

Inscription sur

<https://www.ined.fr/fr/actualites/rencontres-scientifiques/seminaires-colloques-ined/7-2-2019/>

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),
Timothée Giraud (CNRS UMS Riate), Pascal Cristofoli (EHESS)
<https://enseignements-2018.ehess.fr/2018/ue/1010/>

Séance 2 : jeudi 7 février 2019 (9h30-12h)

Organiser et traiter plus facilement ses données avec R : utiliser le package dplyr

Elodie Baril (Ined) et Arnaud Bringé (Ined)

La gestion de données sous R apparaît souvent compliquée et peut constituer un frein pour des utilisateurs habitués au management de données sous d'autres logiciels statistiques ou des débutants en R. Cette séance s'adresse aux personnes souhaitant s'initier à cette syntaxe claire et cohérente, sous formes de verbes permettant de traiter et manipuler facilement les données contenues dans une ou plusieurs tables R.

Ce package, proposé en 2016, fait partie du tidyverse (<https://www.tidyverse.org/>) et a l'avantage de présenter une grammaire unifiée. Il semble aujourd'hui incontournable pour qui souhaite traiter ses données avec R en toute autonomie.

Grâce à sa syntaxe plus « naturelle » dplyr permet de sélectionner des lignes/colonnes, créer des variables, regrouper de données et enfin calculer des statistiques descriptives. Des fonctions servent également à renommer les variables, ordonner les données, fusionner des tableaux très simplement. Elles sont en général plus rapides que leur équivalent sous R de base et sont adaptées au traitement de données de grande dimension.

Nous verrons à l'aide d'un exemple concret comment préparer ses données, expliciter les éléments de la grammaire de dplyr, enchaîner les opérations élémentaires pour constituer un fichier d'analyse et enfin calculer facilement des statistiques agrégées.