Science des données III : cours 1

Introduction & logiciels

Philippe Grosjean & Guyliann Engels

Université de Mons, Belgique
Laboratoire d'Écologie numérique des Milieux aquatiques

http://biodatascience-course.sciviews.org
sdd@sciviews.org

UMONS

Que va-t-on faire dans ce cours SDD III ?

UMONS

Introduction

- Au cours de **biostatistique et probabilités de Ba2**, nous avons abordé :
  - Les statistiques descriptives, qui résument à l'aide de descripteurs ou présentent de manière visuelle (graphique) le contenu de jeux de données,
  - Les statistiques inférentielles, basées sur les tests d'hypothèse pour répondre à des questions « binaires » (est-ce ceci –H_0–, ou son contraire –H_1– ?)
- Au cours de **biostatistique de Ba3**, nous avons étudié diverses techniques multivariées : exploratoires (clusters et ordination) et confirmatoires (ANOVA, régression linéaire, modèle linéaire).
- Dans ce cours, nous aborderons diverses techniques statistiques complémentaires : l'analyse de **séries spatio-temporelles**, la **classification supervisée**, et la **régression non linéaire**.

UMONS

Introduction

- Au cours de **biostatistique et probabilités de Ba2**, nous avons abordé :
  - Les statistiques descriptives, qui résument à l'aide de descripteurs ou présentent de manière visuelle (graphique) le contenu de jeux de données,
  - Les statistiques inférentielles, basées sur les tests d'hypothèse pour répondre à des questions « binaires » (est-ce ceci –H_0–, ou son contraire –H_1– ?)
- Au cours de **biostatistique de Ba3**, nous avons étudié diverses techniques multivariées : exploratoires (clusters et ordination) et confirmatoires (ANOVA, régression linéaire, modèle linéaire).
- Dans ce cours, nous aborderons diverses techniques statistiques complémentaires : l'analyse de **séries spatio-temporelles**, la **classification supervisée**, et la **régression non linéaire**.

UMONS

Références

Syllabus : la matière étant plus complexe que pour les 2 cours précédent, un syllabus est disponible sous forme PDF.

Autre références utiles :

- **R for Data Science** : disponible en ligne à http://r4ds.had.co.nz

- **Venables W.N. & B.D. Ripley, 2002**. Modern applied statistics with S-PLUS (4th ed.). Springer, New York, 495 pp.

UMONS

Les outils logiciels que nous utiliserons

UMONS

R, R Studio & SciViews Box 2018

Comme pour les années précédentes, nous utiliserons un **système complètement configuré**, la SciViews Box. La version 2018 avec des logiciels réactualisés est disponible. Elle contient, entre autres :

- **R** 3.4.4, ainsi que plus d'un millier de packages R additionnels préinstallés (snapshot de CRAN en date du 22/04/2018)
- **R Studio server** 1.1.442 préinstallé complètement, y compris les outils additionnels requis pour compiler les fichiers R Markdown. Une grande nouveauté par rapport à la version précédente : des packages facilitant la lecture des données (`data.io`), le "workflow" (`flow`) et les graphiques (`chart`), ainsi qu'un outil de configuration simplifié de la SciViews Box. Snippets de SciViews.
- **LyX** 2.2.3, customisé avec les outils SciViews pour éditer des documents contenant des chunks R.
- **Python 3.5.2 Scientific** préinstallé avec plus de 200 packages supplémentaires. **Jupyter** avec des notebooks Python 3.5.2 et R 3.4.4. **Spyder** 3.1.4, un IDE pour éditer du code et des scripts en Python.
- Une série de logiciels complémentaires tels **R Commander**, **EqualX**, **Meld**, **DB Browser pour SQLite**, …

UMONS

Installation et utilisation

- **Installation facilitée sous Windows** et **MacOS** (et bientôt sous Linux également)
- **Version 64 bit conseillée** (mais nécessite d'activer les options de virtualisation du processeur dans le BIOS)
- **Version 32 bit compatible** avec tous les PC récents et à privilégier pour les machines moins puissantes (Intel Core i3 ou Pentium, par exemple)
- Nécessite tout de même un **PC suffisamment puissant** : Core i5 haut de gamme, équivalent ou mieux, 4Go de RAM au moins (8Go ou plus conseillé), une carte graphique accélérée, un disque dur rapide avec une vingtaine de Go libres (un disque SSD est un must), et un accès administrateur pour l'installation.

Prise en main

Vous allez maintenant configurer et découvrir cette nouvelle machine virtuelle...

UMONS