An alternative to preference mapping: Regression trees and random forests
Application to preferences for ciders

Léa Koenig[1], Adèle Lepetit[1], Ronan Symoneaux[2], Philippe Courcoux[1]

[1] ONIRIS, Sensometrics and Chemiometrics Lab. BP43222 Nantes, France

E-mail : lea.koenig@oniris-nantes.fr ; adele.lepetit@oniris-nantes.fr ; philippe.courcoux@oniris- nantes.fr

[2] Groupe ESA, UR GRAPPE – Angers – France

E-Mail : r.symoneaux@groupe-esa.com

Introduction

The aim of this study is to evaluate the use of multivariate regression trees and random forests for modelling preference data as an alternative method to preference mapping. Regression trees and Random forests are recursive methods which allow explaining a response from a set of explanatory variables. In the context of preference study, these techniques are evaluated for the identification of drivers of liking and the prediction of preference.

Random forests

Random Forest methods (Breiman, 2001), are used to determine the importance of the different explanatory variables in order to select only the most important ones. A succession of regression trees are built using a double randomization (bootstrapping on products and randomized selection of variables). This selection of attributes is used for eliciting the main drivers of liking.

Multivariate regression trees

Multivariate regression trees (De'ath, 2002) are used to explain and predict multivariate response as a function of several explanatory variables. It produces a decision tree allowing the modelization of the multivariate response as a set of decision rules. At each node, the variable and the cutoff point are chosen as the most discriminant for all consumers (minimization of the intra-group inertia).

Application to ciders

Data came from a CASDAR project on the preference for ciders funded by the French Ministry of Agriculture. This project involved ESA Angers, Adria of Normandie, INRA and IFPC. A sensory profile was performed on 19 ciders (different origins) by a panel of 10 trained subjects on 47 attributes (aroma, odor and color). On the same ciders, a hedonic study was realized by 341 consumers on an 11 points scale.
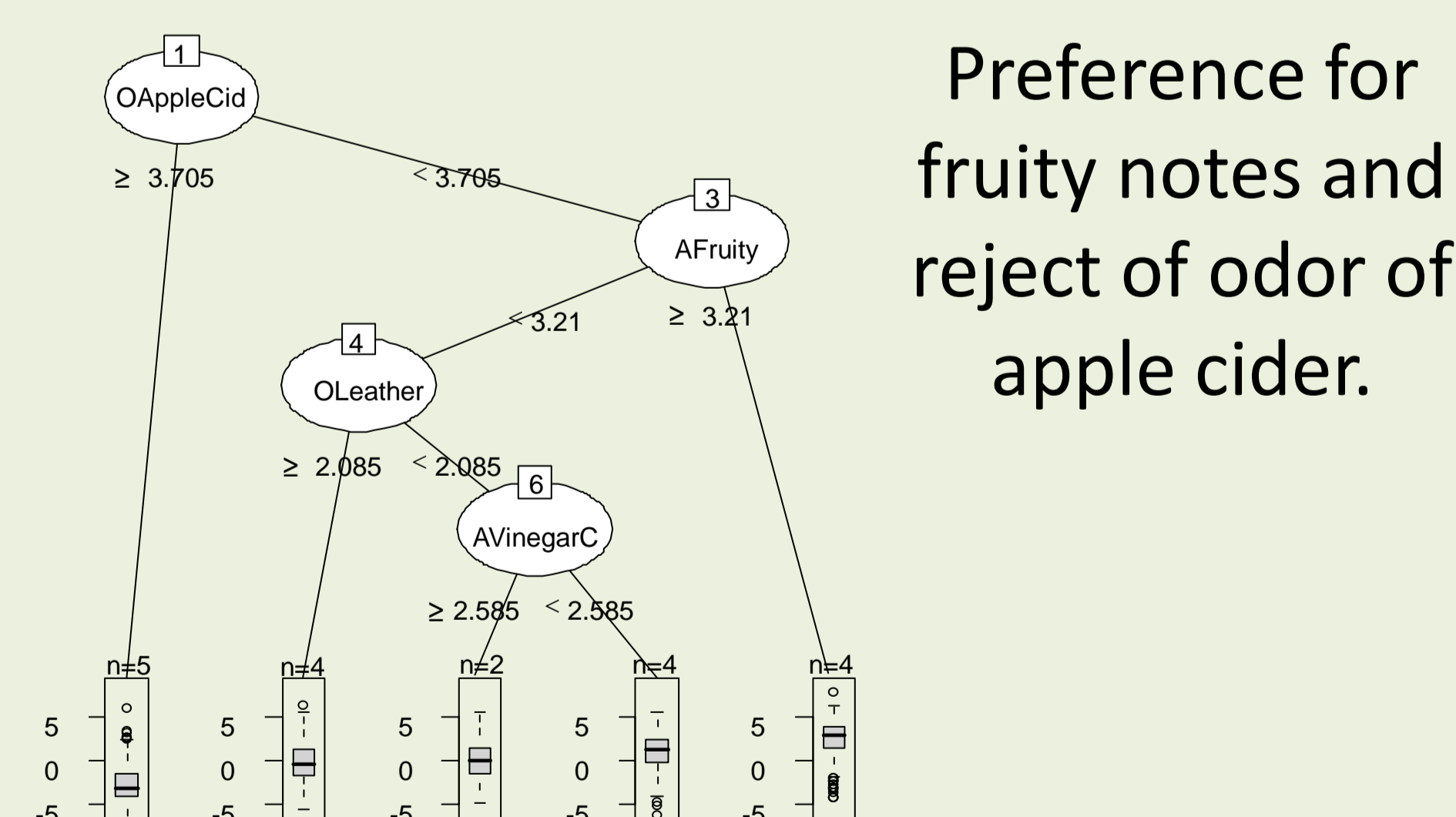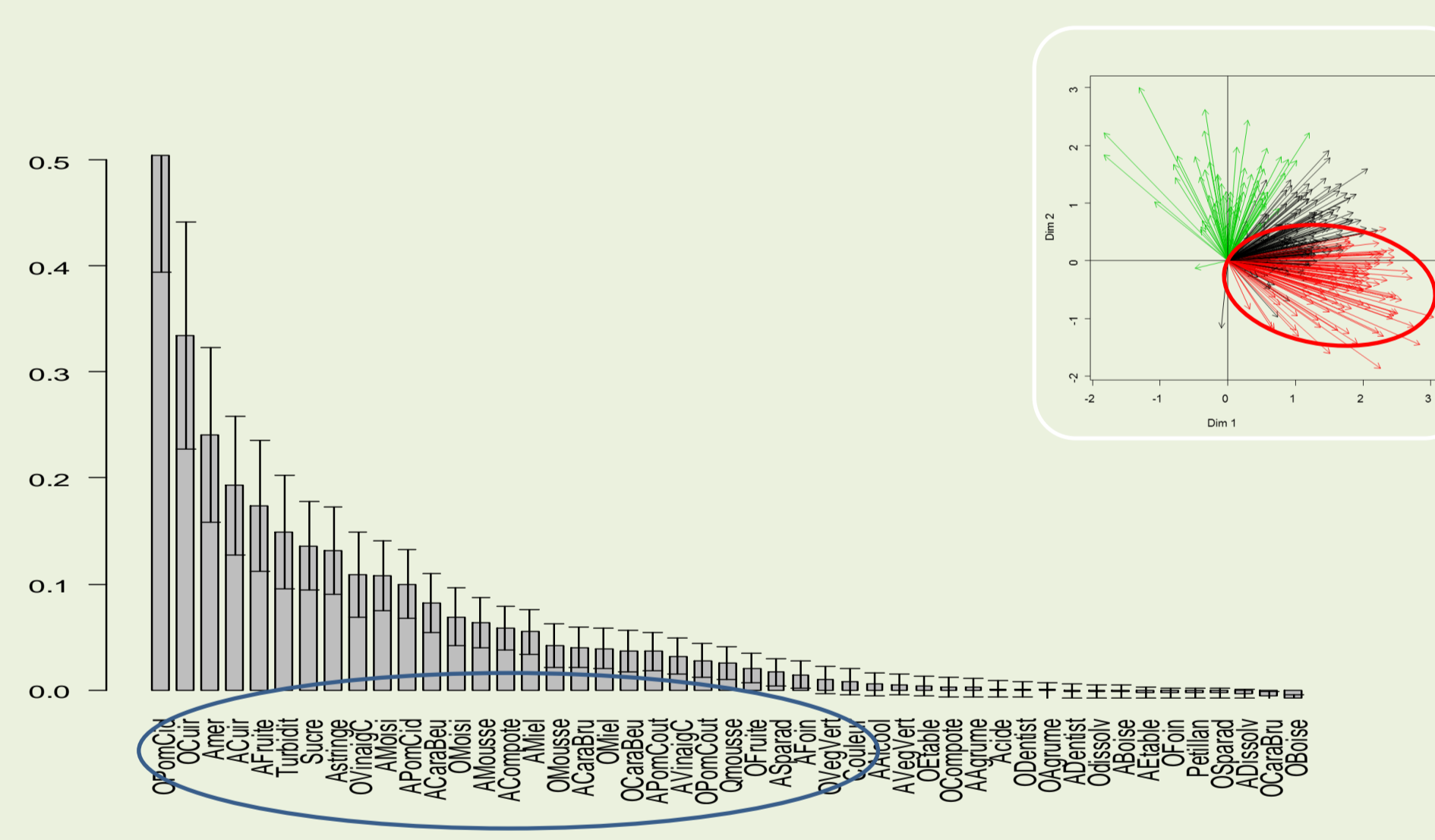
A segmentation of the panel was performed by using k-means method, in order to study the different types of consumers. For each of the three classes, individual preferences were explained by the sensory attributes selected by random forests.
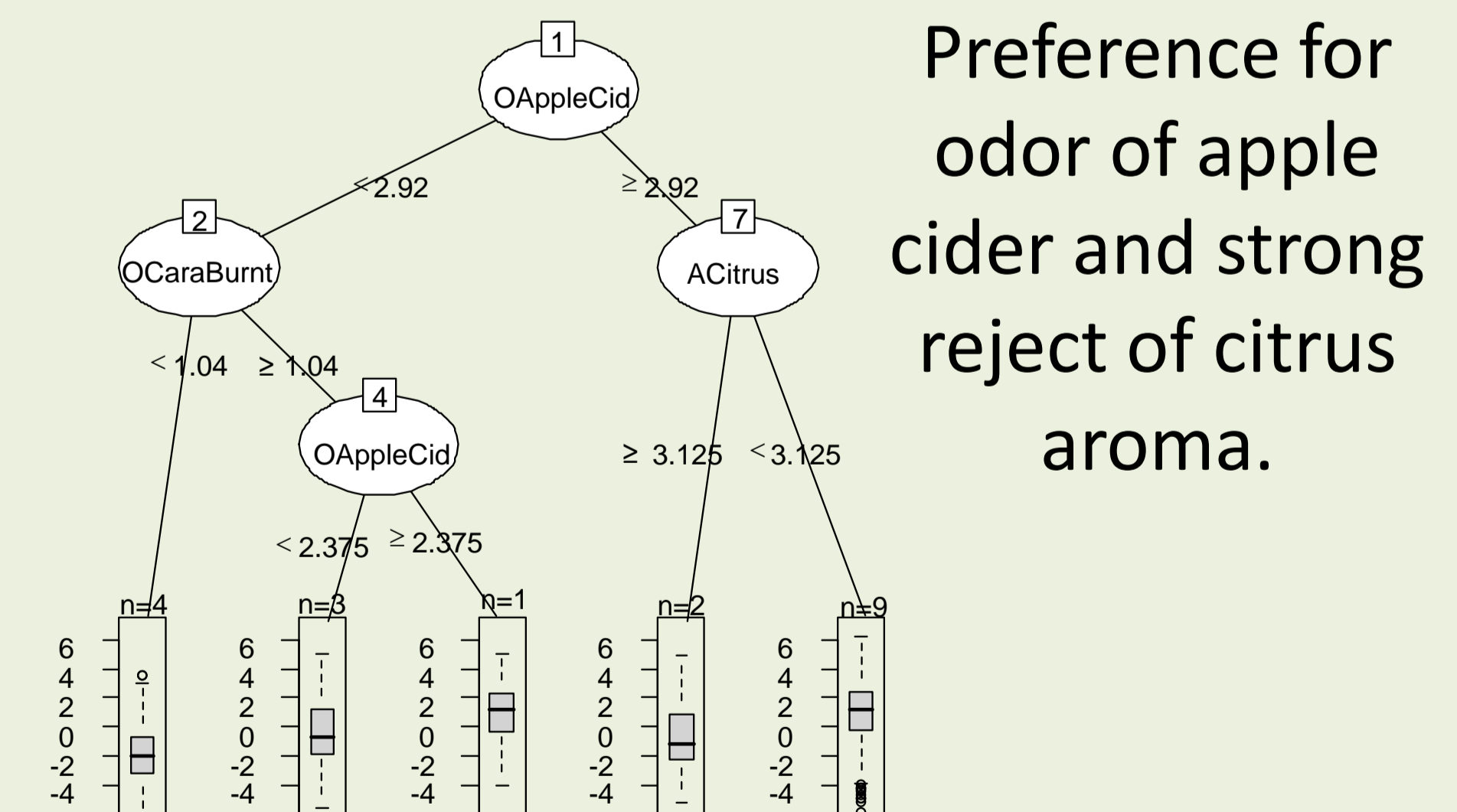
Class 1 : « Consumers closed to average consumer»
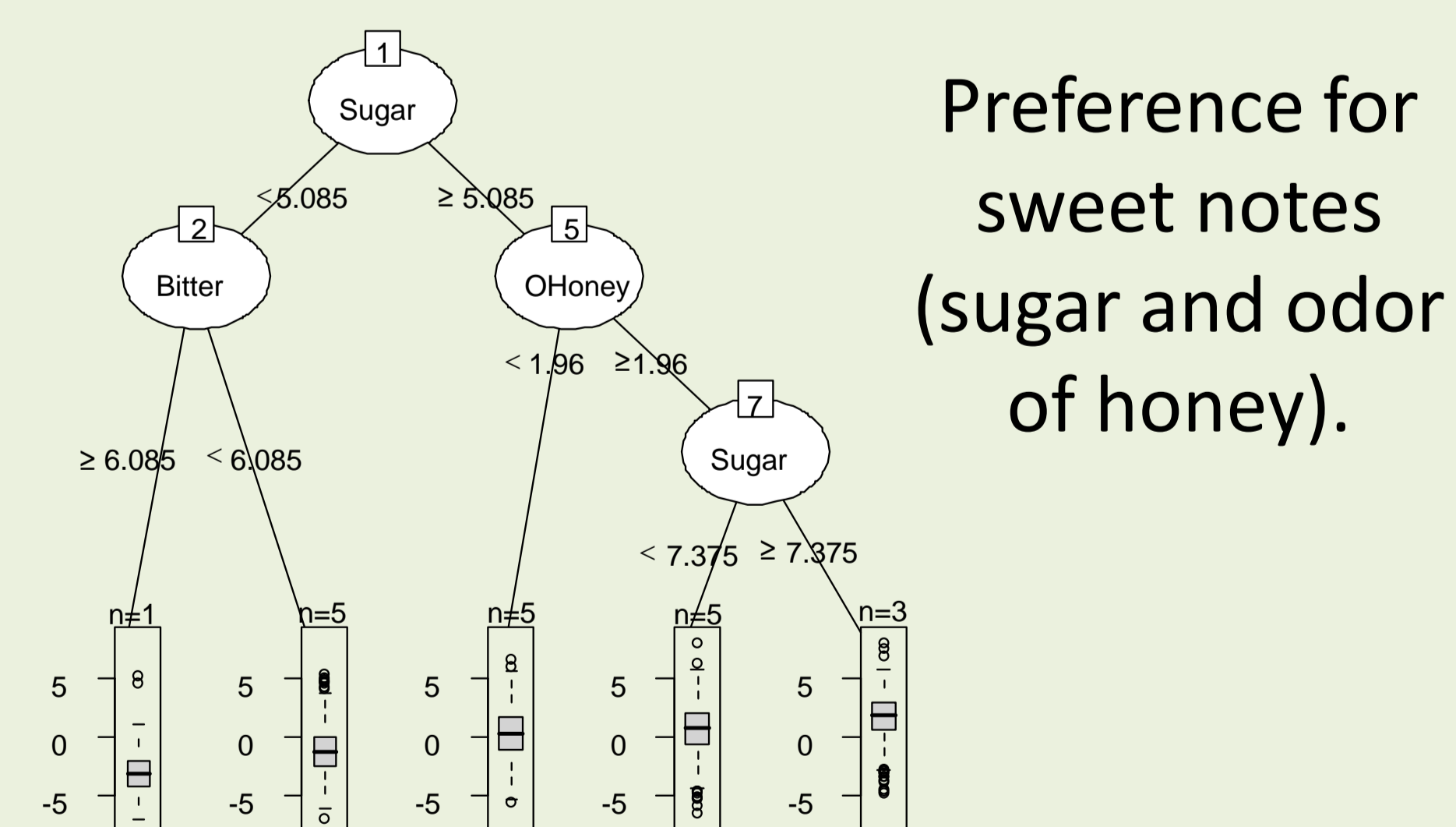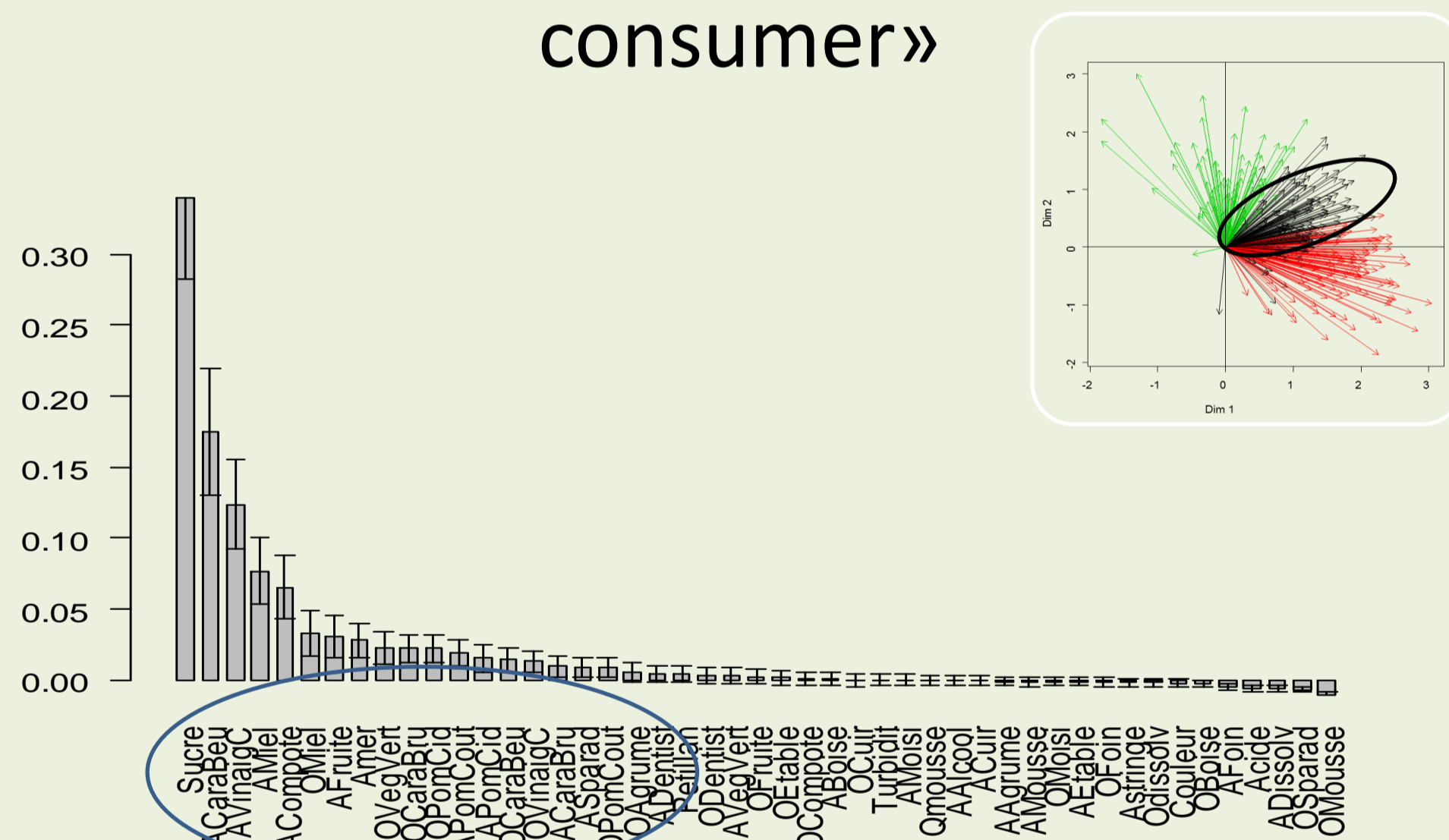
Preference for sweet notes (sugar and odor of honey).

Class 2 : « New consumers »

Preference for fruity notes and reject of odor of apple cider.

Class 3 : « Consumers familiar with ciders»

Preference for odor of apple cider and strong reject of citrus aroma.

Conclusion :

The use of regression trees is clearly suitable in the context of a preference study. Random forests give a way to select reliable attributes for modeling preference. Multivariate trees allow an easy understanding and interpretation of the preference of a panel. It can also deal with non-linearity and interactions between attributes.

References :

Breiman L. (1984). Classification and Regression Trees. Taylor & Francis, 368p.

Breiman L. (2001). Random Forest. *Machine Learning*. Vol 45, pp.5-32

De'ath G. (2002). Multivariate Regression Trees: A New Technique for Modeling Species-Environment Relationships. *Ecology*, Vol 83, No 4, pp 1105-1117

Romano R., Davino C., Naes T. (2014). Classification trees in consumer studies for combining both product attributes and consumer preferences with additional consumer characteristics. *Food Quality and Preference.* Vol 33, 27-36