Formation R pour la data analyse

Objectifs : S’approprier les outils, et les méthodes permettant de rendre son processus d’analyse de données sous R, aisé, reproductible, et performant.

Compétences visées : - Adopter une organisation efficace de son travail sous R
- Simplifier la manipulation de données avec le package dplyr
- Manipuler aisément les champs catégoriels (package forecats), textuels (package stringr) et de date (package lubridate)
- Assembler des tables par jointure (par colonnes) et par fusion (par lignes)
- Construire des graphiques élaborés grâce au package ggplot2
- Réaliser des documents automatisés à l'aide de Quarto
- Découvrir la programmation fonctionnelle au travers du package purrr

Durée : 3 jour(s) (21 heures)

Public : Data Analysts, Data Scientists, Statisticiens et plus généralement toute personne intéressée par la Data Analyse avec R.

Pré-requis : Pour suivre ce stage dans de bonnes conditions, il est recommandé d'avoir suivi en amont la formation R – Prise en main, analyses statistiques et graphiques

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

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.

Accessibilité : Vous souhaitez suivre notre formation Formation par ville 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

Tarifs :
- Présentiel : 1650 € HT
- Distanciel : 1500 € HT
(-10% pour 2 inscrits, -20% dès 3 inscrits)

Option(s) :
- Forfait déjeuners : 60 € HT

Nos prochaines sessions

Distance
du 11 au 13 juin 2025
du 8 au 10 décembre 2025

Lyon
du 22 au 24 septembre 2025

Paris
du 4 au 6 juin 2025
du 29 sept. au 1 oct. 2025
du 3 au 5 décembre 2025

Toulouse
du 12 au 14 mai 2025
du 17 au 19 novembre 2025

Programme :

- Organiser son travail sous R (2h)

- Travailler en projet R : notion de working directory, workspace, history
- Architecture de son projet R : data, plots, images, scripts, ...
- Bonnes pratiques pour la création de fichiers de données
- Importation et exportation de fichiers avec le package here
- Mettre à jour ses packages
- Mettre à jour R et R Studio

- Manipuler facilement ses données avec le package dplyr (3h)

- 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()
- Renommer ses variables avec la fonction rename()
- Calcul de paramètres par sous-groupes : fonctions group_by() et summarise()
- Passage du format wide au format long
- Exercices

- Manipulation des variables catégorielles avec le package forecats (1h)

- Inspecter les variables catégorielles avec les fonctions levels(), fct_count et fct_unique
- Modifier l'ordre des modalités
- Modifier le nom des modalités
- Exercices

- Manipuler les chaînes de caractères avec le package stringr (1h)

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

- Manipuler des données de date : utilisation du package lubridate (1h)

- 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 heures
- Exercices

- Assemblage de tables (2h)

- Les différentes jointures (par colonne): left join, right join, inner join et full join
- Assemblage par lignes
- Exemple d'applications pour l'analyse de données
- Exercices

- Réaliser des représentations graphiques performantes avec le package ggplot2 (4h)

- Le principe des couches successives de ggplot2

- Réalisation des graphiques de base : scatterplot, barplots, line plot, boxplots
- Gestion des couleurs, titres, axes et légendes
- Représentation des séries temporelles
- Utilisation du format long et facetting
- Ajouter du texte sur un graphique (ex : équation)
- Exporter son graphique : format et résolution
- Utilisation des addins esquisse et Colour Picker

- Générer dynamiquement son rapport d'analyse avec Quarto (4h)

- Principe, formats de sorties (html, docx, pdf)
- La structure d'un script .qmd
- Mise en forme des parties texte
- Gestion de l'affichage du code et de son résultat
- Importation de données
- Insertion de graphiques, de tables, d'images
- Gestion de la table des matières et numérotation
- Les scripts paramétrés : automatisation de documents répétitifs

- Introduction à la programmation fonctionnelle avec le package purrr (2h)

- Les list
- Les fonctions map()
- Nested data
- Exercices

Date de dernière modification : 16 avril 2025