

ggplot2

Jean Souris

10/12/2020

Introduction

Avant de commencer cette démonstration, je comptais remercier le Github suivant, qui m’a grandement aidé lors de la réalisation de ce travail :

Il existe plusieurs méthodes pour mettre en page différentes représentations graphiques de nos données.

Je vais utiliser les ressources ci-dessous qui serviront à la représentation graphiques de nos données :

<https://juba.github.io/tidyverse/08-ggplot2.html>

```
library(tidyverse)
```

```
## -- Attaching packages ----- tidyverse 1.3.0 --
```

```
## v ggplot2 3.3.2      v purrr  0.3.4
## v tibble  3.0.4      v dplyr  1.0.2
## v tidyr   1.1.2      v stringr 1.4.0
## v readr   1.4.0      v forcats 0.5.0
```

```
## -- Conflicts ----- tidyverse_conflicts() --
```

```
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
```

Je vais maintenant importer une base de donnée nommée ‘questionr’, qui contient les résultats partiels concernant les communes de plus de 2000 habitants de France métropolitaine :

```
#Je met un"#" devant cet argument car sinon il ne me laisse pas Kniter ce fichier en PDF :
```

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

Ensuite, je vais la selectionner juste en dessous :

```
library(questionr)
data(rp2012)
```

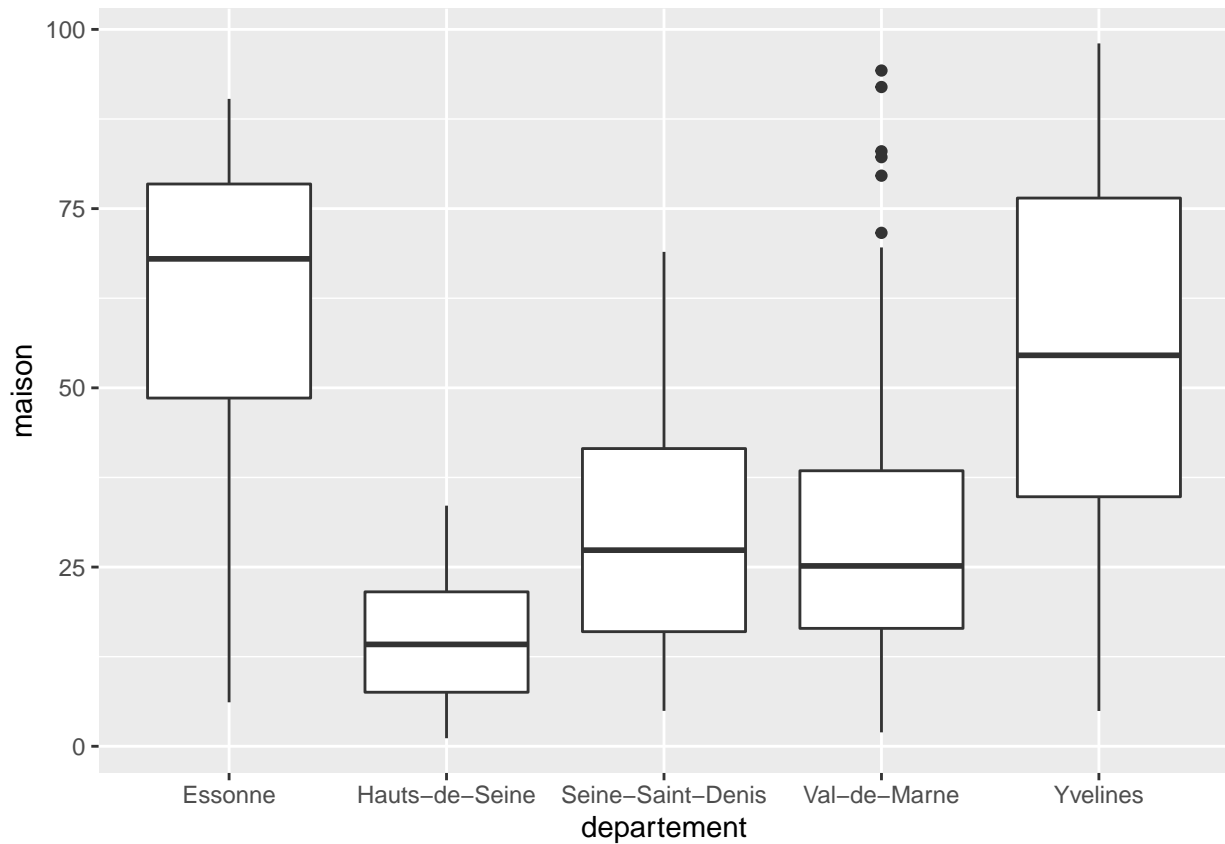
Mais, avant de créer mon grapahique, je dois créer une sorte de raccourcis qui me permet de “filtrer” ma base de donnée initiale :

```
rp <- filter(rp2012, departement %in% c("Hauts-de-Seine", "Yvelines", "Essonne", "Seine-Saint-Denis", "
```

La boîte à moustache :

La représentation graphique que j'ai choisi s'intitule "geom_boxplot" qui, en français représente une boîte à moustache en statistiques:

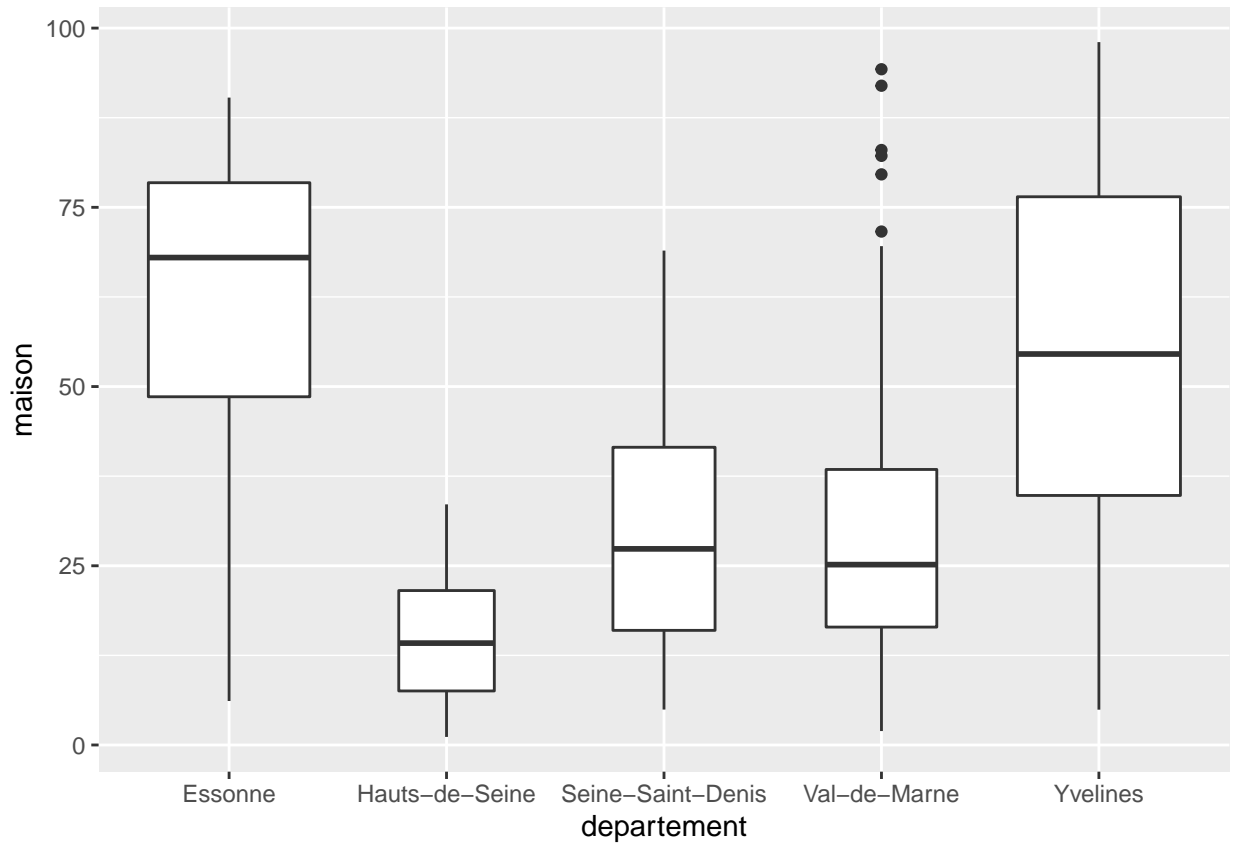
```
ggplot(rp) + geom_boxplot(aes (x = departement, y = maison))
```



Voici donc une boîte à moustache me représentant la répartition du taux de maisons en fonction des départements que j'ai sélectionnés.

Mais, grâce à ggplot2, je peux également en modifier sa couleur, tout comme la largeur de chaque boîte proportionnellement :

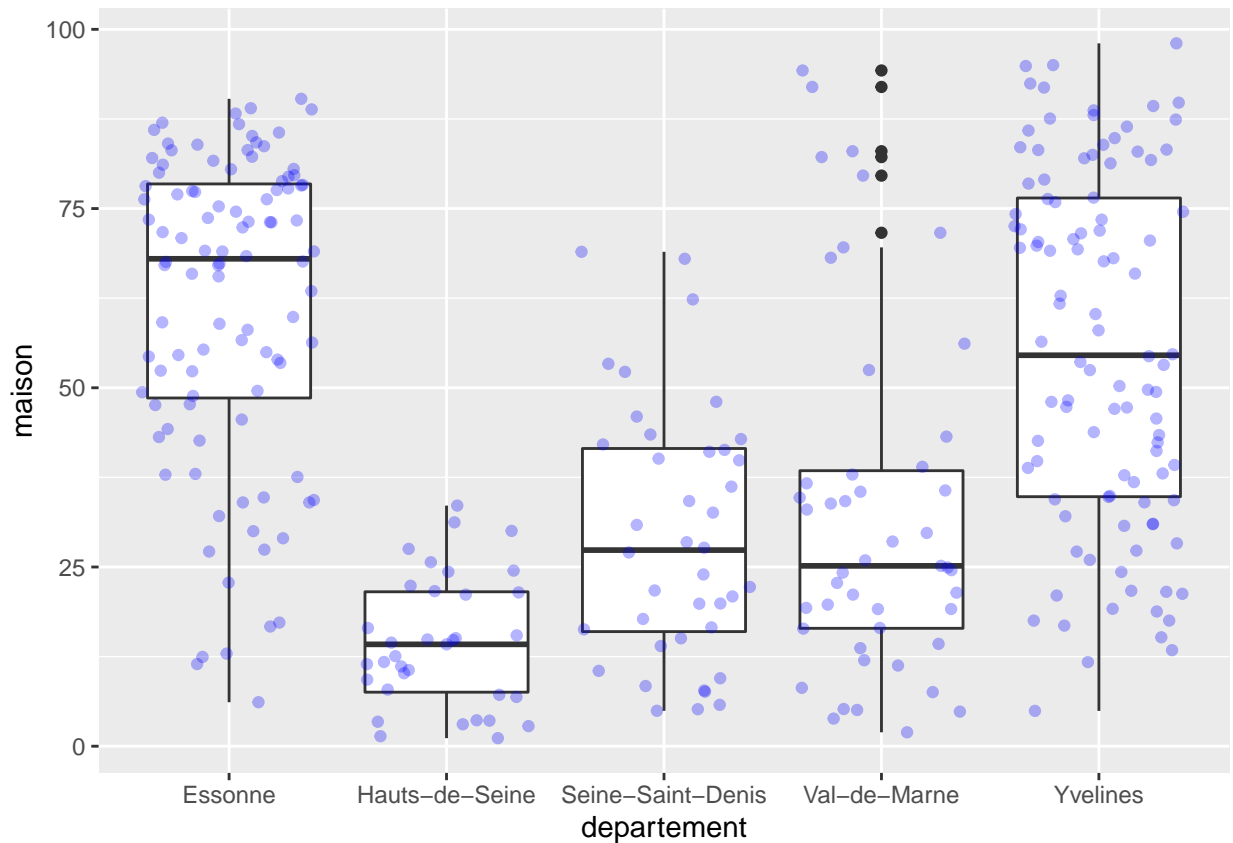
```
ggplot(rp) + geom_boxplot(aes(x = departement, y = maison), varwidth = TRUE)
```



J'obtiens maintenant une boîte à moustache, avec la taille de chaque boîte proportionnelle avec leur nombre de maison par département et ce, grâce à l'objet 'varwidth' que nous avons rajouté.

Pour en finir avec les boîtes à moustaches, il m'est possible d'accumuler plusieurs fonctions geom afin de superposer plusieurs statistiques ensembles, ou, comme dans ce cas-ci pour rendre notre boîte à moustache plus précise :

```
ggplot(rp) + geom_boxplot(aes(x = departement, y = maison)) + geom_jitter(aes(x = departement, y = maison))
```



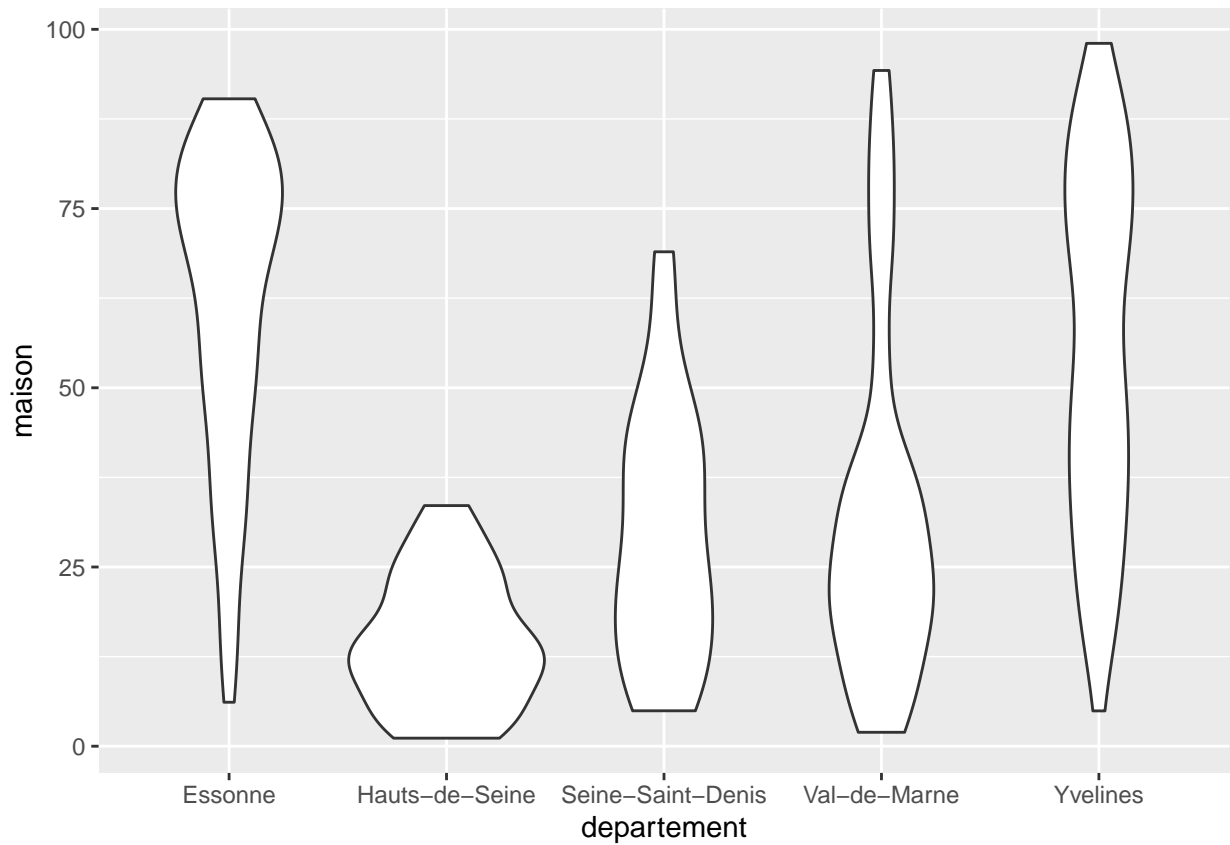
Dans ce graphique, grâce au geom “jitter”, qui est plus précis que le geom “point” pour cette demo, nous pouvons avoir une idée plus précise de la répartition des maisons par département.

Le Diagramme de Violon :

Je vais maintenant passer au diagramme de violon, aussi appelé `geom_violin` dans R.

Globalement, elle transmet les mêmes statistiques que la boîte à moustache, mais sous une forme déjà proportionnelle en fonction des boîtes, qui sont sur une même échelle graphique.. Pour en avoir un bon exemple, nous allons prendre pour chaque futurs graphes les mêmes statistique que les boîtes à moustaches vues.

```
ggplot(rp) + geom_violin(aes (x = departement, y = maison))
```



Voici donc la première représentation de notre boîte à moustache mais sous forme de diagramme de violon.

Tout comme la boîte à moustache, il nous est possible d'en modifier la couleur, que ce soit celle des contours, avec l'objet "col", ou bien l'intérieur avec "fill" :

```
ggplot(rp) + geom_violin(aes (x = departement, y = maison), col = "red", fill = "lightblue")
```

