

SICSS-Montréal : Formation intensive en R

Visseho Adjiwanou, PhD.

19 May 2021

Plan de cours

- Jour 1: Introduction à R
- Introduction générale
- RStudio et RMarkdown
- Tidyverse
- Jour 2: Transformation des données
- Jour 3: Visualisation
- Jour 4: Analyse/modélisation
- Jour 5: Programmation / Travailler avec les données de texte

1. Description

Le paysage de l'analyse quantitative en sciences sociales tel que nous le connaissons est en train de changer drastiquement devant nous avec d'une part, la multitude des données, notamment des données numériques et administratives qui deviennent de plus en plus disponibles, et d'autre part, avec l'intégration de nouvelles méthodes d'analyses issues des sciences computationnelles pour traiter ces données. Ces changements conduisent à une redéfinition de la nature même des recherches que le chercheur en sciences sociales est amené à mener. Mais également de son outil d'analyse. de nouveaux outils jusque-là réservé à quelques initiés commencent à devenir la norme afin de permettre aux chercheurs en sciences sociales de prendre plein avantage de ce nouvel environnement. Ce cours vise à vous introduire à l'un de ces outils, le logiciel R.

2. Objectifs

A la suite de ce cours, vous serez amené capable de : - conduire une analyse descriptive avec R; - conduire la plupart part des modèles de régressions - être capable de présenter une visualisation adéquate des résultats - travailler avec les données de texte - travailler avec plusieurs types de données

3. Matériel

3.1. Logiciels

Pour pouvoir utiliser ce langage, nous allons recourir à R Studio qui est un environnement de développement intégré. Finalement, R Studio permet de sortir les documents en version Word ou pdf. Pour ce faire, vous aurez besoin aussi d'installer sur votre machine latex. Nous n'allons pas apprendre à utiliser Latex, mais une connaissance antérieure est un atout. Cependant, vous pouvez tout simplement utiliser une application tinytex pour convertir vos fichiers en pdf.

Voici la procédure à suivre pour l'installation de R, de RStudio, et de Latex¹.

3.2.1. Installation de R

Suivez les instructions selon votre système d'exploitation: <http://cran.cnr.berkeley.edu/>

3.2.2. Installer RStudio (Desktop)

- Sur le site de RStudio <https://www.rstudio.com/>
- Ou directement à partir de ce lien: <https://www.rstudio.com/products/rstudio/download/>

3.2.3. Installer tinytex

C'est un package à installer. Voici la procédure décrite ici : <https://yihui.org/tinytex/>

3.2.4. Installer Latex (Optionnel)

- Pour les utilisateurs de Windows: installez **la complète version de MiKTeX** : <https://miktex.org/2.9/setup>
- Cliquez sur Net Installer
- Téléchargez la version
- Installez la en double-cliquant sur le fichier téléchargé et suivez les instructions.
- Pour les utilisateurs de Mac, vous pouvez utiliser TeXShop : <https://pages.uoregon.edu/koch/texshop/>
- Vous pouvez aussi utiliser MiKTeX.

¹Tout ceci vous sera précisé à nouveau lors du premier cours, si vous n'avez pas pu les installer

3.3. Références obligatoires

- Wickham, Hadley & Grolemund, Garrett. 2017. R for Data Science: Import, Tidy, Transform, Visualize, and Model data. Boston. O'Reilly. Pp.492. <https://r4ds.had.co.nz/>
- John, Fox & Sanford, Weisberg. 2019. An R Companion to Applied regression, 3rd edition. Sage.

3.4. Ressources en ligne

En français:

- <https://juba.github.io/tidyverse/index.html>: Une très bonne ressource en français
- <http://dimension.usherbrooke.ca/dimension/v2ssrcadre.html>

En anglais:

- <https://clanfear.github.io/CSSS508/>: Vous avez une série de cours et de code ici
- <https://jrnold.github.io/qss-tidy/>: tous les codes du livre de Kosuke Imai sont réécrits ici en version tidyverse.
- <https://r4ds.had.co.nz/>: le livre de Wickham en ligne
- <https://www.bitbybitbook.com/en/1st-ed/preface/>: le livre de Salganik en ligne
- <https://data.hypotheses.org/>

Comme vous vous en rendrez compte assez rapidement, il y a des millions de ressources en ligne. Votre tâche souvent va consister à trouver la bonne ressource pour la tâche à faire. Aussi, google.com vous sera d'un atout formidable.

4. Slack

Slack est un forum de discussion en classe où vous pouvez poser des questions sur le matériel, interagir avec vos camarades. Vous pouvez aussi poser directement vos questions sur Moodle. Sachez que vos questions peuvent aider d'autres personnes aussi à comprendre la matière. De même, d'autres personnes peuvent avoir de meilleures réponses à vos questions.

6. Calendrier

Lundi 7 juin: Introduction à R, RStudio et RMarkdown
Mardi 8 juin: Transformation des données avec dplyr
Mercredi 9 juin: Visualisation avec ggplot
Jeudi 10 juin: Modélisation
Vendredi 11 juin: Programmation avec R

7. Liste de lectures et d'exercices

Jour 1: <https://www.bitbybitbook.com/fr/1st-ed/introduction/> Jour 2: Joshua Blumenstock : Predicting poverty and wealth from mobile phone metadata Jour 3: Pierotti, Rachel. (2013). "Increasing Rejection of Intimate Partner Violence: Evidence of Global Cultural Diffusion." American Sociological Review, 78: 240-265. Jour 4: Shelley, Clark et al. 2019. "The Impact of Childcare on Poor Urban Women's Economic Empowerment in Africa". Demography, 56, 4.

8. Importants packages

1. tidyverse

- dplyr
- ggplot

2. summarytools :

- <https://cran.r-project.org/web/packages/summarytools/vignettes/Introduction.html>

3. rdhs

4. purr

- <https://jennybc.github.io/purrr-tutorial/>

5. Stargazer

6. Stringer