Programmation

Dans cette partie, on présente quelques références sur les bases de la programmation dans différents langages. Ce cours ne spécifie pas de langage à utiliser, vous êtes donc libre de choisir celui qui vous convient le mieux pour faire les différents exercices, ainsi que votre projet. Dans tous les cas, il y a un ensemble de bonnes pratiques qui permet d'avoir un code lisible, compréhensible et réutilisable. Vous trouvez un (rapide) guide des bonnes pratiques en programmation à ce lien.

1 R

R est un langage de programmation spécialisé dans l'analyse statistique et la visualisation de données. Il est gratuit, open-source et disponible sur Windows, macOS et Linux. Régulièrement mis à jour, de nouveaux packages sont disponibles chaque jour (cf. CRAN).

Voici un guide qui présente les bases de R : Apprendre R en Y minutes et un guide de bonnes pratiques en R. Enfin, un point important du langage R est l'opérateur *pipe*. Celui-ci permet de rendre le code plus clair. Vous pouvez trouver une explication de cette opérateur ici.

2 Python

Python est un langage de programmation généraliste. Il est gratuit, open-source et disponible sur Windows, macOS et Linux. Bien que généraliste, il y a une importante communauté autour du l'analyse de données, machine learning en Python. Ainsi, les méthodes usuelles ont déjà été implementées, e.g. dans le package sklearn, mais si besoin, il est possible d'aller voir le code pour le modifier.

Voici un guide qui présente les bases de Python : Apprendre Python en Y minutes et un guide de bonnes pratiques en Python.

3 Julia

Julia est aussi un langage de programmation généraliste. C'est un langage bien plus récent que ses deux compères, R et Python (début des années 2010 pour Julia vs. début des années 1990 pour R et Python). Il peut donc sembler moins mature sur certains points, mais la plupart des méthodes d'analyse de données classiques ont été implémentées. De même que R et Python, ce langage est gratuit, open-source et disponible sur les principaux systèmes d'exploitation. Julia a l'avantage d'être plus rapide que R et Python. Il inclut aussi un support natif d'appel à des librairies en C ou en Fortran, et un support non-natif d'appel à des librairies en R et Python.

Voici un guide qui présente les bases de Julia : Learn Julia in Y minutes (il n'y a pas de version française pour l'instant) et un guide de bonnes pratiques en Julia (en anglais).

4 SAS

SAS est un langage de programmation propriétaire spécialisé dans l'analyse de données. Bien qu'utilisé dans certaines industries, e.g. pharmaceutique, son utilisation tend à diminuer. De plus, la license à renouveler tous les ans, l'absence de communauté en ligne (pour avoir de l'aide, à part la documentation officielle, il n'y a pas grand chose) et l'impossibilité de voir le code des différentes procédures font que je déconseille l'utilisation de SAS dans le cadre de ce cours. Cependant, si vous le voulez, l'université peut fournir des licenses moyennant paiement.