



L'objectif de ce séminaire est d'échanger autour du logiciel de statistique libre, gratuit et multiplateforme *R* (<http://www.r-project.org/>). Il s'adresse aux praticiens impliqués dans le traitement quantitatif des données en sciences humaines et sociales (utilisateurs de données, chercheurs, ingénieurs ou étudiants) qu'ils aient ou non déjà utilisé le logiciel *R*. Chaque séance est organisée autour de la présentation d'une expérience de traitement de données avec le logiciel (fonction spécifique et/ou packages). Le cadre de ces réunions est informel et les participants doivent se sentir libres d'intervenir afin de confronter leurs expériences.

Élisabeth Morand et Bénédicte Garnier (Ined),
Florence Maillouchon (CMH-Cnrs), Pascal Cristofoli (EHESS)
<http://www.ehess.fr/fr/enseignement/enseignements/2015/ue/1010/>

Séance 2 : jeudi 21 janvier 2016 de 9h à 12h

Analyse Textuelle avec R par *Mónica Bécue-Bertaut*
(*Universidad Politècnica de Catalunya*) et *Annie Morin*
(*Université de Rennes I*)

Sous le nom d'analyse textuelle, on signifie l'analyse de corpus de textes, ou documents, par des méthodes statistiques multidimensionnelles. Ces corpus peuvent être de très grande taille (plus de 500000 occurrences) mais c'est loin d'être toujours le cas.

TextoMineR, package R, permet d'analyser les textes par diverses méthodes statistiques multidimensionnelles, parmi lesquelles l'analyse des correspondances (AC) est privilégiée. Ce package permet aussi d'utiliser l'analyse factorielle multiple pour tableaux de contingence (AFMTC), l'analyse de correspondances pour tableaux lexicaux agrégés généralisés (CA-GALT) ainsi que les méthodes de classification hiérarchique, avec ou sans restriction de contiguïté, et

l'extraction des mots et phrases caractéristiques de catégories de documents. Ainsi, non seulement le contenu verbal des documents est dévoilé, mais en outre la structure du corpus est appréhendée. Replacer le contenu verbal dans la structure temporelle, hiérarchique, etc. du corpus est essentiel pour l'interprétation.

De nombreux exemples montrent que le même type de codification et les mêmes méthodes permettent de traiter des problématiques bien différentes telles que celles posées par les questions ouvertes et l'analyse de questions ouvertes en différentes langues ; les bases bibliographiques ; les bases de sentences judiciaires ; les scripts de film ; les réquisitoires, plaidoiries ou autres textes rhétoriques ; les descriptions verbales en sensométrie : les corpus de tweets.

Nombre de places limité à 30 - **inscription indispensable à**

<http://www.ined.fr/fr/actualites/rencontres-scientifiques/seminaires-colloques-ined/russ-21-janvier-2016/>