

# Gaspard Palay TutoDataTable Mazzucato Jupiter

Gaspard PALAY

1/30/2021

## 1. Critères d'évaluation

1. Comportement du Rmd lors de son exécution
2. Qualité rédactionnelle du dossier
3. Didactisme et pertinence du dossier
4. Qualité et lisibilité du Rmarkdown
5. Qualité des applications permettant d'illustrer le package

## 2. Lien vers le document commenté

En cliquant **ici**, vous trouverez le lien menant au GitHub de Thomas Massé qui héberge les travail réalisé.

## 3. Auteurs du document commenté

Le document évalué dans le cadre de ce rendu a été produit par Claire MAZZUCATO et Adrien JUPITER, étudiants en MSc Data Management à Paris School of Business.

## 4. Synthèse du document

Le document illustre l'utilisation du package data.tanle et s'intéresse à son utilisation, à sa popularité, et à l'alternative de la data frame traditionnelle.

Les auteurs expliquent aux lecteurs que ce package est très populaire du à sa capacité de gérer une très grande quantité de données plus rapidement que le ferait les packages traditionnels.

Les auteurs composent un travail didactique se décomposant de la manière suivante : - Installation du package - Usage du package - Conversion des datasets - Fonction de filtre du package - Sélection / renommage / regroupement / jonction / des tables et colonnes

Les auteurs utilisent le dataset mtcars pour illustrer leurs propos.

## 5. Extrait commenté des parties de code

```
# dataframe syntaxe
mtcars[mtcars$cyl == 6 & mtcars$gear == 4, ]
# datatable syntaxe
mtcars_dt[cyl==6 & gear==4, ]
```

Cette partie du code filtre les colonnes cyl et gear portant les valeurs 6 et 4

```
setnames(mtcars_dt, 'vs', 'engine_type')  
colnames(mtcars_dt)
```

Cette partie du code renomme grace à la fonction intégrée setnames de data.frame la colonne par 'engine\_types'

```
dcast.data.table(mtcars_dt, cyl ~ carb, fun.aggregate = mean, value.var = 'mpg')
```

Cette autre partie crée un tableau croisé dynamique montrant le kilométrage moyen mpg des cylindres par rapport au carburateur Carb.

## 6. Evaluation du travail suivant les 5 critères

### 1. Comportement du Rmd lors de son exécution

L'exécution du code Rmd se réalise parfaitement et sans conditions. Le code RMD appelle bien les librairie nécessaire et les installe aussi si l'utilisateur ne les possède pas au préalable.

### 2. Qualité de la rédaction du dossier

La rédaction de ce document est de très bonne qualité. Il ne comporte aucune faute d'orthographe et synthaxique.

Le document est très bien rédigé.

### 3. Didactisme et pertinence du dossier

Le dossier se retrouve facilement sur son github qui est d'ailleurs très bien structuré. Le fichier ne comporte pas de table des matières.

Les descriptions sont bien explicitées et bien illustrées grace au dataset mtcars.

L'auteur arrive à faire adhérer le lecteur à sa production. Il transmet une connaissance nécessaire à la reproduction de son travail. Le document est pertinent pusique chacun des élèves de la promotion manipule au quotidien des dataframe.

Le document est didactique car il se veut sous la forme d'un tutoriel.

### 4. Qualité et lisibilité du RMarkdown

Le RMarkdown est bien écrit. Les tableaux sont bien structurés très lisibles. Il est lisible et les chunk sont très bien paramétrés.

### 5. Qualité des applications permettant d'illustrer le package

Les auteurs utilisent un dataset populaire pour appuyer leur démonstration. Ils illustrent toutes les fonctions avec leur utilisation sur le dataset précité.

## 7. Conclusion

Il s'agit d'un bon travail selon moi. L'auteur est arrivé au bout de son travail et synthétisant les principales fonction de `data.table` en le comparant au package pré intégré à R : `data.frame`. Les termes qu'il adopte sont très bien choisis et illustrent avec pertinence ses propos. Le document est enregistré en local sur mon ordinateur et me servira très probablement à la réalisation de travaux personnels.

Vous retrouvez ce document sur mon **GitHub**.