Getting started with ClustVarLV. version 1.4.1

Evelyne Vigneau

2015-05-29

beginners with ClustVarLV

The ClustVarLV package is dedicated to the CLV method for the Clustering of Variables Around Latent Variables (Vigneau & Qannari,2003),

```r
library(ClustVarLV)
```

For illustration, we consider the "apples_sh" dataset which includes the sensory characterization and consumers preference for 12 varieties of apples (Daillant-Spinnler et al.,1996).

```r
data(apples_sh)
# 43 sensory attributes of 12 varieties of apple from southern hemisphere
senso<-apples_sh$senso
# Scores of liking given fy 60 consumers for each of the 12 varieties of apple
pref<-apples_sh$pref
```

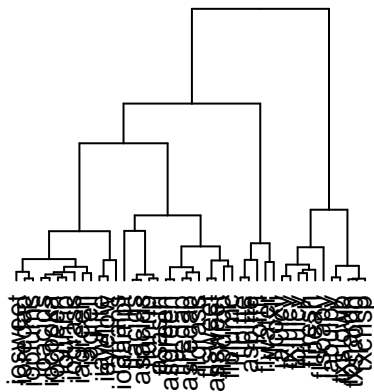## Clustering of the sensory attributes

The aim is to find groups of sensory attributes correlated, or anti-correlated, to each others. Herein "directional" groups are sought. Each group is associated with a latent component which makes it possible to identify the underlying sensory dimensions.

```r
resclv_senso <- CLV(X = senso, method = "directional", sX = TRUE)
# option sX=TRUE means that each attribute will be auto-scaled (standard deviation =1)

# Print of the 'clv' object
print(resclv_senso)

# Dendrogram of the CLV hierarchical clustering algorithm :
plot(resclv_senso,"dendrogram")
```
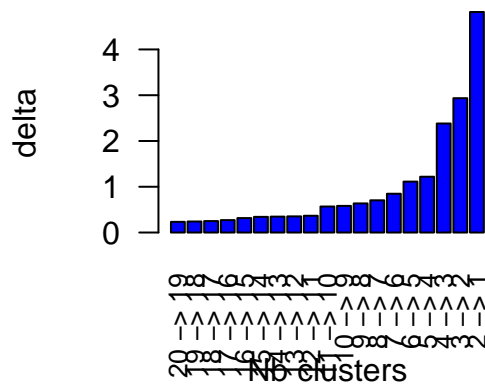
CLV Dendrogram


```
# Graph of the variation of the clustering criterion
plot(resclv_senso,"delta")
```

Variation of criterion (after consolidation

The graph of the variation of the clustering criterion between a partition into K clusters and a partition into (K-1) clusters (after consolidation) is useful for determining the number of clusters to be retained. Because the criterion clearly jumps when passing from 4 to 3 groups, a partition into 4 groups is retained.

```
# Summary the CLV results for a partition into 4 groups
summary(resclv_senso,K=4)
```

```
## $number
## clusters
##  1  2  3  4
## 12 14 12  5
##
```
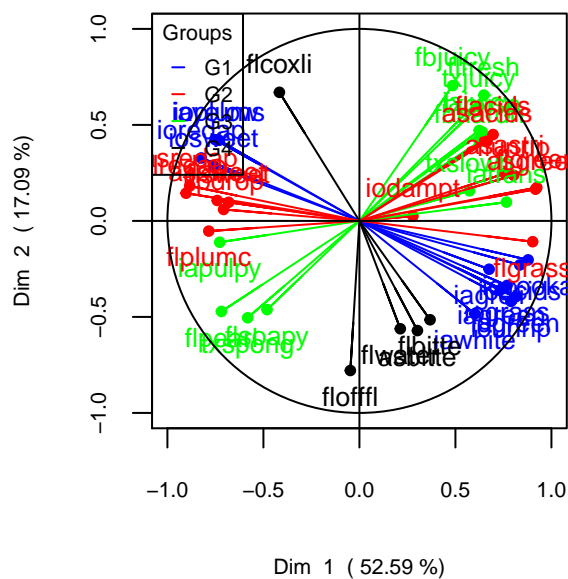
```
## $prop_within
##       Group.1 Group.2 Group.3 Group.4
## [1,]  0.8355  0.7337   0.734  0.7289
##
## $prop_tot
## [1] 0.7616
##
## $groups
## $groups[[1]]
##         cor in group  |cor|next group
## iogreen         0.98             0.74
## ioredap        -0.97             0.80
## ioacids         0.96             0.74
## iounrip         0.96             0.68
## iocooka         0.96             0.81
## iagreen         0.92             0.60
## ioplums        -0.90             0.75
## iograss         0.89             0.72
## iayelow        -0.89             0.63
## iagreli         0.89             0.55
## iosweet        -0.86             0.79
## iawhite         0.76             0.60
##
## $groups[[2]]
##         cor in group  |cor|next group
## asgreen         0.94             0.80
## flgreen         0.93             0.81
## flredap        -0.93             0.88
## flunrip         0.93             0.64
## asredap        -0.93             0.83
## asastri         0.92             0.58
## assweet        -0.90             0.63
## flsweet        -0.86             0.60
## flacids         0.86             0.62
## flgrass         0.85             0.82
## flplumc        -0.83             0.66
## asacids         0.83             0.56
## flpdrop        -0.76             0.59
## iodampt         0.33             0.27
##
## $groups[[3]]
##         cor in group  |cor|next group
## txcrisp         0.97             0.54
## txjuicy         0.94             0.58
## txspong        -0.94             0.53
## fbhardn         0.93             0.55
## iajuicy         0.90             0.52
## flfresh         0.90             0.66
## iapulpy        -0.83             0.64
## flpearl        -0.81             0.77
## iatrans         0.79             0.66
## fbjuicy         0.79             0.53
## txslowb         0.78             0.45
## flsoapy        -0.64             0.53
```

```
## 
## $groups[[4]]
##         cor in group  |cor|next group
## asbitte          0.95              0.34
## flbitte          0.90              0.41
## flcoxli         -0.84              0.54
## flofffl          0.81              0.29
## flwater          0.75              0.31
## 
## 
## $set_aside
## NULL
## 
## $cormatrix
##       Comp1 Comp2 Comp3 Comp4
## Comp1  1.00  0.76  0.43  0.43
## Comp2  0.76  1.00  0.67  0.19
## Comp3  0.43  0.67  1.00  0.01
## Comp4  0.43  0.19  0.01  1.00
```
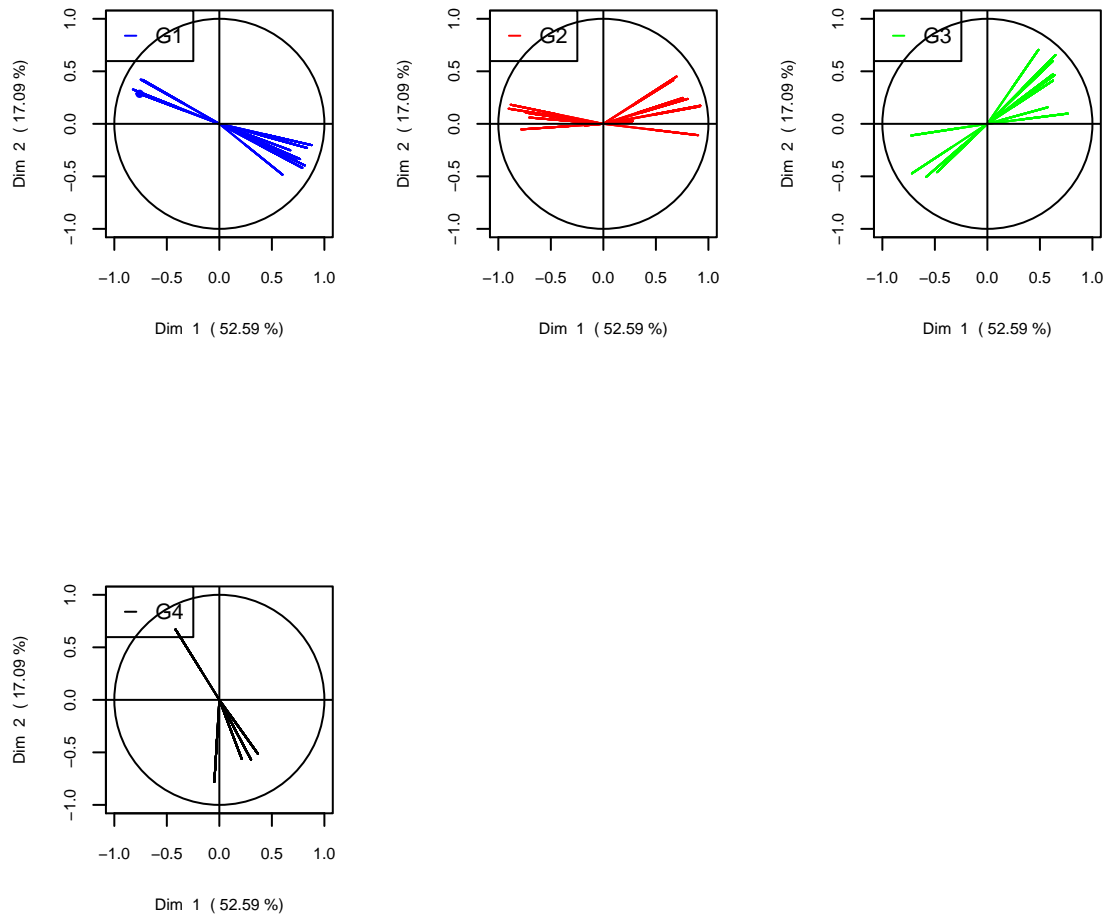
The function plot_var() allows us to describe the groups of variables into a two dimensional space obtained by Principal Components Analysis. Several options are available for the choice of the axes, for adding labels, producing a plot without colours but symbols, having only one plot or a plot by groups of variables.

```
# Representation of the group membership for a partition into 4 groups
plot_var(resclv_senso,K=4,label=T,cex.lab=0.8)
```


or

```
plot_var(resclv_senso,K=4,beside=T)
```


Additional functions :

```
# Extract the group membership of each variable
get_partition(resclv_senso,K=4,type="vector")
# or
get_partition(resclv_senso,K=4,type="matrix")

# Extract the group latent variables
get_comp(resclv_senso,K=4)
```

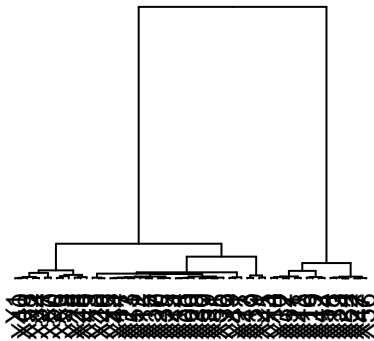## Clustering of the consumers' preference data

The aim is to find segments of consumers. Herein "local" groups are sought. Each group latent variable represents a synthetic direction of preference. If, simultaneously, the aim is to explain these directions of

preference by means of the sensory attributes of the products, the sensory data has to be included as external data.

```
res.segext<- CLV(X = pref, Xr = senso, method = "local", sX=TRUE, sXr = TRUE)

print(res.segext)

plot(res.segext,"dendrogram")
```
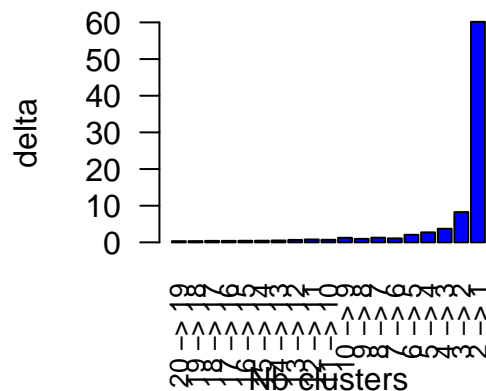
CLV Dendrogram


```
plot(res.segext,"delta")
```

Variation of criterion (after consolidation

Two or three segments may be explored. To Compare the partitions into two or three segments :

```
table(get_partition(res.segext,K=2),get_partition(res.segext,K=3))
```

```
##
##      1  2  3
##    1 12 28  0
##    2  2  0 18
```

Each latent variable being a linear combination of the external variables (sensory), it is possible to extract the associated loadings

```
get_load(res.segext,K=3)
```

```
##                 Comp1        Comp2        Comp3
## iosweet   0.09589740 -0.11243600   0.170065231
## ioacids  -0.04386473  0.14575014  -0.181554627
## iogreen  -0.11714632  0.09817949  -0.188648499
## ioredap   0.08408320 -0.10912674   0.210758453
## iograss  -0.12869347  0.09064836  -0.164391644
## iounrip  -0.11022465  0.10832505  -0.159098176
## iocooka  -0.03972714  0.15610353  -0.180077680
## iodampt   0.01716467  0.06379372  -0.104010800
## ioplums   0.12919401 -0.06974771   0.194629392
## iawhite  -0.14814587  0.05571348  -0.072367992
## iagreen  -0.06489321  0.11467716  -0.167953543
## iayelow   0.10745652 -0.09712042   0.135004928
## iagreli   0.01142059  0.12511441  -0.172633043
## iajuicy   0.20647838  0.21459247  -0.113849158
## iatrans   0.08181864  0.20027109  -0.113725956
## iapulpy  -0.02530062 -0.16927896   0.123476168
## fbjuicy   0.23053897  0.21393868  -0.085459713
## fbhardn   0.14217811  0.21350356  -0.074432978
## txcrisp   0.18273207  0.23060437  -0.081569097
## txjuicy   0.21365042  0.23119349  -0.100979076
## txslowb   0.05108176  0.15451433  -0.073779800
## txspong  -0.17462895 -0.21172483   0.075861428
## flgreen   0.08240198  0.21148600  -0.212099157
## flredap   0.06608129 -0.14564504   0.190123018
## flsweet   0.07624093 -0.08040511   0.191963762
## flacids   0.12759887  0.17878462  -0.214577876
## flbitte  -0.28768482 -0.01853208  -0.002779502
## flgrass  -0.05981788  0.15996944  -0.166073967
## flfresh   0.25544419  0.25096548  -0.132150873
## flpdrop   0.09307223 -0.10540908   0.103396259
## flwater  -0.24348387 -0.05511837   0.021650008
## flofffl  -0.32171652 -0.12038473   0.098986497
## flplumc  -0.02766669 -0.15985868   0.159218105
## flunrip   0.03659315  0.16572691  -0.227213615
## flcoxli   0.27848883  0.04023097   0.064057522
## flpearl  -0.14561103 -0.20872572   0.152113914
## flsoapy  -0.22462943 -0.17028960   0.122183453
## assweet   0.11042842 -0.08378596   0.195131491
## asacids   0.10209954  0.16407373  -0.208978301
```

```
## asbitte -0.32002510 -0.04858914  0.007748031
## asgreen  0.08426373  0.20758903 -0.222199420
## asredap  0.08574768 -0.13062593  0.184568786
## asastri  0.02003463  0.15128160 -0.216116849
```

Using the CLV_kmeans function

This procedure is less time consuming when the number of variables is large. The number of clusters needs to be fixed (e.g.3).

The initialization of the algorithm can be made at random, "nstart" times :

```
res.clvkm.rd<-CLV_kmeans(X = pref, Xr = senso, method = "local", sX=TRUE,
                         sXr = TRUE, clust=3, nstart=100)
```

or the initialization can be defined by the user, for instance on the basis of the clusters obtained by cutting the CLV dendrogram to get 3 clusters

```
res.clvkm.hc<-CLV_kmeans(X = pref, Xr = senso, method = "local", sX=TRUE,
                         sXr = TRUE, clust=res.segext[[3]]$clusters[1,])
```

It is possible to compare the partitions according to the procedure used :

```
table(get_partition(res.segext,K=3),get_partition(res.clvkm.hc,K=3))
```

```
##
##      1  2  3
##   1 14  0  0
##   2  0 28  0
##   3  0  0 18
```

In this case, the CLV solution is the same that the CLV_kmeans solution with an initialization based on the partition obtained by cutting the dendrogram.

```
table(get_partition(res.segext,K=3),get_partition(res.clvkm.rd,K=3))
```

```
##
##      1  2  3
##   1 13  0  1
##   2  0  0 28
##   3  1 17  0
```

Partitions are very close.

Warning : Changes with respect to the earlier versions of the ClustVarLV package

The changes are illustrated on the basis of the examples given above.

from version 1.4.0	for earlier versions
resclv_senso <- CLV(X = senso,method = "directional", sX = TRUE) plot(resclv_senso,"dendrogram"); plot(resclv_senso,"delta")	resclv_senso <- CLV(X = senso,method=1, sX = TRUE, graph=TRUE)
summary(resclv_senso,K=4)	descript_gp(resclv_senso,X=senso,K=4)
plot_var(resclv_senso,K=4)	gpmb_on_pc(resclv_senso,X=senso,K=4)
get_partition(resclv_senso,K=4,type="vector")	resclv_senso[[4]]$clusters[2,]
get_comp(resclv_senso,K=4)	resclv_senso[[4]]$comp
get_load(res.segext,K=3)	res.segext[[3]]$loading
res.clvkm.rd<-CLV_kmeans(X = pref, Xr = senso, method = "local", sX=TRUE, sXr = TRUE, clust=3, nstart=100)	res.clvkm.rd<-CLV_kmeans(X = pref, Xr = senso, method = 2, sX=TRUE, sXr = TRUE, init=3, nstart=100)

References

Vigneau E., Qannari E.M. (2003). Clustering of variables around latents componenets. *Comm. Stat*, 32(4), 1131-1150.

Daillant-Spinnler B., MacFie H.J.H, Beyts P., Hedderley D. (1996). Relationships"Relationships between perceived sensory properties and major preference directions of 12 varieties of apples from the southern hemisphere. *Food Quality and Preference*, 7(2), 113-126.