

Formation R - Data Analyse avancée

Objectifs : Perfectionner son utilisation du langage R dans l'optique de structurer, automatiser et restituer efficacement des analyses de données reproductibles.

Compétences visées : - Structurer un projet R reproductible en organisant scripts, données et bonnes pratiques de codage

- Mettre en œuvre des traitements de données avancés avec le tidyverse, incluant pivots, jointures spécifiques, facteurs et chaînes de caractères
- Automatiser des traitements analytiques en concevant des fonctions R robustes et des itérations avec purrr
- Analyser et optimiser les performances de traitements de données volumineuses à l'aide de data.table et duckDB
- Produire des visualisations et tableaux de synthèse professionnels, statiques ou interactifs, adaptés à la restitution analytique
- Générer et automatiser des rapports d'analyse reproductibles à l'aide de Quarto et de rapports paramétrés

Durée : 3 jour(s) (21 heures)

Public : Toute personne utilisant déjà le langage R pour la réalisation d'analyses de données, de reporting ou d'études statistiques.

Pré-requis : Une pratique minimale de R et des packages du tidyverse (dplyr, ggplot2) est requise.

Méthode pédagogique : Pédagogie active mêlant exposés, exercices et applications pratiques dans le logiciel R.

Tarifs :

- Présentiel : 1950 € HT
 - Distanciel : 1800 € HT
- (-10% pour 2 inscrits, -20% dès 3 inscrits)

Option(s) :

- Forfait déjeuners : 75 € HT

Nos prochaines sessions

Programme :

- Rappels R et bonnes pratiques générales

- Projets R et chemins relatifs
- Organisation de son projet et ses répertoires
- Importations de fichiers et problématiques courantes

- L'approche tidyverse vs R base
- Bonnes pratiques de codage
- Outils d'aide (lintr / styler)
- Séparer son code en plusieurs fichiers
- Versionner son code avec git (principes généraux uniquement)

Pratique : passer en revue un script R « fonctionnel mais perfectible » et l'améliorer en mettant en place des bonnes pratiques de codage et d'organisation

- Tidyverse avancé

- Rappels sur les fonctions clés du tidyverse
- Principes de la « tidy data » et fonctions de pivot de données
- Jointures spécifiques (semi_join, anti_join)
- Traitements sur de multiples colonnes avec across
- Sélection de multiples colonnes avec le tidyselect
- Facteurs : packageforcats et bonnes pratiques sur les facteurs
- Chaînes de caractères et expressions régulières

Pratique : réaliser divers traitements analytiques courants impliquant notamment gestion de facteurs, pivots de données et traitements sur de multiples colonnes d'un dataframe

- Automatisation et fonction

- Rappels sur la création et l'appel de fonctions
- Intégration de messages et warning dans une fonction
- Approche pratique du tidyeval
- Bonnes pratiques dans l'écriture des fonctions
- Itérations avec purrr

Pratique : refactoriser un script R en encapsulant dans des fonctions diverses étapes du traitement de données

- Gestion de données massives

- Evaluer les temps de traitements de manipulations de données
- Fondamentaux du package data.table
- Traitements de données courants avec data.table
- Data.table vs dplyr : avantages et inconvénients
- Connexion à duckDB pour des traitements accélérés

Pratique : mettre en place des traitements analytiques à partir d'une base de données conséquente, en utilisant différentes approches et en comparant leurs performances

- Dataviz & tableaux

- Rappels ggplot2 : fondamentaux, thèmes personnalisés
- Crédit de visualisations interactives avec ggiraph
- Autres packages pour la dataviz : spécificités et avantages / inconvénients
- Crédit de tableaux de synthèse formatés avec GT
- Personnalisation de tableaux GT
- Autres packages pour les tableaux : spécificités et avantages / inconvénients
- Introduction à la cartographie avec sf

Pratique : créer des graphiques interactifs et des tableaux de synthèse, avec un rendu propre et professionnel

- Reporting avec Quarto

- Fondamentaux de Quarto : objectifs, principes, différences avec R markdown
- Options de code des chunks
- Personnalisation du YAML
- Intégration de tableaux et graphiques dans un rapport Quarto
- Trucs et astuces Quarto
- Rapports paramétrés
- Génération de rapports Quarto en ligne de commande

Pratique : automatiser la création de plusieurs rapports d'analyse avec un rendu propre et professionnel

Modalités d'évaluation : Un formulaire d'auto-évaluation proposé en amont de la formation nous permettra d'évaluer votre niveau et de recueillir vos attentes. Ce même formulaire soumis en aval de la formation fournira une appréciation de votre progression.

Des exercices pratiques seront proposés à la fin de chaque séquence pédagogique pour l'évaluation des acquis.

En fin de formation, vous serez amené(e) à renseigner un questionnaire d'évaluation à chaud.

Une attestation de formation vous sera adressée à l'issue de la session.

Trois mois après votre formation, vous recevrez par email un formulaire d'évaluation à froid sur l'utilisation des acquis de la formation.

Qualification du formateur : Les formations DATA VALUE sont animées par des intervenants experts, alliant expérience opérationnelle et compétences pédagogiques reconnues. Ils disposent de plus de 10 ans d'expérience professionnelle et assurent une veille constante sur les pratiques, outils et normes de leur domaine afin de garantir des contenus actualisés et adaptés aux besoins des stagiaires.

Délais d'accès et modalités de formation : Les inscriptions peuvent être effectuées jusqu'à 48 heures avant le début de la formation.

Vous pourrez trouver des informations plus détaillées concernant les modalités de nos formations à cette adresse : <https://www.datavalue.fr/modalites-acces-formation>.

Accessibilité : Vous souhaitez suivre notre formation R - Data Analyse avancée et êtes en situation de handicap ? Merci de nous contacter afin que nous puissions envisager les adaptations nécessaires et vous garantir de bonnes conditions d'apprentissage

Date de dernière modification : 29 janvier 2026