

Science des données V : module 1



Modularisation

Philippe Grosjean & Guyliann Engels

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



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

Pourquoi modulariser son analyse ?

Les dérives du Notebook

- Le R Notebook (et R Markdown) sont très pratiques pour rassembler les analyses (les instructions R) et la partie narrative du rapport.
- Mais si les analyses nécessitent beaucoup de code, la partie narrative est morcelée
- Les instructions ne sont **pas réutilisables** d'un notebook à l'autre

Solution

Extraire les instructions volumineuses dans des scripts et récupérer à l'aide de la fonction `source()`. Rendre son code réutilisable en écrivant des **fonctions()**, de préférence optimisées.

Ranger avant d'arranger...

- Avant de modulariser vos analyses, pensez à **faire du ménage** !
 - Utilisez-vous un projet (et un dépôt Git) ?
 - Les sous-répertoires et noms de fichiers sont-ils logiques et organisés ?
 - Utilisez-vous des **chemins relatifs** pour que le projet soit portable ?
 - Les chunks dans le notebook sont-ils morcelés de manière logique ?

Sortir le code du Notebook

- Si les chunks restent trop volumineux, ou si les instructions doivent être réutilisées, **il faut les sortir**
- Un script R est tout indiqué pour les accueillir

Exercice

Sur base du document Markdown fourni, extrayez le code volumineux dans un script R.