DellaDATA
Claire DELLA - VEDOVA

Utilisation de R – niveau 2

R pour la data analyse (3 jours – 21 h)

A qui s'adresse cette formation ?

Cette formation s'adresse aux personnes qui utilisent déjà R et souhaitent **aller au-delà des bases**, structurer leur flux de travail et adopter de bonnes pratiques pour l'analyse de données et le reporting reproductible.

Elle est particulièrement adaptée aux :

- personnes amenées à analyser régulièrement des données avec R dans le cadre de leur activité professionnelle ou de leurs études
- utilisateurs et utilisatrices de R ayant acquis des bases par une formation d'initiation ou par la pratique, et souhaitant gagner en efficacité et en fiabilité
- ingénieur(e)s, chargé(e)s d'études, doctorant(e)s, chercheur(se)s ou personnels techniques produisant des analyses sous R
- biostatisticien(ne)s, data managers et chargé(e)s d'études travaillant sur des données de santé ou de recherche
- professionnel(le)s intervenant dans des institutions de santé ou agences publiques (par exemple agences sanitaires, autorités de santé, structures exploitant des données médico-administratives telles que le PMSI)
- personnes souhaitant optimiser la manipulation de données avec le tidyverse et produire des analyses reproductibles et des rapports dynamiques avec Quarto

Cette formation s'inscrit comme une **suite logique** à une formation de prise en main de R.

Elle peut également être complétée par des ateliers thématiques d'une journée ou des formations plus spécialisées.

Ce que cette formation n'est pas

Cette formation n'a pas vocation à :

- enseigner l'ensemble du langage R ni couvrir tous les usages possibles de R en data science
- se substituer à un cursus académique complet en statistique, en informatique ou en data science
- aborder des développements avancés relevant de la programmation logicielle (optimisation de performances, développement de packages, programmation orientée objet avancée)
- proposer des méthodes « clés en main » sans réflexion sur la structuration du code, la reproductibilité et la qualité des analyses

Elle vise au contraire à :

- consolider et structurer l'utilisation de R pour l'analyse de données
- adopter des bonnes pratiques de manipulation, de transformation et de visualisation des données
- fiabiliser les analyses et les rendre reproductibles
- poser des bases solides pour un usage professionnel de R, et pour poursuivre vers des formations ou ateliers plus spécialisés

Prérequis

Il est nécessaire d'avoir **déjà utilisé le logiciel R** et de disposer de bases opérationnelles.

Les participant(e)s doivent notamment :

- savoir travailler dans RStudio et utiliser des scripts R
- connaître les principales structures de données (vecteurs, data frames)
- savoir importer des données tabulaires (CSV, Excel)
- comprendre et exécuter du code R simple

Cette formation s'adresse à des personnes ayant suivi une formation d'initiation à R ou ayant acquis un niveau équivalent par la pratique.

La formation **ne reprend pas les bases de prise en main de R** (installation, découverte de l'interface, notions élémentaires du langage), mais vise à approfondir et structurer l'analyse de données avec R.

Objectifs et compétences visées

Cette formation vise à permettre aux participant(e)s de **structurer et optimiser leur utilisation de R pour l'analyse de données**, en adoptant des bonnes pratiques de travail, de manipulation et de nettoyage des données, ainsi que de reporting reproductible.

À l'issue de la formation, les participant(e)s seront capables de :

- organiser et structurer efficacement leur travail sous R à l'aide de projets clairs et de scripts lisibles
- mettre en place un **versionnage des scripts** afin de sécuriser et tracer l'évolution des analyses
- importer des données de manière robuste, en **automatisant et en accélérant les importations** de fichiers tabulaires
- nettoyer des données de façon systématique et reproductible
- (uniformisation des variables, gestion des doublons, corrections et contrôles de cohérence)
- enchaîner les **étapes de manipulation des données** nécessaires à l'analyse (sélection, transformation, agrégation, jointures, changement de format)
- gérer différents types de variables (facteurs, chaînes de caractères, dates) de manière fiable
- produire des analyses descriptives et des visualisations adaptées aux données et aux objectifs
- automatiser la production de résultats et de **rapports dynamiques avec Quarto**
- améliorer la lisibilité, la maintenabilité et la réutilisation de leur code R

Cette formation permet ainsi de **gagner en efficacité, en fiabilité et en reproductibilité** dans le traitement et l'analyse de données sous R, et constitue un socle solide pour un usage professionnel avancé du logiciel.

Programme de la formation

Module 1 – Structurer son travail sous R

- Création et structuration de projets R
- Architecture des projets et bonnes pratiques
- Versionnage des scripts en local avec Git

Module 2 – Reporting reproductible avec Quarto

- Principes du reporting reproductible
- Premiers éléments : texte, code, tableaux et graphiques
- Structuration d'un document Quarto
- Numérotation, table des matières, légendes et renvois

Module 3 – Importation avancée des données

- Gestion des problèmes d'encodage
- Accélération des importations
- Automatisation des importations de fichiers CSV et Excel

Module 4 – Manipulations avancées des data frames

- Introduction au méta-package tidyverse
- Création de sous-ensembles, ajout et transformation de variables
- Création de tables agrégées par groupe
- Autres manipulations utiles

Module 5 – Nettoyage de données

- Uniformisation des noms des variables
- Création de labels
- Gestion des doublons
- Nettoyage et corrections diverses

Module 6 – Manipulations supplémentaires

- Jointure de tables à l'aide d'une clé primaire
- Assemblage de tables de même structure
- Passage en format long et wide

Module 7 – Travailler avec différents types de variables et s'initier à la programmation fonctionnelle

- Variables catégorielles avec forcats
- Chaînes de caractères avec stringr et glue
- Dates et heures avec lubridate
- Initiation à la programmation fonctionnelle avec purrr (quelques exemples)

Module 8 – Réalisation de graphiques avec ggplot2

- Principe du package
- Premiers graphiques (scatterplots, etc.)
- Gestion des couleurs, formes, tailles et courbes de tendance
- Découpage en sous-fenêtres
- Autres types de graphiques

Utilisation raisonnée de l'intelligence artificielle

Tout au long de la formation, des outils d'intelligence artificielle peuvent être utilisés **de manière transversale** pour accompagner l'écriture, la compréhension et l'amélioration de scripts R.
L'intelligence artificielle est mobilisée comme un **outil d'assistance au travail d'analyse de données** (aide à la formulation de code, à la compréhension de scripts existants, à l'amélioration de traitements), dans une démarche critique et encadrée.
Elle ne se substitue ni à la compréhension des méthodes, ni au raisonnement analytique.

Modalités pratiques

⏱️ Durée : 3 jours (21 heures)

💻 Format : distanciel (visioconférence), présentiel possible selon les besoins

👥 Effectif : Groupes à taille limitée (environ 5), favorisant les échanges et les questions

📘 Pédagogie et organisation :
Alternance d'apports théoriques et de mises en pratique
Supports remis aux participant(e)s. Temps dédiés aux questions et à l'interprétation

🧰 Logiciels : R et R Studio, Git

📝 Évaluation des acquis : Exercices pratiques, mises en situation et échanges oraux tout au long de la formation.

🏷️ Tarif
1 500 € HT par participant(e) pour la formation de 3 jours (21 h) - Supports pédagogiques inclus.
Les tarifs en intra-entreprise sont établis sur devis, en fonction du format et des modalités de la formation.
Organisme exonéré de TVA en vertu de l'article 261-4-4°a du CGI.

✔️ Suivi et assiduité : Émargement par demi-journée.

⚙️ Adaptation et accessibilité :
Les exemples peuvent être adaptés au contexte des participants en intra-entreprise.
Toute situation de handicap ou besoin spécifique peut être signalé en amont afin d'étudier les adaptations possibles

Informations de contact

Organisme de formation : **DellaDATA**
NDA : 93040100204 auprès du préfet de région PACA
SIRET : 90838795400013
Organisme exonéré de TVA en vertu de l'article 261-4-4°a du CGI.
Adresse : 1087 CH de Ste Roustagne - 04100 Manosque
Formatrice et interlocutrice : Claire Della Vedova (claire@delladata.fr / 06.64.98.21.42)