

Note book GGplot_2

Utilisation de GGplot_2

ggplot2 est une extension du tidyverse qui permet de générer des graphiques avec une syntaxe cohérente et puissante. Elle nécessite l'apprentissage d'un "mini-langage" supplémentaire, mais permet la construction de graphiques complexes de manière efficace.

Une des particularités de ggplot2 est qu'elle part du principe que les données relatives à un graphique sont stockées dans un tableau de données (data frame, tibble ou autre). -

Les prérequis : Il faut tout d'abord avoir un data frame du type :

	mpg	cyl	disp	hp	drat	wt	qsec	vs	am	gear	carb
Mazda RX4	21.0	6	160	110	3.90	2.620	16.46	0	1	4	4
Mazda RX4 Wag	21.0	6	160	110	3.90	2.875	17.02	0	1	4	4
Datsun 710	22.8	4	108	93	3.85	2.320	18.61	1	1	4	1
Hornet 4 Drive	21.4	6	258	110	3.08	3.215	19.44	1	0	3	1
Hornet Sportabout	18.7	8	360	175	3.15	3.440	17.02	0	0	3	2
Valiant	18.1	6	225	105	2.76	3.460	20.22	1	0	3	1

Un graphique ggplot2 s'initialise à l'aide de la fonction `ggplot()`. Les données représentées graphiquement sont toujours issues d'un tableau de données (data frame), qu'on passe en argument `data` à la fonction :

`ggplot(data = rp) ## Ou, équivalent ggplot(rp)`

Une fois le tableau initialisé, il est temps de jouer avec les données. La manipulation passe par deux axes clés : - la définition des variables sur leur axe respectifs ; - le type de graphique à utiliser.

Il existe une multitude de graphiques qui ont chacun une fonction bien précise.

Le nuage de point utilisable avec la fonction : `geom_point()` (qui permet de modifier : taille, couleur et forme des points). `geom_point(size, color, shape)`

une fois l'initialisation terminée, il faut utiliser la fonction suivante : `ggplot(mtcars, aes(x=wt, y=mpg))+ geom_point()`

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

`geom_point(size, color, shape) ggplot(mtcars, aes(x=wt, y=mpg))+ geom_point() head(mtcars)`

Pour la question des points, tout est dans la definition du "geom_point", personnalisable à l'infini

`ggplot(rp) + geom_point(aes(x = dipl_sup, y = cadres), color = "darkgreen", size = 3, alpha = 0.3)`

LES HISTOGRAMMES: Pour cela, il faut utiliser la commande : `geom_histogram` qui se configure de la manière suivante

`geom_histogram(aes(x = mpg))`

Définir X permet d'afficher la donnée que l'on souhaite représenter.

LES BOITES A MOUSTACHE : utilisable avec la commande : "geom_boxplot" :

`ggplot(data) + geom_boxplot(aes(x = cyl , y = wt))`

exemple toujours avec le tableau ci dessus Personnalisable avec les commandes :

`} fill = "", color = ""`

Toujours avec les boîtes à moustaches, la commande “varwidth” permet de varier la longueur des boîtes en fonction des effectifs de la catégorie. Pour l’appliquer, on fait :

```
ggplot(data) + geom_boxplot(aes(x = cyl , y = wt), varwidth = TRUE)
```

Pour plus d’information : <https://juba.github.io/tidyverse/08-ggplot2.html#ressources-1>
(<https://juba.github.io/tidyverse/08-ggplot2.html#ressources-1>)