

# Formation R pour la data analyse

(3 jours)

Claire Della Vedova (claire.della-vedova@datavalue.fr)  
Data Value (formation@datavalue.fr)

## Objectif

S'approprier les outils et les méthodes permettant de rendre son processus d'analyse de données avec R, aisé, reproductible et performant. Acquérir une grande dextérité dans la manipulation des données grâce aux packages dplyr, tidyr et lubridate. Apprendre à réaliser des rapports d'analyses, dynamiques et automatisés, avec les packages R markdown, et des graphiques performants avec le package ggplot2.

## Public

A toute personne amenée dans le cadre de son activité professionnelle, ou ses études, à analyser des données avec le logiciel R, et souhaitant rendre son processus de travail plus performant.

## Prérequis

Il est nécessaire d'avoir une connaissance de base du logiciel R : savoir qu'il existe différentes classes de données (numeric, factor, character), savoir manipuler à minima les vecteurs et les data frame : création, extraction d'un élément, savoir importer des données, etc. Dans l'idéal, le participant a suivi la formation Bien Débuter avec R, ou atteint un niveau équivalent par la pratique.

## Méthode

La formation se déroule selon une alternance de théorie, puis de mise en pratique, ensemble en pas à pas, puis d'exercices à réaliser en autonomie.

Les parties théoriques sont présentées à l'aide de diapositives qui vous seront transmises au format pdf, et que vous pourrez conserver.

Les mises en pratiques, en pas à pas, s'appuient sur des tutoriels qui vous seront transmis sous format de pages html, que vous pourrez conserver, comme des aides mémoires.

## Programme (3 jours)

### Introduction

- Origine et évolution de R
- Place de R dans la data analyse

## Rappels

- Les outils de R Studio
- Organiser son travail sous R : projet R et architecture
- Bonnes pratiques pour construire un dataset
- Importations, vérification et exportation de données
- Importation et exportation de scripts et d'environnements
- Mises à jours des packages, de RStudio et de R

## Manipulations de données avec le package dplyr

- Introduction au package tidyverse et à la notion de pipe
- Filtrer des lignes avec la fonction filter()
- Sélectionner des colonnes (variable) avec la fonction select()
- Création de nouvelles variables avec la fonction mutate()
- Calcul de paramètres par sous groupes : fonctions group\_by() et summarise()
- Autres fonctions intéressantes

## Manipulations des variables catégorielles avec le package forecats

- Inspection des variables catégorielles avec les fonctions levels(), fct\_count et fct\_unique()
- Modifier l'ordre des modalités
- Modifier le nom des modalités
- Réaliser des regroupements de modalités

## Manipulations des chaînes de caractères avec le package stringr

- Détection de patterns
- Découpage
- Gestion des longueurs
- Remplacement

## Manipulations des données de date : utilisation du package lubridate

- Convertir les données au format YYYY-MM-DD et HH:MM:SS
- Décomposer les éléments d'année, de mois et de jour
- Calculer des différences de dates et les exprimer en jours, ou heure

## Assemblage et pivot de tables

- Jointure de tables: left join, right join, inner join et full join
- Combinaison de tables par les lignes
- Les formats wide et long
- Passage d'un format wide à long et inversement

## Représentations graphiques avec le package ggplot2

- Le principe des couches successives de ggplot2
- Réalisation des graphiques de base : scatterplot, barplots, line plot, boxplots

- Représentation des séries temporelles
- Utilisation du format long et facetting
- Gestion des couleurs, titres, axes et légendes

## **Création de rapports dynamiques et automatisés avec rmarkdown**

- Principe, formats de sorties (html, docx, pdf)
- Les différents éléments d'un fichier Rmd : en-tête, chunk, etc. . .
- Gestion des éléments de texte : gras, italique, titre
- Gestion des tables
- Gestion des images
- Gestion des graphiques
- Gestion des options des éléments de code
- Importation de données
- Gestion de la table des matières et numérotation

## **Introduction à la programmation fonctionnelle avec le package purrr (module optionnel - en fonction du rythme de la formation, mais supports pédagogiques transmis dans tous les cas)**

- Les lists
- Les fonctions `apply()`, `lapply()`, `tapply()`
- Les fonctions `map()`
- Les Nested data