

# Introduction à R/RStudio

Avril 2020

Au fil du temps, R est devenu un des outils les plus utilisés dans plusieurs domaines, autant en sciences naturelles qu'humaines, pour la manipulation des données, l'analyse statistique, et la création d'éléments de visualisation de haute qualité. Bien que l'utilisation de R au sein de Statistique Canada soit relativement récente, il s'y trouve une communauté d'utilisateurs chevronnés de R et plusieurs équipes commencent à s'y intéresser.

Le but de ce cours est d'offrir aux utilisateurs débutants une base solide dans les meilleures pratiques pour la préparation préalable (« data wrangling ») et l'analyse des données, tout en mettant l'accent sur des stratégies afin d'assurer la reproductibilité. Apprendre un nouveau langage de programmation peut être ardu et difficile. Même si vous avez de l'expérience en programmation, un nouveau langage amène son lot de défis qui peut vous faire sentir comme un réel débutant : nouvel environnement, nouveau style, nouvelle syntaxe. Tout au long de ce cours, les participants se feront proposer plusieurs exercices qu'ils pourront faire à leur rythme, et auront aussi accès à de l'aide personnalisée à leurs besoins.

Ce cours est conçu pour les débutants en R (particulièrement ceux avec une expérience en SAS), et les gestionnaires qui aimeraient apprendre comment R et Rstudio pourraient être utilisés par leur équipe. Des ressources additionnelles favorisant l'amélioration et la bonification des acquis seront suggérées au travers le matériel et la documentation fournis pour le cours.

## **Objectifs du cours:**

À la fin du cours, les utilisateurs seront en mesure de:

- 1) Effectuer des requêtes de base sur les données en utilisant dplyr et baseR
- 2) Effectuer des manipulations sur les données et produire des mesures descriptives
- 3) Joindre et concaténer des bases de données
- 4) Produire des graphiques sommaires pour l'exploration de données
- 5) Produire des graphiques de haute qualité prêts pour la publication en utilisant ggplot2

## **Élément du cours:**

### **Jour 1 :**

- Trucs et astuces pour utiliser Slack et Zoom
- GitHub et le matériel du cours
- Introduction à R et l'environnement Rstudio
  - Pourquoi R ?
- Installation de R et Rstudio : aide personnalisée si nécessaire !
- Attention : veuillez installer les logiciels sur un ordinateur ne nécessitant PAS l'accès au réseau de Statistique Canada pour fonctionner en ligne.

### **Jour 2 :**

- On se lance ! R comme calculatrice
- Création de variables
- Sous-ensembles de données (« subsetting ») et requêtes
- Introduction aux fonctions

### **Jour 3 :**

- Introduction aux bases de données
- Graphiques dans R

### **Jour 4 :**

- Installation et chargement des « packages »
- Apprentissage interactif avec le package SWIRL
- Importer des données (.csv, .sas)

### **Jour 5 :**

- Introduction au package dplyr
  - select, filter, summarize, mutate
- Le 'piping' : combiner plusieurs opérations en une

### **Jour 6 :**

- Joindre des bases de données
- Graphiques de haute qualité : ggplot2

### **Jour 7 :**

- Révision
- Encore plus de R : sujets spéciaux