

Réf : ET_DAT_R

Langage R

Manipulation de données pour la Data Science

Objectif

R est un langage statistique dédié à la Data Science très riche en fonctionnalités.

L'ensemble de ressources proposé s'adresse à toute personne qui souhaite apprendre à extraire des données issues de sources variées (CSV, Excel, base de données SQL, API Web) avec ce langage pour être en mesure ensuite de les manipuler et les organiser.

Vous apprendrez notamment à filtrer, trier, regrouper, résumer, fusionner ou nettoyer des données à l'aide de différents packages.

Public

Informaticiens

Durée estimée

pour le suivi des modules indispensables

5 à 7 heures

Contenu pédagogique

► Les modules indispensables

Langage R - Les fondamentaux de la programmation R pour la Data Science



Vidéo

Ce module vous propose la consultation d'une vidéo d'une durée de 2h23.

- **Découverte de R**
 - Présentation de R
 - Installation et ressources
 - Le système de packages
 - IDE RStudio et création d'un premier script
- **Les bases du langage R**
 - Les variables et les types
 - Les formats : les vecteurs
 - Les formats : les matrices
 - Les formats : les DataFrame
 - Les formats : les listes
 - Les conditions (if ... else)
 - Les boucles
 - Les fonctions
 - Les fonctions de type apply
- **R et la data**
 - Récupérer les données d'un fichier au format CSV
 - Récupérer d'autres formats de données : Excel, SQL, API
- **Analyses statistiques et premiers graphiques**
 - Analyse de variables numériques
 - Analyse de variables catégorielles
 - Autres méthodes d'analyses statistiques
- **Restitution de résultats : aperçu des possibilités**
 - Les graphiques

- Les cartes
- La création de reporting avec le package markdown
- La création de tableaux de bord web avec le package shiny

Langage R - Accédez à tous vos types de données



Vidéo

Ce module vous propose la consultation d'une vidéo d'une durée de 1h52.

- **Quelques rappels sur R et son environnement**
 - Pourquoi R ?
 - Installation de l'environnement
 - Ressources liées à R
- **Données au format CSV**
 - Accéder à des données au format CSV
 - Mise en pratique - Data Visualisation : graphique en bâtons
- **Données au format Excel**
 - Accéder à des données au format Excel
 - Mise en pratique - Data Visualisation : nuage de points
- **Données d'une base de données SQL**
 - Accéder à des données provenant d'une base de données SQL
 - Mise en pratique - Data Visualisation : nuage de mots
- **Données provenant d'une API**
 - Accéder à des données provenant d'une API
 - Mise en pratique - Data Visualisation : graphique dynamique en bâtons pour le web
- **Données provenant d'une base de données NoSQL MongoDB**
 - Accéder à des données provenant d'une base de données NoSQL
 - Mise en pratique - Data Visualisation : tableau dynamique pour le web

Langage R - Manipulez vos données avec les packages dplyr et tidyr



Vidéo

Ce module vous propose la consultation d'une vidéo d'une durée de 1h45.

- **Rappels sur l'accès avec R à différents formats de données**
 - Accès à des données au format CSV
 - Accès à d'autres formats de données (Excel, bases de données SQL, API...)
- **Manipulation à l'aide du package dplyr**
 - Introduction au package "dplyr" : syntaxe et "pipe"
 - Manipulation d'observations (lignes)
 - Manipulation de variables (colonnes)
 - Création des tableaux de données résumées
 - Jointures de données
- **Introduction au nettoyage de données avec le package tidyr**
 - Remodeler des données
 - Compléter les valeurs manquantes