

Le travail de Wenjun Zhao est disponible ici

BENOUCIEF Amine

22/12/2020

```
{r setup, include=FALSE} knitr::opts_chunk$set(echo = TRUE)
```

SYNTHESE DU TRAVAIL EN QUESTION

dplyr est une extension facilitant le traitement et la manipulation de données contenues dans une ou plusieurs tables. Elle propose une syntaxe claire et cohérente, sous formes de verbes, pour la plupart des opérations de ce type. Wenjun a su bien expliquer les principales fonctions. Cependant des explications a cote des lignes de code aurait ete plus facile pour le lecteur afin de faciliter la comprehension. De meme, pas de bibliographie nous ne savons pas si elle s'est inspiree d'un travail ou elle a cree elle meme ses exemples (si oui il fallait le preciser).

Overview

dplyr est une grammaire de manipulation de données, fournissant un ensemble cohérent de verbes qui vous aident à résoudre les défis de manipulation de données les plus courants:

mutate () ajoute de nouvelles variables qui sont des fonctions de variables existantes

select () sélectionne les variables en fonction de leurs noms.

filter () sélectionne les observations en fonction de leurs valeurs.

summary () réduit plusieurs valeurs à un seul résumé.

arrange () modifie l'ordre des lignes.

Tout cela se combine naturellement avec *group_by ()* qui vous permet d'effectuer n'importe quelle opération «par groupe». Vous pouvez en savoir plus sur eux dans *vignette ("dplyr")*. En plus de ces verbes à table unique, dplyr fournit également une variété de verbes à deux tables, que vous pouvez découvrir en *vignette ("two-table")*.

dplyr est conçu pour résumer la manière dont les données sont stockées. Cela signifie qu'en plus de travailler avec des trames de données locales, vous pouvez également travailler avec des tables de base de données distantes, en utilisant exactement le même code R. Installez le package dbplyr puis lisez *vignette ("databases", package = "dbplyr")*.

Installation

Le moyen le plus simple d'obtenir dplyr est d'installer tout le tidyverse:

```
install.packages ("tidyverse")
```

Alternativement, installez simplement dplyr:

```
install.packages ("dplyr")
```

Ou la version de développement de GitHub:

```
install.packages ("devtools")  
devtools :: install_github ("tidyverse / dplyr")
```

Utilisation

```
library(dplyr)  
starwars %>%  
  filter(species == "Droid")
```

```
library(dplyr)  
starwars %>%  
  select(name, ends_with("color"))
```

```
library(dplyr)  
starwars %>%  
  mutate(name, bmi = mass / ((height / 100) ^ 2)) %>%  
  select(name:mass, bmi)
```

```
library(dplyr)  
starwars %>%  
  arrange(desc(mass))
```

```
library(dplyr)  
starwars %>%  
  group_by(species) %>%  
  summarise(  
    n = n(),  
    mass = mean(mass, na.rm = TRUE)  
  ) %>%  
  filter(n > 1)
```

EVALUATION DU TRAVAIL EN QUESTION

Critère 1 : Visuel sur pdf 4/4 Agreable a lire.

Critère 2 : Originalite du code 4/4 Wenjun a bien montre comment utiliser les fonctions importantes.

Critère 3 : Fonctionnalité du code 4/4 tout fonctionne.

Critère 4 : Lisibilité du code 3/4 Lisible mais peut perturber le lecteur.

Critère 5 : Explications données 3/4 Explications claire sur le code, cependant il aurait du etre place a cote des lignes de code au lieu du debut.

CONCLUSION

Globalement bon travail qui explique bien le package dplyr. Les commentaires aurait pu etre mieux placer.