

Formation Langage R : Data Science

■ Durée :	2 jours (14 heures)
■ Tarifs inter-entreprise :	1 975,00 € HT (standard) 1 580,00 € HT (remisé)
■ Public :	Développeurs R
■ Pré-requis :	Avoir suivi la formation Langage R : fondamentaux ou notions équivalentes
■ Objectifs :	Acquérir les bases solides pour importer, nettoyer et manipuler des données avec R. Explorer et visualiser des données de façon efficace. Produire des analyses statistiques descriptives et inférentielles. Implémenter des algorithmes simples de machine learning supervisés et non supervisés. Savoir utiliser les principaux packages R pour la Data Science.
■ Modalités pédagogiques, techniques et d'encadrement :	• Formation synchrone en présentiel et distanciel. • Méthodologie basée sur l'Active Learning : 75 % de pratique minimum. • Un PC par participant en présentiel, possibilité de mettre à disposition en bureau à distance un PC et l'environnement adéquat. • Un formateur expert.
■ Modalités d'évaluation :	• Définition des besoins et attentes des apprenants en amont de la formation. • Auto-positionnement à l'entrée et la sortie de la formation. • Suivi continu par les formateurs durant les ateliers pratiques. • Évaluation à chaud de l'adéquation au besoin professionnel des apprenants le dernier jour de formation.
■ Sanction :	Attestation de fin de formation mentionnant le résultat des acquis
■ Référence :	PRO102048-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

Importation et manipulation des données

Importer des données de formats variés

CSV avec readr, base R

Excel avec readxl

JSON avec jsonlite

Connexion bases de données (dbplyr, DBI)

Aperçu des packages : haven (SPSS, Stata)

Atelier pratique : Importer différents formats de données

Maîtriser la manipulation des données dans R

Filtrage, sélection, mutation

Agrégation et group_by

Tri et réorganisation

Jointures (left_join, inner_join, etc.)

Gestion des valeurs manquantes : is.na, replace_na

Nettoyage des formats de colonnes (dates, facteurs, strings)

Atelier pratique : pipeline complet avec dplyr et tidyr

Visualisation des données

Principes de la data visualisation

Importance de la visualisation exploratoire

Types de graphiques adaptés aux données

Représentations graphiques classiques et avancées

Création de graphiques avec ggplot2

Syntaxe de base

Types : histogrammes, boxplots, scatterplots, bar charts

Facets et thèmes

Atelier pratique : Créer des graphiques statiques avec ggplot2

Visualisations interactives

Introduction à plotly pour ggplot

Cartes interactives avec leaflet (optionnel)

Atelier pratique : Transformer un ggplot en graphique interactif

Statistiques descriptives et inférentielles

Écrire des statistiques descriptives à l'aide de R

Moyenne, médiane, écart-type

Résumés et quantiles

Tableaux de fréquence

Outils pour la description statistique

Atelier pratique : Résumer un jeu de données réelles

Tests d'hypothèses

t-tests (une ou deux échantillons)

ANOVA

Test du chi-deux

Corrélation de Pearson et Spearman

Packages : base R, broom

Atelier pratique : Réaliser et interpréter des tests statistiques

Modélisation statistique et Machine Learning

Analyse descriptive des datasets multivariés et techniques de réduction dimensionnelle

ACP (Analyse en Composantes Principales)

Packages : FactoMineR, factoextra

Analyse Factorielle des Correspondances (AFC)

Atelier pratique : Réaliser une ACP et interpréter les axes

Algorithmes supervisés avec études de cas sur R

Régression linéaire simple et multiple (lm)

Régression logistique (glm)

Analyse des résidus
Visualisation des résultats
Exemple pratique sur un jeu de données

Atelier pratique : Ajuster et interpréter des modèles supervisés

Algorithmes non supervisés avec études de cas sur R

K-means clustering
Préparation des données
Choix du nombre de clusters
Visualisation des clusters

Atelier pratique : Segmenter des données avec k-means

Introduction à l'Analyse des séries temporelles

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

Atelier pratique : Prévoir des valeurs avec ARIMA