

Les petits tuto rapido: Cas Ggplot2 de Teddy LEANDRE

Tarik HAKAM

22/12/2020

1. Critères d'évaluation

1. Comportement du Rmd à l'exécution
2. Qualité de la rédaction du dossier
3. Accessibilité, 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 Teddy LEANDRE hébergeant le fruit de sa réalisation.

3. Auteurs du document commenté

Le document évalué dans le cadre de ce rendu a été produit par Teddy LEANDRE, étudiant en Master 2 Data Management à Paris School of Business.

4. Synthèse du document

Il s'agit d'un document illustrant l'utilisation du package **ggplot2** à travers diverses applications et à l'aide du dataset **iris**.

Ggplot2 est un package R conçu spécialement pour la data visualization de données et permettant une meilleure analyse exploratoire des données. Il fournit des tracés élégants et esthétiques créés de manière itérative et édités ultérieurement.

A travers ce document, l'auteur illustre son propos à l'aide d'un seul type de plot, en y ajoutant des modifications telles que :

- l'ajout de marqueurs :
 - ajout de texte
 - repérage de texte
 - mise en évidence d'une zone particulière
 - ajout d'un segment
- changement de titre

5. Extrait commenté des parties de code

Tout d'abord, l'auteur commence par installer et charger le package `ggplot2`, ainsi que le dataset `iris` afin de s'essayer à personnaliser les axes et le titre d'un plot.

Ici, il décide d'utiliser la fonction `plot` sur le data set `iris3`.

Puis, il décide d'affecter le dataset `iris` à une fonction `p` en choisissant de définir les valeurs attribuer aux axes `x` et `y` et au titre ("Type d'espèces").

`color=Species`, permet d'attribuer des couleurs en fonction des espèces.

```
plot(iris3)

p <- ggplot(iris, aes(x = Sepal.Length, y = Petal.Length, title("Type d'espèces"),
                     colour=Species)) + geom_point()
```

Ensuite, l'auteur souhaite tracer des marqueurs à des coordonnées définies sur les axes `x` et `y`.

Ici, il s'agit :

- D'annoter le plot `p` en y intégrant une chaîne de caractères, labellisée `"text"` aux coordonnées $x = 6$, $y = 5$

```
p + annotate("text", x = 6, y = 5, label = "text")
```

- D'annoter le plot `p` en y intégrant plusieurs répétitions d'une même chaîne de caractères, labellisée `"text"` à plusieurs coordonnées à la fois, telles que :

$\Rightarrow x = 4, y = 5$

$\Rightarrow x = 5, y = 6$

$\Rightarrow x = 6, y = 7$

```
p + annotate("text", x = 4:6, y = 5:7, label = "text")
```

- De mettre en évidence une zone particulière de forme quadrilatérale (ici rectangulaire : "rect") sur plot `p` de coordonnées : $x_{min} = 5$, $x_{max} = 7$, $y_{min} = 4$ et $y_{max} = 6$, telle que :

```
p + annotate("rect", xmin = 5, xmax = 7, ymin = 4, ymax = 6, alpha = .5)
```

Avec une transparence de la figure intégrée régit par la fonction `alpha = .5`.

- D'ajouter un segment sur le graphique à points (ou nuage de points) aux coordonnées suivantes : $x = 5$, $xend = 7$, $y = 4$ et $yend = 5$, de couleur "black".

```
p + annotate("segment", x = 5, xend = 7, y = 4, yend = 5, colour = "black")
```

Pour finir, l'auteur choisit de modifier le titre du graphique ainsi que les labels (ou noms) des 2 axes x et y tels que :

- Titre : Graphique sur les différentes espèces de pétale et de sépale
- Label x : Longueur des sépales
- Label y : Longueur des pétales

```
p + ggtitle("Graphique sur les différentes espèces de pétale et de sépale") +  
  xlab("Longueur des sépales") + ylab("Longueur des pétales")
```

6. Evaluation du travail suivant les 5 critères précités

1. Comportement du Rmd à l'exécution

L'exécution du code Rmd se réalise parfaitement à condition que l'on retire lzs lignes de code 93 à 96 qui renvoient à des fichiers .png et .ico non disponible.

2. Qualité de la rédaction du dossier

La rédaction de ce document est de bonne qualité. Certes redondante, mais aboutie en devinant les intentions de l'auteur.

Elle suit cette logique de se vouloir être un tutoriel virgularisant les fonctions du package ggplot2 de manière efficace.

3. Accessibilité, didactisme et pertinence du dossier

La lecture de ce dossier est aisée et accessible.

Les descriptions sont bien explicitées et bien illustrées.

L'auteur arrive à faire adhérer le lecteur à sa production. Le document a vocation à transmettre une connaissance et le but, pour ma part, est atteint.

Le sujet est pertinent et à vocation à servir de mémo pour de futures réalisations.

4. Qualité et lisibilité du RMarkdown

Le RMarkdown est bien écrit.

Il est lisible et aéré.

Il s'agit d'un code bien réalisé.

5. Qualité des applications permettant d'illustrer le package

Les applications sont simples, de qualités et maîtrisées.

7. Conclusion

Selon moi, il s'agit globalement d'un bon travail.

L'auteur fournit un dossier recherché, documenté et accessible.

On y apprend des choses nouvelles.

Ce document apporte une bonne approche pour découvrir ce package.

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