

# Analyse de DPLYR par WENJUN ZHAO

Wenjun ZHAO

12/20/2020

## I. Introduction

Vous pouvez retrouver mon travail sur mon Github, <https://github.com/claudia0524/PSBX>.

Pour ce dossier, nous avons étudié mon travail, étudiant en M2DM au sein de PSB.

Nous avons trouvé son dossier sur son Github : <https://github.com/claudia0524/PSBX>

Maintenant, nous allons établir mes 5 critères d'évaluations, qui seront les mêmes pour tout les dossiers que j'ai vus étudier pour ce devoir:

### Les cinq critères d'évaluation

1. *Rmd se comporte bien à l'exécution*
2. *Les aspects intéressants, didactique, complet*
3. *La qualité Rmarkdown, la qualité de l'écriture*
4. *Didactique, conformité aux exigences vues plus haut et comporte du calcul symbolique et du calcul numérique*
5. *La qualité du LaTeX et des illustrations, la qualité de l'écriture, le choix des ressources internet, la compréhension personnelle des concepts*

## II. Synthèse du travail en question

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.

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")*.

Nous allons maintenant étudier mes codes dans le chapitre suivant.

## III. Extrait commenté des parties de code

### Installation

Le moyen le plus simple d'obtenir dplyr est d'installer tout le tidyverse(ZHAO, n.d.):

```
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)
```

```
##
## 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
```

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

```
## # A tibble: 6 x 14
##   name height mass hair_color skin_color eye_color birth_year sex gender
##   <chr> <int> <dbl> <chr>      <chr>      <chr>      <dbl> <chr> <chr>
## 1 C-3PO 167 75 <NA>      gold        yellow        112 none masculin
## 2 R2-D2 96 32 <NA>      white, bl~ red          33 none masculin
## 3 R5-D4 97 32 <NA>      white, red red          NA none masculin
## 4 IG-88 200 140 none metal        red          15 none masculin
## 5 R4-P~ 96 NA none silver, r~ red, blue    NA none féminin
## 6 BB8 NA NA none none          black        NA none masculin
## # ... with 5 more variables: homeworld <chr>, species <chr>, films <list>,
## #   vehicles <list>, starships <list>
```

```
library(dplyr)
```

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

```
## # A tibble: 87 x 4
##   name hair_color skin_color eye_color
##   <chr> <chr>      <chr>      <chr>
## 1 Luke Skywalker blond fair blue
## 2 C-3PO <NA> gold yellow
## 3 R2-D2 <NA> white, blue red
## 4 Darth Vader none white yellow
## 5 Leia Organa brown light brown
## 6 Owen Lars brown, grey light blue
## 7 Beru Whitesun lars brown light blue
## 8 R5-D4 <NA> white, red red
```

```
## 9 Biggs Darklighter black light brown
## 10 Obi-Wan Kenobi auburn, white fair blue-gray
## # ... with 77 more rows
```

```
library(dplyr)
```

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

```
## # A tibble: 87 x 4
##   name      height mass  bmi
##   <chr>      <int> <dbl> <dbl>
## 1 Luke Skywalker    172    77  26.0
## 2 C-3PO             167    75  26.9
## 3 R2-D2              96    32  34.7
## 4 Darth Vader       202   136  33.3
## 5 Leia Organa       150    49  21.8
## 6 Owen Lars         178   120  37.9
## 7 Beru Whitesun lars 165    75  27.5
## 8 R5-D4              97    32  34.0
## 9 Biggs Darklighter 183    84  25.1
## 10 Obi-Wan Kenobi   182    77  23.2
## # ... with 77 more rows
```

```
library(dplyr)
```

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

```
## # A tibble: 87 x 14
##   name height mass hair_color skin_color eye_color birth_year sex gender
##   <chr> <int> <dbl> <chr>      <chr>      <chr>      <dbl> <chr> <chr>
## 1 Jabb~    175  1358 <NA>      green-tan~ orange        600 herm~ mascu~
## 2 Grie~    216   159 none      brown, wh~ green, y~      NA male mascu~
## 3 IG-88    200   140 none      metal      red          15 none mascu~
## 4 Dart~    202   136 none      white      yellow       41.9 male mascu~
## 5 Tarf~    234   136 brown      brown      blue         NA male mascu~
## 6 Owen~    178   120 brown, gr~ light      blue         52 male mascu~
## 7 Bossk    190   113 none      green      red          53 male mascu~
## 8 Chew~    228   112 brown      unknown    blue        200 male mascu~
## 9 Jek ~    180   110 brown      fair       blue         NA male mascu~
## 10 Dext~    198   102 none      brown      yellow       NA male mascu~
## # ... with 77 more rows, and 5 more variables: homeworld <chr>, species <chr>,
## #   films <list>, vehicles <list>, starships <list>
```

```
library(dplyr)
```

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

```
## `summarise()` ungrouping output (override with `.groups` argument)

## # A tibble: 9 x 3
##   species     n mass
##   <chr>   <int> <dbl>
## 1 Droid         6  69.8
## 2 Gungan        3   74
## 3 Human       35  82.8
## 4 Kaminoan      2   88
## 5 Mirialan      2  53.1
## 6 Twi'lek       2   55
## 7 Wookiee       2  124
## 8 Zabrak        2   80
## 9 <NA>          4   48
```

## IV. Evaluation du travail en question

1. *Rmd se comporte bien à l'exécution (2/4)*
2. *Les aspects intéressant, didactique, complet (3/4)*
3. *La qualité Rmarkdown, la qualité de l'écriture (3/4)*
4. *Didactique, conformité aux exigences vues plus haut et comporte du calcul symbolique et du calcul numérique (3/4)*
5. *La qualité du LaTeX et des illustrations, la qualité de l'écriture, le choix des ressource internet, la compréhension personnelle des concepts (3/4)*

## V. Conclusion

En général, ce travail exécuse bien dans l'environement de R. Et il nous explique clairement comment fonctionner le package Dplyr dans un rmd, On comprend rapidement le but principal de ce package. L'aspect est intéressant, complet et propre. D'ailleurs, il a expliqué chaque code, on peut les comprendre facilement. Il a présenté des fonctions qui peut être utile pour analyser des données dans les rapports. La qualité Rmarkdown, la qualité de l'écriture et La qualité du LaTeX sont ainsi bien.

## VI. Bibliographie

ZHAO, Wenjun. n.d. "Travail Dplyr." <https://github.com/claudia0524/PSBX>.