Three-way clustering around latent variables approach with constraints to improve configurations' interpretability

Véronique Cariou [a], Tom F. Wilderjans [b,c]

[a] StatSC, ONIRIS, INRA, 44322 Nantes, France

[b] Methodology and Statistics Research Unit, Leiden University, The Netherlands

[c] Research Group of Quantitative Psychology and Individual Differences, Leuven, Belgium

XXVIe Rencontres de la Société Francophone de Classification

3 – 5 sept 2019, Nancy

Introduction

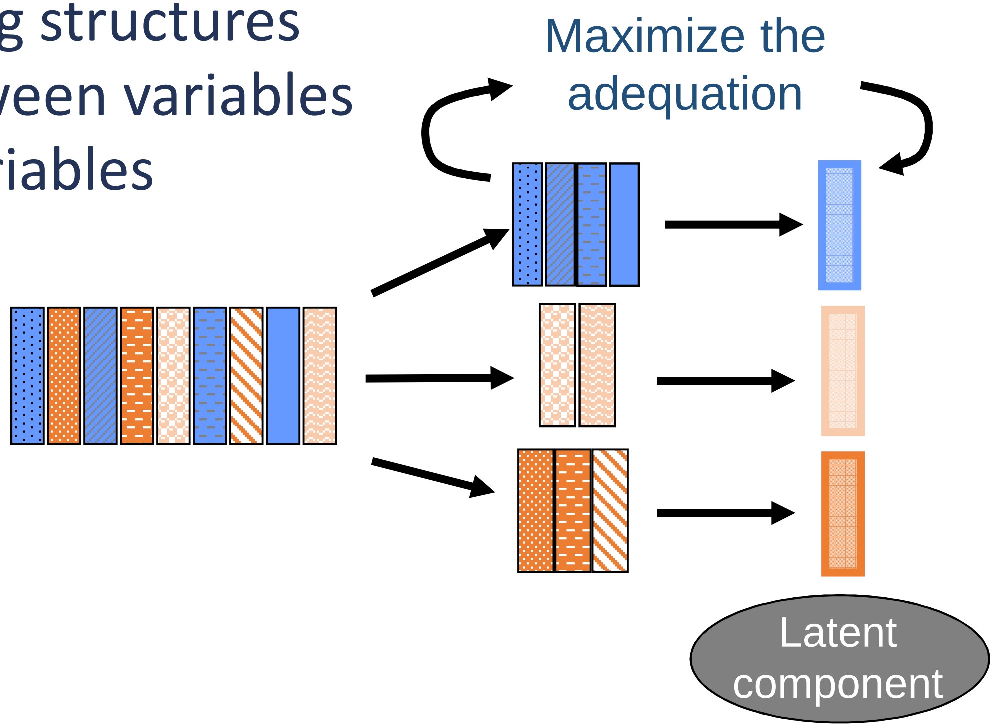
Clustering of variables around latent components with CLV:

- ✓ Understand the underlying structures
- ✓ Detect redundancies between variables
- ✓ Reduce the number of variables

Maximize

$$T^{(Q)} = \sum_{q=1}^{Q} \sum_{j=1}^{J} \delta_{qj} \, cov^2(\mathbf{x_j}, \mathbf{c_q})$$

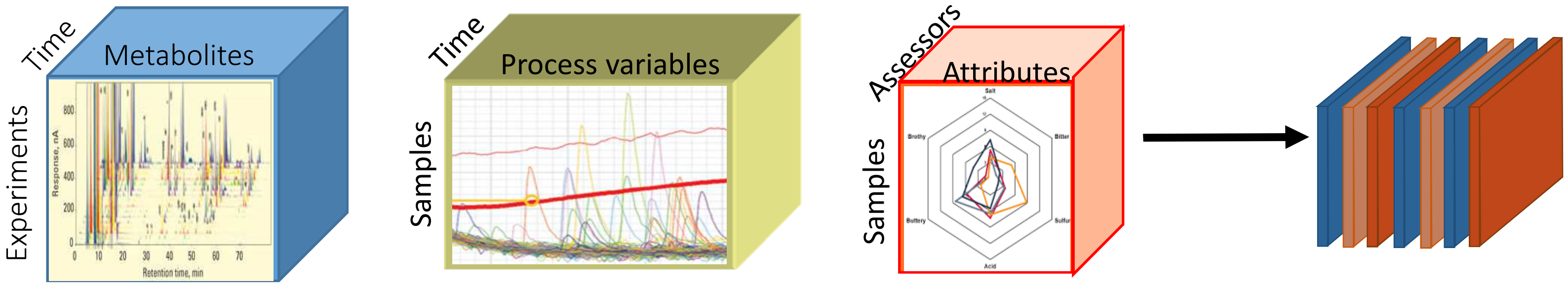
with $var(\mathbf{c_q}) = 1$



Vigneau, E., & Qannari, E. M. (2003). Clustering of variables around latent component. *Communications in Statistics – Simulation and Computation*, **32**,1131–1150

Introduction

Extension of CLV to three-way arrays with CLV3W:



Connexion with clusterwise models and other clustering techniques

Basford, K. E., et McLachlan, G. J. 1985. The mixture method of clustering applied to three-way data. *Journal of Classification*, **2**(1), 109–125

Vichi, M. 1999. One-mode classification of a three-way data matrix. Journal of Classification, **16**(1), 27–44

Llobell, F., Cariou, V., Vigneau, E., Labenne, A., et Qannari, E. M. 2018. Analysis and clustering of multiblock datasets by means of the statis and clustatis methods. application to sensometrics. *Food Quality and Preference*.

Overview

Clustering around latent components for three-way data

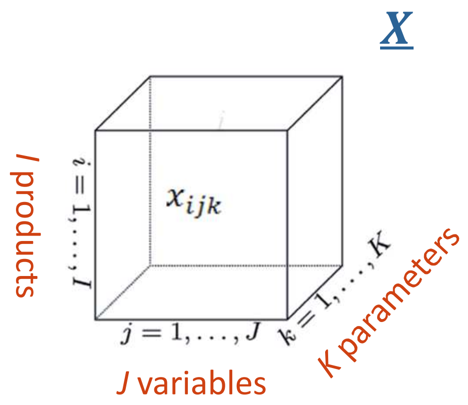
Adding a non negativity constraint on loadings

Application on sensory data for consumers' segmentation

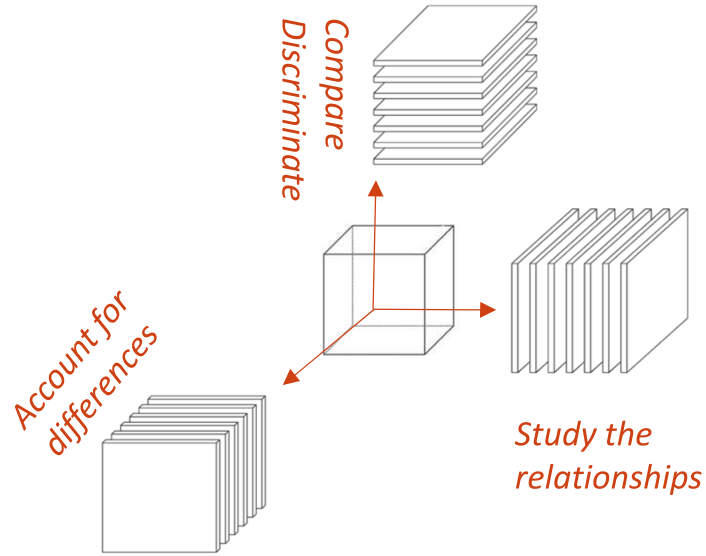
Imposing a global weighting scheme

Conclusion and perspectives

Clustering around latent components for Three-way data



x_{ijk}: score associated with the i^{th} product rated by the j^{th} consumer along with the k^{th} parameters



$\boldsymbol{X}_j \; - j^{th}$ lateral slice of $\underline{\boldsymbol{X}}$ – contains the scores of I samples on the j^{th} variable according to K parameters.

Clustering around latent components for Three-way data

Three-way partitioning of the 2nd mode:

✓Maximise the adequation between each lateral slice and the estimated one associated with its group

Minimize the Loss function:

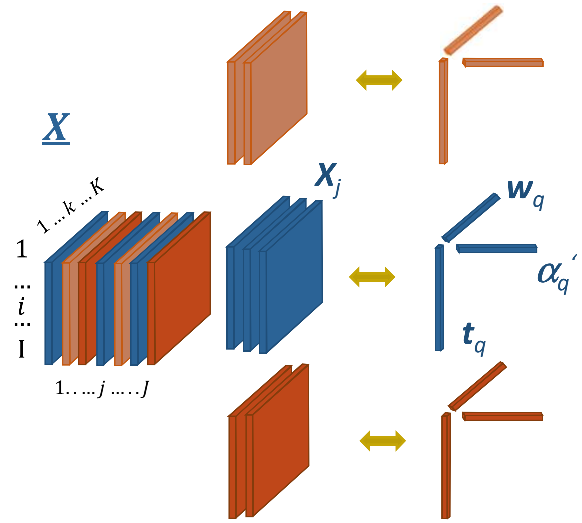
$$f = \sum_{j=1}^{J} \sum_{q=1}^{Q} \delta_{jq} \left\| \mathbf{X}_j - \alpha_{jq}\left(\mathbf{t}_q \mathbf{w}_q^{\mathrm{T}}\right) \right\|_F^2$$

1st mode latent component: $\boldsymbol{t}_q$

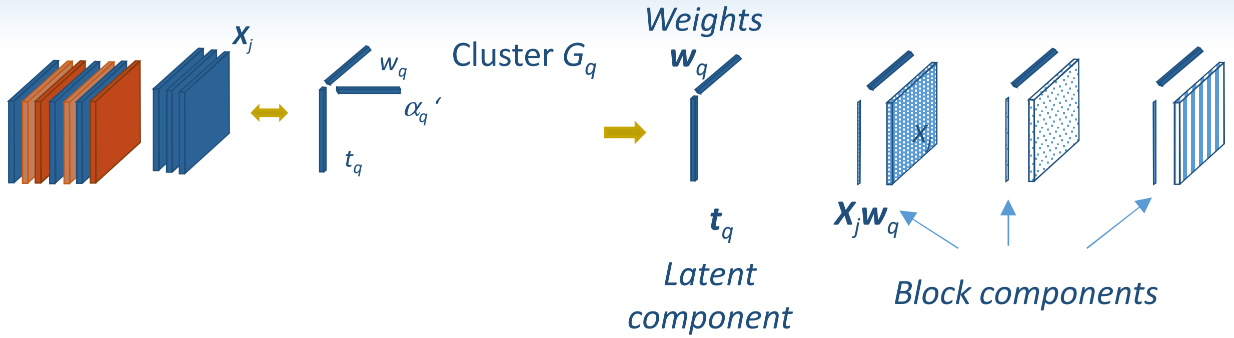
2nd mode loadings of variables: α_q

3rd mode attributes' weights: $\boldsymbol{w}_q$

cluster dummy: δ_{jq}



Clustering around latent components for Three-way data



Minimize the Loss function: ↔ Maximise the squared covariance function:

$$f = \sum_{j=1}^{J} \sum_{q=1}^{Q} \delta_{jq} \left\| \boldsymbol{X}_j - \alpha_{jq} \left(\boldsymbol{t}_q \boldsymbol{w}_q^{\intercal} \right) \right\|_F^2 \qquad g = \sum_{j=1}^{J} \sum_{q=1}^{Q} \delta_{jq} cov^2 \left(\boldsymbol{X}_j \boldsymbol{w}_q, \boldsymbol{t}_q \right)$$

with $\left\| \boldsymbol{w}_q \right\|$=1 and $\left\| \boldsymbol{t}_q \right\|$=1

Krijnen,W. P. 1993. The analysis of three-way arrays by constrained PARAFAC methods. DSWO Press
Wilderjans T. F., et Ceulemans, E. 2013. Clusterwise Parafac to identify heterogeneity in three-way data. *Chemometrics and Intelligent Laboratory Systems*, **129**, 87–97

Initialisation

- Using a multi-start procedure obtained by randomly assigning the variables to *Q* clusters

- Using a rational initial partitioning obtained by applying a Hierarchical Ascendant Algorithm based on criterion *f* and Ward's aggregation criterion

  - The variation of the criterion from step *l* to step *l*+1 corresponds to the aggregation of two clusters, say G_{A} and G_{B}.
  - It can be written as:

$$\Delta f = -\sum_{\boldsymbol{X}_{j\in G_A}} \left\| \boldsymbol{X}_j - \alpha_{jG_A}\left(\boldsymbol{t}_{G_A}\boldsymbol{w}_{G_A}^{\top}\right)\right\|_F^2 - \sum_{\boldsymbol{X}_{j\in G_B}} \left\| \boldsymbol{X}_j - \alpha_{jG_B}\left(\boldsymbol{t}_{G_B}\boldsymbol{w}_{G_B}^{\top}\right)\right\|_F^2$$

$$+ \sum\nolimits_{\boldsymbol{X}_{j\in(G_A\cup G_B)}} \left\| \boldsymbol{X}_j - \alpha_{j(G_A\cup G_B)}\left[\boldsymbol{t}_{(G_A\cup G_B)}\boldsymbol{w}_{(G_A\cup G_B)}^{\top}\right]\right\|_F^2$$

Algorithm

Starting from an initial partitioning into Q clusters

Iterate the following two steps until convergence:

(1) computing the cluster-specific parameters conditional on the cluster memberships

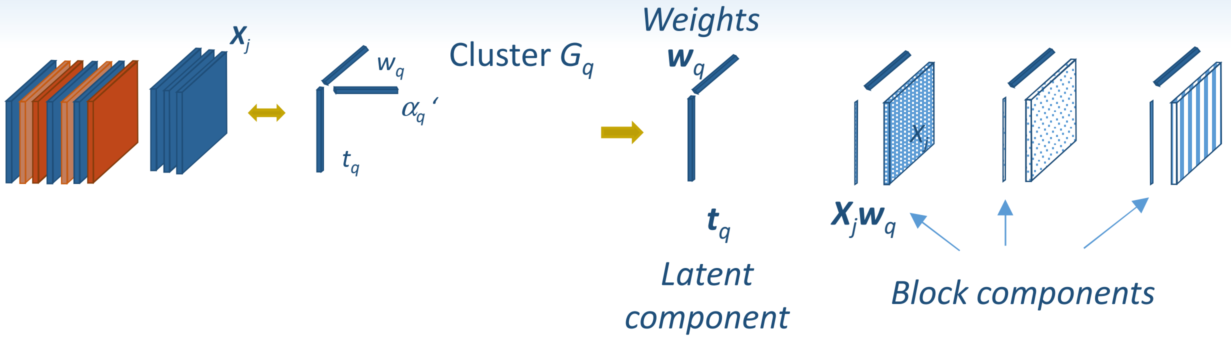
(2) updating the cluster membership δ_{jq} of each variable conditional on the cluster-specific parameters (i.e., $\boldsymbol{t}_q$ and $\boldsymbol{w}_q$)

For the assignment, variable j is assigned to the cluster G_q for which

f_{jq} is minimal: $f_{jq} = \left\|\boldsymbol{X}_j - \alpha_{jq}\left(\boldsymbol{t}_q \boldsymbol{w}_q^{\top}\right)\right\|_F^2$ with $\left\|\boldsymbol{t}_q\right\|$=1, $\left\|\boldsymbol{w}_q\right\|$=1

Determination of α_{jq} by means of a linear regression

Adding a non negativity constraint



Minimize the Loss function: ↔ Maximise the covariance function:

$$f = \sum_{j=1}^{J} \sum_{q=1}^{Q} \delta_{jq} \left\| \boldsymbol{X}_j - \alpha_{jq} \left(\boldsymbol{t}_q \boldsymbol{w}_q^{\intercal} \right) \right\|_F^2$$

$$g = \sum_{j=1}^{J} \sum_{q=1}^{Q} \delta_{jq} cov^2 \left(\boldsymbol{X}_j \boldsymbol{w}_q, \boldsymbol{t}_q \right)$$

with $\left\| \boldsymbol{w}_q \right\|=1$, $\alpha_{jq} \geq 0$ and $\left\| \boldsymbol{t}_q \right\|=1$

Determination of α_{jq} by means of a non negative linear regression

Consumers' segmentation

Emotion ratings associated to coffee aromas

- ✓ 84 undergraduate students from Oniris
- ✓ 12 coffee aromas

Earth, Hay, Cedar; Vanilla, Coriander seeds, Flower Coffee, Apricot, Lemon, Honey, Basmati Rice, Hazelnut, Medicine

- ✓ 15 emotions rated with a 5-point Likert scale

\- *Angry, Unpleasant, Irritated, Disgusted, Upset*

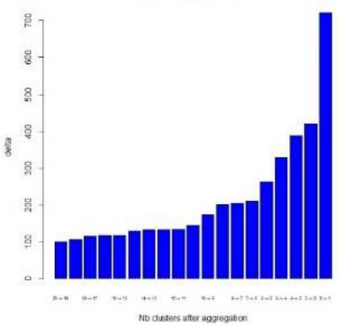
\+ *Happy, Well, Amused*

Energic, Calm, Nostalgic, Free, Excited, Unique, Surprised

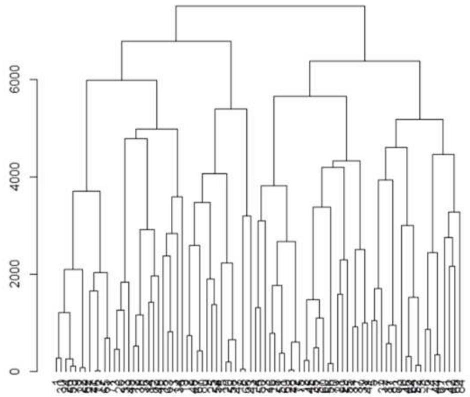
within a cluster

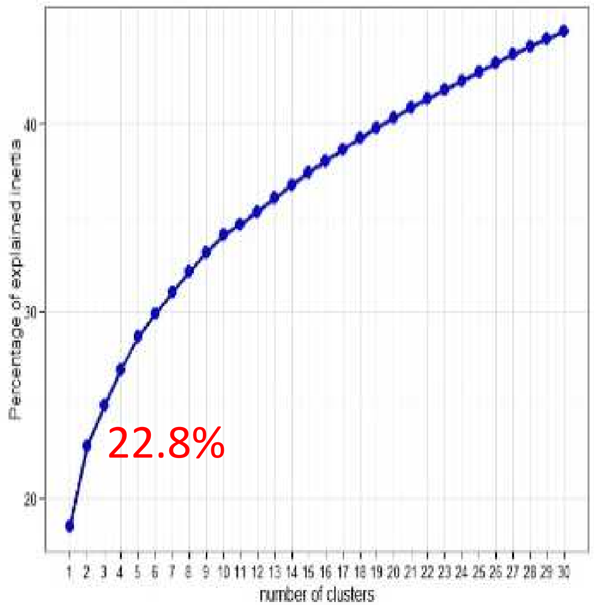
consumer agreement on the product ratings

Consumers' segmentation

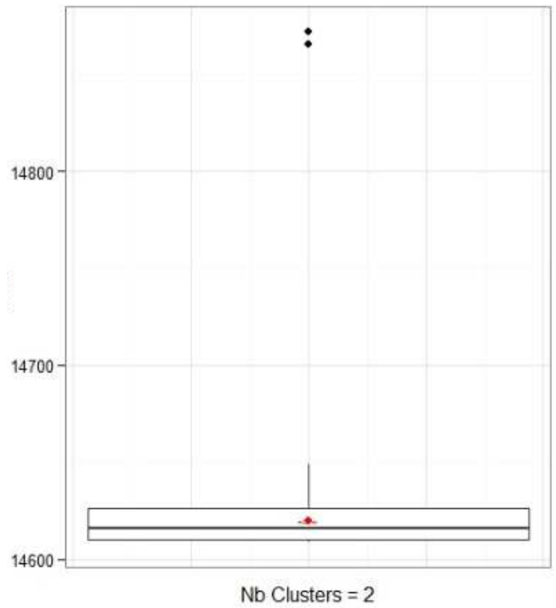


Evolution of the aggregation criterion





% of explained inertia / number of clusters



Sensitivity to initalization

Consumers’ segmentation

Functions in Package ClustVarLV

```
res <- CLV3W_kmeans(coffee,K=2,mode.scale=2, NN=TRUE,init=50,cp.rand=5)
summary(res)
```

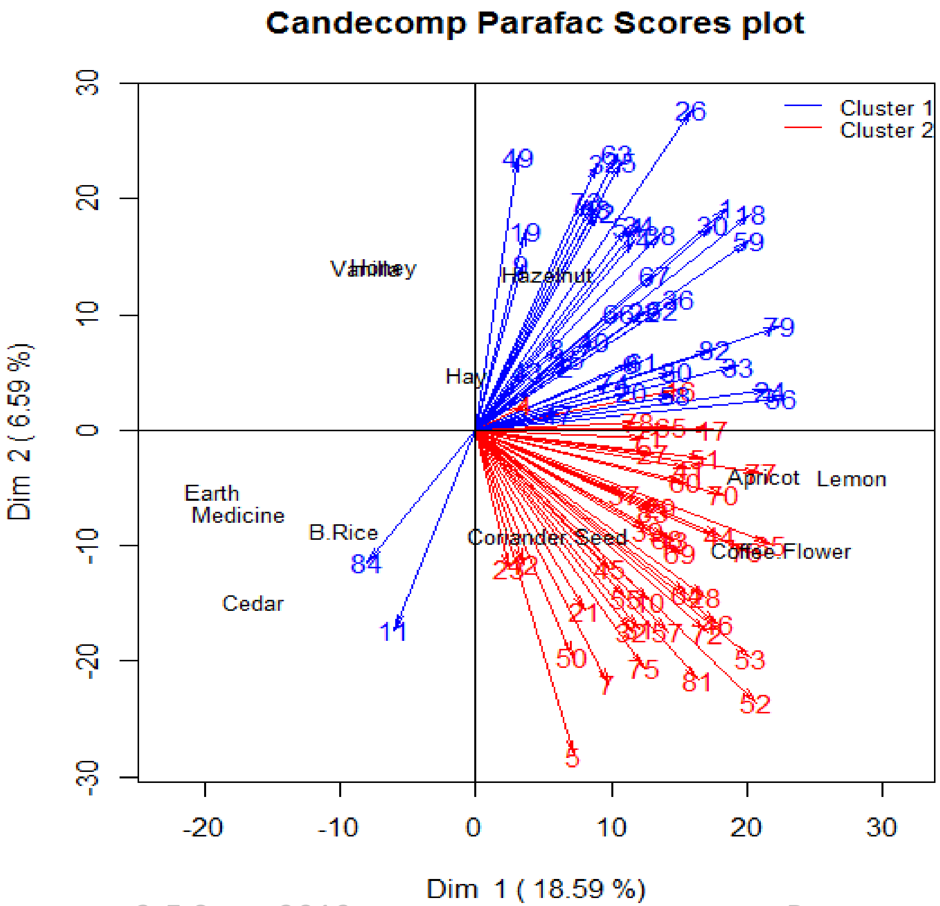
```
$size
clusters
 1  2
42 42

$prop_explained_per_cluster
[1] 21.308 24.316

$prop_explained_total
[1] 22.812

$comp
              Comp1  Comp2
Vanilla      -0.306  0.061
B.Rice       -0.058 -0.218
Lemon         0.524  0.435
Coffee.Flower 0.422  0.278
```

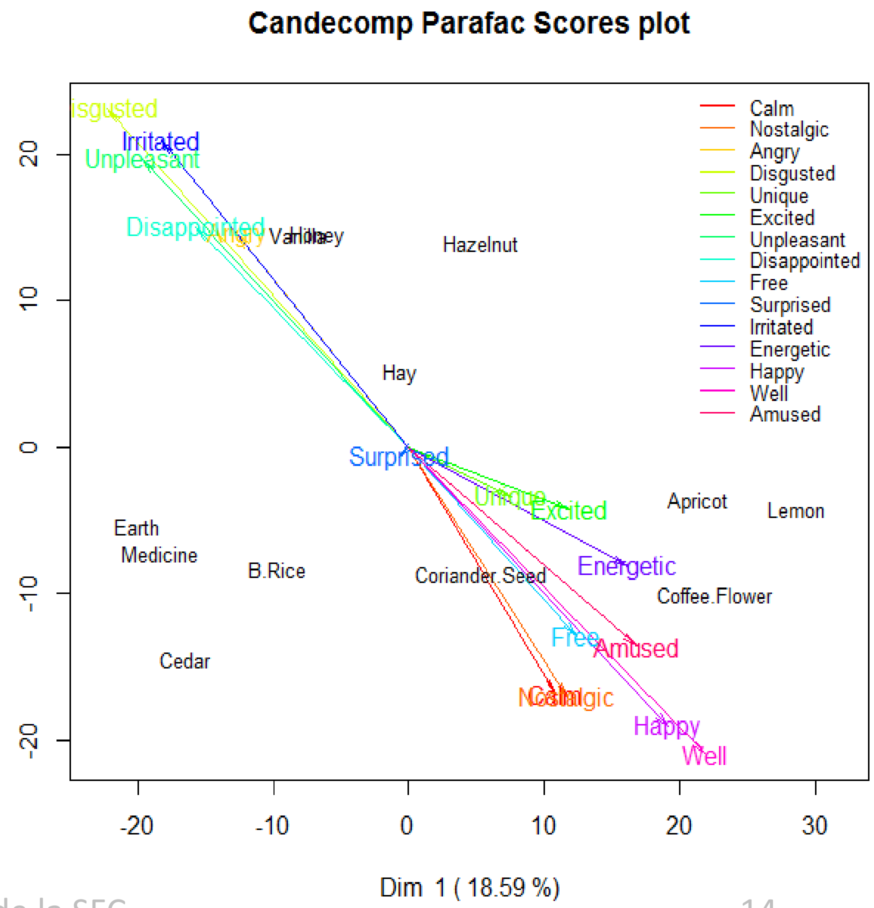
```
$weight
              Comp1  Comp2
Calm          0.215  0.179
Nostalgic     0.188  0.227
Angry        -0.224 -0.222
Disgusted    -0.377 -0.371
Unique        0.094  0.127
Excited       0.138  0.207
Unpleasant   -0.337 -0.327
Disappointed -0.232 -0.281
Free          0.229  0.188
Surprised    -0.008 -0.009

$cormatrix
       Comp.1 Comp.2
Comp.1  1.000  0.586
Comp.2  0.586  1.000
```

```
$groups
$groups$`Cluster 1`
   loading cor in.group cor next.group
1    0.218        0.824          0.504
2    0.078        0.605          0.343
3    0.152        0.600          0.135
6    0.134        0.623          0.408
8    0.086        0.384          0.157
9    0.072        0.528         -0.055
11   0.000       -0.434         -0.115
12   0.128        0.654          0.226
14   0.154        0.692          0.345
15   0.090        0.366          0.160
18   0.236        0.808          0.505
19   0.091        0.376         -0.043
20   0.129        0.536          0.402
22   0.144        0.639          0.382
```

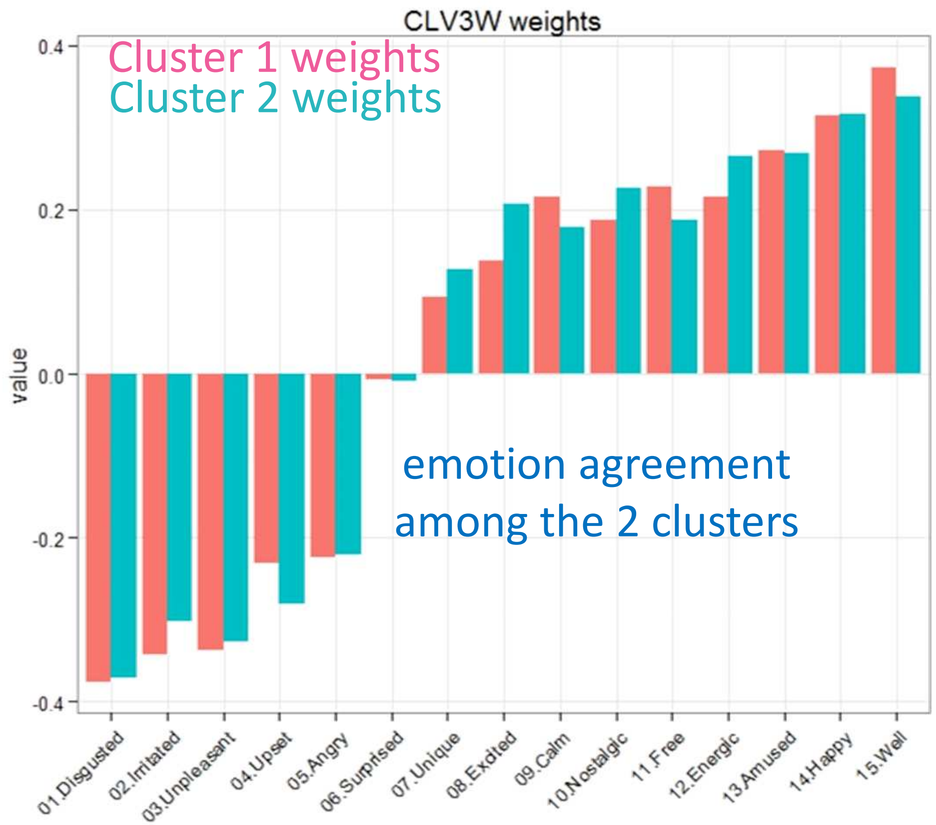
Consumers' segmentation

```
plot_var.clv3w(res.clv3w,K=2,labels=TRUE,cex.lab=0.8,beside=FALSE,mode3=TRUE)
```

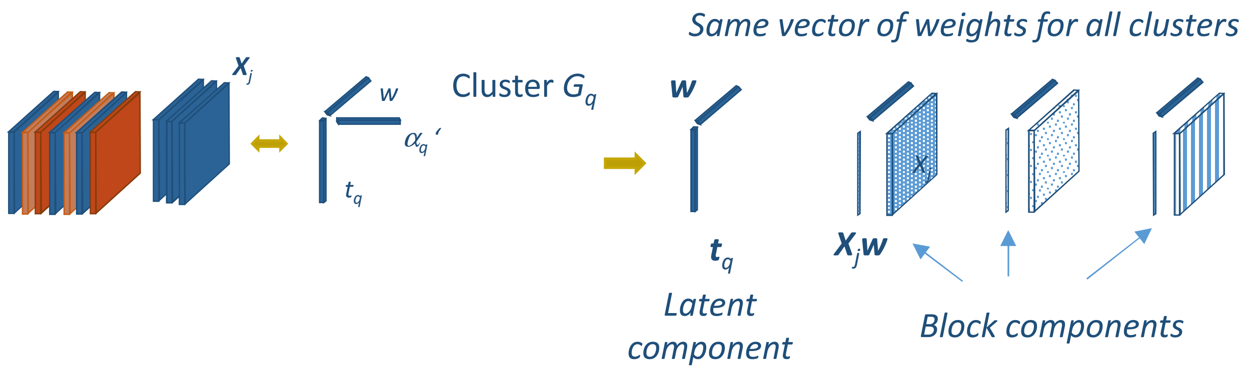




Consumers' segmentation



Imposing a global weighting scheme



Minimize the Loss function: ⟷ Maximise the squared covariance function:

$$f = \sum_{j=1}^{J}\sum_{q=1}^{Q} \delta_{jq} \left\| \boldsymbol{X}_j - \alpha_{jq}\left(\boldsymbol{t}_q \boldsymbol{w}^{\intercal}\right) \right\|_F^2 \qquad g = \sum_{j=1}^{J}\sum_{q=1}^{Q} \delta_{jq} cov^2\left(\boldsymbol{X}_j \boldsymbol{w}, \boldsymbol{t}_q\right)$$

with $\|\boldsymbol{w}\|$=1 and $\|\boldsymbol{t}_q\|$=1

Algorithm

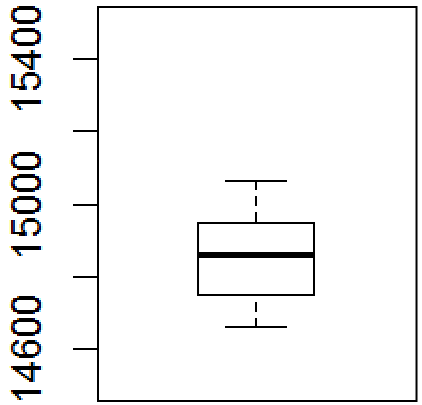
Iterate the following two steps until convergence:

(1) computing parameters conditional on the cluster memberships:

- initializing $\boldsymbol{w}$, the two steps are alternated until convergence:
- Each $\boldsymbol{t}_{\mathrm{q}}$ is updated as the left singular vector corresponding to the largest singular value of the $(I \times \#G_q)$ matrix of the block components associated with the variables of G_q: $\left[\cdots|\boldsymbol{X}_j \boldsymbol{w}|\cdots\right]_{\boldsymbol{X}_j \in G_q}$
- Global $\boldsymbol{w}$ is updated as the left singular vector associated with the largest singular value from of the $(K \times J)$ matrix $\left[\cdots|\boldsymbol{X}_j^{\mathrm{T}} \boldsymbol{t}_q|\cdots\right]_{j=1,\ldots,J}$

(2) updating the cluster membership δ_{jq} of each variable conditional on the cluster-specific $\boldsymbol{t}_q$ and the global weighting scheme $\boldsymbol{w}$

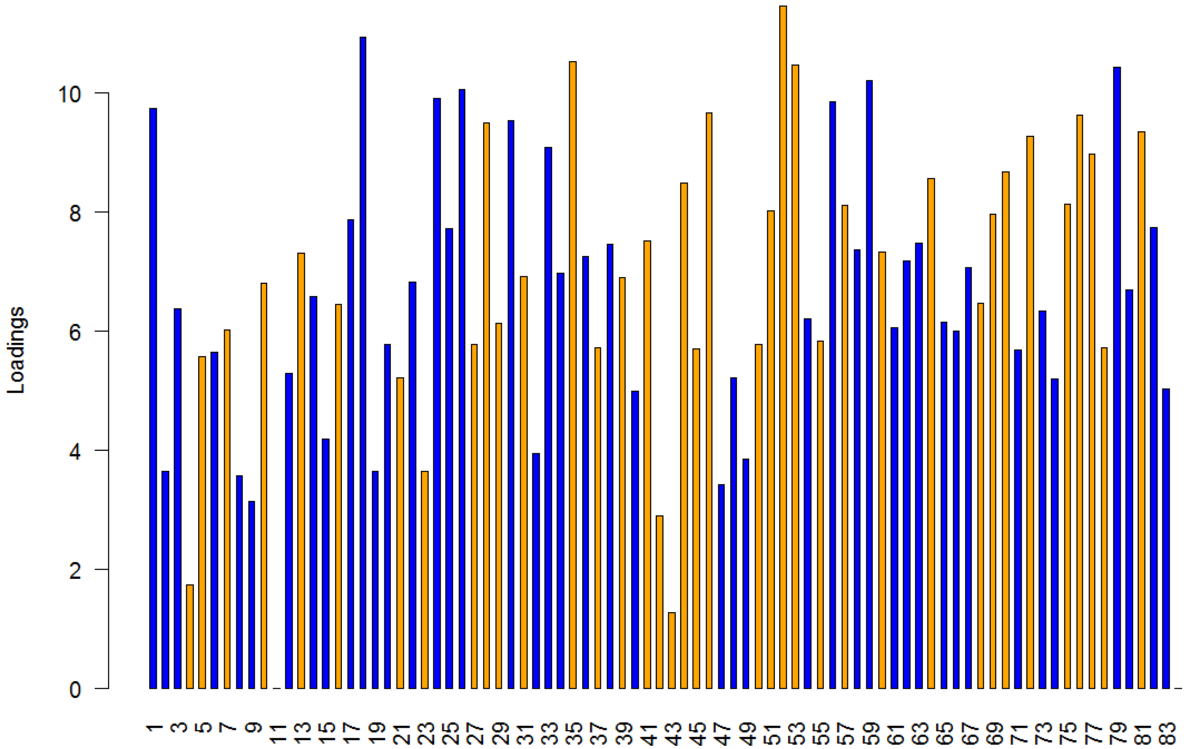
Consumers' segmentation



Sensitivity to initalization

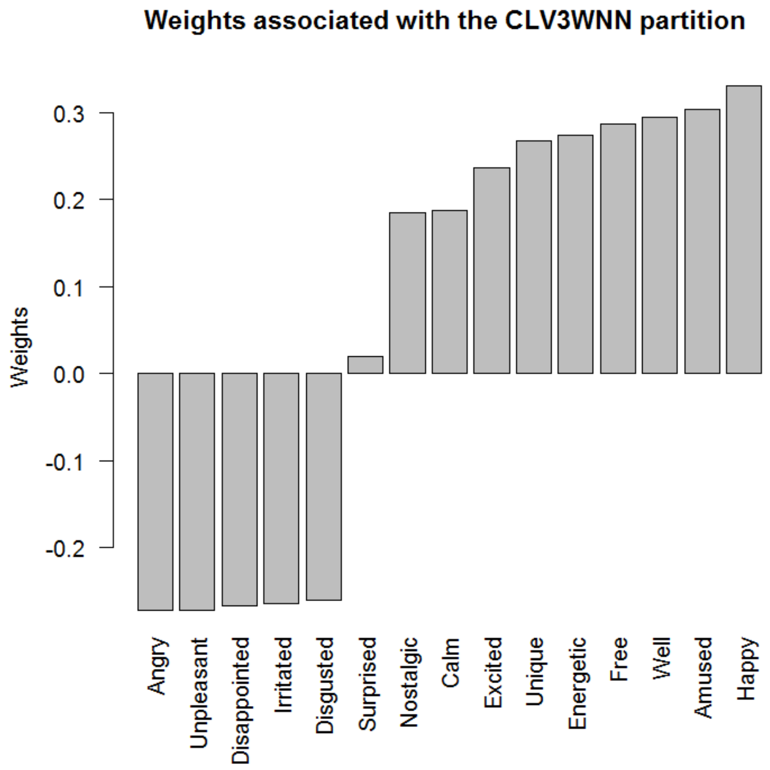
Inertia explained with NN constraint: 22.8%

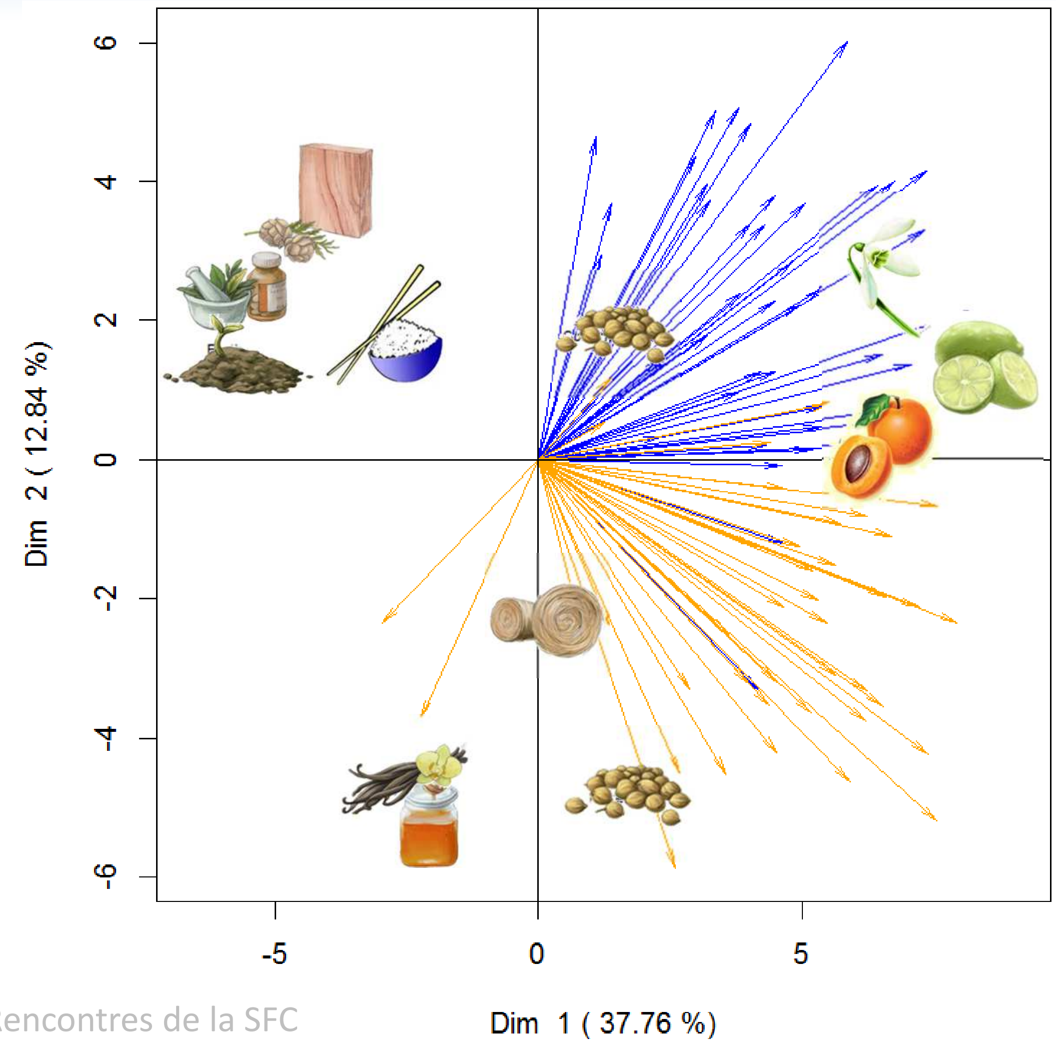
Inertia explained with NN and Global Weights: 22.5%



Size of cluster 1: 44 ; Size of cluster 2: 38
2 consumers set aside

Consumers' segmentation





Conclusion

Cluster analysis of three-way data with CLV3W providing:

- ✓ Latent component associated with the 1st mode samples
- ✓ Loadings reflecting the adequation of each slice in its cluster
- ✓ Weights deriving a block component for each slice

Available in Cran R

Package 'ClustVarLV'

June 28, 2019

Title Clustering of Variables Around Latent Variables

Version 2.0.0

Author Evelyne Vigneau [aut, cre], Mingkun Chen [ctb], Veronique Cariou [aut]

Easier interpretation with constraints on loadings and weights

Perspectives

Choice of the number of clusters:

- ✓ Adaptation of Hartigan criterion $H_Q = (J - Q - 1)\left(\frac{f_Q}{f_{Q+1}} - 1\right)$

Index corresponding to the largest difference between two consecutive ones

Choice between full model or common weights constraint

Further comparisons needed with other strategies such as simultaneous decomposition and clustering models

Minimize the Loss function: $f = \sum_{j=1}^{J}\sum_{q=1}^{Q} \delta_{jq} \left\| \boldsymbol{X}_j - \boldsymbol{t}_q \boldsymbol{w}^{\top} \right\|_F^2$ with $\left\|\boldsymbol{w}_q\right\|$=1

Vichi, M., Rocci, R., et Kiers, H. A. 2007. Simultaneous component and clustering models for three-way data : within and between approaches. *Journal of Classification*, **24**(1), 71–98

Thank you for your attention