

Gaspard_Palay_ggplot2_ElGhaldy_LIU

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 Soukaine ElGhady qui héberge les travail réalisé.

3. Auteurs du document commenté

Le document évalué dans le cadre de ce rendu a été produit par Soukaine ElGhady et Jiayue Liu, tous deux étudiants en MSc Data Management à Paris School of Business.

4. Synthèse du document

Le document illustre l'utilisation du package **ggplot2** et s'intéresse plus particulièrement à la création d'un nuage de point et de ses fonctionnalités. Il explique les différentes manières de paramétrer cette fonction de ggplot2.

Ggplot2 est un système de création déclarative de graphiques, basé sur La Grammaire du Graphique. L'utilisateur fournit les données, dit à ggplot2 comment associer les variables à l'esthétique, quelles primitives graphiques utiliser, et il s'occupe des détails.

A travers ce document, l'auteur commence tout d'abord par expliquer les fonctions de base de ggplot et du nuage de point. Il explique comment paramétrer la fonction à savoir :

- Nuage de points simples
- Annoter les points
- Ajouter des lignes de régression
- Changer l'apparence des points et des traits
- Nuage de points avec plusieurs groupes
- Changer la couleur, le type, la taille des points automatiquement
- Ajouter des droites de régression
- Changer la couleur, le type, la taille des points manuellement

L'auteur montre comment manipuler la fonction de manière plus élaborée à savoir :

- Ajouter la densité marginale
- Nuage de points avec estimation de la densité
- Nuage de points avec ellipse
- Nuage de points avec distribution marginale

- Nuages de points personnalisés

Le document est bien structuré et une table des matières est présente en début de document ce qui facilite sa lecture

5. Extrait commenté des parties de code

Tout d'abord, l'auteur commence par installer et charger le package `ggplot2`, et prépare les données sur lesquelles il va travailler. Il utilise le jeu de données `mtcars`.

```
mtcars$cyl <- as.factor(mtcars$cyl)
head(mtcars)
```

Ici, il convertit la colonne `cyl` en variable de type facteur

Il présente ensuite comment effectuer des nuages de point simples en paramétrant les couleurs, la taille, la forme grâce à la fonction `geom_point(size, color, shape)`.

Le paramètre `size` permet de changer la taille du plot, `shape` sa forme, `aes(size=qsec)` la taille des points.

La fonction `geom_text()` permet d'annoter les points ... etc.

L'auteur présente ensuite comment ajouter des lignes de régression grâce à la fonction en fournissant au lecteur un format simplifié, à savoir `geom_smooth(method="auto", se=TRUE, fullrange=FALSE, level=0.95)` tout en expliquant succinctement les paramètres de la fonction intégrée.

La partie d'après démontre comment changer l'apparence des points et des traits. En tant que lecteur, je comprends les choses suivantes :

Changer la couleur et la forme des points Changer le type de trait et la couleur

```
ggplot(mtcars, aes(x=wt, y=mpg)) +
  geom_point(shape=18, color="blue")+
  geom_smooth(method=lm, se=FALSE, linetype="dashed",
              color="darkred")
```

Changer la couleur de remplissage de l'intervalle de confiance

```
ggplot(mtcars, aes(x=wt, y=mpg)) +
  geom_point(shape=18, color="blue")+
  geom_smooth(method=lm, linetype="dashed",
              color="darkred", fill="blue")
```

Le document présente ensuite succinctement comment : 1. Créer un nuage de points à plusieurs groupes 2. Changer la couleur, le type, la taille des points automatiquement en fonction des niveaux de `cyl` 3. Ajouter d'autres droites de régression 4. Changer la couleur, le type, la taille des points manuellement 5. Ajouter la densité marginale 6. Nuage de point avec estimation de la densité 2d 7. Nuage de points avec ellipse 8. Nuage de point avec distribution marginale

Enfin, le document illustre la manière de créer un nuage de point totalement personnalisé en paramétrant manuellement les couleurs, en supprimant l'intervalle de confiance ... etc.

6. Evaluation du travail suivant les 5 critères

1. Comportement du Rmd lors de son exécution

L'exécution du code Rmd se réalise parfaitement à condition que l'on retire les dernières lignes de code qui renvoient à des fichiers `.png` (cheatsheet). Il faut aussi préalablement installer en local tous les packages nécessaires.

2. Qualité de la rédaction du dossier

La rédaction de ce document est de bonne qualité. Il comporte quelques oublie d'espacement entre des mots, et très peu de fautes d'orthographe.

Le document est très bien rédigé même s'il comporte très peu d'explications écrites puisque celui ci est majoritairement composé de chunk et d'annotations via des `#commentaires`.

3. Didactisme et pertinence du dossier

Le dossier se retrouve facilement sur les github respectifs des deux auteurs. Leurs github sont d'ailleurs très bien structurés. Le fichier comporte une table des matières ce qui est très agréable en préambule de la lecture d'un document. Le lecteur sait tout de suite à quoi s'attendre sans avoir besoin de parcourir l'entiereté du document au préalable.

Les descriptions sont bien explicitées et bien illustrées.

L'auteur arrive à faire adhérer le lecteur à sa production. L'objectif de se document, très bien explicité au début du deocument, n'est pas de passer en revu toutes les fonctionnalités de ggplot2 mais de s'intereser uniquement au parametrage des nuages de points. A mon humble avis, il fait le tour de l'ensemble des fonctionnalités et est très pédagogique pour un novice.

Les parametrages couleurs se retrouvent dans plusieurs sections, elles auaient pu peut-être être regroupés dans une seul et unique.

Le document est enregistré en local sur mon ordinateur et me servira très prochainement à la réalisation de plusieurs travaux personnels tant il m'a plu.

4. Qualité et lisibilité du RMarkdown

Le RMarkdown est bien écrit. Les plots sont bien organisés et les sauts de ligne permettent d'aérer le document.

Il est lisible et les chunk sont très bien paramétrés.

5. Qualité des applications permettant d'illustrer le package

L'auteur s'est appuyé sur le dataset mtcars et il a bien fait car celui ci est déjà nettoyé. Les différents plots sont bien paramétrés et très impressionnants.

7. Conclusion

Il s'agit globalement d'un très bon travail selon moi. Les deux auteurs sont arrivés à trouver une entente pour leur production et cela se ressent à la lecture. On ne distingue pas les parties de chacuns.

On y apprend beaucoup de choses.

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