Le travail de JINGWEN SU **est disponible ici**

Théo Marié

20/12/2020

SYNTHESE DU TRAVAIL EN QUESTION

DPLYR est un package interessant par ses fonctionnalité de traitement et de modifications de bases de donnés, c'est indispensable dans le mode professionel du data analyst.Le travail que Jingwen présente est remarquable, car il est très pédagogue. Elle nous explique bien tout ce qu'elle fait et à quoi ça sert. J'ai pris du plaisir à lire son travail, et cela m'a aidé à mieux comprendre le package dplyr.

EXTRAIT COMMENTE DES PARTIES DU CODE

En général, le jeu de données original que nous analysons contient beaucoup de variables (colonnes). Le premier problème que nous devons résoudre est de restreindre la portée pour trouver les données (variables) dont nous avons besoin. select () nous permet de sous-ensemble rapidement l'ensemble de données par nom de variable.

```
library(dplyr)
library(MASS)#birthwt

select=dplyr::select

# select(birthwt,age:smoke)
head(select(birthwt,age:smoke))
```

```
##    age lwt race smoke
## 85  19 182    2     0
## 86  33 155    3     0
## 87  20 105    1     1
## 88  21 108    1     1
## 89  18 107    1     1
## 91  21 124    3     0
```

```
# select(birthwt,-(age:smoke))
head(select(birthwt,-(age:smoke)))
```

```
##    low ptl ht ui ftv  bwt
## 85   0   0  0  1   0 2523
## 86   0   0  0  0   3 2551
## 87   0   0  0  0   1 2557
## 88   0   0  0  1   2 2594
## 89   0   0  0  1   0 2600
## 91   0   0  0  0   0 2622
```

Ici, on étudie la fonction SELECT.La première fonction selectionne dans le tableau birthwt, les colonnes de age à smoke. Ensuite, elle réitère l'opération en mettant un "-", ce qui va prendre toutes les colonnes du tableau birthwt sauf celles comprises entre age et smoke.

```
summarize(birthwt,delay=mean(age,na.rm = TRUE))
```

```
##     delay
## 1 23.2381
```

On étudie ici la fonction SUMMARIZE, qui réduit un bloc de données à un tableau d'une seule ligne. Et effectiement, après avoir summarize le tableau birthwt, en voulant uniquement la colonne delay, et son minimum, on obtient le résultat attendu.

EVALUATION DU TRAVAIL EN QUESTION

Critère 1 : Visuellement apréciable sur pdf 3/4

Critère 2 : idées pour faire le code 3/4

Critère 3 : Fonctionnalité du code 4/4

Critère 4 : lisibilité du code 4/4

Critère 5 : explications données 4/4

CONCULSION

Un très bon travail, ou Jingwen nous explique des fonctionnalités simple mais efficaces, avec une notion de pédagogie elle aussi très efficace.