External preference segmentation with additional information on consumers.

Evelyne VIGNEAU
Mathilde CHARLES
Mingkun CHEN

External preference mapping and panel segmentation

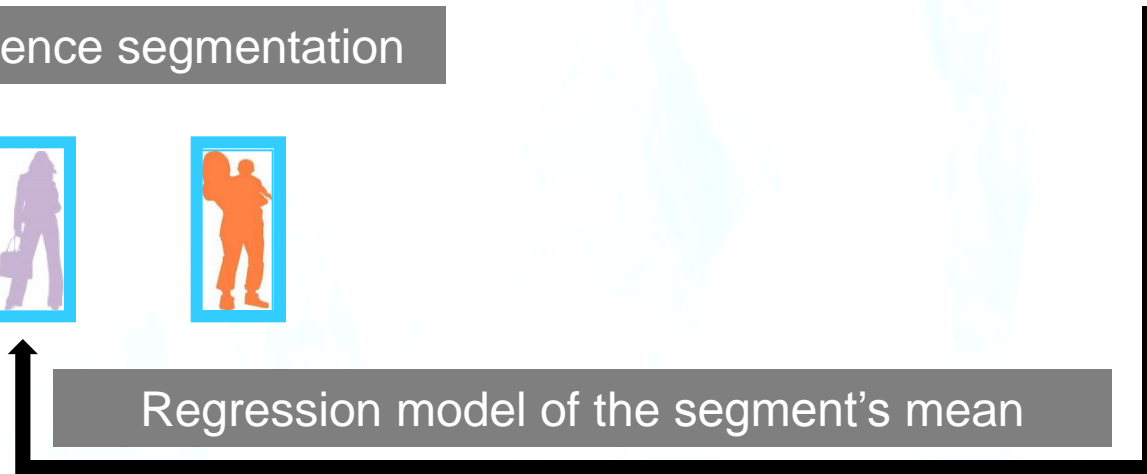Consumers

Sensory experts

products

liking scores

Sensory profiles

Internal preference segmentation

Regression model of the segment's mean

Clustering around Latent Variables (CLV) with co-variables measured on samples

objective : merge together consumers who have similar drivers of preference

Y
Liking scores

segmentation

Modelisation

X
Sensory profiles

Simultaneously, define :

- groups of consumers and

- in each group, the linear combination of the sensory attributes which explains at much as possible the liking scores.

➤*Alternative strategy :* Fuzzy C-means (FCM) on the residuals of a regression on sensory or design variables (Naes,Kubberod,Sivertsen, FQP, 2001 - Johansen, Herlseth, Naes, FQP, 2010, Menichelli, Olsen, Meyer, Naes, FQP, 2012)……………………………………. *see next presentation.*

CLV with co-variables measured on samples

c_1

Y
c_2
X

c_3
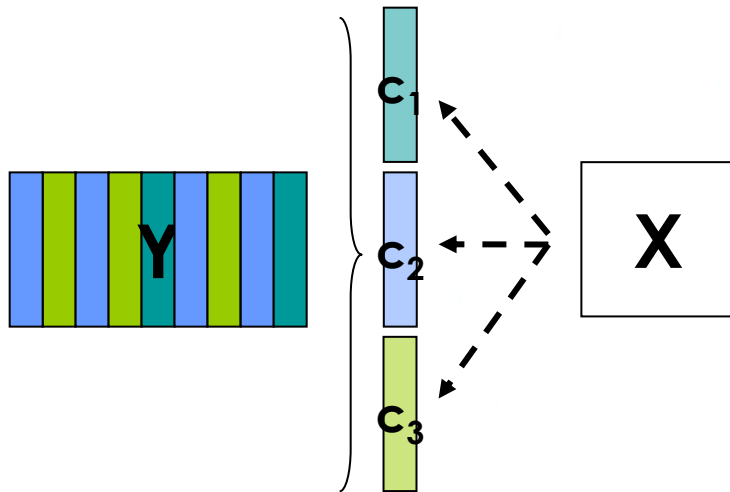
maximize

$$S = \sum_{k=1}^{K} \sum_{j=1}^{p} \delta_{kj}\, cov(\, y_j, t_k \,)$$

with $\quad t_k = X\, a_k \qquad a'_k\, a_k = 1$

y_j : scores of likings for consumer j ($j=1,...,p$)
c_k : latent variable in group k ($k=1,...,K$)
δ_{kj}=1 if x_j belongs to G_k, =0 otherwise

Solution

K segments

in each segment k,

t_k is the first PLS regression component of $\overline{y}_k$ on X

Specificity of « statistical » clustering method in hedonic studies (crisp algorithm)

Each consumer belongs to one, and only one, group

BUT not all consumers are well represented by their group's mean (« non typical » or « spurious » likings)

OR some consumers are almost between two groups (degree of neighborhood between segments)

possible difficulties in the interpretation of each segment
… specifically by means of external consumers attributes
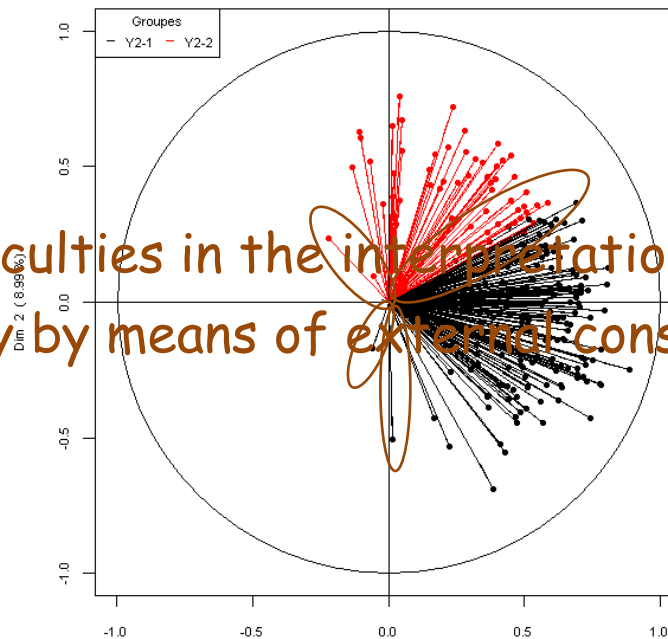
Taking account of additional information on consumers

Working at the individual level :

➢ and sum up the consumers in a segment.

➢ after discarding those consumers with low cluster contribution and/or high between-cluster position
- R2 with the own cluster, R2 with the next nearest cluster (~silhouette indices)
- cluster membership's values from fuzzy clustering.
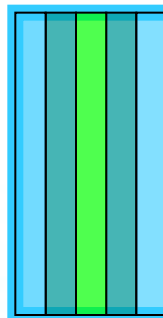
working on the segment level

➢ On the basis of latent components = **central tendencies** in the hedonic space associated with the consumers attributes and to the products attributes .
⇨ **L-CLV approach**

Clustering around Latent Variables (CLV) with co-variables measured on products **and additional information on consumers** L-shaped data

Information collected by means of a questionnaire on consumers

Z'
$(M \times p)$

Liking scores

Y
$(n \times p)$

Sensory attributes

X
$(n \times Q)$

L-CLV

maximize $\qquad \tilde{S}_X^Z = \sum_{k=1}^{K} \text{cov}(c_k, t_k)$

with

$$c_k = P_k \, u_k \qquad u_k' \, u_k = 1$$

$$P_k = Y_k \, Z_k$$

$P_k \leftrightarrow$ interaction between Y and Z

$$t_k = X \, a_k \qquad a_k' \, a_k = 1$$

$$\tilde{S}_X^Z = \frac{1}{n} \sum_{k=1}^{K} u_k' \, Z_k' \, Y_k' \, X \, a_k$$
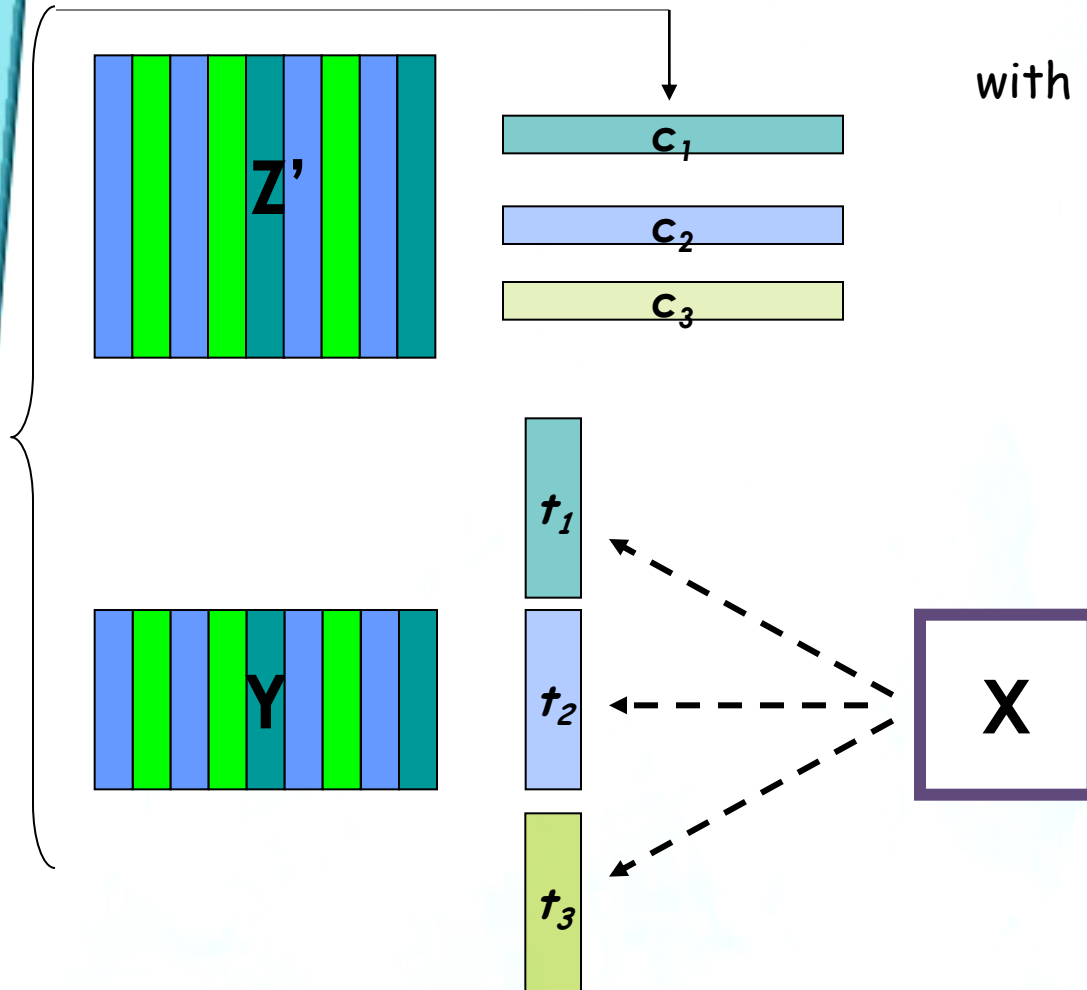
Apple Case study (COSI-VEG 2010-2013)
31 apple cultivars locally produced (Loire Valley, France)

Consumer questionnaire

- Frequency of Consumption,
- Apple cultivars known
- Important sensory attributes,
- Modalities of consumption (peeled/during meal/ …)
- Purchase criteria
- Supply location
- 14 questions « eater style » (likert scale)
- 7 questions « opinion on apple » (likert scale)
- Age, gender, professional activity….

Hedonic test

224 regular apple consumers

-Liking score on a 9-points

-5 sessions during 3 weeks

Sensory descriptive analysis

15 assessors, 15 attributes

Crunchy	A_Pineapple/Banana
Juicy	A_Sweet/Rose
Fondant	A_Woody/Earthy
	A_Rustic
Sweet	A_Lemon
Acid	A_White flowers
	A_Ripe fruit
Odour intensity	A_Green
Aroma intensity	

Apple Case study (COSI-VEG 2010-2013)
31 apple cultivars locally produced (Loire Valley, France)

Consumers attributes
categorical : dummy variables, globally scaled
numerical : centered and unit scaled

224

Z'

65

Liking scores
centered and unit scaled

224

Y

31

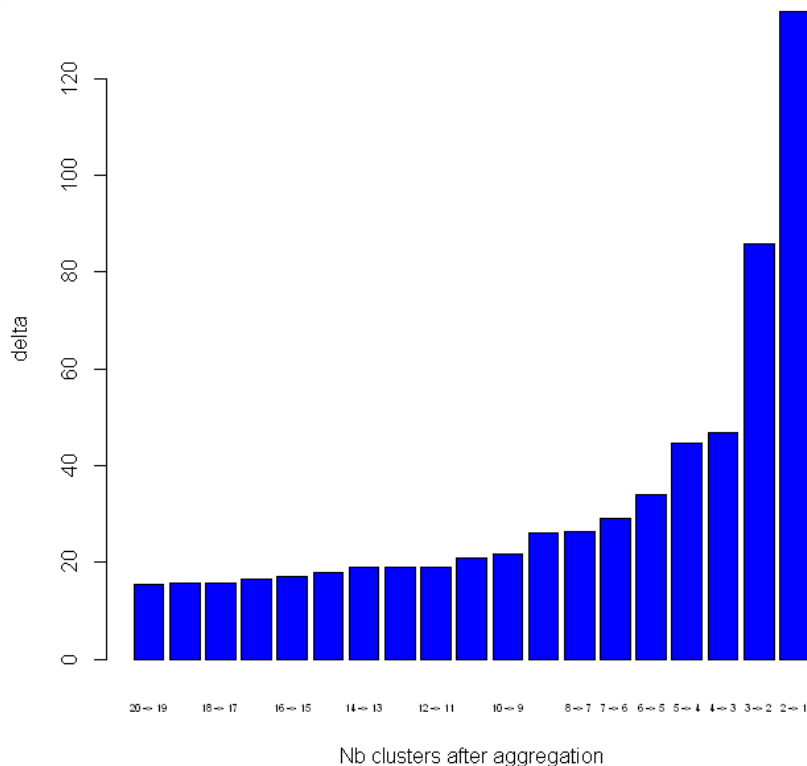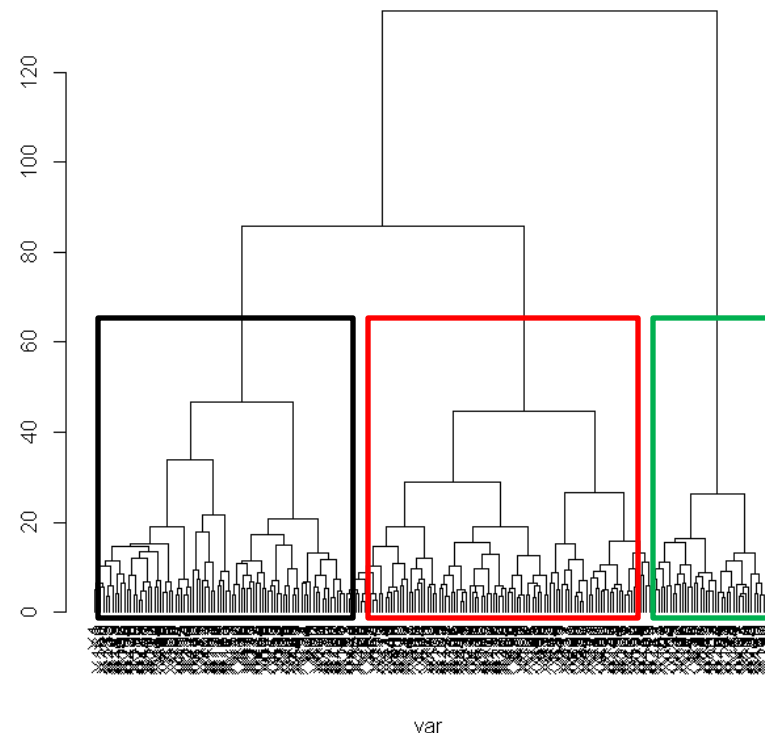*Sensory attributes*
centered and unit scaled

30

X | X²

31

L-CLV : choice of the number of segments
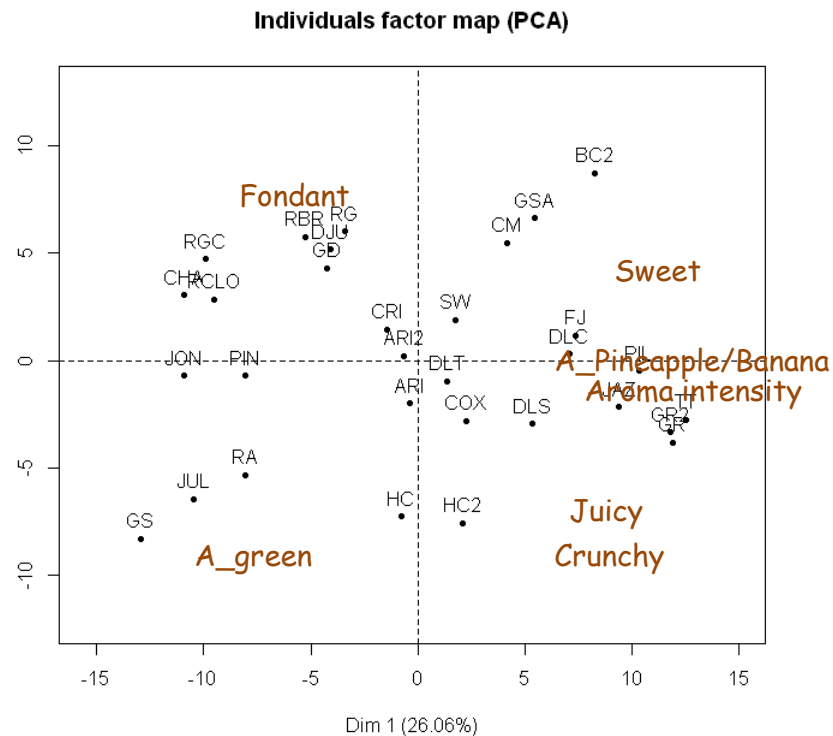
Evolution of the aggregation criterium

CLV Dendrogram

Segment L3-1   82 consumers   (37%)
Segment L3-2   96 consumers   (43%)
Segment L3-3   46 consumers   (20%)
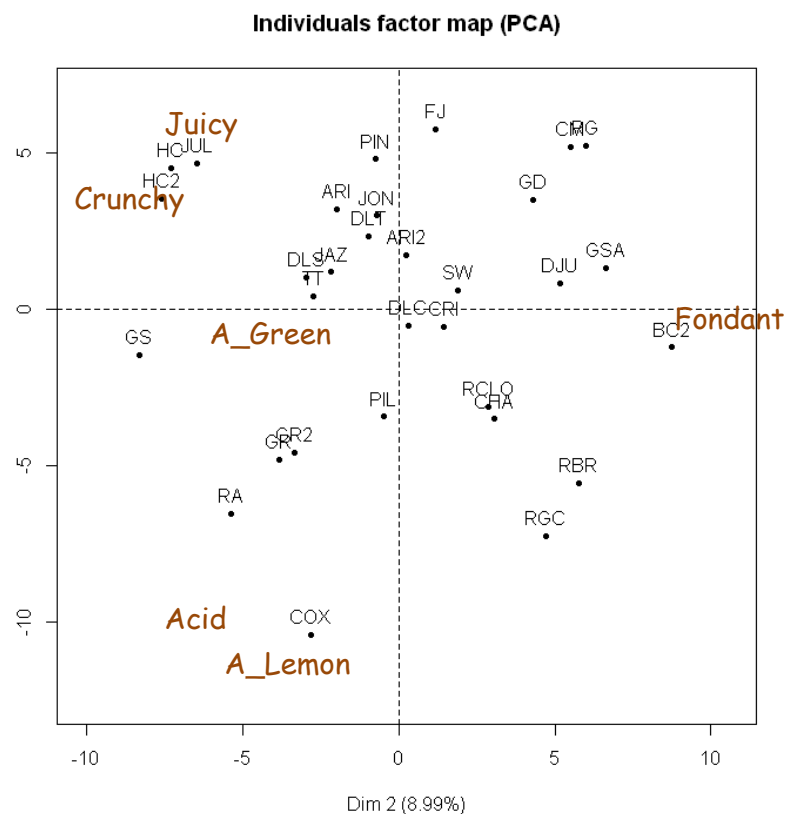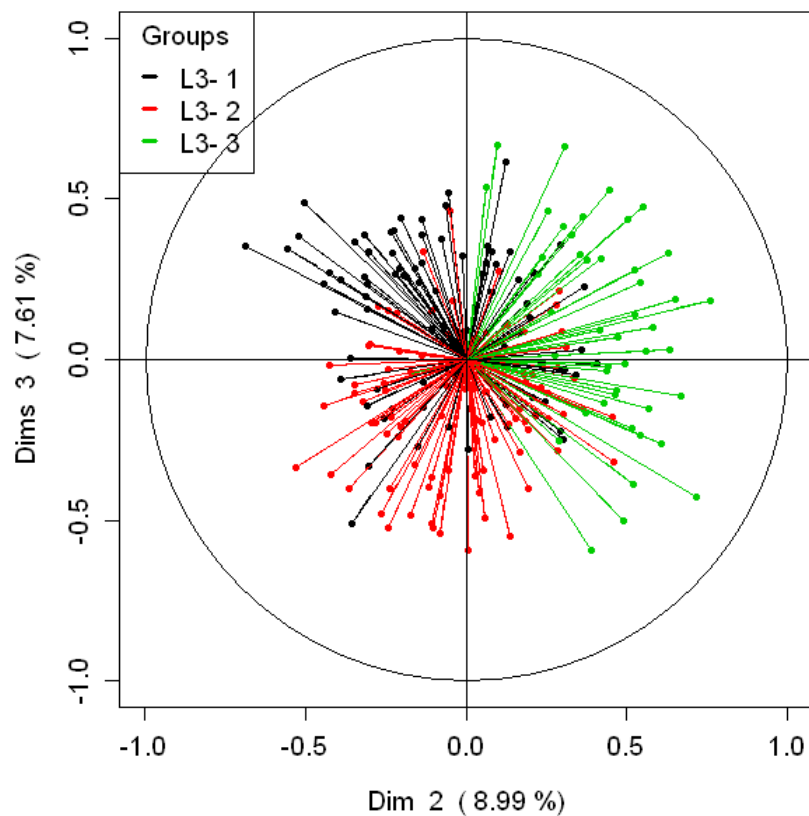
L-CLV :

Consumers segments represented on the internal preference mapping
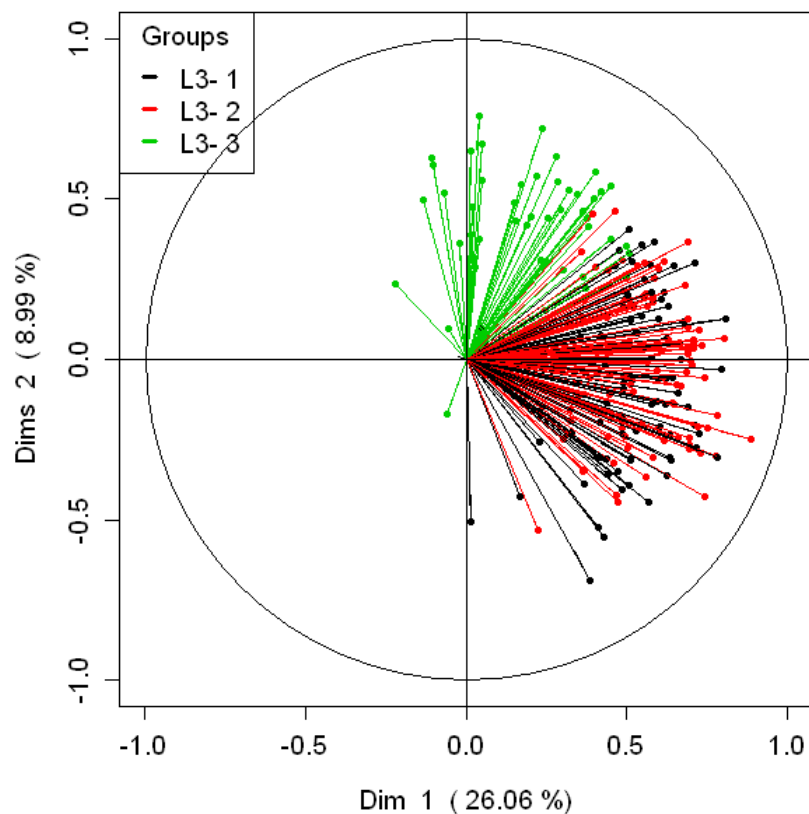
L-CLV:

Consumers segments represented on the internal preference mapping

L-CLV :Loadings associated to the sensory attributes (a_k)

|	t_1	t_2	t_3		t_1	t_2	t_3
Crunchy	0.308	0.202	-0.214	Crunchy _sq	-0.138	-0.098	-0.102
Juicy	0.307	0.154	-0.073	Juicy _sq	-0.150	-0.083	-0.047
Fondant	-0.222	-0.147	0.297	Fondant _sq	-0.122	-0.134	-0.020
Sweet	0.330	0.376	0.278	Sweet _sq	0.025	-0.033	-0.154
Acid	-0.068	0.060	-0.298	Acid _sq	-0.120	-0.077	-0.237
Odour Int	0.203	0.189	0.038	Odour Int _sq	-0.007	-0.022	-0.190
Aroma Int	0.282	0.344	0.066	Aroma Int _sq	0.091	0.040	0.018
A_Pineapple/Banana	0.320	0.390	0.067	A_Pineapple/Banana _sq	0.166	0.205	-0.179
A_Sweet/Rose	0.322	0.322	0.292	A_Sweet/Rose _sq	0.125	0.143	0.097
A_Woody/Earthy	-0.234	-0.270	-0.138	A_Woody/Earthy _sq	-0.079	-0.089	-0.078
A_Rustic	-0.232	-0.293	0.055	A_Rustic _sq	-0.005	-0.059	0.082
A_Lemon	-0.071	0.077	-0.242	A_Lemon _sq	-0.212	-0.116	-0.207
A_White flowers	0.049	-0.033	-0.095	A_White flowers _sq	0.038	0.009	-0.044
A_Ripe fruit	0.020	0.090	0.258	A_Ripe fruit _sq	0.004	-0.021	0.174
A_Green	-0.147	-0.202	-0.350	A_Green _sq	-0.089	-0.113	-0.248

L-CLV : Loadings associated to the sensory attributes (a_k)
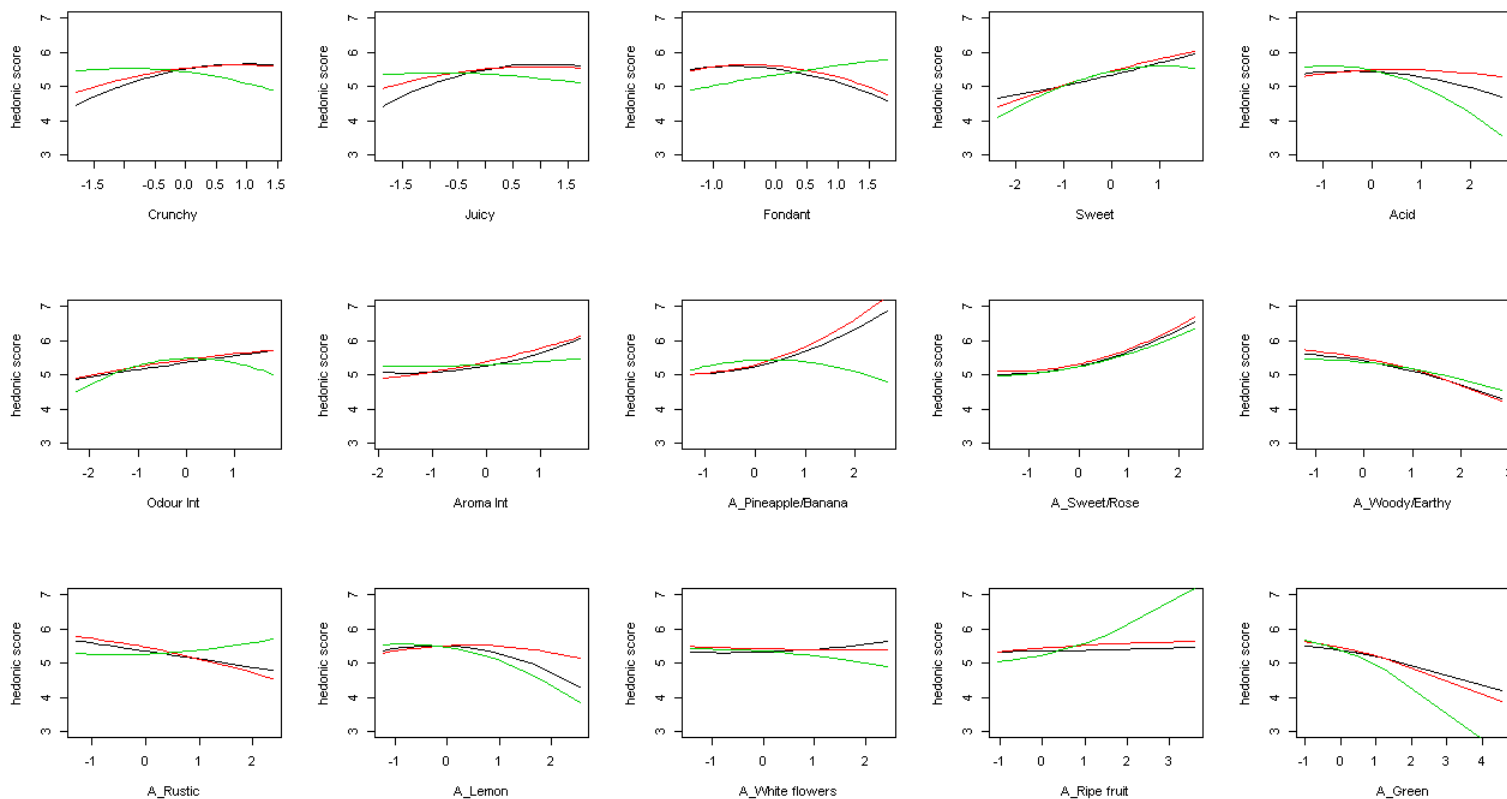
- Segments L3-1 and L3-2 : similar sensory keydrivers (texture crunchy and juicy, sweet flavor, « pineapple/banana » aroma) … slight differences for acidity, « lemon » aroma.

- Segment L3-3 : do not reject fondant texture, appreciate more « rustic » and « ripe-fruit » aroma than « pineapple/banana » aroma, clearly reject acidity, « green » aroma.

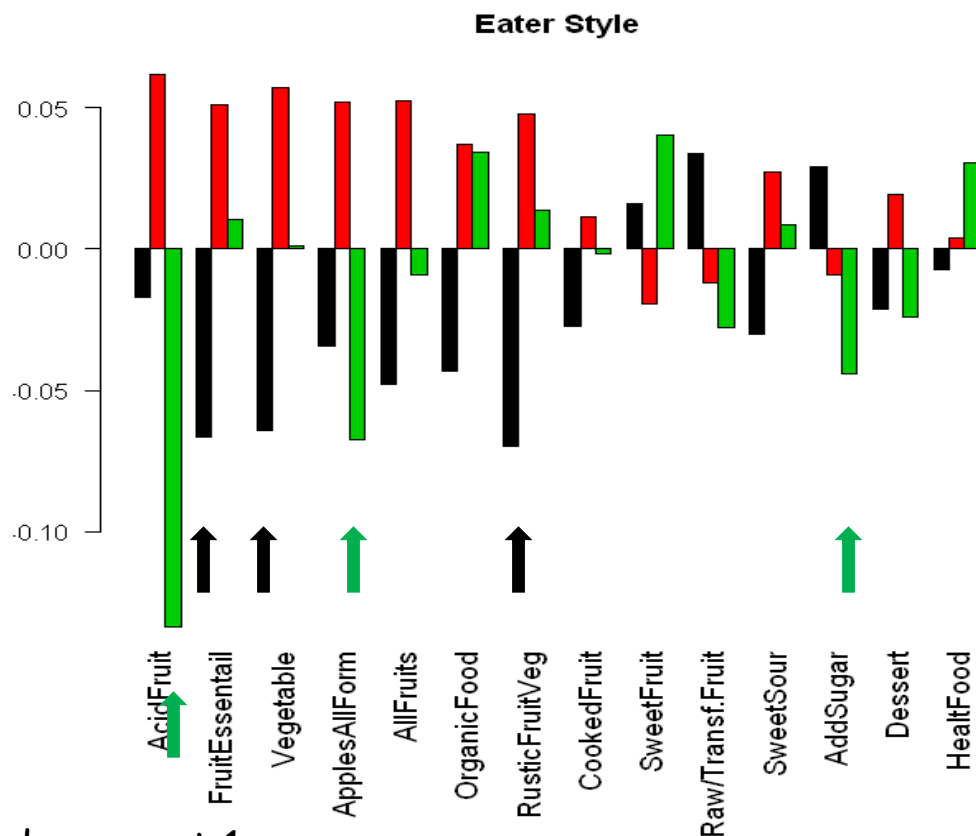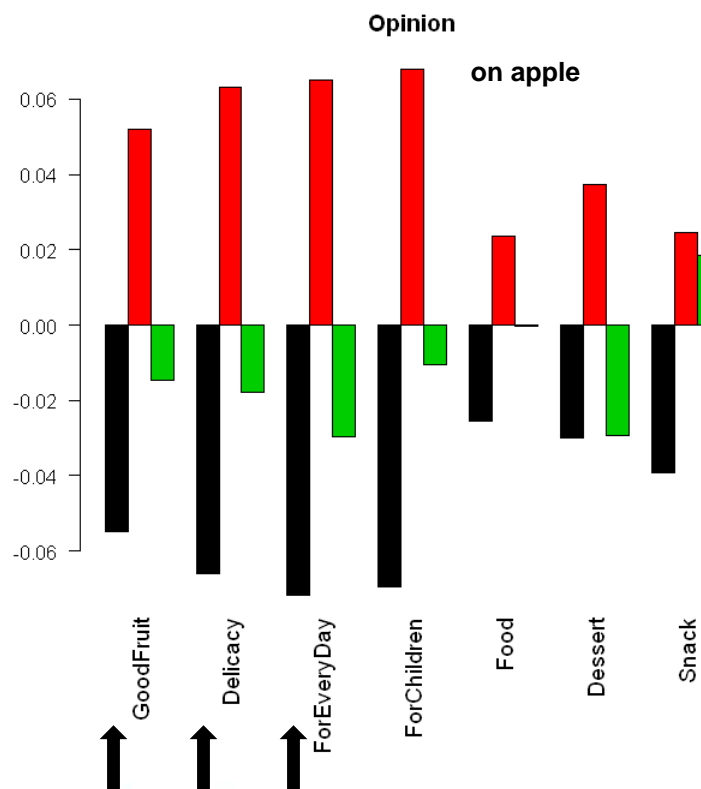# L-CLV : Loadings associated to the consumers attributes (u_k)
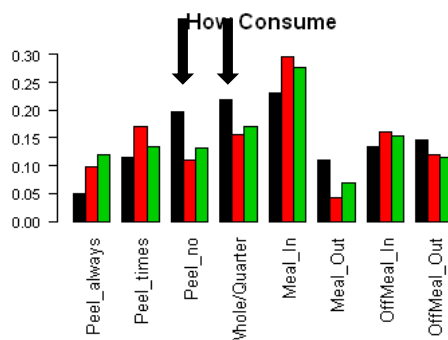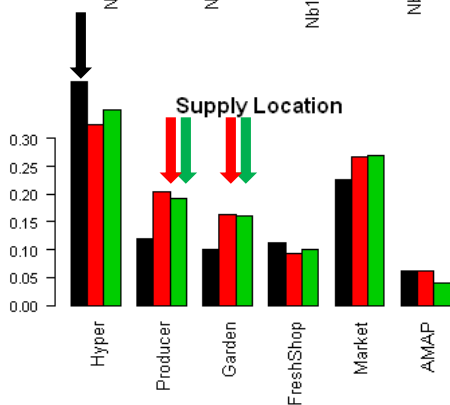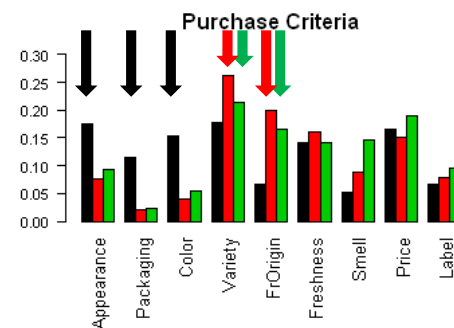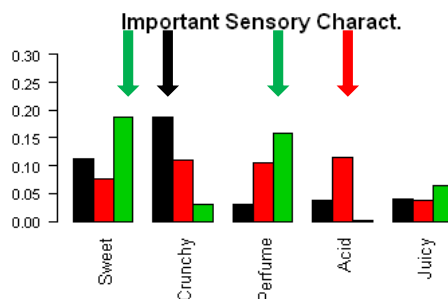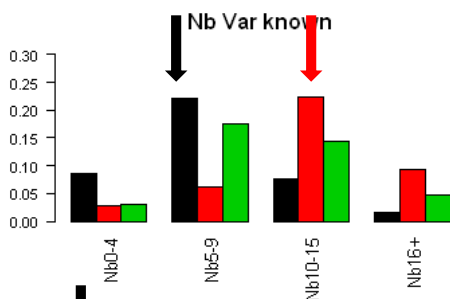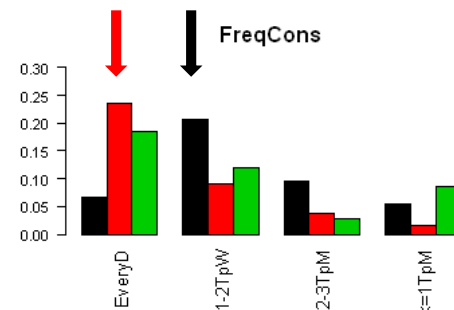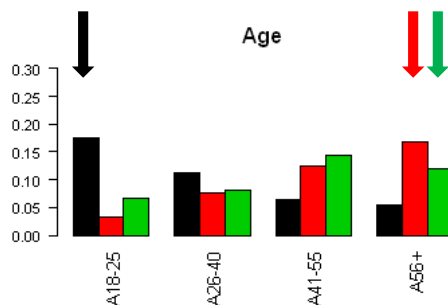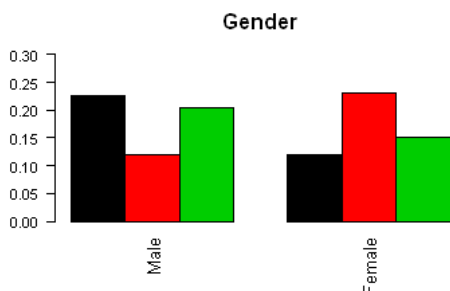
the most discriminant modalities associated to the consumers attributes are highlighted.
(chi-square test/v.test)

L-CLV : Loadings (u_k) associated to agreement measurements
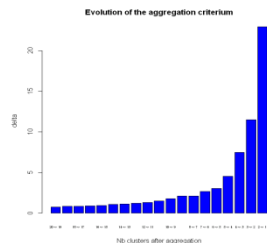
(likert-type scale, centered data)

• Clear opposition between segment 2 and segment 1
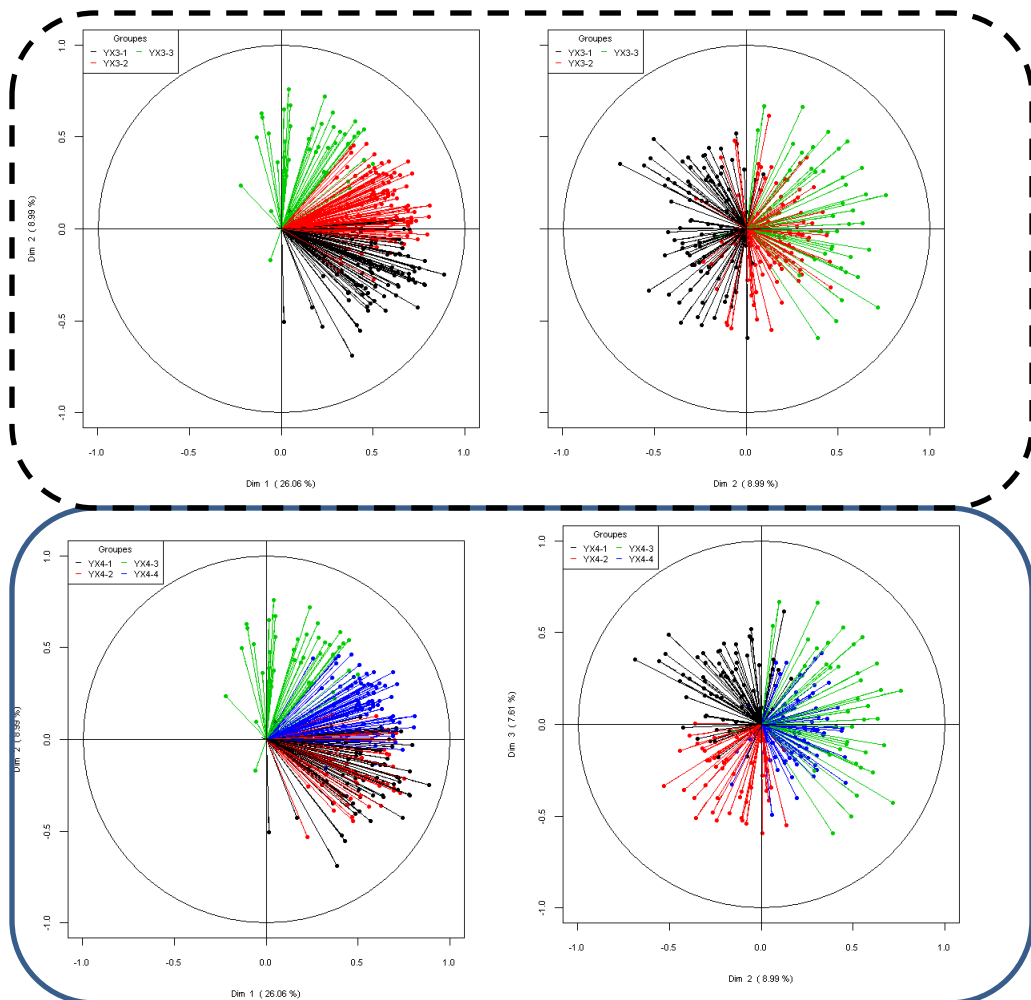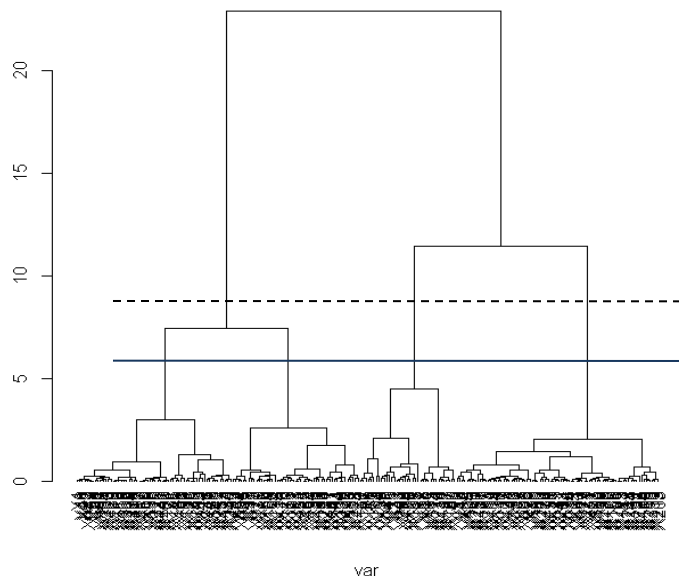Segment 1 : apple is not a « good fruit », not for every day, don't eat at lot of fruits and vegetables

• Segment 3 : don't like acid fruits, prefer sweet fruits but don't add sugar, are very not fond of apples

Comparison with CLV on Y with external X-block
… without Z-block

Comparison regarding the sensory key-drivers

L- CLV on Y with external X-block and Z-block
Partition in 3 groups

	n	Main sensory drivers
L3-1	82	+ Sweetness + Juicy + Crunchy + A_Pineapple/banana + A_Sweet/rose
L3-2	96	++ A_Pineapple/banana ++ Sweetness + Aroma intensity 0 A_Lemon
L3-3	46	-- A_green -- Acidity - A_lemon + Fondant 0 A_Pineapple/banana

Correlation coeff. between the latent variables t_k
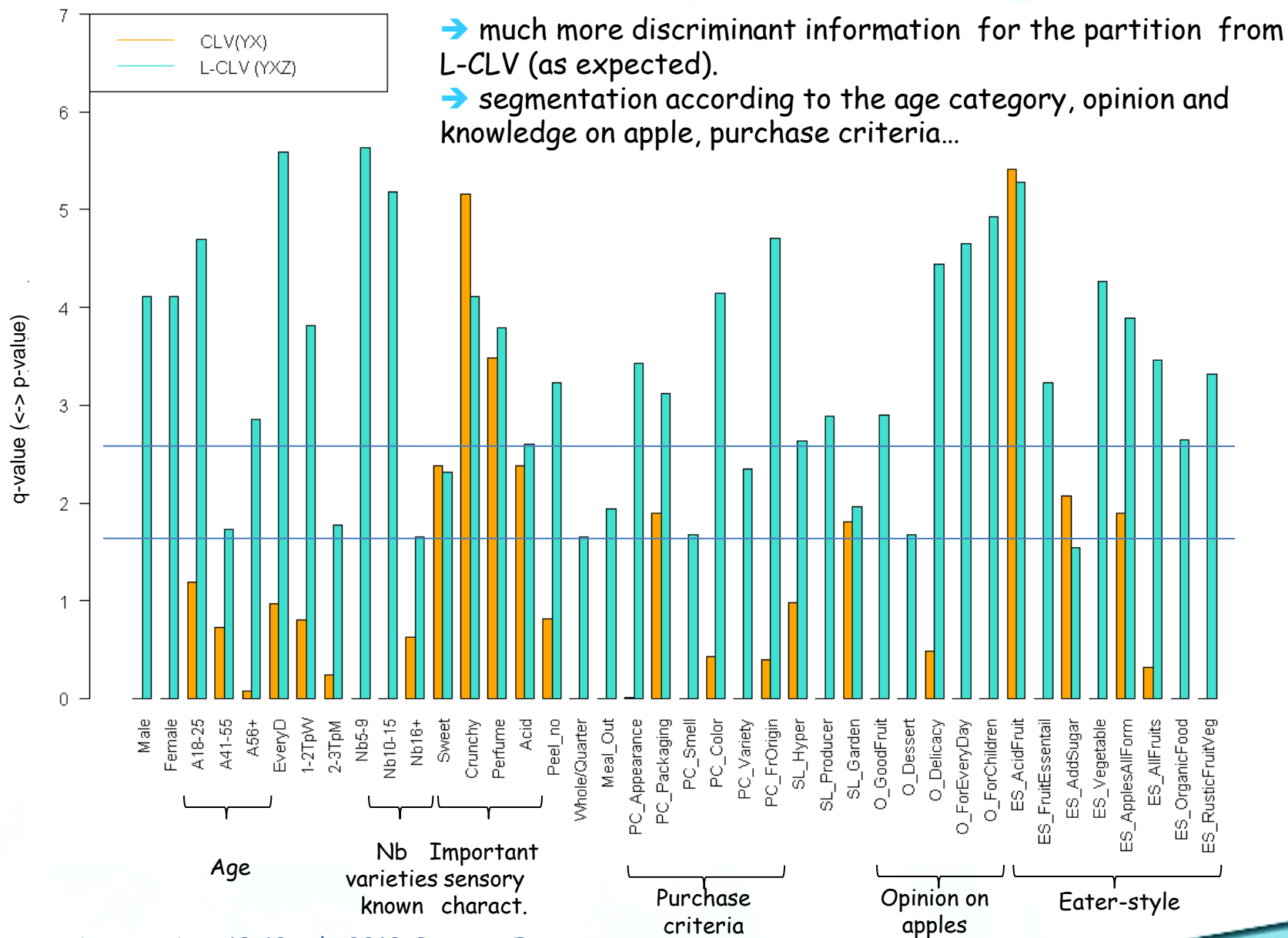
0.99

0.92

0.95

0.92

0.99

CLV on Y with external X-block
(without Z-block)
Partition in 4 groups

	n	Main sensory drivers
YX4-1	67	++ Juicy ++ Crunchy + Sweetness + A_Pineapple/banana
YX4-2	45	++A_Pineapple/banana - A_Rustic + Aroma intensity + A_Lemon
YX4-3	40	-- A_green -- Acidity - A_lemon + Fondant 0 A_Pineapple/banana
YX4-4	72	++ Sweetness ++ A_Sweet/rose + A_Pineapple/banana - A_green

Comparison regarding the consumers attributes

➔ much more discriminant information for the partition from L-CLV (as expected).
➔ segmentation according to the age category, opinion and knowledge on apple, purchase criteria…

Conclusion

In external preference mapping/segmentation,
by taking into account only the external information on products,
no relevant information is necessarily gained with
the subsequent use of the consumers attributes.

Taking into account simultaneously external information
on products attributes and consumers attributes
makes it possible to reveal a segmentation of consumers
interpretable in terms of sociological and behavioural parameters
in relation with the sensory key-drivers.

L-CLV method is suitable for this purpose

(marketing research) .

Thank you for your attention !

Comment : Clustering in hedonic studies

STATISTICAL PACKAGE CLUSTERING MAY NOT BE BEST FOR GROUPING CONSUMERS TO UNDERSTAND THEIR MOST LIKED PRODUCTS

RENOO YENKET[1], EDGAR CHAMBERS IV[1,3] and DALLAS E. JOHNSON[2]

Comment : Clustering in hedonic studies

STATISTICAL PACKAGE CLUSTERING MAY NOT BE BEST FOR GROUPING CONSUMERS TO UNDERSTAND THEIR MOST LIKED PRODUCTS

RENOO YENKET[1], EDGAR CHAMBERS IV[1,3] and DALLAS E. JOHNSON[2]

Each cons

BUT not al their group's mean
(« non typical » or « spurious » likings) ......low « cluster contribution »

OR some consumers are almost between two groups (degree of neighborhood between segments)
...... high « between-cluster position »

Cluster contribution

$$R^2_{own,j} = \max\left(0, r(y_{j \in Gk}, c_k)^2\right)$$

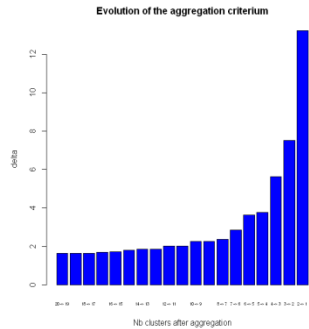$$R^2_{nearest,j} = \max\left(0, \max_{\ell \neq k} r(y_{j \in Gk}, c_\ell)\right)^2$$

between-cluster position

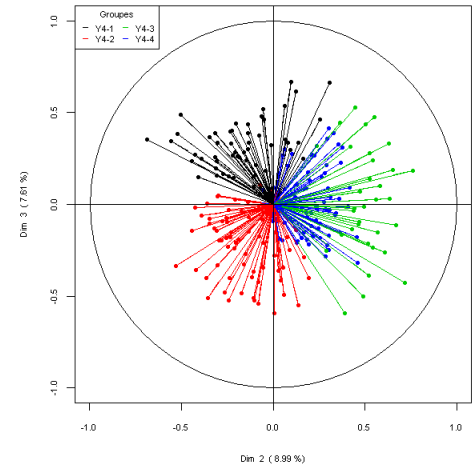$$1 - R^2 ratio_j = \frac{1 - R^2_{own,j}}{1 - R^2_{nearest,j}}$$

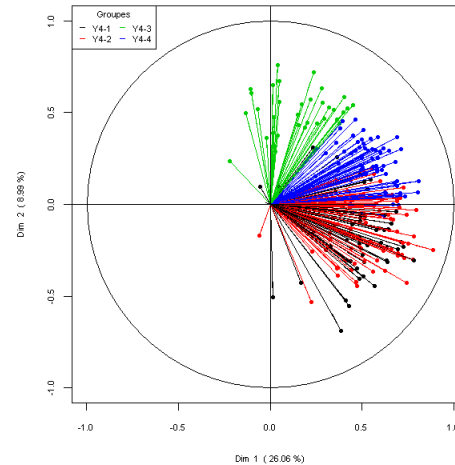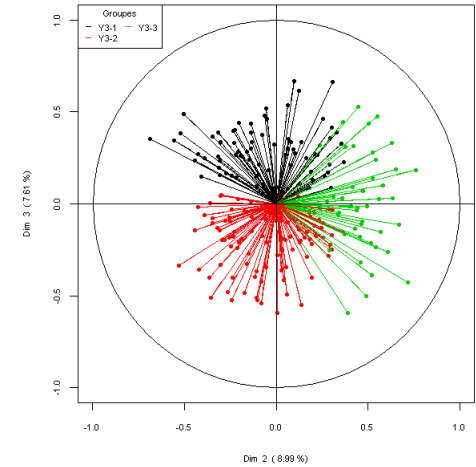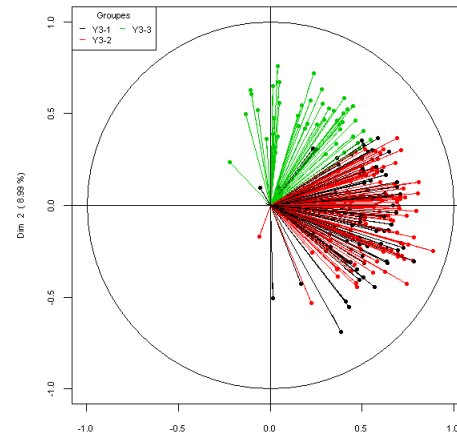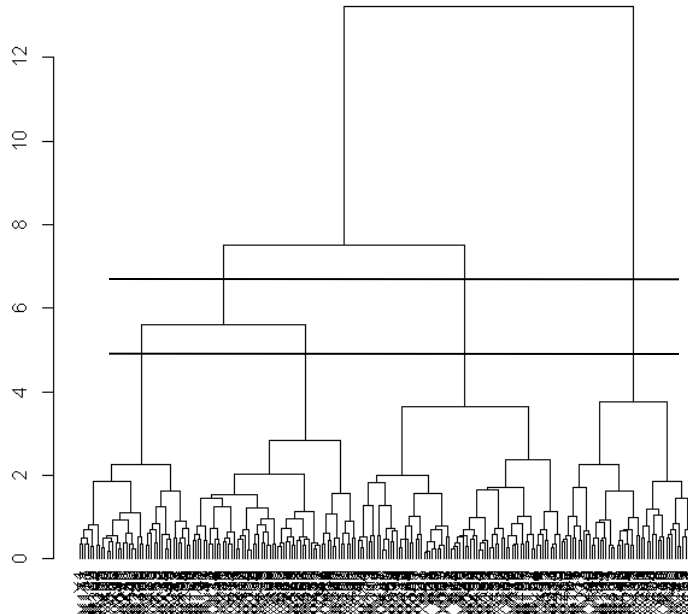# Alternative approaches

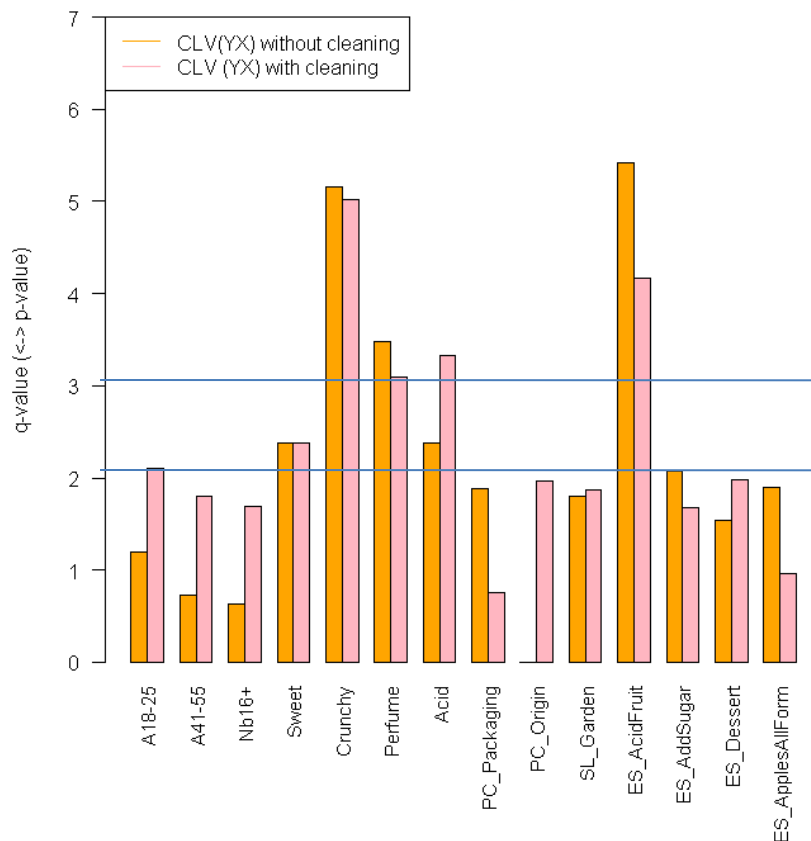CLV with Y   without external X-block and Z-block

Interpretation of the segments from CLV on Y with external X-block, without or with cleaning

cleaning = discard « spurious » and/ or « between-clusters » consumers, select the consumers near their group's center.

Here selection criterion 1-R² ratio <0.90

$$R^2_{own,j} = \max\left(0, r(y_{j\in Gk}, c_k)^2\right) \qquad R^2_{nearest,j} = \max\left(0, \max_{\ell \neq k} r(y_{j\in Gk}, c_\ell)\right)^2 \qquad 1 - R^2 ratio_j = \frac{1 - R^2_{own,j}}{1 - R^2_{nearest,j}}$$

⇨ 115 selected / 109 excluded

No clear improvement by the cleaning process
The segments can't be better been explained by the consumers attributes,
except « Age category » and « Origin » for purchase criteria which become more discriminant.