

Tutoriel {data.table}

Toby Dylan Hocking*

Résumé du tutoriel

{data.table} est un package R qui permet le traitement efficace de grands jeux de données. Il inclut des fonctions pour la lecture et l'écriture de fichiers CSV, l'agrégation, la restructuration, la jointure, etc. Par rapport aux autres logiciels avec ces fonctions, l'utilisation de {data.table} peut être 10x plus efficace en temps de calcul et en utilisation de mémoire. Il est aussi très facile à installer (il ne dépend pas d'autres packages) et à mettre à jour (les nouvelles versions de {data.table} sont compatibles avec les anciennes). Sur CRAN depuis 2006, il est très utilisé, avec plus de 1500 autres packages qui en dépendent.

Depuis 2024, il y a même des traductions en français, pour les messages d'erreur, ainsi que les vignettes.

Ce tutoriel va discuter de :

- l'utilisation du package (fonctions de base et avancées),
- son efficacité relative aux autres packages (tidyverse, pandas, etc),
- la manière d'écrire un package qui dépend de {data.table},
- quelques remarques sur la communauté de développeurs de {data.table} qui est très ouverte aux contributions du public.

Prérequis : maîtrise de base de la programmation en langage R.

Durée : 2h

À réaliser avant de venir au tutoriel :

Installer R version 3.3.0 ou supérieure et ensuite utiliser la commande

```
install.packages("data.table")
```

Informations complémentaires :

- Page web en anglais : <http://r-datatable.com/>
- Vignette d'introduction en français : <https://rdatatable.gitlab.io/data.table/articles/fr/datatable-intro.html>
- Code source du tutoriel : <https://github.com/tdhock/2023-10-LatinR-data.table>

* Professeur Agrégé, Département d'informatique, Université de Sherbrooke, Québec, Canada,
Toby.Dylan.Hocking@USherbrooke.ca