

Atelier 3 Instructions: Rapports Rmd

2024-01-31

Introduction

Bienvenue !

Dans l'exercice de cette semaine, vous allez appliquer vos connaissances en R acquises lors des deux leçons de préparation de cette semaine :

Consultez les notes de leçon suivantes pendant l'atelier :

- Structures de Données
- R Markdown

La date limite pour rendre l'exercice est le vendredi 2 février à 23h59 UTC.

Téléchargement et ouverture de l'exercice

1. **Téléchargez le dossier de l'atelier** en cliquant sur le bouton “Télécharger le code de l'atelier” sur la page de l'atelier du sujet ATELIER 3 de la page du cours.
2. **Décompressez le dossier week_03_workshop.zip**. Si vous êtes sur PC, allez dans votre dossier Téléchargements, faites un clic droit sur le fichier zip téléchargé et cliquez sur “**Extraire tout**”.
3. **IMPORTANT** : cliquez sur le fichier projet RStudio appelé “week_03_workshop.Rproj” dans le dossier décompressé pour ouvrir le projet dans RStudio.

Votre tâche consistera à créer un court rapport basé sur R Markdown comparant la distribution de **DEUX VARIABLES CATÉGORIELLES** dans le jeu de données *India TB pathways*. Vous pourriez comparer, par exemple, le niveau d'éducation et l'habitude de fumer, ou l'emploi et le lieu de visite, ou la consommation d'alcool et le diabète - n'importe quelles deux variables de votre choix.

Le rapport doit contenir ces trois choses :

- a. Un graphique créé avec `{ggplot2}`/`{esquisse}`
- b. Un tableau créé avec `{flextable}` ou un autre package de tableau dans R.
- c. Au moins une utilisation de *code R en ligne* dans le Rmd.

Étapes pour compléter l'exercice

- 1) **Créez un nouveau fichier Rmd** (pas un script R !) dans le dossier “rmd” et donnez-lui un nom approprié (par exemple, `tb_rapport_VOTRE_NOM.Rmd`).
- 2) **Copiez l'en-tête YAML** du rapport d'exemple dans votre nouveau fichier.
- 3) **Créez un nouveau chunk de code** pour le chargement des packages et **chargez les packages nécessaires**. (Reportez-vous au rapport d'exemple pour des indices sur les packages que vous pourriez avoir besoin.) Assurez-vous d'utiliser les bonnes options de chunk !
- 4) **Ajoutez une courte description** du jeu de données India TB, similaire à celle des accidents de moto en Colombie dans le rapport d'exemple, sous un nouveau *titre de section*. Les informations sur le jeu de données se trouvent dans le dossier “data”.
 - Rappelez-vous que les *titres de section* apparaîtront en grand texte dans votre rapport final, évitez donc d'utiliser des noms comme “Nettoyage des données”. Gardez ces notes cachées dans des chunks de code à la place.
- 5) **Créez un NOUVEAU chunk de code** dans lequel vous importez le jeu de données India TB. **Donnez au dataframe un titre court et descriptif**. N'oubliez pas d'utiliser la fonction `here()` lors du chargement du CSV, sinon le chargement des données pourrait ne pas fonctionner.
 - Astuce : Vous pouvez utiliser Ctrl+Alt+I pour insérer un nouveau chunk de code dans votre rmd.
- 6) **Inspectez le dataframe et choisissez deux variables catégorielles pour votre analyse**. Une fois que vous et votre groupe avez choisi vos variables souhaitées, vous pouvez procéder à la création des trois exigences pour l'exercice (créer graphique, créer tableau, et utilisation de code en ligne).
- 7) **Créez un graphique avec esquisse**. Rappelez-vous que les noms de variables dans esquisse sont codés par couleur en fonction du type de données. Les variables catégorielles sont *orange*.
 - Astuce : Un bon moyen de visualiser deux variables catégorielles est avec un *graphique à barres empilées*. Vous pouvez en créer un dans esquisse en ajoutant une variable à “X” ou “Y” et la deuxième variable à “Fill”.
- 8) **Créez un tableau en utilisant {flextable} ou {gt}**. Vous devriez résumer les MÊMES deux variables que vous avez choisies pour le tracé ci-dessus. N'oubliez pas d'utiliser les options de chunk appropriées pour que seul le joli tableau soit imprimé (voir le rapport d'exemple pour des indices).
 - Remarque: Vous pouvez créer un tableau de fréquence en utilisant la fonction `tabyl()`, mais cela crée un dataframe qui n'est pas très présentable dans un rapport.
- 9) **Mettez en évidence des nombres spécifiques** des données en utilisant **du code en ligne**. Voir le rapport d'exemple Rmd pour des indices !
- 10) **Tricotez votre rapport !** Cette partie finale et la plus importante.
 - Remarque: Avant de tricoter, assurez-vous de supprimer ou de commenter toutes les instances de `esquisser()` ou `View()`. C'est parce que les fonctions qui ouvrent de nouvelles fenêtres interrompent le processus de tricotage. Vous devrez peut-être essayer plusieurs fois avant de tricoter avec succès votre rapport. Soyez patient et demandez de l'aide aux instructeurs si vous êtes bloqué.

Pour soumettre votre travail, veuillez télécharger **JUSTE VOTRE Rmd** sur la page de soumission. Mais assurez-vous que le Rmd peut être tricoté avec succès avant de le télécharger.

Défi Bonus (Facultatif)

Vous avez terminé tôt ? Essayez un ou les deux exercices de défi ci-dessous.

Défi 1 : Utiliser patchwork pour combiner des images

OBJECTIF : Utiliser les opérations `{patchwork}` dans votre rapport Rmd et assembler deux graphiques !

Le package `{patchwork}` vous permet de combiner plusieurs graphiques en un seul, ce qui peut faciliter la comparaison des graphiques et garder votre visualisation de données ordonnée et organisée.

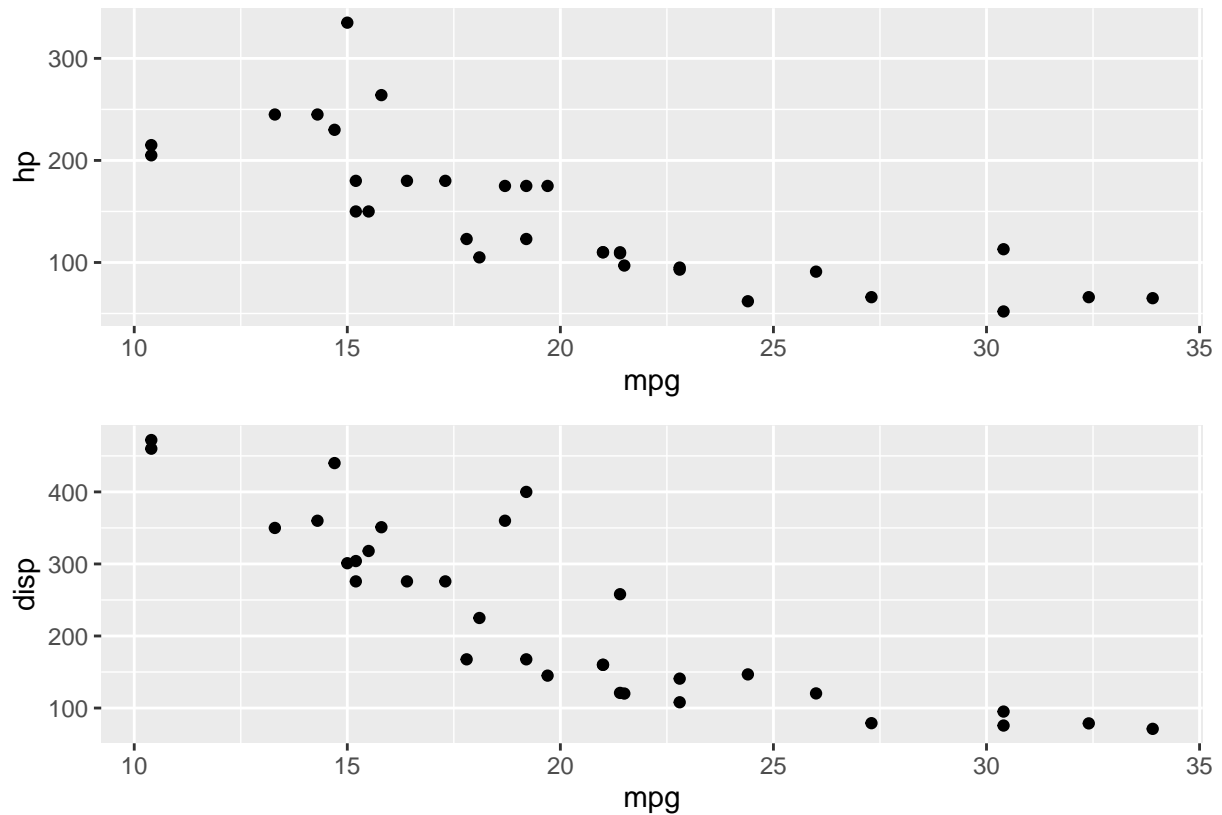
Prenons un exemple simple. Supposons que nous ayons deux graphiques appelés `plot1` et `plot2` :

```
# charger les packages requis
pacman::p_load(tidyverse, patchwork)

# Créer plot1
plot1 <- ggplot(mtcars, aes(x = mpg, y = hp)) +
  geom_point()

# Créer plot2
plot2 <- ggplot(mtcars, aes(x = mpg, y = disp)) +
  geom_point()

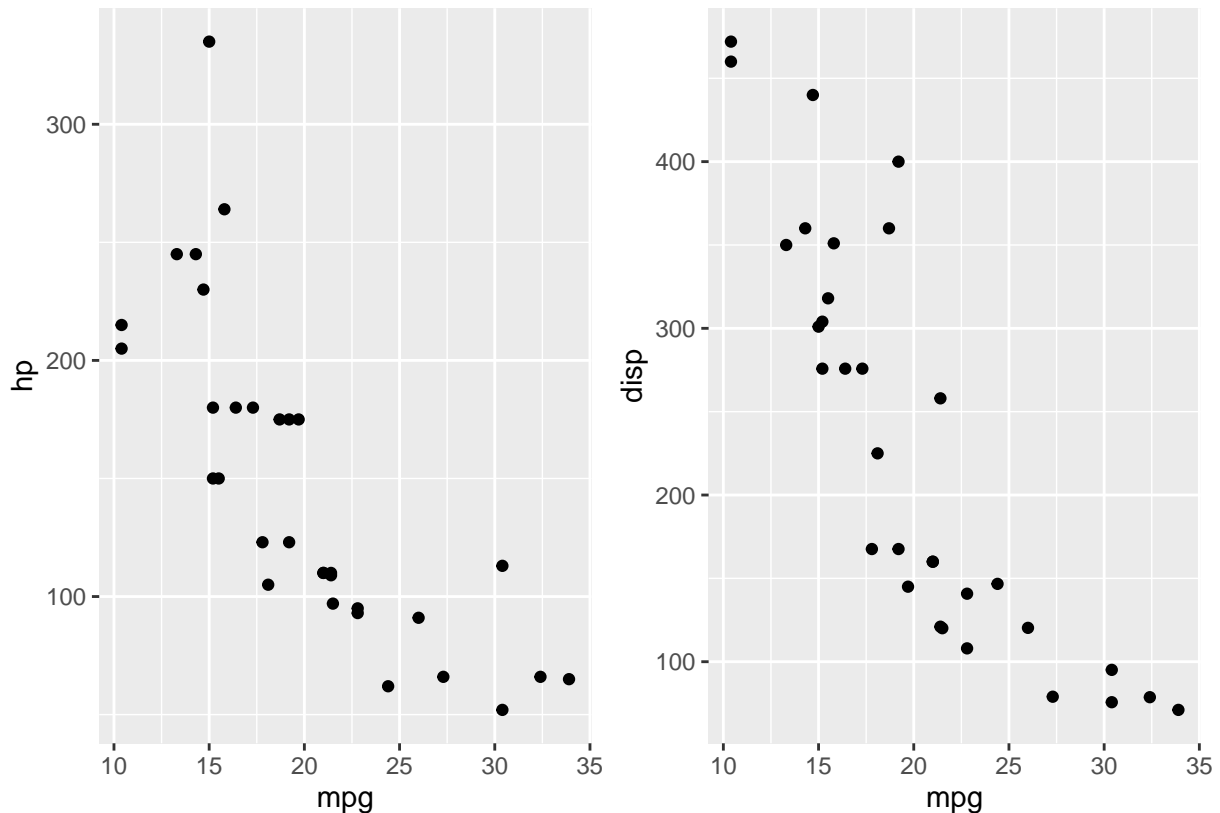
# Combiner les graphiques verticalement avec patchwork
combined_plot <- plot1 / plot2
combined_plot
```



Dans cet exemple, l'opérateur / est utilisé pour disposer les graphiques verticalement, avec `plot1` en haut et `plot2` en bas.

Si vous vouliez que les graphiques soient disposés horizontalement, vous utiliseriez l'opérateur + à la place :

```
# Combiner les graphiques verticalement avec patchwork
combined_plot_2 <- plot1 + plot2
combined_plot_2
```



Défi 2 : Utiliser des crochets carrés dans le code en ligne

OBJECTIF : Extraire des nombres d'un dataframe en utilisant la notation des crochets carrés, et les afficher dans votre rapport tricoté avec du code en ligne.

Dans R markdown, nous utilisons une syntaxe spécifique pour inclure du code R en ligne, qui est du code R qui s'exécute dans le texte de votre document. Pour utiliser du code R en ligne, nous incluons simplement le code R entre des backticks et le précédons d'un `r` comme ceci

Nous avons trouvé qu'il y avait `r votre_code_ici` cas dans la Catégorie B.

Une utilisation courante du code R en ligne est d'accéder à des éléments spécifiques d'un objet de données tel qu'une matrice ou un dataframe. Pour accéder à un élément spécifique, nous utilisons des crochets `[]` avec les indices de ligne et de colonne. Par exemple, si nous avons une matrice `m`, nous pouvons accéder à l'élément à la ligne 2, colonne 3 comme suit : `m[2, 3]`.

Prenons un exemple avec un tableau Markdown. Supposons que nous ayons un tableau comme suit :

Catégorie	Compte
A	10
B	20
C	30

Si nous voulons afficher le compte pour la Catégorie B dans notre texte, nous pourrions écrire quelque chose comme ceci dans notre script Rmarkdown (pas dans un chunk de code) :

```
Nous avons trouvé qu'il y avait `r counts[2, 2]` cas dans la Catégorie B.
```

Lorsque ce document R Markdown est tricoté, il s'affichera comme suit :

“Nous avons trouvé qu'il y avait 20 cas dans la Catégorie B.”