



# Combining statistics and semantic for an automated data analysis of Free-Comment sensory description of products

Michel Visalli, Benjamin Mahieu

[michel.visalli@inrae.fr](mailto:michel.visalli@inrae.fr)

[benjamin.mahieu@oniris-nantes.fr](mailto:benjamin.mahieu@oniris-nantes.fr)



<https://www.pangbornsymposium.com/>

# Overview

- I. **The collected data and their encoding for statistical analysis**
- II. Cleaning information and extracting sensory descriptors
- III. Semantic-based hierarchical structuration of extracted descriptors
- IV. Aggregation of descriptors: semantic-constrained classification
- V. Application: automated vs. manual encoding
- VI. Conclusion

# I. The collected data and their encoding for statistical analysis

Consumer	Product	Description
C_1	P_1	good taste, good vegetable smell but too compact
...	...	...
C_N <sub>C</sub>	P_N <sub>P</sub>	strong taste, nice color, vegetables are varied

# I. The collected data and their encoding for statistical analysis

Consumer	Product	Description	Instructions (facultative)
C_1	P_1	good taste, good vegetable smell but too compact	Sensory modality (visual aspect, texture, flavor, etc.)
...	...	...	Quality vs. Defect
C_N <sub>C</sub>	P_N <sub>P</sub>	strong taste, nice color, vegetables are varied	

# I. The collected data and their encoding for statistical analysis

Consumer	Product	Description	Instructions (facultative)
C <sub>1</sub>	P <sub>1</sub>	good taste, good vegetable smell but too compact	Sensory modality (visual aspect, texture, flavor, etc.)  Quality vs. Defect
...	...	...	
C <sub>N<sub>C</sub></sub>	P <sub>N<sub>P</sub></sub>	strong taste, nice color, vegetables are varied	

Consumer	Product	D <sub>1</sub>	D <sub>2</sub>	D <sub>2</sub>	...	D <sub>N<sub>D</sub></sub>	Instructions (facultative)
C <sub>1</sub>	P <sub>1</sub>	1	1	1	...	0	Sensory modality (visual aspect, texture, flavor, etc.)  Quality vs. Defect
...	...	...	...	...	...	...	
C <sub>N<sub>C</sub></sub>	P <sub>N<sub>P</sub></sub>	1	0	1	...	0	

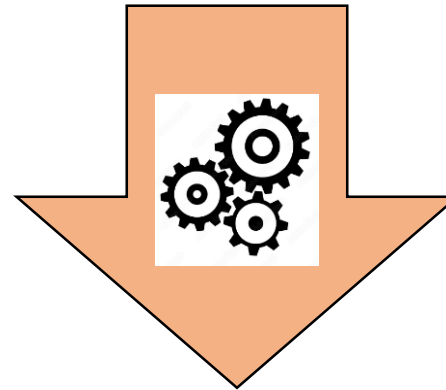
# I. The collected data and their encoding for statistical analysis

Consumer	Product	Description	Instructions (facultative)
C_1	P_1	good taste, good vegetable smell but too compact	Sensory modality (visual aspect, texture, flavor, etc.)  Quality vs. Defect
...	...	...	
C_N <sub>C</sub>	P_N <sub>P</sub>	strong taste, nice color, vegetables are varied	

Consumer	Product	D_1	D_2	D_2	...	D_N <sub>D</sub>	Instructions (facultative)
C_1	P_1	1	1	1	...	0	Sensory modality (visual aspect, texture, flavor, etc.)  Quality vs. Defect
...	...	...	...	...	...	...	
C_N <sub>C</sub>	P_N <sub>P</sub>	1	0	1	...	0	
		Descriptor cited or not					

# I. The collected data and their encoding for statistical analysis

Consumer	Product	Description	Instructions (facultative)
C_1	P_1	good taste, good vegetable smell but too compact	Sensory modality (visual aspect, texture, flavor, etc.)  Quality vs. Defect
...	...	...	
C_N <sub>C</sub>	P_N <sub>P</sub>	strong taste, nice color, vegetables are varied	

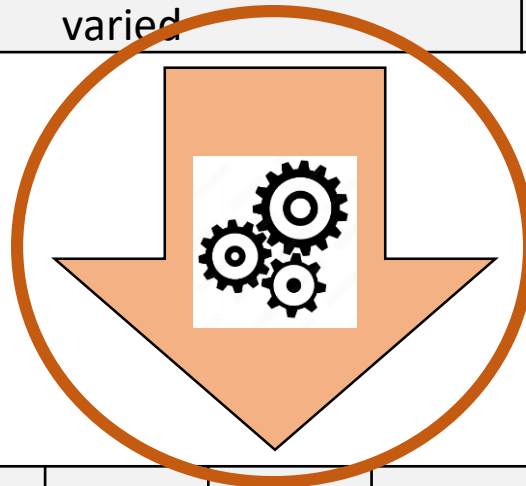


Consumer	Product	D_1	D_2	D_2	...	D_N <sub>D</sub>	Instructions (facultative)
C_1	P_1	1	1	1	...	0	Sensory modality (visual aspect, texture, flavor, etc.)  Quality vs. Defect
...	...	...	...	...	...	...	
C_N <sub>C</sub>	P_N <sub>P</sub>	1	0	1	...	0	

Descriptor cited or not

# I. The collected data and their encoding for statistical analysis

Consumer	Product	Description	Instructions (facultative)
C_1	P_1	good taste, good vegetable smell but too compact	Sensory modality (visual aspect, texture, flavor, etc.)  Quality vs. Defect
...	...	...	
C_N <sub>C</sub>	P_N <sub>P</sub>	strong taste, nice color, vegetables are varied	



Consumer	Product	D_1	D_2	D_2	...	D_N <sub>D</sub>	Instructions (facultative)
C_1	P_1	1	1	1	...	0	Sensory modality (visual aspect, texture, flavor, etc.)  Quality vs. Defect
...	...	...	...	...	...	...	
C_N <sub>C</sub>	P_N <sub>P</sub>	1	0	1	...	0	
		Descriptor cited or not					



# Overview

- I. The collected data and their encoding for statistical analysis
- II. Cleaning information and extracting sensory descriptors**
- III. Semantic-based hierarchical structuration of extracted descriptors
- IV. Aggregation of descriptors: semantic-constrained classification
- V. Application: automated vs. manual encoding
- VI. Conclusion

## II. Cleaning information and extracting sensory descriptors

**Rule-based data extraction** = application a particular set of rules and patterns to capture specific expressions and **transform text into data**

Rule creation: Based on a **domain-specific lexicon** including **food related attributes grouped by categories**

Rule application:

- **Normalization** of Free-Comment to prepare raw text for NLP
- **Extraction of attributes** using **regular expressions** (sequences of characters that specify a search pattern used by string-searching algorithms)
- **Contextualization of attributes** using **semantic rules**

Rule refinement: **Iterative refinement** of rules by repetitive processing to improve accuracy and performance based on **panel leader feedback**

### Example of Free-Comment given by a consumer

*This compote has a taste of peaches with large pieces giving it an "homemade" aspect. It is not acid but too sweet and cloying.*

### Conversion of text to lower case

*t*his compote has a taste of peaches with large pieces giving it an "homemade" aspect. It is not acid but too sweet and cloying.

## II. Cleaning information and extracting sensory descriptors

**Removal of words less than 3 characters** (except those in a predefined list)

*this compote has ~~a~~ taste ~~of~~ peaches with large pieces giving ~~it~~ ~~an~~ "homemade" aspect. ~~It is~~ not acid but too sweet and cloying.*

### Removal of special characters and multiple spaces

*this compote has taste peaches with large pieces giving "homemade" aspect. not acid but too swet and cloying.*

## II. Cleaning information and extracting sensory descriptors

**Spellchecking** using hunspell (free spell checker and morphological analyzer library, available in many programming languages)

*this compote has taste peaches with large pieces giving homemade aspect. not acid but too **sweet** and cloying.*

## II. Cleaning information and extracting sensory descriptors

### Search for words/expressions of interest

- **attributes** (related to food products)
- **quantifiers** (related to perceived intensities, sizes or levels of appropriateness of the attributes)
- **context words** (additional information on the attributes)

*this compote has taste peaches with large pieces giving homemade aspect. not acid but too sweet and cloying.*



## II. Cleaning information and extracting sensory descriptors

**Lemmatization** = replacement of words/expressions by their canonical forms (singular, masculine)

*this compote has taste peaches with large pieces giving homemade aspect. not acid but too-much sweet and cloying.*

## II. Cleaning information and extracting sensory descriptors

**Association** of quantifiers and context words with attributes  
(based on semantic-rules, recursive search)

*this compote has taste+peach with large+piece giving  
homemade+aspect. not+acid but too-much+sweet and  
cloying.*

# Overview

- I. The collected data and their encoding for statistical analysis
- II. Cleaning information and extracting sensory descriptors
- III. Semantic-based hierarchical structuration of extracted descriptors**
- IV. Aggregation of descriptors: semantic-constrained classification
- V. Application: automated vs. manual encoding
- VI. Conclusion

### III. Semantic-based hierarchical structuration of extracted descriptors

**Objective = to “make sense” of the extracted data and give information to the clustering algorithm to aggregate attributes without human intervention**

Attributes are transformed into **concepts** organized according to a **hierarchical structure** defined in the domain-specific lexicon

- **Peach** is a **flavor attribute**, belonging to the “**fruity**” category, and “**yellow fruit**” sub-category → **Flavor>Fruity>Yellow-fruit>Peach** (can be aggregated with “nectarine”)
- **Acid** is a **basic taste attribute**, associated with the **negative quantifier “not”** → **Taste>Not-acid** ( $\neq$  Taste>Acid)
- **Homemade** is a non-categorized **attribute**, associated with the **context-word “appearance”** → **NC>Homemade-appearance**

### III. Semantic-based hierarchical structuration of extracted descriptors

*This compote has a taste of peaches with large pieces giving it an "homemade" aspect. It is not acid but too sweet and cloying.*



Expression	Attribute	Modality	Hierarchy	Quantifier	Context	Concept
compote	<del>compote</del>	-	-	-	-	
taste peaches	peach	flavor	fruity>yellow-fruit	-	taste	flavor>fruity>yellow-fruit>peach-taste
large pieces	piece	texture	fracturable	large	-	texture>fracturable>piece-large
homemade aspect	homemade	-	-	-	aspect	NC>homemade-aspect
too sweet	sweet	taste	-	too-much	-	taste>too-much-sweet
not acid	acid	taste	-	not	-	taste>not-acid
cloying	cloying	valence	not-pleasant	-	-	valence>not-pleasant>cloying

# III. Semantic-based hierarchical structuration of extracted descriptors

	A	B	C	D	E	F
1	RegularExpressionFR	AttributeFR	FullConceptFR	ProductCategory	AttributeEN	
360	solvants?	solvant	flaveur:chimique>solvant		solvent	
361	verniss	vernis	flaveur:chimique>vernis		varnish	
362	acescent?c?e?s?	acescent	saveur:acide		sour	
363	vinaigré?e?s?	vinaigre	saveur:acide		vinegar	
364	ac[ée]tiques?	acétique	flaveur:chimique>vinaigre>acétique		acetic	
365	acides? ac[ée]tiques?	acide acétique	saveur:acide		acetic acid	
366	d[ée]fauts?	défaut	flaveur:défaut		defect	
367	anomalie	anomalie	flaveur:défaut>anomalie		anomaly	
368	bouchons?	bouchon	flaveur:défaut>bouchon	vin	corked	
369	bouchonn[ée]e?s?	bouchonné	flaveur:défaut>bouchon	vin	corked	
370	probl[ée]mes?	problème	flaveur:défaut>problème		problem	
371	empyreumatiques?	empyreumatique	flaveur:empyreumatique		empyreumatic	
372	bouillons?	bouillon	flaveur:empyreumatique>bouillon		broth	
373	brais[ée]e?s?	braisé	flaveur:empyreumatique>braisé		braised	
374	brul[ée]e?s?	brûlé	flaveur:empyreumatique>brûlé		burnt	
375	benzol	benzol	flaveur:empyreumatique>brulé>benzol		benzol	
376	bitum[ée]e?s?	bitume	flaveur:empyreumatique>brulé>bitume		tar	
377	caoutchouc	caoutchouc	flaveur:empyreumatique>brûlé>caoutchouc		rubber	
378	crame[ée]e?s?	cramé	flaveur:empyreumatique>brûlé>cramé		burnt	
379	goudron	goudron	flaveur:empyreumatique>brûlé>goudron		tar	
380	suie	suie	flaveur:empyreumatique>brûlé>suie		soot	
381	caramels?	caramel	flaveur:empyreumatique>caramel		caramel	
382	caram[ée]lis[ée]e?s?	caramélisé	flaveur:empyreumatique>caramel		caramelized	
383	confiture lait	confiture de lait	flaveur:empyreumatique>caramel>confiture-de-lait		milk jam	
384	chocolat[ée]?e?s?	chocolat	flaveur:empyreumatique>chocolat		chocolate	
385	chocolats? chauds?	chocolat chaud	flaveur:empyreumatique>chocolat>chocolat-chaud		hot chocolate	
386	chocolats? lait	chocolat au lait	flaveur:empyreumatique>chocolat>chocolat-lait		milk chocolate	
387	chocolats? noirs?	chocolat noir	flaveur:empyreumatique>chocolat>chocolat-noir		dark chocolate	
388	cigarettes?	cigarette	flaveur:empyreumatique>fumé		cigarette	
389	fumage	fumage	flaveur:empyreumatique>fumé		smoked	
390	fum[ée]e?s?	fumé	flaveur:empyreumatique>fumé		smoked	
391	fumet	fumet	flaveur:empyreumatique>fumé		smoked	
392	bacon	bacon	flaveur:empyreumatique>fumé>bacon		bacon	
393	barbecue	barbecue	flaveur:empyreumatique>fumé>barbecue		barbecue	
394	grill[ée]?e?s?	grillé	flaveur:empyreumatique>grillé		grilled	
395	biscottes?	biscotte	flaveur:empyreumatique>grillé>biscotte		rusk	
396	pains?	pain	flaveur:empyreumatique>grillé>pain		bread	

**Domain-specific lexicon**  
(food products)

**Includes generic and specific attributes**

# III. Semantic-based hierarchical structuration of extracted descriptors

	A	B	C	D	E	F
1	RegularExpressionFR	AttributeFR	FullConceptFR	ProductCategory	AttributeEN	
360	solvants?	solvant	flaveur:chimique>solvant		solvent	
361	verniss	vernis	flaveur:chimique>vernis		varnish	
362	acescent?c?e?s?	acescent	saveur:acide		sour	
363	vinaigré?e?s?	vinaigre	saveur:acide		vinegar	
364	ac[ée]tiques?	acétique	flaveur:chimique>vinaigre>acétique		acetic	
365	acides? ac[ée]tiques?	acide acétique	saveur:acide		acetic acid	
366	d[ée]fauts?	défaut	flaveur:défaut		defect	
367	anomalie	anomalie	flaveur:défaut>anomalie		anomaly	
368	bouchons?	bouchon	flaveur:défaut>bouchon	vin	corked	
369	bouchonn[ée]				corked	
370	probl[ée]mes				problem	
371	empyreumatique				empyreumatic	
372	bouillons?				broth	
373	brais[ée]e?s?				braised	
374	brul[ée]e?s?				burnt	
375	benzol				benzol	
376	bitum[ée]e?s?				tar	
377	caoutchouc				rubber	
378	crame[ée]e?s?				burnt	
379	goudron				tar	
380	suie				soot	
381	caramels?				caramel	
382	caram[ée]lis[ée]				caramelized	
383	confiture lait				milk jam	
384	chocolat[ée]?				chocolate	
385	chocolats? chauds?	chocolat chaud	flaveur:empyreumatique>chocolat>chocolat-chaud		hot chocolate	
386	chocolats? lait	chocolat au lait	flaveur:empyreumatique>chocolat>chocolat-lait		milk chocolate	
387	chocolats? noirs?	chocolat noir	flaveur:empyreumatique>chocolat>chocolat-noir		dark chocolate	
388	cigarettes?	cigarette	flaveur:empyreumatique>fumé		cigarette	
389	fumage	fumage	flaveur:empyreumatique>fumé		smoked	
390	fum[ée]e?s?	fumé	flaveur:empyreumatique>fumé		smoked	
391	fumet	fumet	flaveur:empyreumatique>fumé		smoked	
392	bacon	bacon	flaveur:empyreumatique>fumé>bacon		bacon	
393	barbecue	barbecue	flaveur:empyreumatique>fumé>barbecue		barbecue	
394	grill[ée]e?s?	grillé	flaveur:empyreumatique>grillé		grilled	
395	biscottes?	biscotte	flaveur:empyreumatique>grillé>biscotte		rusk	
396	pains?	pain	flaveur:empyreumatique>grillé>pain		bread	

Derived from existing lexicons and ontologies and updated by panel leaders through experimentations

1900 attributes organized into categories and sub-categories of attributes:

- **Sensory categories:** Appearance, Flavor, Taste, Texture, Trigeminal
- **Affective categories:** Positive/Negative valence
- **Other categories:** Balance, Complexity, Familiarity, Evoked context, etc.

160 quantifiers organized into:

- Appropriateness: Too-much/Right/Not-enough
- Intensity: Not/Few/Very
- Size: Small/Large

# Overview

- I. The collected data and their encoding for statistical analysis
- II. Cleaning information and extracting sensory descriptors
- III. Semantic-based hierarchical structuration of extracted descriptors
- IV. Aggregation of descriptors: semantic-constrained classification**
- V. Application: automated vs. manual encoding
- VI. Conclusion



## IV. Aggregation of descriptors: semantic-constrained classification

### Objective:

Find a **partition of the descriptors** such that the resulting **product discrimination is maximized** under the **constraint** that **each element of the partition contains descriptors with the same RootConcept**. Product discrimination is measured thanks to an approximation of the pvalue of the  $\chi^2_{mr}$  test<sup>1</sup>.

<sup>1</sup>Bilder, C. R., Loughin, T. M., & Nettleton, D. (2000). Multiple Marginal Independence Testing for Pick Any/C Variables. *Communications in Statistics - Simulation and Computation*, 29(4), 1285–1316. <https://doi.org/10.1080/03610910008813665>

## IV. Aggregation of descriptors: semantic-constrained classification

### Objective:

Find a **partition of the descriptors** such that the resulting **product discrimination is maximized** under the **constraint** that **each element of the partition contains descriptors with the same RootConcept**. Product discrimination is measured thanks to an approximation of the pvalue of the  $\chi_{mr}^2$  test<sup>1</sup>.

### 2 steps approach:

1 - Hierarchical clustering

2 – Partitioning (transfer) algorithm

<sup>1</sup>Bilder, C. R., Loughin, T. M., & Nettleton, D. (2000). Multiple Marginal Independence Testing for Pick Any/C Variables. *Communications in Statistics - Simulation and Computation*, 29(4), 1285–1316. <https://doi.org/10.1080/03610910008813665>

# IV. Aggregation of descriptors: semantic-constrained classification

## Objective:

Find a **partition of the descriptors** such that the resulting **product discrimination is maximized** under the **constraint** that **each element of the partition contains descriptors with the same RootConcept**. Product discrimination is measured thanks to an approximation of the pvalue of the  $\chi^2_{mr}$  test<sup>1</sup>.

## 2 steps approach:

### 1 - Hierarchical clustering

1. Each descriptor forms a cluster
2. Merge clusters resulting in the best increase in product discrimination among mergers that satisfy the constraint
3. Repeat until no merge lead to an increase in product discrimination

### 2 – Partitioning (transfer) algorithm

<sup>1</sup>Bilder, C. R., Loughin, T. M., & Nettleton, D. (2000). Multiple Marginal Independence Testing for Pick Any/C Variables. *Communications in Statistics - Simulation and Computation*, 29(4), 1285–1316. <https://doi.org/10.1080/03610910008813665>

# IV. Aggregation of descriptors: semantic-constrained classification

## Objective:

Find a **partition of the descriptors** such that the resulting **product discrimination is maximized** under the **constraint** that **each element of the partition contains descriptors with the same RootConcept**. Product discrimination is measured thanks to an approximation of the pvalue of the  $\chi^2_{mr}$  test<sup>1</sup>.

## 2 steps approach:

### 1 - Hierarchical clustering

### 2 – Partitioning (transfer) algorithm

1. Each descriptor forms a cluster
2. Merge clusters resulting in the best increase in product discrimination among mergers that satisfy the constraint
3. Repeat until no merge lead to an increase in product discrimination

→ **Provides a reasonable estimation of the number of clusters and initialization of the partitioning algorithm**

<sup>1</sup>Bilder, C. R., Loughin, T. M., & Nettleton, D. (2000). Multiple Marginal Independence Testing for Pick Any/C Variables. *Communications in Statistics - Simulation and Computation*, 29(4), 1285–1316. <https://doi.org/10.1080/03610910008813665>

# IV. Aggregation of descriptors: semantic-constrained classification

## Objective:

Find a **partition of the descriptors** such that the resulting **product discrimination is maximized** under the **constraint** that **each element of the partition contains descriptors with the same RootConcept**. Product discrimination is measured thanks to an approximation of the pvalue of the  $\chi^2_{mr}$  test<sup>1</sup>.

## 2 steps approach:

### 1 - Hierarchical clustering

1. Each descriptor forms a cluster
2. Merge clusters resulting in the best increase in product discrimination among mergers that satisfy the constraint
3. Repeat until no merge lead to an increase in product discrimination

### 2 – Partitioning (transfer) algorithm

1. Consider each descriptor in turn
2. Consider moving it from its current cluster to every others and perform the move if it increases product discrimination.
3. Repeat until convergence

→ **Provides a reasonable estimation of the number of clusters and initialization of the partitioning algorithm**

<sup>1</sup>Bilder, C. R., Loughin, T. M., & Nettleton, D. (2000). Multiple Marginal Independence Testing for Pick Any/C Variables. *Communications in Statistics - Simulation and Computation*, 29(4), 1285–1316. <https://doi.org/10.1080/03610910008813665>

# IV. Aggregation of descriptors: semantic-constrained classification

## Objective:

Find a **partition of the descriptors** such that the resulting **product discrimination is maximized** under the **constraint** that **each element of the partition contains descriptors with the same RootConcept**. Product discrimination is measured thanks to an approximation of the pvalue of the  $\chi^2_{mr}$  test<sup>1</sup>.

## 2 steps approach:

### 1 - Hierarchical clustering

1. Each descriptor forms a cluster
2. Merge clusters resulting in the best increase in product discrimination among mergers that satisfy the constraint
3. Repeat until no merge lead to an increase in product discrimination

→ Provides a reasonable estimation of the number of clusters and initialization of the partitioning algorithm

### 2 – Partitioning (transfer) algorithm

1. Consider each descriptor in turn
2. Consider moving it from its current cluster to every others and perform the move if it increases product discrimination.
3. Repeat until convergence

→ Refined the solution from the hierarchical clustering

<sup>1</sup>Bilder, C. R., Loughin, T. M., & Nettleton, D. (2000). Multiple Marginal Independence Testing for Pick Any/C Variables. *Communications in Statistics - Simulation and Computation*, 29(4), 1285–1316. <https://doi.org/10.1080/03610910008813665>

# Overview

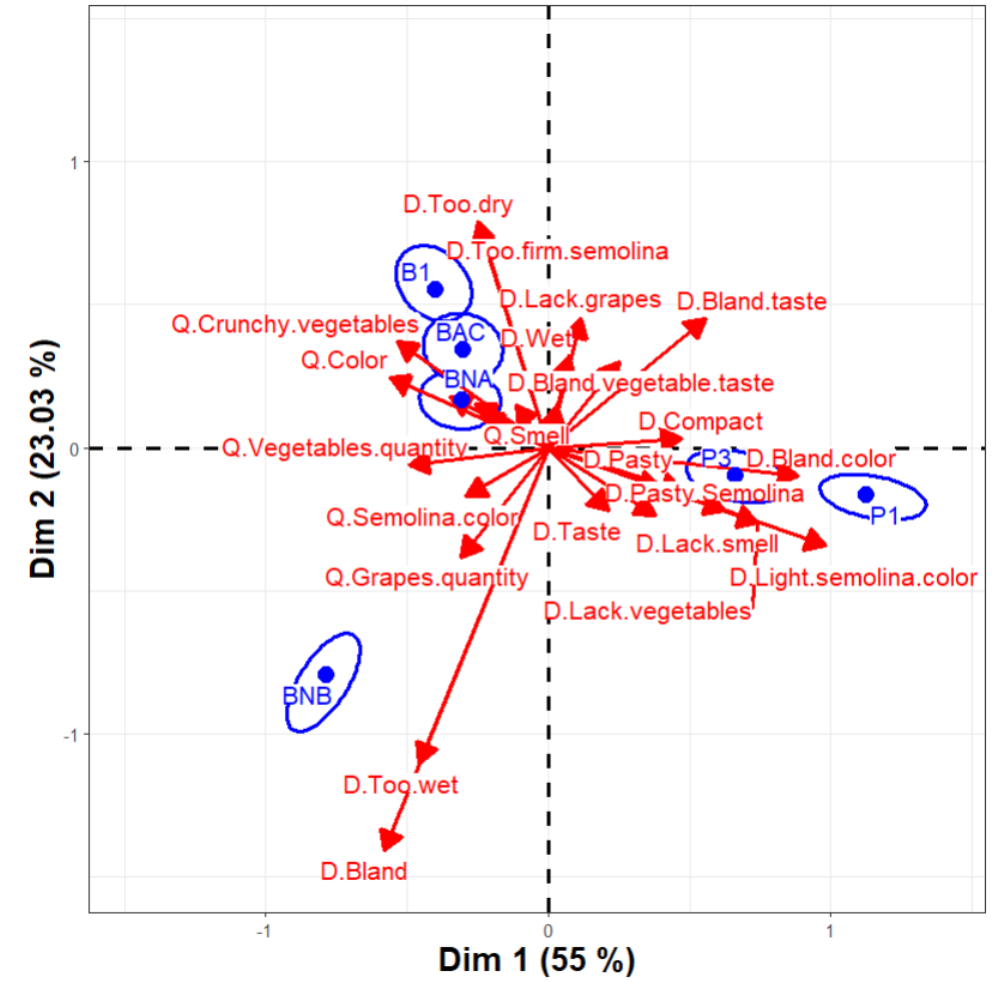
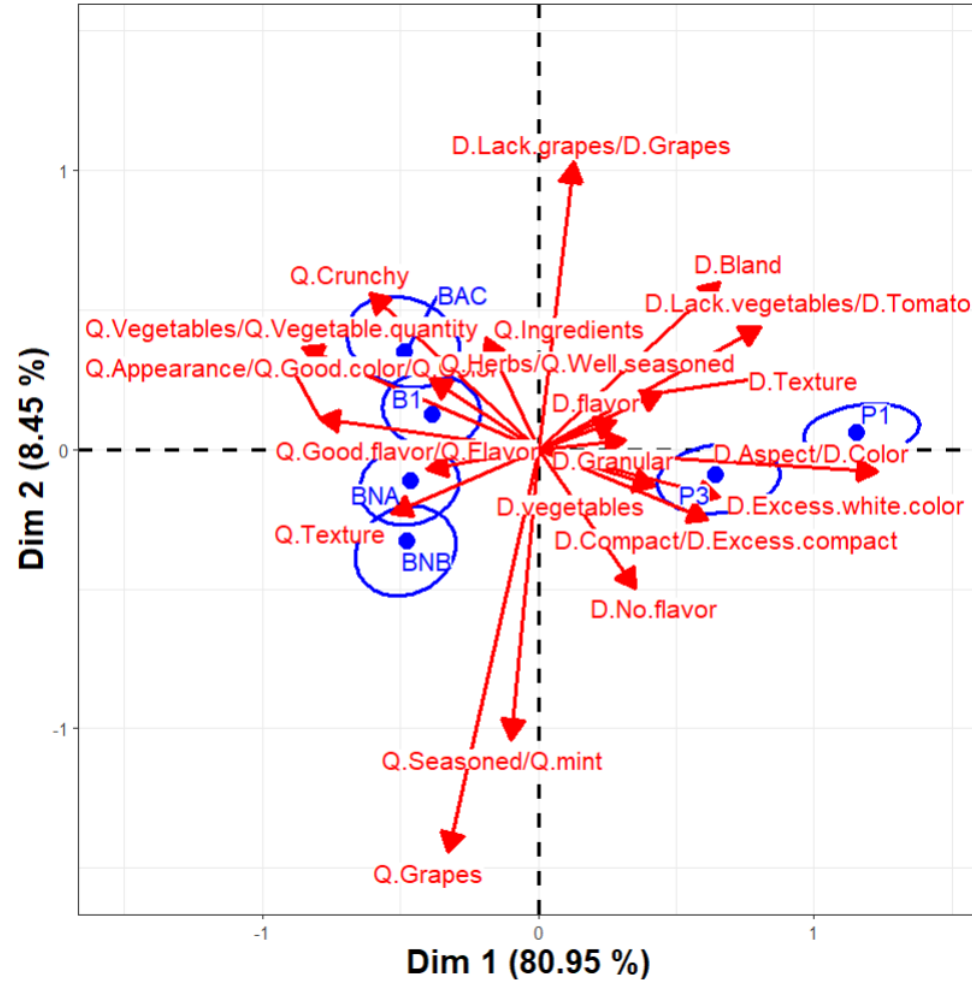
- I. The collected data and their encoding for statistical analysis
- II. Cleaning information and extracting sensory descriptors
- III. Semantic-based hierarchical structuration of extracted descriptors
- IV. Aggregation of descriptors: semantic-constrained classification
- V. Application: automated vs. manual encoding**
- VI. Conclusion

# V. Application : automated vs manual encoding process

## Automated encoding process

## Manual encoding process

Only descriptors having an association with at least one product are shown



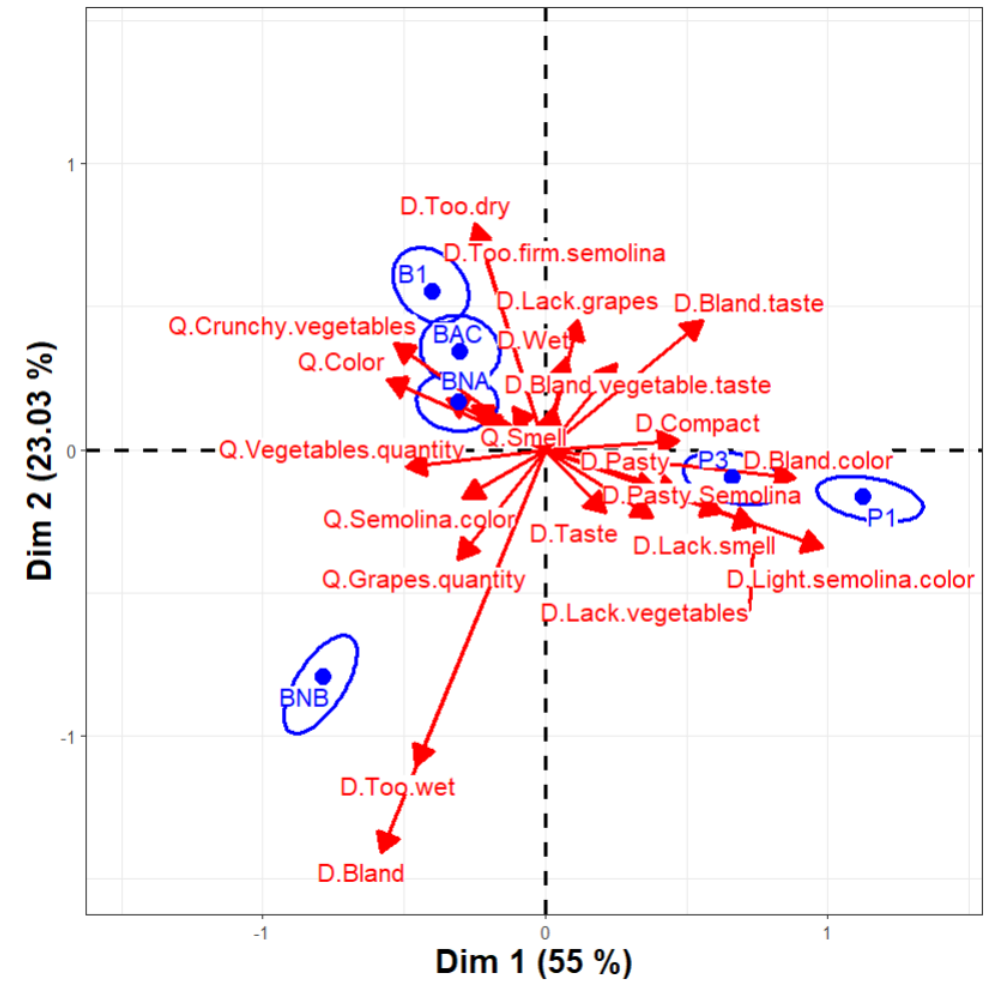
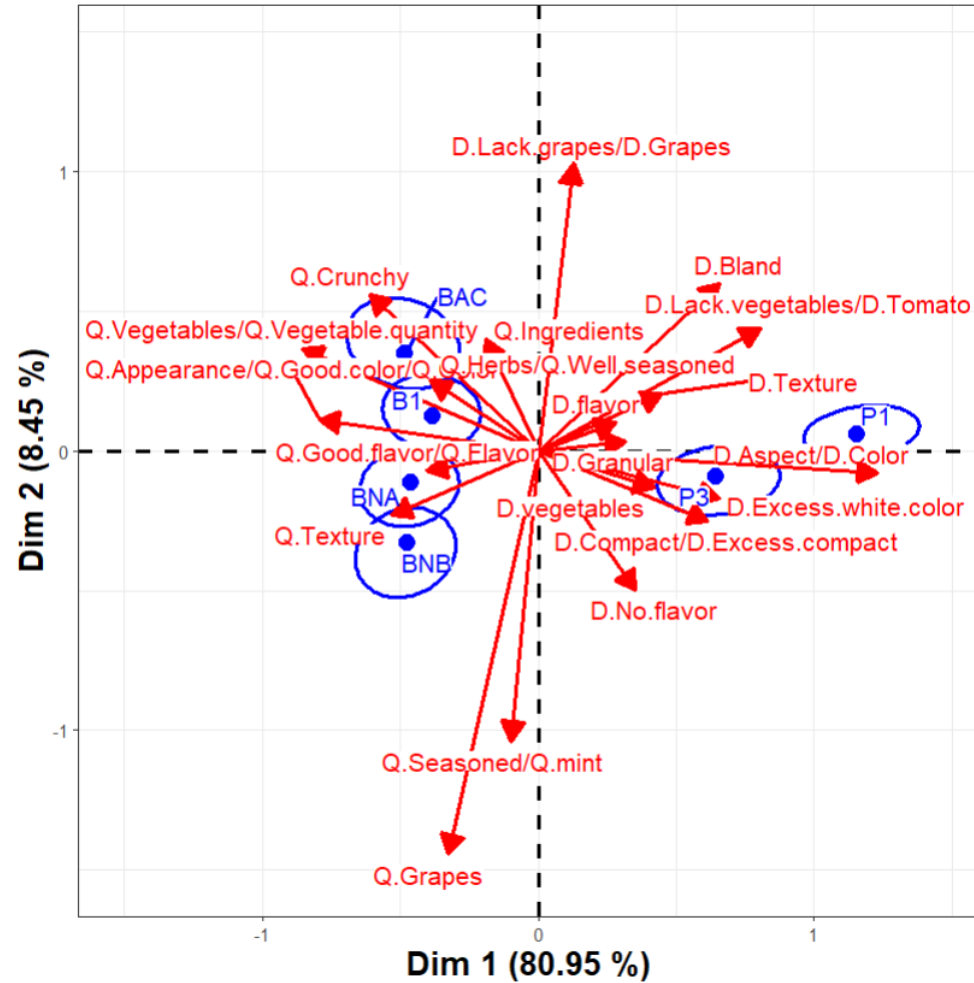


# V. Application : automated vs manual encoding process

## Automated encoding process

## Manual encoding process

Only descriptors having an association with at least one product are shown



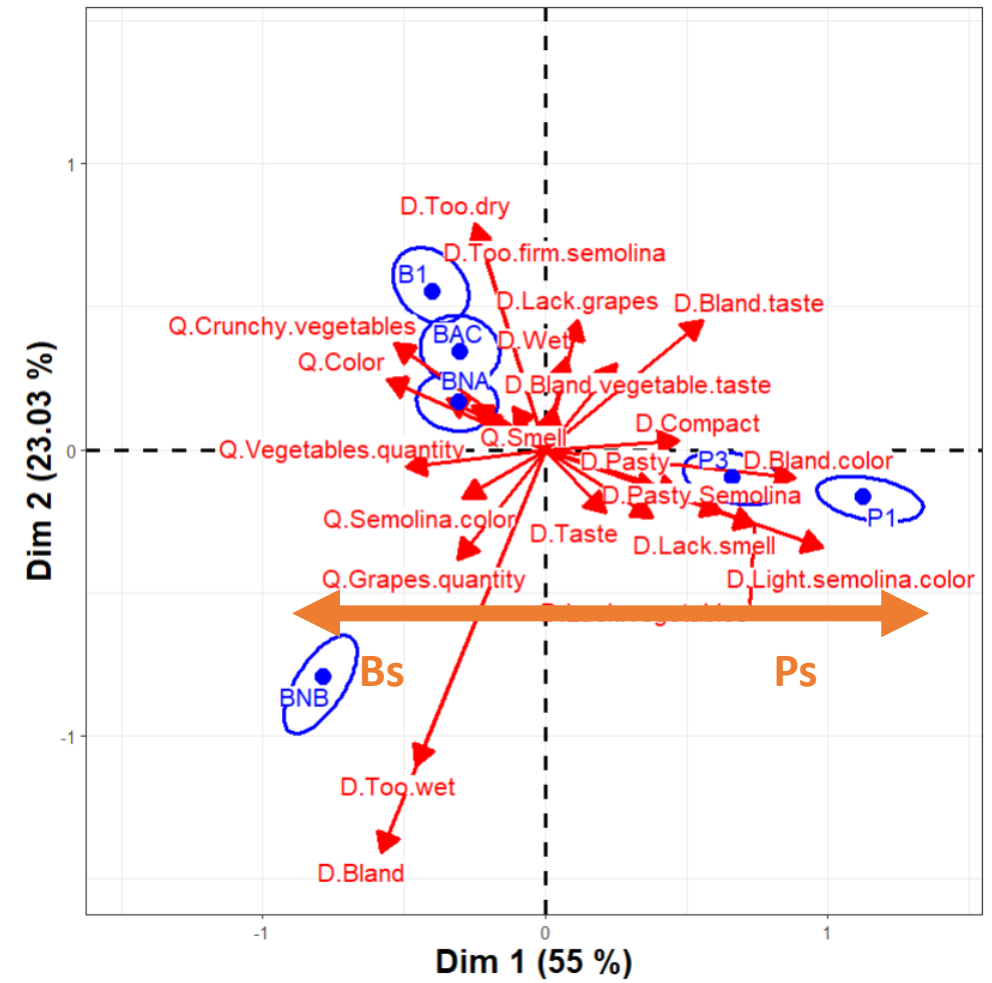
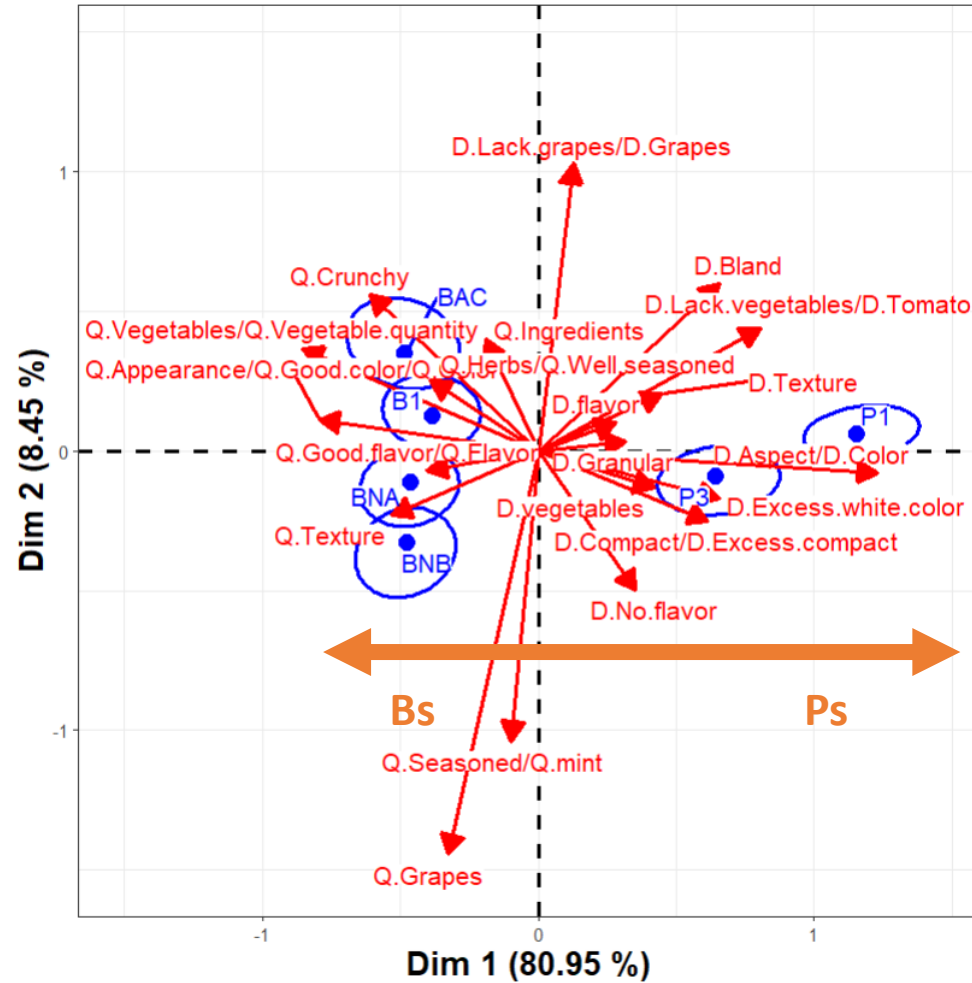
➤ Same first dimension

# V. Application : automated vs manual encoding process

Automated encoding process

Manual encoding process

Only descriptors having an association with at least one product are shown



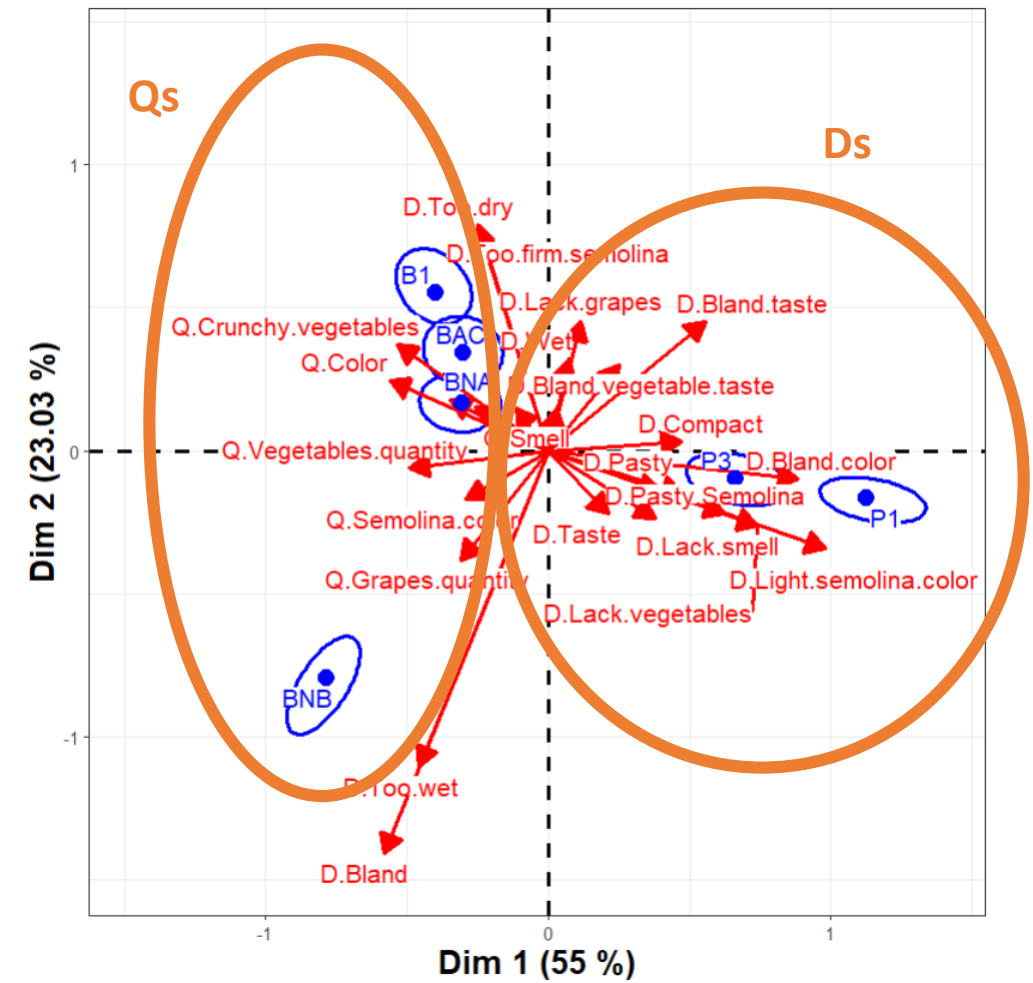
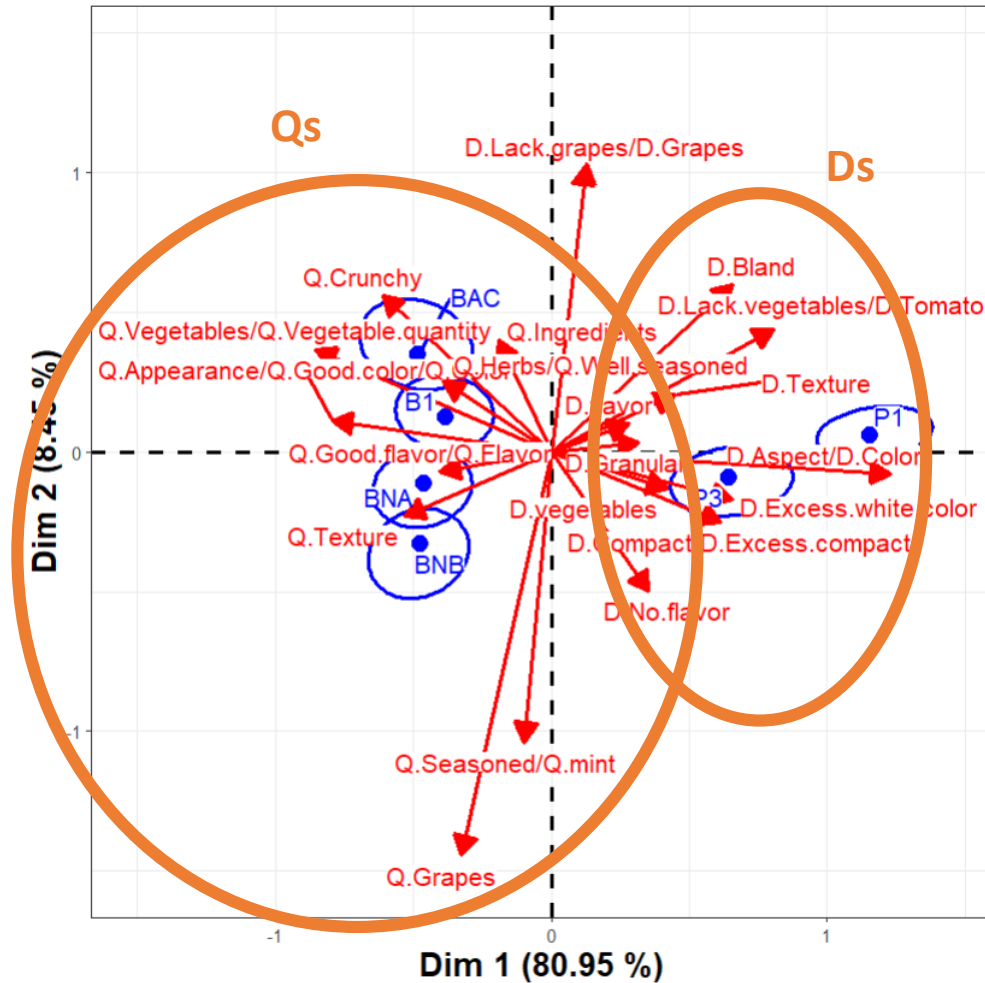
➤ Same first dimension  
 → Product configuration ←

# V. Application : automated vs manual encoding process

## Automated encoding process

## Manual encoding process

Only descriptors having an association with at least one product are shown



➤ Same first dimension

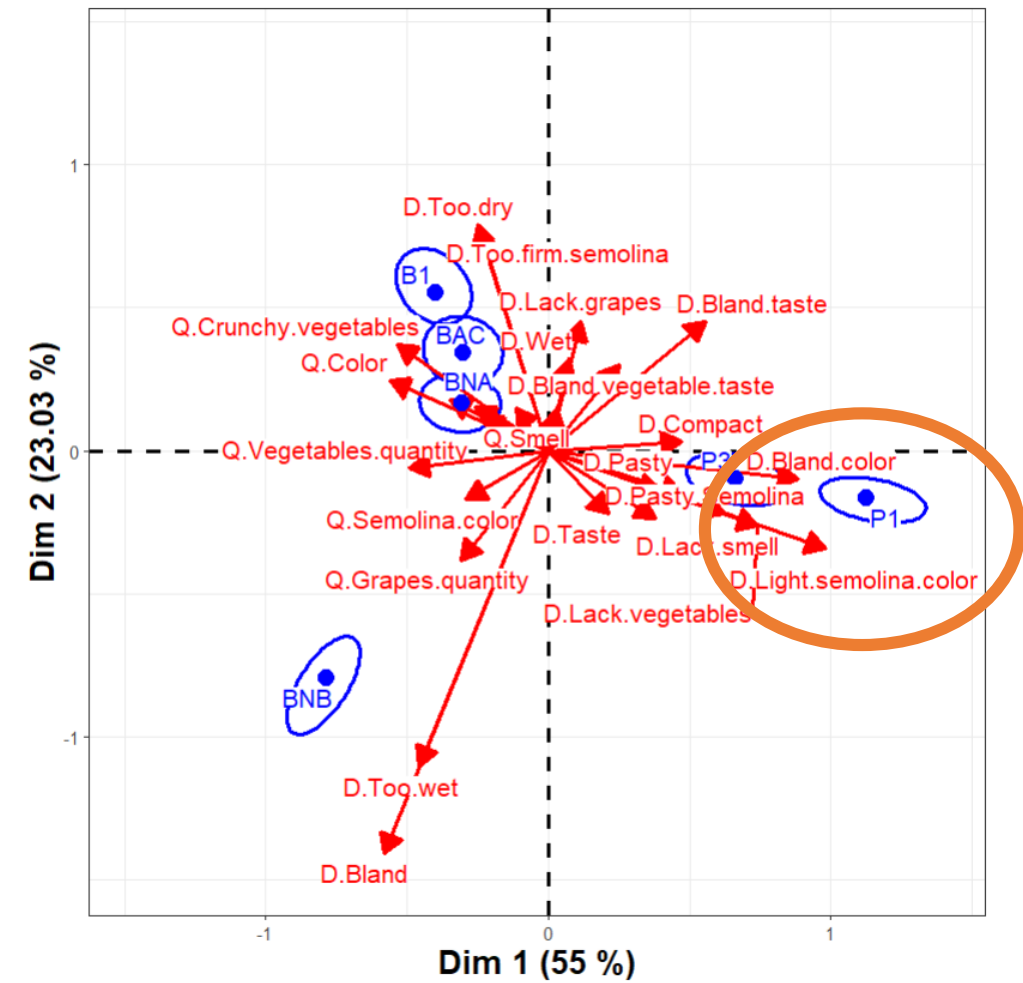
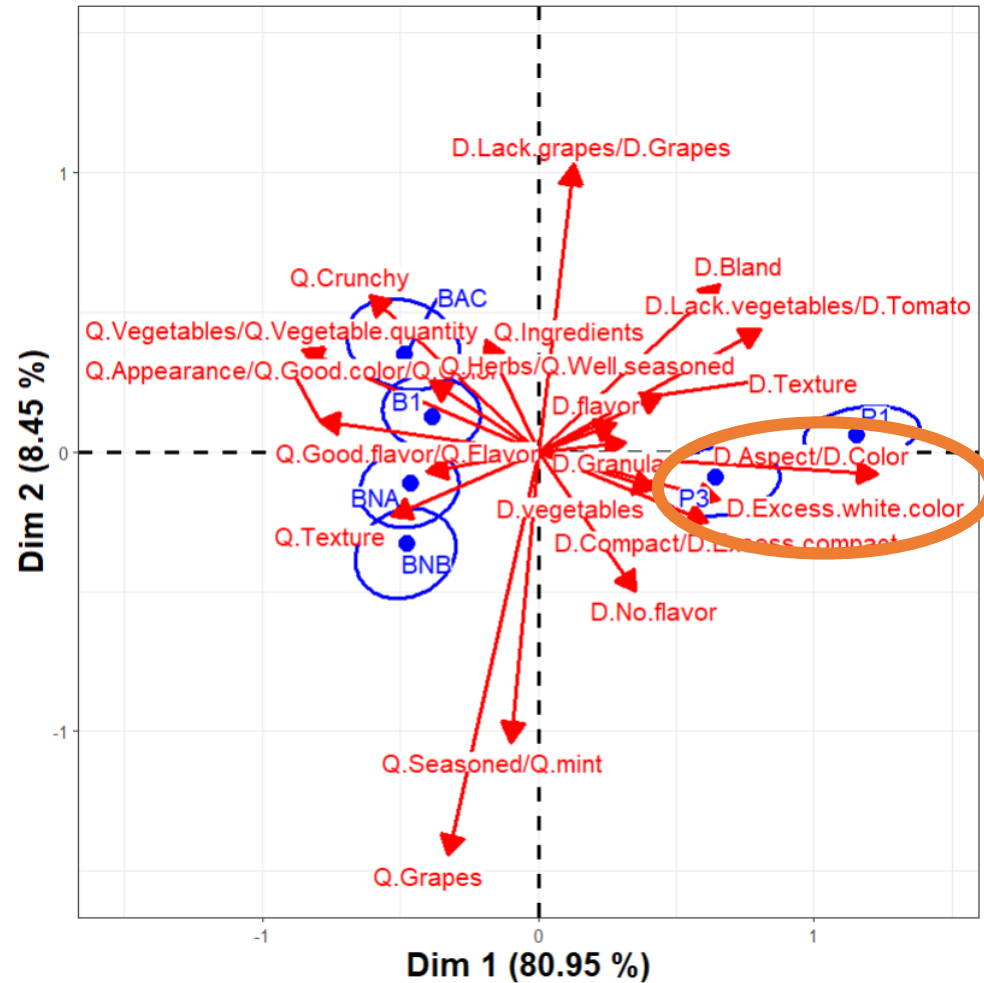
➔ Almost no quality in positive parts : qualities lie on the negative part ➔

# V. Application : automated vs manual encoding process

## Automated encoding process

## Manual encoding process

Only descriptors having an association with at least one product are shown



➤ Same first dimension

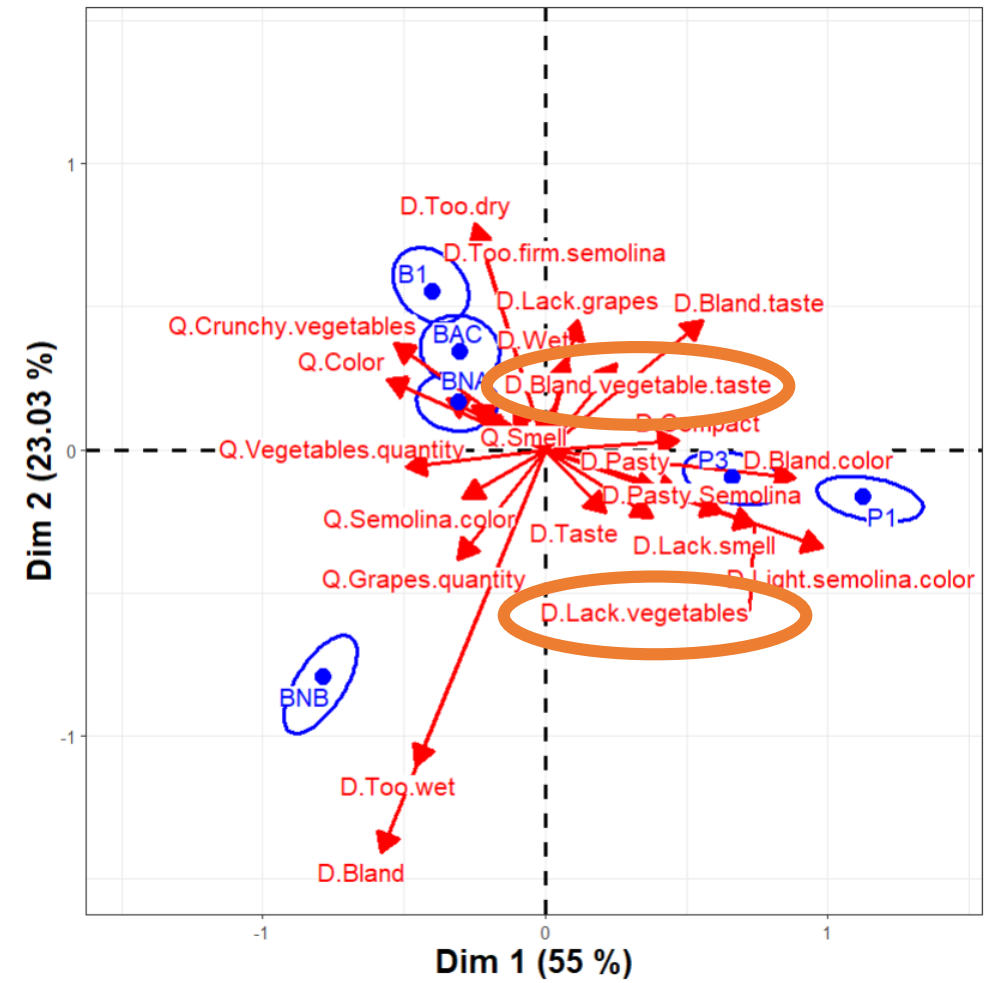
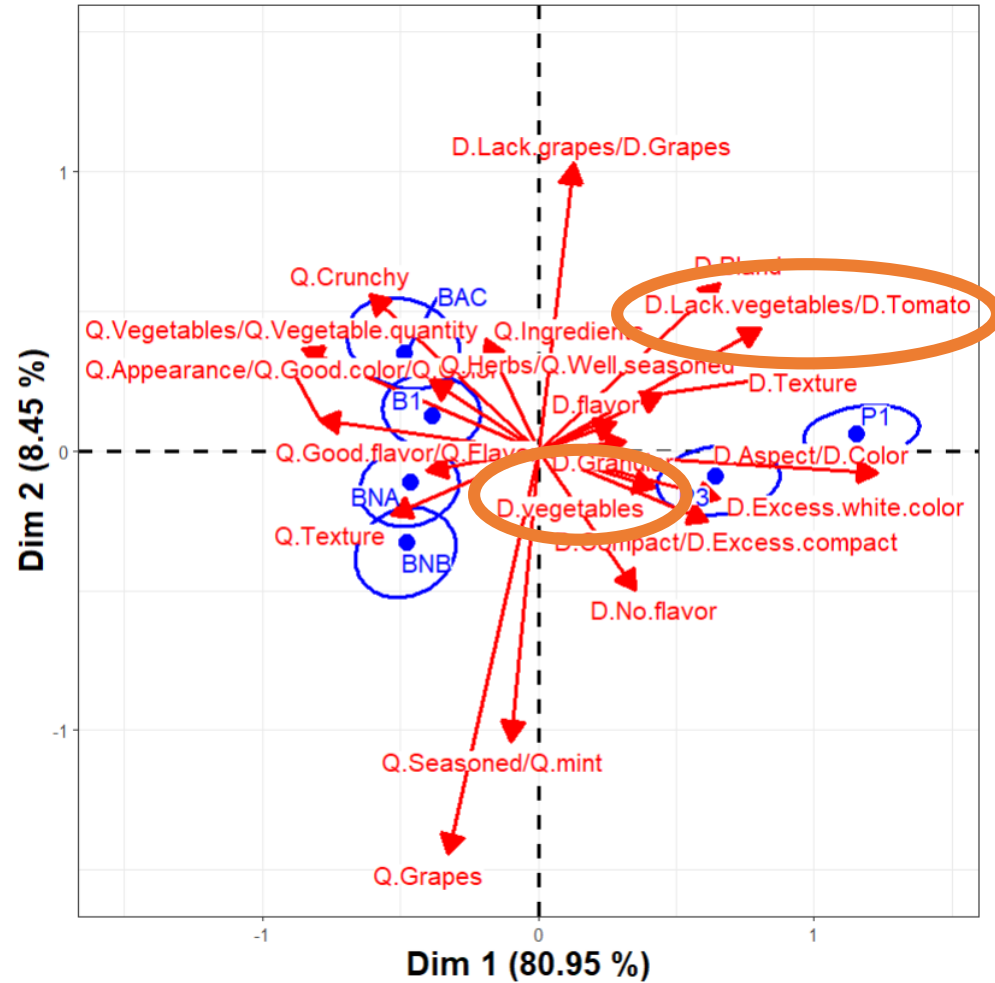
➔ Ps show color lightness defect ←

# V. Application : automated vs manual encoding process

## Automated encoding process

## Manual encoding process

Only descriptors having an association with at least one product are shown



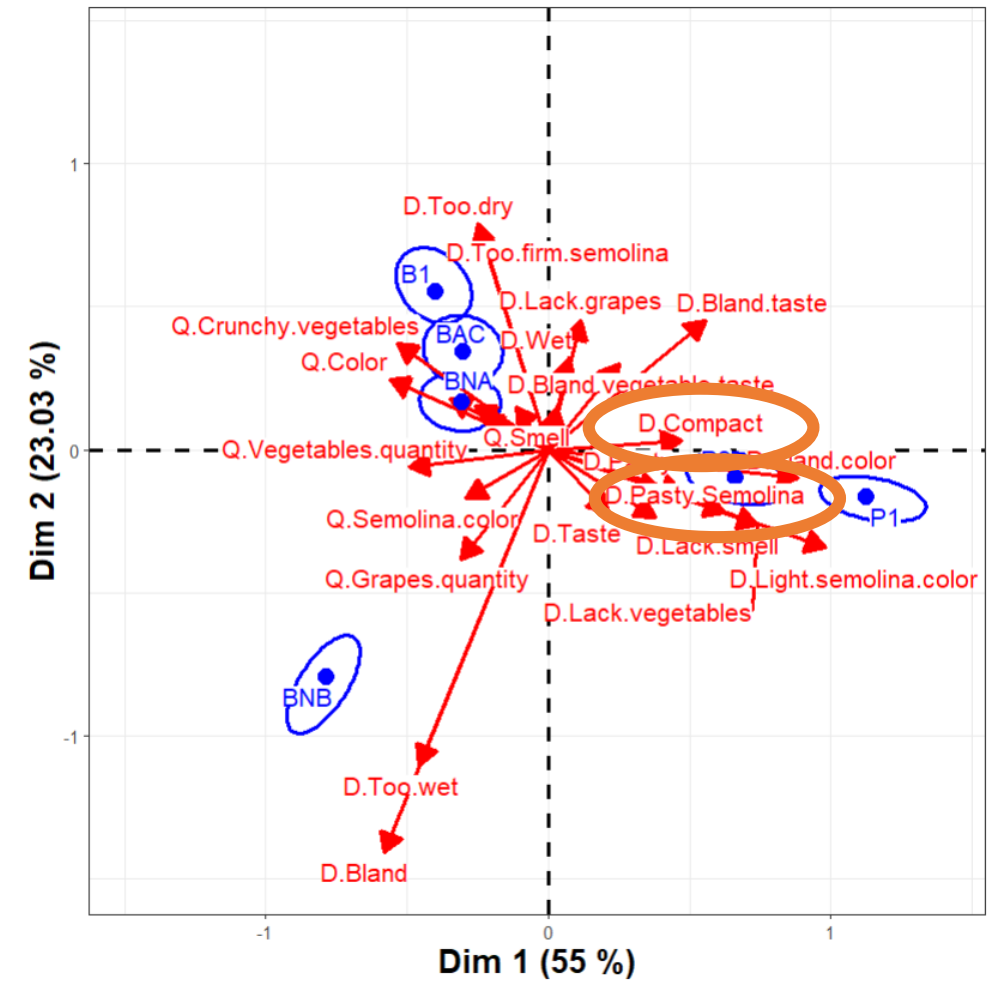
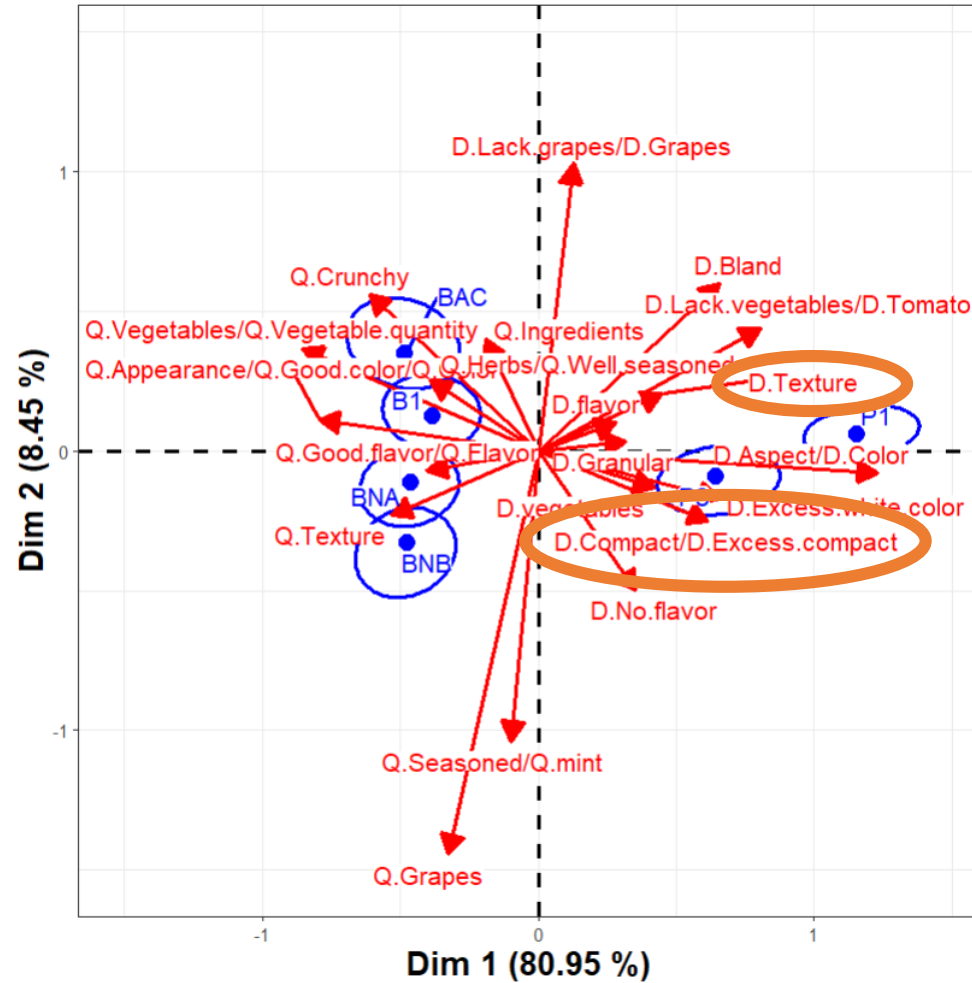
➤ Same first dimension  
 → Ps show vegetables defects ←

# V. Application : automated vs manual encoding process

## Automated encoding process

## Manual encoding process

Only descriptors having an association with at least one product are shown



➤ Same first dimension

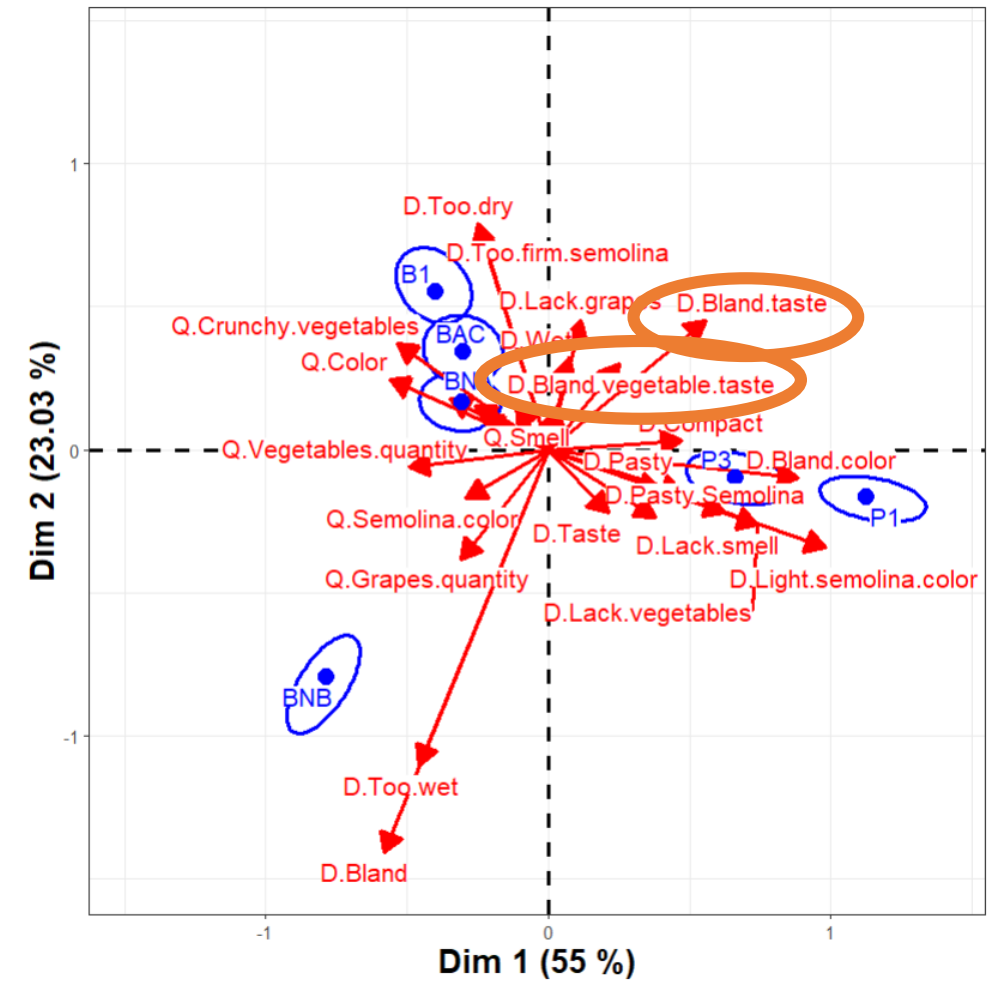
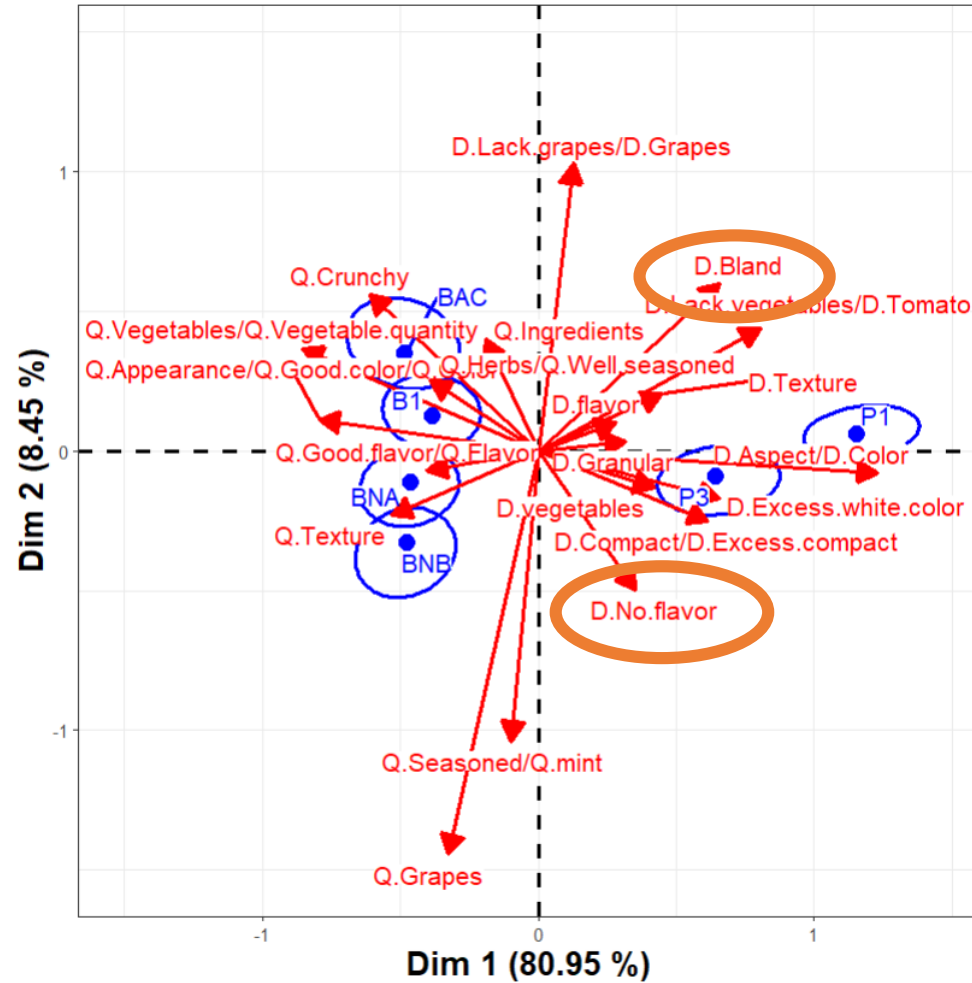
➔ Ps show compactness defects ←

# V. Application : automated vs manual encoding process

## Automated encoding process

## Manual encoding process

Only descriptors having an association with at least one product are shown



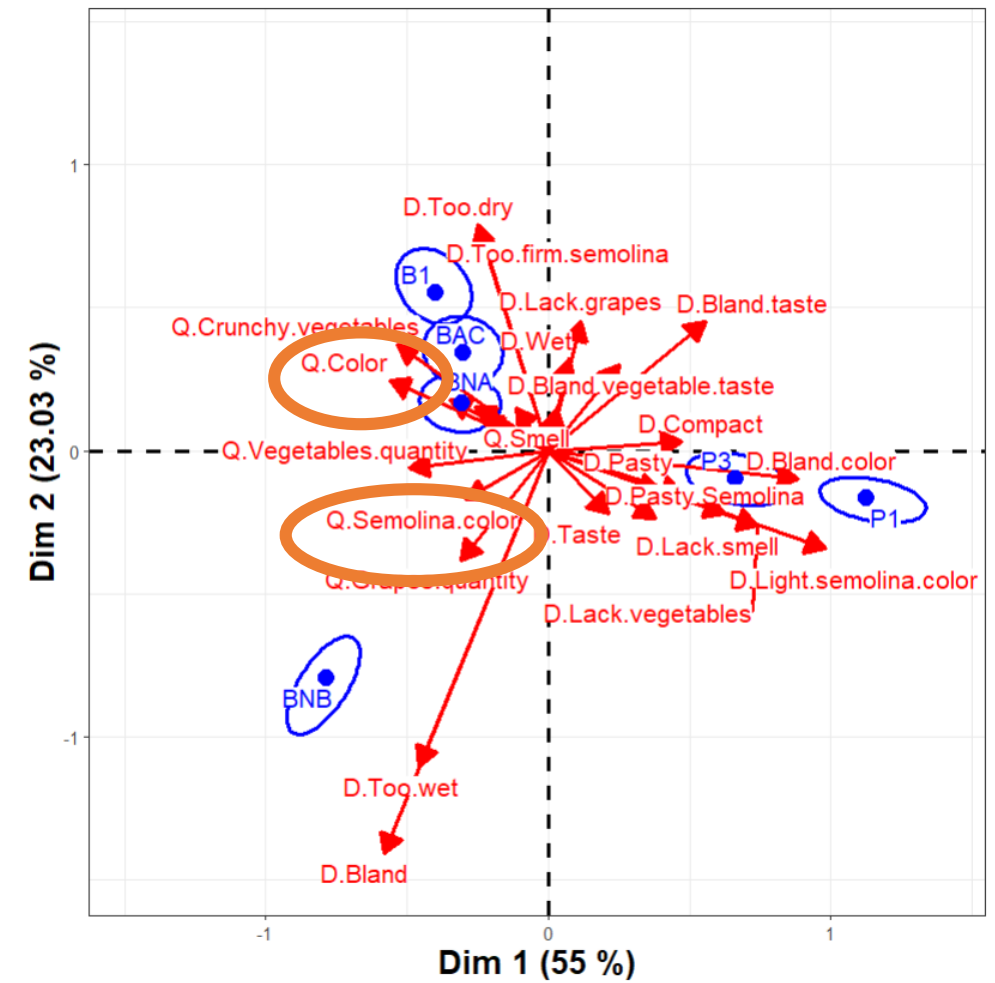
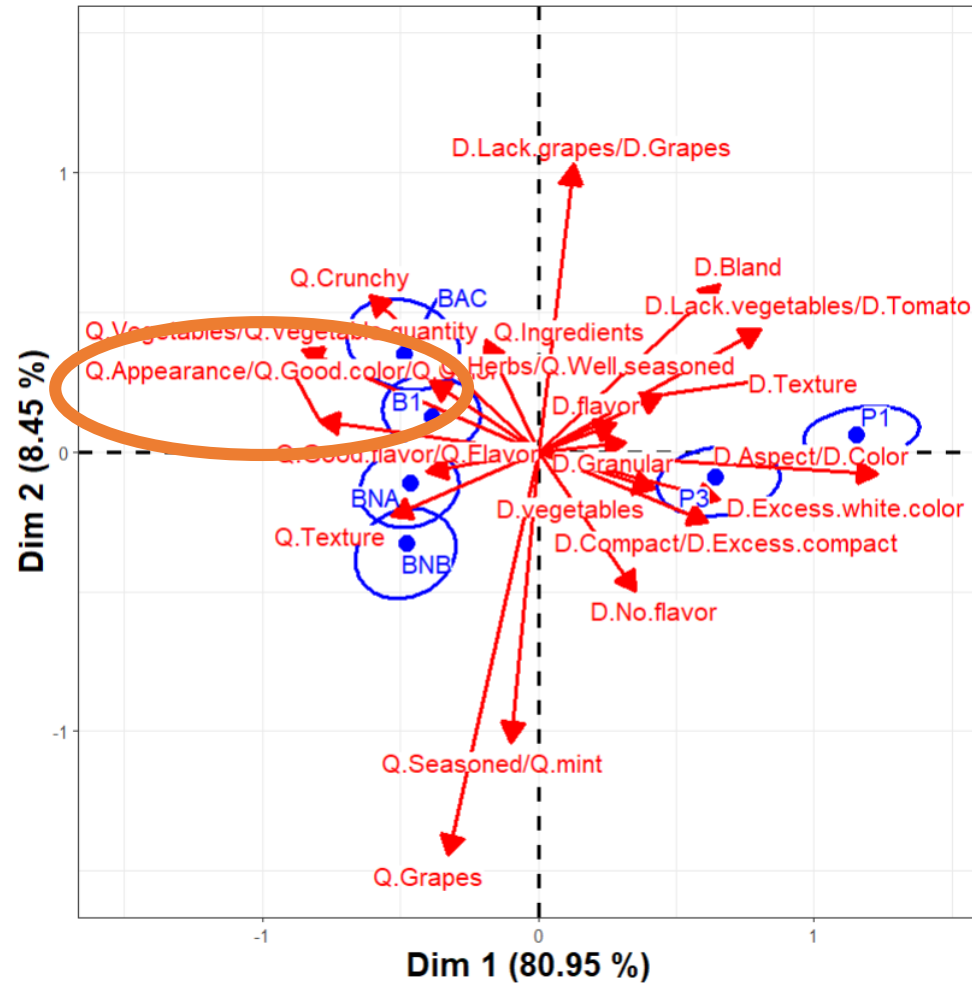
➤ Same first dimension  
 → Ps show blandness defects ←

# V. Application : automated vs manual encoding process

Automated encoding process

Manual encoding process

Only descriptors having an association with at least one product are shown



➤ Same first dimension  
→ Bs show appearance/color qualities ←

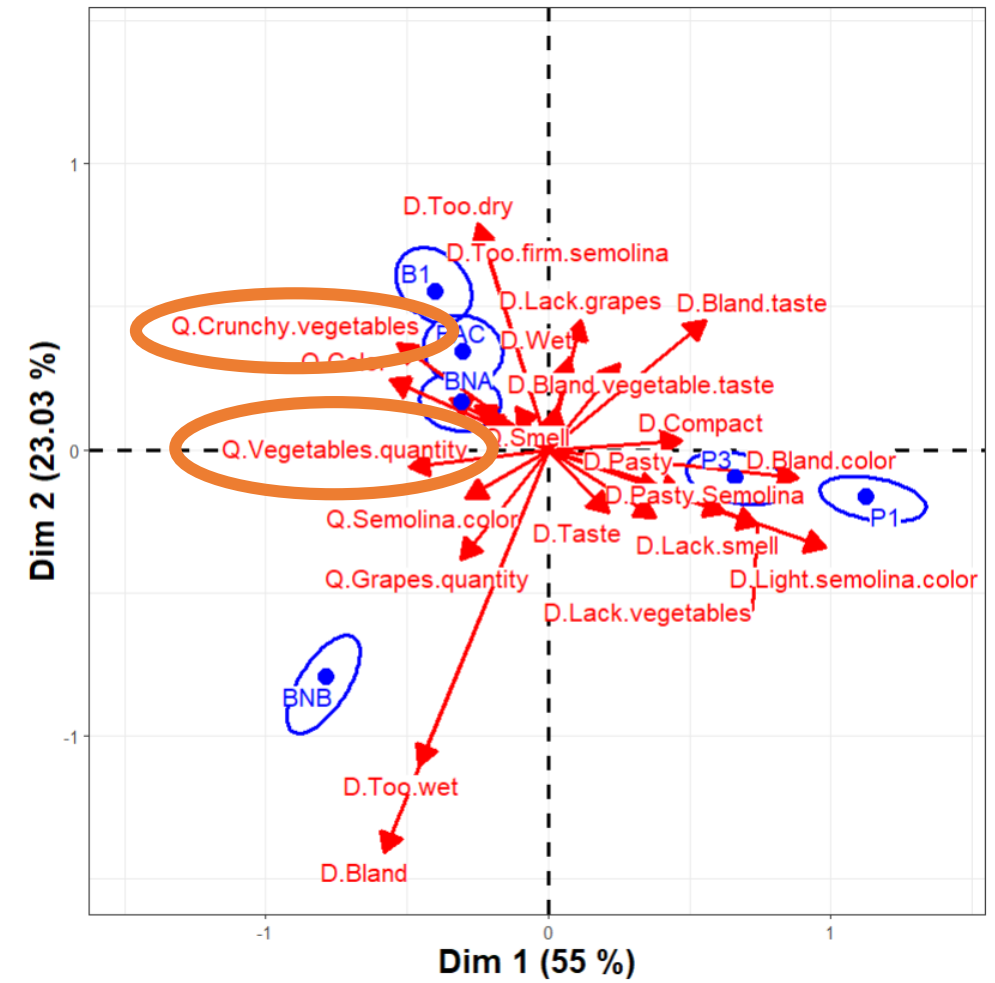
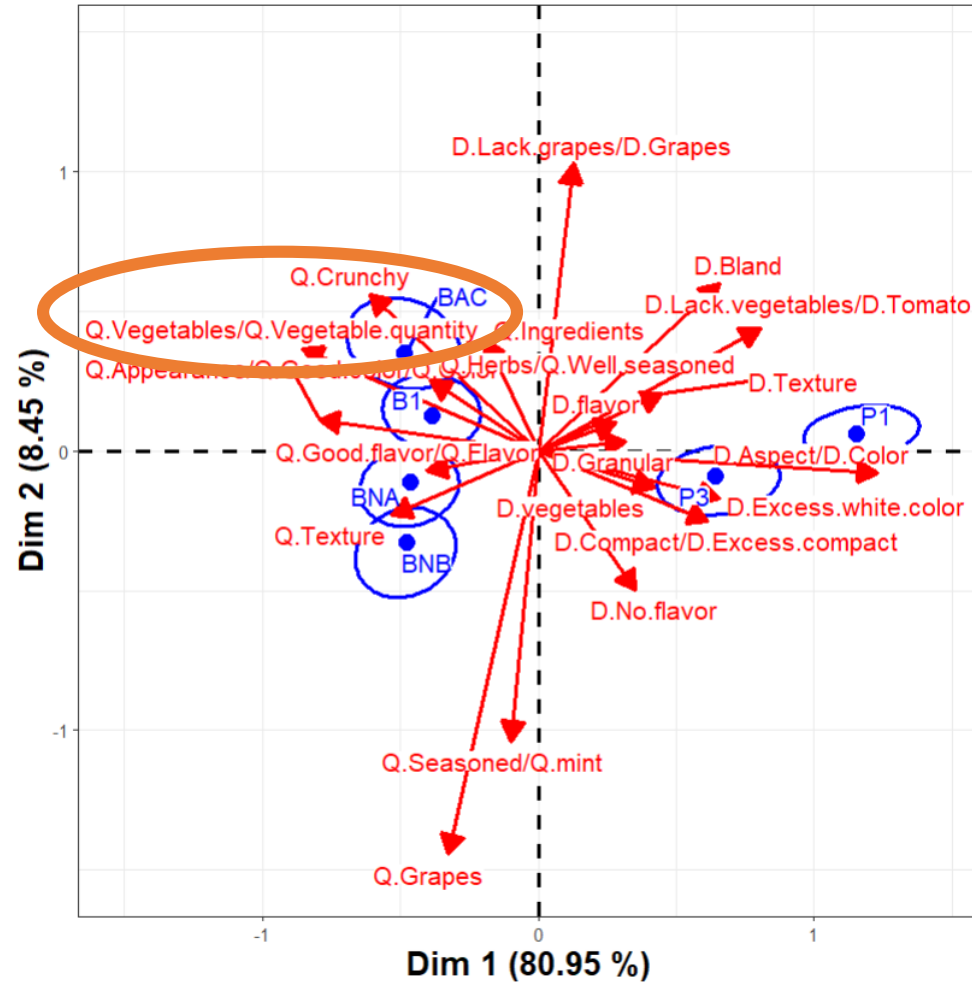


# V. Application : automated vs manual encoding process

Automated encoding process

Manual encoding process

Only descriptors having an association with at least one product are shown



➤ Same first dimension

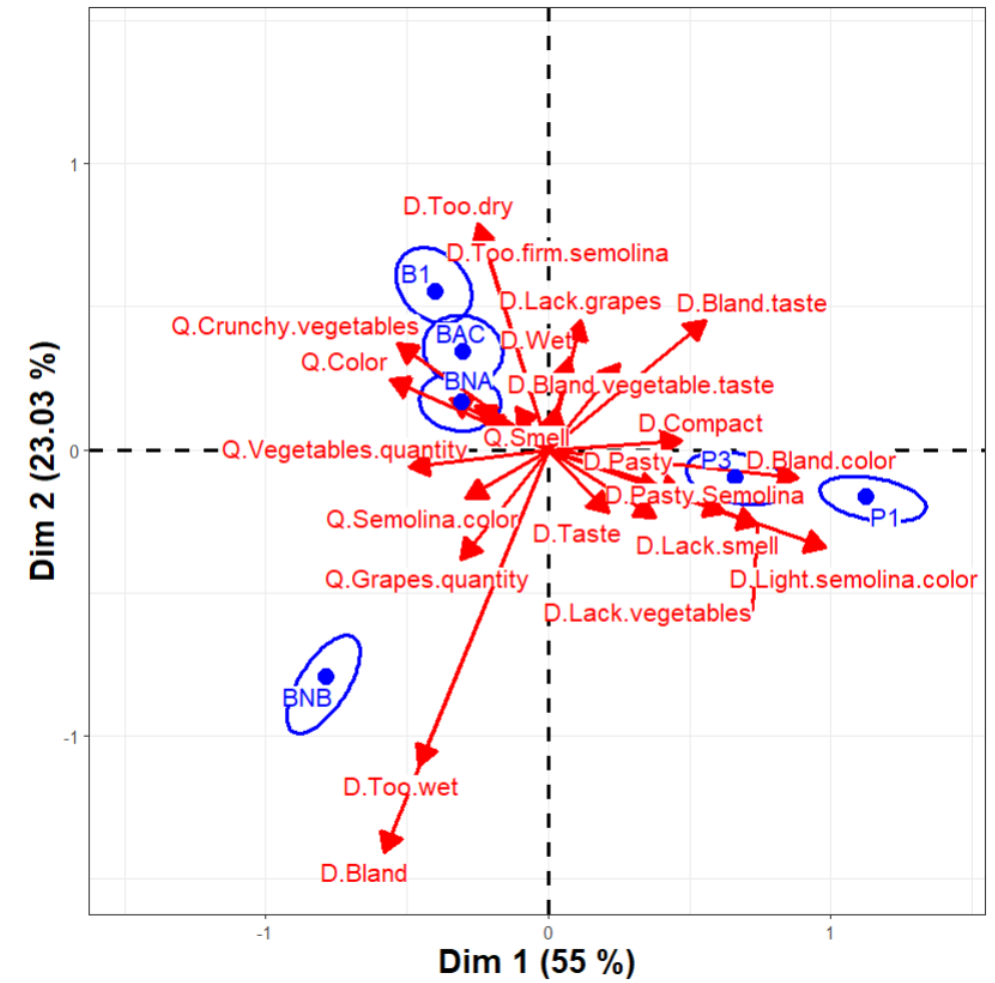
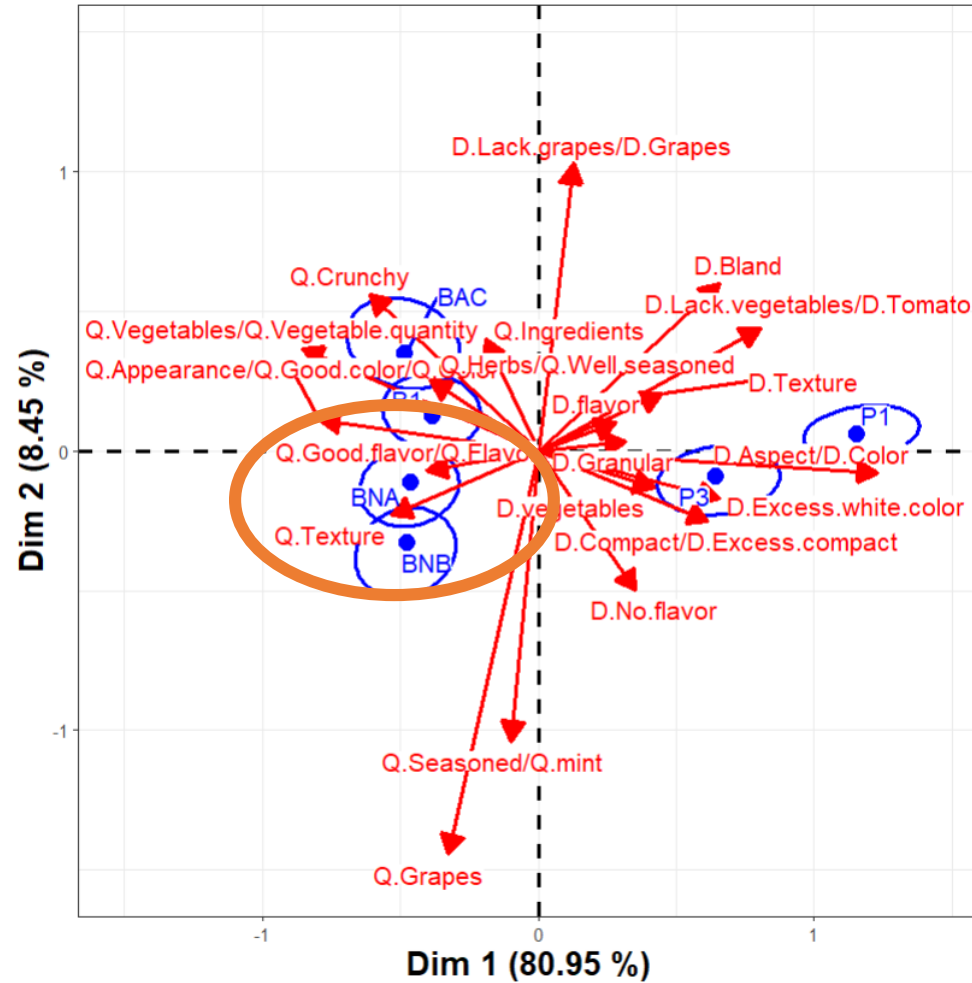
➔ Bs show vegetables quantity and crunchiness qualities ←

# V. Application : automated vs manual encoding process

Automated encoding process

Manual encoding process

Only descriptors having an association with at least one product are shown



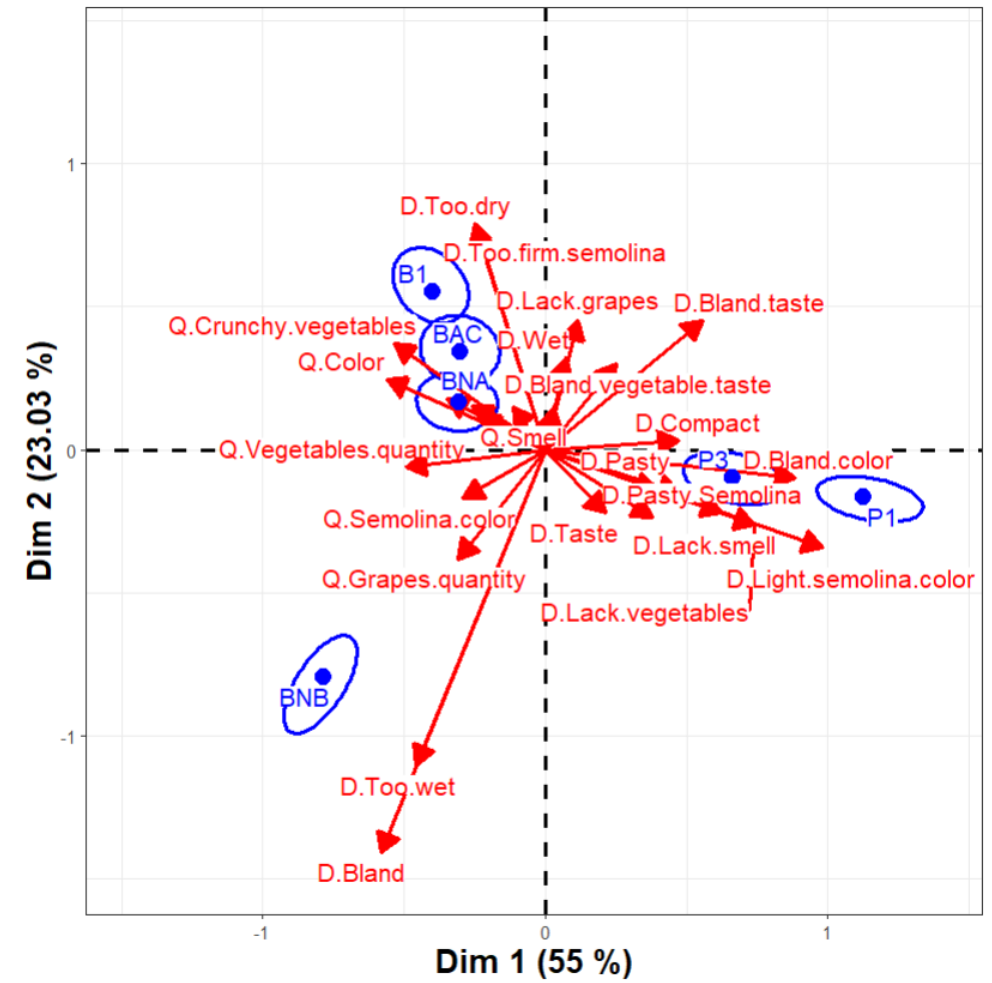
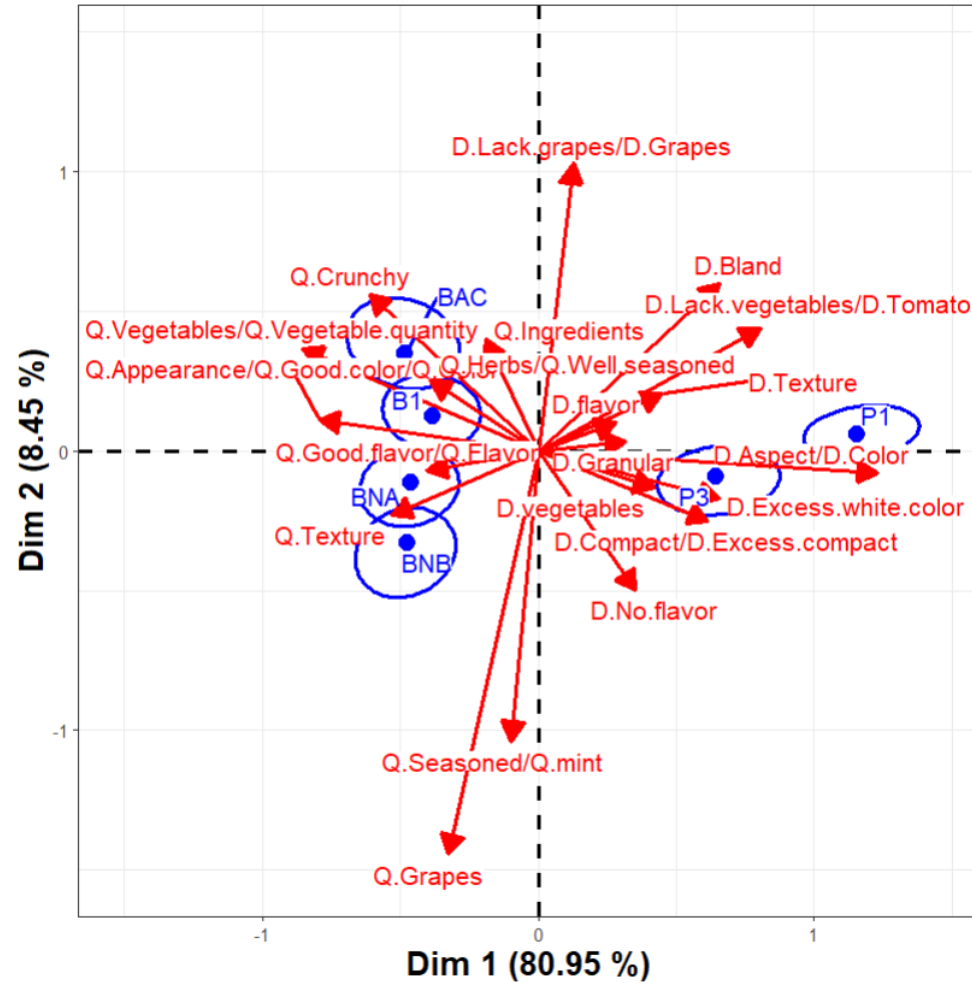
➤ A additional flavor and texture quality information on the first dimension with automation

# V. Application : automated vs manual encoding process

Automated encoding process

Manual encoding process

Only descriptors having an association with at least one product are shown



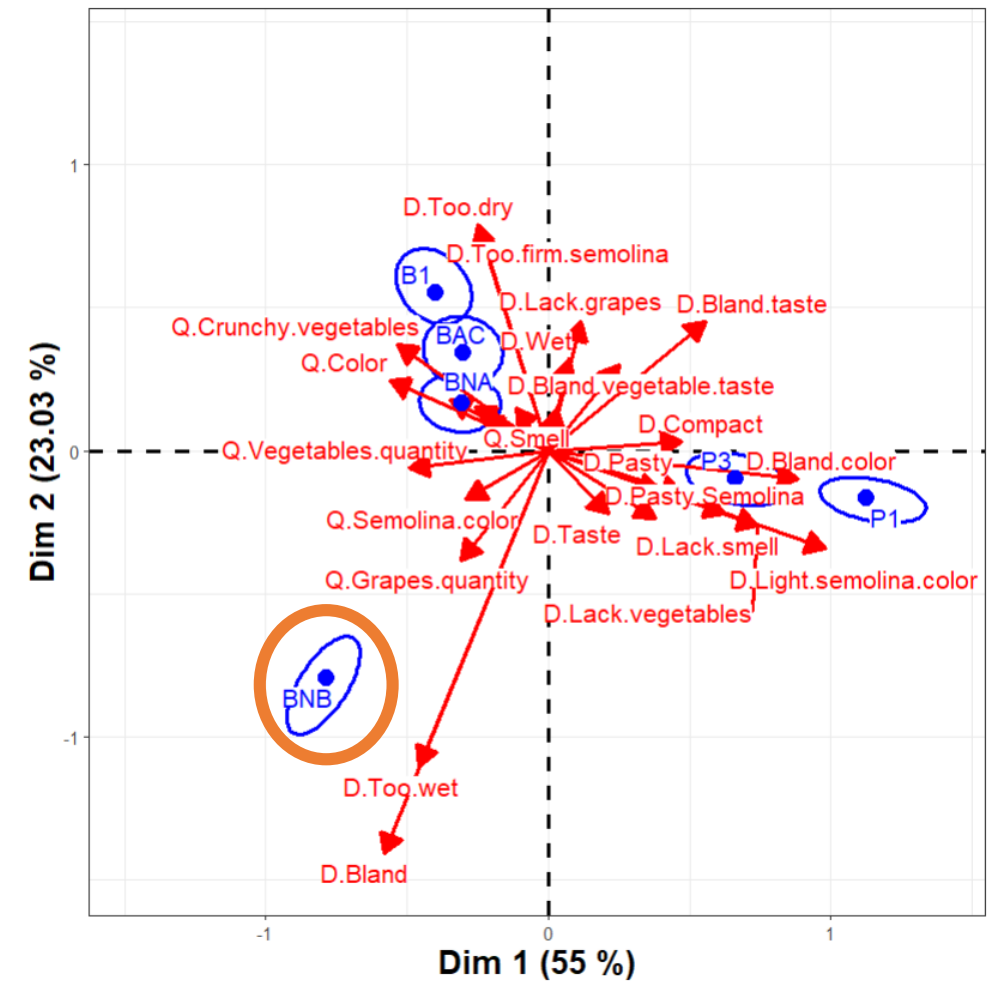
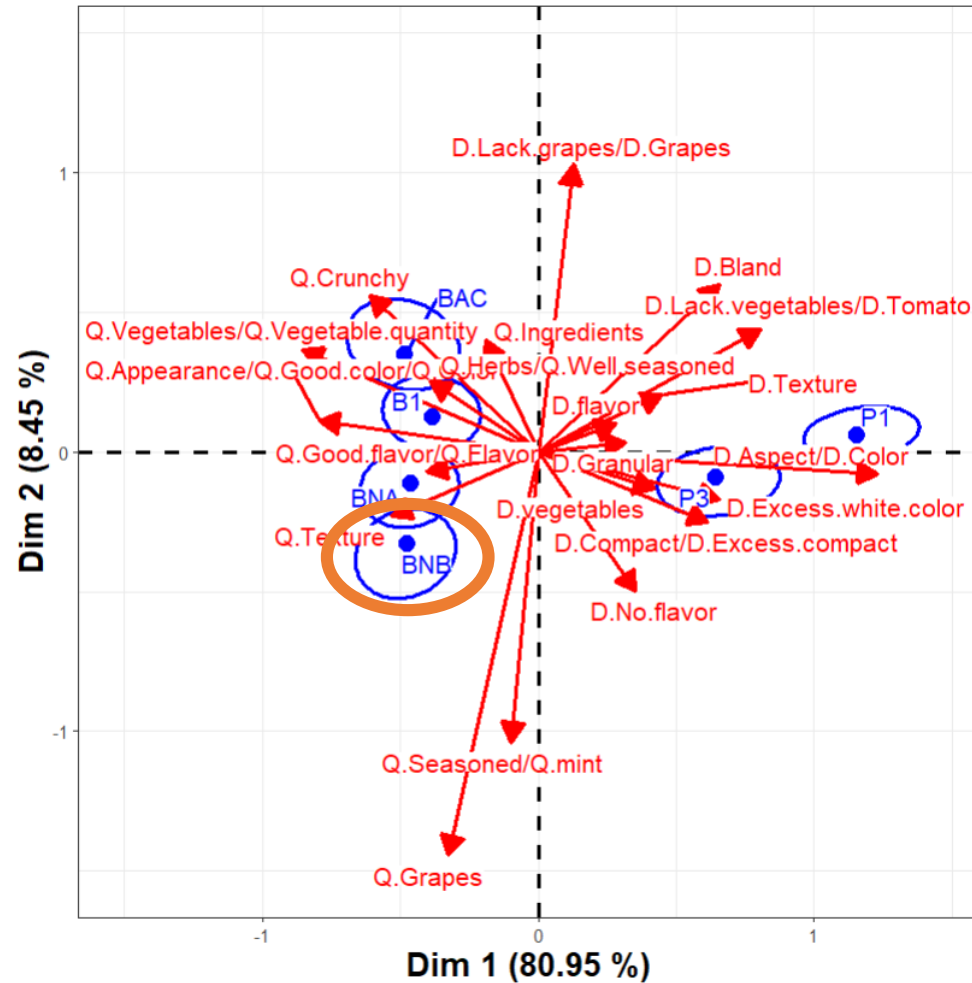
➤ Differences on second dimension

# V. Application : automated vs manual encoding process

## Automated encoding process

## Manual encoding process

Only descriptors having an association with at least one product are shown



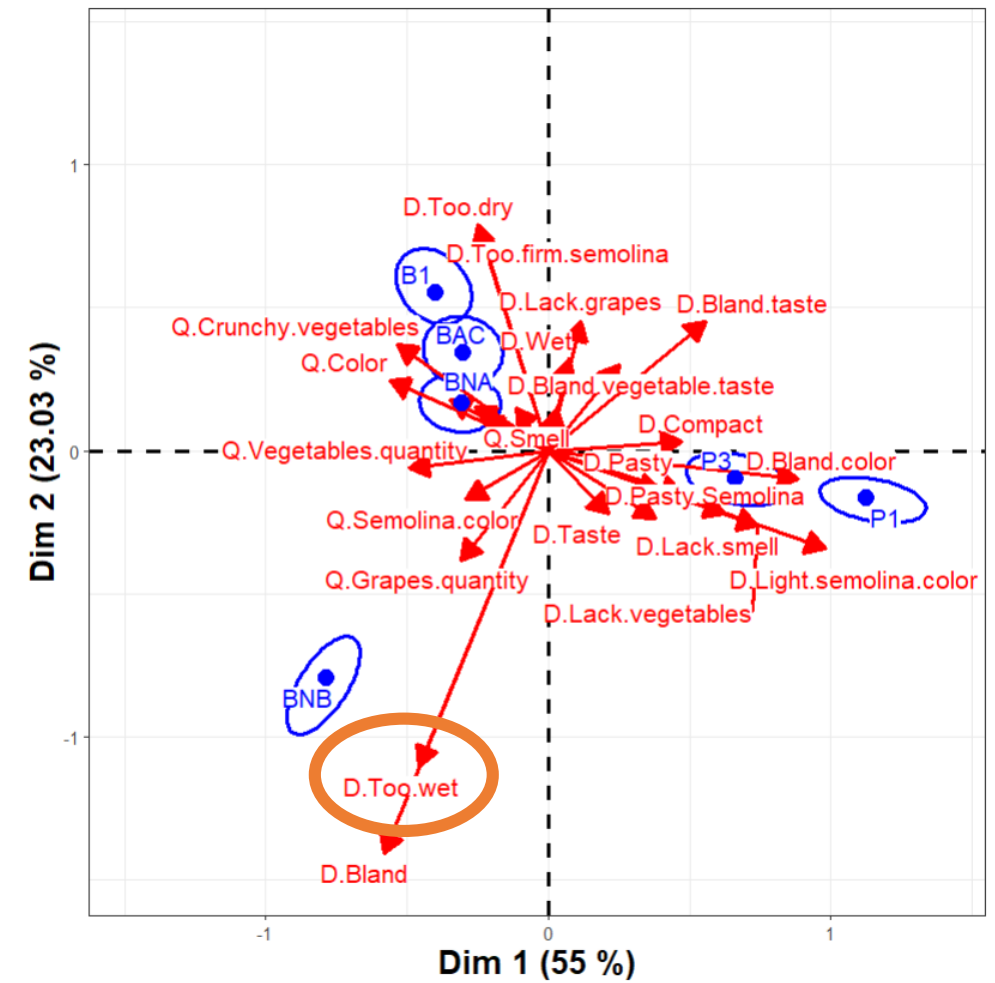
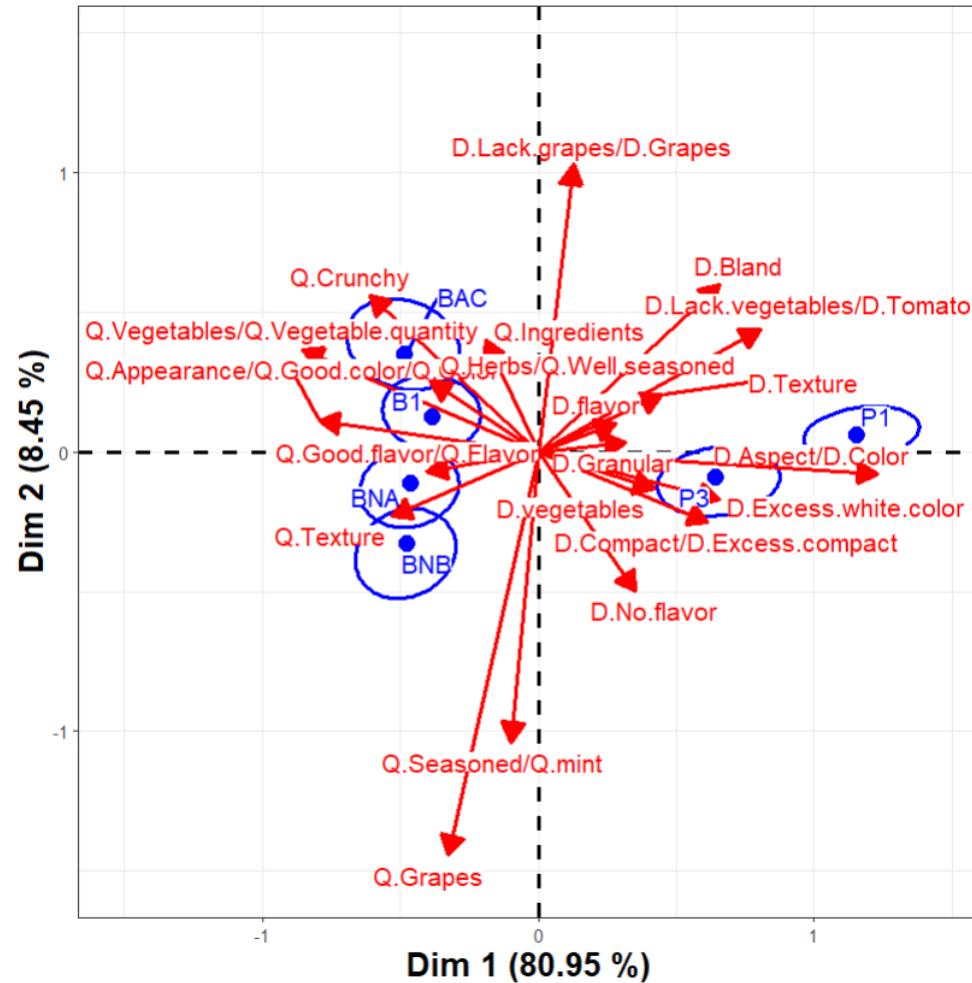
➤ Differences on second dimension  
 → BNB less “particular” with automation ←

# V. Application : automated vs manual encoding process

Automated encoding process

Manual encoding process

Only descriptors having an association with at least one product are shown



➤ Differences on second dimension

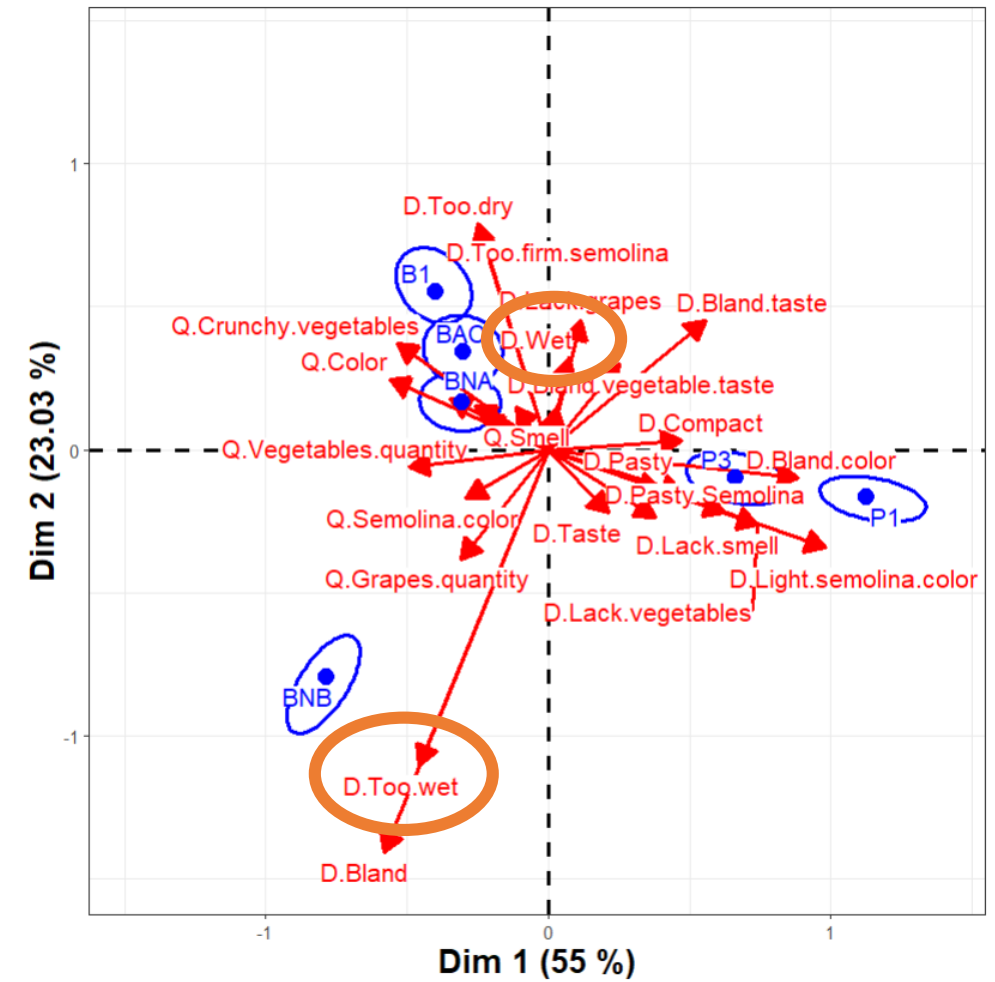
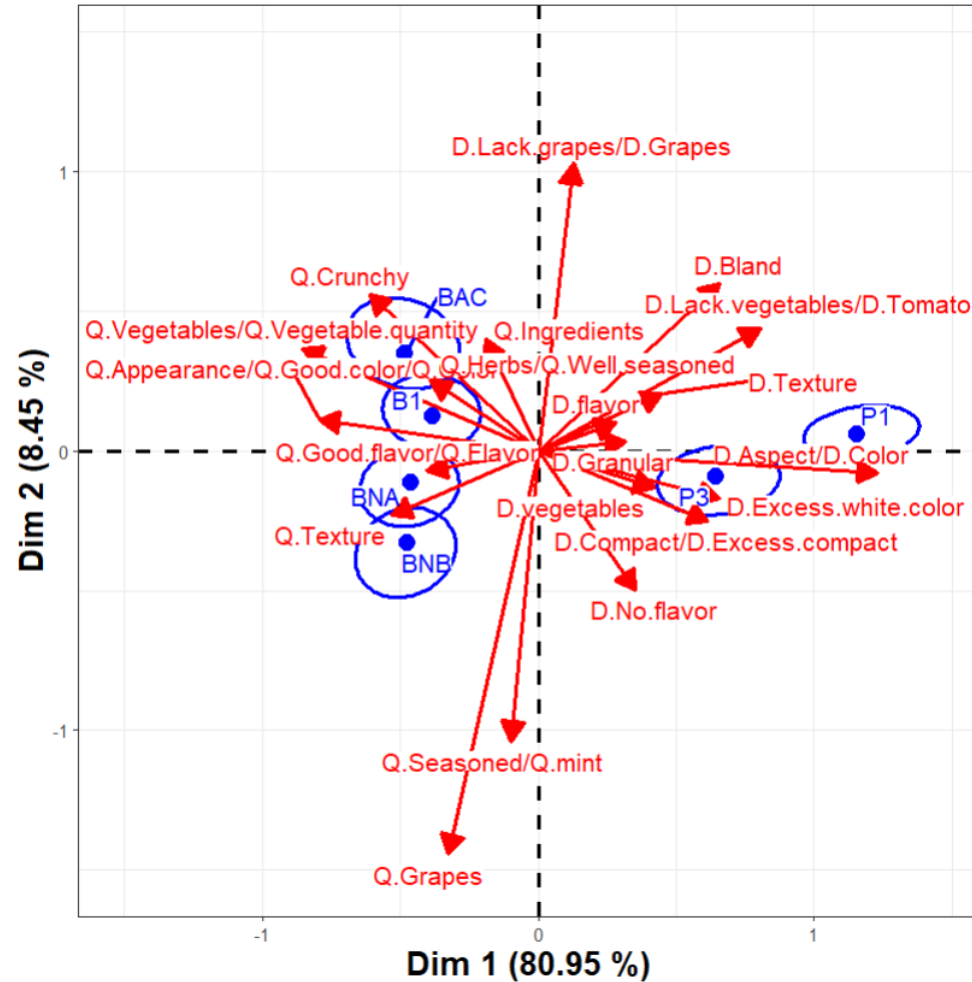
➔ Lack wet defect information with automation ➔

# V. Application : automated vs manual encoding process

## Automated encoding process

## Manual encoding process

Only descriptors having an association with at least one product are shown



➤ Differences on second dimension

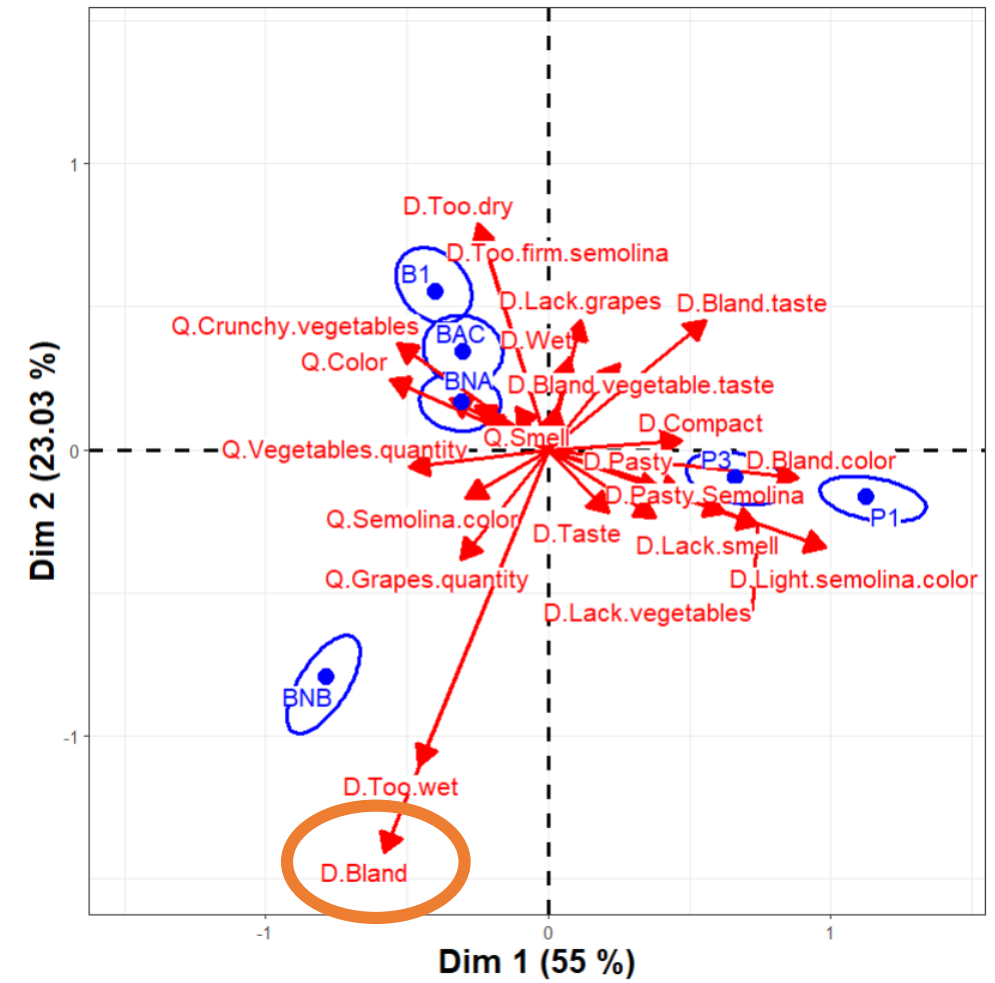
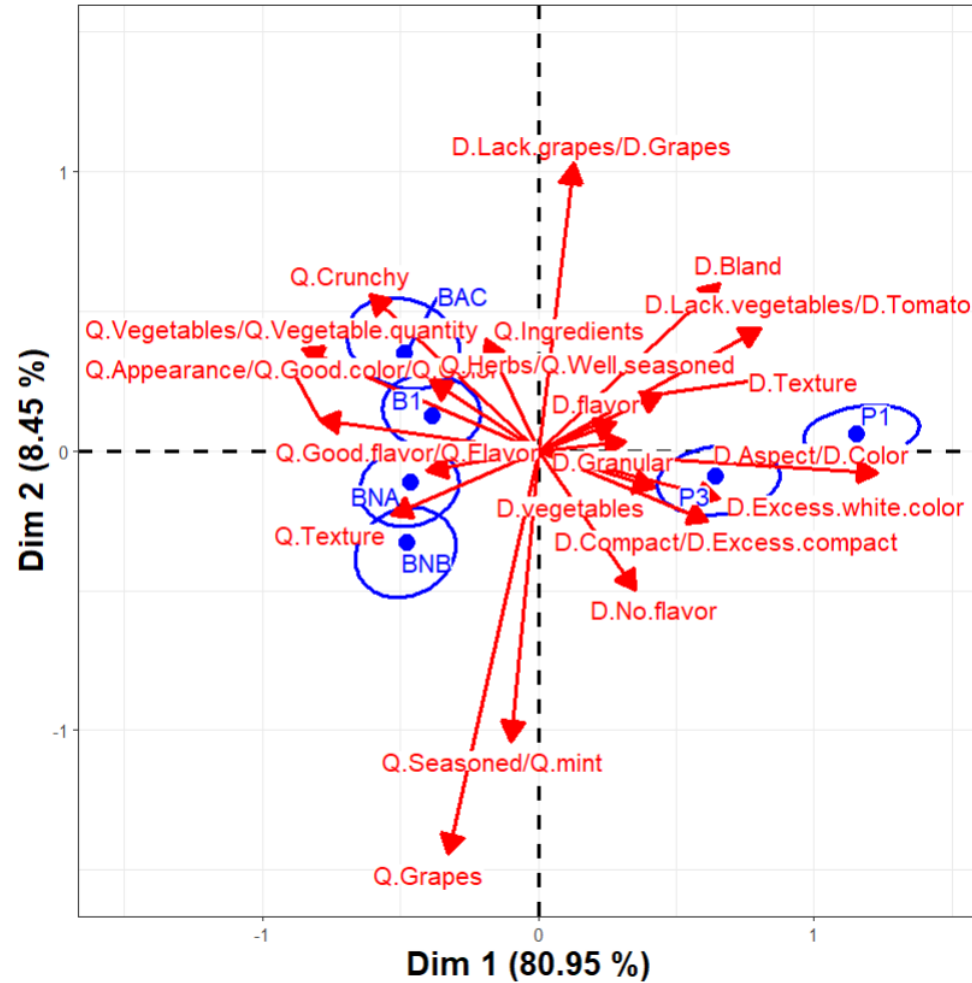
➔ Lack wet defect information with automation ← **BUT...**

# V. Application : automated vs manual encoding process

## Automated encoding process

## Manual encoding process

Only descriptors having an association with at least one product are shown



➤ Differences on second dimension

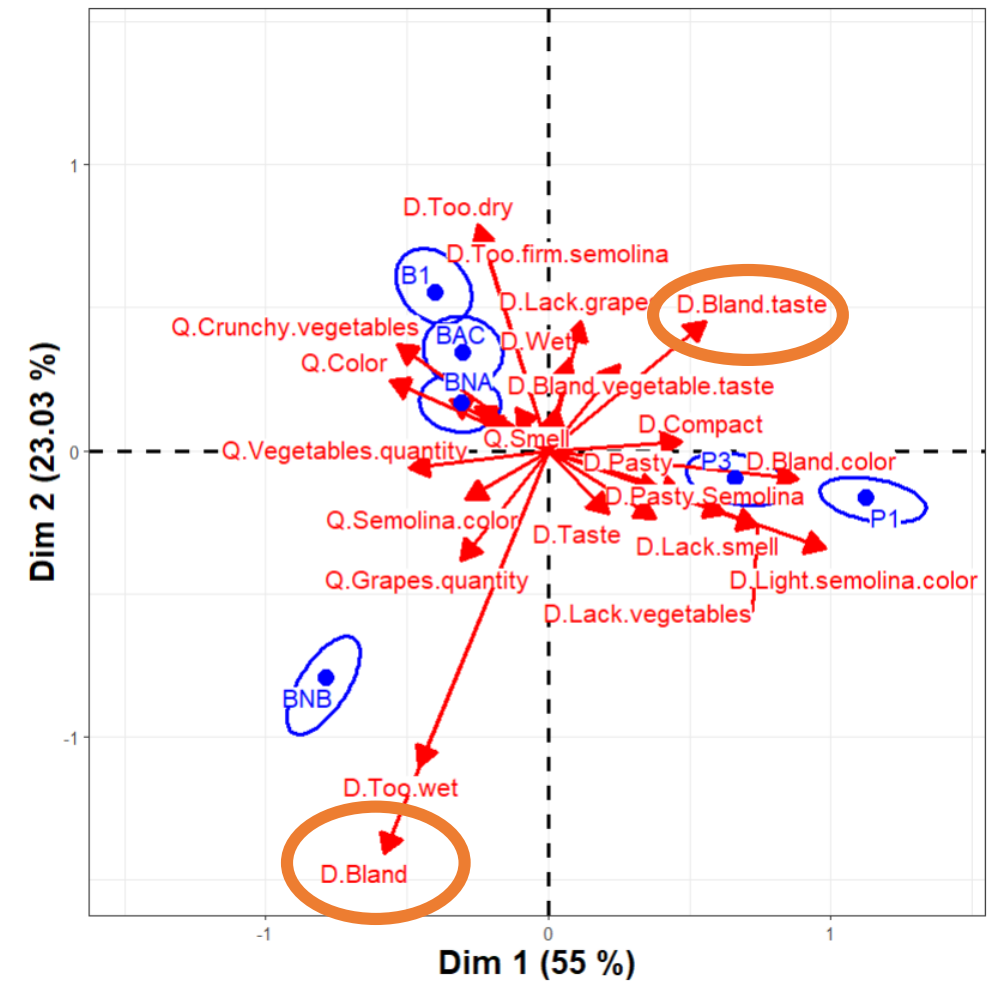
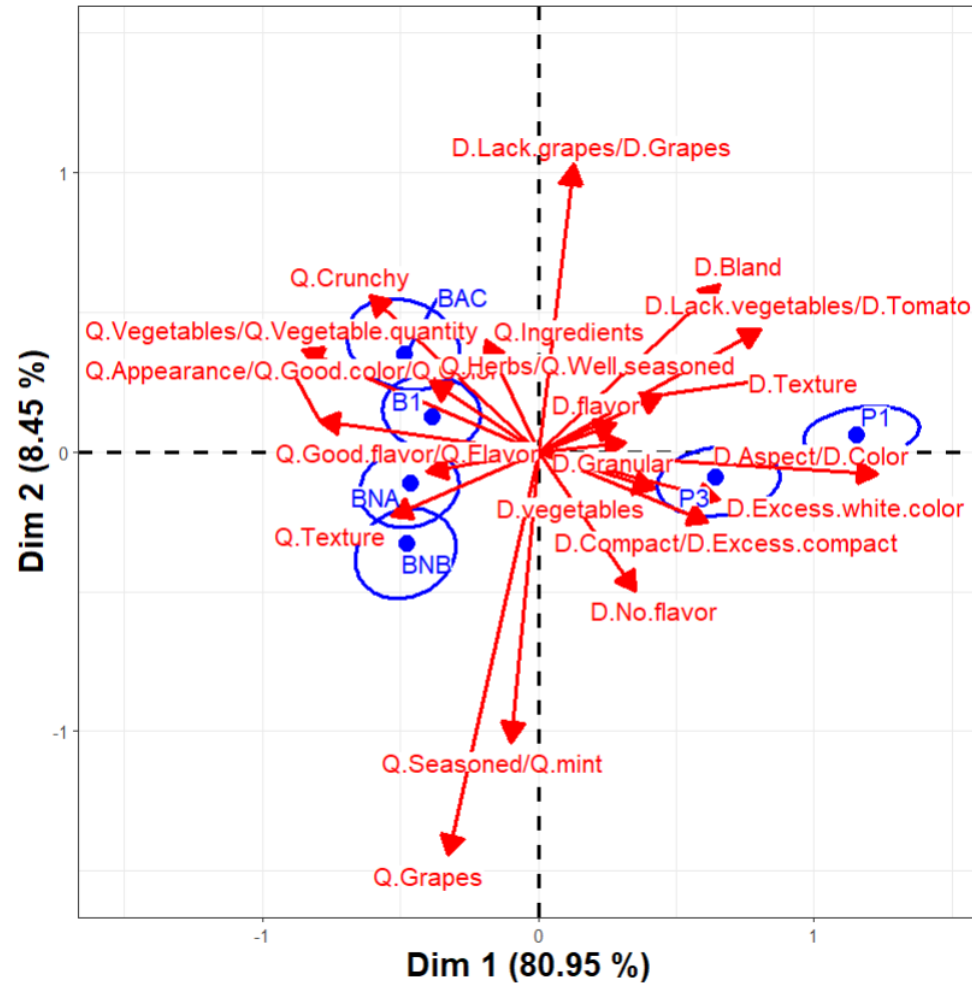
➔ Lack bland defect information with automation ➔

# V. Application : automated vs manual encoding process

## Automated encoding process

## Manual encoding process

Only descriptors having an association with at least one product are shown



➤ Differences on second dimension

➔ Lack bland defect information with automation ← BUT...

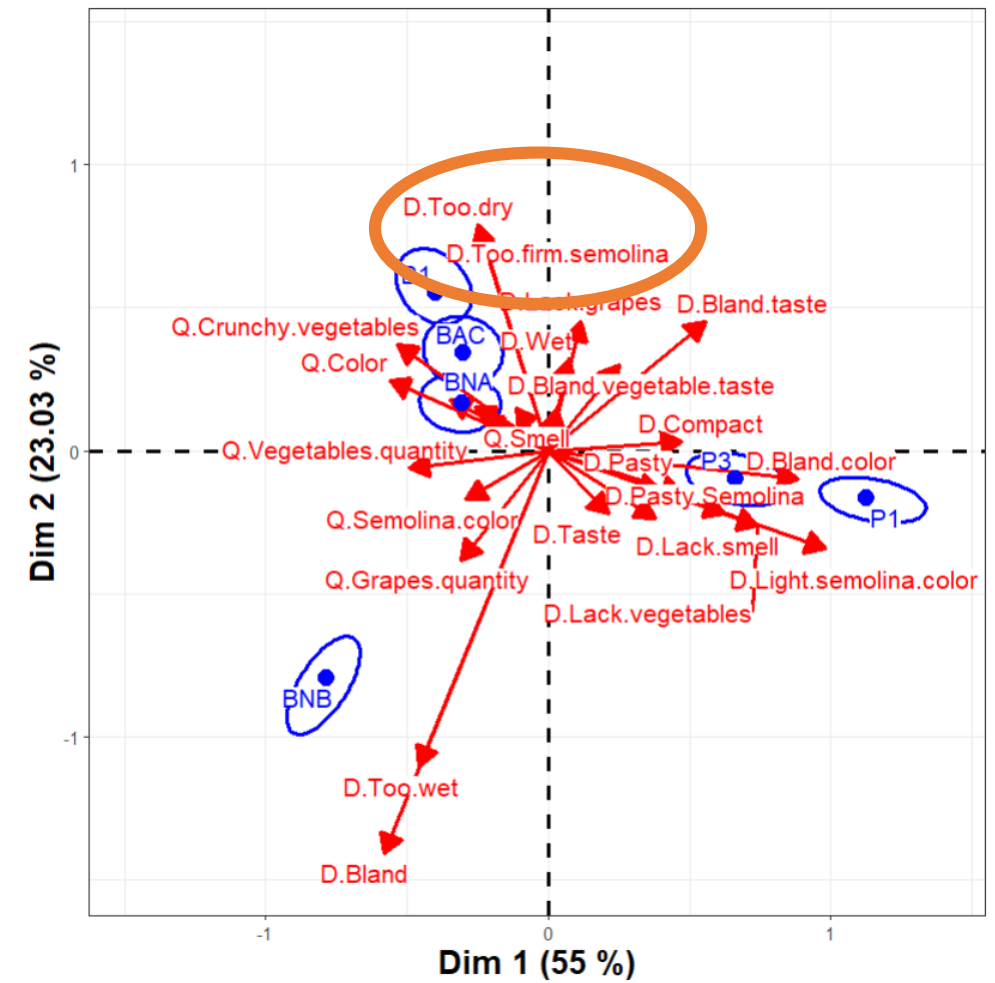
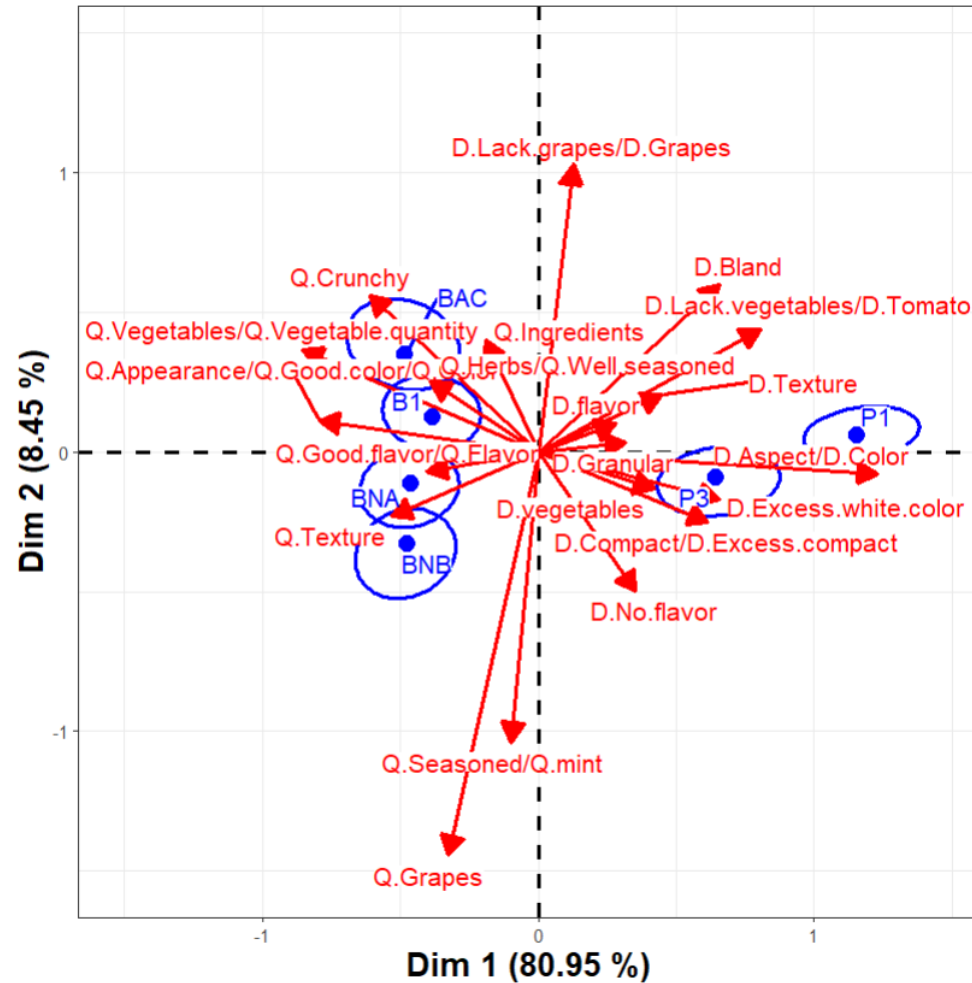


# V. Application : automated vs manual encoding process

Automated encoding process

Manual encoding process

Only descriptors having an association with at least one product are shown



➤ Differences on second dimension

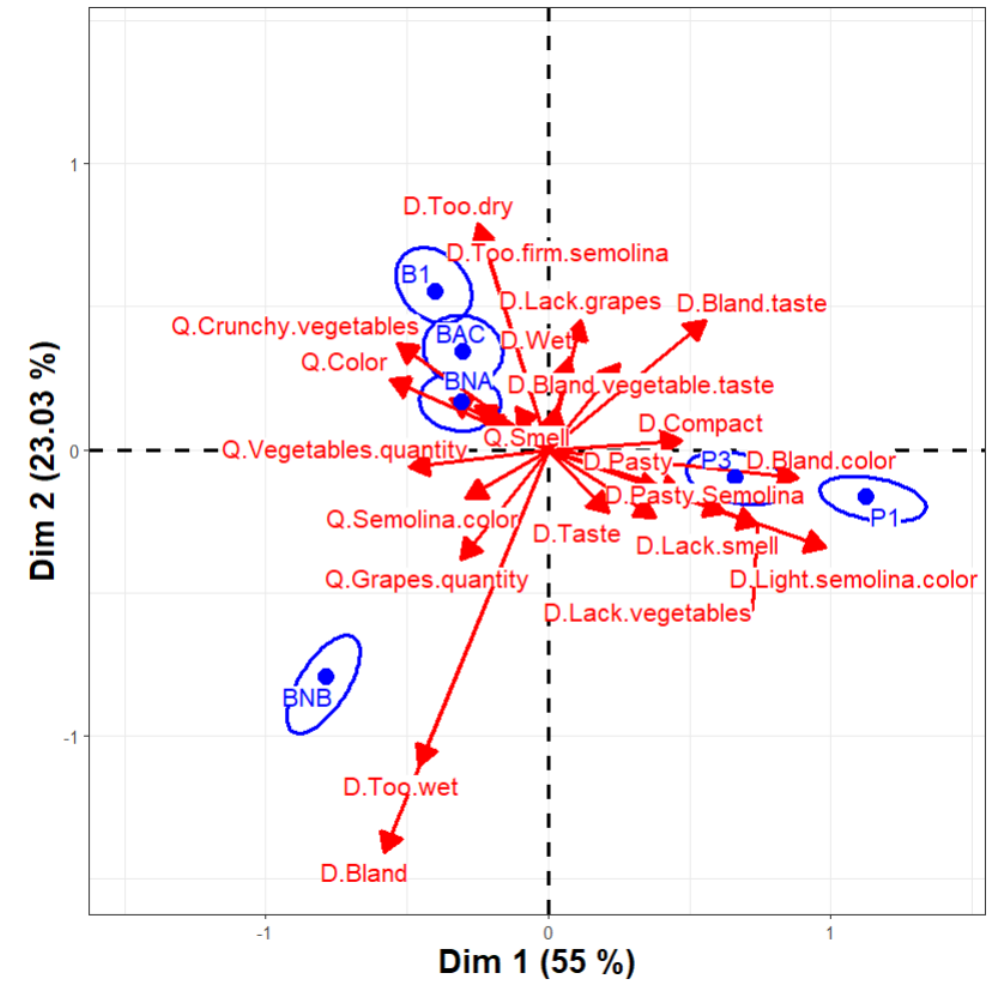
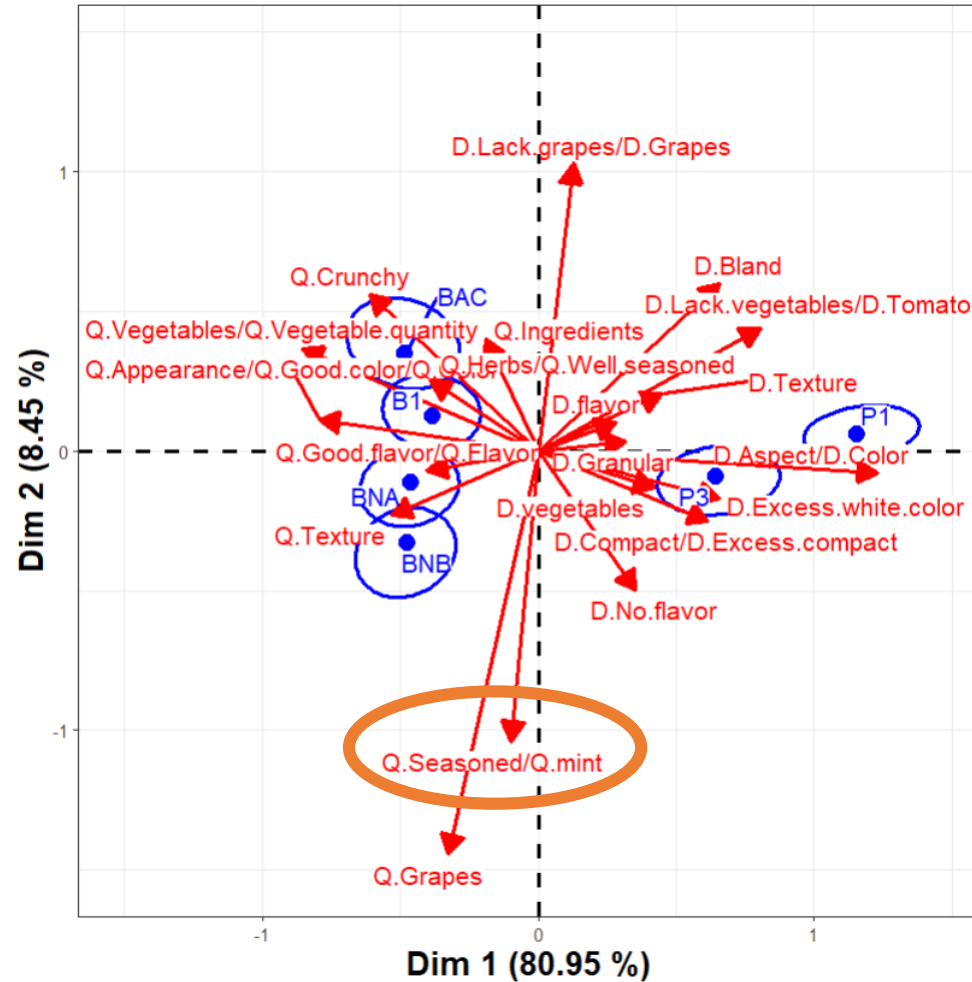
➔ Lack too dry/firm defect information with automation ➔ (without “but”)

# V. Application : automated vs manual encoding process

## Automated encoding process

## Manual encoding process

Only descriptors having an association with at least one product are shown



➤ Differences on second dimension

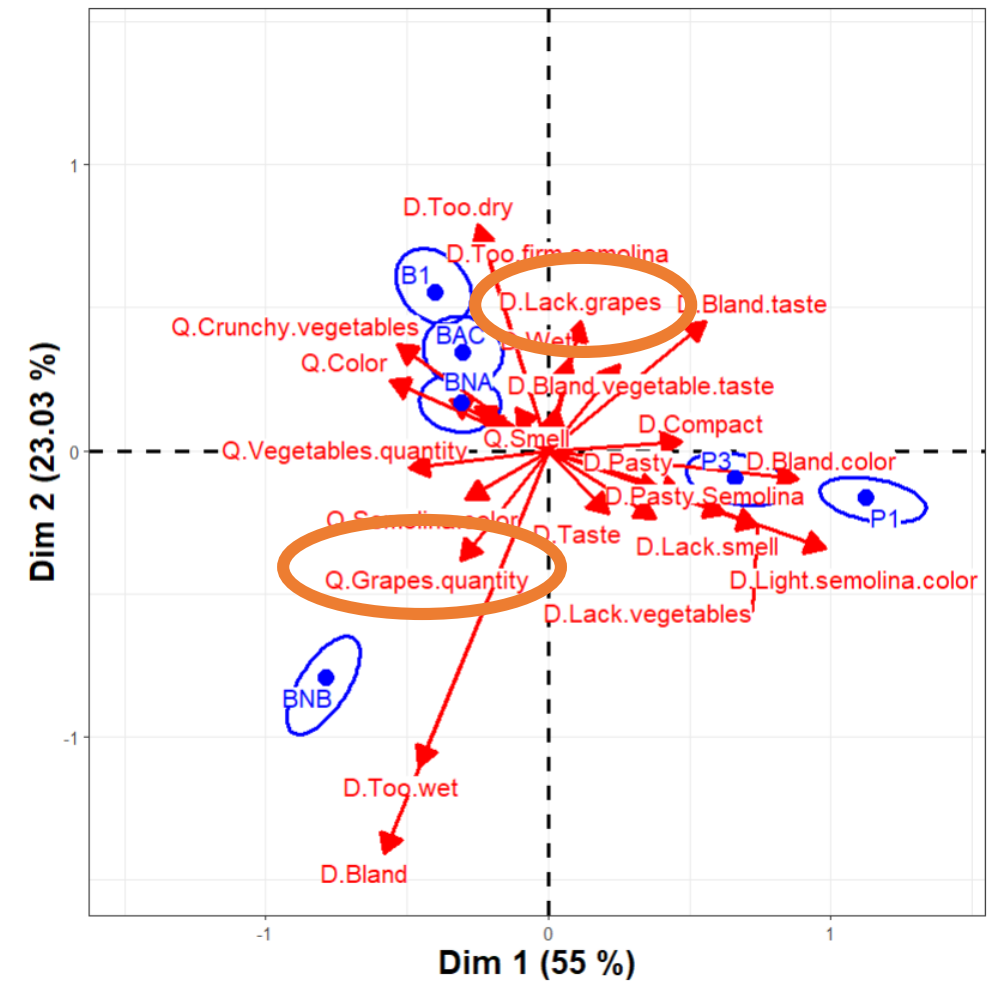
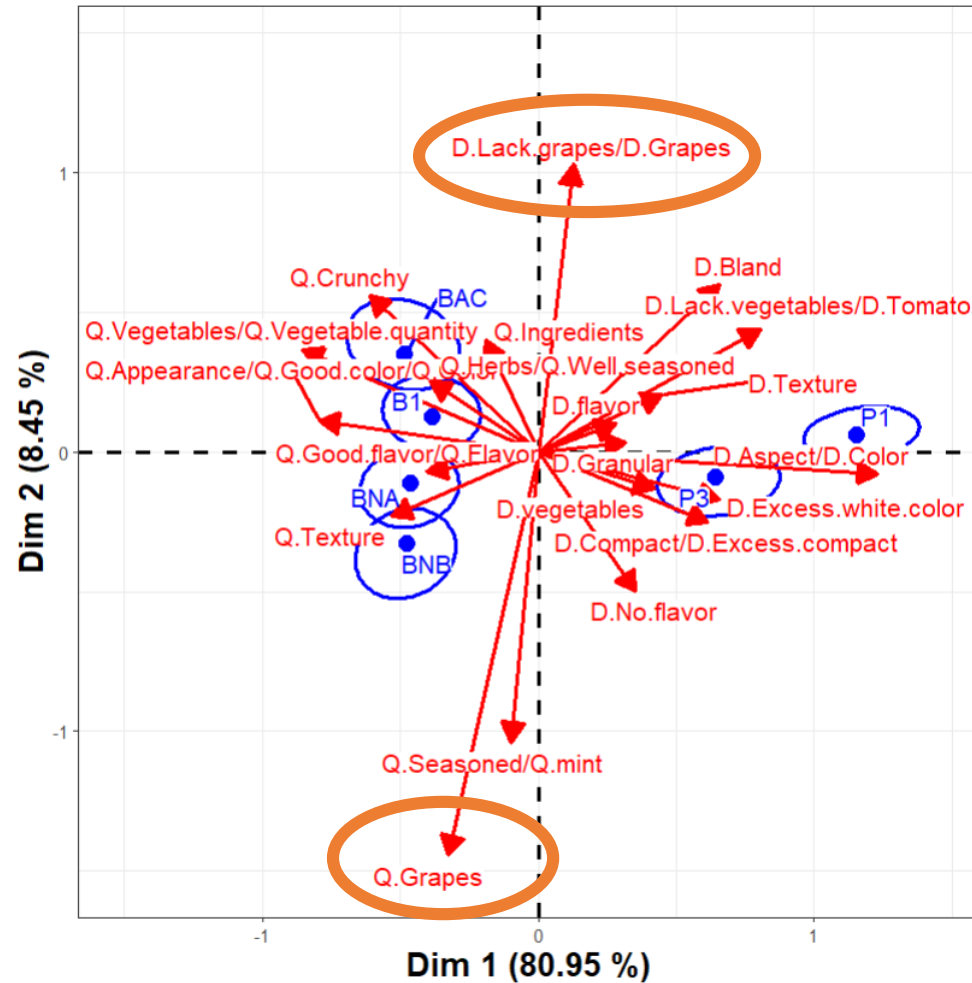
➔ Lack seasoning quality information with manual coding ➜ (still without "but")

# V. Application : automated vs manual encoding process

Automated encoding process

Manual encoding process

Only descriptors having an association with at least one product are shown



➤ Similarities on second dimension

→ Grapes qualities on negative part and defects on positive parts ←

# Overview

- I. The collected data and their encoding for statistical analysis
- II. Cleaning information and extracting sensory descriptors
- III. Semantic-based hierarchical structuration of extracted descriptors
- IV. Aggregation of descriptors: semantic-constrained classification
- V. Application: automated vs. manual encoding
- VI. Conclusion**

# Conclusion

“Old-school” approach

- **Expert system** rather than machine learning (no need for large datasets to train AI)
- (Relatively) **easy to implement**
- **More control on extracted data** compared to pretrained models for natural language such as ChatGPT

Limitations (WIP)

- Only in French (at the moment)
- Issues with some contexts not always correctly handled
- Issues of classification with attributes having several meanings (e.g. sweet)
- **Some loss of information** compared to manual preprocessing

Promising results

- Computation time **about 5 minutes** for a dataset (one full day for the panel leader)
- **Reproducible data preprocessing**
- **Combination of semantics and statistics with no human intervention**

Will be available as a **web app** free for non-commercial use

# Acknowledgements

Cécile MURSIC

Techni'*Sens*

Matthieu CAMILLERI

Ronan SYMONEAUX

