

R, RStudio et workflow

ou comment bien commencer son travail d'analyse

Benjamin Louis

MàJ: 18/11/2021



C'est quoi R ?

- Langage de programmation et environnement intégré pour la science des données
- Développé dans les années 90 par Ross Ihaka et Robert Gentleman
- Depuis 1997 : *R-core team* (internationale)
- Logiciel libre et gratuit (*licence GNU GPL*)
- Langage interprété exécutable pas à pas (ligne de code)
- Version : 3.6.1 (05/07/2019)
- Téléchargement sur le CRAN : <https://cran.r-project.org/>

Forces

- Libre et gratuit
- Multiplateforme
- Manipulation et visualisation de données
- *On vient pour le langage, on reste pour la communauté !*
- > 13000 packages (ensemble de fonction)
- Couplage : SQL, WinBUGS, QGis, Excel, ...

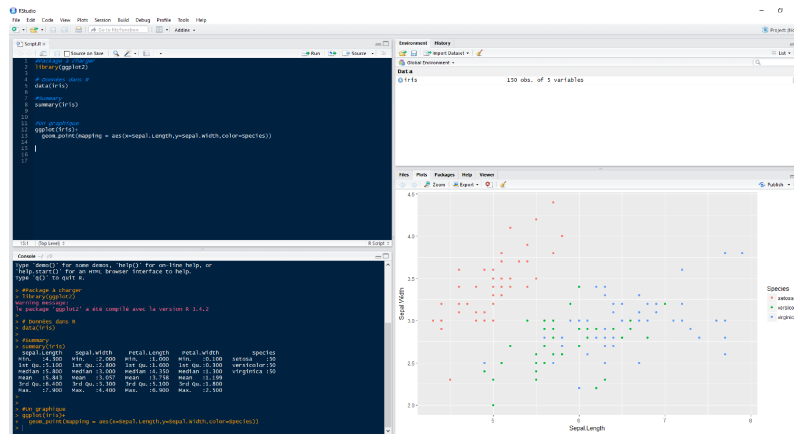
Faiblesses

- Difficile à apprendre
- Gestion de la mémoire (de moins en moins vrai)
- Interface de base minimale



RStudio

- Entreprise fondée par J.J. Allaire en 2009
- Création de produits libres (et gratuits) ainsi que des services
- 2011 : RStudio IDE, interface utilisateur ergonomique, configurable et intégrant de nombreux outils utiles
- Téléchargement : <http://www.rstudio.com>



Session, environnement et assignation

Une session de travail R commence à l'ouverture de R/RStudio et se termine à sa fermeture.

on **assigne** un nom à des objets ou des valeurs à l'aide de `<-`

```
ma_variable1 <- 1
ma_variable2 <- "Hello"
ma_variable3 <- TRUE
ma_variable4 <- 2/3
```

On peut lister les objets créés avec la fonction `ls()`

```
ls()
```

```
## [1] "dir"          "file"          "ma_variable1" "ma_variable2" "ma_variable3" "ma_variable4"
```

Ou en regardant l'onglet *Environment* dans RStudio.

Projet RStudio

Working directory

- R comprend la notion de dossier de travail où il va chercher/sauvegarder des fichiers.

```
getwd()
```

```
## [1] "D:/MEGA/PRO - Statisticien Independant/GIT PROJECTS/FORMATIONS/Workfl"
```

- Chemins d'accès dépendent du système d'exploitation, de l'organisation des dossiers, ...

vs RStudio project

- Construction d'espaces de travail portables et partageables
- Premier pas vers la reproductibilité !

File --> New project

Organiser son projet

```
## D:/MEGA/PRO - Statisticien Independant/GIT PROJECTS/FORMATIONS/stat
## +-- data
## |   \-- processed-data.csv
## +-- data-raw
## |   +-- data-raw-processing.R
## |   \-- data-raw.csv
## +-- docs
## |   \-- protocol.docx
## +-- mon_projet.Rproj
## +-- R
## |   +-- 1-exploratory_analysis.R
## |   \-- 2-statistics_analysis.R
## \-- reports
##     \-- my_reports.Rmd
```

Package {here}

```
library(here)  
here()
```

```
## [1] "D:/[...]/static/mon_projet"
```

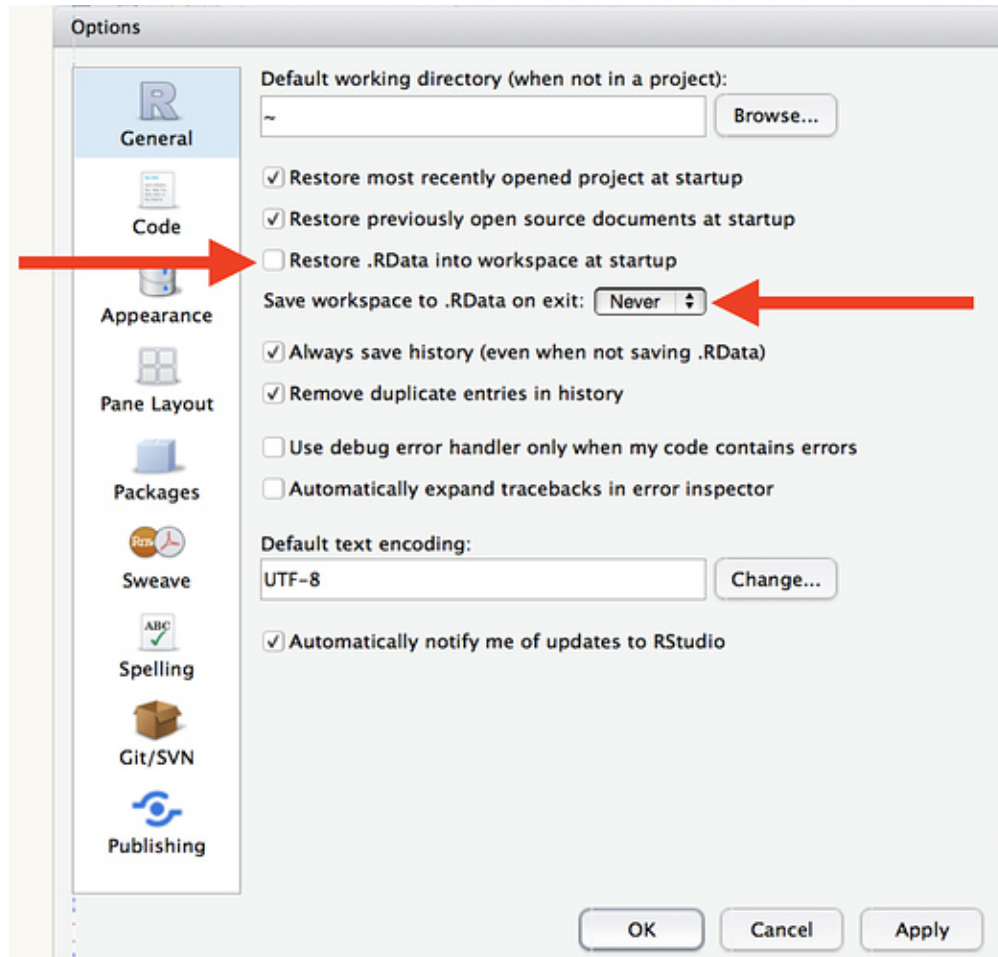
```
here("R")
```

```
## [1] "D:/[...]/static/mon_projet/R"
```

```
here("data-raw", "data-raw-processing.R")
```

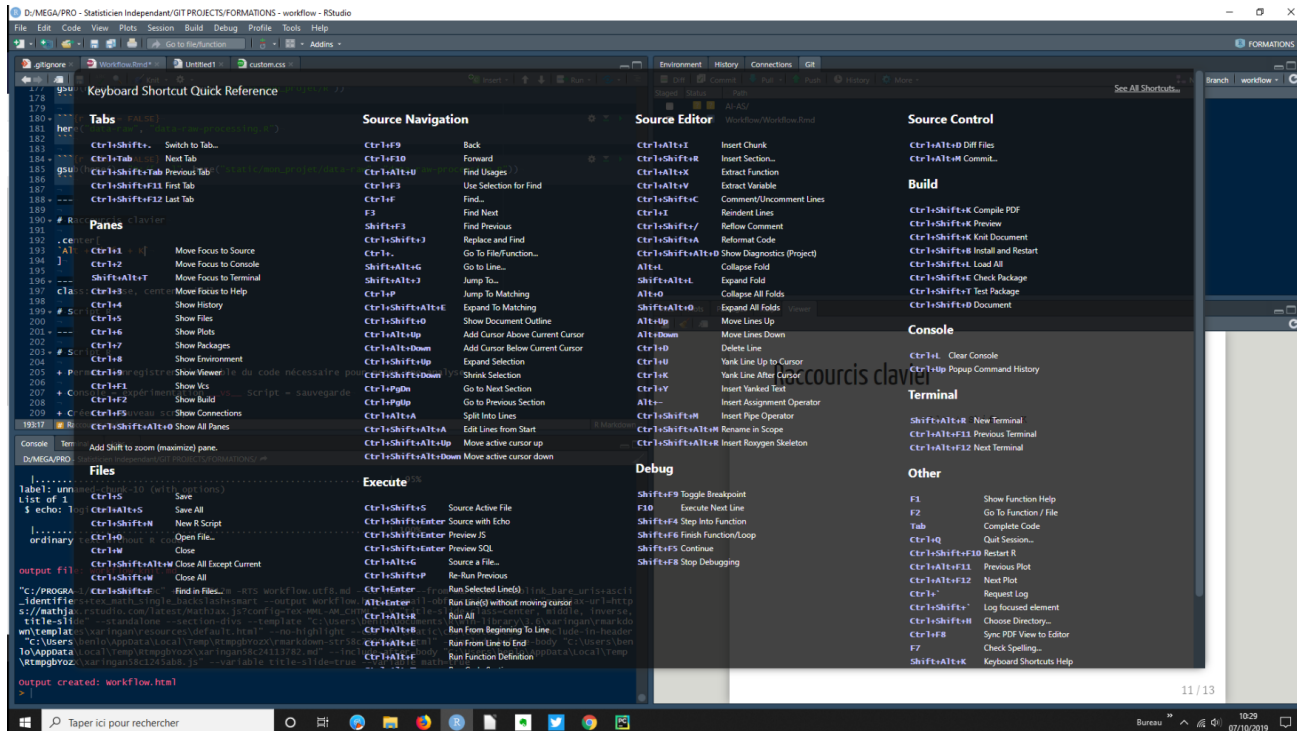
```
## [1] "D:/[...]/static/mon_projet/data-raw/data-raw-processing.R"
```

Conseil !



Raccourcis clavier

Alt + Shift + K



Script R

Script R

- Permet d'enregistrer l'ensemble du code nécessaire pour mener une analyse
- Console = expérimentation **vs** Script = sauvegarde
- Créer un nouveau script :

File --> New file --> R Script ou **Ctrl + Shift + N**

- Lancer la ligne de code active ou l'ensemble des lignes de code sélectionnées : **Ctrl + Enter**
- Lancer l'ensemble du script : **Ctrl + Shift + S**
- Ligne de commentaire commence par **#**

Script R : quelques conseils

- Commencer par une description sous forme de commentaire
- Mettre l'ensemble des packages nécessaires en début de script
- Ne jamais commencer par `rm(list = ls())` ni `setwd()`
(<https://www.tidyverse.org/articles/2017/12/workflow-vs-script/>)
- Utiliser des commentaires
- Attention aux scripts trop longs : scinder l'analyse en plusieurs scripts
(`source()`)
- Utiliser un style cohérent (<http://stat405.had.co.nz/r-style.html>)
- Éviter les copier-coller --> automatiser avec des fonctions !
- Ne pas hésiter à confronter votre code à des collègues !

Exercices

- Ouvrir une session R via RStudio
- Créer un projet RStudio et lui donner un nom
- Créer un premier script
- Donner une description au script (e.g. *"Premier script d'exerice...."*)
- Ne pas oublier de sauvegarder !!!