



R CONSOLIDATION

Objectifs pédagogiques :

- ▶ Savoir mettre en place son environnement de travail R
- ▶ Savoir utiliser une documentation à bon escient tout au long d'un projet statistique
- ▶ Maîtriser l'ingestion de données (connaître différents formats d'import, connaître les fonctions basiques d'import des principaux packages R d'ingestion, découvrir l'import de fichiers volumineux)
- ▶ Maîtriser les fonctions basiques et la syntaxe du package dplyr.
- ▶ Savoir manipuler et présenter des données classiques.

Prérequis

- Appétence pour l'approche analytique et quantitative
- Connaissances de base en programmation (concepts de boucles, conditions ...)
- Connaissances des notions de statistiques, d'analyse de données, d'économétrie et de Machine Learning

Modalités pédagogiques :

Public :

- ▶ Data scientists avec connaissances de base en R (avoir déjà utilisé R, connaître dplyr)

Déroulé de la formation :

- ▶ Durée : 8 demi-journées

Modalités d'évaluation des participants :

- ▶ L'évaluation se fera sous forme d'exercices avec fil directeur lors d'une séance dédiée.

Moyens spécifiques de la formation :

- ▶ Formateur avec une expérience significative en R et une formation de data scientist (> 2 ans)

Moyens pédagogiques :

- Support de présentation
- Exercices pratiques
- Bases de données pour faire les exercices
- ▶ Dispositif de suivi de la formation : feuille de présence

Panorama des séquences :

Séquence 1 : Devenir autonome en R

Environnement de travail





Travailler avec des projets sur R Studio
Architecture de dossier, documentation, log

Séquence 2 : Manipuler des données simples

Bonnes pratiques générales (dépendances, commentaires, etc.)

Les objets R

Manipuler une table (visualisation, opérations de base)

Séquence 3 : Manipuler des données complexes

Tri de données et analyse avec le package tidyr

Jointures

Traitement de données textuelles avec le package stringr et introduction aux expressions régulières

Séquence 4 : Produire et exporter des résultats

Export de données

Package gt

Quarto

Séquence 5 : Lire des données aux formats variés

Lire des fichiers .csv, .xlsx, .txt, SAS, Parquet

Bases de données et API

Sparklyr

Séquence 6 : Projet statistique - Modulariser ses analyses

Code robuste

Paramétrage

dplyr

Boucles et fonctions

Séquence 7 : Projet statistique - Structurer ses analyses

Création d'un fichier main et dépendances

Tidy data

Introduction à Git

Séquence 8 : Evaluation

► **Contact** : formations@veltys.com

► **01 87 44 69 70**