

# Visualisation de données avec FactoMineR

François Husson

<https://husson.github.io>

Unité de mathématiques appliquées, Agrocampus Ouest, Rennes

Webinaire tuto@mate – 1er avril 2019

## Présentation

- Recherche : analyse de données, tableaux multiples, données manquantes
- Enseignement : cursus d'ingénieur, master *science des données*
- MOOC en analyse de données et MOOC en Sensométrie
- Formation continue : statistique avec R, analyse de données



2018



2nd ed: 2017

1st ed: 2011



2nd ed: 2016

1st ed: 2009



2nd ed: 2013

1st ed: 2005



2013



3rd ed: 2012

2nd ed: 2010

1st ed: 2008



2012

Packages:

FACTOMINER<sup>R</sup>

- miss

MDA

- SensoMine<sup>R</sup>

- Factoshiny -

FactoInvestigate - RcmdrPlugin.FactoMineR

# Plan

Présentation des méthodes

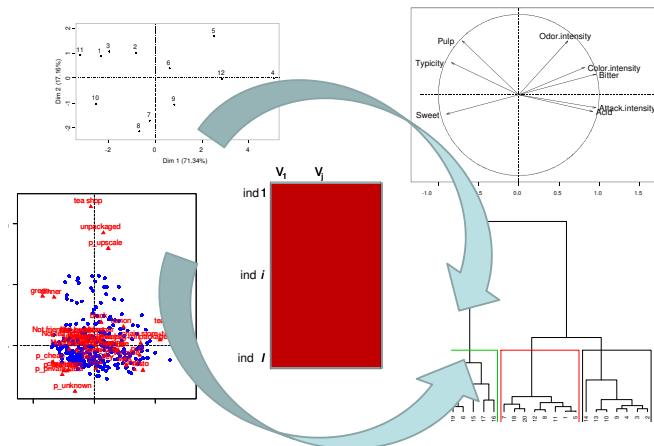


Factoshiny pour l'interface et les graphes interactifs

FactoInvestigate pour des rapports automatisés

Conclusion

# Les méthodes d'analyse de données

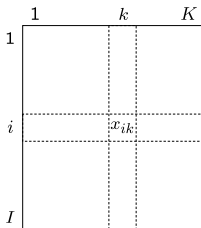


## Objectifs :

- Descriptif - exploratoire : visualisation de données
- Synthèse - résumé de grands tableaux individus  $\times$  variables

# L'analyse en Composantes Principales (ACP)

L'ACP s'intéresse à des tableaux de données rectangulaires avec des **individus** en lignes et des **variables quantitatives** en colonnes



# L'analyse en Composantes Principales (ACP)

L'ACP s'intéresse à des tableaux de données rectangulaires avec des **individus** en lignes et des **variables quantitatives** en colonnes

|     |   |          |     |
|-----|---|----------|-----|
|     | 1 | $k$      | $K$ |
| 1   |   |          |     |
| $i$ |   | $x_{ik}$ |     |
| $I$ |   |          |     |

- Économie : valeur de l'**indicateur**  $k$  dans la **région**  $i$
- Psychologie : degré d'accord de l'**individu**  $i$  avec l'**affirmation**  $k$
- Sociologie : **tps passé** à l'**activité**  $k$  par les individus de la **CSP**  $i$
- Enquête PISA : note de l'**élève**  $i$  dans la **discipline**  $k$

## Les données vins

- 10 individus : vins blancs du Val de Loire



- Quels vins se ressemblent ? Peut-on faire des groupes de vins ?
- Comment caractériser un vin ?
- Quels descripteurs se ressemblent ?

## Les données vins

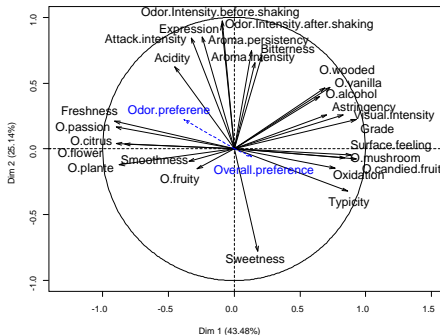
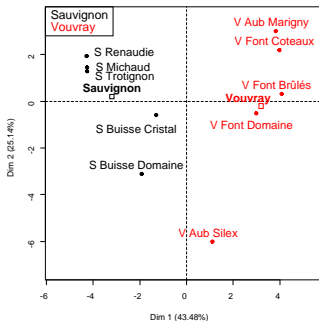
- 10 individus : vins blancs du Val de Loire
- 27 variables quantitatives : descripteurs sensoriels
  - mais aussi 2 variables d'appréciation
  - et 1 variable qualitative : label des vins (Vouvray - Sauvignon)

|                  | O.fruity | O.passion | O.citrus | ... | Sweetness | Acidity | Bitterness | Astringency | Aroma.intensity | Aroma.persistency | Visual.intensity | Odor.preference | Overall.preference | Label     |
|------------------|----------|-----------|----------|-----|-----------|---------|------------|-------------|-----------------|-------------------|------------------|-----------------|--------------------|-----------|
| S Michaud        | 4.3      | 2.4       | 5.7      | ... | 3.5       | 5.9     | 4.1        | 1.4         | 7.1             | 6.7               | 5.0              | 6.0             | 5.0                | Sauvignon |
| S Renaudie       | 4.4      | 3.1       | 5.3      | ... | 3.3       | 6.8     | 3.8        | 2.3         | 7.2             | 6.6               | 3.4              | 5.4             | 5.5                | Sauvignon |
| S Trotignon      | 5.1      | 4.0       | 5.3      | ... | 3.0       | 6.1     | 4.1        | 2.4         | 6.1             | 6.1               | 3.0              | 5.0             | 5.5                | Sauvignon |
| S Buisse Domaine | 4.3      | 2.4       | 3.6      | ... | 3.9       | 5.6     | 2.5        | 3.0         | 4.9             | 5.1               | 4.1              | 5.3             | 4.6                | Sauvignon |
| S Buisse Cristal | 5.6      | 3.1       | 3.5      | ... | 3.4       | 6.6     | 5.0        | 3.1         | 6.1             | 5.1               | 3.6              | 6.1             | 5.0                | Sauvignon |
| V Aub Silex      | 3.9      | 0.7       | 3.3      | ... | 7.9       | 4.4     | 3.0        | 2.4         | 5.9             | 5.6               | 4.0              | 5.0             | 5.5                | Vouvray   |
| V Aub Marigny    | 2.1      | 0.7       | 1.0      | ... | 3.5       | 6.4     | 5.0        | 4.0         | 6.3             | 6.7               | 6.0              | 5.1             | 4.1                | Vouvray   |
| V Font Domaine   | 5.1      | 0.5       | 2.5      | ... | 3.0       | 5.7     | 4.0        | 2.5         | 6.7             | 6.3               | 6.4              | 4.4             | 5.1                | Vouvray   |
| V Font Brûlés    | 5.1      | 0.8       | 3.8      | ... | 3.9       | 5.4     | 4.0        | 3.1         | 7.0             | 6.1               | 7.4              | 4.4             | 6.4                | Vouvray   |
| V Font Coteaux   | 4.1      | 0.9       | 2.7      | ... | 3.8       | 5.1     | 4.3        | 4.3         | 7.3             | 6.6               | 6.3              | 6.0             | 5.7                | Vouvray   |

- Quels vins se ressemblent ? Peut-on faire des groupes de vins ?
- Comment caractériser un vin ?
- Quels descripteurs se ressemblent ?



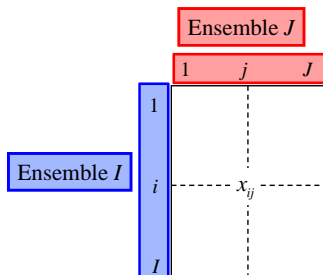
# Représentation des individus et des variables



⇒ Utilisation d'information supplémentaire

- la variable qualitative *cépage*
- les variables quantitatives d'*appréciation*

# L'analyse des correspondances (AFC)



$x_{ij}$  : nombre d'individus appartenant  
à l'élément  $i$  de l'ensemble  $I$   
à l'élément  $j$  de l'ensemble  $J$

Personnages de Mots

Phèdre (Racine)

CSP

classes d'âge

Nombre de fois que le personnage  
 $i$  a utilisé le mot  $j$

nombre d'individus de la classe  
d'âge  $j$  et de la CSP  $i$

Profession de foi Mots

Nombre de fois où la profession de  
foi  $i$  utilise le mot  $j$

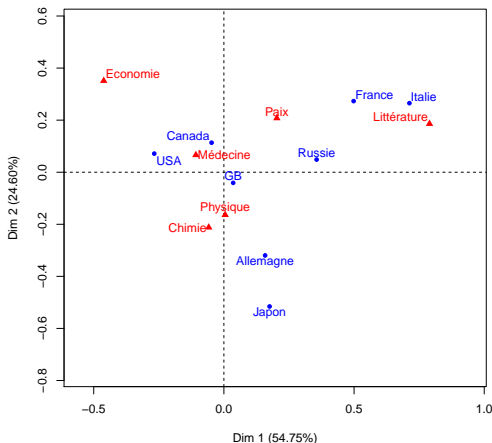
⇒ Exemples où le test d'indépendance du  $\chi^2$  peut être appliqué

## Données sur les prix Nobel

|           | Chimie | Economie | Littérature | Médecine | Paix | Physique | Somme |
|-----------|--------|----------|-------------|----------|------|----------|-------|
| Allemagne | 24     | 1        | 8           | 18       | 5    | 24       | 80    |
| Canada    | 4      | 3        | 2           | 4        | 1    | 4        | 18    |
| France    | 8      | 3        | 11          | 12       | 10   | 9        | 53    |
| GB        | 23     | 6        | 7           | 26       | 11   | 20       | 93    |
| Italie    | 1      | 1        | 6           | 5        | 1    | 5        | 19    |
| Japon     | 6      | 0        | 2           | 3        | 1    | 11       | 23    |
| Russie    | 4      | 3        | 5           | 2        | 3    | 10       | 27    |
| USA       | 51     | 43       | 8           | 70       | 19   | 66       | 257   |
| Somme     | 121    | 60       | 49          | 140      | 51   | 149      | 570   |

Y a-t'il un lien entre les pays et les catégories de prix ? Certains pays ont-ils des spécificités ?

## Exemple des prix Nobel



- opposition sciences - autres  
dans une moindre mesure, opposition physique/chimie - science économique
- positions des pays illustrent leur spécificité dans l'obtention des prix Nobel

AFC donne une visualisation synthétique qui aide la compréhension du tableau (a fortiori avec de grands tableaux)

# L'Analyse des correspondances multiples (ACM)

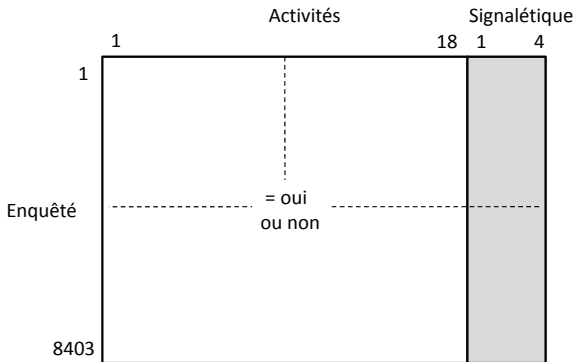
Pour analyser des questionnaires (tableau individus - variables qualitatives)

## L'Analyse des correspondances multiples (ACM)

Pour analyser des questionnaires (tableau individus - variables qualitatives)

- Extrait d'une enquête de l'Insee de 2003 sur la construction des identités, appelée « Histoire de vie »
- 8403 individus
- 2 sortes de variables :
  - *Parmi les loisirs suivants, indiquez ceux que vous pratiquez régulièrement* : Lecture, Ecouter de la musique, Cinéma, Spectacle, Exposition, Ordinateur, Sport, Marche, Voyage, Jouer de la musique, Collection, Activité bénévole, Bricolage, Jardinage, Tricot, Cuisine, Pêche, nombre d'heures moyen par jour à regarder la TV
  - le signalétique (4 questions) : sexe, âge, profession, statut matrimonial

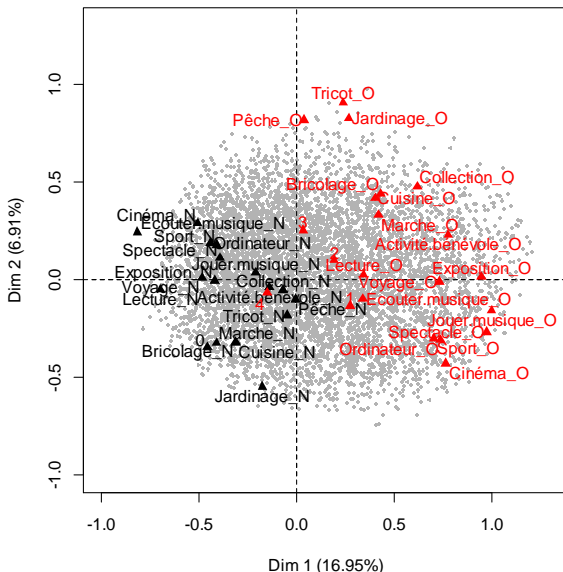
## Exemple : les données loisirs



ACM : loisirs en actif, signalétique en supplémentaire

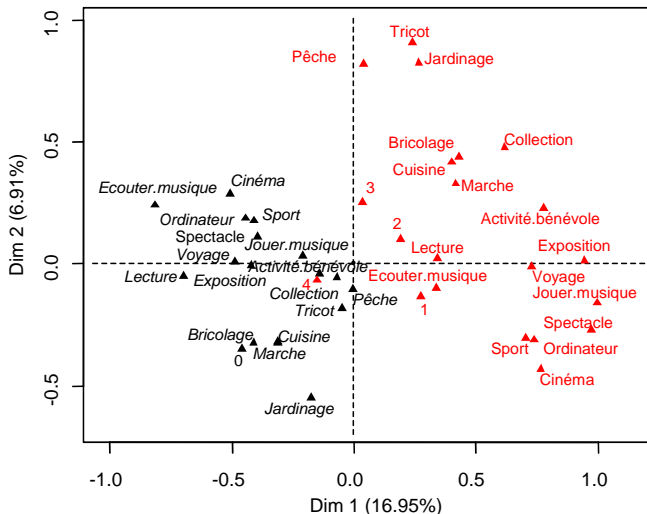
- 1 individu = profil d'activités
- Principales dimensions de variabilité des profils d'activités
- Liaisons entre ces dimensions et le signalétique

## Représentation simultanée

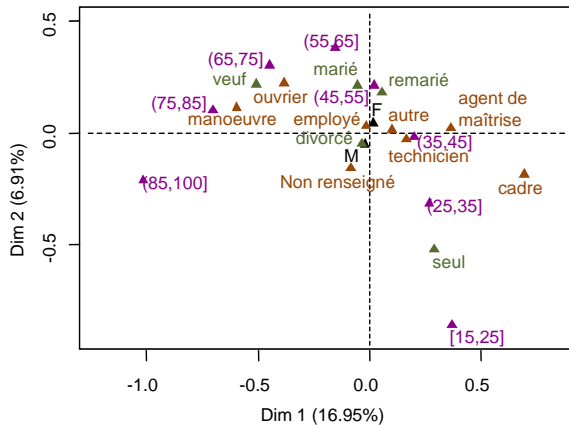
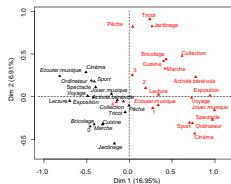




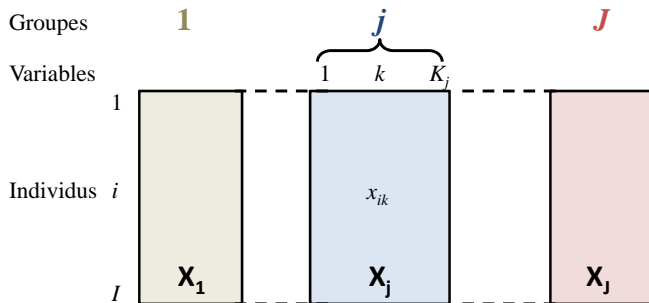
# Représentation des modalités



# Représentation des modalités

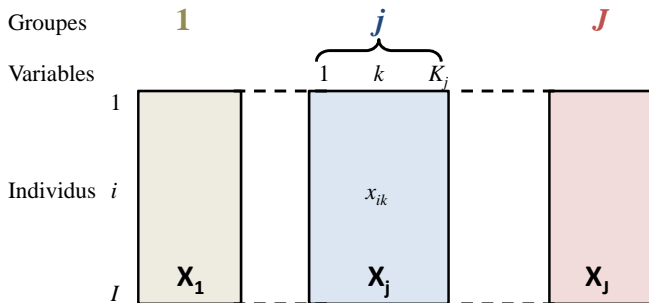


# L'Analyse Factorielle Multiple (AFM)



Exemples avec des variables **quantitatives et/ou qualitatives**  
**et/ou des tableaux de contingence** :

# L'Analyse Factorielle Multiple (AFM)



Exemples avec des variables **quantitatives et/ou qualitatives** et/ou des **tableaux de contingence** :

- enquête avec plusieurs thèmes (et plusieurs questions par thème)
- tableau pays  $\times$  indicateurs économique, sur plusieurs années
- questionnaire avec échelles de likert et questions qualitatives
- analyse textuelle d'un mouvement social par les journaux, à

## Description sensorielle de vins : comparaison de jurys

- 10 vins blancs du Val de Loire : 5 Vouvray - 5 Sauvignon
- description sensorielle de 3 jurys : œnologue, conso., étudiant

|        | Expert<br>(27) | Conso<br>(15) | Etudiant<br>(15) |
|--------|----------------|---------------|------------------|
| Vin 1  |                |               |                  |
| Vin 2  |                |               |                  |
| ...    |                |               |                  |
| Vin 10 |                |               |                  |

- Comment caractériser les vins ?
- Les vins sont-ils décrits de la même façon par les différents jurys ? Y-a t'il des spécificités par jury ?
- Peut-on comparer les typologies des vins d'un jury à l'autre ?

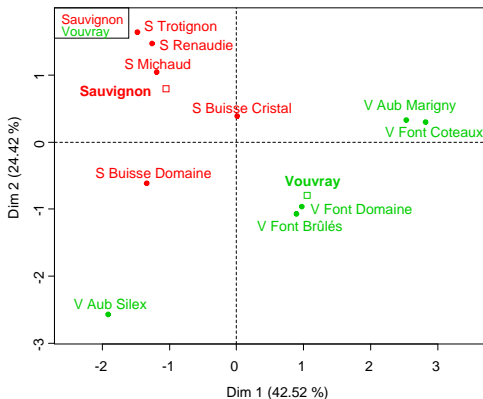
## Description sensorielle de vins : comparaison de jurys

- 10 vins blancs du Val de Loire : 5 Vouvray - 5 Sauvignon
- description sensorielle de 3 jurys : œnologue, conso., étudiant

|        | Expert<br>(27) | Conso<br>(15) | Etudiant<br>(15) | Appréciation<br>(60) | Cépage<br>(1) |
|--------|----------------|---------------|------------------|----------------------|---------------|
| Vin 1  |                |               |                  |                      |               |
| Vin 2  |                |               |                  |                      |               |
| ...    |                |               |                  |                      |               |
| Vin 10 |                |               |                  |                      |               |

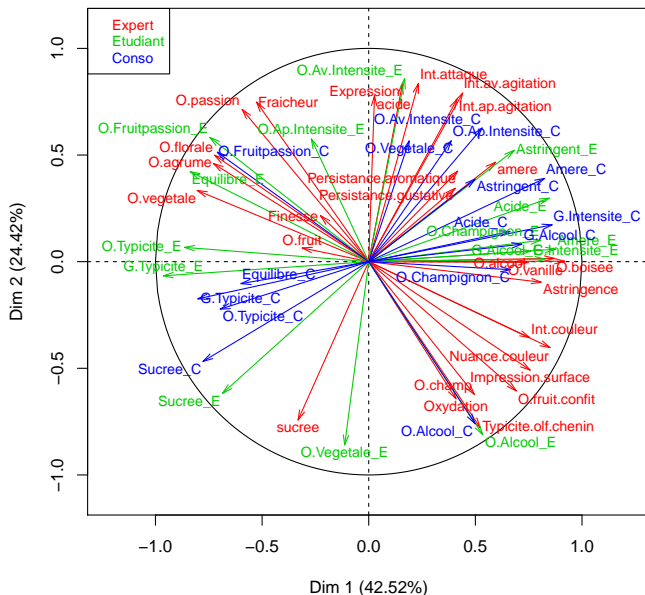
- Comment caractériser les vins ?
- Les vins sont-ils décrits de la même façon par les différents jurys ? Y-a t'il des spécificités par jury ?
- Peut-on comparer les typologies des vins d'un jury à l'autre ?

## Représentation des individus



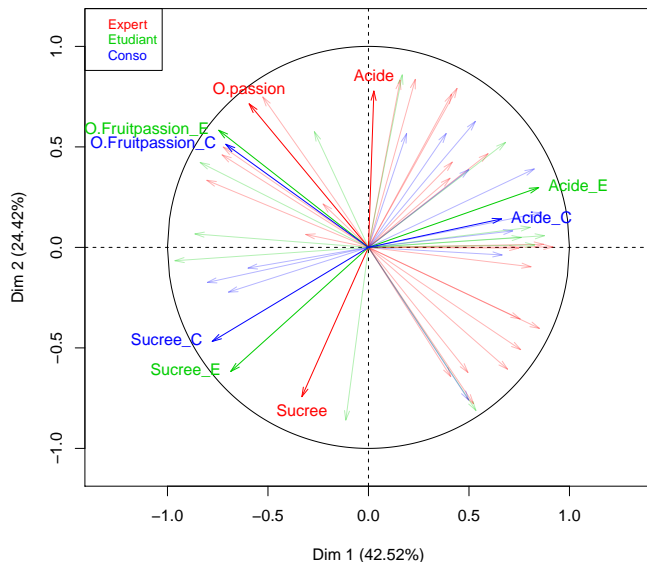
- Les deux cépages sont bien séparés
- Les Vouvray sont plus différents du point de vue sensoriel
- Plusieurs groupes de vins, ...

# Représentation des variables

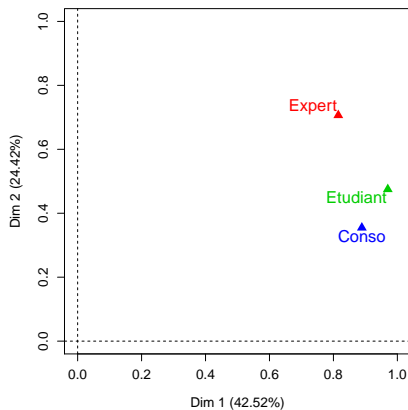




# Représentation des variables



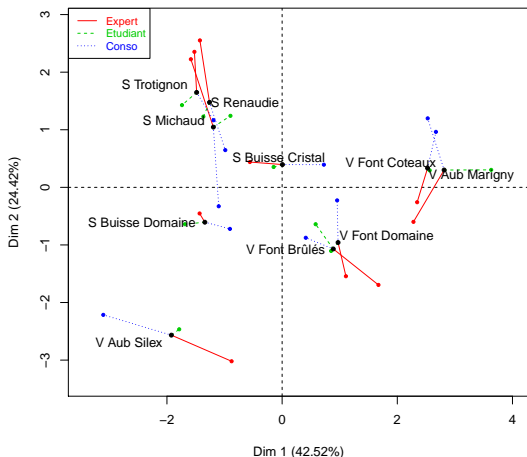
## Représentation des groupes



- 1ère dimension commune à tous les groupes
- 2ème dimension due au groupe Expert
- 2 groupes sont proches quand ils induisent la même structure

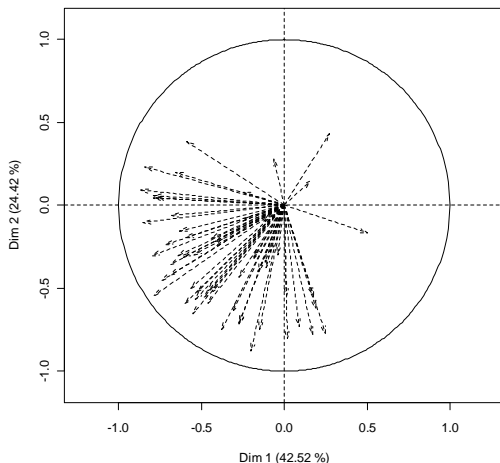
⇒ Ce graphe fournit une comparaison synthétique des groupes  
⇒ Les positions relatives des individus sont-elles similaires d'un groupe à l'autre ?

## Représentation des points partiels



- Point partiel = représentation d'un individu vu par un groupe
- Un individu est au barycentre de ses points partiels
- Un individu est homogène si ses points partiels sont proches

## Représentation de variables supplémentaires

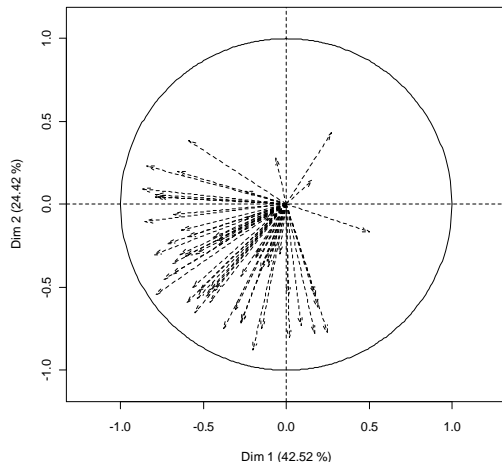


Les préférences sont liées à la description sensorielle

## Représentation de variables supplémentaires



Le vin préféré est  
*Vouvray Aubussière*  
*Silex*



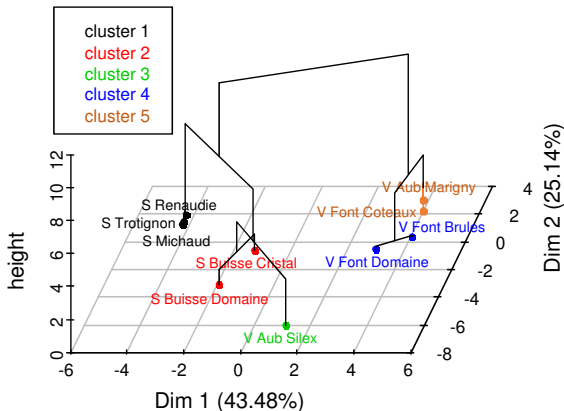
Les préférences sont liées à la description  
sensorielle



## Classification et plan factoriel

Représentation de l'arbre et des classes sur un plan factoriel

### Hierarchical clustering on the factor map



# Plan

Présentation des méthodes



Factoshiny pour l'interface et les graphes interactifs

FactoInvestigate pour des rapports automatisés

Conclusion



# FACTOMINER<sup>R</sup> en quelques mots

## Le package

- propose des méthodes d'analyses factorielles et de classification
- de nombreux indicateurs (qualité de représentation, contribution, description automatique des axes, ...)
- possibilité d'ajouter des éléments supplémentaires
- interface graphique (en français et en anglais)
- gestion des données manquantes (package missMDA)
- module graphique (package Factoshiny)
- rapport automatisé (package FactoInvestigate)
- aides à l'utilisateur (site internet, vidéos, livres, MOOC)

# FACTOMINER<sup>®</sup> en quelques mots

Différentes méthodes pour différents formats de données :

| Données   | Méthodes  | Fonction |
|---|---|----------|
| Variables quantitatives                           | An. en composantes principales                                    | PCA      |
| Table de contingence                              | An. des correspondances   | CA       |
| Variables qualitatives                            | An. des correspondances multiples                                 | MCA      |
| Données mixtes                                    | An. factorielle de données mixtes                                 | FAMD     |
| Groupes de variables                              | An. factorielle multiple  | MFA      |
| Hierarchie sur les variables                      | An. factorielle multiple hiérarchique                             | HMFA     |
| Groupes d'individus                               | An. factorielle multiple duale                                    | DMFA     |
| Tableau de contingence et variables contextuelles | An. des correspondances généralisée sur tableaux lexicaux agrégés | CaGalt   |

Méthodes de classification et méthodes outils complémentaires :

| Méthodes  | Fonction        |
|---|-----------------|
| Classification ascendante hiérarchique                        | HCPC            |
| Description d'une variable qualitative (ex. var. de classe)   | catdes          |
| Description d'une variable quantitative (ex. d'une dimension) | condes, dimdesc |

## Menu déroulant – Interface graphique – Package complémentaire

- **Factoshiny** : interface et graphes interactifs

⇒ faciliter l'utilisation des packages pour les utilisateurs

- **FactoInvestigate** : rapport automatisé

⇒ propose une interprétation des résultats

- **missMDA** : gestion des données manquantes

⇒ aller plus loin que les méthodes standards du package

## Aides à l'utilisateur : site internet

- <http://factominer.free.fr>
- en anglais et en français
- exemples, aides sur les fonctions, références, etc.

The screenshot shows a web browser window with the address bar displaying `factominer.free.fr/index_fr.html`. The website has a dark navigation bar with the following menu items: Accueil, Méthodes FactoMineR, Enseignement MOOC, livres, Améliorations graphiques, Valeurs manquantes missMDA, Rapport automatique, Google group, and Plus. Below the navigation bar is a large banner with the **FACTOMINER** logo. To the right of the logo are flags for the United Kingdom and France. The main content area is titled "FactoMineR en quelques mots" and contains a paragraph describing the package as an R tool for multidimensional exploratory data analysis, developed by François Husson, Julie Josse, Sébastien Lê, and J. Mazet. Below this is a section titled "Pourquoi utiliser FactoMineR?" with a list of six points: 1. Permet de mettre en œuvre des méthodes d'analyse de données (ACP, correspondances, etc.). 2. Permet l'ajout d'information supplémentaire (individus, variables). 3. Fournit une vue géométrique et des sorties graphiques. 4. Fournit des aides à l'interprétation (description automatique des axes, indicateurs). 5. Permet de prendre en compte des structures sur les données (hiérarchie, etc.). 6. Beaucoup de matériel pédagogique (MOOC, livres, etc.) est disponible. On the right side, there is a "Menu de l'accueil" with links to: Présentation de FactoMineR, Nouvelles, Installation de FactoMineR, Comment citer FactoMineR?, and Historique de FactoMineR. Below that is a "Liens utiles" section with links to: Département de Mathématiques d'Agrocampus Rennes, R Project, and FactoInvetigate.

FactoMineR : analyse de

← → ↻ ① factominer.free.fr/index\_fr.html ☆ ⓘ ⋮

Accueil Méthodes FactoMineR Enseignement MOOC, livres Améliorations graphiques Valeurs manquantes missMDA Rapport automatique Google group Plus

**FACTOMINER**

🇬🇧 🇫🇷

> **FactoMineR en quelques mots**

FactoMineR est un package R dédié à l'analyse exploratoire multidimensionnelle de données (à la Française). Il a été développé et il est maintenu par François Husson, Julie Josse, Sébastien Lê, d'Agrocampus Rennes, et J. Mazet.

> **Pourquoi utiliser FactoMineR?**

1. Il permet de mettre en œuvre des méthodes d'analyse de données telles que l'analyse en composantes principales (ACP), l'analyse des correspondances (AC), l'analyse des correspondances multiples (ACM) ainsi que des analyses plus avancées.
2. Il permet l'ajout d'information supplémentaire telle que des individus et/ou des variables supplémentaires.
3. Il fournit un point de vue géométrique et de nombreuses sorties graphiques.
4. Il fournit de nombreuses aides à l'interprétation (description automatique des axes, nombreux indicateurs, ...).
5. Il peut prendre en compte diverses structures sur les données (structure sur les variables, hiérarchie sur les variables, structure sur les individus).
6. Beaucoup de matériel pédagogique (MOOC, livres, etc.) est disponible pour

> **Menu de l'accueil**

- Présentation de FactoMineR
- Nouvelles
- Installation de FactoMineR
- Comment citer FactoMineR?
- Historique de FactoMineR

> **Liens utiles**

- Département de Mathématiques d'Agrocampus Rennes
- R Project
- FactoInvetigate

## Aides à l'utilisateur : un Google group

- <https://groups.google.com/group/factominer-users/>
- possibilité de poser des questions et/ou répondre
- en français ou en anglais

The screenshot shows a web browser window displaying the Google Groups page for "FactoMineR users". The URL in the address bar is <https://groups.google.com/forum/?hl=fr#forum/factominer-users>. The page header includes the Google logo, a search bar, and navigation links. Below the header, the group name "FactoMineR users" is displayed, along with the number of subjects (32 sur 405) and a "G+1" button. A table of messages is shown, with columns for message details, number of replies, and dates.

| Message  | Replies   | Date       |
|--|-----------|------------|
| La 2ème session du MOOC "Analyse de données multidimensionnelles" débute le 1er mars | 1 message | 19 janv.   |
| Nouveau module graphique   | 11        | 17/01/2015 |
| Select the best sample using a reference   | 1         | 16 mai     |
| salut  | 1         | 15 mai     |
| PCAshiny scale.unit=F impossible ?   | 2         | 12 mai     |
| ACM et questions à choix multiples (plusieurs modalités dans la même question)       | 6         | 2 mai      |
| Plot CA neatly (1)   | 1         | 30 avr.    |
| Interpreting MCA results   | 11        | 22 avr.    |
| Installing FactoMineR on Linux (1)   | 1         | 21 avr.    |
| Estimation of PC for MFA   | 4         | 21 avr.    |
| General questions for FAMD   | 1         | 1 avr.     |

# Plan

Présentation des méthodes



Factoshiny pour l'interface et les graphes interactifs

FactoInvetigate pour des rapports automatisés

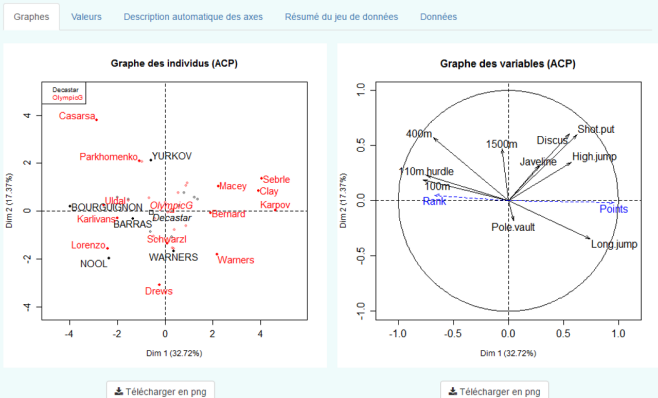
Conclusion

# Graphiques interactifs avec le package Factoshiny

- Réaliser des analyses sans besoin de maîtriser le code
- Visualisation en temps réel des modifications apportées

```
> res <- PCAshiny(decathlon) ## analyse factorielle sur les données
> res <- PCAshiny(res.pca)    ## graphe sur un objet résultat de FactoMineR
> res2 <- PCAshiny(res)      ## objet résultat de Factoshiny
```

## ACP sur le jeu de données decathlon



# Graphiques interactifs avec le package Factoshiny

Une petite démonstration :

```
library(Factoshiny)
vins <- read.table("https://husson.github.io/img/vins_expert.csv",
                  header=TRUE, sep=";", row.names=1)
res <- PCAshiny(vins)
```



# Plan

Présentation des méthodes



Factoshiny pour l'interface et les graphes interactifs

FactoInvestigate pour des rapports automatisés

Conclusion

Rapport automatisé avec le package FactoInvestigate

Propose une interprétation des résultats basée sur l'objet résultat

```
> res.pca <- PCA(MesDonnees, ...)
> library(FactoInvestigate)
> Investigate(res.pca)
```

<http://factominer.free.fr/reporting>

## Analyse en Composantes Principales

### Jeu de données decathlon

Ce jeu de données contient 41 individus et 13 variables. 2 variables quantitatives sont illustratives, 1 variable qualitative est illustrative

### 1. Observation d'individus extrêmes

L'analyse des graphes ne révèle aucun individu singulier.

## 2. Distribution de l'inertie

L'inertie des axes factoriels indique d'une part si les variables sont structurées et suggère d'autre part le nombre judicieux de composantes principales à étudier.

Les 2 premiers axes de l'ACP expliquent **80,09%** de l'inertie totale du jeu de données : cela signifie que 80,09% de la variabilité totale du nuage des individus (ou des variables) est représentée dans ce plan. C'est un pourcentage assez important, et le premier plan représente donc conventionnellement la variabilité contenue dans une grande part du jeu de données étudié. Cette valeur est supérieure à la valeur référence de **37,78%** la variabilité expliquée par ce plan est donc significative (cette dernière référence est la quartile 0,95 de la distribution des pourcentages d'inertie obtenus en simulant 1671 jeux de données aléatoires de dimensions comparables sur la base d'une distribution normale).

Du fait de ces observations, il serait tout de même probablement préférable de considérer également dans l'analyse les dimensions supérieures ou égales à la troisième.

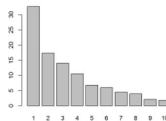
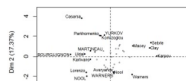


Figure 2. Decomposition of the total inertia on the components of the ACP

Une estimation du nombre pertinent d'axes à interpréter suggère de restreindre l'analyse à la description des 3 premiers axes. Ces composantes révèlent un taux d'inertie supérieur à celle du quantile 0.95 de distributions aléatoires (54,14% contre 51,44%). Cette observation suggère que seuls ces axes sont porteurs d'une véritable information. En conséquence, la description de l'analyse sera restreinte à ces seuls axes.

### 3. Description du plan 1:2



# Rapport automatisé avec le package FactoInvestigate

Une petite démonstration :

```
library(FactoInvestigate)
vins <- read.table("https://husson.github.io/img/vins_expert.csv",
                  header=TRUE, sep=";", row.names=1)
res.pca <- PCA(vins, quanti.sup=29:30, quali.sup=1)
Investigate(res.pca)
```

# Plan

Présentation des méthodes



Factoshiny pour l'interface et les graphes interactifs

FactoInvestigate pour des rapports automatisés

Conclusion

## Un exemple en linguistique

- |                           |   |
|---------------------------|---|
| - Aragon (23 textes) :    | FeuJoie, Perpétuel, Destinées, Snark, Peinture, ...     |
| - Balzac (49 textes) :    | <i>Chouans, Physiologie, Vendetta, Gobseck, ...</i>     |
| - Corneille (34 textes) : | <i>Mélite, Clitandre, Veuve, Galerie, Suivante, ...</i> |
| - ...                     |   |



## Un exemple en linguistique

- Aragon (23 textes) : FeuJoie, Perpétuel, Destinées, Snark, Peinture, ...
- Balzac (49 textes) : Chouans, Physiologie, Vendetta, Gobseck, ...
- Corneille (34 textes) : Méliite, Clitandre, Veuve, Gelerie, Suivante, ...
- ...

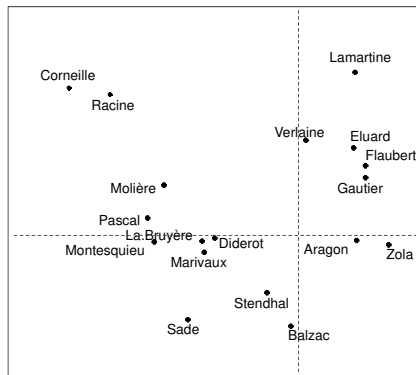
On conserve les  
mots cités au  
moins 100 fois

978 mots



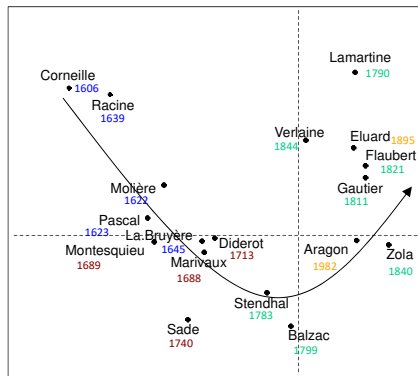
|           |      |      |      |     |     |
|-----------|------|------|------|-----|-----|
| accord    | 264  | 0    | 88   | 44  | ... |
| affaire   | 1029 | 2040 | 74   | 154 | ... |
| âge       | 545  | 629  | 92   | 108 |     |
| ah        | 219  | 0    | 0    | 0   |     |
| air       | 2093 | 2009 | 95   | 191 |     |
| allemagne | 366  | 0    | 0    | 0   |     |
| allemand  | 476  | 0    | 0    | 0   |     |
| amant     | 303  | 760  | 566  | 0   |     |
| âme       | 478  | 2190 | 1101 | 240 |     |
| ami       | 1090 | 2583 | 307  | 407 |     |
| amour     | 1374 | 3286 | 1791 | 167 |     |
| an        | 1812 | 3009 | 112  | 182 |     |
| anglais   | 315  | 0    | 0    | 0   |     |
| . . .     |      |      |      |     |     |

## Un exemple en linguistique



Avec l'AFC, les auteurs sont d'autant plus proches qu'ils emploient les mots dans les mêmes proportions, i.e. qu'ils s'intéressent aux mêmes sujets et ont les mêmes préoccupations

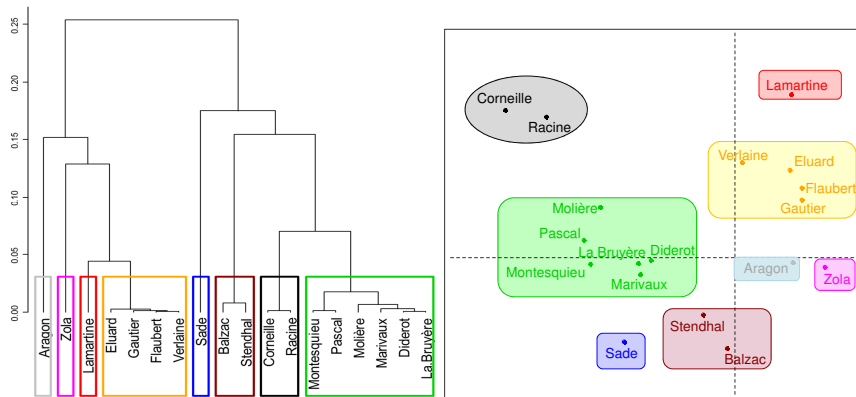
## Un exemple en linguistique



- Corneille et Racine sont proches et très éloignés de Zola. Ce sont 2 auteurs classiques du 17ème tandis que Zola est un naturaliste du 19ème
- Évolution du vocabulaire selon les siècles

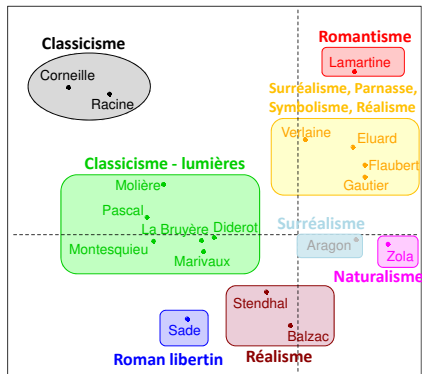
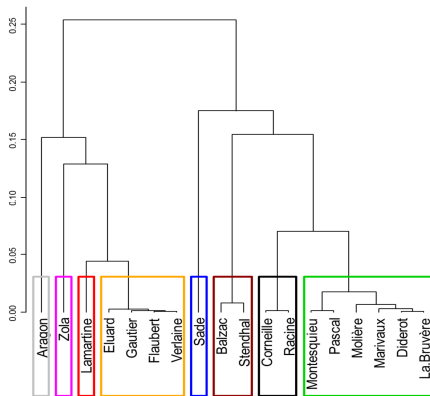


## Un exemple en linguistique



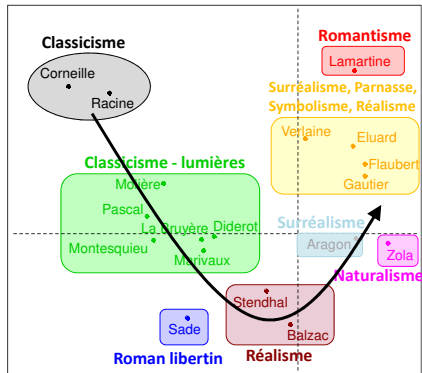
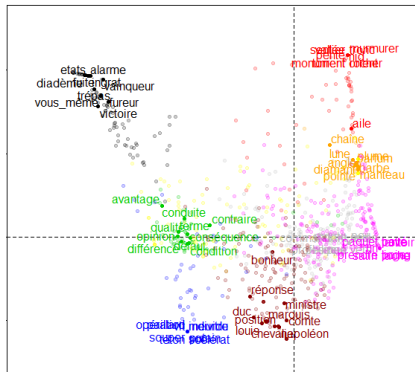
- La classification retrouve des classes d'auteurs connues

## Un exemple en linguistique



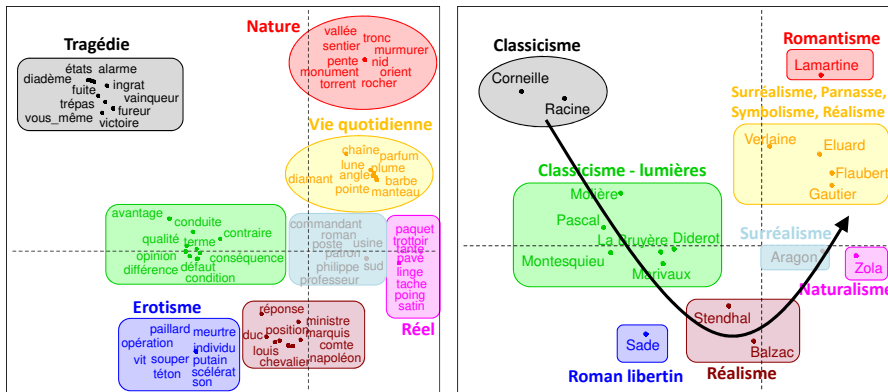
- Stendhal et Balzac (réalistes) sont très éloignés de Lamartine (romantique). On retrouve ici que les auteurs réalistes ont un point commun : s'éloigner des excès romantiques !
- Points communs naturalistes / réalistes : montrer la société telle qu'elle est, le roman devient le miroir de la société

## Un exemple en linguistique



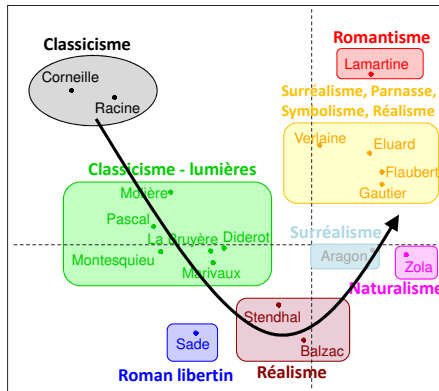
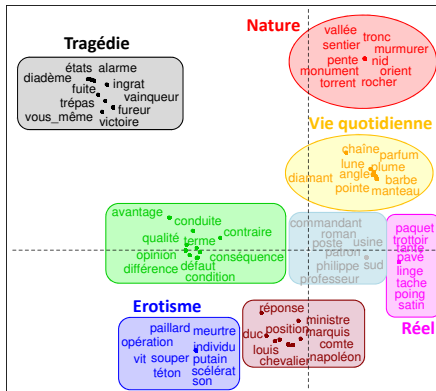
Les mots permettent de caractériser les sujets de prédilection des auteurs et les courants littéraires

# Un exemple en linguistique



- Le naturalisme est la suite logique du réalisme : le naturalisme montre le milieu où vit le protagoniste pour expliquer son comportement de façon "scientifique"
- Évolution du vocabulaire selon les courants littéraires

## Un exemple en linguistique



- Le naturalisme est la suite logique du réalisme : le naturalisme montre le milieu où vit le protagoniste pour expliquer son comportement de façon "scientifique"
- Évolution du vocabulaire selon les courants littéraires

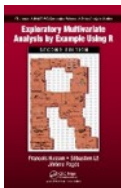
Une vidéo pour en savoir plus sur cet exemple

## Matériel sur FactoMineR

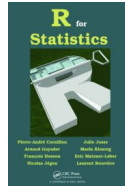
- FactoMineR : pour mettre en œuvre les méthodes
  - Factoshiny : pour un menu déroulant et graphes interactifs
  - missMDA : pour la gestion des données manquantes
  - FactoInvestigate : pour les rapports automatisés
- 
- site FactoMineR : <http://factominer.free.fr>
  - site F. Husson : <https://husson.github.io>
- 
- 2 articles dans J. of stat. software ([FactoMineR](#), [missMDA](#))
  - 2 articles dans R journal ([CA-galt](#), [MFACT](#))

## Aides à l'utilisateur

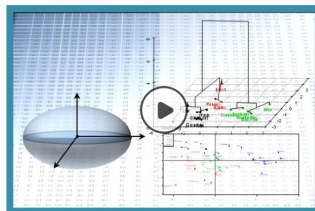
*Analyse de données avec R (2<sup>e</sup> ed)*



*R pour la stat. et sc. des données*



MOOC analyse de données  
multidimensionnelles



Une chaîne Youtube : <https://www.youtube.com/HussonFrancois>

**R** missMDA : données manquantes

**R** RcmdrPlugin.FactoMineR: menu déroulant  
Factoshiny : graphes interactifs



**FACTOMINER**

