Analyse de données pour une science reproductible

# Préface

2019-09-16

<https://psyteachr.github.io/msc-data-skills/intro.html>

# Informations générales

Ce cours donne un aperçu des compétences nécessaires pour une recherche reproductible et réplicable et pour une approche ouverte et transparente de la science (*open science*) à l'aide du langage de programmation statistique R. Les étudiants apprendront la visualisation, le nettoyage, et la préparation des données, l'itération et les fonctions, les simulations de probabilité et de données, les modèles linéaires généraux et les flux de travail reproductibles. L'apprentissage est renforcé par des exercices hebdomadaires impliquant de travailler avec différents types de données.

## Objectifs du cours

Ce cours vise à enseigner aux étudiants les principes de base de la recherche reproductible et à fournir une formation pratique en traitement et analyse de données dans le langage de programmation statistique R.

## Acquis d'apprentissage visés

À la fin de ce cours, les étudiants seront en mesure de :

* S'appuyer sur un éventail de compétences et de techniques spécialisées pour formuler un plan de recherche adapté à divers types de questions en psychologie et en neurosciences.
* Rédiger des scripts en R pour organiser et transformer des ensembles de données en utilisant les meilleures pratiques acceptées.
* Expliquer les bases de la probabilité et son rôle dans l'inférence statistique.
* Analyser de façon critique les données et présenter des statistiques descriptives et inférentielles de façon reproductible.

## Plan du cours

## Exercices

Des exercices sont disponibles à la fin de la page Web de chaque leçon. Celles-ci ne sont ni notées ni obligatoires, mais si vous pouvez travailler sur chacune d'entre elles (à l'aide de ressources en ligne, bien sûr), vous pourrez facilement compléter les évaluations notées.

Téléchargez tous les exercices et fichiers de données ci-dessous sous forme d'archive ZIP.

01 intro : Introduction à R, fonctions, Rmarkdown

02 données : Vecteurs, données tabulaires, importation de données, *pipes*

Compétences essentielles : Vous devez être en mesure de faire ces exercices pour pouvoir progresser en classe au-delà des deux premiers cours.

03 ggplot : Visualisation des données

04 tidyr : Données en ordre

05 dplyr : Querelles de données

06 jointures : Relations de données

07 fonctions : Fonctions et itérations

08 simulation : Simulation

09 glm : GLM

## *‘Packages’* utilisés dans ce cours

tidyverse

cowsay

goodshirt

ukbabynames

cowplot

plotly

MASS

faux

## Ressources externes

Des ressources diverses, comme des conseils sur l'installation, ou des résultats des démonstrations de codage en direct, seront ajoutées aux annexes tout au long du semestre.

En attendant, voici quelques ressources (en anglais) qui pourront vous être utiles.

### Manuels (gratuits) en ligne

* [Glasgow Psychology RStudio](http://rstudio1.psy.gla.ac.uk/)
* [Learning Statistics with R](https://learningstatisticswithr-bookdown.netlify.com/) par Danielle Navarro
* [R for Data Science](http://r4ds.had.co.nz/) par Garrettt Grolemund et Hadley Wickham

### Cours et tutoriels en ligne

* [swirl](http://swirlstats.com/)
* [R for Reproducible Scientific Analysis](http://swcarpentry.github.io/r-novice-gapminder/)
* [codeschool.com](http://tryr.codeschool.com/)
* [datacamp](https://www.datacamp.com/courses/free-introduction-to-r)
* [Improving your statistical inferences](https://www.coursera.org/learn/statistical-inferences/) sur Coursera

### Antisèches

* Vous pouvez accéder à l’aide de RStudio dans le menu “**help**”
* Vous pouvez également télécharger et imprimer l’antisèche de RStudio : [RStudio Cheat Sheets](https://www.rstudio.com/resources/cheatsheets/)

### Autres

* Un guide stylistique pour le code dans R : <https://style.tidyverse.org>
* Le hashtag [#rstats sur twitter](https://twitter.com/search?q=%2523rstats)