

Manipulations des données avec R

Structurer et transformer les données avec dplyr et lubridate (1 jour – 7 h)

A qui s'adresse cette formation ?

Cette formation s'adresse aux personnes travaillant avec des données sous R et souhaitant gagner en efficacité et en fiabilité dans la manipulation des tables de données.

Elle est particulièrement adaptée aux :

- data analystes, statisticien(ne)s, biostatisticien(ne)s,
- professionnel(le)s de la santé (médecins DIM, chargé(e)s d'études),
- chargé(e)s de projet, ingénieur(e)s, technicien(ne)s,
- enseignant(e)s, chercheur(se)s, doctorant(e)s et étudiant(e)s,

dans tous les domaines d'application (santé, environnement, agronomie, sciences sociales, industrie, etc.).

Ce que cette formation n'est pas

Cette formation n'a pas vocation à :

- couvrir l'ensemble des fonctionnalités avancées du tidyverse,
- se substituer à une formation longue en data science ou en statistiques,
- fournir des scripts "clé en main" sans compréhension des mécanismes,

Elle vise au contraire à développer de bons réflexes, à structurer les manipulations de données et à renforcer l'autonomie des participant(e)s dans la préparation des données sous R.

Prérequis

- Connaître les bases du logiciel R : exécution de scripts, manipulation d'objets simples, importation de données, lecture et compréhension d'un code R existant.
- Aucun prérequis spécifique en dplyr, tidyr ou lubridate n'est nécessaire.

Objectifs et compétences visées

Cette formation vise à permettre aux participant(e)s de manipuler efficacement des tables de données sous R en s'appuyant sur les principaux packages du tidyverse.

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

- comprendre la logique de manipulation des données avec dplyr, tidyr et lubridate,
- filtrer des observations et sélectionner des variables pertinentes,
- créer, transformer et recoder des variables,
- regrouper des données et produire des résumés,
- joindre plusieurs tables à l'aide de clés communes,
- restructurer des données entre formats long et wide,
- manipuler des données de dates et de temps de manière fiable.

Ces compétences permettent de sécuriser les étapes de préparation des données et de gagner en efficacité dans tout projet d'analyse sous R.

Programme de la formation

Module 1 – Organisation et introduction au tidyverse

Objectif : comprendre l’environnement de travail et la place des packages de manipulation.

- Organisation d’un projet R et architecture de travail
- Introduction aux packages dplyr, tidyr et lubridate
- Place de ces packages dans l’écosystème tidyverse
- Utilisation du pipe

Module 2 – Manipulation des tables avec dplyr et tidyr

Objectif : maîtriser les opérations courantes de manipulation des données.

- Principales fonctions de manipulation des tables
- Gestion des conflits de packages
- Filtrer des observations
- Sélectionner des variables
- Créer ou transformer des variables
- Regrouper des observations et les résumer
- Combinaison de tables par jointure
- Pivoter une table en formats long et wide

Module 3 – Gestion des dates et des temps avec lubridate

Objectif : manipuler des données temporelles sans erreurs courantes.


- Problèmes fréquents liés aux données de dates et de temps
- Principales fonctions du package lubridate
- Conversion de chaînes de caractères en dates et temps
- Extraction des composantes (année, mois, jour, heure...)
- Arrondis des dates et des temps
- Ajouter, soustraire et calculer des délais


Utilisation raisonnée de l’intelligence artificielle

Tout au long de la formation, des outils d’intelligence artificielle sont utilisés comme **supports pédagogiques** afin de clarifier la logique des manipulations de données avec R, d’aider à comprendre la syntaxe des fonctions, et d’accompagner la structuration et l’écriture du code utilisant **dplyr**, **tidyr** et **lubridate**.

L'IA est mobilisée comme **outil d'assistance à l'apprentissage et à la structuration des manipulations de données**, dans une démarche **critique, encadrée et complémentaire des apprentissages**, sans se substituer à la compréhension des mécanismes, ni aux choix méthodologiques réalisés par les participant(e)s.

Modalités pratiques

 **Durée** : 1 jour (7 heures)

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

 **Effectif** : groupes à taille limitée (environ 5 participant(e)s)

 **Pédagogie et organisation** :

La formation repose sur une alternance d'apports méthodologiques et de mises en pratique.

Des supports pédagogiques sont remis aux participant(e)s.

Des temps dédiés aux questions et à l'appropriation des concepts sont prévus tout au long de la journée.

- **Logiciels utilisés** : R et RStudio
- **Évaluation des acquis** : exercices pratiques, mises en situation et échanges oraux
- **Suivi et assiduité** : émargement par demi-journée

 **Cadre administratif et tarif**

- **Tarif inter-entreprise** : 500 € HT par participant(e) (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.

 **Adaptation et accessibilité**

Les exemples peuvent être adaptés au contexte des participant(e)s 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)