

5.1 – Le Système R – Niveau 1

Démarrer avec RStudio, Tidyverse et Rmarkdown pour manipuler ses données et faire ses graphiques

Texte d'accroche : L'objectif de la formation est d'acquérir les bases de Rstudio, du Tidyverse et de Rmarkdown, pour la manipulation de tableaux, la production de graphiques et la production de rapports. Cette formation se concentre sur les étapes de l'analyse de données en amont de l'analyse statistique et est basée sur le package et langage Tidyverse, conçu spécifiquement pour la manipulation de données et la production de graphiques.

Mots clés : Rstudio, Tidyverse, Rmarkdown

Durée totale de la formation (heures) : 14h

Organisation des séances (durée, horaires etc.) : 2 séances de 7h : de 9h à 12h30 et de 13h30 à 17h.

Les instructions pour l'installation des logiciels et packages, les données qui servent de support pour la formation, le pdf à imprimer contenant le code de la formation (hors exercices) et servant à prendre des notes, ainsi que le document Rmarkdown servant de modèle pour taper le code lors de la formation sont envoyés à chaque participante une semaine avant la formation.

Les personnes en situation de handicap peuvent me contacter afin que nous puissions trouver ensemble comment adapter la formation pour répondre à leurs besoins.

Mode de formation : présentiel – distanciel - hybride

Nombre de participants : De 6 à 15 personnes

Niveau (fondamental, intermédiaire ou avancé) : Fondamental. Cette formation est conçue pour les personnes n'ayant jamais utilisé R et Rstudio (ou ayant tout oublié).

Programme détaillé :

J1 - Séquence 1 : RStudio, bases de R et Rmarkdown

- L'interface RStudio: les 4 fenêtres, les menus déroulants
- La création d'objet et l'assignation, les fonctions, les classes (vecteurs numériques, vecteurs de caractères, dataframe) d'objets et les types de variables (facteur, date)
- Création et utilisation d'un document Rmarkdown

J1 - Séquence 2 : la syntaxe du Tidyverse

- La syntaxe d'un graphique avec le package ggplot2
- La syntaxe de la manipulation de tableaux et la manipulation de colonnes avec le package dplyr



J1 - Séquence 3 : vision d'ensemble de la chaîne du traitement de données

- Importer ses données, visualiser les données brutes, modifier le type des variables, transformer les données (calculs de moyennes, sélection de lignes), visualiser les données transformées

J1 - Séquence 4 : les tableaux de données à nombreuses variables

- Opérations conditionnelles, gestion des données manquantes

- Représentation de combinaisons de variables sur le même graphique, représentation des écart types

J2 - Rappels du Jour 1

- Exercices de lecture de code pour réviser les fonctions de base

J2 - Séquence 5 : l'amélioration des graphiques

- Les 4 types de calques de ggplot2 et les différents types de modification (couleurs, axes, légendes, polices)

J2 - Séquence 6 : fusion de tableaux

- Les deux manières essentielles de joindre des tableaux (collés et fusionnés)

J2 - Séquence 7 : restructurer et nettoyer ses tableaux

- Restructurer ses tableaux (passer du format large à long) avec le package tidyr

- Nettoyer une chaîne de caractère (package stringr) et organiser l'ordre des facteurs (package forcats)

Réactivation mémoire et cartographie des connaissances

- Réactivation mémoire individuelle des connaissances acquises et cartographie des connaissances communes (carte mentale avec le logiciel xmind)

Evaluations de la formation :

- autoévaluation individuelle de l'atteinte des objectifs (graphiques de réussite aux quiz et exercices)

- évaluation de satisfaction de la formation (questionnaire en ligne)

Compétences acquises à l'issue de la formation :

- A l'issue la formation, les participant-e-s seront capable de :
- Installer R et RStudio, installer un package, utiliser les quatre fenêtres de RStudio, créer des vecteurs numériques et de caractères et des tableaux de données
- Charger un tableau de données, vérifier et déclarer les types des variables (facteur, date)
- Effectuer les manipulations de données les plus courantes: filtrer des lignes, sélectionner des colonnes, faire des opérations (moyennes, écart-types) groupées



selon certaines conditions (modalités, traitements, année, etc.), rajouter des colonnes, faire un calcul entre colonnes

- Améliorer l'esthétique des graphiques pour la publication (couleurs, axes, police, légendes)
- Fusionner des tableaux de tailles différentes
- Restructurer et nettoyer les tableaux (corriger des chaînes de caractères, réorganiser les niveaux d'un facteur)
- Générer avec Rmarkdown un rapport html à partir du code

Méthode pédagogique :

La formation utilise l'interface conviviale de RStudio, qui facilite l'utilisation de R et le format de fichiers Rmarkdown, qui permet de générer automatiquement des rapports (html, word, pdf ou même des diaporamas) à partir du code. La méthode pédagogique consiste en un enchaînement de séquences progressives constituées d'une démonstration (apport théorique), suivie d'un quiz puis d'un exercice et mobilise des acquis des sciences cognitives en matière d'apprentissage pour renforcer la mémorisation. A cela s'ajoute une méthode pédagogique conçue pour les débutant-e-s et consistant à dédoubler la séquence de démonstration de la manière suivante :

- Une première démonstration, pendant lequel le formateur présente les connaissances et compétences à acquérir en tapant le code, pendant que les participant-e-s, sans taper le code, concentrent leur attention sur la compréhension des informations, les questions et la prise de note sur le document pdf envoyé en amont et à imprimer et contenant le code tapé par le formateur (ce qui facilite grandement la prise de notes).
- Une deuxième démonstration, pendant lequel le code est retapé cette fois-ci tous ensemble, afin que les participant-e-s pratiquent et fassent des erreurs corrigées par le formateur.

Cette méthode est utilisée le premier jour, la démonstration ne se faisant ensuite qu'en un seul temps une fois les participant-e-s à l'aise avec le langage et sa syntaxe. Cette méthode est très adaptée aux formations à des outils informatiques, qui sont la plupart du temps basées sur le faire ensemble (le formateur montre pas à pas l'enchaînement des opérations à accomplir) et dans lesquelles les participants sont très vite dépassés par la nécessité de faire plusieurs tâches simultanément : regarder le tableau pour voir les opérations tapées par le formateur, regarder ensuite son écran pour les taper, écouter les explications du formateur, prendre des notes, corriger ses erreurs. Ce mélange des tâches entraîne une forte diminution de la capacité d'écoute et de compréhension des apprenants, et une forte diminution de l'efficacité pédagogique de la formation. Plus de détails sur :

https://oswaldoforey.github.io/Website_bookdown/.

Langue de la formation : Français | Anglais