 'centering' of the variables, here, we will do two types of normalisations:

- variables (rows) with z-scores;
- individuals (columns) with vector length normalisation (Euclidean norm).

```r
data = normalisations(data)
```

And let's go.

Clustering methods: hierarchical clustering

```r
library(cluster)
maCAH = agnes(t(data), metric = "manhattan", method="ward")
#plot(maCAH, which.plots = 2)
# Et une fonction un peu plus esthétique
maCAH_save = maCAH
cahPlotCol(maCAH, k = 3)
```

```
## Warning: `guides(<scale> = FALSE)` is deprecated. Please use `guides(<scale> =
## "none")` instead.
```
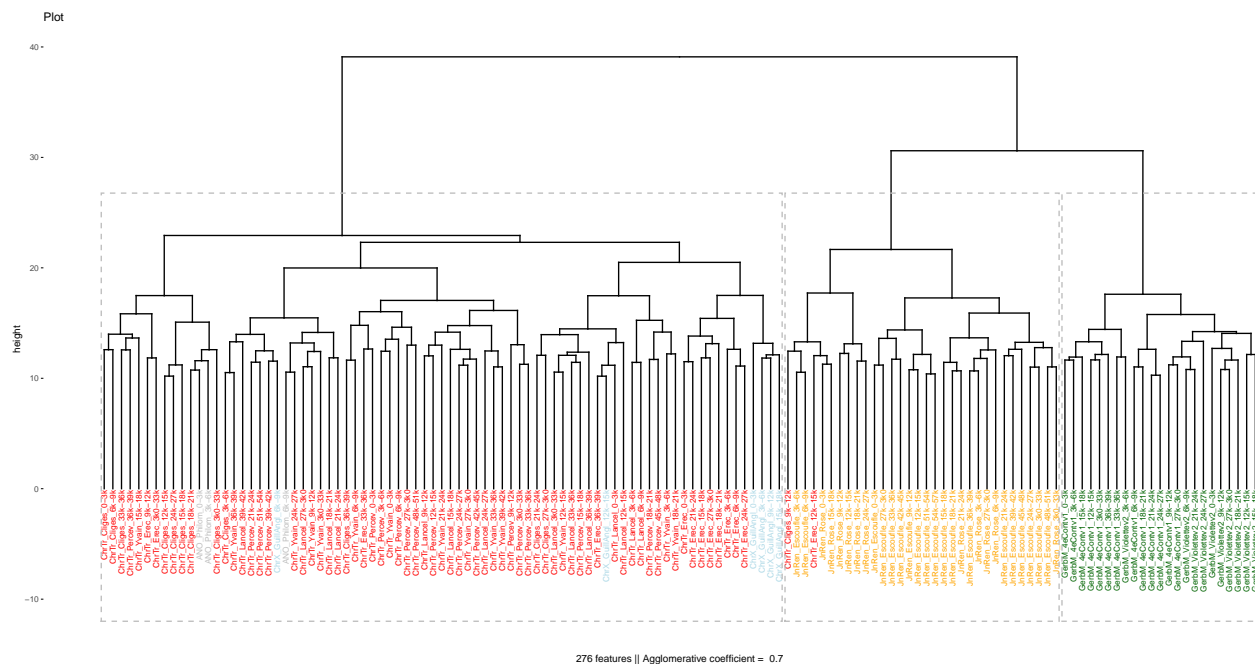

Class description

Creating groups: cutting an AHC

If we want to describe a classification, in terms of the specificities of each group, several solutions are possible. We can start from supervised clustering from K-means or K-medoids, or from the cut of a CAH.

The number of classes should be chosen according to the dendrogram, and a height graph can be used (corresponding to the heights at which the branches of the dendrogram are separated).

```r
maCAH = maCAH_save
maCAH2 = as.hclust(maCAH)
plot(maCAH2$height, type="h", ylab="heights")
```
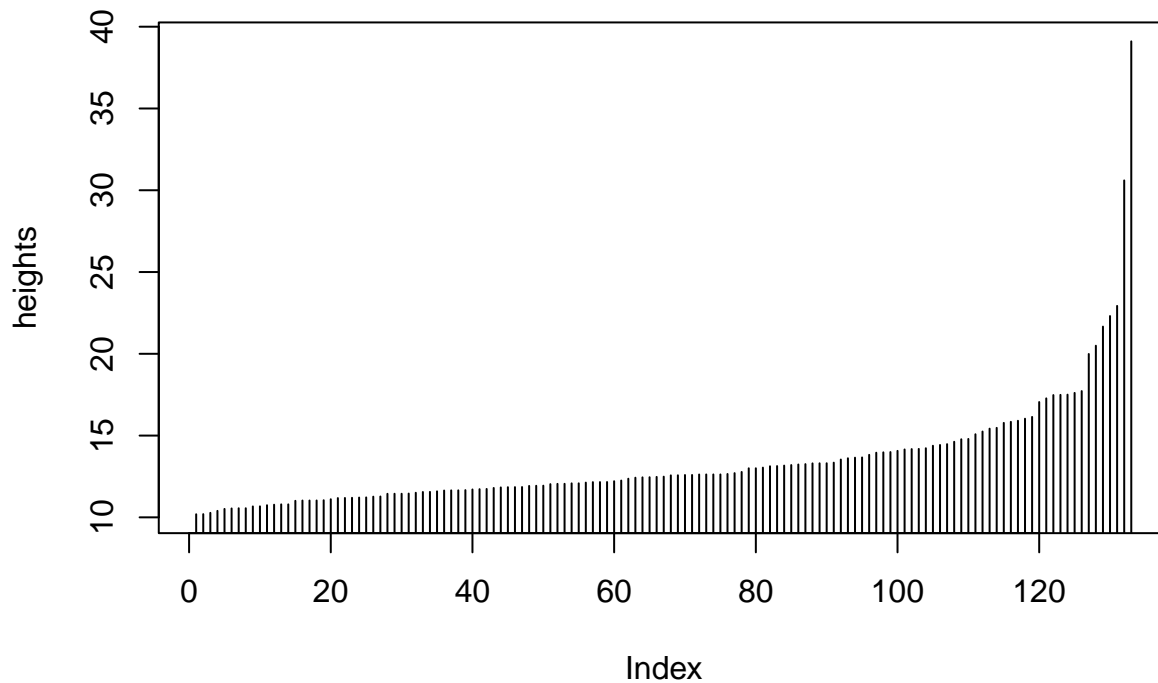
```
#cahPlotCol(maCAH, k = 4)
```

And let's cut

```
#Je découpe en classes
classes = cutree(maCAH, k = "4") # ="4"
#J'ajoute les classes à mon tableau
myClassif = t(data)
myClassif = cbind(as.data.frame(myClassif), as.factor(classes))
colnames(myClassif[length(myClassif)])[] = "Classes"
```

```
myClassif[length(myClassif)]
```

```
##                      as.factor(classes)
## ChrTr_Cliges_0-3k                      1
## ChrTr_Cliges_3k-6k                     2
## ChrTr_Cliges_6k-9k                     1
## ChrTr_Cliges_9k-12k                    3
## ChrTr_Cliges_12k-15k                   1
## ChrTr_Cliges_15k-18k                   1
## ChrTr_Cliges_18k-21k                   1
## ChrTr_Cliges_21k-24k                   2
## ChrTr_Cliges_24k-27k                   1
## ChrTr_Cliges_27k-3k0                   2
## ChrTr_Cliges_3k0-33k                   1
## ChrTr_Cliges_33k-36k                   1
## ChrTr_Cliges_36k-39k                   2
## ChrTr_Erec_0-3k                        2
## ChrTr_Erec_3k-6k                       2
## ChrTr_Erec_6k-9k                       2
## ChrTr_Erec_9k-12k                      1
## ChrTr_Erec_12k-15k                     3
## ChrTr_Erec_15k-18k                     2
```

```
## ChrTr_Erec_18k-21k                 2
## ChrTr_Erec_21k-24k                 2
## ChrTr_Erec_24k-27k                 2
## ChrTr_Erec_27k-3k0                 2
## ChrTr_Erec_3k0-33k                 1
## ChrTr_Erec_33k-36k                 2
## ChrTr_Erec_36k-39k                 2
## GerbM_4eContv1_0-3k                4
## GerbM_4eContv1_3k-6k               4
## GerbM_4eContv1_6k-9k               4
## GerbM_4eContv1_9k-12k              4
## GerbM_4eContv1_12k-15k             4
## GerbM_4eContv1_15k-18k             4
## GerbM_4eContv1_18k-21k             4
## GerbM_4eContv1_21k-24k             4
## GerbM_4eContv1_24k-27k             4
## GerbM_4eContv1_27k-3k0             4
## GerbM_4eContv1_3k0-33k             4
## GerbM_4eContv1_33k-36k             4
## GerbM_4eContv1_36k-39k             4
## GerbM_Violettev2_0-3k              4
## GerbM_Violettev2_3k-6k             4
## GerbM_Violettev2_6k-9k             4
## GerbM_Violettev2_9k-12k            4
## GerbM_Violettev2_12k-15k           4
## GerbM_Violettev2_15k-18k           4
## GerbM_Violettev2_18k-21k           4
## GerbM_Violettev2_21k-24k           4
## GerbM_Violettev2_24k-27k           4
## GerbM_Violettev2_27k-3k0           4
## JnRen_Escoufle_0-3k                3
## JnRen_Escoufle_3k-6k               3
## JnRen_Escoufle_6k-9k               3
## JnRen_Escoufle_9k-12k              3
## JnRen_Escoufle_12k-15k             3
## JnRen_Escoufle_15k-18k             3
## JnRen_Escoufle_18k-21k             3
## JnRen_Escoufle_21k-24k             3
## JnRen_Escoufle_24k-27k             3
## JnRen_Escoufle_27k-3k0             3
## JnRen_Escoufle_3k0-33k             3
## JnRen_Escoufle_33k-36k             3
## JnRen_Escoufle_36k-39k             3
## JnRen_Escoufle_39k-42k             3
## JnRen_Escoufle_42k-45k             3
## JnRen_Escoufle_45k-48k             3
## JnRen_Escoufle_48k-51k             3
## JnRen_Escoufle_51k-54k             3
## JnRen_Escoufle_54k-57k             3
## JnRen_Rose_0-3k                    3
## JnRen_Rose_3k-6k                   3
## JnRen_Rose_6k-9k                   3
## JnRen_Rose_9k-12k                  3
## JnRen_Rose_12k-15k                 3
```

```
## JnRen_Rose_15k-18k                3
## JnRen_Rose_18k-21k                3
## JnRen_Rose_21k-24k                3
## JnRen_Rose_24k-27k                3
## JnRen_Rose_27k-3k0                3
## JnRen_Rose_3k0-33k                3
## ANO_Philom_0-3k                   1
## ANO_Philom_3k-6k                  1
## ANO_Philom_6k-9k                  2
## ChrTr_Lancel_0-3k                 2
## ChrTr_Lancel_3k-6k                2
## ChrTr_Lancel_6k-9k                2
## ChrTr_Lancel_9k-12k               2
## ChrTr_Lancel_12k-15k              2
## ChrTr_Lancel_15k-18k              2
## ChrTr_Lancel_18k-21k              2
## ChrTr_Lancel_21k-24k              2
## ChrTr_Lancel_24k-27k              2
## ChrTr_Lancel_27k-3k0              2
## ChrTr_Lancel_3k0-33k              2
## ChrTr_Lancel_33k-36k              2
## ChrTr_Lancel_36k-39k              2
## ChrTr_Lancel_39k-42k              2
## ChrTr_Percev_0-3k                 2
## ChrTr_Percev_3k-6k                2
## ChrTr_Percev_6k-9k                2
## ChrTr_Percev_9k-12k               2
## ChrTr_Percev_12k-15k              2
## ChrTr_Percev_15k-18k              2
## ChrTr_Percev_18k-21k              2
## ChrTr_Percev_21k-24k              2
## ChrTr_Percev_24k-27k              2
## ChrTr_Percev_27k-3k0              2
## ChrTr_Percev_3k0-33k              2
## ChrTr_Percev_33k-36k              2
## ChrTr_Percev_36k-39k              1
## ChrTr_Percev_39k-42k              2
## ChrTr_Percev_42k-45k              2
## ChrTr_Percev_45k-48k              2
## ChrTr_Percev_48k-51k              2
## ChrTr_Percev_51k-54k              2
## ChrTr_Yvain_0-3k                  2
## ChrTr_Yvain_3k-6k                 2
## ChrTr_Yvain_6k-9k                 2
## ChrTr_Yvain_9k-12k                2
## ChrTr_Yvain_12k-15k               2
## ChrTr_Yvain_15k-18k               1
## ChrTr_Yvain_18k-21k               2
## ChrTr_Yvain_21k-24k               2
## ChrTr_Yvain_24k-27k               2
## ChrTr_Yvain_27k-3k0               2
## ChrTr_Yvain_3k0-33k               2
## ChrTr_Yvain_33k-36k               2
## ChrTr_Yvain_36k-39k               2
```

```
## ChrTr_Yvain_39k-42k                          2
## ChrX_GuillAngl_0-3k                          2
## ChrX_GuillAngl_3k-6k                         2
## ChrX_GuillAngl_6k-9k                         2
## ChrX_GuillAngl_9k-12k                        2
## ChrX_GuillAngl_12k-15k                       2
## ChrX_GuillAngl_15k-18k                       2
```

Describing classes

Let's try to understand now what put *Istorietta* and *Novella* in the same class.

Let's do a v-test,

```
#Je charge FactoMineR
library(FactoMineR)
#Et je décrit les classes
mesClasses = catdes(myClassif, num.var = length(myClassif))
# $\eta^2$
mesClasses$quanti.var[1:20,]
```

```
##                 Eta2       P-value
## mi         0.6067071 3.247561e-26
## lues       0.4205129 2.374671e-15
## ne         0.3729971 3.760159e-13
## ainc       0.3434885 7.182392e-12
## tres       0.3378623 1.240865e-11
## puis       0.3203420 6.605563e-11
## envers     0.2916684 9.268530e-10
## par        0.2834025 1.943543e-09
## i          0.2813232 2.338137e-09
## car        0.2786663 2.958565e-09
## vos        0.2774791 3.285672e-09
## et         0.2731556 4.806280e-09
## piec       0.2667659 8.394990e-09
## que        0.2633077 1.132837e-08
## ensement   0.2571459 1.925019e-08
## rien       0.2507798 3.312733e-08
## son        0.2362306 1.124333e-07
## si         0.2350380 1.241377e-07
## de         0.2332258 1.442518e-07
## tantost    0.2318712 1.613470e-07
```

```
# v-test pour la classe 1
mesClasses$quanti$`2`
```

```
##             v.test Mean in category  Overall mean sd in category Overall sd
## vos       5.359460      0.030460899  1.120168e-03    0.063450038 0.06127847
## si        5.083251      0.027648124 -2.215684e-04    0.057726741 0.06136894
## i         5.048499      0.028875440  7.210587e-04    0.055317092 0.06242257
## et        4.915092      0.027696561  3.416521e-04    0.047462632 0.06229621
## que       4.757119      0.027049811  1.162612e-03    0.049811999 0.06091146
## endroit   4.654196      0.026290375  5.040180e-04    0.070295033 0.06201594
## tu        4.615309      0.024597557  8.189366e-05    0.068271908 0.05945671
## donc      4.420470      0.022049750 -6.270784e-04    0.062040787 0.05742114
## rien      4.391401      0.024091507  2.477895e-04    0.054892920 0.06077553
```

12

```
## tantost       4.258762     0.023322408  1.558072e-04   0.069419302 0.06088873
## se            4.234164     0.023071845  2.775055e-04   0.053905472 0.06025836
## parmi         4.151459     0.022768168 -1.184356e-04   0.068925033 0.06170758
## maintenant    4.092973     0.023463641  4.804627e-04   0.069615063 0.06285346
## jusque        3.989228     0.021857703 -1.384222e-04   0.060036872 0.06171849
## ci            3.952084     0.021676979  3.129010e-04   0.063337580 0.06050844
## enmi          3.934641     0.022318493  5.060378e-04   0.074104625 0.06205223
## ton           3.844895     0.020517787 -9.882763e-05   0.067550324 0.06001928
## environ       3.693252     0.021327809  5.727686e-04   0.073439403 0.06290317
## que_que       3.480142     0.019257324  1.128466e-04   0.074939986 0.06157500
## ne            3.425065     0.018635083  1.765005e-04   0.050633591 0.06032363
## trestot       3.372928     0.019644306  1.254563e-03   0.064185483 0.06102763
## meisme        3.368693     0.018005234 -4.128993e-04   0.062553253 0.06119869
## mie           3.309156     0.016494313 -1.509190e-03   0.060359318 0.06089725
## caienz        3.277302     0.018970486  4.657882e-04   0.081308161 0.06320093
## iqui          3.171573     0.016956785 -8.760715e-05   0.083012934 0.06015404
## mien          3.027660     0.014866353 -1.092468e-03   0.061942868 0.05899994
## ainsi         2.867704     0.015691472  6.005674e-05   0.064601888 0.06101294
## je            2.810644     0.014745309 -5.028395e-04   0.058841380 0.06072524
## volentiers    2.809312     0.016687879  1.201658e-03   0.068013263 0.06170258
## encui         2.792873     0.016210468  3.682679e-04   0.083288744 0.06349248
## des           2.762604     0.017096793  1.374131e-03   0.071018693 0.06370380
## es            2.752415     0.015379703  2.130242e-04   0.072178548 0.06167861
## ja            2.625830     0.014789005  1.803563e-04   0.051459931 0.06227324
## oil           2.581665     0.012584239 -6.323925e-04   0.068270922 0.05730319
## mais          2.553392     0.012564929 -1.142150e-03   0.054113890 0.06008768
## atot          2.465618     0.012709992 -5.181472e-04   0.074960843 0.06005248
## certes        2.363889     0.012084972 -5.361797e-04   0.066289743 0.05976265
## enjusque      2.356413     0.011572064 -7.421027e-04   0.082198688 0.05849404
## aie           2.338659     0.012605241 -4.226268e-04   0.080885263 0.06235402
## pis           2.310551     0.011182819 -7.771464e-04   0.069779594 0.05793917
## ca            2.305968     0.012077357 -1.414693e-04   0.061966815 0.05931086
## autresi       2.232883     0.011789982 -4.959580e-04   0.067487245 0.06158861
## apres         2.231109     0.010932331 -1.245446e-03   0.069165881 0.06109492
## dementre      2.102039     0.010040377 -7.922205e-04   0.081422136 0.05768326
## aillors       2.070503     0.010397556 -7.598855e-04   0.070314483 0.06031797
## eluec         2.041302     0.008297709 -1.316803e-03   0.074488814 0.05272032
## ice           1.981802     0.010164113 -6.027700e-04   0.068215404 0.06081179
## iluec        -2.009794    -0.011430977 -6.313230e-04   0.051663876 0.06014732
## icest        -2.016077    -0.011141893  1.506647e-04   0.044278726 0.06269648
## entor        -2.057124    -0.011405057 -2.773734e-04   0.051353397 0.06054834
## envers       -2.327639    -0.013073954 -1.013764e-04   0.038963980 0.06238333
## enz          -2.367556    -0.013720456 -1.319323e-03   0.047002601 0.05862989
## manois       -2.373573    -0.013120984 -2.078817e-04   0.001622933 0.06089561
## ai           -2.403256    -0.013656049 -1.730094e-03   0.031284952 0.05554578
## le           -2.416834    -0.013252835 -1.687586e-04   0.055476950 0.06059744
## dont         -2.711546    -0.015906348 -1.206050e-03   0.047883434 0.06068301
## tres         -2.765064    -0.015600444 -1.737427e-04   0.031151328 0.06244906
## cant         -2.786984    -0.016665910 -1.353938e-03   0.058250402 0.06149711
## deci         -2.807025    -0.015605093 -6.477949e-04   0.021774340 0.05964374
## las          -2.901549    -0.016888836 -1.272078e-03   0.043108165 0.06024471
## ome          -2.913014    -0.014866453  1.357601e-03   0.048377769 0.06234115
## jadis        -2.938838    -0.016262649  2.907956e-04   0.022451770 0.06304791
## poi          -2.959967    -0.016251744  6.848927e-06   0.042564888 0.06148285
```

13

```
## sore        -3.259225     -0.018019317 -8.899155e-04   0.029367732 0.05882824
## ensement    -3.283226     -0.018108190 -8.794942e-04   0.020945895 0.05873671
## dusque      -3.410004     -0.019195722 -1.349331e-03   0.036619599 0.05858056
## orendroit   -3.717312     -0.020334948 -7.245119e-04   0.021044346 0.05904950
## puis        -3.731630     -0.021536969 -1.224134e-03   0.036261880 0.06092983
## maint       -3.924552     -0.022042516 -1.172905e-03   0.032286701 0.05952264
## lues        -4.065541     -0.022943106  7.760116e-05   0.022926596 0.06338088
## fi          -4.091927     -0.022819552 -2.082586e-04   0.031361276 0.06185224
## piec        -4.131570     -0.022474014 -6.722174e-04   0.002779808 0.05906567
## a           -4.265022     -0.022351659  3.823202e-04   0.051889234 0.05966397
## mi          -4.411298     -0.024340980 -1.054734e-03   0.017372052 0.05908687
## ainc        -4.476988     -0.024980078 -1.402815e-03   0.022988802 0.05894750
## lor         -4.534106     -0.024915500  2.764698e-04   0.031526587 0.06219112
## estre       -4.587131     -0.025254950 -3.262122e-04   0.057952469 0.06082990
## chascun     -4.716923     -0.026285374 -9.237044e-04   0.047058988 0.06018344
## son         -5.025995     -0.027355590 -2.930194e-04   0.052533134 0.06027052
## de          -5.260008     -0.028776604  8.257859e-04   0.040742828 0.06299387
## par         -5.688059     -0.032394786 -3.183964e-04   0.051560498 0.06312179
##                    p.value
## vos        8.347120e-08
## si         3.710293e-07
## i          4.452947e-07
## et         8.874098e-07
## que        1.963753e-06
## endroit    3.252477e-06
## tu         3.925115e-06
## donc       9.848633e-06
## rien       1.126224e-05
## tantost    2.055622e-05
## se         2.294033e-05
## parmi      3.303621e-05
## maintenant 4.258777e-05
## jusque     6.628881e-05
## ci         7.747351e-05
## enmi       8.332105e-05
## ton        1.206040e-04
## environ    2.214040e-04
## que_que    5.011473e-04
## ne         6.146523e-04
## trestot    7.437340e-04
## meisme     7.552546e-04
## mie        9.357766e-04
## caienz     1.048043e-03
## iqui       1.516160e-03
## mien       2.464550e-03
## ainsi      4.134621e-03
## je         4.944250e-03
## volentiers 4.964744e-03
## encui      5.224221e-03
## des        5.734221e-03
## es         5.915756e-03
## ja         8.643808e-03
## oil        9.832494e-03
## mais       1.066795e-02
```

```
## atot        1.367772e-02
## certes      1.808422e-02
## enjusque     1.845241e-02
## aie         1.935311e-02
## pis         2.085765e-02
## ca          2.111244e-02
## autresi      2.555667e-02
## apres       2.567388e-02
## dementre     3.554987e-02
## aillors     3.840527e-02
## eluec       4.122085e-02
## ice         4.750145e-02
## iluec       4.445301e-02
## icest       4.379194e-02
## entor       3.967430e-02
## envers      1.993126e-02
## enz         1.790602e-02
## manois      1.761691e-02
## ai          1.624981e-02
## le          1.565617e-02
## dont        6.697027e-03
## tres        5.691169e-03
## cant        5.320118e-03
## deci        5.000143e-03
## las         3.713224e-03
## ome         3.579590e-03
## jadis       3.294452e-03
## poi         3.076716e-03
## sore        1.117171e-03
## ensement     1.026264e-03
## dusque      6.496200e-04
## orendroit   2.013541e-04
## puis        1.902451e-04
## maint       8.689126e-05
## lues        4.792119e-05
## fi          4.278027e-05
## piec        3.602946e-05
## a           1.998825e-05
## mi          1.027526e-05
## ainc        7.570349e-06
## lor         5.784792e-06
## estre       4.493791e-06
## chascun     2.394383e-06
## son         5.008279e-07
## de          1.440488e-07
## par         1.284913e-08
```

```r
plot(mesClasses, level = 0.01, barplot = TRUE)
```
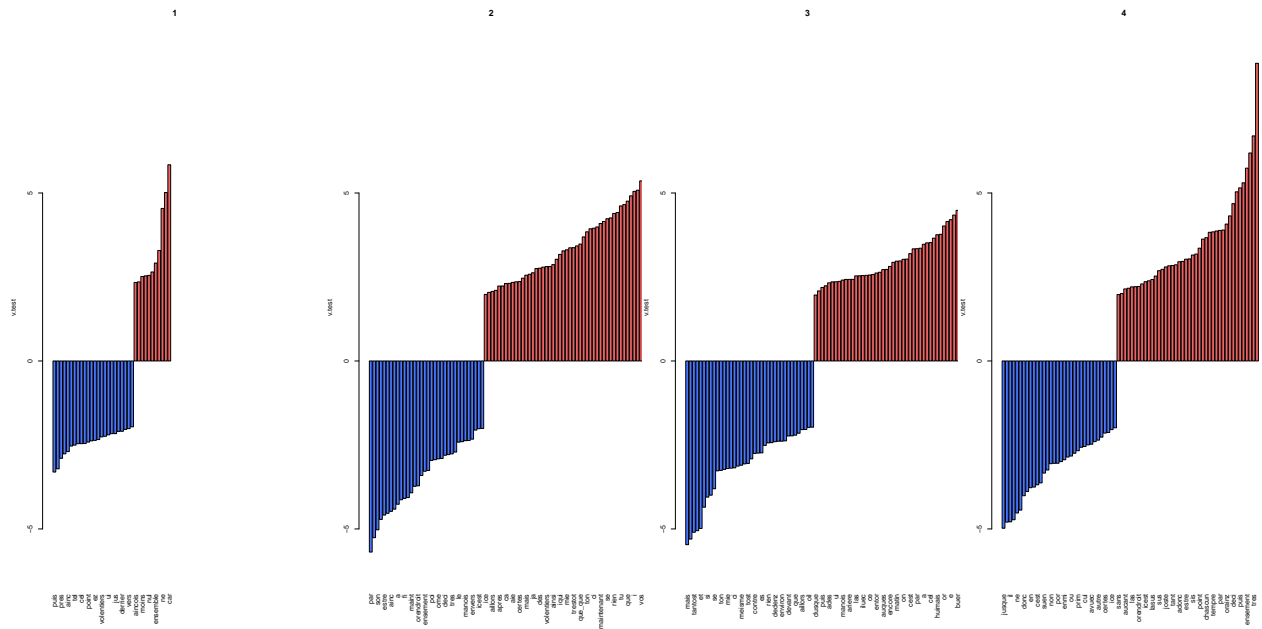
```r
classesDesc(maCAH, data, k = 4)
```

```
## 
## Link between the cluster variable and the quantitative variables
## ===================================================================
##                 Eta2      P-value
## mi         0.60670712 3.247561e-26
## lues       0.42051293 2.374671e-15
## ne         0.37299713 3.760159e-13
## ainc       0.34348854 7.182392e-12
## tres       0.33786234 1.240865e-11
## puis       0.32034202 6.605563e-11
## envers     0.29166842 9.268530e-10
## par        0.28340251 1.943543e-09
```

```
## i         0.28132316 2.338137e-09
## car       0.27866633 2.958565e-09
## vos       0.27747913 3.285672e-09
## et        0.27315563 4.806280e-09
## piec      0.26676592 8.394990e-09
## que       0.26330766 1.132837e-08
## ensement  0.25714594 1.925019e-08
## rien      0.25077976 3.312733e-08
## son       0.23623057 1.124333e-07
## si        0.23503805 1.241377e-07
## de        0.23322582 1.442518e-07
## tantost   0.23187118 1.613470e-07
## mais      0.22941743 1.975263e-07
## se        0.22774457 2.266471e-07
## poi       0.22438012 2.985641e-07
## jusque    0.21919699 4.553024e-07
## maintenant 0.21302270 7.496282e-07
## fi        0.21074267 9.001838e-07
## chascun   0.20409148 1.530141e-06
## enz       0.20367873 1.581092e-06
## il        0.20246725 1.740462e-06
## ai        0.19873398 2.337423e-06
## maint     0.19829976 2.418740e-06
## tu        0.19786456 2.503028e-06
## donc      0.19322010 3.602910e-06
## orendroit 0.19023399 4.547858e-06
## combien   0.18899178 5.009098e-06
## dusque    0.18548582 6.573017e-06
## trestot   0.17883411 1.096559e-05
## estre     0.17332274 1.669540e-05
## lor       0.17200562 1.845069e-05
## ja        0.17190378 1.859373e-05
## deci      0.17079322 2.022573e-05
## endroit   0.16724721 2.643498e-05
## por       0.16680662 2.732653e-05
## buer      0.16279863 3.691548e-05
## sovent    0.15895092 4.919217e-05
## a         0.15527292 6.462963e-05
## parmi     0.15114260 8.765682e-05
## e         0.14665439 1.218140e-04
## o         0.14619194 1.259994e-04
## cest      0.14495938 1.378548e-04
## ou        0.13430070 2.979540e-04
## ton       0.13246346 3.398621e-04
## enmi      0.13071115 3.851805e-04
## cel       0.13011717 4.018444e-04
## ci        0.13000463 4.050804e-04
## orainz    0.12929877 4.259650e-04
## mie       0.12856532 4.487814e-04
## or        0.12828454 4.578285e-04
## adonc     0.12572410 5.490031e-04
## des       0.12458918 5.948867e-04
## cui       0.12162313 7.332379e-04
## en        0.12026707 8.065315e-04
```

```
## come       0.11683558 1.025477e-03
## las        0.11494000 1.170302e-03
## point      0.11490915 1.172817e-03
## nul        0.11333440 1.308446e-03
## sore       0.11256929 1.379757e-03
## cant       0.11240797 1.395270e-03
## tempre     0.11113717 1.523574e-03
## suen       0.11036288 1.607320e-03
## huimais    0.10997474 1.650977e-03
## avuec      0.10965140 1.688224e-03
## environ    0.10890180 1.777761e-03
## jadis      0.10789299 1.905605e-03
## non        0.10574606 2.208251e-03
## ariere     0.10484206 2.349276e-03
## meisme     0.10381111 2.520835e-03
## le         0.10211444 2.830113e-03
## tel        0.09923446 3.441744e-03
## ce         0.09917191 3.456363e-03
## es         0.09541504 4.454609e-03
## mien       0.09466047 4.686539e-03
## que_que    0.09430134 4.801009e-03
## tost       0.09391935 4.925755e-03
## sempre     0.09307490 5.212813e-03
## atant      0.08846098 7.092585e-03
## dedenz     0.08609820 8.295451e-03
## sis        0.08561985 8.562026e-03
## neporcant  0.08546251 8.651511e-03
## sor        0.08518211 8.813239e-03
## ome        0.08397044 9.546488e-03
## entor      0.08385654 9.618382e-03
## caienz     0.08181397 1.100045e-02
## ainsi      0.08161262 1.114667e-02
## ambedos    0.08103703 1.157516e-02
## tant       0.08050352 1.198654e-02
## tot        0.08007807 1.232472e-02
## moins      0.07873092 1.345761e-02
## aincois    0.07669683 1.536143e-02
## iqui       0.07581997 1.626026e-02
## volentiers 0.07430900 1.792973e-02
## encore     0.07415671 1.810690e-02
## soz        0.07193773 2.088847e-02
## atot       0.07178873 2.108935e-02
## certes     0.07151744 2.145988e-02
## manois     0.07090437 2.232052e-02
## autre      0.07032446 2.316520e-02
## matin      0.06954947 2.434225e-02
## on         0.06954364 2.435133e-02
## pres       0.06953102 2.437097e-02
## lendemain  0.06946005 2.448172e-02
## petitet    0.06945828 2.448448e-02
## ui         0.06942912 2.453014e-02
## auques     0.06930973 2.471791e-02
## joste      0.06898824 2.523043e-02
## qui        0.06862502 2.582180e-02
```

```
## foiz         0.06790703 2.703033e-02
## iluec        0.06757093 2.761461e-02
## rechief_     0.06658723 2.939537e-02
## je           0.06625095 3.002909e-02
## ensemble     0.06541562 3.166066e-02
## dont         0.06405448 3.450332e-02
## prim         0.06352656 3.567056e-02
## jus          0.06091837 4.201567e-02
## encui        0.06086042 4.216827e-02
## aie          0.05971690 4.529005e-02
## nos          0.05839427 4.917680e-02
##
## Description of each cluster by quantitative variables
## =====================================================
## $`1`
##               v.test Mean in category  Overall mean sd in category Overall sd
## car         5.840120       0.09510483  7.319929e-04     0.084985104 0.06365382
## combien     5.013557       0.08297548  4.654078e-04     0.180592324 0.06482764
## ne          4.540723       0.06971307  1.765005e-04     0.065066331 0.06032363
## cui         3.287981       0.05393837  8.087358e-04     0.063854875 0.06365121
## ensemble    2.914940       0.04554834 -8.410045e-05     0.059541678 0.06166565
## avuec       2.649216       0.04172054  7.708665e-04     0.065524344 0.06088808
## nul         2.548642       0.04101212  8.837323e-04     0.072516755 0.06202147
## ja          2.533683       0.04023516  1.803563e-04     0.101518351 0.06227324
## moins       2.514954       0.03893063 -1.631001e-04     0.069266689 0.06123169
## suen        2.353823       0.03652777 -2.311615e-05     0.092788715 0.06116786
## aincois     2.336690       0.03714945  4.802917e-04     0.063485487 0.06181575
## ha         -1.960766      -0.03128468 -8.609838e-05     0.046323708 0.06267701
## vers       -2.013021      -0.03169219 -4.607731e-04     0.052172088 0.06111427
## e          -2.036965      -0.03258584 -2.356442e-04     0.015896669 0.06255939
## derrier    -2.093220      -0.03340340 -4.563894e-04     0.003665439 0.06200124
## lues       -2.096717      -0.03365881  7.760116e-05     0.017333542 0.06338088
## jus        -2.163239      -0.03451623 -7.919305e-04     0.030951851 0.06140979
## ou         -2.164355      -0.03399320 -2.386716e-04     0.054551621 0.06143316
## ui         -2.198762      -0.03525572  1.143130e-04     0.049788721 0.06336603
## ainz       -2.242304      -0.03477644 -1.358346e-04     0.049931424 0.06085414
## volentiers -2.260261      -0.03420319  1.201658e-03     0.030769495 0.06170258
## mi         -2.339297      -0.03614423 -1.054734e-03     0.003966197 0.05908687
## ez         -2.363578      -0.03750602 -2.481779e-04     0.004115631 0.06209364
## aval       -2.378365      -0.03728365 -3.278488e-04     0.026250438 0.06120732
## point      -2.414885      -0.03807028 -4.874275e-04     0.033269629 0.06130453
## soz        -2.457764      -0.03867846 -1.280515e-03     0.032033920 0.05993862
## cel        -2.461481      -0.03928960 -9.504197e-04     0.060430316 0.06135438
## si         -2.463004      -0.03859357 -2.215684e-04     0.072191118 0.06136894
## tel        -2.505479      -0.03976567 -1.361219e-03     0.052899832 0.06037957
## adonc      -2.532725      -0.03924841  1.047382e-04     0.004306827 0.06120553
## ainc       -2.697876      -0.04177555 -1.402815e-03     0.015379987 0.05894750
## ariere     -2.764282      -0.04347311 -2.840075e-04     0.036575998 0.06154476
## pres       -2.898186      -0.04681669 -8.612511e-04     0.048526396 0.06246115
## sor        -3.211056      -0.05119791 -1.030976e-03     0.045275802 0.06154161
## puis       -3.306580      -0.05236990 -1.224134e-03     0.051306427 0.06092983
##                  p.value
## car         5.216312e-09
## combien     5.343294e-07
```

```
## ne          5.606164e-06
## cui         1.009085e-03
## ensemble    3.557567e-03
## avuec       8.067871e-03
## nul         1.081431e-02
## ja          1.128706e-02
## moins       1.190478e-02
## suen        1.858145e-02
## aincois     1.945533e-02
## ha          4.990633e-02
## vers        4.411246e-02
## e           4.165353e-02
## derrier     3.632953e-02
## lues        3.601864e-02
## jus         3.052279e-02
## ou          3.043713e-02
## ui          2.789485e-02
## ainz        2.494173e-02
## volentiers  2.380507e-02
## mi          1.932008e-02
## ez          1.809944e-02
## aval        1.738960e-02
## point       1.574020e-02
## soz         1.398050e-02
## cel         1.383647e-02
## si          1.377784e-02
## tel         1.222856e-02
## adonc       1.131797e-02
## ainc        6.978339e-03
## ariere      5.704828e-03
## pres        3.753283e-03
## sor         1.322480e-03
## puis        9.444254e-04
##
## $`2`
##              v.test Mean in category  Overall mean sd in category Overall sd
## vos        5.359460      0.030460899  1.120168e-03      0.063450038 0.06127847
## si         5.083251      0.027648124 -2.215684e-04      0.057726741 0.06136894
## i          5.048499      0.028875440  7.210587e-04      0.055317092 0.06242257
## et         4.915092      0.027696561  3.416521e-04      0.047462632 0.06229621
## que        4.757119      0.027049811  1.162612e-03      0.049811999 0.06091146
## endroit    4.654196      0.026290375  5.040180e-04      0.070295033 0.06201594
## tu         4.615309      0.024597557  8.189366e-05      0.068271908 0.05945671
## donc       4.420470      0.022049750 -6.270784e-04      0.062040787 0.05742114
## rien       4.391401      0.024091507  2.477895e-04      0.054892920 0.06077553
## tantost    4.258762      0.023322408  1.558072e-04      0.069419302 0.06088873
## se         4.234164      0.023071845  2.775055e-04      0.053905472 0.06025836
## parmi      4.151459      0.022768168 -1.184356e-04      0.068925033 0.06170758
## maintenant 4.092973      0.023463641  4.804627e-04      0.069615063 0.06285346
## jusque     3.989228      0.021857703 -1.384222e-04      0.060036872 0.06171849
## ci         3.952084      0.021676979  3.129010e-04      0.063337580 0.06050844
## enmi       3.934641      0.022318493  5.060378e-04      0.074104625 0.06205223
## ton        3.844895      0.020517787 -9.882763e-05      0.067550324 0.06001928
## environ    3.693252      0.021327809  5.727686e-04      0.073439403 0.06290317
```

```
## que_que     3.480142      0.019257324  1.128466e-04   0.074939986 0.06157500
## ne          3.425065      0.018635083  1.765005e-04   0.050633591 0.06032363
## trestot     3.372928      0.019644306  1.254563e-03   0.064185483 0.06102763
## meisme      3.368693      0.018005234 -4.128993e-04   0.062553253 0.06119869
## mie         3.309156      0.016494313 -1.509190e-03   0.060359318 0.06089725
## caienz      3.277302      0.018970486  4.657882e-04   0.081308161 0.06320093
## iqui        3.171573      0.016956785 -8.760715e-05   0.083012934 0.06015404
## mien        3.027660      0.014866353 -1.092468e-03   0.061942868 0.05899994
## ainsi       2.867704      0.015691472  6.005674e-05   0.064601888 0.06101294
## je          2.810644      0.014745309 -5.028395e-04   0.058841380 0.06072524
## volentiers  2.809312      0.016687879  1.201658e-03   0.068013263 0.06170258
## encui       2.792873      0.016210468  3.682679e-04   0.083288744 0.06349248
## des         2.762604      0.017096793  1.374131e-03   0.071018693 0.06370380
## es          2.752415      0.015379703  2.130242e-04   0.072178548 0.06167861
## ja          2.625830      0.014789005  1.803563e-04   0.051459931 0.06227324
## oil         2.581665      0.012584239 -6.323925e-04   0.068270922 0.05730319
## mais        2.553392      0.012564929 -1.142150e-03   0.054113890 0.06008768
## atot        2.465618      0.012709992 -5.181472e-04   0.074960843 0.06005248
## certes      2.363889      0.012084972 -5.361797e-04   0.066289743 0.05976265
## enjusque    2.356413      0.011572064 -7.421027e-04   0.082198688 0.05849404
## aie         2.338659      0.012605241 -4.226268e-04   0.080885263 0.06235402
## pis         2.310551      0.011182819 -7.771464e-04   0.069779594 0.05793917
## ca          2.305968      0.012077357 -1.414693e-04   0.061966815 0.05931086
## autresi     2.232883      0.011789982 -4.959580e-04   0.067487245 0.06158861
## apres       2.231109      0.010932331 -1.245446e-03   0.069165881 0.06109492
## dementre    2.102039      0.010040377 -7.922205e-04   0.081422136 0.05768326
## aillors     2.070503      0.010397556 -7.598855e-04   0.070314483 0.06031797
## eluec       2.041302      0.008297709 -1.316803e-03   0.074488814 0.05272032
## ice         1.981802      0.010164113 -6.027700e-04   0.068215404 0.06081179
## iluec      -2.009794     -0.011430977 -6.313230e-04   0.051663876 0.06014732
## icest      -2.016077     -0.011141893  1.506647e-04   0.044278726 0.06269648
## entor      -2.057124     -0.011405057 -2.773734e-04   0.051353397 0.06054834
## envers     -2.327639     -0.013073954 -1.013764e-04   0.038963980 0.06238333
## enz        -2.367556     -0.013720456 -1.319323e-03   0.047002601 0.05862989
## manois     -2.373573     -0.013120984 -2.078817e-04   0.001622933 0.06089561
## ai         -2.403256     -0.013656049 -1.730094e-03   0.031284952 0.05554578
## le         -2.416834     -0.013252835 -1.687586e-04   0.055476950 0.06059744
## dont       -2.711546     -0.015906348 -1.206050e-03   0.047883434 0.06068301
## tres       -2.765064     -0.015600444 -1.737427e-04   0.031151328 0.06244906
## cant       -2.786984     -0.016665910 -1.353938e-03   0.058250402 0.06149711
## deci       -2.807025     -0.015605093 -6.477949e-04   0.021774340 0.05964374
## las        -2.901549     -0.016888836 -1.272078e-03   0.043108165 0.06024471
## ome        -2.913014     -0.014866453  1.357601e-03   0.048377769 0.06234115
## jadis      -2.938838     -0.016262649  2.907956e-04   0.022451770 0.06304791
## poi        -2.959967     -0.016251744  6.848927e-06   0.042564888 0.06148285
## sore       -3.259225     -0.018019317 -8.899155e-04   0.029367732 0.05882824
## ensement   -3.283226     -0.018108190 -8.794942e-04   0.020945895 0.05873671
## dusque     -3.410004     -0.019195722 -1.349331e-03   0.036619599 0.05858056
## orendroit  -3.717312     -0.020334948 -7.245119e-04   0.021044346 0.05904950
## puis       -3.731630     -0.021536969 -1.224134e-03   0.036261880 0.06092983
## maint      -3.924552     -0.022042516 -1.172905e-03   0.032286701 0.05952264
## lues       -4.065541     -0.022943106  7.760116e-05   0.022926596 0.06338088
## fi         -4.091927     -0.022819552 -2.082586e-04   0.031361276 0.06185224
## piec       -4.131570     -0.022474014 -6.722174e-04   0.002779808 0.05906567
```

21

```
## a          -4.265022  -0.022351659  3.823202e-04   0.051889234 0.05966397
## mi         -4.411298  -0.024340980 -1.054734e-03   0.017372052 0.05908687
## ainc       -4.476988  -0.024980078 -1.402815e-03   0.022988802 0.05894750
## lor        -4.534106  -0.024915500  2.764698e-04   0.031526587 0.06219112
## estre      -4.587131  -0.025254950 -3.262122e-04   0.057952469 0.06082990
## chascun    -4.716923  -0.026285374 -9.237044e-04   0.047058988 0.06018344
## son        -5.025995  -0.027355590 -2.930194e-04   0.052533134 0.06027052
## de         -5.260008  -0.028776604  8.257859e-04   0.040742828 0.06299387
## par        -5.688059  -0.032394786 -3.183964e-04   0.051560498 0.06312179
##                      p.value
## vos       8.347120e-08
## si        3.710293e-07
## i         4.452947e-07
## et        8.874098e-07
## que       1.963753e-06
## endroit   3.252477e-06
## tu        3.925115e-06
## donc      9.848633e-06
## rien      1.126224e-05
## tantost   2.055622e-05
## se        2.294033e-05
## parmi     3.303621e-05
## maintenant 4.258777e-05
## jusque    6.628881e-05
## ci        7.747351e-05
## enmi      8.332105e-05
## ton       1.206040e-04
## environ   2.214040e-04
## que_que   5.011473e-04
## ne        6.146523e-04
## trestot   7.437340e-04
## meisme    7.552546e-04
## mie       9.357766e-04
## caienz    1.048043e-03
## iqui      1.516160e-03
## mien      2.464550e-03
## ainsi     4.134621e-03
## je        4.944250e-03
## volentiers 4.964744e-03
## encui     5.224221e-03
## des       5.734221e-03
## es        5.915756e-03
## ja        8.643808e-03
## oil       9.832494e-03
## mais      1.066795e-02
## atot      1.367772e-02
## certes    1.808422e-02
## enjusque  1.845241e-02
## aie       1.935311e-02
## pis       2.085765e-02
## ca        2.111244e-02
## autresi   2.555667e-02
## apres     2.567388e-02
## dementre  3.554987e-02
```

```
## aillors     3.840527e-02
## eluec       4.122085e-02
## ice         4.750145e-02
## iluec       4.445301e-02
## icest       4.379194e-02
## entor       3.967430e-02
## envers      1.993126e-02
## enz         1.790602e-02
## manois      1.761691e-02
## ai          1.624981e-02
## le          1.565617e-02
## dont        6.697027e-03
## tres        5.691169e-03
## cant        5.320118e-03
## deci        5.000143e-03
## las         3.713224e-03
## ome         3.579590e-03
## jadis       3.294452e-03
## poi         3.076716e-03
## sore        1.117171e-03
## ensement    1.026264e-03
## dusque      6.496200e-04
## orendroit   2.013541e-04
## puis        1.902451e-04
## maint       8.689126e-05
## lues        4.792119e-05
## fi          4.278027e-05
## piec        3.602946e-05
## a           1.998825e-05
## mi          1.027526e-05
## ainc        7.570349e-06
## lor         5.784792e-06
## estre       4.493791e-06
## chascun     2.394383e-06
## son         5.008279e-07
## de          1.440488e-07
## par         1.284913e-08
##
## $`3`
##              v.test Mean in category  Overall mean sd in category Overall sd
## lues       7.432426       0.07300456  7.760116e-05    0.090846838 0.06338088
## ainc       5.993744       0.05329410 -1.402815e-03    0.085818381 0.05894750
## piec       5.588973       0.05043315 -6.722174e-04    0.089631823 0.05906567
## ai         5.104455       0.04216336 -1.730094e-03    0.089897669 0.05554578
## son        4.762659       0.04414490 -2.930194e-04    0.048266300 0.06027052
## buer       4.483014       0.04465774  7.059136e-04    0.098141931 0.06332973
## de         4.342790       0.04317705  8.257859e-04    0.072617816 0.06299387
## e          4.205535       0.04049423 -2.356442e-04    0.101030083 0.06255939
## por        4.150096       0.03857561 -9.683274e-04    0.054023452 0.06154922
## or         4.018021       0.03830009  3.234897e-04    0.071409318 0.06105266
## orendroit  3.770621       0.03374446 -7.245119e-04    0.090464500 0.05904950
## huimais    3.757853       0.03611431 -2.655292e-05    0.091457791 0.06212400
## lor        3.656308       0.03547872  2.764698e-04    0.077611886 0.06219112
## cel        3.523703       0.03251869 -9.504197e-04    0.061729450 0.06135438
```

```
## sempre     3.516382      0.02851283 -9.845056e-04  0.105575048 0.05418604
## a          3.473814      0.03246850  3.823202e-04  0.058955505 0.05966397
## jadis      3.356658      0.03305330  2.907956e-04  0.090910368 0.06304791
## par        3.342869      0.03234775 -3.183964e-04  0.054456069 0.06312179
## o          3.337712      0.03305395  5.699007e-04  0.072297592 0.06286689
## cest       3.195129      0.03053971  1.484300e-04  0.071006305 0.06144143
## tel        3.033653      0.02699545 -1.361219e-03  0.059204659 0.06037957
## on         3.026679      0.03019895  7.613598e-04  0.060873332 0.06282559
## le         2.977872      0.02776694 -1.687586e-04  0.059276894 0.06059744
## matin      2.971334      0.02901206  5.040591e-04  0.081157651 0.06197494
## ou         2.936827      0.02769194 -2.386716e-04  0.058966796 0.06143316
## encore     2.814237      0.02792735  1.012135e-03  0.079349141 0.06177857
## non        2.725671      0.02567863  2.777969e-05  0.063093810 0.06078955
## auques     2.719138      0.02593740  2.001416e-04  0.081334333 0.06114088
## il         2.641227      0.02547260  5.509920e-05  0.064368290 0.06216239
## entor      2.618108      0.02426345 -2.773734e-04  0.070551788 0.06054834
## ome        2.574238      0.02620167  1.357601e-03  0.075951036 0.06234115
## ce         2.559692      0.02304635 -5.938111e-04  0.064138103 0.05965727
## el         2.548908      0.02481374  3.912810e-04  0.078035328 0.06189220
## iluec      2.542842      0.02304613 -6.313230e-04  0.066460405 0.06014732
## comfait    2.536353      0.01980304 -1.068248e-03  0.106093619 0.05315453
## las        2.532571      0.02234792 -1.272078e-03  0.082206145 0.06024471
## chascun    2.430028      0.02171688 -9.237044e-04  0.049632629 0.06018344
## ariere     2.428010      0.02284947 -2.840075e-04  0.063950565 0.06154476
## estre      2.427895      0.02253748 -3.262122e-04  0.053912271 0.06082990
## manois     2.408062      0.02249353 -2.078817e-04  0.094705724 0.06089561
## nos        2.369114      0.02366874  3.932747e-04  0.069957473 0.06346192
## ui         2.358038      0.02324596  1.143130e-04  0.068583323 0.06336603
## pas        2.355156      0.02222685 -9.319817e-05  0.069390430 0.06121759
## ades       2.325868      0.01996805 -9.526394e-04  0.083831923 0.05810207
## maint      2.238094      0.01945048 -1.172905e-03  0.055912663 0.05952264
## puis       2.189135      0.01942500 -1.224134e-03  0.065311069 0.06092983
## aucun      2.085238      0.01916836 -7.368930e-04  0.076655437 0.06166133
## dusque     1.963475      0.01645715 -1.349331e-03  0.067025978 0.05858056
## encui     -1.971527     -0.01901041  3.682679e-04  0.001901022 0.06349248
## oil       -1.977005     -0.01817063 -6.323925e-04  0.044182274 0.05730319
## adonc     -2.040477     -0.01922926  1.047382e-04  0.038143949 0.06120553
## aillors   -2.044258     -0.01984882 -7.598855e-04  0.035217348 0.06031797
## je        -2.150860     -0.02072282 -5.028395e-04  0.063580035 0.06072524
## que       -2.208703     -0.01966481  1.162612e-03  0.061001284 0.06091146
## envers    -2.230456     -0.02164216 -1.013764e-04  0.035942046 0.06238333
## devant    -2.233445     -0.02024449  6.552606e-04  0.044788428 0.06044586
## des       -2.378667     -0.02208426  1.374131e-03  0.039989927 0.06370380
## environ   -2.388401     -0.02268559  5.727686e-04  0.030885670 0.06290317
## i         -2.389928     -0.02237435  7.210587e-04  0.054925822 0.06242257
## dedenz    -2.403494     -0.02271998 -2.883155e-04  0.038781679 0.06028640
## soi       -2.432170     -0.02440553 -9.233953e-04  0.063216078 0.06236548
## rien      -2.438433     -0.02269461  2.477895e-04  0.047680729 0.06077553
## atot      -2.513871     -0.02388893 -5.181472e-04  0.002388868 0.06005248
## es        -2.735924     -0.02591087  2.130242e-04  0.022646326 0.06167861
## car       -2.742781     -0.02629606  7.319929e-04  0.046444602 0.06365382
## contre    -2.752595     -0.02593343 -6.030255e-04  0.037841625 0.05944297
## ainsi     -2.914341     -0.02746711  6.005674e-05  0.036053395 0.06101294
## tost      -3.049823     -0.02975053 -3.795220e-04  0.047885323 0.06220782
```

```
## sovent      -3.064049      -0.02840109  1.123506e-04   0.028934527 0.06011109
## meisme      -3.103562      -0.02981659 -4.128993e-04   0.042441871 0.06119869
## endroit     -3.129932      -0.02954550  5.040180e-04   0.026393421 0.06201594
## ci          -3.181006      -0.02948459  3.129010e-04   0.036386048 0.06050844
## mien        -3.190808      -0.03023663 -1.092468e-03   0.032988937 0.05899994
## mie         -3.201299      -0.03168947 -1.509190e-03   0.052597661 0.06089725
## neporcant   -3.228553      -0.03178713 -9.719695e-04   0.029318143 0.06165344
## ton         -3.256965      -0.03036122 -9.882763e-05   0.031645006 0.06001928
## ne          -3.270540      -0.03036611  1.765005e-04   0.040976147 0.06032363
## se          -3.804313      -0.03521142  2.775055e-04   0.043941728 0.06025836
## tu          -3.995618      -0.03669577  8.189366e-05   0.014433212 0.05945671
## si          -4.052618      -0.03872359 -2.215684e-04   0.042655313 0.06136894
## trestot     -4.352984      -0.03987109  1.254563e-03   0.042080524 0.06102763
## et          -4.986366      -0.04774727  3.416521e-04   0.062043300 0.06229621
## maintenant -5.051955      -0.04867682  4.804627e-04   0.024624399 0.06285346
## tantost     -5.099135      -0.04790961  1.558072e-04   0.004790911 0.06088873
## vos         -5.301580      -0.04917340  1.120168e-03   0.036543728 0.06127847
## mais        -5.467919      -0.05200572 -1.142150e-03   0.037975960 0.06008768
##                   p.value
## lues        1.066241e-13
## ainc        2.050647e-09
## piec        2.284159e-08
## ai          3.317490e-07
## son         1.910582e-06
## buer        7.359626e-06
## de          1.406848e-05
## e           2.604650e-05
## por         3.323356e-05
## or          5.868905e-05
## orendroit   1.628420e-04
## huimais     1.713774e-04
## lor         2.558737e-04
## cel         4.255604e-04
## sempre      4.374704e-04
## a           5.131163e-04
## jadis       7.889053e-04
## par         8.291718e-04
## o           8.447121e-04
## cest        1.397683e-03
## tel         2.416120e-03
## on          2.472564e-03
## le          2.902575e-03
## matin       2.965094e-03
## ou          3.315895e-03
## encore      4.889312e-03
## non         6.417085e-03
## auques      6.545234e-03
## il          8.260623e-03
## entor       8.841876e-03
## ome         1.004612e-02
## ce          1.047650e-02
## el          1.080607e-02
## iluec       1.099549e-02
## comfait     1.120139e-02
```

```
## las        1.132295e-02
## chascun    1.509764e-02
## ariere     1.518191e-02
## estre      1.518674e-02
## manois     1.603748e-02
## nos        1.783078e-02
## ui         1.837180e-02
## pas        1.851494e-02
## ades       2.002558e-02
## maint      2.521489e-02
## puis       2.858704e-02
## aucun      3.704773e-02
## dusque     4.959099e-02
## encui      4.866364e-02
## oil        4.804103e-02
## adonc      4.130288e-02
## aillors    4.092806e-02
## je         3.148722e-02
## que        2.719534e-02
## envers     2.571719e-02
## devant     2.551964e-02
## des        1.737535e-02
## environ    1.692189e-02
## i          1.685168e-02
## dedenz     1.623925e-02
## soi        1.500865e-02
## rien       1.475109e-02
## atot       1.194140e-02
## es         6.220539e-03
## car        6.092136e-03
## contre     5.912498e-03
## ainsi      3.564404e-03
## tost       2.289765e-03
## sovent     2.183633e-03
## meisme     1.912061e-03
## endroit    1.748467e-03
## ci         1.467646e-03
## mien       1.418753e-03
## mie        1.368095e-03
## neporcant  1.244184e-03
## ton        1.126102e-03
## ne         1.073424e-03
## se         1.421983e-04
## tu         6.452572e-05
## si         5.064775e-05
## trestot    1.342970e-05
## et         6.152565e-07
## maintenant 4.373117e-07
## tantost    3.412089e-07
## vos        1.148049e-07
## mais       4.553490e-08
##
## $`4`
##              v.test Mean in category  Overall mean sd in category Overall sd
```

```
## mi          8.860288      0.09867177 -1.054734e-03   0.075548267 0.05908687
## tres         6.699704      0.07952532 -1.737427e-04   0.082675138 0.06244906
## envers       6.191128      0.07347020 -1.013764e-04   0.093633469 0.06238333
## ensement     5.739909      0.06334288 -8.794942e-04   0.105331314 0.05873671
## poi          5.302837      0.06211292  6.848927e-06   0.082777579 0.06148285
## puis         5.153040      0.05858469 -1.224134e-03   0.058713626 0.06092983
## enz          5.035529      0.05491946 -1.319323e-03   0.078470174 0.05862989
## deci         4.682404      0.05255146 -6.477949e-04   0.105954611 0.05964374
## fi           4.318187      0.05066958 -2.082586e-04   0.084632142 0.06185224
## orainz       4.075043      0.04772238 -1.328464e-05   0.096827771 0.06149489
## maint        3.897498      0.04301871 -1.172905e-03   0.090623851 0.05952264
## par          3.885349      0.04639926 -3.183964e-04   0.060966998 0.06312179
## dusque       3.868600      0.04182037 -1.349331e-03   0.078304501 0.05858056
## tempre       3.843215      0.03847080 -1.182135e-03   0.123234862 0.05416378
## sovent       3.830098      0.04396908  1.123506e-04   0.098510052 0.06011109
## chascun      3.669706      0.04114703 -9.237044e-04   0.074696205 0.06018344
## cant         3.627231      0.04113752 -1.353938e-03   0.048659205 0.06149711
## point        3.357684      0.03872323 -4.874275e-04   0.081233373 0.06130453
## atant        3.183078      0.03607355 -7.231649e-04   0.095692039 0.06068621
## sis          3.153477      0.03551504 -6.843108e-04   0.108210857 0.06026141
## lendemain    3.038398      0.03113288 -6.593563e-04   0.127883913 0.05492939
## estre        3.029665      0.03477996 -3.262122e-04   0.053306422 0.06082990
## foiz         2.964439      0.03457912  1.114868e-04   0.078964591 0.06103757
## adonc        2.953588      0.03454071  1.047382e-04   0.074926573 0.06120553
## petitet      2.862830      0.03225530 -5.320008e-04   0.100828392 0.06012269
## tant         2.842788      0.03358045  3.046525e-05   0.074946971 0.06195496
## dedenz       2.833031      0.03224606 -2.883155e-04   0.077375680 0.06028640
## joste        2.799203      0.02647022 -1.685000e-03   0.087484447 0.05280230
## sore         2.727872      0.02967911 -8.899155e-04   0.079828331 0.05882824
## sus          2.687564      0.03155611 -3.783354e-04   0.078204982 0.06237761
## desor        2.527606      0.02785510 -6.782375e-04   0.066483721 0.05926133
## lasus        2.424991      0.02514968 -8.581839e-04   0.111017396 0.05630187
## rechief_     2.386776      0.02774250 -2.117305e-04   0.064151632 0.06148430
## icest        2.359137      0.02832594  1.506647e-04   0.084433942 0.06269648
## issi         2.292783      0.02852595  5.446814e-04   0.095117114 0.06406674
## orendroit    2.217995      0.02422422 -7.245119e-04   0.061692736 0.05904950
## entretant    2.211721      0.02255370 -5.897820e-04   0.130121193 0.05493215
## las          2.206046      0.02404451 -1.272078e-03   0.057390281 0.06024471
## qui          2.165746      0.02495938 -5.710659e-04   0.056368839 0.06188412
## aucant       2.144321      0.02330456 -6.497723e-04   0.116050088 0.05864385
## icel         2.006525      0.02379383  2.376075e-04   0.067906242 0.06162961
## sans         1.976922      0.02378167  4.369380e-04   0.074722284 0.06199086
## et          -1.989383     -0.02326595  3.416521e-04   0.055964862 0.06229621
## ice         -2.042487     -0.02426300 -6.027700e-04   0.020130624 0.06081179
## lez         -2.127542     -0.02466199 -2.066088e-04   0.034963140 0.06034267
## certes      -2.146468     -0.02497194 -5.361797e-04   0.032488281 0.05976265
## aincois     -2.267523     -0.02622040  4.802917e-04   0.051570970 0.06181575
## autre       -2.356429     -0.02743384 -3.004737e-04   0.047361134 0.06044740
## mon         -2.401235     -0.02794967 -4.928951e-04   0.032971632 0.06002650
## avuec       -2.478135     -0.02797193  7.708665e-04   0.048101198 0.06088808
## des         -2.497496     -0.02893280  1.374131e-03   0.043344630 0.06370380
## cui         -2.545563     -0.03005598  8.087358e-04   0.041719361 0.06365121
## ambedos     -2.569604     -0.02968998 -5.997531e-04   0.045649397 0.05943046
## prim        -2.672400     -0.03095482 -1.658407e-03   0.018641704 0.05754945
```

```
## tot        -2.746142    -0.03243767  4.403260e-04   0.047776862 0.06285077
## ou         -2.829706    -0.03335300 -2.386716e-04   0.046191288 0.06143316
## se         -2.859866    -0.03254976  2.775055e-04   0.047253328 0.06025836
## enmi       -2.940292    -0.03424914  5.060378e-04   0.003474892 0.06205223
## nul        -2.996190    -0.03451462  8.837323e-04   0.046397068 0.06202147
## por        -3.044471    -0.03666322 -9.683274e-04   0.047256381 0.06154922
## ce         -3.046914    -0.03521926 -5.938111e-04   0.044328296 0.05965727
## non        -3.057819    -0.03538112  2.777969e-05   0.033205793 0.06078955
## parmi      -3.245538    -0.03826865 -1.184356e-04   0.016661941 0.06170758
## suen       -3.337055    -0.03890599 -2.311615e-05   0.021898918 0.06116786
## o          -3.629492    -0.04289509  5.699007e-04   0.040368225 0.06286689
## cest       -3.686134    -0.04299396  1.484300e-04   0.043825556 0.06144143
## ja         -3.754082    -0.04435213  1.803563e-04   0.034244452 0.06227324
## en         -3.772330    -0.04280914  6.394603e-04   0.065559443 0.06046365
## come       -3.885236    -0.04372729  6.245793e-04   0.048290043 0.05992704
## donc       -4.011530    -0.04450576 -6.270784e-04   0.016951845 0.05742114
## rien       -4.438971    -0.05114271  2.477895e-04   0.037350190 0.06077553
## ne         -4.525100    -0.05182159  1.765005e-04   0.033682177 0.06032363
## que        -4.726479    -0.05367878  1.162612e-03   0.040169908 0.06091146
## il         -4.787733    -0.05663789  5.509920e-05   0.039393770 0.06216239
## i          -4.796448    -0.05631286  7.210587e-04   0.028964240 0.06242257
## jusque     -4.979225    -0.05867789 -1.384222e-04   0.014023400 0.06171849
##                 p.value
## mi         7.980833e-19
## tres       2.088428e-11
## envers     5.973520e-10
## ensement   9.472755e-09
## poi        1.140164e-07
## puis       2.562970e-07
## enz        4.765318e-07
## deci       2.835297e-06
## fi         1.573160e-05
## orainz     4.600576e-05
## maint      9.719155e-05
## par        1.021832e-04
## dusque     1.094622e-04
## tempre     1.214330e-04
## sovent     1.280923e-04
## chascun    2.428292e-04
## cant       2.864772e-04
## point      7.859834e-04
## atant      1.457185e-03
## sis        1.613378e-03
## lendemain  2.378398e-03
## estre      2.448252e-03
## foiz       3.032355e-03
## adonc      3.141028e-03
## petitet    4.198763e-03
## tant       4.472078e-03
## dedenz     4.610886e-03
## joste      5.122895e-03
## sore       6.374436e-03
## sus        7.197528e-03
## desor      1.148432e-02
```

```
## lasus      1.530877e-02
## rechief_   1.699686e-02
## icest      1.831748e-02
## issi       2.186048e-02
## orendroit  2.655519e-02
## entretant  2.698593e-02
## las        2.738081e-02
## qui        3.033061e-02
## aucant     3.200715e-02
## icel       4.480029e-02
## sans       4.805050e-02
## et         4.665892e-02
## ice        4.110321e-02
## lez        3.337504e-02
## certes     3.183570e-02
## aincois    2.335832e-02
## autre      1.845159e-02
## mon        1.633983e-02
## avuec      1.320711e-02
## des        1.250738e-02
## cui        1.091018e-02
## ambedos    1.018148e-02
## prim       7.531074e-03
## tot        6.030059e-03
## ou         4.659074e-03
## se         4.238198e-03
## enmi       3.279034e-03
## nul        2.733765e-03
## por        2.330901e-03
## ce         2.312035e-03
## non        2.229542e-03
## parmi      1.172288e-03
## suen       8.467133e-04
## o          2.839795e-04
## cest       2.276863e-04
## ja         1.739781e-04
## en         1.617304e-04
## come       1.022307e-04
## donc       6.032656e-05
## rien       9.038997e-06
## ne         6.036687e-06
## que        2.284464e-06
## il         1.686762e-06
## i          1.615036e-06
## jusque     6.383953e-07
```

It is also possible to use other computations of specificities, like Lafon (1984), that needs absolute frequencies.

```
# install.packages("textometry")
freq_abs = data_abs_save
# On splitte le corpus selon les classes
freq_abs_class = matrix(nrow = nrow(freq_abs), ncol = length(unique(classes)), dimnames = list(rownames
for(i in 1:length(unique(classes))){
  # sum the values for each member of the class
  freq_abs_class[, i] = rowSums(freq_abs[, classes == i])
```

```
}
specs = textometry::specificities(freq_abs_class)
# and now we can look at the specificities for class 1
# positive or negative
head(sort(specs[, 2], decreasing = TRUE))
```

```
##      tu     vos      i      et      si      je
## 71.3169 46.1653 21.0711 15.5810 13.4017 13.3224
```

```
head(sort(specs[, 2]))
```

```
##      de     son     par     lor      mi      le
## -34.5720 -30.3887 -22.4133 -21.6038 -19.2344 -18.6369
```

And one can use other varieties of non-parametric tests, such as the Mann-Whitney test, for example, to compare two populations...