

# Programmation R

5 j (35 heures)

Ref : IAD006

## Public

Ingénieurs, Data Analysts, statisticiens, développeurs en environnement statistique

## Pré-requis

Connaissances en statistiques descriptives et inférentielles  
Expérience en programmation appréciée mais non indispensable  
Avoir suivi une formation à Python et/ou à la data science est un plus

## Moyens pédagogiques

Formation réalisée en présentiel ou à distance selon la formule retenue  
Nombreux exercices pratiques et mises en situation, échanges basés sur la pratique professionnelle des participants et du formateur, formation progressive en mode participatif. Vidéoprojecteur, support de cours fourni à chaque stagiaire

## Modalités de suivi et d'évaluation

Feuille de présence émargée par demi-journée par les stagiaires et le formateur  
Exercices de mise en pratique ou quiz de connaissances tout au long de la formation permettant de mesurer la progression des stagiaires  
Questionnaire d'évaluation de la satisfaction en fin de stage  
Auto-évaluation des acquis de la formation par les stagiaires  
Attestation de fin de formation

Cette formation intensive de 5 jours vous permettra de maîtriser le langage R et son utilisation en data science. Vous apprendrez à manipuler les différentes structures de données, réaliser des analyses statistiques, créer des visualisations et mettre en œuvre des algorithmes de machine learning. Grâce à de nombreux exercices, vous découvrirez les packages les plus populaires pour le data wrangling (dplyr, tidyr), la data visualisation (ggplot2) et le machine learning (caret). Vous serez initié aux bonnes pratiques de développement avec R et verrez comment intégrer R à d'autres langages et outils.

À la fin de la formation, vous serez autonome pour réaliser des projets data science complets avec R.

## Objectifs

- Comprendre les concepts fondamentaux du langage R
- Savoir manipuler les vecteurs, matrices, listes et dataframes
- Réaliser des opérations de data wrangling avec dplyr et tidyr
- Maîtriser les bases de la programmation fonctionnelle avec R
- Produire des graphiques avancés et interactifs avec ggplot2 et plotly
- Conduire des analyses statistiques (tests, modèles linéaires...)

Programmation R

Mettre en œuvre des techniques de machine learning supervisées et non supervisées

Développer des applications web interactives avec Shiny

Intégrer R à d'autres langages et outils dans des projets data science

## Programme détaillé

### FONDAMENTAUX DE R

---

#### INTRODUCTION AU LANGAGE R

---

Historique et caractéristiques clés du langage

Comparaison avec les autres langages data science (Python, SAS, MATLAB...)

Installation de R et de l'IDE RStudio

#### SYNTAXE ET OBJETS DE BASE EN R

---

Types de données (numeric, character, logical...)

Opérateurs et fonctions de base

Bonnes pratiques : conventions de nommage, indentation, commentaires...

#### STRUCTURES DE DONNEES EN R

---

Vecteurs et facteurs

Matrices et arrays

Listes

Dataframes

#### IMPORTATION ET EXPORTATION DE DONNEES

---

Lecture/écriture de fichiers plats (CSV, TXT...)

Connexion à une base de données

Scraping de données depuis une page web

#### INTRODUCTION A LA PROGRAMMATION FONCTIONNELLE

---

Concept d'immutabilité

Application de fonctions avec lapply/sapply/tapply

Écriture de fonctions en R

#### DATA WRANGLING ET VISUALIZATION

---

#### MANIPULATION DE DONNEES AVEC DPLYR

---

Opérations principales : select, filter, mutate, arrange, summarise...

Enchaînement d'opérations avec les pipes

Jointures de dataframes

## **RESHAPING DE DONNEES AVEC TIDYR**

---

Passage du format wide au format long avec gather

Passage du format long au format wide avec spread

Séparation et combinaison de colonnes

## **GESTION DES DONNEES MANQUANTES ET DES OUTLIERS**

---

Détection et visualisation des valeurs manquantes/aberrantes

Suppression et imputation des valeurs manquantes

Gestion des outliers

## **DATA VISUALISATION AVEC GGLOT2**

---

Concepts clés de la grammaire des graphiques

Esthétiques, géométries et thèmes

Création de graphiques avancés : small multiples, graphiques interactifs...

## **PROJET DIRIGE : ANALYSE EXPLORATOIRE D'UN JEU DE DONNEES**

---

Import, nettoyage et transformation des données

Analyses univariées et multivariées

Visualisations avancées et interprétation des résultats

## **STATISTIQUES AVEC R**

---

### **STATISTIQUES DESCRIPTIVES AVEC R**

---

Indicateurs de tendance centrale et de dispersion

Quantiles et boxplots

Corrélations et associations

### **TESTS STATISTIQUES PARAMETRIQUES**

---

Tests de comparaison (t-test, ANOVA...)

Tests d'indépendance (chi-2, test exact de Fisher)

Corrélations et régressions linéaires simples

### **TESTS STATISTIQUES NON PARAMETRIQUES**

---

Tests de rang (Wilcoxon, Kruskal-Wallis...)

Corrélations de Spearman

## **MACHINE LEARNING AVEC R**

---

### **INTRODUCTION AU MACHINE LEARNING**

---

Types d'apprentissage : supervisé, non supervisé, par renforcement  
Compromis biais-variance, sur- et sous-apprentissage  
Validation croisée et optimisation des hyperparamètres

### **PREPARATION DES DONNEES POUR LE ML**

---

Encodage des variables catégorielles  
Normalisation et standardisation des variables numériques  
Sélection de variables

### **ALGORITHMES DE ML SUPERVISES**

---

Régression  
Classification  
Arbres de décision et forêts aléatoires

### **ALGORITHMES DE ML NON SUPERVISES**

---

Clustering (K-means, hierarchical clustering...)  
Réduction de dimension (ACP, MDS, t-SNE...)

### **PROJET DIRIGE : PREDICTION SUPERVISEE SUR UN JEU DE DONNEES REEL**

---

Analyse du problème et des données  
Feature engineering et sélection de variables  
Comparaison et optimisation de différents algorithmes  
Interprétation des résultats

### **DEVELOPPEMENT ET INDUSTRIALISATION AVEC R**

---

### **DEVELOPPEMENT D'APPLICATIONS WEB AVEC SHINY**

---

Composants UI et Server  
Inputs et outputs (widgets, graphiques, tables...)  
Mise en page et thèmes  
Exemples d'applications

### **PROJET FIL ROUGE : DEVELOPPEMENT D'UN DASHBOARD SHINY**

---

Analyse des besoins et maquettage  
Importation et préparation des données

Programmation R

Création des différents éléments graphiques

Mise en production du dashboard

## **CONCLUSION ET PERSPECTIVES**

---

Récapitulatif des points clés de la formation

Panorama des packages R pour la data science

Ressources pour approfondir et monter en compétences

Retour d'expérience sur les métiers de la data science

---