

Résumé sur le package ggplot2

Jingwen SU

12/20/2020

Introduction

À la fin du cours, tous les étudiants ont téléchargé les documents pertinents sur Github et partagé les résultats avec nous. Afin d'utiliser ces packages d'installation de R de manière plus complète, je vais lire, analyser et évaluer les articles de mes collègues pour améliorer mes lacunes.

Ceci est une introduction à l'article lu dans cet article. Si vous souhaitez en savoir plus, vous pouvez rechercher des références.

- **Title:** Note book GGplot_3
- **Auteurs:** SERVANT Thomas
- **Lien sur Github:** SERVANT Thomas

Synthèse du travail en question

L'auteur a d'abord présenté brièvement le package d'installation ggplot2, puis a présenté les données à utiliser.

Dans la seconde partie, l'auteur nous a montré le dessin de quatre types de graphes: "Le nuage de point", "Les Histogrammes", "Les boîtes à moustache", "Les graphes en violon".

Dans la troisième partie, l'auteur prend d'autres projets de cours que nous avons réalisés comme exemples pour montrer comment utiliser ggplot2 pour dessiner des images.

Dans la quatrième partie, il a présenté **facet_wrap**, a expliqué et comparé la fonction similaire ggplot2 :: facet_grid (). Et l'a démontré.

Dans la cinquième partie, l'auteur explique comment calculer la corrélation entre deux variables. Et des deux aspects de la façon de calculer et la façon d'images, il montre la corrélation des données différentes.

Contenu principal et explication

1. Base de données

Il existe de nombreuses bases de données intégrées dans notre package d'installation ou dans les données installées. Lorsque nous avons juste besoin de pratiquer ou d'essayer une certaine fonction, ceux-ci peuvent nous aider à éviter de trouver une base de données et de la télécharger. Ce sont rapides. Nous pouvons donc essayer de nous souvenir de quelques noms de bases de données simples afin de pouvoir les appeler directement la prochaine fois que nous recherchons des données.

Ou nous pouvons utiliser les méthodes suivantes pour afficher la base de données:

```
library(ggplot2)
head(mpg)
```

```
## # A tibble: 6 x 11
##   manufacturer model displ  year   cyl trans      drv    cty   hwy fl   class
##   <chr>          <chr> <dbl> <int> <int> <chr>    <chr> <int> <int> <chr> <chr>
## 1 audi          a4      1.8  1999     4 auto(l5)  f      18    29 p   compa~
## 2 audi          a4      1.8  1999     4 manual(m5) f      21    29 p   compa~
## 3 audi          a4      2    2008     4 manual(m6) f      20    31 p   compa~
## 4 audi          a4      2    2008     4 auto(av)   f      21    30 p   compa~
## 5 audi          a4      2.8  1999     6 auto(l5)  f      16    26 p   compa~
## 6 audi          a4      2.8  1999     6 manual(m5) f      18    26 p   compa~
```

2.Facet__wrap

L'auteur présente ici en détail l'utilisation et la différence des deux fonctions:

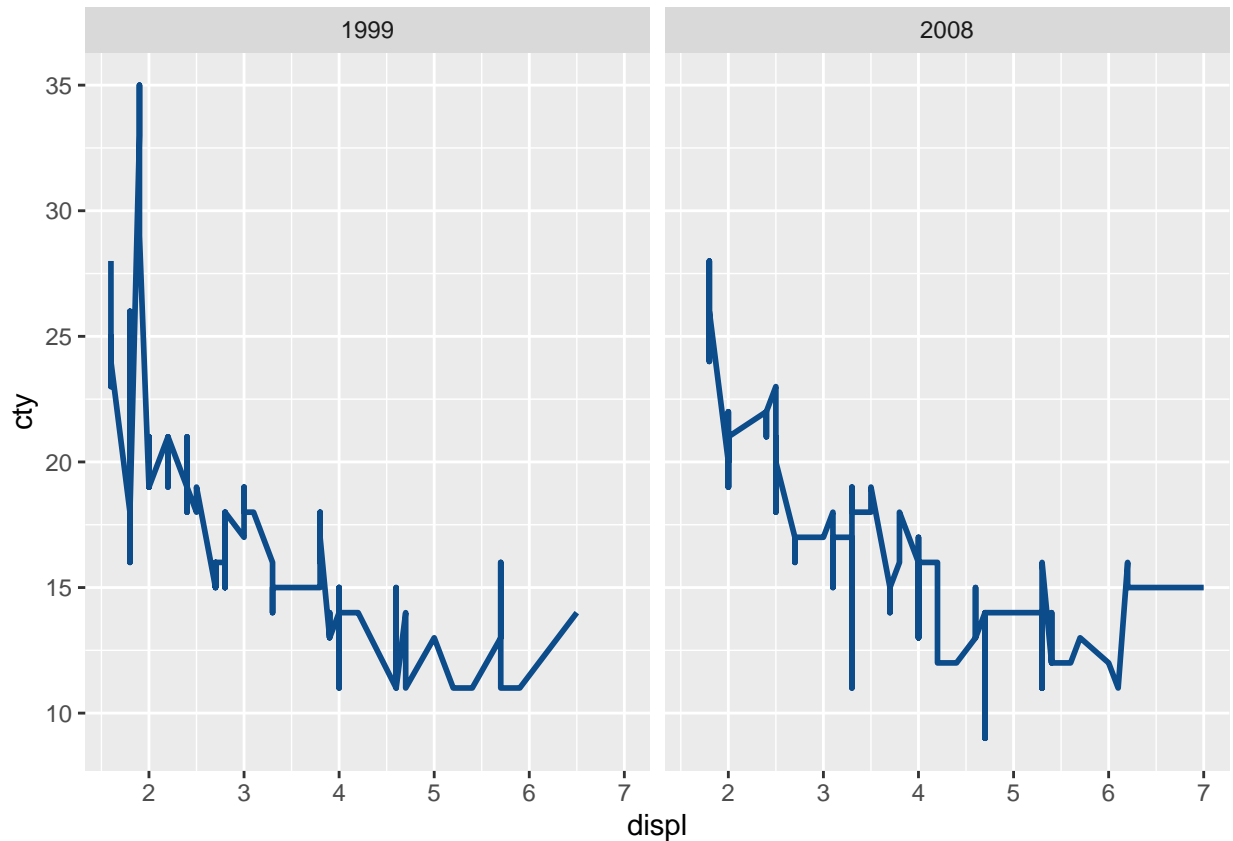
La fonction **Facet__graphprend** est équivalente à **ggplot2::facet__grid()** en ce qu'elle permet de construire une grille de petits multiples où les lignes et les colonnes correspondent à une valeur de données spécifique.

Bien que cela **ggplot2::facet__grid()** puisse être utilisé, cela entraînerait des résultats inattendus car il n'est pas possible de spécifier si vous faites référence à un nœud ou à un attribut d'arête. De plus **ggplot2::facet__grid()**, dessine les bords dans les panneaux même si le panneau ne contient pas les deux nœuds terminaux.

Facet__graphprend en charge tous ces problèmes, vous permettant de définir le type de données auquel les lignes et les colonnes font référence, ainsi que de filtrer les bords en fonction des nœuds de chaque panneau (même lorsque les nœuds ne sont pas dessinés).

```
#install.packages("igraph")

ggplot(mpg) +
  aes(x = displ, y = cty) +
  geom_line(size = 1L, colour = "#0c4c8a") +
  theme_grey() +
  facet_wrap(vars(year))
```



3. Pertinence des données

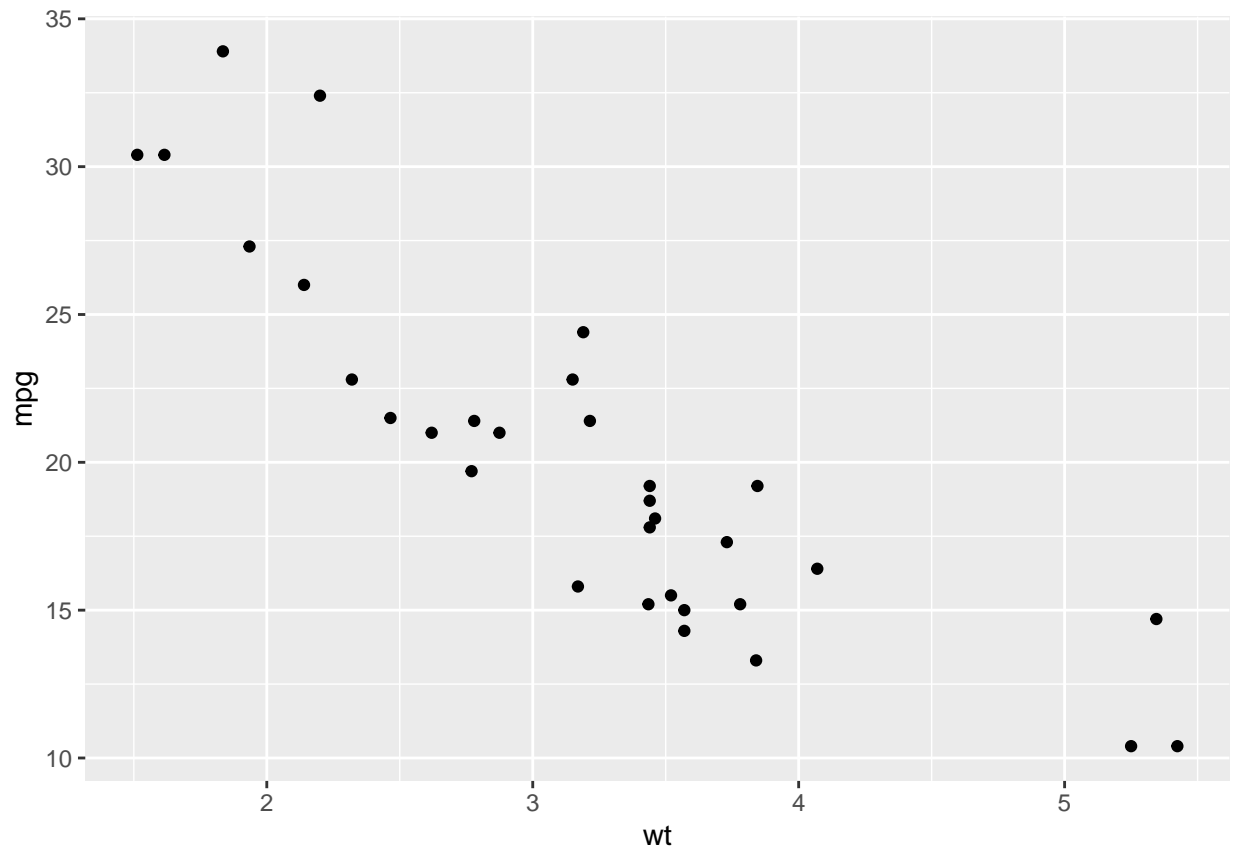
Pour calculer la corrélation entre deux variables on peut utiliser la fonction : “cor” , pour obtenir la matrice.

Une fois la matrice réalisée, nous allons utiliser la fonction “graph_from_adjacency_matrix” qui nous permettra de voir en un coup d’oeil les différentes corrélations. Durant cette période, afin de montrer la corrélation sous forme d’images, nous utiliserons la librairie igraph.

4. Les fonctions

- **Le nuage de point:** : `geom_point` (qui permet de modifier : taille, couleur et forme des points).

```
ggplot(mtcars, aes(x=wt, y=mpg))+ geom_point()
```

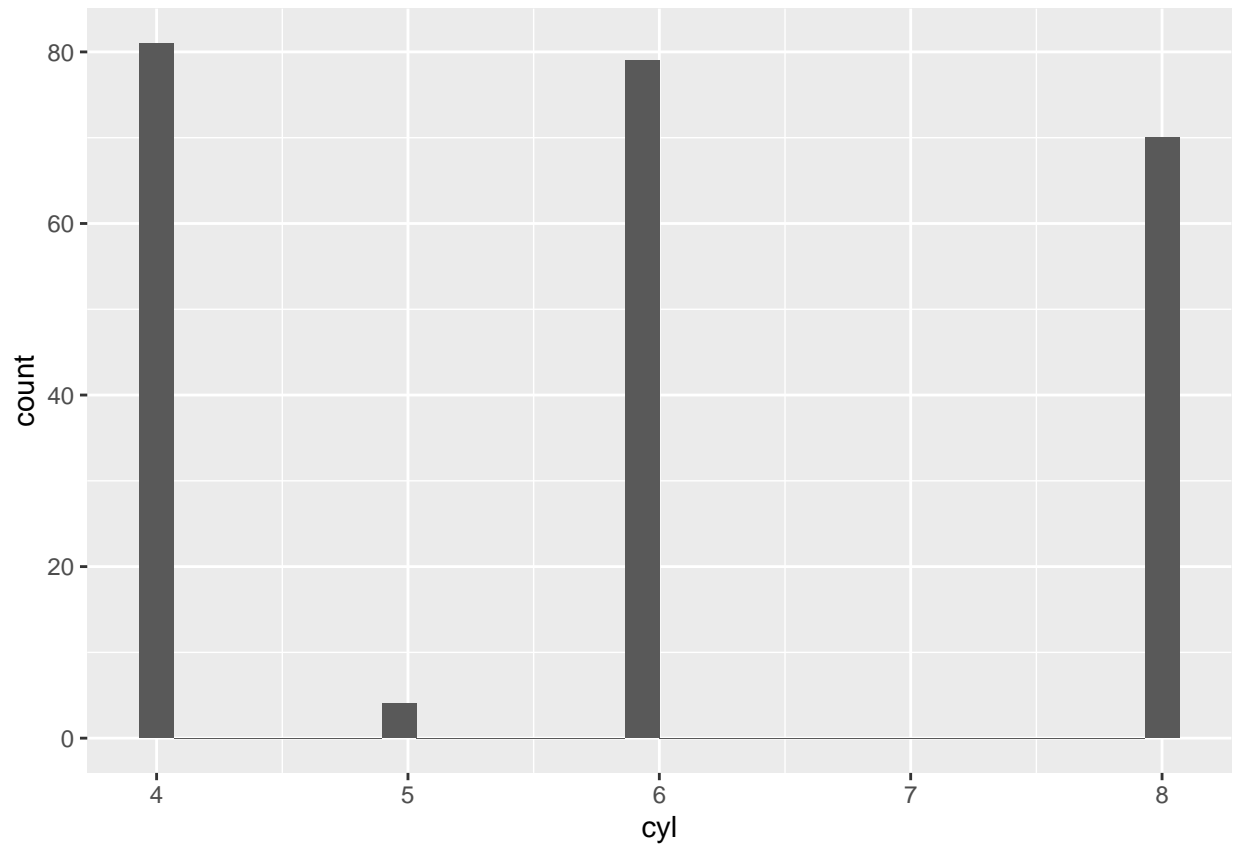


```
 #(definition du DataFrame, definition des axes : aes (x = , y = )+ geom_point()
```

- **Les Histogrammes:** `geom_histogram`

```
ggplot(mpg) + geom_histogram(aes(x = cyl))
```

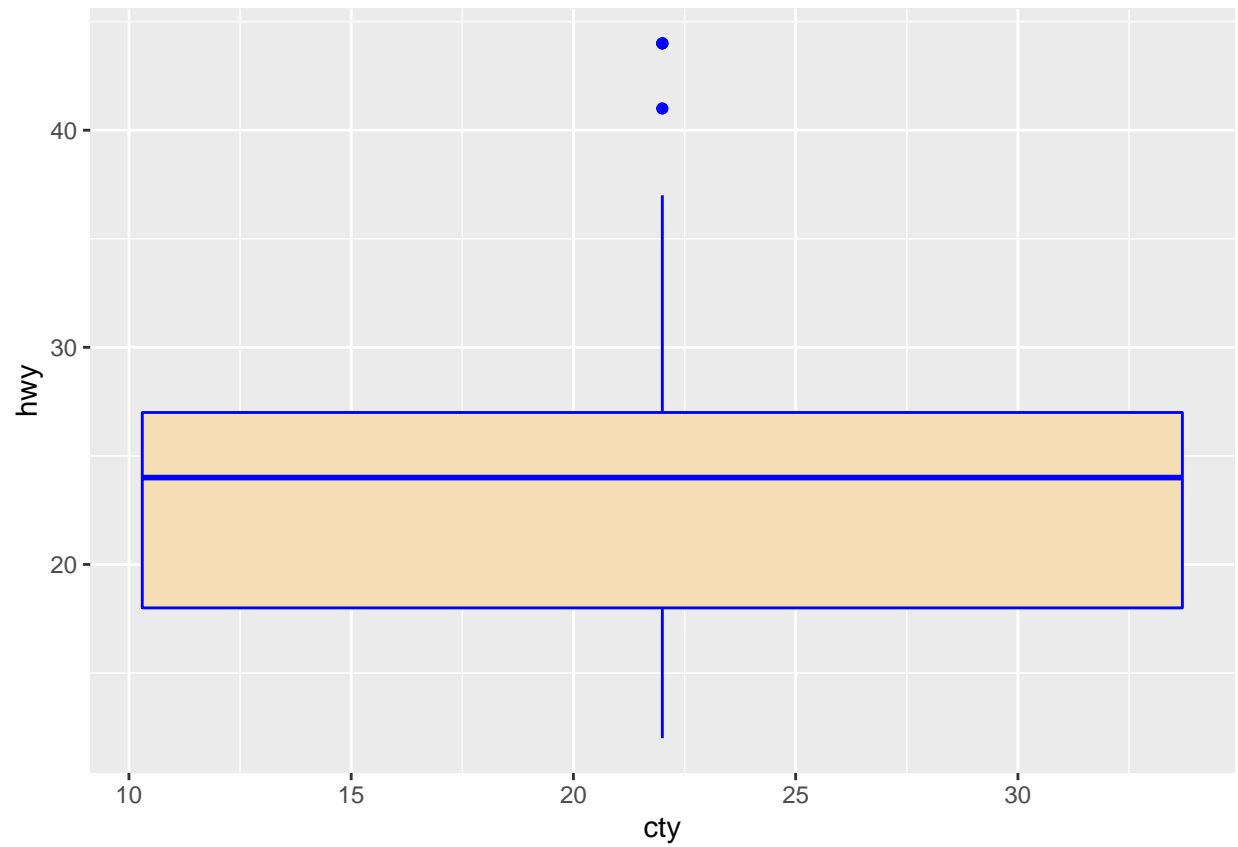
```
## 'stat_bin()' using 'bins = 30'. Pick better value with 'binwidth'.
```



- Les boites a mustache: `geom_boxplot`

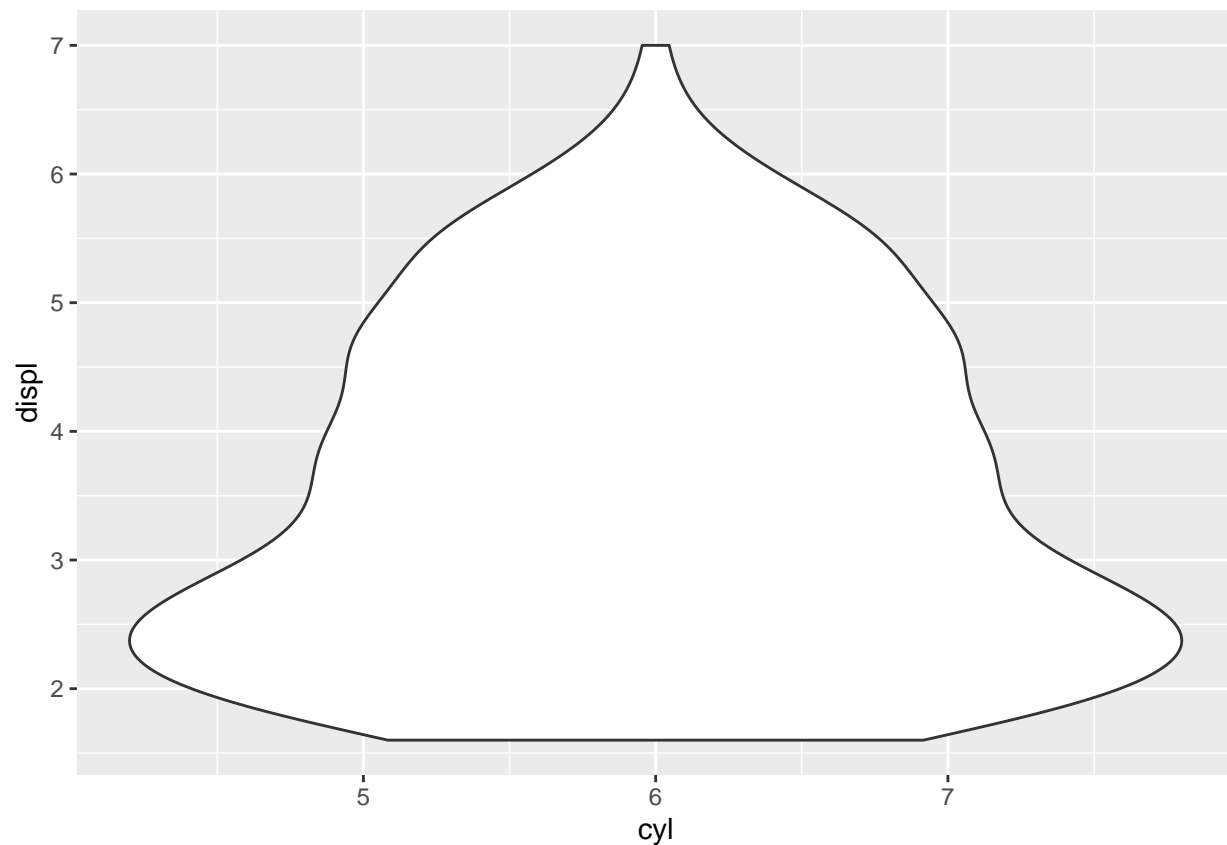
```
ggplot(mpg) + geom_boxplot(aes(x = cty , y = hwy),fill = "wheat", color = "blue")
```

```
## Warning: Continuous x aesthetic -- did you forget aes(group=...)?
```



- **Geom_violin** : geom_violin

```
ggplot(mpg) + geom_violin(aes(x = cyl, y = displ))
```



Evaluation et résumer

Selon mes normes, il s'agit d'un article de haut niveau. Bien que son titre soit erroné, le texte intégral est très riche. Il a présenté en détail tous les points et concepts qui nécessitent une attention particulière. En outre, il a également démontré les données de notre projet. Après une étape, il a approfondi sa compréhension des fonctions connexes et a eu sa propre réflexion. Cela vaut la peine d'être appris. J'aime la plénitude de son contenu.