

Les petits tuto rapido : Cas Dplyr

Teddy Leandre

10/12/2020

I) Petite Introduction

`dplyr` est une itération de `plyr` qui fournit des fonctions flexibles basées sur “`ver`” pour manipuler des données dans R. La dernière version de `dplyr` peut être téléchargée depuis CRAN en utilisant la commande ci-dessous.

```
# install `dplyr` and `magrittr` from CRAN "The comprehensive R archive network"
install.packages("dplyr")
install.packages("magrittr")
```

II) Différentes fonctionnalités de `Dplyr`

Cette section fournit une vue d'ensemble de ce qu'est `dplyr` et pourquoi un développeur peut vouloir l'utiliser. Dans ces exemples on utilisera le dataset `iris` => voir `help("iris")` pour plus d'information. Pour démontrer l'utilisation de `dplyr` on utilisera les fonctions `filter` et `mutate` ainsi que le paquet `magrittr` pour utiliser la fonction pipe `%>%`.

Il est toujours important de charger la bibliothèque pour obtenir les fonctionnalités du paquet.

```
# Load dplyr
library(dplyr)
```

```
##
## Attaching package: 'dplyr'
```

```
## The following objects are masked from 'package:stats':
##
##     filter, lag
```

```
## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

```
# Load magrittr (complementary package to use with dplyr)
library(magrittr)
```

- Pour filtrer des données statistiques en fonction de certains critères, nous utilisons un `filter`.

```
p <- iris
# On filtre les espèces qui correspondent à l'espèce "versicolor" et qui sont supérieur à 3 cm en
# longueur et en largeur
Species_versicolor <- p %>% filter(Species == "versicolor", Sepal.Length > 3, Sepal.Width > 3)
```

- Pour ajouter de nouvelles variables ou modifier des variables existantes, nous utilisons `mutate`.

```
# On ajoute une nouvelle variable nommer `Sepal.dimension` qui mesure l'aire des sépales d'une fleur pour l'espèce "versicolor"
New_species_versicolor <- Species_versicolor %>%
  select(Sepal.Length , Sepal.Width, Species) %>%
  mutate(Sepal.dimension = Sepal.Length * Sepal.Width )
New_species_versicolor
```

##	Sepal.Length	Sepal.Width	Species	Sepal.dimension
## 1	7.0	3.2	versicolor	22.40
## 2	6.4	3.2	versicolor	20.48
## 3	6.9	3.1	versicolor	21.39
## 4	6.3	3.3	versicolor	20.79
## 5	6.7	3.1	versicolor	20.77
## 6	5.9	3.2	versicolor	18.88
## 7	6.0	3.4	versicolor	20.40
## 8	6.7	3.1	versicolor	20.77

“La science, aujourd’hui, cherchera une source d’inspiration au-dessus d’elle ou périra.” **Simone Weil**

C'est pourquoi dans le cadre de mon mini tutoriel je tiens à citer mes sources :  [Ma source](#)



[Mon Github](#)

Revenir à [L'introduction](#)

Note that the `echo = FALSE` parameter was added to the code chunk to prevent printing of the R code that generated the plot.