Quelques outils utiles

Géraldine Derroire – mars 2019

Pour une introduction sur les bonnes pratiques de gestion et d’analyse de données, voir les guide de la British Ecological Society :

* <https://www.britishecologicalsociety.org/wp-content/uploads/2018/12/BES-Reproducible-Code.pdf>
* <https://www.britishecologicalsociety.org/wp-content/uploads/2018/12/BES-Data-Management-GBIF.pdf>

# R et Rstudio

* tutoriels conseillés dans la section pré-requis de ce cours : <https://ericmarcon.github.io/Cours-R-avance/>
* <https://r4ds.had.co.nz/>
* <https://www.rstudio.com/>

# Packages utiles

Deux packages sont intéressants pour manipuler facilement de gros jeux de données. Les deux ont des avantages et des inconvénients. Le mieux est d’en choisir un et de s’y tenir.

* Dplyer : assez intuitifs
  + <https://dplyr.tidyverse.org/>
  + <https://ericmarcon.github.io/Cours-R-avance/Tidyverse.pdf>
  + <https://www.datacamp.com/courses/dplyr-data-manipulation-r-tutorial>
* Data.table : plus rapide
  + <https://github.com/Rdatatable/data.table/wiki>
  + <https://www.datacamp.com/courses/data-table-data-manipulation-r-tutorial>

Pour les graphiques : ggplot2

* <https://ggplot2.tidyverse.org/>
* <https://ericmarcon.github.io/Cours-R-avance/> (section tidyverse)

# RMarkdown

Permets d’écrire des rapports reproductibles (écrire et faire des analyses dans un même document).

* Section « Productivité » du support de cours <https://ericmarcon.github.io/Cours-R-avance/> et les références citées
* <https://m-clark.github.io/Introduction-to-Rmarkdown/>

# Version control

* Section « Productivité » du support de cours <https://ericmarcon.github.io/Cours-R-avance/> et les références citées
* <https://www.hendrikbruns.tk/post/using-rstudio-and-git-version-control/>
* <https://support.rstudio.com/hc/en-us/articles/200532077-Version-Control-with-Git-and-SVN>

Pour travailler à plusieurs, il est utile d’interfacer Git avec GitHub ou Bitbucket. GitHub est le plus couramment utiliser mais les répertoires privés sont payants (ce qui est un problème pour les données non publiques). Bitbucket n’a pas cette contrainte.

* <https://happygitwithr.com/rstudio-git-github.html>
* <https://github.com/>
* <https://bitbucket.org/product>

# Gestion des références bibliographiques

Il est utile d’utiliser un logicel de gestion des références bibliographiques. Si tu n’en as pas, je te conseille Mendeley <https://www.mendeley.com/newsfeed> (voir les tutoriels sur leur site). Mendeley permet d’avoir des librairies partagées.