

Plan de Cours

Titre : initiation au langage de programmation R dans un contexte de recherche en sciences de la santé

Rédigé par : Reda Goudrar, Maxine Joly-Chevrier, Thomas Li, étudiants en médecine de l'Université de Montréal

Supervisé par : Dr Michaël Sauthier, MD, MBI, PhD

Cours préalables : Aucune connaissance préalable sur le logiciel R n'est requise. Par contre, des connaissances en mathématiques de base niveau CEGEP sont requises.

Description : Le projet consiste à créer une plateforme d'enseignement complémentaire sur le langage de programmation R dans un contexte de recherche en santé, disponible à l'ensemble de la communauté étudiante et professorale. À travers une vingtaine de capsules informatives gratuites, différentes notions de traitement de données, statistiques et méthodes d'analyse seront présentées. Le but est d'outiller les participants à ce logiciel qui est l'un des plus employés en recherche médicale actuellement. Le projet est supervisé par Dr Sauthier, pédiatre intensiviste et bioinformaticien, de l'Hôpital Ste-Justine. Pour la réalisation des capsules, ce projet rassemblera des étudiants de divers programmes (sciences de la santé et informatique), professeurs de ces programmes et chercheurs.

Objectifs généraux : Se familiariser avec les manipulations de base sur R dans un contexte de recherche en sciences de la santé.

Méthodes pédagogiques : Capsules vidéo partagées en accès libre sur un site Internet. Tutoriels interactifs sur learnr.

Évaluation : Évaluation formative en fin de module (exercice d'intégration) sur une base volontaire.

Matériel de référence (optionnels) :

- R for Data Science: Import, Tidy, Transform, Visualize, and Model Data. Wickham, H., Grolemund, G. O'Reilly. (2017).

Calendrier : les étudiants peuvent suivre cette formation à leur rythme. La formation est disponible en tout temps. Le temps estimé pour compléter cette formation est estimé à environ 20h.

Informations supplémentaires : Aucun crédit de formation ne sera attribué. Une attestation pourra être remise à la complétion de la formation.

I) Les bases

- 1. Introduction et présentation des applications concrètes de RStudio en recherche**
 - Présentation globale de R
 - Les différents devis des études (ex.: cohorte, cas-témoins, etc.)
 - Présentation d'une étude qui sera utilisée tout au long des capsules afin d'illustrer les différents concepts
 - Référencer à un glossaire ou à un lexique des termes

- 2. Familiarisation avec l'environnement de développement RStudio**
 - Installation de R et RStudio
 - Présentation globale de l'interface
 - Installation des modules
 - Utilisation de la console (calculs arithmétiques de base)

- 3. Présentation de la syntaxe**
 - Commentaires
 - Variables
 - Types de données (entiers, décimaux, caractères, facteurs, dates...)
 - Fonctions
 - Conditions et opérateurs logiques
 - Boucles

- 4. Notions de statistiques**
 - Moyenne, médiane, mode
 - Variance et écart type
 - Quartile et interquartile

- 5. Vecteurs**
 - Création
 - Obtenir un sous-ensemble de vecteur
 - Fonctions vectorielles (action sur chaque position du vecteur, ie `is.na()`)
 - Fonctions de regroupement (résumer le vecteur)
 - Calculs entre plusieurs vecteurs

- 6. Matrices**
 - Créer et nommer les éléments d'une matrice
 - Dimensions d'une matrice
 - Obtenir un sous-ensemble d'une matrice
 - Manipulation d'une matrice (ajout, suppression colonnes, transposition etc)
 - Fonctions de regroupement
 - Calculs avec matrices

7. Tableaux (*data frame*)

- Définition et différences avec une matrice
- Créer
- Importer des données (CSV et Excel)
- Inspecter
- Sélection colonnes ou lignes
- Modification (colonnes et lignes)
- Sauvegarder

II) Les données

8. Regroupement de données

- Principes
- Catégorisation des données (*mutate*, *case_when*)
- Fonctions d'agglomération (*count*, *min*, *max*, etc.)

9. Pivotage

- Format clé-valeur (ou *entity-attribute-value* ou *tidy*)
- Pivotage long vers large
- Pivotage large vers long

10. Combiner des tableaux

- Additionner tableaux (*bind_rows*, *bind_cols*)
- Clés primaires
- Jonctions

11. et 12. Visualiser des données

- Importance de la visualisation (e.g. quartet anscombe)
- Principes de grammar of graphics
- Graphiques (*ggplot*) (exemples avec plusieurs catégories)
 - Nuages de points (*ggplot*)
 - lignes
 - histogrammes
 - box plot/violin plot
 - alluvial plot
- Modifications de base : changer les titres des axes, l'échelle, titre principal, couleurs et légende, ajouter courbe de tendance...
- Sauvegarder (PDF, EPS, PNG) le graphique

III) Statistiques

13. Tester une hypothèse

- Valeur de p
- T-tests, régression linéaire
- Assumptions associated with hypothesis

14. ANOVA

- *One-Way* et *Two-Way* ANOVA
- Faire un test post-hoc
- Présenter les résultats sur un graphique

15. Estimateur de Kaplan-Meier

- Estimateur de Kaplan-Meier de la fonction de survie
- Représentation graphique de la courbe de survie

16. Randomisation d'un groupe de patients et chiffrement