

D'une simple vignette à l'enseignement par MOOC

Quelques idées pour améliorer la visibilité et l'accessibilité de son package R

François Husson

Département de statistique et informatique, Agrocampus Ouest

Cinquièmes Rencontres R



Toulouse
du 22 au 24 juin 2016



Plan

- 1 Introduction
- 2 Quelques conseils de base (ou lapalissades)
- 3  **FACTOMINER**
- 4 Packages complémentaires
- 5 Diffusion
- 6 Enseignement

Introduction

La construction d'un package permet de :

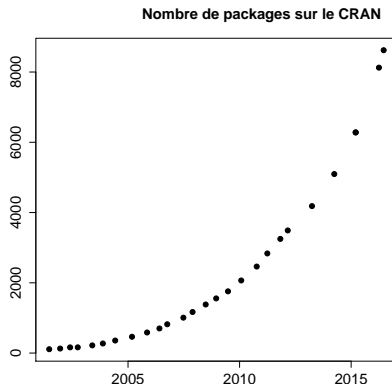
- proposer de nouvelles méthodes statistiques ou démarches méthodologiques
- partager son travail avec l'ensemble de la communauté scientifique
- faciliter la confrontation de méthodes
- rendre disponibles des jeux de données

La création d'un package est chronophage : elle DOIT être bénéfique au concepteur du package ET à la communauté scientifique

De plus en plus de packages

20 juin 2016 :

- CRAN : 8642 packages
- Bioconductor : 2422 packages
- R-Forge : 2006 projets
- GitHub : ??? projets



⇒ la visibilité d'un package est de plus en plus limitée

De nombreux packages se retrouvent inutilisés ... et donc inutiles !!

Plan

- 1 Introduction
- 2 Quelques conseils de base (ou lapalissades)
- 3  **FACTOMINER**
- 4 Packages complémentaires
- 5 Diffusion
- 6 Enseignement

Avant la soumission du package

- Un package pour faire quoi ?
 - quel apport par rapport aux packages existants ?
 - faut-il proposer une fonction aux auteurs d'un autre package ?

- Un package pour qui ?
 - pour qq chercheurs du domaine \implies GitHub ou page internet
 - pour une audience large \implies CRAN ou bioconductor (GitHub)

Avant la soumission du package

La première version du package peut être limitée, mais ce qui est fait doit être bien fait

- le package évoluera et certains choix sont difficiles à modifier
 - le nom du package
 - les noms des principales fonctions
 - les arguments par défaut
- les utilisateurs n'utiliseront pas le package s'ils ne comprennent pas son fonctionnement
 - bien documenter ses fonctions
 - bien choisir ses exemples
 - faire une vignette ("guide des premiers pas")

Après la soumission du package

"Faire vivre" et maintenir le package :

- corriger le package si coquilles
- répondre aux questions des utilisateurs
- inclure de nouveaux développements, de nouvelles options
- améliorer la programmation (Rcpp, parallélisation, etc.)

Construire des packages complémentaires

Faire connaître le package

Plan

- 1 Introduction
- 2 Quelques conseils de base (ou lapalissades)
- 3  FACTOMINER**
- 4 Packages complémentaires
- 5 Diffusion
- 6 Enseignement

FACTOMINER^R 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 (avec le package missMDA)
- module graphique
- aides à l'utilisateur (site internet, vidéos, livres)

FACTOMINER^R 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

Plan

- 1 Introduction
- 2 Quelques conseils de base (ou lapalissades)
- 3 **FACTOMINER**
- 4 Packages complémentaires
- 5 Diffusion
- 6 Enseignement

Menu déroulant – Interface graphique – Package complémentaire

- `RcmdrPlugin.FactoMineR` : menu déroulant
- `Factoshiny` : graphes interactifs

⇒ faciliter l'utilisation des packages pour les utilisateurs

- `missMDA` : gestion des données manquantes

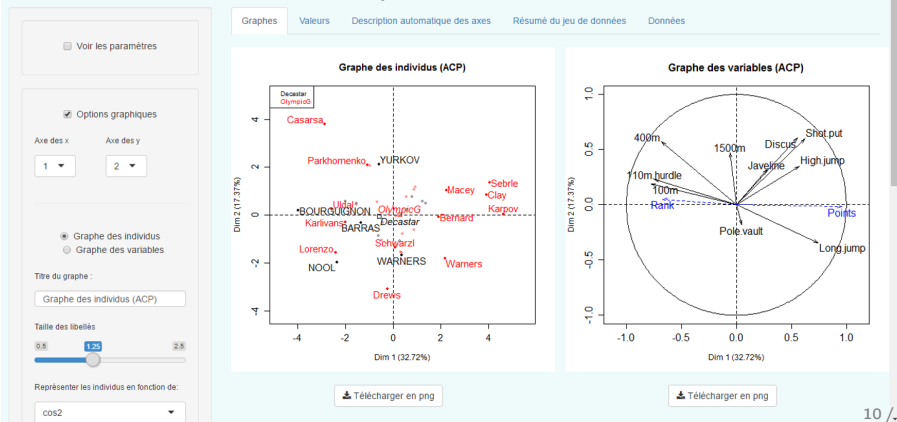
⇒ aller plus loin que les méthodes standards du package

Graphiques interactifs avec le package Factoshiny

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

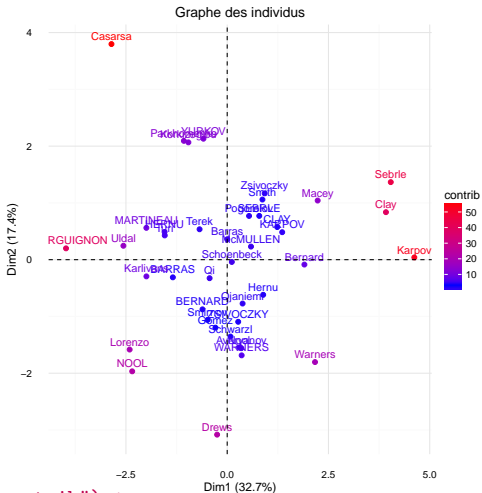
```
> res <- PCAshiny(decathlon) ## analyse factorielle sur les données
> res <- PCAshiny(res.pca)   ## graphe sur un objet résultat de FactoMiner
> res2 <- PCAshiny(res)     ## objet résultat de Factoshiny
```

ACP sur le jeu de données decathlon



De nouveaux packages graphiques

- le package **explor**
 - graphes interactifs
 - possibilité de bouger les libellés
- le package **factoextra**
 - basé sur **ggplot2**
 - construction séquentielle des graphes en ajoutant des couches



```
> library(factoextra)
> fviz_pca_ind(res.pca, col.ind="contrib") +
  labs(title="Graphe des individus") +
  scale_color_gradient2(mid="blue",high="red") +
  theme_minimal()
```

missMDA : package de gestion des données manquantes

Variables

1 j p

Individus

1

i

n

The diagram shows a light blue rectangular matrix representing a data set. The columns are labeled 'Variables' at the top, with indices 1, j , and p . The rows are labeled 'Individus' on the left, with indices 1, i , and n . Inside the matrix, several cells contain a question mark '?' to indicate missing data points. The missing values are scattered across the matrix, with a notable concentration in the lower-right quadrant.

Etude et mise en œuvre des méthodes factorielles en présence de données manquantes : ACP, ACM, AFDM, AFM

- 1 Imputation par analyse factorielle itérative
- 2 Analyse du tableau complété

ACP Itérative

Principe : imputer par des valeurs qui n'influent pas sur les résultats de l'ACP

- 1 initialisation $\ell = 0$: X^0 (imputation par la moyenne)
- 2 itération ℓ :
 - (a) ACP sur le tableau complété $\rightarrow (F^\ell, U^\ell)$;
 S dimensions conservées
 - (b) données manquantes imputées par $F^\ell U^{\ell'}$
 $\implies X^\ell = W * X + (1 - W) * F^\ell U^{\ell'}$
 - (c) moyennes (et écarts-types) sont mis à jour
- 3 étapes d'estimation et d'imputation sont répétées

\implies fournit les axes et composantes principales (mieux que Nipals)

\implies fournit une imputation du jeu de données

Gestion des données manquantes : exemple en ACP

```

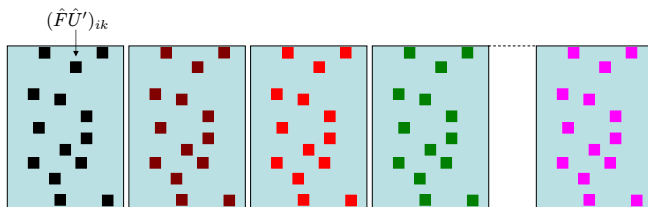
> library(missMDA)
> data(orange)
> nb <- estim_ncpPCA(orange,ncp.max=5)           ## Estime le nb de dimensions
> comp <- imputePCA(orange,ncp=nb,scale=TRUE)    ## Complète le tableau
> res.pca <- PCA(comp$completeObs)             ## Effectue l'ACP

```

> orange					> comp\$completeObs				
Sweet	Acid	Bitter	Pulp	Typicity	Sweet	Acid	Bitter	Pulp	Typicity
NA	NA	2.83	NA	5.21	5.54	4.13	2.83	5.89	5.21
5.46	4.13	3.54	4.62	4.46	5.46	4.13	3.54	4.62	4.46
NA	4.29	3.17	6.25	5.17	5.45	4.29	3.17	6.25	5.17
4.17	6.75	NA	1.42	3.42	4.17	6.75	4.73	1.42	3.42
...					...				
NA	NA	NA	7.33	5.25	5.71	3.87	2.80	7.33	5.25
4.88	5.29	4.17	1.50	3.50	4.88	5.29	4.17	1.50	3.50

Imputation multiple en ACP

⇒ ACP itérative : une méthode d'imputation simple



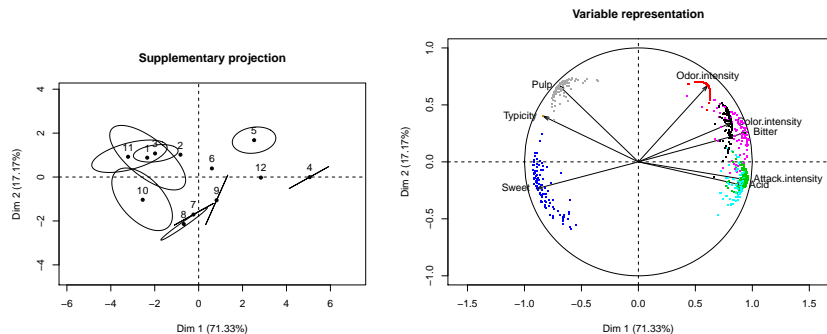
⇒ Une valeur unique ne peut pas refléter la variabilité de prédiction

⇒ Imputation multiple : générer plusieurs valeurs plausibles pour chaque valeur manquante

Visualisation de l'incertitude liée aux données manquantes

```

> library(missMDA)
> mi <- MIPCA(orange, scale = TRUE, ncp=2)
> mi$res.MI          ## sortie pour les tableaux imputés
> plot(mi)
  
```



Permet de ne pas analyser de tableaux avec trop de données manquantes

Gestion des données manquantes : exemple en ACM

```

> library(missMDA)
> data(vnf)
> nb <- estim_ncpMCA(vnf,ncp.max=5)      ## Estime le nb de dimensions
> imp <- imputeMCA(vnf, ncp=nb)         ## Complète le tableau disjonctif
> res <- MCA(vnf,tab.disj=imp$tab.disj) ## ACM utilisant tab disj complété

```

	V1	V2	V3	...	V14
ind 1	a	NA	g	...	u
ind 2	NA	f	g	...	u
ind 3	a	e	h	...	v
ind 4	a	e	h	...	v
ind 5	b	f	h	...	u
ind 6	c	f	h	...	u
ind 7	c	f	NA	...	v
...
ind 1232	c	f	h	...	v

	V1_a	V1_b	V1_c	V2_e	V2_f	V3_g	V3_h	...
ind 1	1	0	0	0.71	0.29	1	0	...
ind 2	0.12	0.29	0.59	0	1	1	0	...
ind 3	1	0	0	1	0	0	1	...
ind 4	1	0	0	1	0	0	1	...
ind 5	0	1	0	0	1	0	1	...
ind 6	0	0	1	0	1	0	1	...
ind 7	0	0	1	0	1	0.37	0.63	...
...
ind 1232	0	0	1	0	1	0	1	...

⇒ Même principe avec FAMD et MFA

Plan

- 1 Introduction
- 2 Quelques conseils de base (ou lapalissades)
- 3  **FACTOMINER**
- 4 Packages complémentaires
- 5 Diffusion
- 6 Enseignement

Aides à l'utilisateur : site internet

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

Accueil Méthodes Classiques Méthodes Avancées Interface Les + Excel F.A.Q. Docs
Contact

FACTOMINER

> Nouveautés

Le nouveau package Factoshiny, complémentaire de FactoMiner, est disponible sur le CRAN. Il permet de paramétrer les principales méthodes de FactoMiner et permet de construire des graphiques de façon interactive. Vous trouverez ici une vidéo de démonstration.

Un MOOC (cours en ligne gratuit) en analyse de données débutera le 2 mars 2015. Vous trouverez de nombreuses vidéos sur les méthodes d'analyse de données, vous pourrez récupérer le fichier pdf avec les transparents, faire des quiz pour vous auto-évaluer, voir comment mettre en oeuvre la méthode sur ordinateur grâce à des vidéos didacticiel sur FactoMiner, et vous aurez la possibilité de faire des exercices. Cliquez ici pour vous inscrire ou pour plus d'information.

Nouveau menu déroulant de FactoMiner disponible en Français ou en anglais grâce au package RcmdrPlugin.FactoMiner

English Version
Version française

> Top Menu

Accueil
Méthodes Classiques
Méthodes Avancées
Interface
Les Plus de Facto
FactoMiner et Excel
F.A.Q.
Documents
Contacts

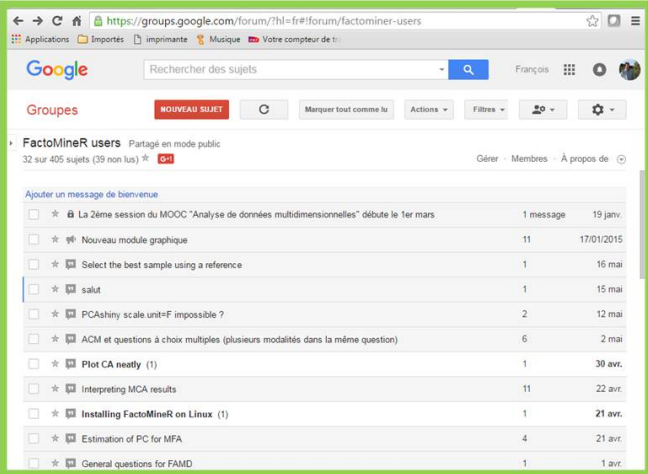
> Liens utiles

Département de Mathématiques d'Agrocampus Rennes
R Project

Démarche en analyse des...
2. Quel est le type d'échantillon ?
3. Quels sont les éléments qualifs ?
4. Quelles est le nature des variables ?
5. Quel est le nature des variables quantitatives ?

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 displaying the Google Groups forum for 'FactoMineR users'. The browser's address bar shows the URL <https://groups.google.com/forum/?hl=fr#!forum/factominer-users>. The page header includes the Google logo, a search bar with the text 'Rechercher des sujets', and the language 'Français'. Below the header, there are navigation buttons for 'Groupes', 'NOUVEAU SUJET', 'Marquer tout comme lu', 'Actions', 'Filtres', and user settings. The main content area shows the group name 'FactoMineR users' with a sub-header 'Partagé en mode public' and '32 sur 405 sujets (39 non lus)'. A list of forum posts is displayed, each with a checkbox, a star icon, a lock icon, a message icon, the subject, the number of messages, and the date.

Message	Nombre de messages	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.

Aides à l'utilisateur : site internet

- <http://mixomics.qfab.org/>
- en anglais
- interface, études de cas, démo, etc.

mixOmics Web Interface


About
Interface Guide
Case Studies
Demo
Contact
Start Wizard

What should I know before using this interface

- mixOmics does not build/infer relationships between the biological entities based on prior biological knowledge. It is an entirely data driven, holistic approach. It indicates which entities are likely to be related to each other (in a statistical sense) that would require further investigation. It indicates the similarity/differences between the samples.
- The two data sets must come from the same biological material (i.e. the same biological samples). The samples/individuals should be displayed in rows and biological entities or variables in columns in two separate files. Check that the samples are matching on each row (see figure below)

X



Y



- This web interface provides a user friendly tool to the mixOmics R package. More flexible options are available if using the original R package (also refer to our [tutorial](#))
- If you are not sure about the biological questions you would like to be answered through this interface, please refer to [which method to use?](#) or [examples](#)

Which method to use?

Depending on your biological question, one or several methodologies can be applied. Below are listed some typical analysis frameworks.

We call variables the expression or abundance of entities that are measured (genes, metabolites, proteins, SNP, ...) and samples or observations or units an individual, a patient, a cell on which the experiment is performed

In mixOmics, the data should be displayed with samples in rows and variables in columns. (see figure above)

I want to explore one single data set (e.g. microarray data)

- I would like to identify the trends or patterns in my data, experimental bias or identify if my samples "naturally" cluster according to the biological conditions: **Principal Component Analysis (PCA)**
- In addition to the above, I would like to select the variables that contribute the most to the variance (the information) in the data set **sparse Principal Component Analysis (sPCA)**


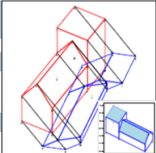

I want to unravel the common information contained in two data sets, where two types of variables are measured on the same samples (e.g. metabolomics and transcriptomics data)

I have one single data set (e.g. microarray data) and I am interested in classifying my samples into known classes

I have one single data set (e.g. microarray data) and I have one continuous response variable or outcome for each sample. I would like to predict the response with my data

Aides à l'utilisateur : site internet

- <http://pbil.univ-lyon1.fr/ade4/>
- en anglais et en français
- exemples, aides sur les fonctions, références, adelist, etc.

		<input type="text" value="Recherche dans le site"/> <input type="text" value="Mélange de lois normales"/>
<p>Accueil</p> <p>Page d'entrée Page de liens English Section Espace invités Maintenance</p> <p>Cours</p> <p>Introductions Biologie et modélisation (L) Tests d'hypothèse Analyse des données Fiches de stage Écologie et Statistique Évolution moléculaire Modélisation Divers</p> <p>Fiches de TD</p> <p>Le logiciel R Biologie et modélisation (L) Statistique descriptive La variabilité Les tests d'hypothèses Courbes de réponse Analyse multivariée (1) Analyse multivariée (2) Structures spatiales Sciences Financières et Assurances Divers</p> <p>Annales</p> <p>Biostatistiques (L) Biologie et modélisation (L) Analyse des données (M) aMIG (M) Algèbre & Statistique</p>	<p>Enseignements de Statistique en Biologie</p> <p>Fondateurs</p> <p>A.B. Dufour, D. Chessel, J.R. Lobry</p> <p>Contributeurs</p> <p>M. Bailly-Bechet, S. Dray, S. Mousset, J. Thioulouse</p> <p>Ingénieurs</p> <p>S. Penel et A. Siberchicot</p> <p>Notes de cours, illustrations, exercices, problèmes, fiches de Travaux Dirigés, Jeux de données pour la pratique de la statistique</p>	
		
		
	<p><i>Avec la participation de</i></p> <p>I. Amat, D. Charif, S. Charles, D. Clot, E. Desouhant, C. Gautier, T. Jombart, F. Menu, A. Necşulea, M.P.L. Nguyen, S. Ollier, S. Pavoine, S. Penel, G. Perrière, D. Pontier, N. Rochette, M. Royer, E. Vautrin, A. Viallefont.</p>	

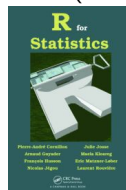
Aides à l'utilisateur : diffusion scientifique

- Suggéré le package dans un des [CRAN Task views](#)
- Exposés aux rencontres R et à useR !
- 2 articles dans R journal ([CA-galt](#), [MFACT](#))
- 2 articles dans J. of stat. software ([FactoMineR](#), [missMDA](#))
- Des livres :

Analyse de données avec R (2^e ed)



Statistique avec R (3^e ed)



Plan

- 1 Introduction
- 2 Quelques conseils de base (ou lapalissades)
- 3 **FACTOMINER**
- 4 Packages complémentaires
- 5 Diffusion
- 6 **Enseignement**

Aides à l'utilisateur : des vidéos

- disponibles depuis [ma page d'enseignement](#)
- disponibles depuis les aides de FactoMineR

Analyse de données avec FactoMineR

L'analyse de données avec R et FactoMineR. Comment faire une ACP, une ACM, une AFM ou encore une classification ? Quelles aides à l'interprétation ? Comment construire des graphiques lisibles ? commen...



Cours d'ACP : théorie et pratique (ancienne version)

de François Husson
38 635 vues • il y a 2 ans



Démarche en analyse des données (ancienne version)

de François Husson
19 348 vues • il y a 3 ans



ACP avec le menu déroulant de FactoMineR

de François Husson
7 533 vues • il y a 3 ans



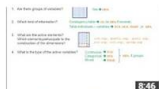
Analyse en Composantes Principales (ACP) avec...

de François Husson
24 748 vues • il y a 3 ans



Exploratory multivariate analysis with R and FactoMineR

This video shows how to perform exploratory multivariate analyses in a French way using R and FactoMineR and how to handle missing values...



Methodology in multivariate exploratory data analysis

de François Husson
7 973 vues • il y a 3 ans



Course on PCA: theory and practice

de François Husson
20 801 vues • il y a 2 ans



Principal component analysis (PCA) with R

de François Husson
25 377 vues • il y a 3 ans



Correspondence analysis with FactoMineR

de François Husson
6 114 vues • il y a 2 ans

Aides à l'utilisateur : chaîne Youtube

- <https://www.youtube.com/HussonFrancois>
- playlist en français (27 vidéos de cours & 18 vidéos logiciel)
- playlist en anglais (11 vidéos)

The screenshot shows the YouTube channel page for François Husson. At the top, there is a search bar with the text "Rechercher" and a magnifying glass icon. To the right of the search bar are buttons for "Mettre en ligne", a notification bell, and a profile picture. Below the search bar, it shows "2 455 abonnés" and "459 453 vues". The main banner image features a man in a blue shirt standing in front of a coastal landscape with purple flowers in the foreground and a blue sea and cliffs in the background. Below the banner, the channel name "François Husson" is displayed, along with a dropdown menu set to "Vous-même" and a "S'abonner" button with "2 455" subscribers. A navigation menu includes "Accueil", "Vidéos", "Playlists", "Chaînes", "Discussion", and "À propos". The "Vidéos" section is active, showing "Pour les abonnés" and "Pour les nouveaux visiteurs". Under "Vidéos à regarder ensuite", there are three video thumbnails: "Equilibre et ACP globale" (16:35, 278 vues), "Cours d'Analyse Factorielle Multiple (partie 4/4 : de François Husson" (16:35, 278 vues), and "Cours d'Analyse Factorielle Multiple (partie 3/4 : étude de François Husson" (19:26, 309 vues). A "Conseils" section on the right lists tips like "Six conseils pour créer une communauté" and "Générer des revenus sur YouTube".

MOOC analyse de données multidimensionnelles



AGRO CAMPUS OUEST

Analyse des données multidimensionnelles

Vous êtes inscrit à ce cours
Voir le contenu du Cours

Les thématiques: Méthodes et statistiques Sciences pour l'agriculture

A propos du cours
Cette 2ème édition augmentée du cours d'analyse de données multidimensionnelles débute le 31 mars 2016.

Fin d'inscription: Inscription close
Début du Cours: 01 mars 2016
Fin du cours: 05 mai 2016
Effort estimé: 05.00 h/semaine



The R Journal



Journal of Statistical Software
April 2016, Volume 70, Issue 1
doi:10.18188/jss.v070.i01

FACTOMINER

MOOC analyse de données multidimensionnelles

MOOC = Massive Open Online Courses

Objectif : comprendre et savoir mettre en œuvre 5 méthodes d'analyse des données (ACP, AFC, ACM, classification et AFM)

Le MOOC a été conçu pour les applications : nombreux exemples et implémentation logiciel (FactoMineR)

Trois sessions déjà jouées :

- mars 2014 sur la plateforme Moodle d'Agrocampus
- mars 2015 sur la plateforme France Université Numérique
- mars 2016 sur la plateforme France Université Numérique

MOOC analyse de données multidimensionnelles

MOOC sur 5 semaines : chaque semaine est focalisée sur l'étude d'une méthode (par exemple : Analyse en Composantes Principales)

Chaque semaine on trouve :

- des vidéos de cours : description de la méthode
- des quiz et exercices sur table
- une vidéo sur la mise en œuvre logicielle
- un exercice sur ordinateur
- (facultatif) des vidéos "Pour aller plus loin"

Egalement un forum et un Wiki

MOOC analyse de données multidimensionnelles

Cours

Contenu du cours
Info Cours
Discussion
Wiki
Progression
Diaporama
Transcription audio
Quiz

Analyse en composantes principales

- 1. Données - problématique Quiz
- 2. Etude des individus et des variables Quiz
- 3. Aides à l'interprétation Quiz

Didacticiel FactoMineR

Exercices Exercice

Pour aller plus loin

» Analyse factorielle des correspondances

» Analyse des correspondances multiples

» Classification

» Pour conclure

CONSEILS

Nous vous recommandons de lancer la vidéo puis de cliquer sur le bouton "HD" en bas à droite des vidéos pour bénéficier d'une lecture en haute définition.

Il est possible de supprimer les sous-titres en cliquant sur la **bulle** en bas à droite des vidéos et sur "Subtitles off".

VIDÉO : DONNÉES - PROBLÉMATIQUE EN ACP

☰

Données - Exemples

Etude des individus Etude des variables Aides à l'interprétation

Les données température

- 15 individus (lignes) : villes de France
- 14 variables (colonnes) :
 - 12 températures mensuelles moyennes (sur 30 ans)
 - 2 variables géographiques (latitude, longitude)

	Jan	Fev	Mars	Avr	Mai	Jun	Juillet	Août	Sept	Oct	Nov	Déc	Long	Lat
Bordeaux	5.6	6.8	10.3	14.1	17.8	19.2	20.9	21.1	19.0	13.8	9.1	6.2	44.3	-10.34
Brest	6.1	5.8	7.8	9.1	11.0	11.6	11.6	11.6	12.7	12.1	9.7	7.1	48.38	-4.20
Clermont	2.6	3.7	7.5	10.0	13.8	16.1	16.4	16.1	15.2	11.2	6.6	3.6	46.47	3.05
Grenoble	1.5	3.2	7.7	10.8	14.5	17.8	20.1	18.8	14.7	11.4	6.5	2.3	45.1	5.43
Lille	2.4	2.9	6.1	9.5	12.6	15.1	17.1	15.7	11.7	6.4	6.1	3.5	50.30	3.64
Lyon	2.1	3.3	7.7	10.9	14.9	17.9	19.1	18.1	16.9	11.4	6.7	3.1	45.76	4.11
Marseille	5.5	6.6	9.1	13.1	16.8	20.8	23.3	22.8	19.9	15.1	10.2	6.9	43.18	5.24
Montpellier	5.6	6.7	9.9	12.8	16.2	20.1	22.7	22.3	19.3	14.6	10.5	6.5	43.50	3.53
Nantes	5.1	6.3	8.4	10.8	13.9	17.2	18.8	16.6	16.4	12.2	8.2	5.5	47.13	-1.10
Nice	7.5	8.5	10.8	13.3	16.7	20.1	22.7	22.5	20.3	16.1	11.6	8.2	43.82	7.15
Paris	3.4	4.1	7.6	10.7	14.3	17.5	19.1	18.7	16.1	11.4	7.1	4.3	48.82	2.2
Rennes	4.8	5.3	7.9	10.1	13.1	16.2	17.9	17.8	15.7	11.6	7.8	5.4	48.05	-1.41
Strasbourg	0.4	1.5	5.5	9.8	14.1	17.2	19.1	18.3	15.1	9.5	4.9	1.3	48.56	7.45
Toulouse	4.7	5.6	9.2	11.6	14.9	18.7	20.9	20.8	18.3	13.3	8.6	5.5	43.30	1.26
Valéry	2.4	3.4	7.1	9.9	13.6	17.1	19.3	18.8	16.1	11.6	6.6	3.4	46.00	3.26

MOOC analyse de données multidimensionnelles

Contenu du cours Info Cours Discussion Wiki

Analyse en composantes principales

Cours

1. Données - problématique Quiz

2. Etude des individus et des variables Quiz

3. Aides à l'interprétation Quiz

Logiciel

Didacticiel **FactoMineR**

Exercices Exercice

Pour aller plus loin

Analyse factorielle des correspondances

Analyse des correspondances multiples

Classification

Pour conclure

Vidéo Diaporama Lignes de code

Données

Données des cours sur les températures dans 15 villes de France

- Le jeu de données
- Le script au format texte
- Le script et les sorties R

Données de didacticiel sur les athlètes au décathlon

- Le jeu de données decathlon
- Le script au format texte
- Le script et les sorties R

MOOC analyse de données multidimensionnelles

Cours

Logiciel

Exercices

Contenu du cours
Info Cours
Discussion
Wiki
Progression

- Analyse en composantes principales

1. Données - problématique Quiz

2. Etude des individus et des variables Quiz

3. Aides à l'interprétation Quiz

Didacticiel FactoMineR

Exercices

Exercice

Pour aller plus loin

> Analyse factorielle des correspondances

> Analyse des correspondances multiples

> Classification

> Pour conclure

Exercice sur table

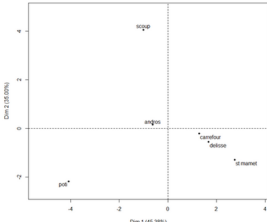
Exercice sur ordinateur

<
☰
☰
>

EXERCICE 1. LECTURE DE GRAPHIQUES D'ACP (6 points possibles)

Lors d'une analyse sensorielle, un jury a évalué six compotes de pomme selon une liste de 11 descripteurs comme l'acidité, l'amertume, la saveur de pomme crue, etc. Les notes ont été regroupées dans un tableau croisant en lignes les compotes de pomme, en colonnes les descripteurs et dans une case du tableau la note pour un produit et un descripteur. Les descripteurs sont de trois sortes, saveur (S), texture (T) ou odeur (O). Par exemple S.sucre correspond à la saveur sucrée. Par ailleurs, une note d'appréciation (note hédonique) a été donnée à chaque compote. Une ACP a été réalisée sur la tableau de données, la note d'appréciation ayant été mise en tant que variable supplémentaire. Les graphes des individus et des variables sont fournis ci-dessous.

Individuelle factor map (PCA)



MOOC analyse de données multidimensionnelles

Contenu du cours Info Cours Discussion Wiki Progression

- Analyse en composantes principales
 1. Données - problématique Quiz
 2. Etude des individus et des variables Quiz
 3. Aides à l'interprétation Quiz
 Didacticiel FactoMineR
 Exercices Exercice
 Pour aller plus loin

- Analyse factorielle des correspondances
 - Analyse des correspondances multiples
 - Classification
 - Pour conclure

Cours

Logiciel

Exercices

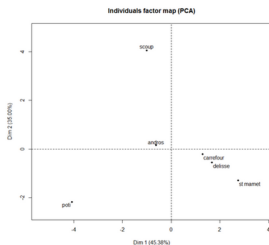
➔

➔



EXERCICE 1. LECTURE DE GRAPHIQUES D'ACP (6 points possibles)

Lors d'une analyse sensorielle, un jury a évalué six compotes de pomme selon une liste de 11 descripteurs comme l'acidité, l'amertume, la saveur de pomme crue, etc. Les notes ont été regroupées dans un tableau croisant en lignes les compotes de pomme, en colonnes les descripteurs et dans une case du tableau la note pour un produit et un descripteur. Les descripteurs sont de trois sortes, saveur (S), texture (T) ou odeur (O). Par exemple S.sucre correspond à la saveur sucrée. Par ailleurs, une note d'appréciation (note hédonique) a été donnée à chaque compote. Une ACP a été réalisée sur la tableau de données, la note d'appréciation ayant été mise en tant que variable supplémentaire. Les graphes des individus et des variables sont fournis ci-dessous.



MOOC analyse de données multidimensionnelles

FORUM

- Discussion entre apprenants
- Aide sur la programmation

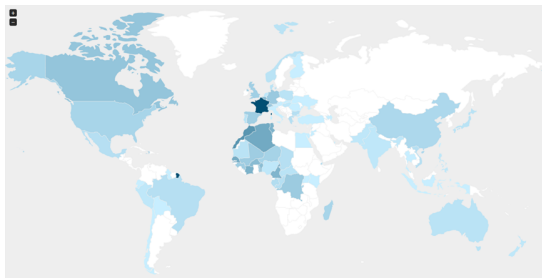
WIKI

- Travail collaboratif
- Jeux de données fournis
- Apprenants proposent une interprétation ou améliorent celle des autres apprenants
- Co-construction converge vers une analyse détaillée

Une évaluation finale

- Exercice sur ordinateur
- Possibilité d'obtention d'une *attestation de suivi avec succès*

Quelle était l'audience de ce MOOC ?



≈ 4500 inscrits en 2016

≈ 5000 en 2015

99 pays

63% de Français

18 à 78 ans ($\bar{x} = 35$)

60% Master, 18% pHD

- Apprenants du monde industriel et académique : Instituts de recherche (inra, inserm, ird, etc.), Universités (France, Turquie, Tunisie, Zaïre, etc.), Entreprises
- Différents champs disciplinaires : biologie, génomique, économie, géographie, linguistique, etc.

⇒ Audience très variée avec des buts différents et des niveaux d'investissement différents

Quels intérêts à utiliser R ?

- Gratuit
- Très connu
- Disponible partout dans le monde (avec la même version)
- Même utilisation quelque soit le système d'exploitation
- Lignes de code \Rightarrow reproductibilité des analyses
- Exemples avec Rmarkdown

Le MOOC pour le grand public

Les apprenants ont des objectifs et niveaux d'implication différents
⇒ MOOC conçu pour être suivi de différentes façons :

- voir seulement les vidéos de cours
- faire les quiz
- mettre en œuvre sur logiciel
- faire les exercices sur table
- faire les exercices sur logiciel
- participer à la rédaction de wiki sur des études de cas

Temps de travail estimé si lecture complète : 5h / semaine
Diverses audiences enrichissent le MOOC : animation wiki, forum

MOOC *versus* enseignement en présentiel

Dans l'apprentissage par un MOOC, les participants peuvent

- gérer personnellement leur temps de travail
- apprendre certaines parties du cours plus que d'autres
- discuter avec d'autres étudiants : apprentissage participatif
- poser des questions sans hésitation sur le forum
- reprendre le cours ultérieurement (ils doivent juste se rappeler qu'ils ont vu la méthode durant le MOOC)

Mais un cours en présentiel a aussi des avantages !

⇒ utilisation du MOOC en **présentiel** et **formation continue** :
une complémentarité d'avenir ?

Le MOOC pour les étudiants d'Agrocampus

Organisation :

- module de 25h d'analyse de données
- 192 étudiants de M1 – 3 enseignants (pour 10 heures)

Contenu :

- mêmes cours, quiz, exercices, forum, wiki
- permanences pour discuter du cours et de leur projet
- évaluation par un quiz supplémentaire et un projet

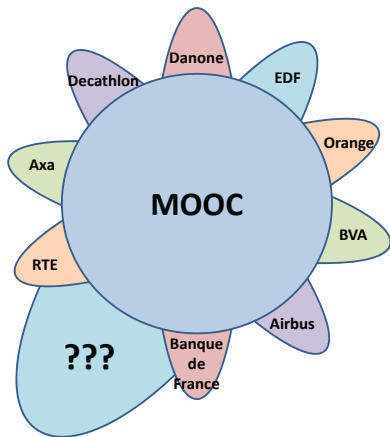
Points positifs :

- étudiants plus autonomes
- questions-discussions avec les étudiants plus intéressantes
- effet positif des discussions et commentaires des apprenants (cours utile)
- plus de temps pour discuter et échanger sur leur projet

Limites :

- certains étudiants préfèrent la douce tranquillité du présentiel
- ceux qui ne travaillent pas sont difficiles à remotiver

Utilisation du MOOC en entreprise



Possibilité d'offrir une même formation à tous ses employés
⇒ facilite la communication

Possibles ajustements

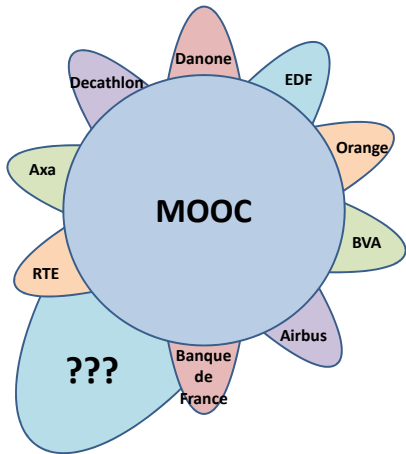
- quiz, exercices
- études de cas
- (forum)

Contributions spécifiques

- animation : mail, annonce, ...
- live, meetup

Entreprises bénéficient à la fois des avantages du MOOC (Massive Open ...) et des avantages du SPOC (Small Private Online Courses)

Comment contribuer à un tel MOOC ? (une proposition)



- contribution financière de base pour les entreprises suivant le MOOC
- tout gratuit pour les étudiants : les enseignants ont passé (beaucoup) de temps pour construire le MOOC
- tout gratuit pour les participants des pays en voie de développement
- contribution intermédiaire pour les entreprises voulant qq adaptations
- contribution plus importante pour les entreprises voulant qq animations, meetup ou live
- possibilité de payer pour "de bonnes vieilles séances de formation continue"

Le MOOC selon moi

Ce que ça apporte

- Tout ce dont j'ai déjà parlé mais aussi ...
- une visibilité accrue (chaîne Youtube 700 vues/j, MOOC \approx 130 000 vues)
- un tutoriel de référence très complet pour un package
- un guide complet pour un ensemble de méthodes
- une ouverture vers de nouveaux publics
- une incitation à repenser l'enseignement et la pédagogie

MAIS

- demande un fort investissement : 1h de cours \approx 20-30h !!!
Séquençage, quiz et exercices sont aussi chronophages
Les MOOC vieilliront-ils bien dans le temps ?

D'un package à un projet sur l'analyse de données

- avr 2006 : package FactoMineR (PCA, CA, MCA, MFA, etc.)
- avr 2007 : article sur FactoMineR (Journal of Statistical Software)
- avr 2007 : site internet
- nov 2007 : package RcmdrPlugin.FactoMineR (menu déroulant)
- juin 2009 & fév 2016 : Analyse de données avec R
- mai 2010 : package missMDA (gestion des données manquantes)
- jan 2011 : Exploratory Multivariate Analysis by Example Using R
- avr 2011 : tutoriels vidéos
- sep 2011 : Google group
- nov 2012 : Análisis de datos con R
- mar 2013 : chaîne Youtube (vidéos de cours et didacticiel)
- juin 2013 : article sur MFACT (R Journal)
- fév 2014 : cours en ligne d'analyse de données (en local)
- fév 2015 : package Factoshiny (graphes interactifs)
- **mar 2015 & 2016 : MOOC d'analyse de données (FUN)**
- juin 2015 : article sur CA-Galt (R Journal)
- avr 2016 : article sur missMDA (Journal of Statistical Software)
- packages factoextra et explor (graphes interactifs)

D'un package à un projet sur l'analyse de données

- avr 2006 : package FactoMineR (PCA, CA, MCA, MFA, etc.)
- avr 2007 : article sur FactoMineR (Journal of Statistical Software)
- avr 2007 : site internet
- nov 2007 : package RcmdrPlugin.FactoMineR (menu déroulant)
- juin 2009 & fév 2016 : Analyse de données avec R
- mai 2010 : package missMDA (gestion des données manquantes)
- jan 2011 : Exploratory Multivariate Analysis by Example Using R
- avr 2011 : tutoriels vidéos
- sep 2011 : Google group
- nov 2012 : Análisis de datos multivariados
- mar 2013 : chaîne YouTube (cours vidéo et cours interactif)
- juin 2013 : article sur MFACT (R Journal)
- fév 2014 : cours en ligne d'analyse de données (en local)
- fév 2015 : package Factoshiny (graphes interactifs)
- mar 2015 & 2016 : MOOC d'analyse de données (FUN)
- juin 2015 : article sur CA-Galt (R Journal)
- avr 2016 : article sur missMDA (Journal of Statistical Software)
- packages factoextra et explor (graphes interactifs)

MERCI