

## Formation Langage R : Fondamentaux + Data Science

■ <b>Durée :</b>	5 jours (35 heures)
■ <b>Tarifs inter-entreprise :</b>	3 475,00 € HT (standard) 2 780,00 € HT (remisé)
■ <b>Public :</b>	Développeurs
■ <b>Pré-requis :</b>	Notions d'algorithmique
■ <b>Objectifs :</b>	<p>Découvrir le langage R et son écosystème pour la data science. Apprendre à écrire du code structuré, lisible et maintenable en R. Importer, nettoyer et préparer des données issues de multiples sources. Maîtriser la manipulation de données avec le tidyverse et les pipelines. Créer des visualisations statiques et interactives impactantes. Produire des documents reproductibles pour le reporting. Réaliser des analyses statistiques descriptives et inférentielles. S'initier à la modélisation statistique et au machine learning avec R.</p>
■ <b>Modalités pédagogiques, techniques et d'encadrement :</b>	<ul style="list-style-type: none"><li>• Formation synchrone en présentiel et distanciel.</li><li>• Méthodologie basée sur l'Active Learning : 75 % de pratique minimum.</li><li>• Un PC par participant en présentiel, possibilité de mettre à disposition en bureau à distance un PC et l'environnement adéquat.</li><li>• Un formateur expert.</li></ul>
■ <b>Modalités d'évaluation :</b>	<ul style="list-style-type: none"><li>• Définition des besoins et attentes des apprenants en amont de la formation.</li><li>• Auto-positionnement à l'entrée et la sortie de la formation.</li><li>• Suivi continu par les formateurs durant les ateliers pratiques.</li><li>• Évaluation à chaud de l'adéquation au besoin professionnel des apprenants le dernier jour de formation.</li></ul>
■ <b>Sanction :</b>	Attestation de fin de formation mentionnant le résultat des acquis
■ <b>Référence :</b>	PRO102049-F

**Note de****■ satisfaction des participants:**

4,50 / 5

**■ Contacts :**

commercial@dawan.fr - 09 72 37 73 73

**■ Modalités d'accès :**

Possibilité de faire un devis en ligne (www.dawan.fr, moncompteformation.gouv.fr, maformation.fr, etc.) ou en appelant au standard.

**■ Délais d'accès :**

Variable selon le type de financement.

**■ Accessibilité :**

Si vous êtes en situation de handicap, nous sommes en mesure de vous accueillir, n'hésitez pas à nous contacter à referenthandicap@dawan.fr, nous étudierons ensemble vos besoins

## Introduction à R et à son écosystème

### Présentation de R et usages

Statistique, data science, visualisation

Exemples de secteurs

Écosystème R

CRAN

Packages : définition, installation, gestion

Atelier : installation et chargement de packages

RStudio et organisation de projet

Création de projets

Organisation des répertoires, chemins relatifs

### Atelier : créer un projet RStudio

### Écrire du code en R

Vecteurs

Listes

Data frames

Matrices

Facteurs

### Atelier : création et manipulation de chaque type

### Programmation en R : structures, fonctions et bonnes pratiques

Normes de codage  
Nommage, commentaires  
Organisation des scripts  
Structures de contrôle  
Conditionnelles : if, else  
Boucles : for, while, repeat  
apply, lapply, map

**Atelier : écrire des structures conditionnelles et des boucles**

## Création de fonctions

Syntaxe de base  
Arguments, valeurs de retour  
Fonctions anonymes  
Bonnes pratiques

**Atelier : automatiser des analyses avec ses propres fonctions**

## Importation et préparation des données

### Techniques d'importation

Formats : .csv, .txt, .xls(x), .sav, JSON  
Packages : readr, readxl, haven, jsonlite  
Connexions bases de données : DBI, dbplyr  
Encodage, séparateur, gestion des NA

**Atelier pratique : importer différents formats**

## Introduction au tidyverse

Notion de pipeline  
Manipulation avec dplyr et tidyr  
Filtrer, sélectionner, trier  
Créer des variables  
Fusionner, joindre  
Reshaping : pivot\_longer, pivot\_wider  
Gestion des valeurs manquantes

**Atelier : pipeline complet avec dplyr et tidyr**

## **Visualisation et reporting avec R**

### **Principes de la data visualisation**

Importance exploratoire

Types de graphiques adaptés

### **Introduction à ggplot2**

Syntaxe de base

Histogrammes, boxplots, scatterplots, bar charts

Facets et thèmes

### **Atelier pratique: graphiques statiques**

### **Visualisations interactives**

plotly pour ggplot

Cartes interactives avec leaflet (optionnel)

### **Atelier pratique : transformer un ggplot en interactif**

### **Production de documents**

RMarkdown, Quarto, Notebooks

Avantages et usages

Atelier : création d'un document reproductible

Gestion de versions avec Git

Intégration RStudio

Initialisation, commits

### **Atelier pratique : prise en main de Git avec RStudio**

## **Analyse statistique et Machine Learning avec R**

### **Statistiques descriptives**

Moyenne, médiane, écart-type

Résumés, quantiles, tableaux de fréquence

### **Atelier pratique: résumer un dataset**

### **Tests d'hypothèses**

t-tests, ANOVA, chi-deux  
Corrélations : Pearson, Spearman  
Packages : base R, broom

### **Atelier pratique : réaliser et interpréter des tests**

### **Modélisation supervisée**

Régressions linéaire et logistique  
Analyse des résidus  
Visualisation des résultats

### **Atelier : ajuster et interpréter des modèles**

### **Analyse multivariée : ACP, AFC (FactoMineR, factoextra)**

### **Atelier pratique : réaliser une ACP**

### **Apprentissage non supervisé**

K-means clustering  
Préparation des données, choix du k  
Visualisation des clusters

### **Atelier pratique : segmentation (identifier des sous-populations aux caractéristiques similaires)**

### **Introduction aux séries temporelles**

Composantes d'une série  
Modèles ARIMA, SARIMA (forecast)  
Visualisation des prévisions