

# Graphiques en R

*Sophie Baillargeon, Université Laval*

*2019-05-24*

## Table des matières

<b>Introduction</b>	<b>2</b>
<b>Système graphique de base en R</b>	<b>4</b>
Procédure de création d'un graphique . . . . .	5
Fonction générique <code>plot</code> . . . . .	5
Différentes fonctions graphiques . . . . .	6
Arguments et paramètres graphiques . . . . .	16
Couleurs . . . . .	20
Ajout d'éléments à un graphique . . . . .	23
Possibilités graphiques spécifiques . . . . .	26
Annotations mathématiques . . . . .	26
Plusieurs graphiques dans une même fenêtre . . . . .	28
Aspects techniques . . . . .	32
Fenêtres graphiques . . . . .	32
Enregistrer un graphique . . . . .	33
Autres fonctions et trucs utiles . . . . .	33
Résumé et point de vue . . . . .	34
<b>Autres systèmes graphiques en R</b>	<b>37</b>
Package <code>lattice</code> . . . . .	38
Package <code>ggplot2</code> . . . . .	41
La grammaire graphique . . . . .	41
Survol des principales fonctions . . . . .	41
Code gabarit minimaliste . . . . .	41
Objet de classe <code>ggplot</code> . . . . .	42
Quelques fonctions de type <code>geom</code> . . . . .	43
Quelques autres fonctions . . . . .	43
Exemples de graphiques de différents types . . . . .	43
Exemples de configuration de graphiques . . . . .	52
Enregistrement d'un graphique - fonction <code>ggsave</code> . . . . .	59
Comparaison entre <code>ggplot2</code> et le système graphique R de base . . . . .	60
<b>Autres possibilités graphiques en R (seulement mentionnées)</b>	<b>60</b>
Graphiques 3D . . . . .	60
Fonction <code>persp</code> du système de base . . . . .	60
Fonction <code>cloud</code> du package <code>lattice</code> . . . . .	61
Graphiques 3D orientables - package <code>rgl</code> . . . . .	61
Cartes géographiques . . . . .	62
Packages <code>maps</code> et <code>mapdata</code> . . . . .	62
Autres packages pour l'intégration de cartes géographiques dans un graphique R . . . . .	63
Graphiques interactifs . . . . .	63
Graphiques interactifs Plotly - package <code>plotly</code> . . . . .	63
Autres packages pour la création de graphiques interactifs en R . . . . .	64
<b>Conclusion</b>	<b>64</b>

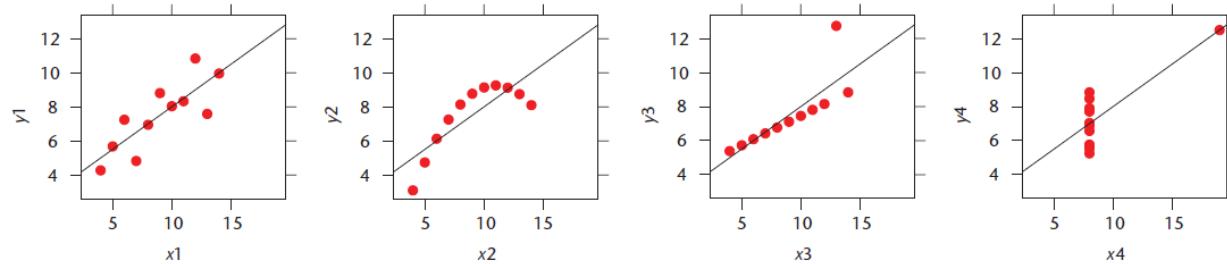
Synthèse	65
Références	70
Annexe	71
Graphiques produits par <code>plot</code> selon les types d'objets donnés en entrée aux arguments <code>x</code> et <code>y</code> . . .	71

---

## Introduction

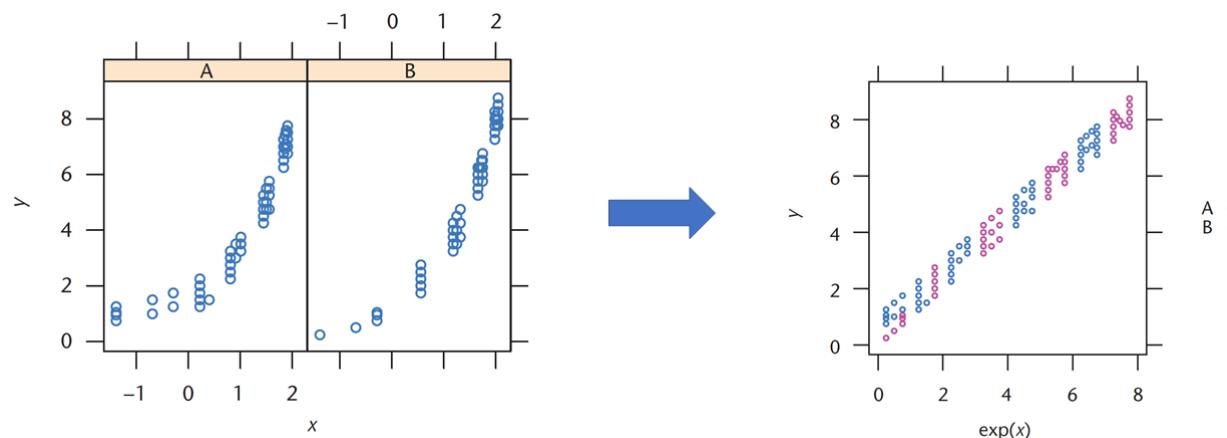
### La visualisation graphique, une étape importante

La communauté statistique insiste depuis longtemps sur l'importance de visualiser graphiquement des données avant de les analyser. En 1973, Anscombe<sup>1</sup> publiait les [quatre jeux de données suivants](#), ayant les mêmes paramètres estimés de régression linéaire simple ( $y = 3 + 0.5x$ ), mais présentant des relations bien différentes !



Plusieurs autres exemples de [jeux de données possédant les mêmes statistiques descriptives, mais des représentations graphiques très différentes](#) ont été publiés depuis<sup>2</sup>.

Certains ont même mis en évidence l'importance de représenter graphiquement des données de plusieurs manières différentes afin de découvrir les surprises qu'elles peuvent cacher. Par exemple, les données suivantes, créées pour enseigner à des étudiants l'importance des graphiques en analyse de données<sup>3</sup>, forment le mot SUPERMAN si elles sont représentées sur une échelle adéquatement transformée et si les observations des deux groupes sont superposées.



<sup>1</sup>Anscombe, F.J. (1973). Graphs in Statistical Analysis. *The American Statistician*, **27**(1). pp. 17-21.

<sup>2</sup>Matejka, J. et Fitzmaurice, G. (2017, Mai). Same Stats, Different Graphs : Generating Datasets with Varied Appearance and Identical Statistics through Simulated Annealing. *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*. ACM. pp. 1290-1294. URL : <https://www.autodeskresearch.com/publications/samestats>

<sup>3</sup>Kozak, M. (2002). Watch out for superman : first visualize, then analyze. *IEEE Computer Graphics and Applications*. **32**(3), pp. 6–9.

## Les graphiques, une force de R

En plus d'être un environnement de calculs statistiques, R permet de produire des graphiques. Il s'agit d'une utilité importante de R. La toute première phrase du [site web de R](#) mentionne cette utilité.

« R is a free software environment for statistical computing and graphics. »

Lors de comparaison entre R et d'autres logiciels permettant de faire de l'analyse de données, les capacités graphiques de R sont souvent considérées comme un de ses meilleurs atouts<sup>45</sup>.

L'installation de base de R comporte beaucoup de fonctions graphiques. La première partie de ces notes présente ces fonctions. Elles permettent de produire des graphiques pour lesquels l'utilisateur garde le plein contrôle sur l'apparence du graphique.

De plus, deux systèmes graphiques alternatifs ont également été développés en R et sont abordés ici :

- **lattice** : package graphique qui met l'emphase sur les représentations multivariables,
- **ggplot2** : package graphique conçu en ayant comme objectif la simplicité d'utilisation et la qualité des graphiques produits.

Pour finir, quelques autres possibilités graphiques en R sont mentionnées rapidement :

- graphiques 3D,
- cartes géographiques,
- graphiques interactifs.

La grande quantité de possibilités graphiques en R contribue à en faire un puissant outil de visualisation.

## Présentation des données utilisées pour les exemples

Plusieurs des exemples de graphiques de ces notes représentent les mêmes données : celles stockées dans le [data frame quakes](#) du package [datasets](#) (chargé par défaut lors de l'ouverture de session).

```
str(quakes)
```

```
## 'data.frame':    1000 obs. of  5 variables:
##   $ lat      : num  -20.4 -20.6 -26 -18 -20.4 ...
##   $ long     : num  182 181 184 182 182 ...
##   $ depth    : int  562 650 42 626 649 195 82 194 211 622 ...
##   $ mag      : num  4.8 4.2 5.4 4.1 4 4 4.8 4.4 4.7 4.3 ...
##   $ stations: int  41 15 43 19 11 12 43 15 35 19 ...
```

Il s'agit d'observations pour 1000 événements sismiques survenus près des îles Fidji depuis 1964. Pour chaque événement, nous avons :

- **lat** et **long** : sa localisation géographique en latitude-longitude,
- **depth** : sa [profondeur](#),
- **mag** : sa [magnitude](#) et
- **stations** : le nombre de stations sismiques ayant rapporté l'événement.

Pour certains exemples, nous aurons besoin de variables catégoriques stockées sous forme de facteur. Ajoutons donc les deux variables supplémentaires suivantes au jeu de données **quakes** :

- **magFactor** : magnitude du séisme arrondie à l'entier inférieur (4, 5 ou 6),
- **region** : région où a eu lieu le séisme ("Ouest" ou "Est").

```
quakes$magFactor <- factor(floor(quakes$mag))
quakes$region <- factor(quakes$long >= 175, labels = c("Ouest", "Est"))
```

<sup>4</sup><http://www.theanalyticslab.nl/2017/03/18/python-r-vs-spss-sas/>

<sup>5</sup><https://www.datacamp.com/community/tutorials/r-or-python-for-data-analysis>

**Remarque** : Nous avons maintenant deux objets nommés `quakes` dans notre chemin de recherche, parce que nous avons ajouté des variables au data frame dans des exemples précédents. Le data frame modifié, soit celui avec les variables ajoutées, est situé dans notre environnement de travail. Alors que le data frame `quakes` original est toujours dans l'environnement du package `datasets`. Celui-là n'a pas été modifié.

Le chemin de recherche de notre session R, pour une commande soumise dans la console, est le suivant.

```
search()
```

```
## [1] ".GlobalEnv"           "package:stats"      "package:graphics"  
## [4] "package:grDevices"     "package:utils"       "package:datasets"  
## [7] "package:methods"        "Autoloads"          "package:base"
```

Lorsque nous demandons à R d'accéder à l'objet nommé `quakes`, il cherche ce nom dans les environnements de son chemin de recherche, un environnement à la fois, en respectant l'ordre des environnements dans la liste précédente. L'environnement "`.GlobalEnv`", soit notre environnement de travail, est le premier de la liste. Lorsque R trouve le nom, il cesse sa recherche. Nous reviendrons sur le fonctionnement des évaluations en R dans un autre cours.

Ainsi, toute commande dans la console contenant le nom `quakes`, accédera à l'objet nommé `quakes` de notre environnement de travail.

```
str(quakes)
```

```
## 'data.frame':   1000 obs. of  7 variables:  
## $ lat      : num  -20.4 -20.6 -26 -18 -20.4 ...  
## $ long     : num  182 181 184 182 182 ...  
## $ depth    : int  562 650 42 626 649 195 82 194 211 622 ...  
## $ mag      : num  4.8 4.2 5.4 4.1 4 4 4.8 4.4 4.7 4.3 ...  
## $ stations : int  41 15 43 19 11 12 43 15 35 19 ...  
## $ magFactor: Factor w/ 3 levels "4","5","6": 1 1 2 1 1 1 1 1 1 ...  
## $ region   : Factor w/ 2 levels "Ouest","Est": 2 2 2 2 2 2 1 2 2 2 ...
```

Il est possible de forcer R à aller chercher l'objet `quakes` dans le package `datasets` avec l'opérateur `:::`.

```
str(datasets::quakes)
```

```
## 'data.frame':   1000 obs. of  5 variables:  
## $ lat      : num  -20.4 -20.6 -26 -18 -20.4 ...  
## $ long     : num  182 181 184 182 182 ...  
## $ depth    : int  562 650 42 626 649 195 82 194 211 622 ...  
## $ mag      : num  4.8 4.2 5.4 4.1 4 4 4.8 4.4 4.7 4.3 ...  
## $ stations: int  41 15 43 19 11 12 43 15 35 19 ...
```

---

## Système graphique de base en R

Les graphiques de base en R sont créés grâce aux fonctions provenant des packages `graphics` et `grDevices` développés par le « R core team ». Ces packages sont intégrés à l'installation de base de R et ils sont chargés par défaut lors de l'ouverture d'une session R. Dans cette section, nous couvrirons les possibilités de ces packages.

## Procédure de création d'un graphique

Un programme pour créer un graphique avec le système graphique de base en R se décompose typiquement en 3 étapes, les suivantes :

1. La configuration des paramètres graphiques (au besoin) :
  - énoncé `par` ou `layout`.
2. L'initialisation d'un graphique :
  - fonction de base : `plot`,
  - ou fonction pour un type spécifique graphiques : `pairs`, `matplot`, `pie`, `barplot`, `mosaicplot`, `hist`, `boxplot`, `qqnorm`, `qqplot`, `curve`, etc.
3. L'ajout séquentiel d'éléments au graphique (au besoin) :
  - fonctions d'ajouts à un graphique déjà initialisé :
    - `points`, `matpoints`, `lines`, `matlines`, `abline`, `segments`, `arrows`, `rect`, `polygon`, `legend`, `text`, `mtext`, `title`, `axis`, `box`, `qqline`, etc.;
    - `matplot`, `barplot`, `hist`, `boxplot`, `curve` avec l'argument `add = TRUE`.

Lorsque les paramètres graphiques sont modifiés avec un appel à la fonction `par` au début du programme, une bonne pratique est d'ajouter aussi à la fin du programme une commande pour redonner aux paramètres graphiques leurs valeurs par défaut.

La mise en forme et les annotations du graphique (p. ex. titre et noms d'axe) sont déterminées par les paramètres graphiques ainsi que par des arguments des fonctions graphiques. Avec un peu de patience et quelques lignes de code, il est possible d'arriver à contrôler parfaitement l'apparence du graphique.

Les prochaines sections présentent les concepts à connaître pour suivre cette procédure de création de graphique et contiennent beaucoup d'exemples. Cependant, la matière n'est pas couverte dans le même ordre que les étapes de la procédure. Même si dans le code de création d'un graphique la configuration des paramètres est l'étape 1, il faut d'abord savoir créer un graphique pour que cette étape ait du sens. De plus, la configuration des paramètres graphique est une étape facultative. Parfois, les valeurs par défaut des paramètres font l'affaire.

Ainsi, nous allons d'abord étudier la principale fonction du système graphique de base en R, soit la fonction `plot`. Dans un deuxième temps, nous ferons un survol de plusieurs autres fonctions de création de graphiques. Ensuite, les arguments et paramètres graphiques qui permettent de contrôler les annotations et la mise en forme seront présentés. Pour terminer la couverture des étapes de la procédure de création, il ne restera plus qu'à parler des fonctions d'ajouts d'éléments à un graphique.

Les exemples présentés seront simple au début, mais se complexifient au fil de la section.

## Fonction générique `plot`

La principale fonction graphique en R est la fonction `plot`. En fait, `plot` est une fonction générique. Il existe plusieurs méthodes associées à cette fonction générique. Pour afficher la liste des méthodes `plot` chargées dans notre session R, il faut soumettre la commande suivante.

```
methods(plot)
```

```
## [1] plot.acf*          plot.data.frame*   plot.decomposed.ts* plot.default
## [5] plot.dendrogram*    plot.density*      plot.ecdf           plot.factor*
## [9] plot.formula*       plot.function      plot.hclust*        plot.histogram*
## [13] plot.HoltWinters*   plot.isoreg*       plot.lm*            plot.medpolish*
## [17] plot.mlm*           plot.ppr*          plot.prcomp*       plot.princomp*
## [21] plot.profile.nls*   plot.raster*       plot.spec*         plot.stepfun
## [25] plot.stl*           plot.table*       plot.ts             plot.tskernel*
## [29] plot.TukeyHSD*
## see '?methods' for accessing help and source code
```

Si nous donnons en entrée à la fonction `plot` des vecteurs en arguments `x` et `y`, c'est la méthode `plot.default` qui est utilisée. Cette méthode trace un diagramme de dispersion.

La fonction `plot` choisit la méthode à utiliser en fonction des classes des arguments `x` et `y` fournis en entrée. C'est une caractéristique « orientée objet » du langage R. Nous reviendrons sur ce concept plus tard. Pour l'instant, il suffit de comprendre que les différentes méthodes produisent différents résultats et acceptent différents arguments. En fait, ces méthodes font souvent appels à d'autres fonctions de création de graphique.

L'annexe contient une panoplie d'exemples de graphiques pouvant être produits avec la fonction générique `plot`, selon les objets qui lui sont données en entrée. Voici ici un résumé des possibilités couvertes dans ces exemples.

### Résumé de graphiques usuels produits par `plot`

Type x	Type y	Graphique produit (méthode utilisée → fonction ou autre méthode appelée)
vecteur	-	diagramme de dispersion en fonction de l'index des observations ( <code>plot.default</code> )
	vecteur	diagramme de dispersion ( <code>plot.default</code> )
facteur	-	diagramme en bâtons ( <code>plot.factor</code> → <code>barplot</code> )
	vecteur	diagrammes en boîtes juxtaposés ( <code>plot.factor</code> → <code>boxplot</code> )
	facteur	diagramme en mosaïque ( <code>plot.factor</code> → <code>spineplot</code> )
data frame	-	dépend du nombre de variables, de leur ordre et de leurs natures
		( <code>plot.data.frame</code> → <code>pairs</code> ou <code>plot.default</code> ou <code>plot.factor</code> ou <code>stripchart</code> )
		dépend de la position des variables dans la formule et de leurs natures
formule	-	( <code>plot.formula</code> → <code>plot.default</code> ou <code>plot.factor</code> ou <code>plot.data.frame</code> )
fonction	-	courbe ( <code>plot.function</code> → <code>curve</code> )

La fonction `plot` peut en faire beaucoup plus que ça grâce à ses nombreuses méthodes. Par exemple, si elle reçoit en entrée la sortie d'un appel à la fonction `lm` qui ajuste un modèle linéaire à des données, elle produit des graphiques de résidus de ce modèle via sa méthode `plot.lm`.

### Différentes fonctions graphiques

En explorant les capacités de la fonction générique `plot`, nous découvrons plusieurs fonctions de création de graphiques. Ces fonctions permettent de créer des graphiques de différents types.

Voici un résumé des principaux types de graphiques pouvant être créés dans le système graphique de base en R, suivi d'exemples.

Représentations d'une variable (observations stockées dans `x`) :

Type de graphique	Appel de fonction	Type de x
diagramme en secteurs ( <i>pie chart</i> )	<code>pie(table(x), ...)</code>	facteur
diagramme en bâtons ( <i>bar plot</i> )	<code>barplot(table(x), ...)</code>	facteur
histogramme	<code>hist(x, ...)</code>	vecteur numérique
courbe de densité à noyau ( <i>kernel density plot</i> )	<code>plot(density(x), ...)</code> → méthode <code>plot.density</code>	vecteur numérique
diagramme en boîte ( <i>boxplot</i> )	<code>boxplot(x, ...)</code>	vecteur numérique
diagramme quantile-quantile théorique normal	<code>qqnorm(x, ...)</code>	vecteur numérique

Représentation d'une expression mathématique : `curve(expr, ...)`

Représentations de deux variables (observations stockées dans `x` et `y`) :

Type de graphique	Appel de fonction	Type de <code>x</code>	Type de <code>y</code>
diagramme en bâtons empilés ou groupés	<code>barplot(table(x, y), ...)</code> avec <code>beside = TRUE</code> pour bâtons groupés	facteur	facteur
diagramme en mosaïque	<code>mosaicplot(tablex, y), ...)</code>	facteur	facteur
diagrammes en boîte juxtaposés	<code>boxplot(y ~ x, ...)</code>	facteur	vecteur numérique
diagramme de dispersion ( <i>scatterplot</i> ) ou en lignes ( <i>line chart</i> )	<code>plot(x, y, ...)</code> → méthode <code>plot.default</code>	vecteur numérique	vecteur numérique
diagramme quantile-quantile empirique	<code>qqplot(x, y, ...)</code>	vecteur numérique	vecteur numérique

Représentation de plus de trois variables numériques :

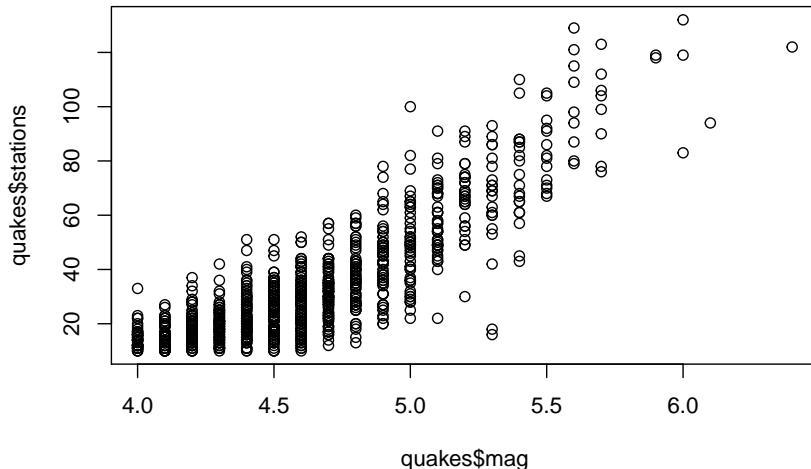
- matrice de diagrammes de dispersion : `pairs(~ x + y + z, ...)`
- diagrammes de dispersion superposés : `matplot(matriceX, matriceY, ...)`

### Exemples de graphiques de différents types

#### Diagramme de dispersion (ou en lignes) - fonctions `plot`, `pairs` et `matplot`

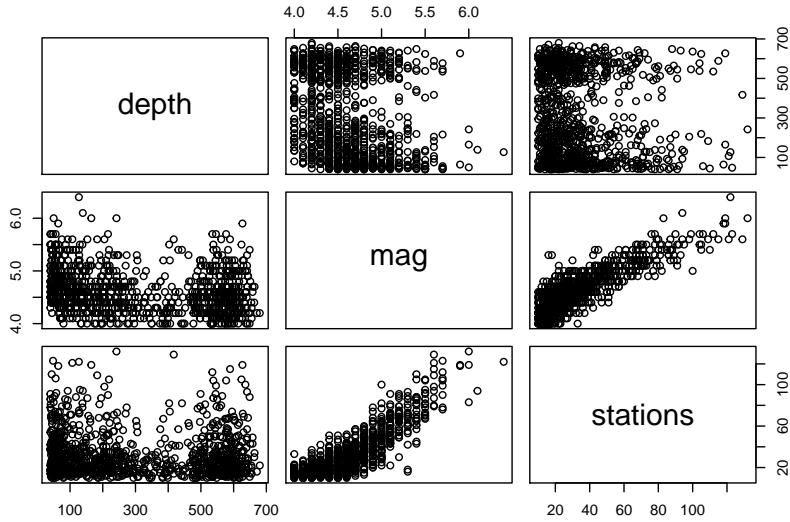
La méthode `plot.default` (via la fonction générique `plot` ou directement) produit un *diagramme de dispersion (scatterplot)*.

```
plot.default(x = quakes$mag, y = quakes$stations)
```



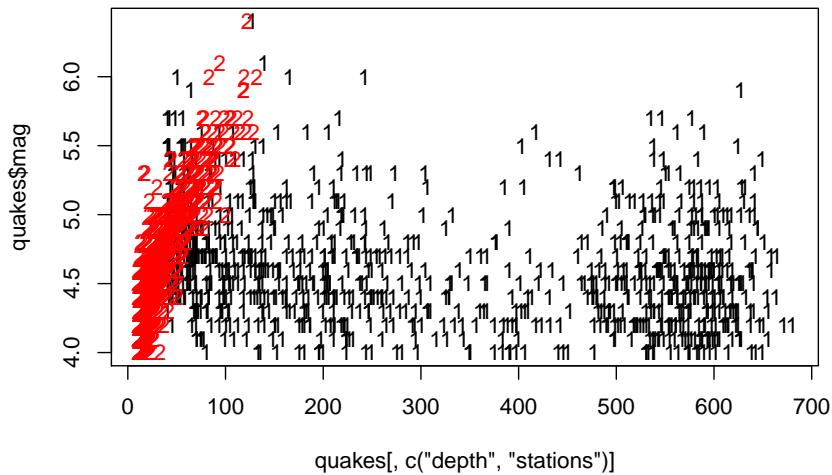
La fonction `pairs` produit une matrice de diagrammes de dispersion entre plusieurs paires de variables.

```
pairs(~ depth + mag + stations, data = quakes)
```



La fonction `matplot` permet de facilement superposer plusieurs diagrammes de dispersion. Elle associe les colonnes de la structure à deux dimensions fournie en `x` à celles de la structure à deux dimensions fournie en `y` (colonne  $i$  de `x` associée à la colonne  $i$  de `y`). La règle de recyclage s'applique si les objets fournis en `x` et `y` ne sont pas de même dimensions, comme dans l'exemple suivant.

```
matplot(x = quakes[, c("depth", "stations")], y = quakes$mag)
```



### Diagramme en secteurs - fonction `pie`

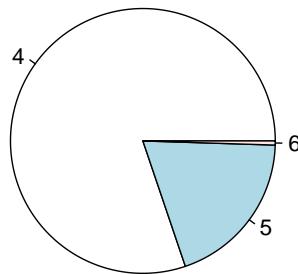
La fonction `pie` permet de créer des [diagrammes en secteurs \(pie charts\)](#), aussi nommés diagrammes circulaires ou diagrammes en pointes de tarte ou camemberts. Ce type de graphique est à utiliser avec parcimonie. Il y a d'ailleurs une mise en garde concernant l'utilisation de ce type de graphique dans la [fiche d'aide de la fonction pie](#).

Ce qui est reproché au diagramme en secteurs est qu'il difficile pour l'œil humain de rapidement comparer les aires des secteurs (qui représentent les fréquences). Il est plus facile de comparer des hauteurs de bâtons.

Donc le diagramme en bâtons serait un meilleur outil pour représenter des fréquences.

Malgré tout, je crois que le diagramme en secteur est utile pour représenter les fréquences d'une **variable à peu de modalités**, dans le cas où **il n'y a pas de fréquences presque égales**, comme dans l'exemple suivant.

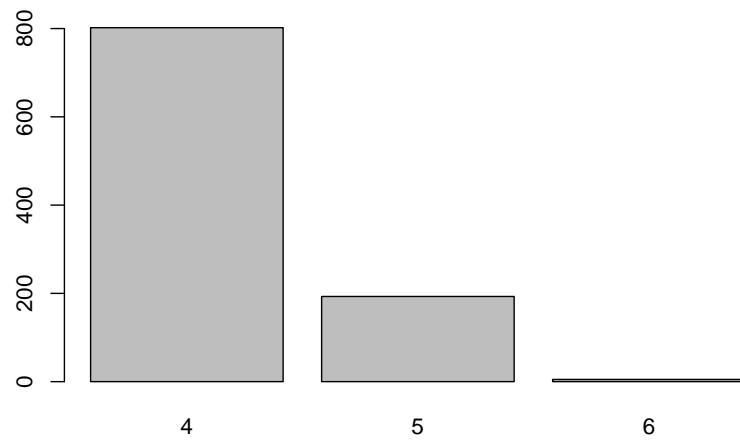
```
pie(table(quakes$magFactor))
```



### Diagramme en bâtons - fonction `barplot`

La fonction `barplot` produit un **diagramme en bâtons (*bar plot*)**, aussi nommé **diagramme à barres**.

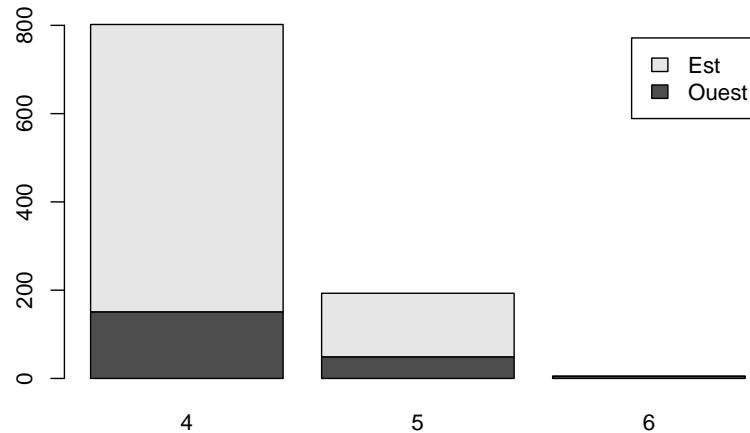
```
barplot(table(quakes$magFactor))
```



Il faut souvent utiliser `barplot` conjointement à la fonction `table` pour calculer les fréquences des observations de la variable à représenter.

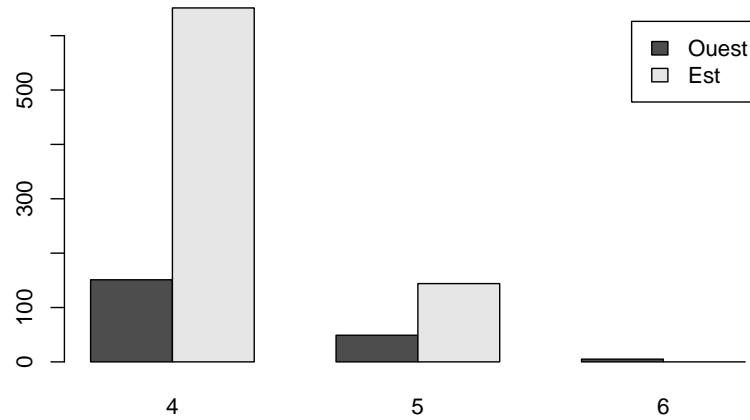
Lorsque ce diagramme est utilisé pour représenter des fréquences croisées entre deux variables, les bâtons pour les différentes variables peuvent être empilés (comportement par défaut de `barplot`) :

```
barplot(table(quakes$region, quakes$magFactor), legend.text = TRUE)
```



ou groupés (grâce à l'argument `beside = TRUE`) :

```
barplot(table(quakes$region, quakes$magFactor), legend.text = TRUE, beside = TRUE)
```



Il est important d'identifier à quelles modalités des variables sont associés les bâtons. Dans le cas d'un diagramme en bâtons bivariés, une légende est souvent nécessaire pour ce faire, car l'axe perpendiculaire aux bâtons ne suffit pas.

#### Diagramme en mosaïque - fonction `mosaicplot`

Une alternative au diagramme à bâtons empilés ou groupés pour représenter des fréquences croisées est le [diagramme en mosaïque](#), qui peut être produit avec la fonction `mosaicplot`.

```
mosaicplot(table(quakes$region, quakes$magFactor))
```

**table(quakes\$region, quakes\$magFactor)**

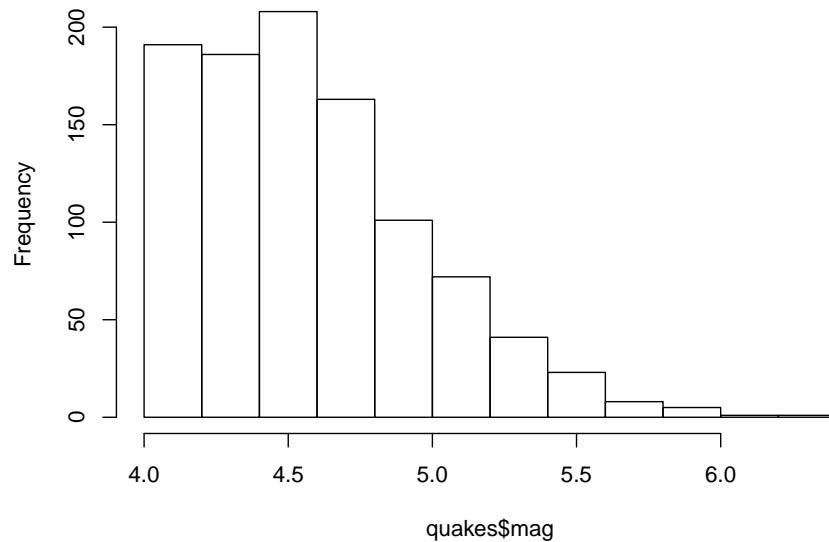


### Histogramme - fonction `hist`

L'[histogramme](#) sert à représenter la distribution empirique d'une variable numérique. La fonction [hist](#) permet d'en produire.

```
hist(quakes$mag)
```

**Histogram of quakes\$mag**



Mine de rien, cette fonction fait beaucoup de travail afin de pouvoir produire un histogramme. Elle commence par briser l'étendue des valeurs de la variable en intervalles disjoints mais contigus, puis elle calcule les fréquences des observations tombant dans les intervalles. Ensuite, elle trace un genre de diagramme en bâtons en insérant aucun espace entre les bâtons (puisque ils réfèrent à des intervalles contigus). La façon dont les intervalles sont formés peut être contrôlée par les arguments `breaks`, `include.lowest` et `right`. La fonction retourne aussi toutes les statistiques qu'elle calcule, mais il faut accompagner l'appel à la fonction `hist` d'une

assignation pour garder une trace de ces statistiques

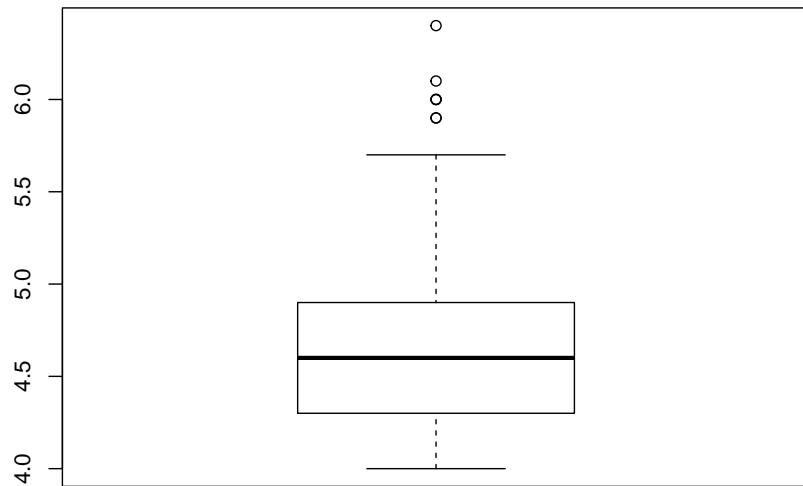
```
stats <- hist(quakes$mag, plot = FALSE)
stats

## $breaks
## [1] 4.0 4.2 4.4 4.6 4.8 5.0 5.2 5.4 5.6 5.8 6.0 6.2 6.4
##
## $counts
## [1] 191 186 208 163 101 72 41 23 8 5 1 1
##
## $density
## [1] 0.955 0.930 1.040 0.815 0.505 0.360 0.205 0.115 0.040 0.025 0.005
## [12] 0.005
##
## $mids
## [1] 4.1 4.3 4.5 4.7 4.9 5.1 5.3 5.5 5.7 5.9 6.1 6.3
##
## $xname
## [1] "quakes$mag"
##
## $equidist
## [1] TRUE
##
## attr(,"class")
## [1] "histogram"
```

### Diagramme en boîte - fonction `boxplot`

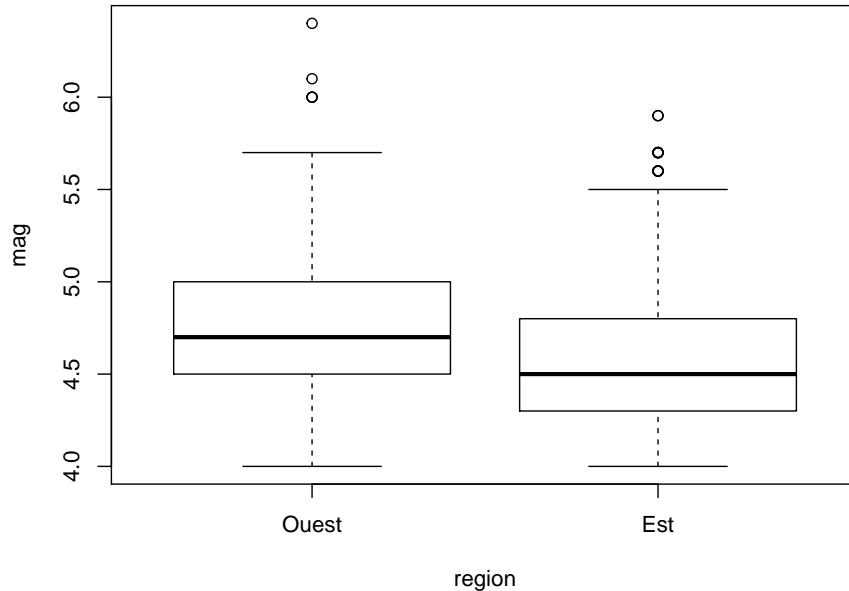
Un autre graphique conçu pour représenter la distribution empirique d'une variable numérique est le [diagramme en boîte \(`boxplot`\)](#). Ce type de graphique peut être produit en R avec la fonction `boxplot`.

```
boxplot(quakes$mag)
```



L'association entre une variable numérique et une variable catégorique peut être représentée par des diagrammes en boîtes juxtaposés.

```
boxplot(mag ~ region, data = quakes)
```

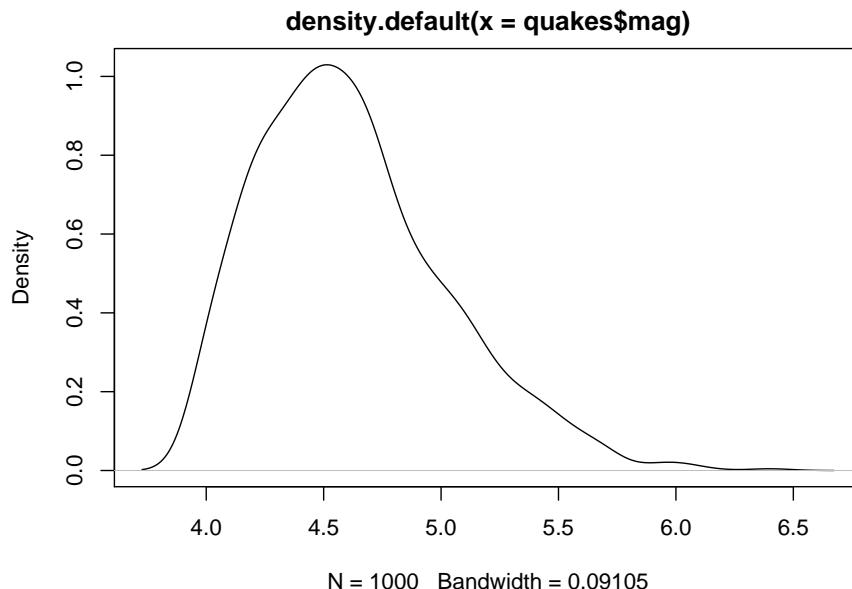


Par exemple, pour produire des diagrammes en boîtes pour les observations de la variable numérique `mag` en fonction de la variable catégorique `region`, la fonction `boxplot` peut être appelée en lui fournissant en entrée une formule contenant dans sa partie de droite la variable numérique et dans sa partie de gauche la variable catégorique.

#### Courbe d'estimation de densité à noyau - méthode `plot.density`

La fonction `density` estime une densité empirique par la [méthode à noyau](#) à partir des observations d'une variable numérique. Une représentation graphique de cette densité empirique avec la méthode `plot.density` (via la fonction générique `plot` ou directement) est une autre alternative pour visualiser les observations d'une variable numérique.

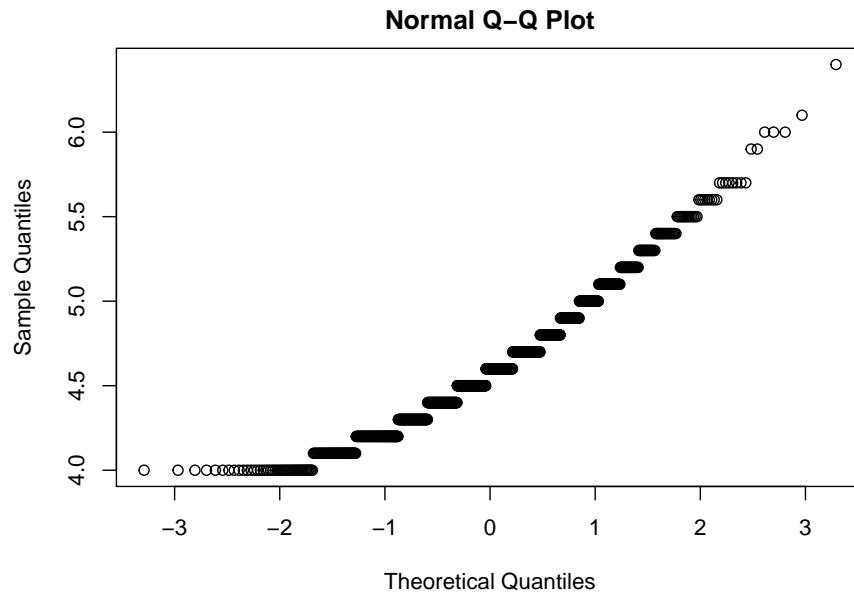
```
plot(density(quakes$mag))
```



## Diagramme quantile-quantile - fonctions `qqnorm` et `qqplot`

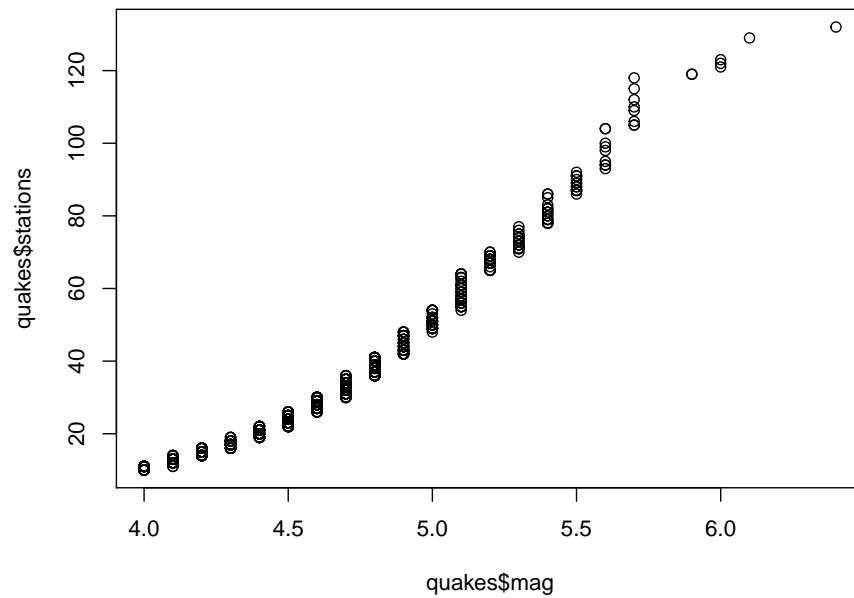
Les [diagrammes quantile-quantile](#) servent à comparer des distributions. Le plus connus des diagrammes quantile-quantile est probablement le diagramme quantile-quantile théorique normal qui permet de comparer une distribution empirique à une distribution théorique normale. Ce type de graphique peut être produit en R avec la fonction `qqnorm` :

```
qqnorm(quakes$mag)
```



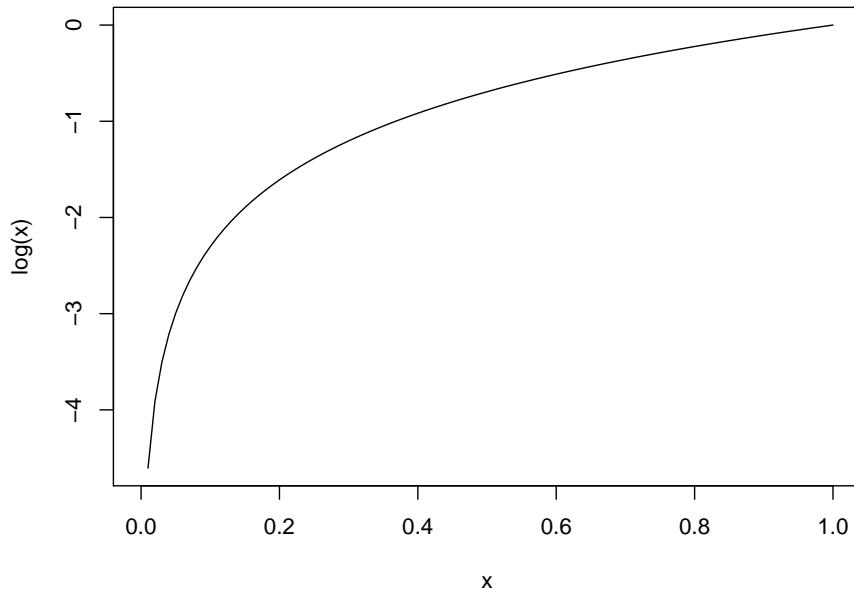
Les diagrammes quantile-quantile empiriques servent quant à eux à comparer deux distributions empiriques. La fonction `qqplot` permet d'en produire.

```
qqplot(quakes$mag, quakes$stations)
```



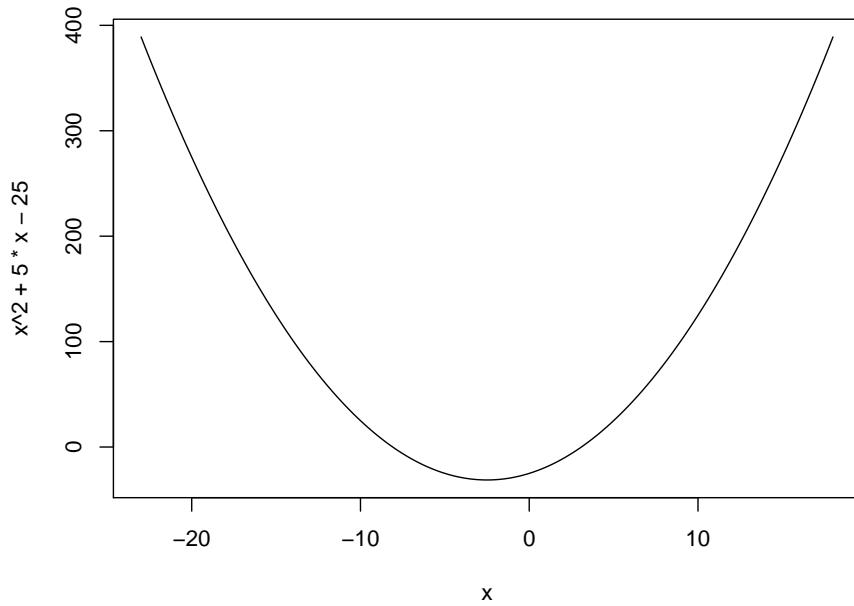
## Représentation graphique d'une expression mathématique - fonction `curve`

La représentation graphique d'une fonction mathématique est possible en R avec la fonction `curve`.  
`curve(log)`



Par défaut, la fonction `curve` fait une évaluation de la valeur de la fonction en 101 points également répartis entre 0 et 1. Il est possible de modifier ces points grâce aux arguments `from`, `to` et `n` (une séquence de points est créée avec la fonction `seq`). De plus, le premier argument n'est pas contraint à être une fonction. Il peut s'agir de n'importe quelle expression écrite comme une fonction de `x`.

`curve(x^2 + 5*x - 25, from = -23, to = 18, n = 200)`



## Arguments et paramètres graphiques

Les fonctions vues jusqu'à présent possèdent toutes des arguments pour contrôler la mise en forme et les annotations des graphiques. La liste complète de ces arguments varient d'une fonction à l'autre, mais certains arguments sont communs à presque toutes les fonctions graphiques. Nous allons voir ici les arguments les plus utiles.

Les arguments permettant de contrôler la mise en forme sont appelés paramètres graphiques. La plupart peuvent être spécifiés soit dans l'appel à la fonction graphique, soit dans un appel à la fonction `par`. Par contre, certains paramètres graphiques peuvent uniquement être fixés avec `par`.

Un paramètre graphique fourni dans un appel à une fonction graphique est effectif seulement pour le graphique produit, alors qu'un paramètre graphique fourni dans un appel à `par` reste effectif jusqu'à ce que nous le modifions de nouveau. Alors, avant de modifier des paramètres graphiques avec `par`, il est bon d'enregistrer les valeurs par défaut des paramètres, comme suit,

```
par.default <- par(no.readonly = TRUE)
```

afin de pouvoir facilement réattribuer ces valeurs par défaut aux paramètres, comme suit,

```
par(par.default)
```

à la fin des nos commandes pour produire un graphique. L'objet `par.default` est une liste contenant les valeurs par défaut de tous les paramètres graphiques modifiables. Il est aussi possible d'enregistrer seulement les paramètres graphiques modifiés par un appel à la fonction `par` en l'accompagnant d'une assignation. Par exemple, l'objet `par.default` obtenu ainsi :

```
par.default <- par(mar = c(1, 1, 2, 1))
```

contient uniquement la valeur par défaut du paramètre `mar`.

Notons que les changements apportés aux paramètres graphiques ne sont pas conservés lors de la fermeture de la session R. À l'ouverture d'une session R, les paramètres graphiques prennent donc leurs valeurs par défaut.

Voici deux tableaux résumant les arguments et paramètres graphiques les plus communs :

Arguments	Éléments graphiques contrôlés
main sub	titre note en bas de page
xlab, ylab xlim, ylim	noms des axes étendue des axes
type	type de représentation ("p" = points, "l" = lignes, "b" = les deux, etc.)

Paramètres	Éléments graphiques contrôlés
ann bty	si FALSE, retire le titre et les noms d'axes si "n", retire le cadre autour de la zone graphique (autres valeurs acceptées : "o", "l", "7", "c", "u" et "]")
lwd lty pch	épaisseur des lignes type des lignes symbole pour les points
font	type de <b>police de caractères</b> dans la zone graphique, (1 = normal, 2 = gras, 3 = italique, etc.)

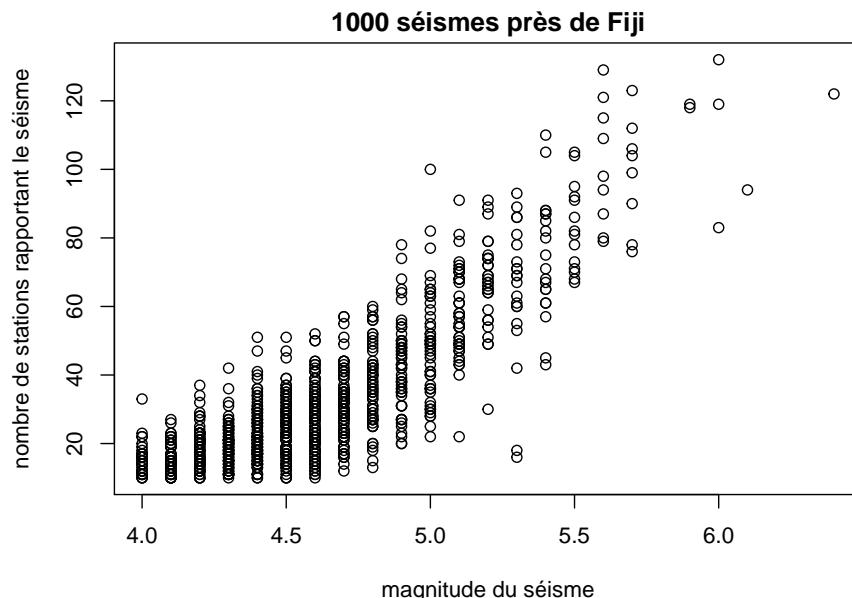
Paramètres	Éléments graphiques contrôlés
<code>font.main, font.sub, font.axis, font.lab family</code>	dans le titre, dans la note en bas de page, dans les marques des axes, dans les noms des axes famille de police de caractères ("serif", "sans", "mono", "symbol", etc.)
<code>cex cex.main, cex.sub, cex.axis, cex.lab</code>	<b>taille des caractères</b> dans la zone graphique, dans le titre, dans la note en bas de page, dans les marques des axes, dans les noms des axes
<code>col col.main, col.sub, col.axis, col.lab bg</code>	<b>couleur</b> des éléments de la zone graphique, du titre, de la note en bas de page, des marques des axes, des noms des axes <b>couleur</b> de l'intérieur du symbole pour les points lorsque <code>pch</code> prend une valeur entre 21 et 25 (pas la même signification dans <code>par</code> )
<code>par(mfrow = c( , )) par(new = TRUE) par(mar = c( , , , )) par(oma = c( , , , ))</code>	division de la fenêtre graphique superposition de graphiques tailles des marges tailles des marges externes (par défaut nulles)

Voici quelques exemples suivis d'un peu plus d'information pour certains arguments et paramètres.

### Exemple : Ajout ou modification des annotations

Ajoutons un titre et modifions les noms des axes du diagramme de dispersion de la variable `stations` de `quakes` en fonction de la variable `mag`.

```
plot(x = quakes$mag, y = quakes$stations,
      xlab = "magnitude du séisme",
      ylab = "nombre de stations rapportant le séisme",
      main = "1000 séismes près de Fiji")
```

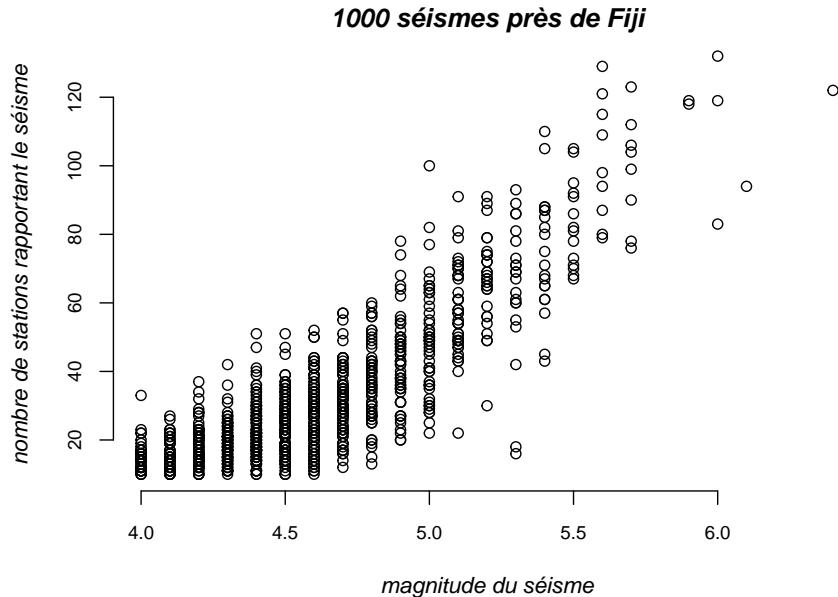


Remarque : Les accents ne causent pas de problèmes.

### Exemple : Modification de la mise en forme

Modifions maintenant quelques éléments de la mise en forme du même graphique.

```
plot(x = quakes$mag, y = quakes$stations,
      xlab = "magnitude du séisme",
      ylab = "nombre de stations rapportant le séisme",
      main = "1000 séismes près de Fiji",
      font.lab = 3,      # nom d'axes en italique
      font.main = 4,     # titre en gras italique
      cex.axis = 0.8,    # marques des axes plus petites
      bty="n")           # pas de cadre autour de la zone graphique
```



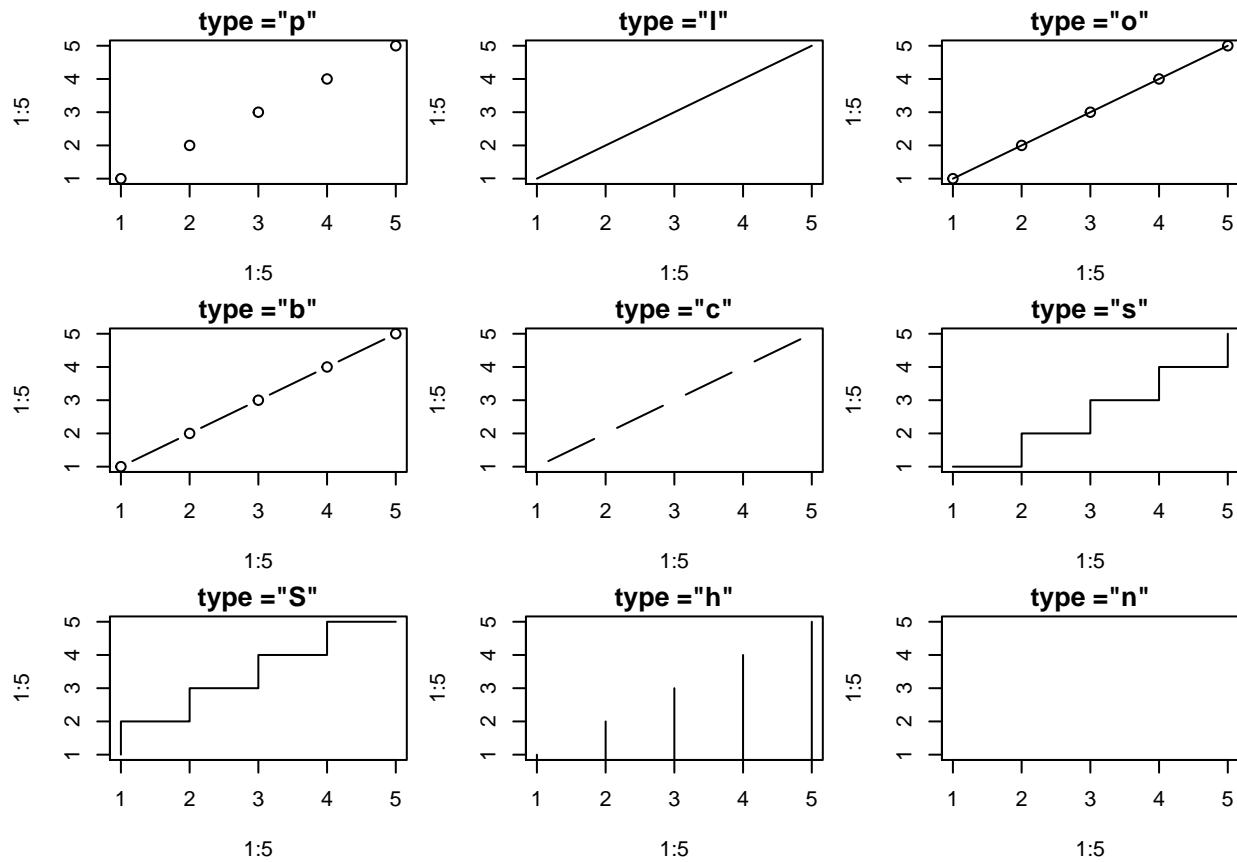
Nous pouvons aussi modifier ces paramètres graphiques par un appel à la fonction `par` comme suit.

```
par.default <- par(font.lab = 3, font.main = 4, cex.axis = 0.8, bty="n")
plot(x = quakes$mag, y = quakes$stations,
      xlab = "magnitude du séisme",
      ylab = "nombre de stations rapportant le séisme",
      main = "1000 séismes près de Fiji")
par(par.default)
```

Utiliser la fonction `par` est pratique si nous avons plusieurs graphiques à produire avec les mêmes paramètres.

## Types de représentation - argument type

Voici un graphique qui représente toutes les valeurs que peut prendre l'argument type.



Ce graphique est une adaptation d'un graphique sur la page web suivante :

<http://www.statmethods.net/graphs/line.html>.

## Types de lignes - paramètre lty

Voici un graphique qui représente toutes les valeurs que peut prendre le paramètre lty.

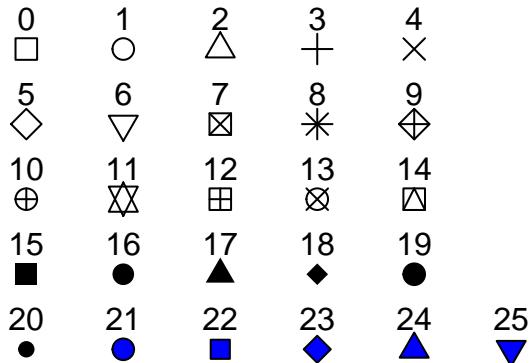
0 ou "blank"	_____
1 ou "solid"	---
2 ou "dashed"	- - -
3 ou "dotted"	....
4 ou "dotdash"	- . - .
5 ou "longdash"	- - - -
6 ou "twodash"	- - - -

Ce graphique est une adaptation d'un graphique sur la page web suivante :

<http://www.sthda.com/french/wiki/les-differentes-types-de-trait-s-dans-r-lty>

## Symboles pour les points - paramètre pch

Voici un graphique qui représente toutes les valeurs numériques que peut prendre le paramètre pch. Ce paramètre accepte aussi comme valeur un caractère quelconque.



Ce graphique est une adaptation d'un graphique sur la page web suivante :

<http://www.sthda.com/french/wiki/les-differentes-types-de-points-dans-r-comment-utiliser-pch>

## Couleurs

Une couleur peut être spécifiée de différentes façons :

- par une chaîne de caractère contenant un nom de couleur,
- par une chaîne de caractères hexadécimaux de la forme "#rrggbb" ou "#rrggbbaa",
- par un nombre entier référant à une palette de couleur.

### Nom de couleur :

La commande suivante affiche tous les noms de couleurs compris par R (sortie non affichée ici). Il y en a 657!

`colors()`

Ces couleurs sont représentées dans le graphique suivant, selon la position de la couleur dans le vecteur retourné par la commande `colors()`<sup>6</sup>. Par exemple, la couleur numéro 121 se nomme :

`colors()[121]`

`## [1] "deepskyblue"`

---

<sup>6</sup>Ce graphique a été produit grâce à un code R partagé sur la page web suivante : <http://research.stowers.org/mcm/efg/R/Color/Chart/>

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25
26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50
51	52	53	54	55	56	57	58	59	60	61	62	63	64	65	66	67	68	69	70	71	72	73	74	75
76	77	78	79	80	81	82	83	84	85	86	87	88	89	90	91	92	93	94	95	96	97	98	99	100
101	102	103	104	105	106	107	108	109	110	111	112	113	114	115	116	117	118	119	120	121	122	123	124	125
126	127	128	129	130	131	132	133	134	135	136	137	138	139	140	141	142	143	144	145	146	147	148	149	150
151	152	153	154	155	156	157	158	159	160	161	162	163	164	165	166	167	168	169	170	171	172	173	174	175
176	177	178	179	180	181	182	183	184	185	186	187	188	189	190	191	192	193	194	195	196	197	198	199	200
201	202	203	204	205	206	207	208	209	210	211	212	213	214	215	216	217	218	219	220	221	222	223	224	225
226	227	228	229	230	231	232	233	234	235	236	237	238	239	240	241	242	243	244	245	246	247	248	249	250
251	252	253	254	255	256	257	258	259	260	261	262	263	264	265	266	267	268	269	270	271	272	273	274	275
276	277	278	279	280	281	282	283	284	285	286	287	288	289	290	291	292	293	294	295	296	297	298	299	300
301	302	303	304	305	306	307	308	309	310	311	312	313	314	315	316	317	318	319	320	321	322	323	324	325
326	327	328	329	330	331	332	333	334	335	336	337	338	339	340	341	342	343	344	345	346	347	348	349	350
351	352	353	354	355	356	357	358	359	360	361	362	363	364	365	366	367	368	369	370	371	372	373	374	375
376	377	378	379	380	381	382	383	384	385	386	387	388	389	390	391	392	393	394	395	396	397	398	399	400
401	402	403	404	405	406	407	408	409	410	411	412	413	414	415	416	417	418	419	420	421	422	423	424	425
426	427	428	429	430	431	432	433	434	435	436	437	438	439	440	441	442	443	444	445	446	447	448	449	450
451	452	453	454	455	456	457	458	459	460	461	462	463	464	465	466	467	468	469	470	471	472	473	474	475
476	477	478	479	480	481	482	483	484	485	486	487	488	489	490	491	492	493	494	495	496	497	498	499	500
501	502	503	504	505	506	507	508	509	510	511	512	513	514	515	516	517	518	519	520	521	522	523	524	525
526	527	528	529	530	531	532	533	534	535	536	537	538	539	540	541	542	543	544	545	546	547	548	549	550
551	552	553	554	555	556	557	558	559	560	561	562	563	564	565	566	567	568	569	570	571	572	573	574	575
576	577	578	579	580	581	582	583	584	585	586	587	588	589	590	591	592	593	594	595	596	597	598	599	600
601	602	603	604	605	606	607	608	609	610	611	612	613	614	615	616	617	618	619	620	621	622	623	624	625
626	627	628	629	630	631	632	633	634	635	636	637	638	639	640	641	642	643	644	645	646	647	648	649	650
651	652	653	654	655	656	657																		

### Chaîne de caractères hexadécimaux de la forme "#rrggbb" ou "#rrggbbaa" :

Dans une chaîne de caractères hexadécimaux de la forme "#rrggbb" ou "#rrggbbaa", les paires des caractères rr, gg, bb et aa sont des digits hexadécimaux spécifiant une valeur entre 00 (minimum) et FF (maximum). Ces digits indiquent respectivement un niveau de rouge, de vert, de bleu et d'opacité. Le niveau d'opacité est facultatif. Par défaut les couleurs sont complètement opaques. Une valeur d'opacité minimale 00 représente une transparence complète et une valeur d'opacité maximale FF représente une opacité complète.

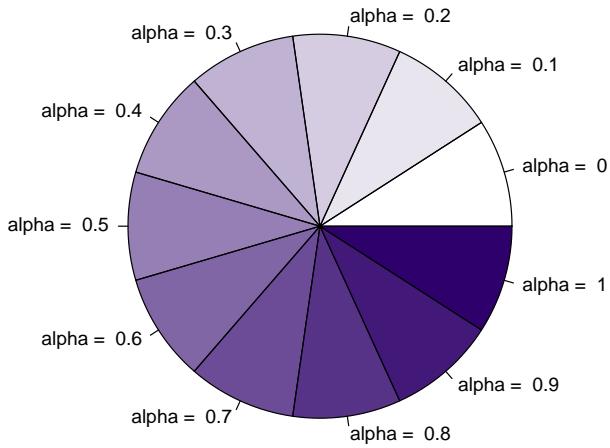
Pour déterminer un code hexadécimal de couleur, les fonctions suivantes sont utiles :

- **rgb** : prend en entrée des niveaux de rouge (**red**), de vert (**green**), de bleu (**blue**) et optionnellement d'opacité (**alpha**), retourne en sortie le code hexadécimal associé ;
- **hsv** : prend en entrée des niveaux de teinte (**hue**), de saturation (**saturation**), de valeur de luminosité (**value**) et optionnellement d'opacité, retourne en sortie le code hexadécimal associé ;
- **hcl** : prend en entrée des niveaux de teinte (**hue**), de chroma (**chroma**), de luminance (**luminance**) et optionnellement d'opacité, retourne en sortie le code hexadécimal associé ;
- **gray** ou **grey** : prend en entrée un niveau de gris entre 0 (noir) et 1 (blanc) et optionnellement d'opacité, retourne en sortie le code hexadécimal associé.

Exemple d'un indigo avec différents niveau d'opacité :

```
opacite <- rgb(red = 0.18, green = 0, blue = 0.42,
                 alpha = seq(from = 0, to = 1, by = 0.1))
opacite

## [1] "#2E006B00" "#2E006B1A" "#2E006B33" "#2E006B4D" "#2E006B66" "#2E006B80"
## [7] "#2E006B99" "#2E006BB3" "#2E006BC" "#2E006BE6" "#2E006BFF"
```



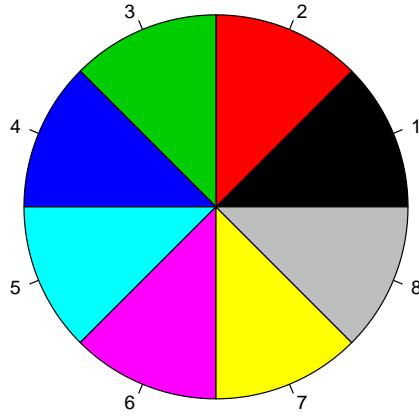
### Nombre entier référant à une palette de couleur :

Une palette de couleur est définie en R et il est possible de spécifier une couleur par un entier référant à une position dans cette palette. Par défaut, la palette de couleur est la suivante :

```
palette()
```

```
## [1] "black"     "red"       "green3"     "blue"      "cyan"      "magenta"   "yellow"    "gray"
```

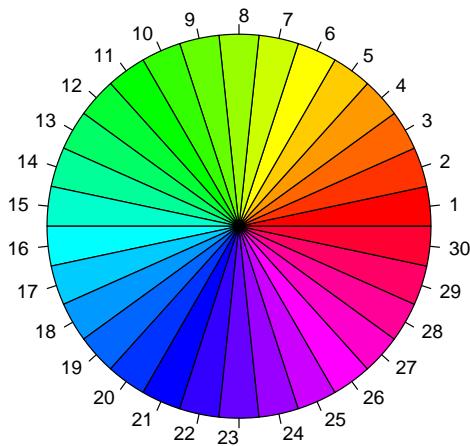
Voici un diagramme en secteur représentant les couleurs de cette palette et les entiers associés.



Il est possible de modifier la palette en lui assignant un nouveau vecteur de noms de couleurs ou de chaînes de caractères hexadécimaux de la forme "#rrggbb" ou "#rrggbbaa" pour représenter des couleurs. Certaines fonctions R sont pratiques pour créer de tels vecteurs contenant des dégradés de couleurs. Il s'agit des fonctions : `rainbow`, `heat.colors`, `terrain.colors`, `topo.colors` et `cm.colors`. Par exemple, utilisons la fonction `rainbow` pour modifier la palette de couleurs en un dégradé de 30 couleurs comme suit.

```
palette(rainbow(30))
```

Maintenant, le diagramme en secteur représentant les couleurs de la palette et les entiers associés devient :



Pour ramener la palette de couleurs à ses valeurs par défaut, il faut soumettre la commande suivante.

```
palette("default")
```

Certains packages R offrent des palettes de couleurs supplémentaires, notamment :

- le package `RColorBrewer` : <https://CRAN.R-project.org/package=RColorBrewer>, qui offre les palettes décrites sur le site web <http://colorbrewer2.org>,
- le package `viridis` ou `viridisLite` : <https://github.com/sjmgarnier/viridis>,
- d'autres packages sont présentés ici : <https://github.com/EmilHvitfeldt/r-color-palettes>.

Le package `shades` offre pour sa part différentes avenues de manipulation des couleurs en R :  
<http://hughjonesd.github.io/tweaking-colours-with-the-shades-package.html>

## Ajout d'éléments à un graphique

Nous pouvons ajouter des éléments étape par étape à un graphique en R, en appelant les unes après les autres des fonctions graphiques. Le tableau suivant présente les principales fonctions graphiques qui ajoutent des éléments à un graphique initialisé.

Fonction(s) R	Élément(s) ajouté(s)
<code>points</code> et <code>matpoints</code>	points selon des coordonnées
<code>lines</code> et <code>matlines</code>	segments de droites reliant des points
<code>abline</code>	droites traversant toute la zone graphique
<code>segments</code>	segments de droites entre des paires de coordonnées
<code>arrows</code>	flèches entre des paires de coordonnées
<code>rect</code>	rectangles
<code>polygon</code>	polygones quelconques
<code>legend</code>	légende
<code>text</code>	texte dans la zone graphique
<code>mtext</code>	texte dans la marge
<code>title</code>	titre
<code>axis</code>	axe
<code>box</code>	boîte autour de la zone graphique
<code>qqline</code>	ligne dans un graphique quantile-quantile théorique

Aussi, certaines fonctions de création de graphique peuvent devenir des fonctions d'ajout d'éléments à un

graphique grâce à l'argument `add`. C'est le cas notamment des fonctions suivantes :

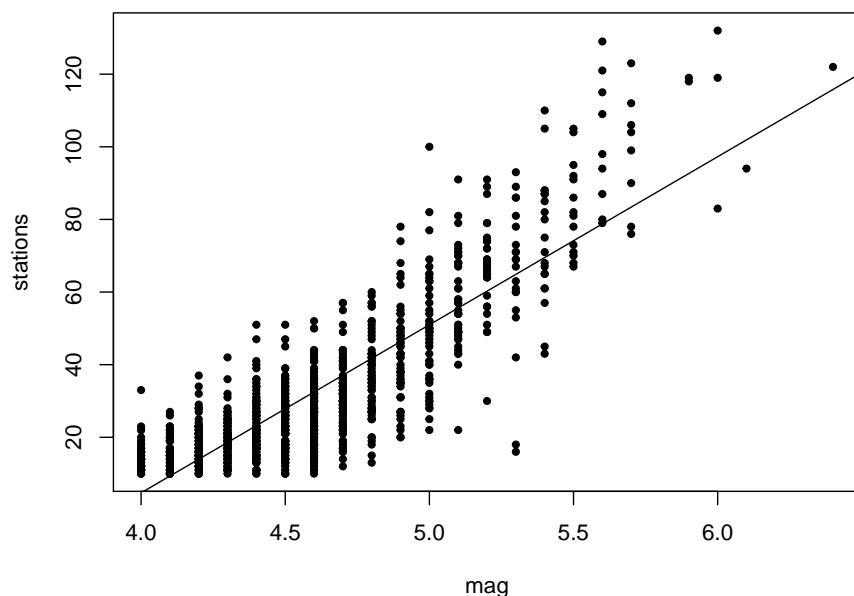
`matplotlib`, `barplot`, `hist`, `boxplot` et `curve`.

En donnant la valeur `TRUE` à l'argument `add` lors de l'appel de ces fonctions, elle ajoute des éléments au graphique de la fenêtre graphique active au lieu de créer un nouveau graphique.

### Exemple : Ajout de lignes

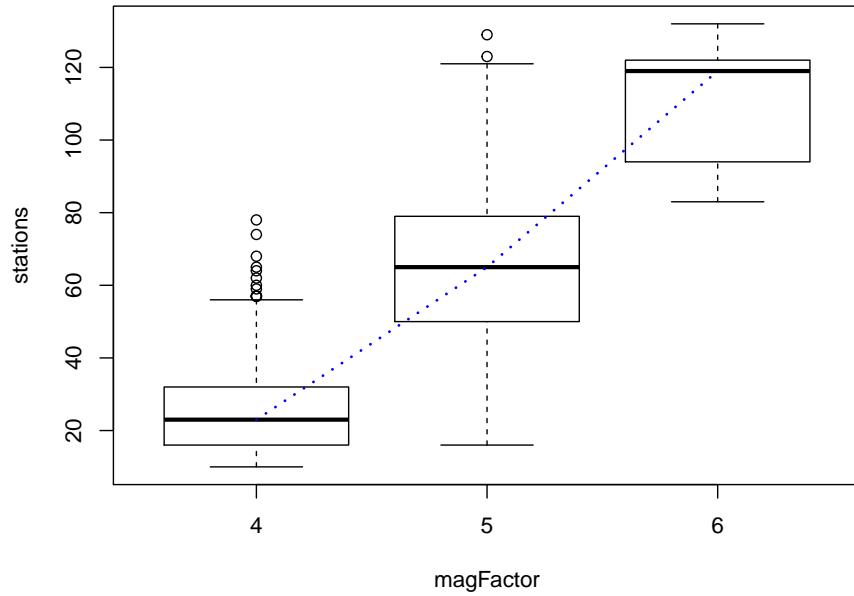
La fonction `abline` permet d'ajouter une droite à un graphique qui traverse toute la zone graphique.

```
lmout <- lm(stations ~ mag, data = quakes) # ajustement d'une droite de régression  
  
plot(stations ~ mag, data = quakes, pch = 20)  
abline(lmout)
```



Avec la fonction `lines`, nous pouvons ajouter des segments de droites reliant des points pour lesquels les coordonnées sont fournies.

```
aggout <- aggregate(stations ~ magFactor, data = quakes, FUN = median)  
aggout$magFactor <- as.numeric(aggout$magFactor)  
  
boxplot(stations ~ magFactor, data = quakes)  
lines(aggout, lty = 3, lwd = 2, col = 4)
```



### Exemple : Ajout de courbes à un histogramme

```

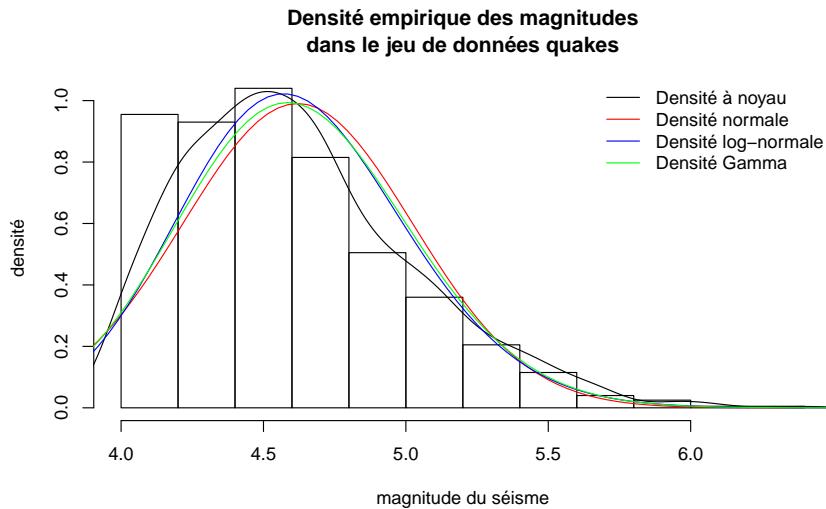
# Initialisation de l'histogramme
hist(quakes$mag, xlab = "magnitude du séisme", ylab = "densité", freq = FALSE,
      main = "Densité empirique des magnitudes\ndans le jeu de données quakes")

# Ajout de la courbe de densité à noyau
lines(density(quakes$mag), xlim = range(quakes$mag))

# Ajout de courbes de densité théoriques avec paramètres estimés à partir des données
moy <- mean(quakes$mag)
et <- sd(quakes$mag)
curve(dnorm(x, mean = moy, sd = et),
      add = TRUE, col = "red", xlim = c(3.5, 7))
curve(dlnorm(x, meanlog = mean(log(quakes$mag)), sdlog = sd(log(quakes$mag))),
      add = TRUE, col = "blue", xlim = c(3.5, 7))
curve(dgamma(x, shape = moy^2/et^2, rate = moy/et^2),
      add = TRUE, col = "green", xlim = c(3.5, 7))

# Ajout d'une légende
legend(x = "topright", lty = 1, bty = "n",
       col = c("black", "red", "blue", "green"),
       legend = c("Densité à noyau", "Densité normale",
                 "Densité log-normale", "Densité Gamma"))

```



## Possibilités graphiques spécifiques

### Annotations mathématiques

Des annotations mathématiques sont des caractères spéciaux communs en science, telle que des exposants, des indices, des fractions, des lettres grecques, etc. Tout élément textuel d'un graphique peut en contenir (p. ex. les valeurs données à un argument `main`, `xlab`, `ylab`, `labels`, `legend`, `text`, etc.). La fiche d'aide intitulée `plotmat` énumère toutes les annotations possibles. Ces annotations impliquent une syntaxe particulière, qui ressemble un peu à du LaTeX. Elles doivent être exprimée sous la forme d'une expression R. La fonction `expression` permet de créer une telle expression.

### Exemple :

Dans le graphique précédent, utilisons des notations mathématiques dans la légende.

```
# Initialisation de l'histogramme
hist(quakes$mag, xlab = "magnitude du séisme", ylab = "densité", freq = FALSE,
      main = "Densité empirique des magnitudes\nndans le jeu de données quakes")

# Ajout de la courbe de densité à noyau
lines(density(quakes$mag), xlim = range(quakes$mag))

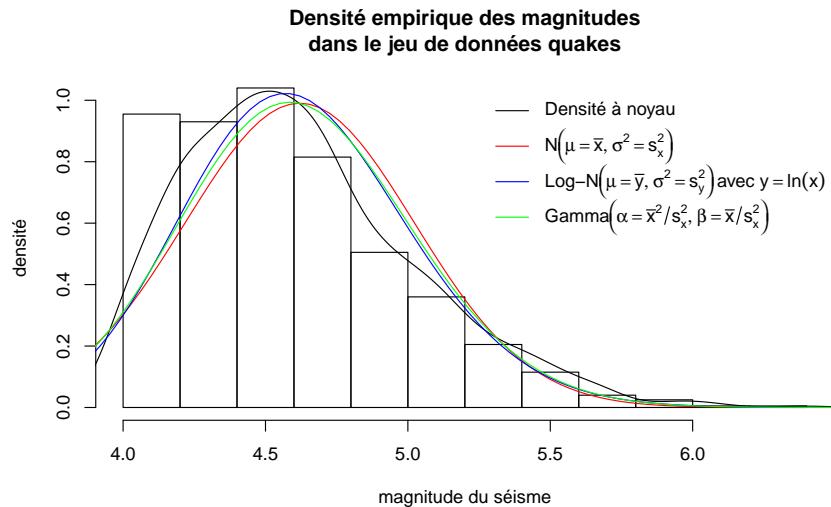
# Ajout de courbes de densité théoriques avec paramètres estimés à partir des données
moy <- mean(quakes$mag)
et <- sd(quakes$mag)
curve(dnorm(x, mean = moy, sd = et),
      add = TRUE, col = "red", xlim = c(3.5, 7))
curve(dlnorm(x, meanlog = mean(log(quakes$mag)), sdlog = sd(log(quakes$mag))),
      add = TRUE, col = "blue", xlim = c(3.5, 7))
curve(dgamma(x, shape = moy^2/et^2, rate = moy/et^2),
      add = TRUE, col = "green", xlim = c(3.5, 7))

# Ajout d'une légende
legend(
  x = "topright", lty = 1, bty = "n",
  col = c("black", "red", "blue", "green"),
  legend = c(
```

```

"Densité à noyau",
expression(
  paste(
    "N",
    bgroup("(",list(mu == bar(x), sigma^2 == s[x]^2),")")
  )
),
expression(
  paste(
    "Log-N",
    bgroup("(",list(mu == bar(y), sigma^2 == s[y]^2),")"),
    " avec ",
    y == ln(x)
  )
),
expression(
  paste(
    "Gamma",
    bgroup("(",list(alpha == bar(x)^2/s[x]^2, beta == bar(x)/s[x]^2),")")
  )
)
)

```



Dans ce code, les trois derniers éléments du vecteur fourni à l'argument `legend` de la fonction du même nom sont des expressions. Ces éléments sont créés par des appels à la fonction `expression`. Penchons-nous sur un de ces appels pour tenter de mieux le comprendre.

```
expression(
  paste(
    "Log-N",
    bgroup("(", list(mu == bar(y), sigma^2 == s[y]^2), ")"),
    " avec ",
    y == ln(x)
  )
)
```

L'appel à la fonction `expression` contient un appel à la fonction `paste`. Nous sommes contraints à cette syntaxe parce que le libellé doit contenir à la fois des chaînes de caractères ordinaires (ici "Log-N" et " avec ") et des expressions mathématiques. Mais la fonction `paste` n'agit pas ici tout à fait comme d'habitude. Dans des expressions mathématiques, `paste` ne travaille pas de façon vectorielle et ne possède pas d'arguments. La fonction ne fait que juxtaposer des expressions.

La première expression mathématique dans le `paste` précédent est `bgroup("(",list(mu == bar(y), sigma^2 == s[y]^2),")")`. Cette expression permet de créer l'annotation ressemblant à  $(\mu = \bar{y}, \sigma^2 = s_y^2)$ . Certains caractères de cet élément ont une signification particulière, par exemple :

- `bgroup("(", x, ")")` permet d'encadrer des éléments de parenthèses dont la hauteur s'adapte à la hauteur des éléments ;
- `list(x, y)` sépare des éléments par une virgule (pas la même signification que la fonction `list`)
- `mu` devient la lettre grecque  $\mu$ ,
- `==` devient un signe d'égalité,
- `^2` devient un exposant,
- `[y]` devient un indice,
- etc.

#### Autres exemples :

- <https://andyphilips.github.io/blog/2017/08/16/mathematical-symbols-in-r-plots.html>

#### Autres options :

Certains packages permettent de carrément mettre des équations LaTeX dans les annotations de graphiques R, notamment les deux packages suivants.

- <https://CRAN.R-project.org/package=latex2exp>
- <https://CRAN.R-project.org/package=tikzDevice>

#### Plusieurs graphiques dans une même fenêtre

Il est possible de diviser la fenêtre graphique en sous-fenêtres. Les deux principaux outils offerts dans le système de base pour effectuer cette division sont les suivants :

- arguments `mfrow` ou `mfcol` de la fonction `par` : produit une grille de sous-fenêtres de tailles égales ;
- fonction `layout` : permet de contrôler les dimensions des sous-fenêtres.

#### Exemple : Sous-fenêtres de tailles égales - paramètre `mfrow` ou `mfcol` de `par`.

Réunissons plusieurs représentations de la variable `mag` du jeu de données `quakes` dans un seul graphique.

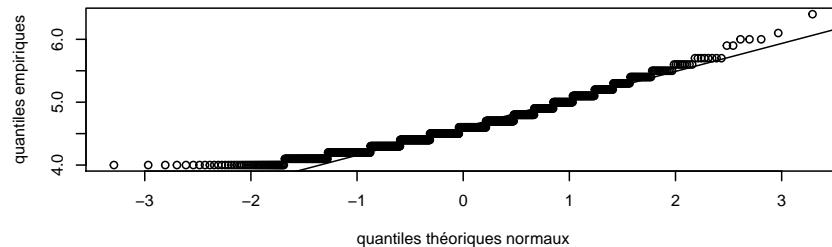
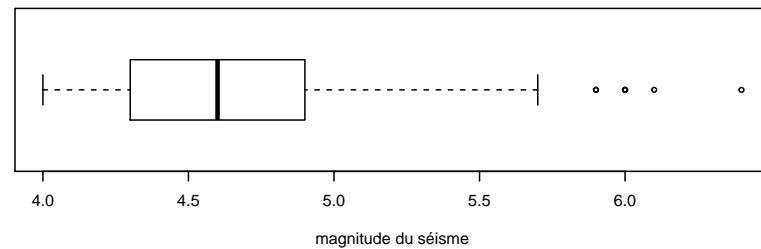
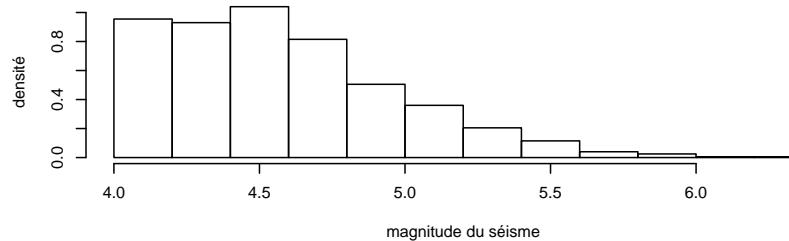
```
# Modification du paramètre graphique mfrow de façon à diviser la
# fenêtre graphique en 3 sous-fenêtres, une en dessous de l'autre
par.default <- par(mfrow = c(3,1))

# Graphique dans la première sous-fenêtre
hist(quakes$mag, xlab = "magnitude du séisme", ylab = "densité", freq = FALSE,
     main = "")

# Graphique dans la deuxième sous-fenêtre
boxplot(quakes$mag, horizontal = TRUE, xlab = "magnitude du séisme")

# Graphique dans la troisième sous-fenêtre
qqnorm(quakes$mag, xlab = "quantiles théoriques normaux",
       ylab = "quantiles empiriques", main = "")
qqline(quakes$mag)
```

```
# Réattribution des valeurs par défaut aux paramètres graphiques
par(par.default)
```



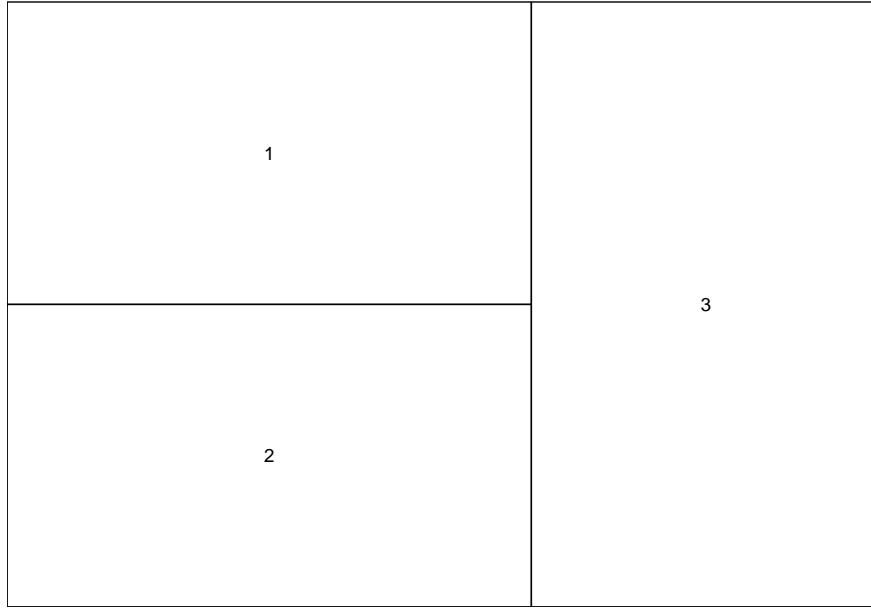
### Exemple : Sous-fenêtres de tailles inégales - fonction layout

La fonction `layout` découpe la fenêtre graphique en quadrillage. Les hauteurs et largeurs des bandes sont déterminées par les arguments `widths` et `heights`. La matrice à donner en argument `mat` contient un entier non négatif pour chaque sous-fenêtre créée par le quadrillage. Dans les commandes qui suivent l'appel à la fonction `layout`, le premier graphique produit est affiché dans la sous-fenêtre associée au chiffre 1, le deuxième dans la sous-fenêtre associée au chiffre 2 et ainsi de suite. L'attribution du chiffre zéro à une sous-fenêtre signifie qu'aucun graphique ne doit être affiché dans cette sous-fenêtre. L'attribution d'un même entier positif à plusieurs sous-fenêtres adjacentes permet de les fusionner en une seule sous-fenêtre.

La fonction `layout.show` permet de visualiser les sous-fenêtres créées.

Voici un exemple de découpage en sous-fenêtre avec `layout`.

```
layout(matrix(c(1,2,3,3), nrow = 2), widths = c(3,2))
layout.show(n = 3)
```



Maintenant, reprenons l'exemple précédent et répartissons les graphiques dans les sous-fenêtre de la disposition que nous venons de créer avec `layout`.

Comme pour les paramètres modifiés avec la fonction `par`, le découpage est effectif tant qu'il n'est pas modifié de nouveau avec `layout` ou jusqu'à la fin de la session R. Une bonne pratique est donc de redonner à `layout` sa valeur par défaut à la fin du code de création du graphique avec l'instruction `layout(matrix(1))`.

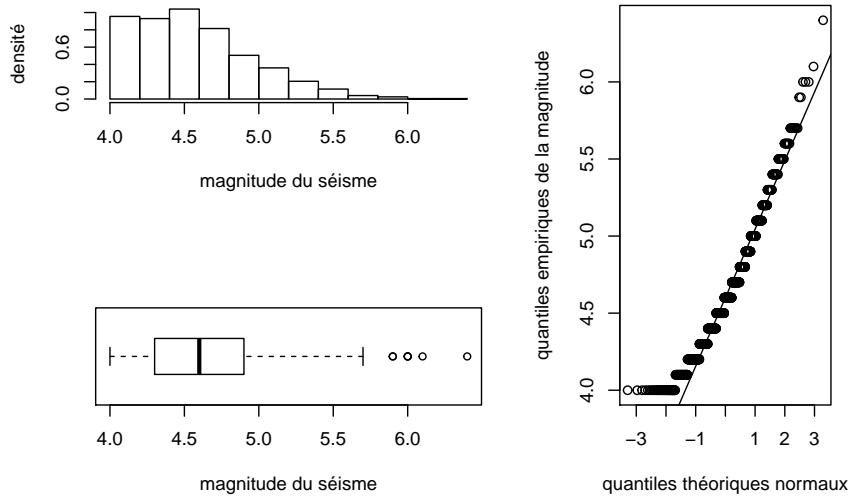
```
# Configuration des sous-fenêtres
layout(matrix(c(1,2,3,3), nrow = 2), widths = c(3,2))

# Graphique dans la première sous-fenêtre
hist(quakes$mag, xlab = "magnitude du séisme", ylab = "densité", freq = FALSE,
     main = "")

# Graphique dans la deuxième sous-fenêtre
boxplot(quakes$mag, horizontal = TRUE, xlab = "magnitude du séisme")

# Graphique dans la troisième sous-fenêtre
qqnorm(quakes$mag, xlab = "quantiles théoriques normaux",
       ylab = "quantiles empiriques de la magnitude", main = "")
qqline(quakes$mag)

# Réattribution de la valeur par défaut de layout
layout(matrix(1))
```



Nous pourrions améliorer ce graphique en ajustant les marges. Nous pourrions également mettre des sous-titres et un titre global.

```
# Enregistrement des valeurs par défaut aux paramètres graphiques
par.default <- par(no.readonly = TRUE)

# Configuration des sous-fenêtres
layout(matrix(c(1,2,3,3), nrow = 2), widths = c(3,2))

# Pour créer un espace pour le titre global
par(oma = c(0, 0, 3, 0))

# Graphique dans la première sous-fenêtre
par(mar = c(3.1, 4.1, 2.1, 2.1), cex.main = 1)
hist(quakes$mag, ylab = "densité", freq = FALSE, main = "Histogramme")

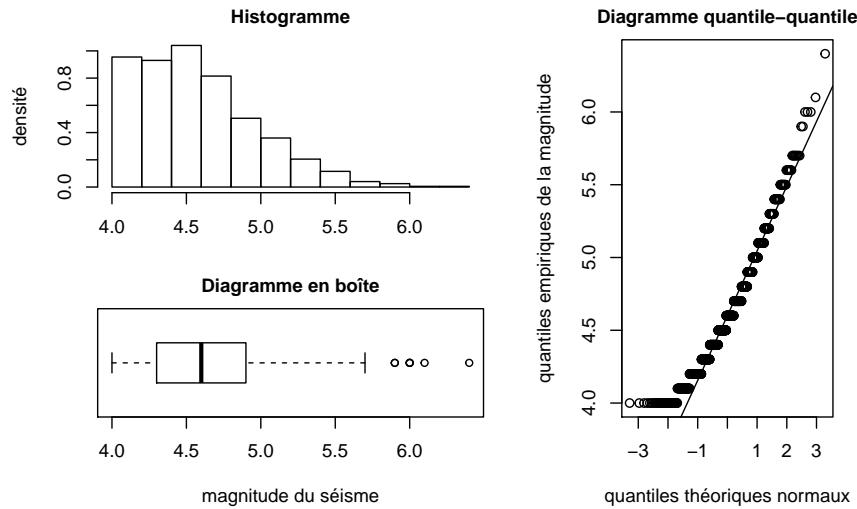
# Graphique dans la deuxième sous-fenêtre
par(mar = c(5.1, 4.1, 2.1, 2.1))
boxplot(quakes$mag, horizontal = TRUE, xlab = "magnitude du séisme",
        main = "Diagramme en boîte")

# Graphique dans la troisième sous-fenêtre
qqnorm(quakes$mag, xlab = "quantiles théoriques normaux",
       ylab = "quantiles empiriques de la magnitude",
       main = "Diagramme quantile-quantile")
qqline(quakes$mag)

# Ajout du titre global
mtext("Densité empirique des magnitudes dans le jeu de données quakes",
      outer = TRUE, cex = 1.2, line = 1.5)

# Réattribution des valeurs par défaut aux paramètres graphiques
par(par.default)
layout(matrix(1))
```

## Densité empirique des magnitudes dans le jeu de données quakes



Autres exemples d'utilisation de layout :

- <http://www.statmethods.net/advgraphs/layout.html>
- [http://sas-and-r.blogspot.ca/2012/09/example-103-enhanced-scatterplot-with.html?utm\\_source=feedburner&utm\\_medium=feed&utm\\_campaign=Feed%3A+SASandR+\(SAS+and+R\)&utm\\_content=Google+Reader](http://sas-and-r.blogspot.ca/2012/09/example-103-enhanced-scatterplot-with.html?utm_source=feedburner&utm_medium=feed&utm_campaign=Feed%3A+SASandR+(SAS+and+R)&utm_content=Google+Reader)

## Aspects techniques

Créer un graphique en R signifie soumettre un programme R qui génère le graphique. Pour produire de nouveau le même graphique, il suffit de conserver le programme permettant de générer le graphique et de le soumettre de nouveau dans la console. Le fait de produire le graphique par un programme plutôt que par un menu dans lequel nous sélectionnons des options permet d'automatiser le travail et de sauver du temps lorsque nous avons à produire plusieurs graphiques similaires.

### Fenêtres graphiques

En R, les graphiques sont par défaut tous créés dans la même fenêtre. RStudio conserve un historique des graphiques, ce qui permet de réafficher un graphique produit antérieurement, mais au cours d'une même session R.

Nous pouvons aussi choisir d'ouvrir une nouvelle fenêtre graphique avec la commande `dev.new()` ou avec

- `windows()` sur Windows,
- `quartz()` sur Mac OS X / OS X / macOS,
- `X11()` sur Unix / Linux.

Les dimensions de la fenêtre graphique courante sont les suivantes

```
par("din")
```

```
# ou  
dev.size()
```

```
## [1] 6.5 4.5
```

Note : Ce paramètre graphique n'est pas modifiable (R.O. = *Read Only* dans la documentation de la fonction `par`).

Comment modifier les dimensions de la fenêtre graphique ?

1. Redimensionner la fenêtre graphique avec la souris. En Windows, dans une fenêtre ouverte avec `windows()`, les modes de redimensionnement sont les suivants :
  - mode R : dimensions fenêtres = dimensions du graphique ;
  - ajuster à la fenêtre : dimensions du graphique ajustées aux dimensions de la fenêtre, mais ratio largeur/hauteur conservé ;
  - taille fixe : dimensions du graphique fixes, donc non affectées par les dimensions de la fenêtre.
2. Ouvrir une nouvelle fenêtre avec X11 (Unix / Linux), `windows` (Windows) ou `quartz` (Mac) et spécifier ses dimensions avec les arguments `width` et `height`. Exemple :

```
windows(width = 10, height = 7.5)
```

Commandes utiles :

- `dev.list()` : pour voir la liste de toutes les fenêtres graphiques ouvertes,
- `dev.cur()` : pour connaître la fenêtre graphique courante,
- `dev.off()` : pour fermer la fenêtre graphique courante (peut aussi être fermée avec la souris).

Pour d'informations sont offertes dans la fiche d'aide R nommée `dev`.

### Enregistrer un graphique

En RStudio, nous pouvons enregistrer un graphique dans la fenêtre graphique par le menu Export. Nous pouvons aussi le faire, comme dans l'exemple suivant, avec les fonctions `bmp`, `postscript`, `pdf`, `png`, `tiff`, `svg` ou `jpeg`, selon le format désiré.

```
# Ouverture de la connexion avec un fichier
png("test.png")

# Code pour créer le graphique
plot(x = quakes$mag, y = quakes$stations)
# potentiellement plusieurs lignes de code ici

# Fermeture de la connexion avec le fichier
dev.off()
```

En fait, ces fonctions redirigent l'affichage de graphiques vers un fichier. L'instruction `dev.off()` est nécessaire pour mettre un terme à la communication entre R et le fichier ouvert.

### Autres fonctions et trucs utiles

À titre de référence, voici une liste de quelques autres fonctions utiles avec des graphiques produits à l'aide du système de base en R.

- `locator` : pour identifier avec la souris une position ;
- `identify` : pour identifier avec la souris des observations ;
- `clip` : pour restreindre la zone d'ajouts dans un graphique ;
- `jitter` : pour ajouter un peu de bruits à des valeurs numériques, ce qui peut permettre de visualiser plus facilement des observations superposées.

Aussi, il est bon de savoir que plusieurs fonctions graphiques de base peuvent, tout comme `hist`, retourner des valeurs en plus de créer un graphique. Les valeurs retournées contiennent, par exemple :

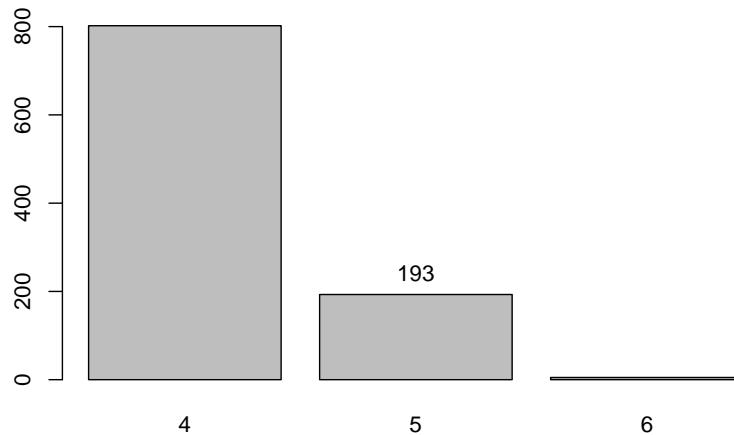
- les coordonnées de certains éléments dans le graphique, par exemple :
  - dans un histogramme produit avec `hist` : les coordonnées des limites et des centres des bâtons (en d'autres mots des intervalles),
  - dans un diagramme en bâtons produit avec `barplot` : les coordonnées centrales des bâtons sur l'axe de la variable catégorique ;

- etc. ;
- des statistiques calculées pour produire le graphique, par exemple :
  - dans un histogramme produit avec `hist` : les fréquences des observations dans les intervalles,
  - dans des diagrammes en boîte produits par la fonction `boxplot` : les statistiques représentées dans les diagrammes,
  - etc.

**Exemple d'utilisation des valeurs rentrées par la fonction `barplot`.**

```
# Calcul des fréquences
freq <- table(quakes$magFactor)
# Production du diagramme en bâtons
out_barplot <- barplot(freq)
out_barplot

##      [,1]
## [1,]  0.7
## [2,]  1.9
## [3,]  3.1
# Ajout de texte au-dessus du deuxième bâton
text(x = out_barplot[2, 1], # coordonnée en x du centre du bâton tirée de out_barplot
      y = freq[2],
      labels = freq[2],
      pos = 3)
```



## Résumé et point de vue

La fonction générique `plot` choisit un bon type de graphique à produire selon le **nombre** et la **nature des variables**.

Sinon, les fonctions graphiques de base en R font très peu de choix pour l'utilisateur.

Les graphiques initialisés sont typiquement minimalistes. Nous y ajoutons ce que nous voulons. Et les possibilités sont presque infinies !

- Avantage : Nous avons le plein contrôle sur l'apparence du graphique.

- Désavantage : Il faut parfois travailler fort pour arriver à nos fins.

### Faiblesse du système graphique de base :

Les représentations multivariables ne sont pas simples à produire. Il n'est pas toujours facile, avec les fonctions graphiques de base, de représenter trois variables et plus sur le même graphique.

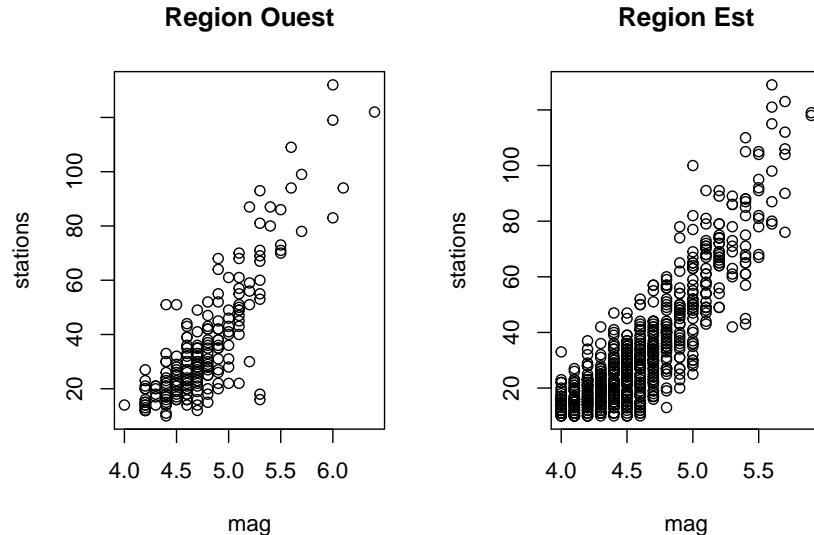
### Exemple - Représenter deux variables numériques et une variable catégorique

Tentons de représenter la relation entre les variables numériques `stations` et `mag` de `quakes` en fonction de la variable catégorique `region`.

#### Option 1 : Sous-fenêtres selon la variable catégorique

Graphiques par niveau du facteur placés côte-à-côte.

```
par.default <- par(mfrow = c(1,2))
plot(stations ~ mag, data = quakes, subset = region == "Ouest", main = "Region Ouest")
plot(stations ~ mag, data = quakes, subset = region == "Est", main = "Region Est")
par(par.default)
```



Il faudrait idéalement :

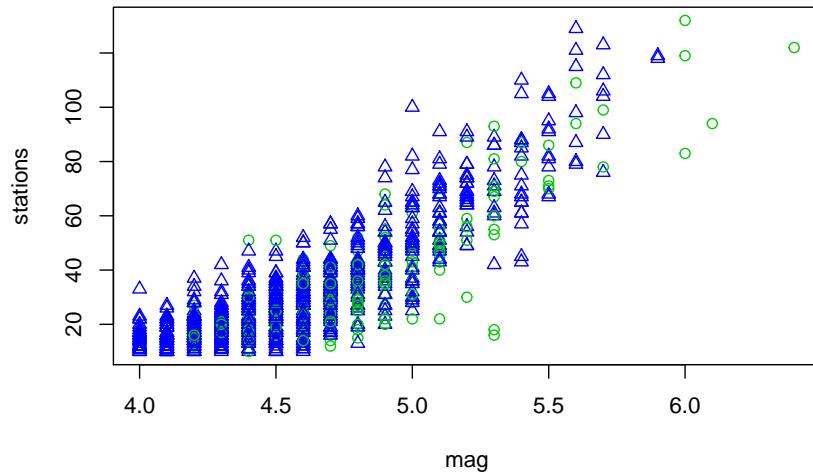
- ajuster les axes afin que tous les graphiques couvrent la même région,
- éviter la redondance dans les noms des axes.

#### Option 2 : Superposition avec aspect distinctif selon la variable catégorique

Faisons varier le symbole utilisé pour les points et sa couleur selon le niveau du facteur `region`.

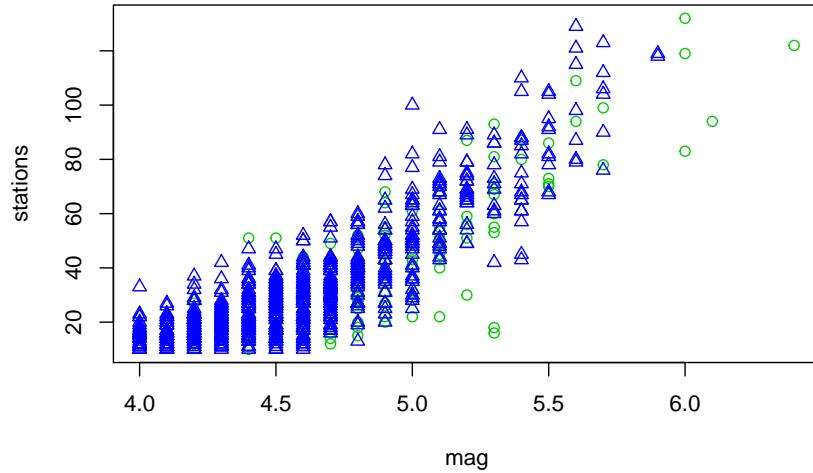
Nous pouvons donner aux arguments graphiques un vecteur de valeurs, plutôt qu'une seule valeur. Ce vecteur doit être de même longueur que le nombre d'observations à représenter (sinon il sera recyclé ou tronqué).

```
plot(stations ~ mag, data = quakes,
      pch = as.numeric(region), col = as.numeric(region) + 2)
```



Nous pouvons aussi ajouter les points pour chaque groupe dans des étapes séparées.

```
# Initialisation d'un graphique vide, mais avec les bonnes étendues
plot(stations ~ mag, data = quakes, type = "n")
# Ajout des points pour la région Ouest
points(stations ~ mag, data = quakes, subset = region == "Ouest", pch = 1, col = 3)
# Ajout des points pour la région Est
points(stations ~ mag, data = quakes, subset = region == "Est", pch = 2, col = 4)
```

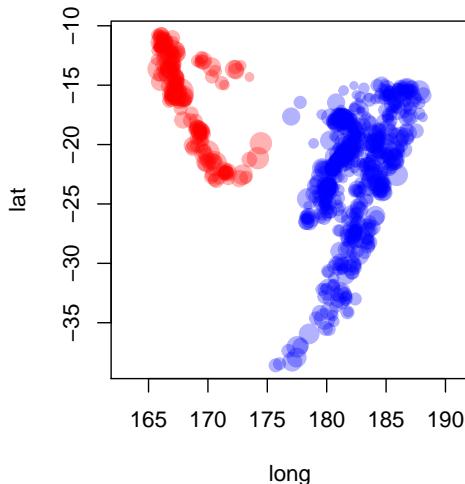


Il faudrait ajouter une légende pour identifier à quoi réfère les couleurs et les symboles.

### Exemple - Représenter trois variables numériques et une variable catégorique

Tentons maintenant de représenter dans un même graphique les variables suivantes de `quakes` : `lat`, `long`, `mag` et `region`.

```
par.default <- par(pty = "s")
# pty = "s" permet d'avoir une région graphique carrée
plot(lat ~ long, data = quakes, asp = 1,
      # asp = 1 permet d'avoir des axes sur la même échelle,
      # ce qui est préférable ici, car les variables sont des coordonnées géographiques
      cex = 3*(mag - min(mag))/(max(mag) - min(mag)) + 1, pch = 20,
      # la taille du symbole dépend de la valeur de la variable mag
      col = ifelse(region == "Ouest", rgb(1, 0, 0, 0.3), rgb(0, 0, 1, 0.3)))
      # la couleur du symbole dépend du facteur région
      # des couleurs transparentes sont utilisées
par(par.default)
```



Il faudrait ajouter une légende pour identifier à quoi réfère les couleurs différents (`region`) et les tailles différentes (`mag`).

Nous avons réussi à créer des représentations multivariables avec le système graphique de base, mais les autres systèmes graphiques en R permettent de créer ce genre de graphiques plus facilement.

---

## Autres systèmes graphiques en R

Deux packages offrent une alternative aux fonctions graphiques de base. Il s'agit des packages `lattice` et `ggplot2`. Ils sont présentés dans cette section.

Tout d'abord, voici une comparaison rapide des différents systèmes graphiques en R.

1. Système graphique de base (packages `graphics` et `grDevices`)
  - fait partie de l'installation de base de R depuis le début  
(mais a été amélioré au fil du temps).
2. Package `lattice`
  - publié pour la première fois en 2001 ;

- développé par Deepayan Sarkar ;
- basé sur le « trellis » système de S-PLUS, qui implémente le système graphique présenté dans : Cleveland, W. S. (1993). *Visualizing data*. Hobart Press.

### 3. Package `ggplot2`

- publié pour la première fois en 2006 sous le nom de `ggplot` ;
- amélioré de façon importante et renommé `ggplot2` en 2007 ;
- créé par Hadley Wickham et maintenant développé par une équipe de plusieurs personnes ;
- basé sur le système graphique présenté dans : Wilkinson, L. (2005). *The grammar of graphics*, Second Edition. Springer.

**Note :** À partir d'ici, le texte est plutôt minimaliste. Le matériel est tiré d'une présentation que j'ai donné. Je n'ai pas eu le temps d'enrober les informations de phrases complètes.

## Package `lattice`

Force = graphiques conditionnels à la valeur d'un ou de plusieurs facteurs

Caractéristiques :

- Les fonctions prennent idéalement une formule en entrée.
  - L'opérateur `|` sert à créer des panneaux (*panel*) de graphiques côte-à-côte, \* chacun des sous-graphique est conditionnel à la valeur de facteur(s), il représente donc seulement le sous-ensemble des observations ayant une modalité particulière pour ce(s) facteur(s).
- L'argument `groups` sert à superposer des éléments avec un aspect (couleur, forme, taille, etc.) qui varie selon les niveaux d'un facteur.
- Un graphique est créé par un seul appel à une fonction du package.
  - Les fonctions ont beaucoup de paramètres.
  - Il n'est pas possible d'ajouter des éléments à un graphique comme dans le système de base.

Semble avoir été quelque peu éclipsé par `ggplot2`.

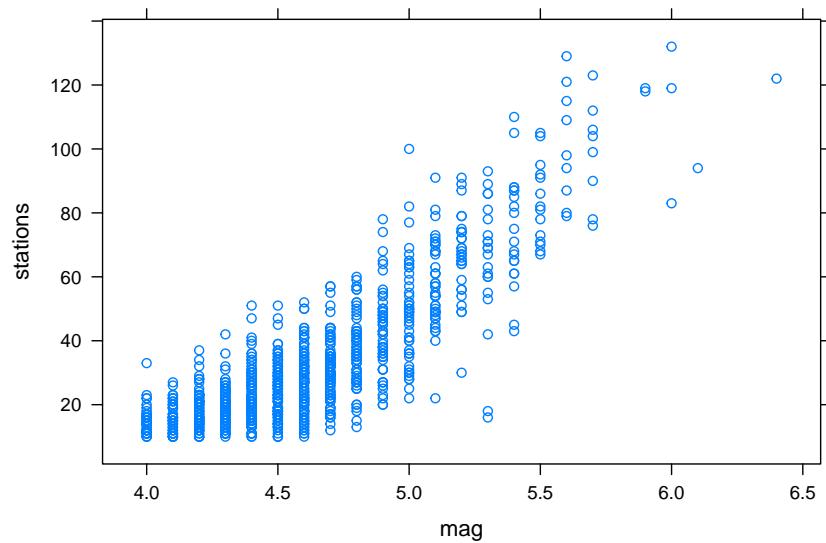
Je ne présente que quelques exemples, avec la fonction `xyplot` servant à créer des diagrammes de dispersion.

### Exemples : Diagrammes de dispersion - fonction `xyplot`

```
# Chargement du package
library(lattice)
```

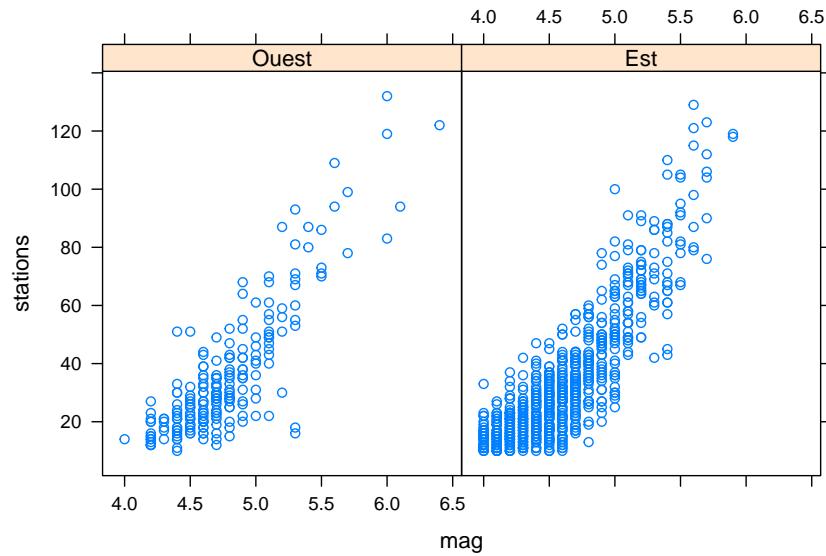
### Diagramme de dispersion

```
xyplot(stations ~ mag, data = quakes)
```



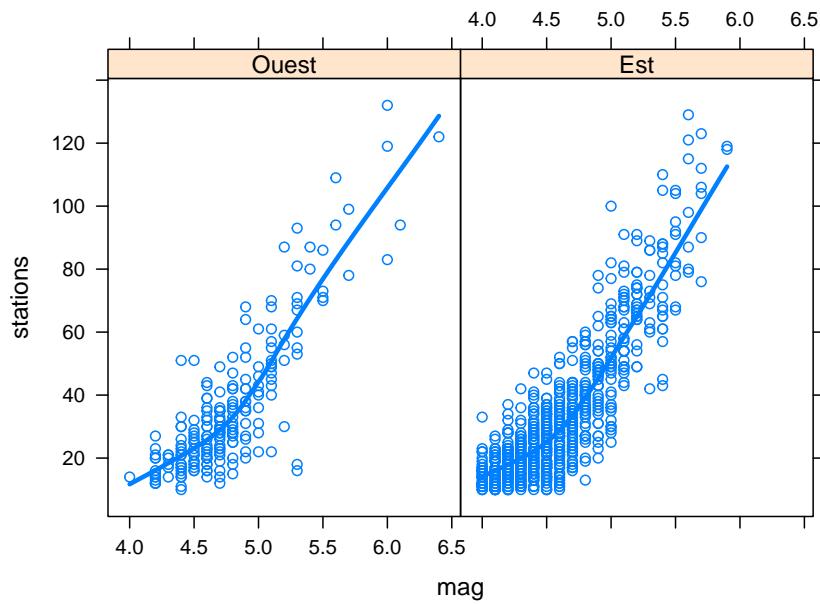
Diagrammes de dispersion juxtaposés

```
xyplot(stations ~ mag | region, data = quakes)
```



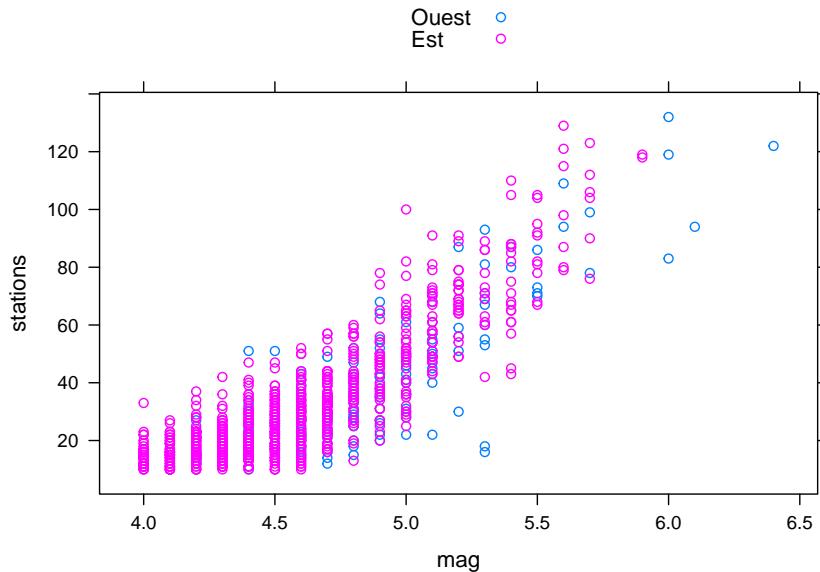
Ajout d'une courbe de lissage :

```
xyplot(stations ~ mag | region, data = quakes,
       type = c("p", "smooth"), lwd = 3)
```



### Diagrammes de dispersion superposés

```
xypplot(stations ~ mag, data = quakes, groups = region, auto.key = TRUE)
```



L'argument `auto.key` permet d'ajouter automatiquement une légende.

## Package `ggplot2`

Objectifs du créateur :

- Reprendre les forces de chacun des systèmes graphiques R précédents :
  - système de base : création de graphiques par couches (ajouts séquentiels d'éléments) ;
  - package `lattice` : représentations multivariables simples.
- Améliorer les systèmes :
  - automatiser certaines configurations graphiques (p. ex. les légendes),
  - faciliter l'ajout de transformations statistiques communes (p. ex. courbes de lissage, barres d'erreur),
  - fournir une esthétique par défaut pensée de façon à transmettre plus efficacement les informations contenues dans le graphique (p. ex. choix de couleur).

Dans le package `ggplot2`, tout a été repensé pour être plus simple d'utilisation et surtout pour que le graphique produit transmette plus efficacement l'information qu'il contient.

- <https://ggplot2.tidyverse.org/>
- <https://r4ds.had.co.nz/data-visualisation.html>

### La grammaire graphique

Principes de base de la grammaire graphique (d'où le `gg` dans `ggplot`) :

Graphique statistique = représentation de **données**, dans un **système de coordonnées** spécifique, divisée en éléments de base :

- **éléments géométriques** (*geoms*) : points, lignes, barres, etc. ;
- **propriétés visuelles** (*aesthetics*) des éléments géométriques : axes, couleurs, formes, tailles, etc.
- **transformations statistiques**, si désiré : courbe de régression ou de lissage, région d'erreur, etc.

Un graphique est spécifié en **associant des variables**, provenant des données, à **des propriétés visuelles** des éléments géométriques du graphique.

### Survol des principales fonctions

- `ggplot` : initialisation d'un objet de classe `ggplot`
- `qplot` : initialisation rapide (`q` pour *quick*) d'un objet `ggplot`
  - Utilisation plus intuitive que `ggplot` pour des gens familiers avec `plot`.
  - N'offre cependant pas toutes les possibilités de `ggplot`.
  - *Ne sera pas couvert ici.*
- `+` : opérateur pour l'ajout de couches ou la modification de configurations dans un objet `ggplot`
- fonctions de type `geom_` (p. ex. `geom_point`, `geom_boxplot`, `geom_bar`, etc.) : spécification de couches à ajouter à un graphique
- `aes` : création d'un `mapping`, soit une association entre des propriétés visuelles et des variables
- `ggsave` : enregistrement d'un graphique

Avec `ggplot2`, produire un graphique = afficher un objet `ggplot`.

### Code gabarit minimaliste

```
ggplot(data = <DATA>) +
  <GEOM_FUNCTION>(mapping = aes(<MAPPINGS>))
```

- source : [http://r4ds.had.co.nz/data-visualisation.html#a-graphing-template](https://r4ds.had.co.nz/data-visualisation.html#a-graphing-template)
- éléments entre `<` et `>` à remplacer :

- <DATA> = jeu de données, obligatoirement stocké dans un **data frame**, dans lequel les variables catégoriques doivent être des **facteurs** (dont les libellés des niveaux ont avantage à être informatifs, car ils apparaîtront dans le graphique) ;
- <GEOM\_FUNCTION> = le nom d'une fonction de type `geom_` pour ajouter une couche au graphique ;
- <MAPPINGS> = arguments à fournir à la fonction `aes`, (les arguments acceptés varient un peu selon la <GEOM\_FUNCTION>) ;

Les arguments `data` et `mapping` peuvent être fournis :

- à la fonction `ggplot` : leur valeur est alors commune à toutes les couches (il est tout de même possible de forcer l'utilisation d'autres données ou d'autres associations pour des couches spécifiques) ;
- ou à une <GEOM\_FUNCTION> : leur valeur est alors spécifique à la couche produite.

### Syntaxes équivalentes au gabarit minimalistique

```
# Version « tout sur la même ligne » du gabarit :
ggplot(data = <DATA>) + <GEOM_FUNCTION>(mapping = aes(<MAPPINGS>))
```

Étant donné que le graphique produit par ce gabarit ne comprend qu'une seule couche, les syntaxes suivantes sont équivalentes à celle du gabarit.

```
# argument data spécifique (fourni dans l'appel à la <GEOM_FUNCTION>)
ggplot() + <GEOM_FUNCTION>(data = <DATA>, mapping = aes(<MAPPINGS>))
```

```
# argument mapping global (fourni dans l'appel à ggplot)
ggplot(data = <DATA>, mapping = aes(<MAPPINGS>)) + <GEOM_FUNCTION>()
```

### Objet de classe `ggplot`

Un graphique produit en `ggplot2` est en fait un objet de classe `ggplot`.

Il faut commencer par l'initialiser :

```
monGraph <- ggplot(data = <DATA>)
```

puis y ajouter des couches :

```
monGraph <- monGraph + <GEOM_FUNCTION>(mapping = aes(<MAPPINGS>))
```

Il est possible d'observer le contenu de l'objet :

```
str(monGraph)
summary(monGraph)
```

et enfin de produire un graphique à partir de l'objet, simplement en l'affichant :

```
monGraph # ou print(monGraph)
```

## Quelques fonctions de type `geom_`

Fonction	Type de graphique	Élément(s) ajouté(s)
<code>geom_point</code>	diagramme de dispersion	points selon des coordonnées
<code>geom_line</code>	diagramme en lignes	segments de droites reliant des points
<code>geom_bar</code>	diagramme en bâtons	barres disjointes, de hauteurs spécifiées ou calculées = fréquences des niveaux d'un facteur
<code>geom_histogram</code>	histogramme	barres collées, de hauteurs calculées = fréquences d'observations d'une variable numérique tombant dans des intervalles joints ( <i>bin</i> )
<code>geom_boxplot</code>	diagramme en boîte	<i>boxplots</i>
<code>geom_density</code>	courbe de densité à noyau	courbe de la densité estimée par noyau ( <i>kernel density</i> )
<code>geom_qq</code>	diagramme quantile-quantile théorique	points pour les couples de quantiles empiriques et théoriques

La plupart de ces fonctions cachent des transformations statistiques (p. ex. `geom_bar` et `geom_histogram` calculent des fréquences, `geom_boxplot` calcule des quantiles, `geom_density` estime une densité, etc.)

## Quelques autres fonctions

Pour ajouter des couches ou modifier des configurations dans un objet `ggplot` initialisé :

- fonctions `ggtitle`, `xlab`, `ylab` : ajouter un titre, modifier les noms d'axes ;
- fonctions de type `coord_` : modifier des configurations reliées au système de coordonnées ;
- fonctions de type `theme_` : modifier des configurations reliées à l'apparence du graphique ;
- fonctions de type `scale_` : modifier les échelles de certaines propriétés visuelles (p. ex. couleurs, formes, tailles, etc.)
- fonctions de type `facet_` : créer des panneaux (*panel*) de graphiques côte-à-côte
  - chacun des sous-graphique est conditionnel à la valeur de facteur(s), il représente donc seulement le sous-ensemble des observations ayant une modalité particulière pour ce(s) facteur(s) ;
- fonctions de type `stat_` : ajouts d'éléments tirés d'un calcul mathématique ou statistique, p. ex. :
  - fonction `stat_function` pour tracer une fonction mathématique.

Il y a beaucoup d'autres fonctions : <http://ggplot2.tidyverse.org/reference/>

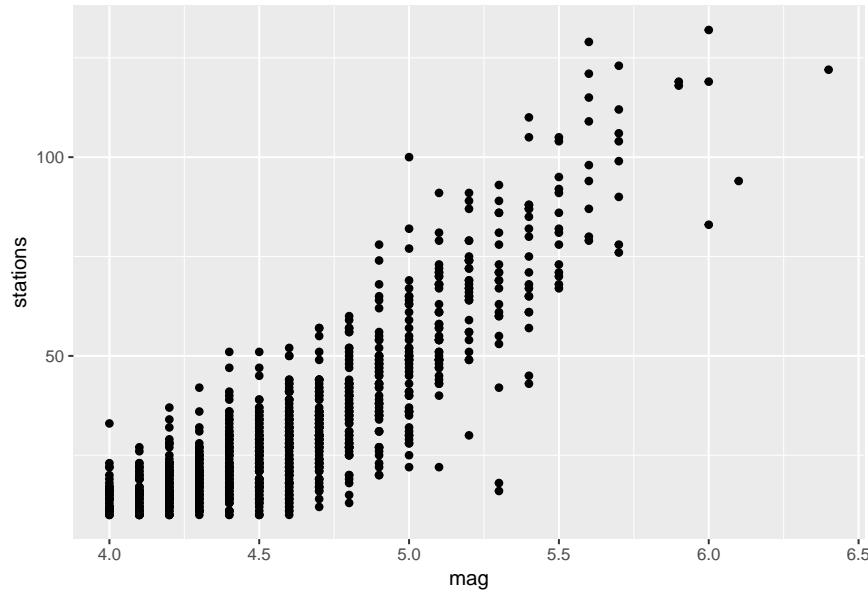
## Exemples de graphiques de différents types

Production de graphiques équivalent à certains de ceux produits avec le R de base

```
# Chargement du package
library(ggplot2)
```

### Diagramme de dispersion - fonction `geom_point`

```
ggplot(data = quakes) + geom_point(mapping = aes(x = mag, y = stations))
```



### Observation d'un objet de type ggplot

Création de l'objet

```
diag <- ggplot(data = quakes) + geom_point(mapping = aes(x = mag, y = stations))
# ou
diag <- ggplot(data = quakes)
diag <- diag + geom_point(mapping = aes(x = mag, y = stations))
```

Contenu de l'objet

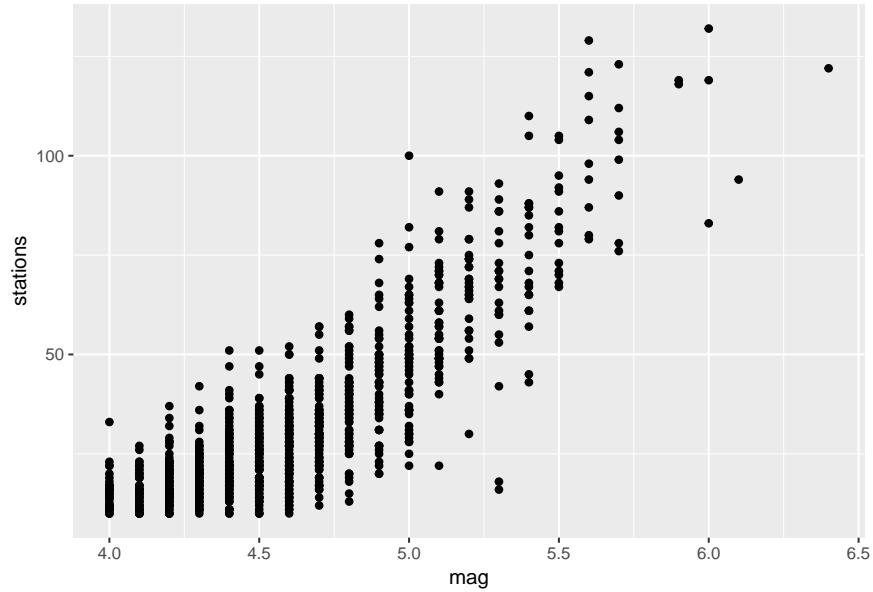
```
str(diag)
# sortie non affichée, car trop longue
summary(diag)
```

```
## data: lat, long, depth, mag, stations, magFactor, region [1000x7]
## faceting: <ggproto object: Class FacetNull, Facet, gg>
##   compute_layout: function
##   draw_back: function
##   draw_front: function
##   draw_labels: function
##   draw_panels: function
##   finish_data: function
##   init_scales: function
##   map_data: function
##   params: list
##   setup_data: function
##   setup_params: function
##   shrink: TRUE
##   train_scales: function
##   vars: function
##   super:  <ggproto object: Class FacetNull, Facet, gg>
## -----
```

```
## mapping: x = ~mag, y = ~stations  
## geom_point: na.rm = FALSE  
## stat_identity: na.rm = FALSE  
## position_identity
```

Production d'un graphique à partir de l'objet = affichage de l'objet

```
diag  
# ou  
#print(diag)
```



Avec la commande suivante, le même graphique est produit, mais il n'est pas enregistré dans un objet R.

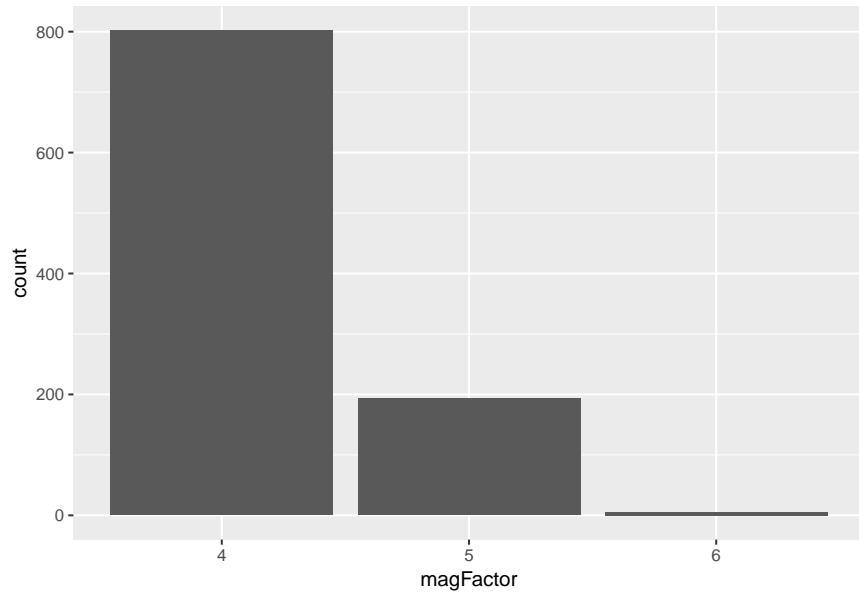
```
ggplot(data = quakes) + geom_point(mapping = aes(x = mag, y = stations))
```

### Diagramme en secteurs

Pas simple à produire avec ggplot2. Nous allons y revenir plus loin.

### Diagramme en bâtons - fonction geom\_bar

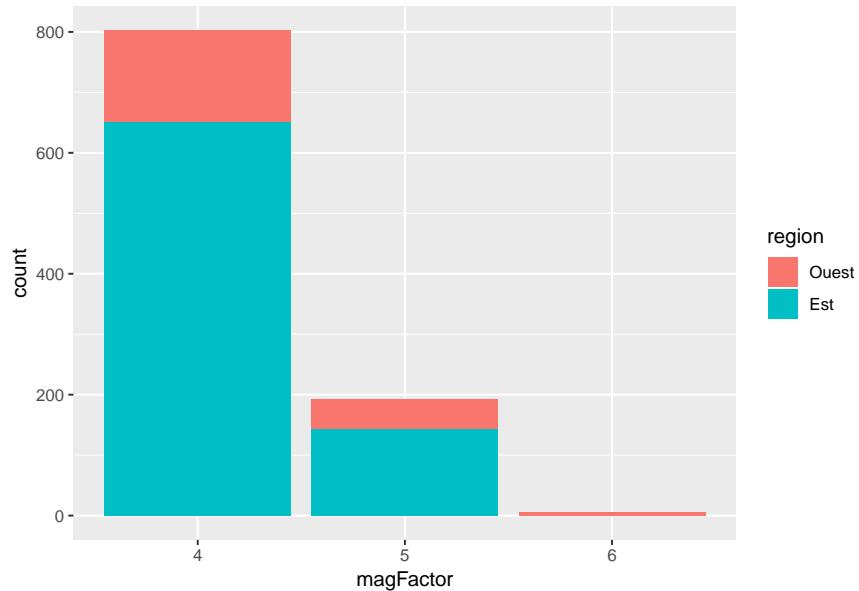
```
ggplot(data = quakes) + geom_bar(mapping = aes(x = magFactor))
```



**Remarque :** Pas besoin d'utiliser `table` pour calculer les fréquences. La fonction `geom_bar` les calcule pour nous. Si nous avons déjà en main les fréquences, il faut plutôt utiliser la fonction `geom_col` pour produire le diagramme en bâtons.

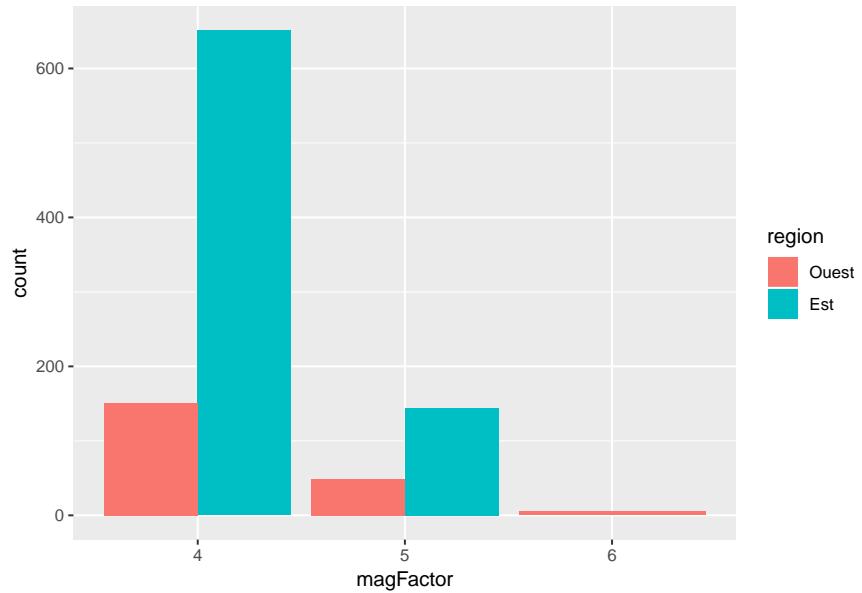
### Deux variables - bâtons empilés

```
ggplot(data = quakes) + geom_bar(mapping = aes(x = magFactor, fill = region))
```



### Deux variables - bâtons groupés

```
ggplot(data = quakes) +
  geom_bar(mapping = aes(x = magFactor, fill = region), position = "dodge")
```



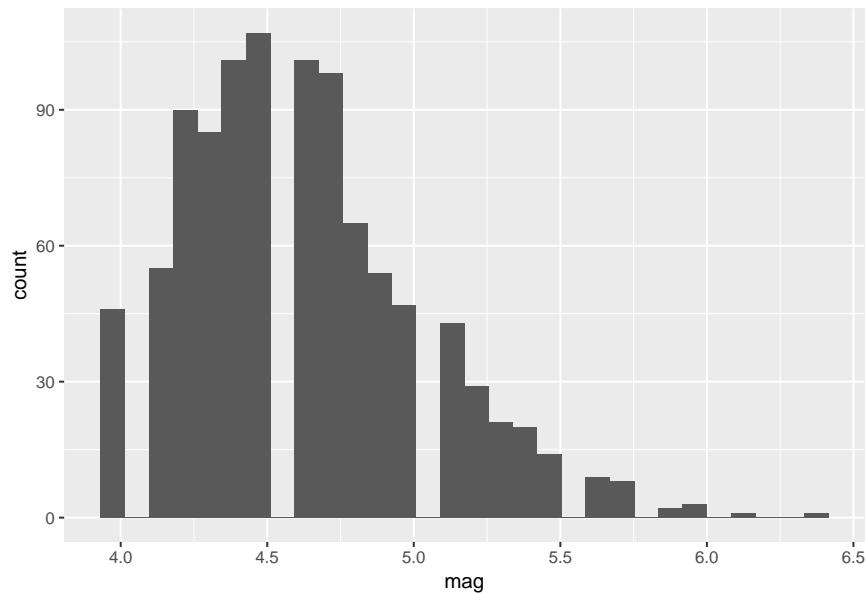
### Diagramme en mosaïque

Pas de fonction conçue pour produire ce type de graphique (mais des solutions sont proposées sur le web).

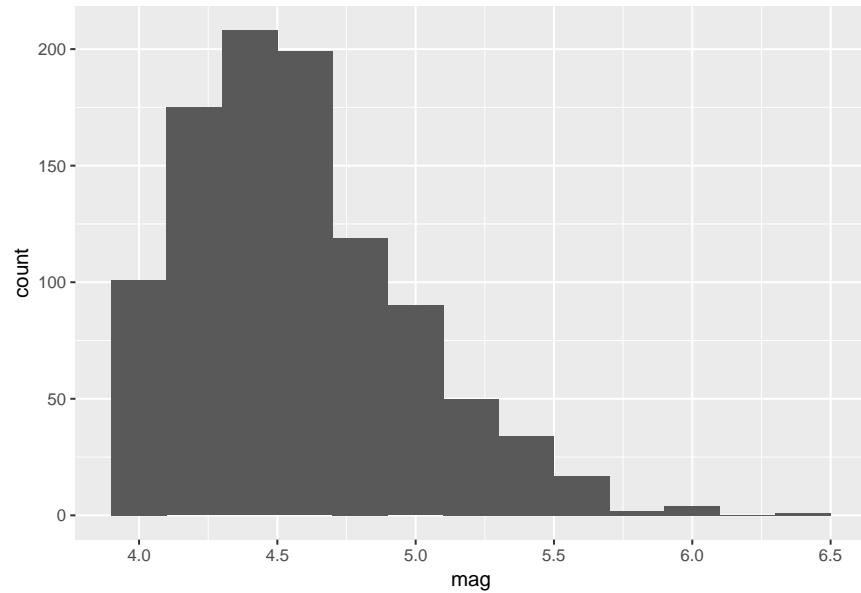
### Histogramme - fonction `geom_histogram`

```
ggplot(data = quakes) + geom_histogram(mapping = aes(x = mag))
```

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

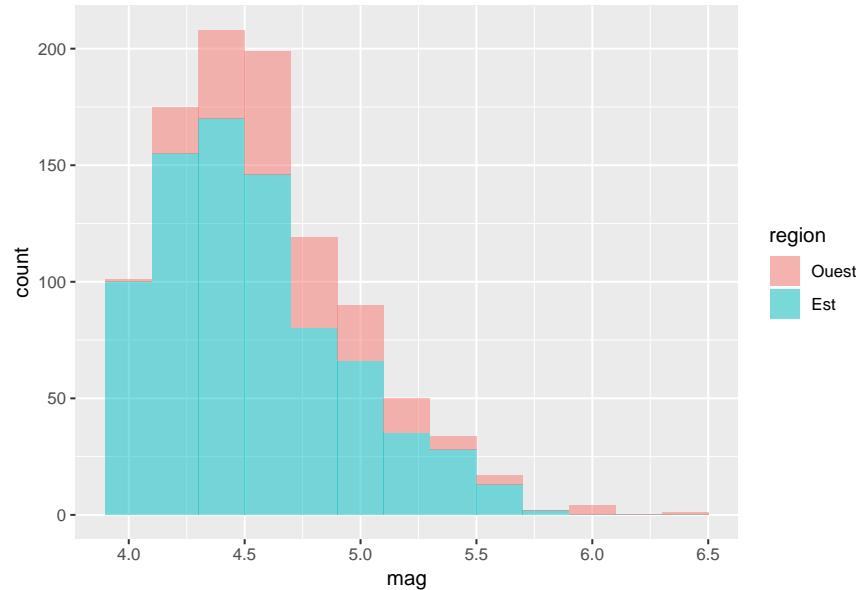


```
ggplot(data = quakes) +
  geom_histogram(mapping = aes(x = mag), binwidth = 0.2)
```



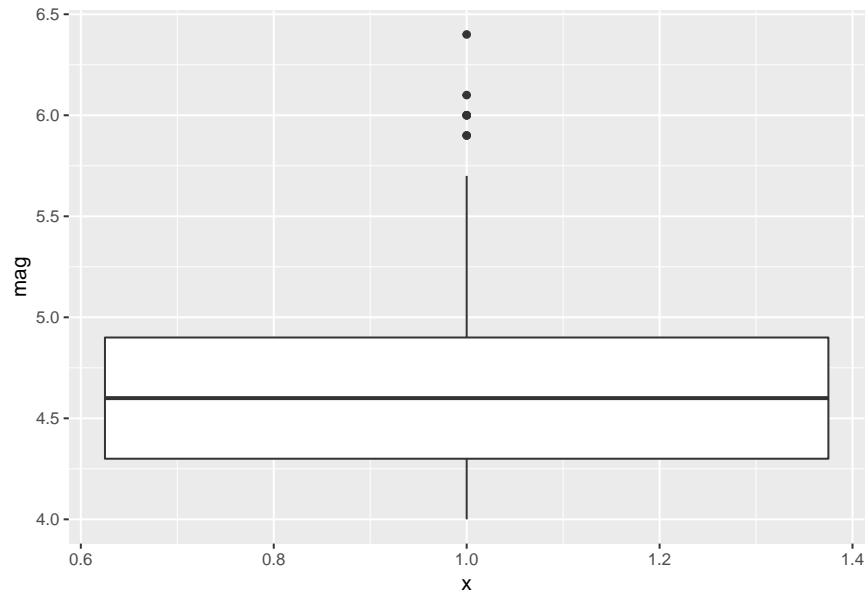
#### Deux variables - diagrammes superposés

```
ggplot(data = quakes) +
  geom_histogram(mapping = aes(x = mag, fill = region), binwidth = 0.2, alpha = 0.5)
```



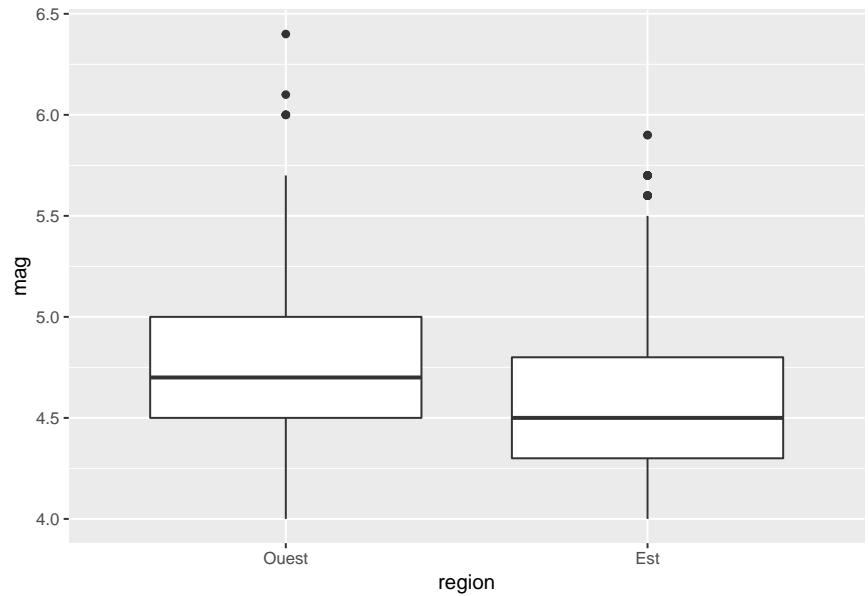
#### Diagramme en boîte - fonction geom\_boxplot

```
ggplot(data = quakes) +
  geom_boxplot(mapping = aes(x = 1, y = mag))
```



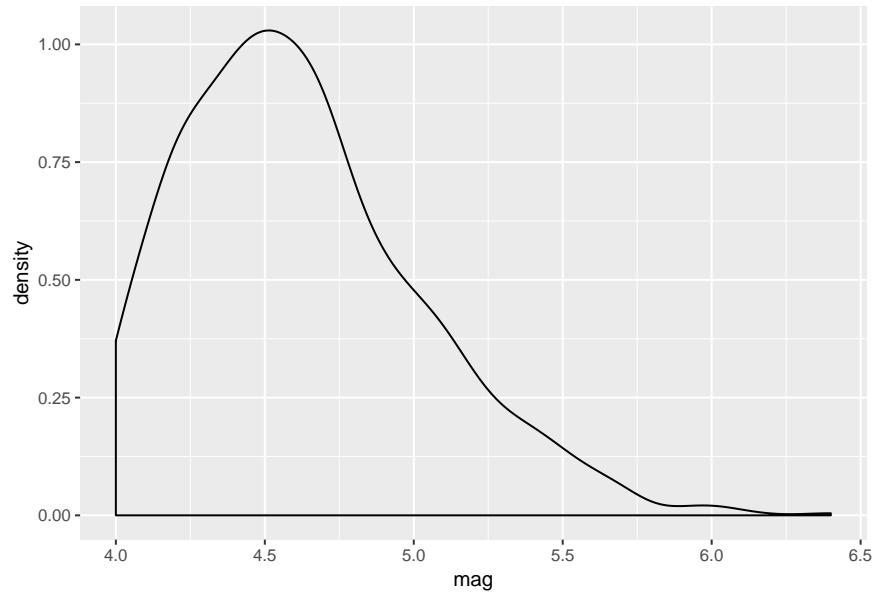
### Deux variables - diagrammes juxtaposés

```
ggplot(data = quakes) +
  geom_boxplot(mapping = aes(x = region, y = mag))
```



### Courbe d'estimation de densité à noyau - fonction geom\_density

```
ggplot(data = quakes) +
  geom_density(mapping = aes(x = mag))
```



Deux variables - courbes de densité superposés

```
ggplot(data = quakes) +
  geom_density(mapping = aes(x = mag, fill = region), alpha = 0.5)
```

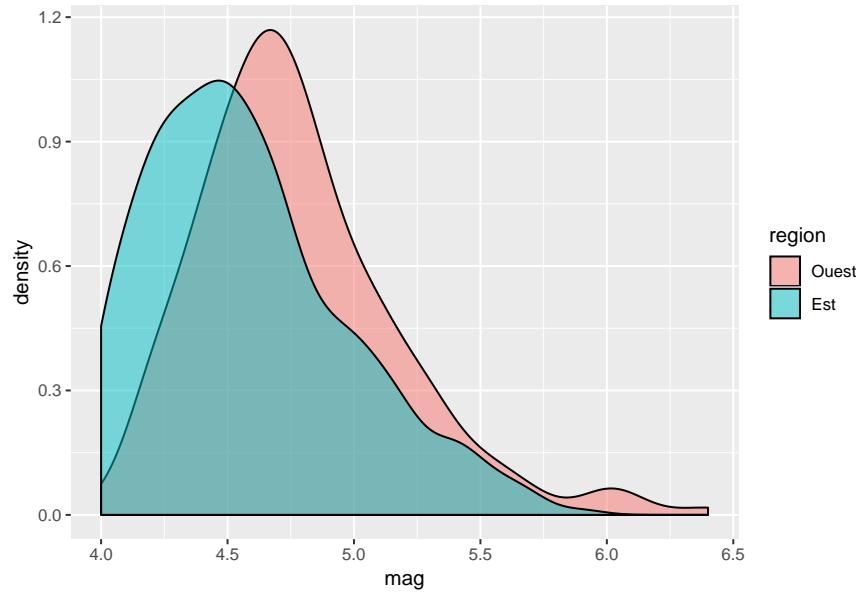
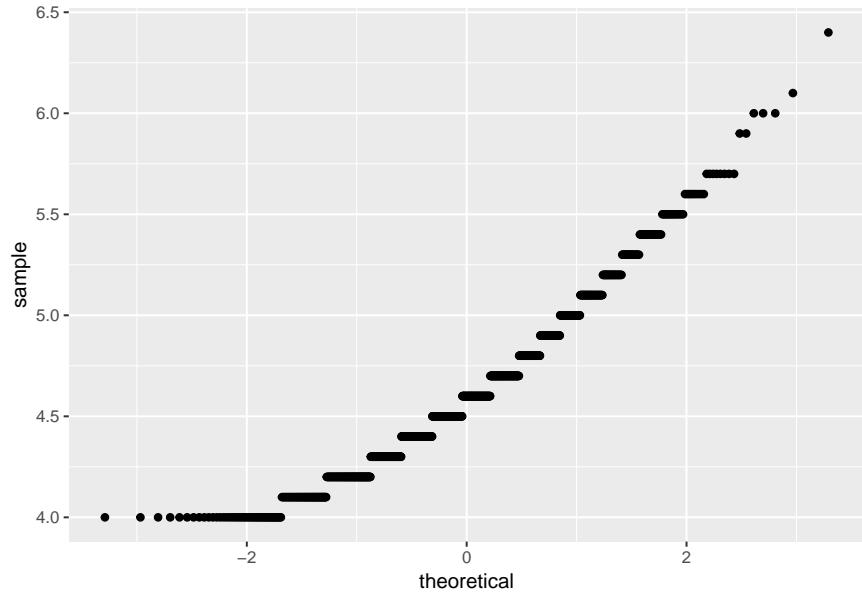


Diagramme quantile-quantile - fonction `geom_qq`

Une variable - diagramme quantile-quantile théorique normal

```
ggplot(data = quakes) +
  geom_qq(mapping = aes(sample = mag))
```



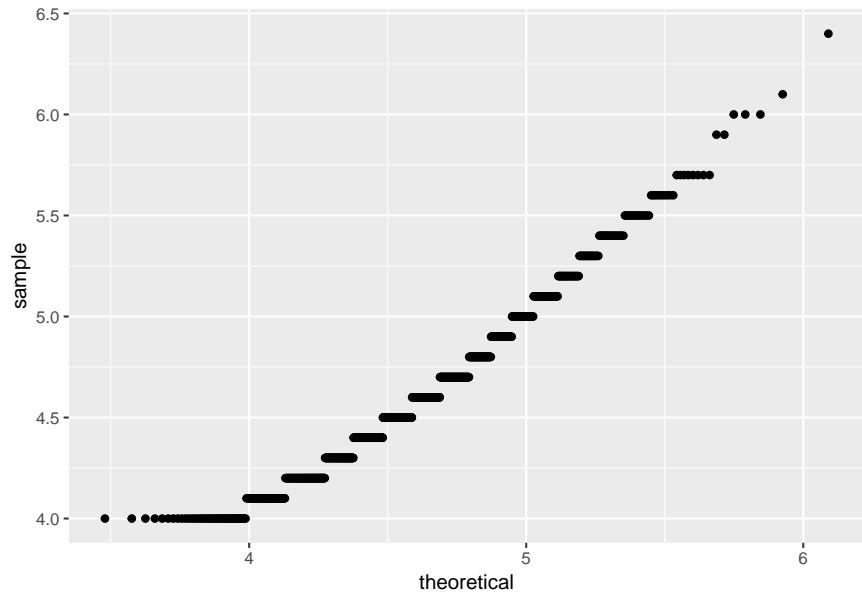
Code équivalent : utiliser l'argument `stat` de `geom_point` plutôt que `geom_qq`.

```
ggplot(data = quakes) +
  geom_point(mapping = aes(sample = mag), stat = "qq")
```

### Une variable - diagramme quantile-quantile théorique non normal

Par exemple pour la comparaison à une distribution log-normale.

```
ggplot(data = quakes) +
  geom_qq(mapping = aes(sample = mag), distribution = qlnorm,
         dparams = list(meanlog = mean(log(quakes$mag)), sdlog = sd(log(quakes$mag))))
```



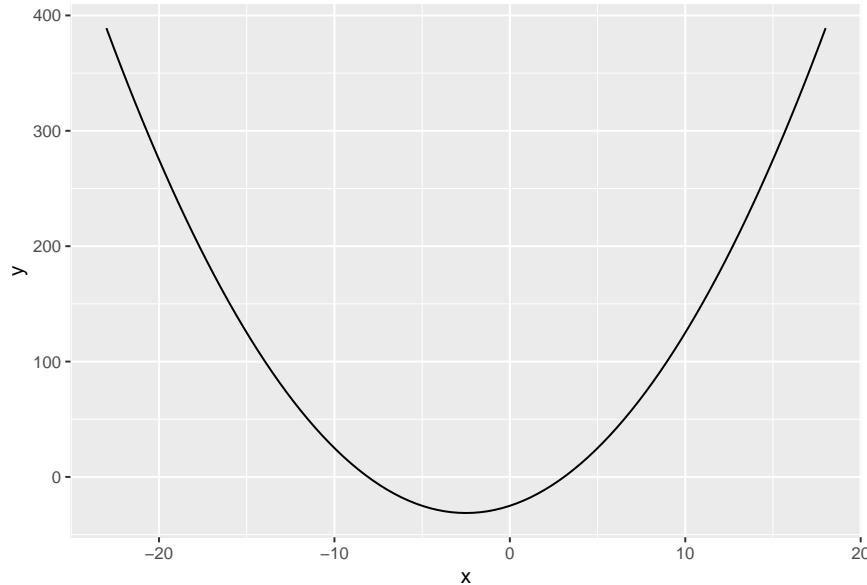
Avec le système graphique R de base c'est aussi faisable, mais c'est plus compliqué.

### Deux variables - diagramme quantile-quantile empirique

Pas de fonction conçue pour produire ce type de graphique (mais des solutions sont proposées sur le web).

### Représentation graphique d'une fonction `stat_function`

```
ggplot(data = data.frame(x = c(-23, 18)), mapping = aes(x)) +  
  stat_function(fun = function(x) {x^2 + 5*x - 25})
```



Ce type de graphique est plus simple à produire avec le système graphique de base, parce que :

- `ggplot2` est conçu pour faire des graphiques à partir de données,
- la fonction `curve` du système de base accepte une expression en entrée.

### Exemples de configuration de graphiques

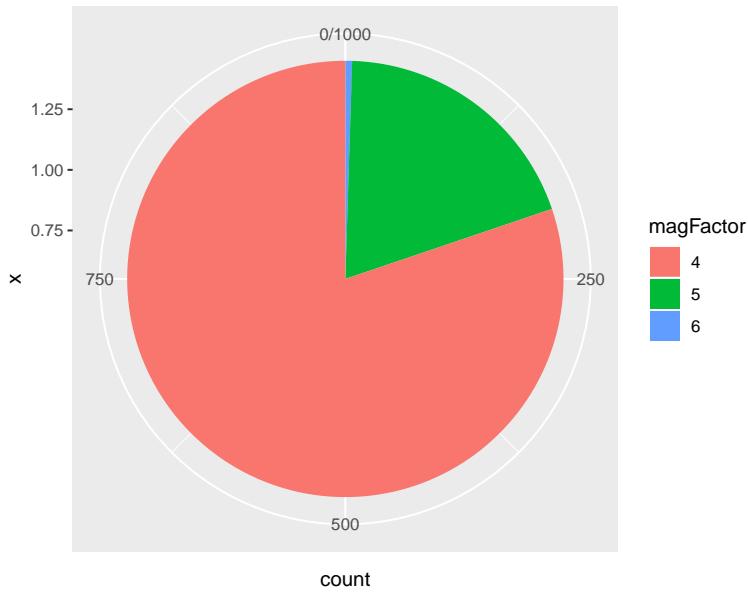
#### Diagramme en secteurs

Les auteurs de `ggplot2` n'offrent pas de fonction conviviale pour la création de diagramme en secteurs, probablement parce qu'ils ne recommandent pas leur utilisation. Malgré tout, tenter de tracer un diagramme en secteurs avec `ggplot2` aide à comprendre davantage les possibilités du package.

#### Coordonnées polaires - fonction `coord_polar`

Pour produire un diagramme en secteurs avec `ggplot2`, il faut d'abord produire un diagramme à bâtons empilés, puis demander l'utilisation d'un système de coordonnées polaires par un appel à la fonction `coord_polar`. C'est d'ailleurs pour illustrer l'utilisation de ce type de système de coordonnées que j'ai choisi d'inclure l'exemple suivant.

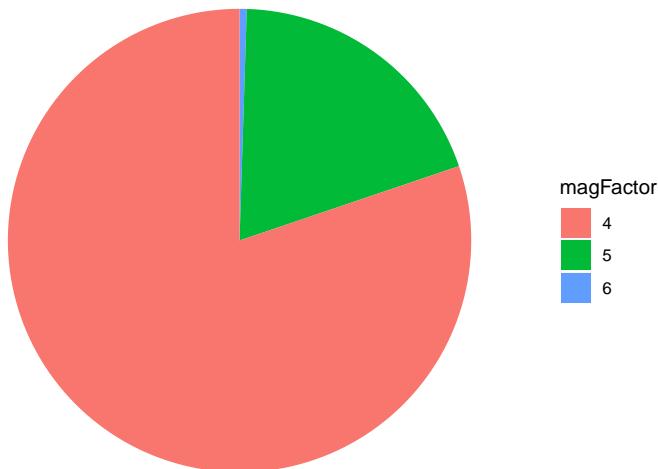
```
ggplot(data = quakes) +  
  geom_bar(mapping = aes(x = 1, fill = magFactor)) +  
  coord_polar("y")
```



### Configurations graphiques - fonction `theme`

Pour améliorer le résultat, il faut aller modifier des configurations graphiques avec la fonction `theme`.

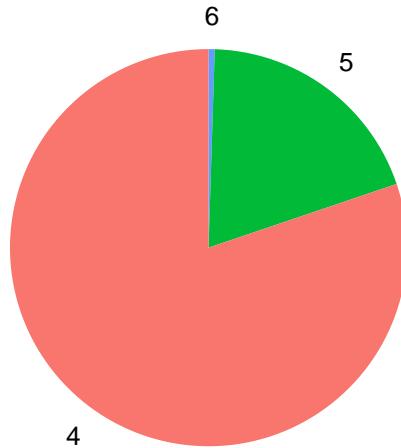
```
ggplot(data = quakes) +
  geom_bar(mapping = aes(x = 1, fill = magFactor)) +
  coord_polar("y") +
  theme(axis.title = element_blank(),
        axis.text = element_blank(),
        axis.ticks = element_blank(),
        panel.grid = element_blank(),
        panel.background = element_blank())
```



### Annotations textuelles - fonction `geom_text`

Nous pourrions même remplacer la légende par des annotations textuelles vis-à-vis les secteurs.

```
ggplot() +
  geom_bar(data = quakes, mapping = aes(x = 1, fill = magFactor)) +
  geom_text(data = as.data.frame(table(magFactor = quakes$magFactor)),
            mapping = aes(x = 1.6,
                           y = Freq/2 + c(rev(cumsum(rev(Freq)))[-1], 0),
                           label = magFactor),
            size = 5) +
  coord_polar("y") +
  theme(legend.position="none", # pour retirer la légende
        axis.title = element_blank(),
        axis.text = element_blank(),
        axis.ticks = element_blank(),
        panel.grid = element_blank(),
        panel.background = element_blank())
```



### Échelle de couleurs - fonctions `scale_colour_xxx` et `scale_fill_xxx`

Le deuxième terme dans le nom de la fonction fait référence aux arguments de `aes` :

- `fill` = couleur de remplissage,
- `colour` = couleur de points et de lignes.

Le troisième terme, ici désigné par `xxx`, peut être :

- `manual` : couleurs spécifiques individuellement,
- `brewer` : utiliser une palette de couleur offerte dans le package R `RColorBrewer`,
- `viridis_d` ou `viridis_c` : utiliser une palette de couleur offerte dans le package R `viridisLite`,
- `grey` : utiliser une palette (dégradé) de couleurs grises,
- etc.

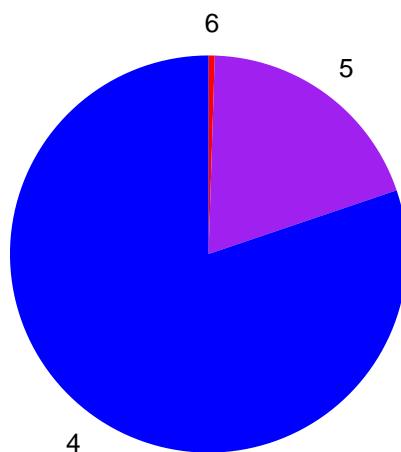
Nous allons changer la palette de couleur utilisée dans le diagramme en secteur.

```
ggplot() +
  geom_bar(data = quakes, mapping = aes(x = 1, fill = magFactor)) +
  geom_text(data = as.data.frame(table(magFactor = quakes$magFactor)),
```

```

mapping = aes(x = 1.6,
              y = Freq/2 + c(rev(cumsum(rev(Freq)))[-1], 0),
              label = magFactor),
size = 5) +
coord_polar("y") +
theme(legend.position="none", # pour retirer la légende
      axis.title = element_blank(),
      axis.text = element_blank(),
      axis.ticks = element_blank(),
      panel.grid = element_blank(),
      panel.background = element_blank()) +
scale_fill_manual(values=c("blue", "purple", "red"))

```



## Diagramme en bâtons horizontaux

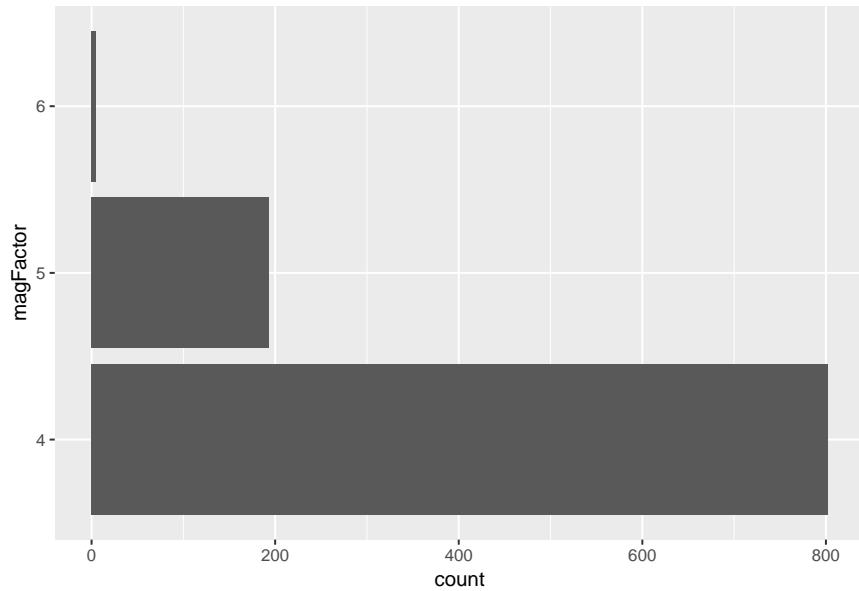
### Intervertir des axes - fonction coord\_flip

Une autre configuration possible en lien avec le système de coordonnées : intervertir les axes.

```

ggplot(data = quakes) +
  geom_bar(mapping = aes(x = magFactor)) +
  coord_flip()

```



### Ajout de courbes à un histogramme

```

moy <- mean(quakes$mag)
et <- sd(quakes$mag)

ggplot(data = quakes, mapping = aes(x = mag)) +

  # Ajout de l'histogramme :
  geom_histogram(mapping = aes(y = ..density..), binwidth = 0.2,
                 colour = "black", fill = "white") +
  xlab("magnitude du séisme") + ylab("densité") +
  ggtitle("Densité empirique des magnitudes dans le jeu de données quakes") +
  theme_minimal()

  # Ajout de la courbe de densité à noyau
  geom_density(aes(colour = "c1")) +
  stat_function(aes(colour = "c2"),
                fun = dnorm,
                args = list(mean = moy, sd = et),
                xlim = c(3.9, 6.5)) +
  stat_function(aes(colour = "c3"),
                fun = dlnorm,
                args = list(meanlog = mean(log(quakes$mag)),
                            sdlog = sd(log(quakes$mag))),
                xlim = c(3.9, 6.5)) +
  stat_function(aes(colour = "c4"),
                fun = dgamma,
                args = list(shape = moy^2/et^2, rate = moy/et^2),
                xlim = c(3.9, 6.5)) +
  theme_minimal()

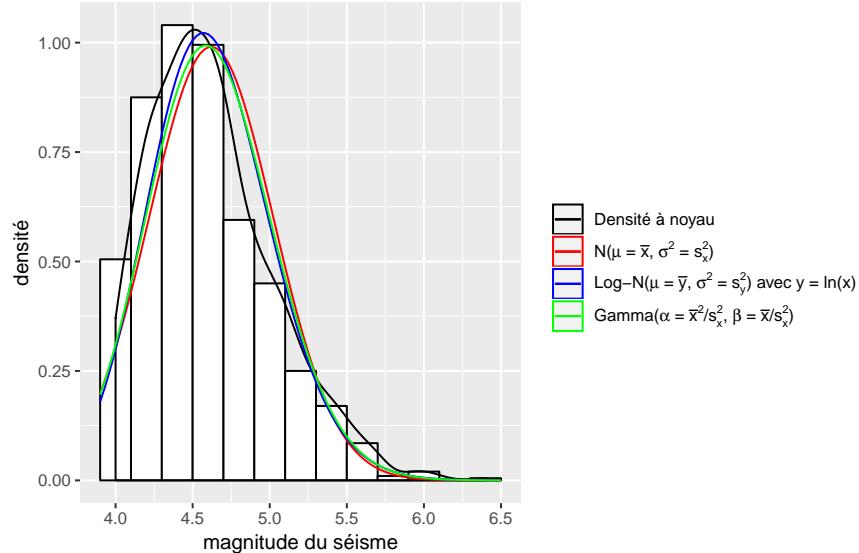
  # Ajout manuel d'une légende
  scale_colour_manual(name = "",
```

```

values = c(c1 = "black", c2 = "red", c3 = "blue", c4 = "green"),
labels = c(c1 = "Densité à noyau",
           c2 = expression(paste("N(", mu, " = ", bar(x), ", ",
           sigma^2, " = ", s[x]^2, ")")),
           c3 = expression(paste("Log-N(", mu, " = ", bar(y), ", ",
           sigma^2, " = ", s[y]^2, ") avec y = ln(x)")),
           c4 = expression(paste("Gamma(", alpha, " = ", bar(x)^2, "/",
           beta, " = ", bar(x), "/", s[x]^2, ")")))) +
theme(legend.text.align = 0)

```

Densité empirique des magnitudes dans le jeu de données quakes



#### Remarque :

- Il est compliqué avec `ggplot2` d'ajouter une légende manuellement comme nous avons dû le faire dans ce graphique.
- Les annotations mathématiques sont toujours possibles avec `ggplot2`, comme dans le système graphique de base.

#### Division de la fenêtre en sous-fenêtres

Cette fonctionnalité n'est pas offerte en `ggplot2`, sauf pour des graphiques conditionnels à la valeur d'un facteur.

Par contre, certains packages offrent des fonctions pour le faire, notamment

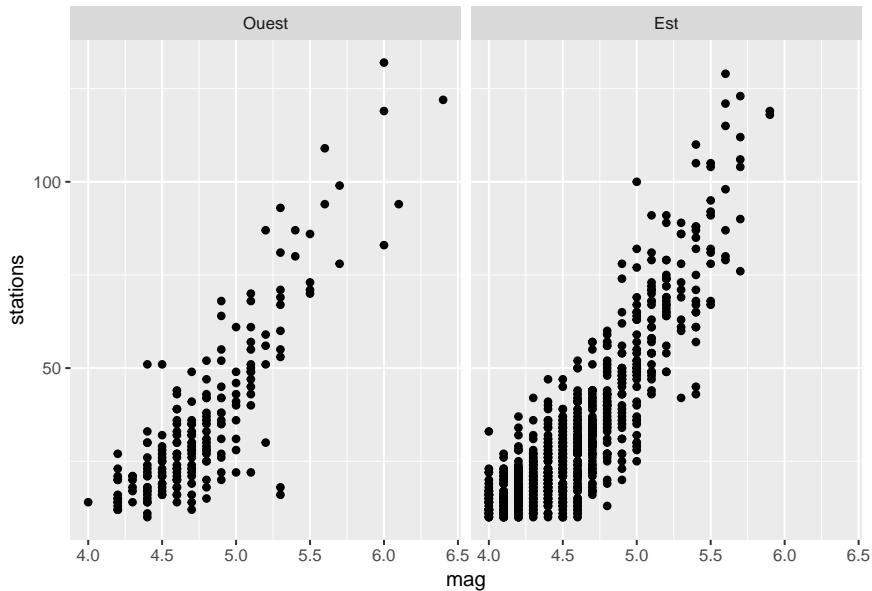
- des opérateurs du package `patchwork`,
- la fonction `plot_grid` du package `cowplot`,
- la fonction `grid.arrange` du package `gridExtra`.

Je ne vais pas illustrer ici l'utilisation de ces packages.

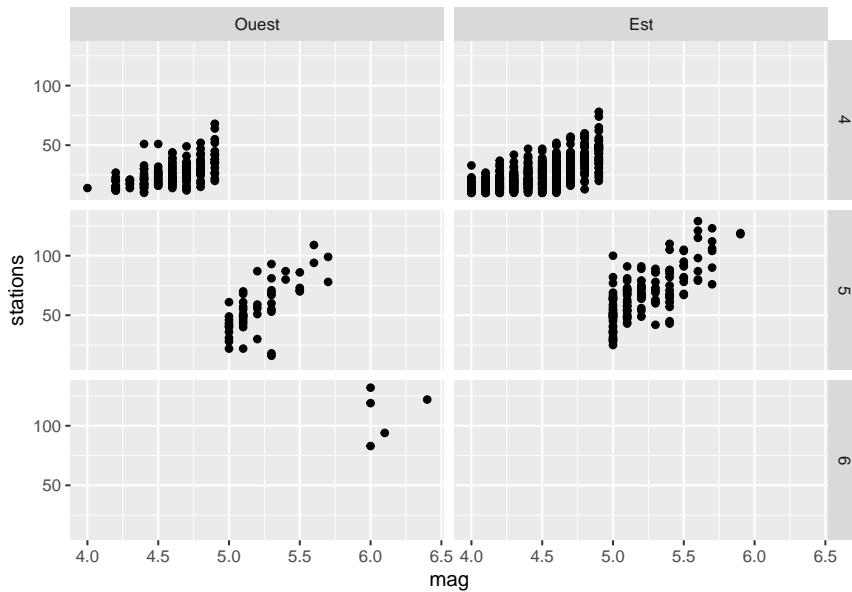
#### Représentations multivariables

##### Sous-fenêtres selon une ou des variable(s) catégorique(s) - fonctions `facet_wrap` et `facet_grid`

```
ggplot(data = quakes, mapping = aes(x = mag, y = stations)) +
  geom_point() +
  facet_wrap(facets = ~ region)
```

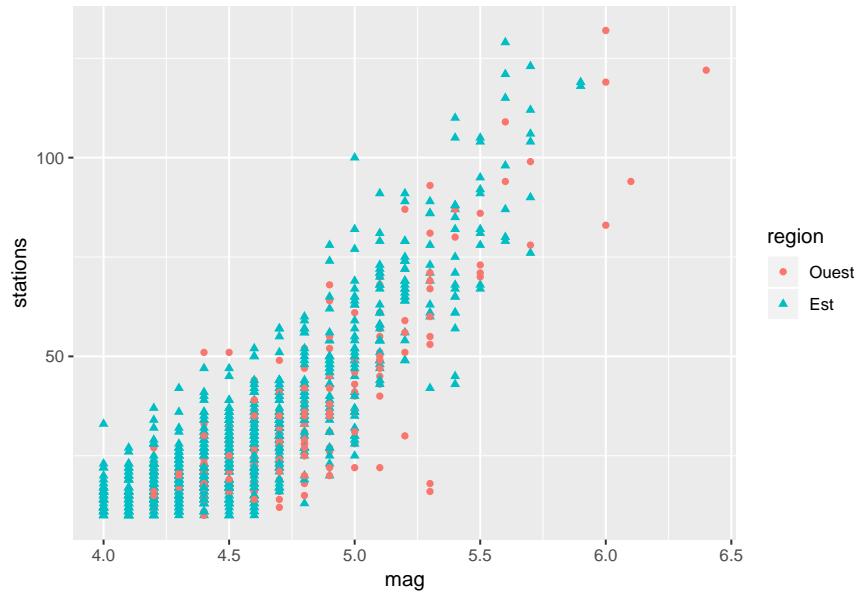


```
ggplot(data = quakes, mapping = aes(x = mag, y = stations)) +
  geom_point() +
  facet_grid(facets = magFactor ~ region)
```



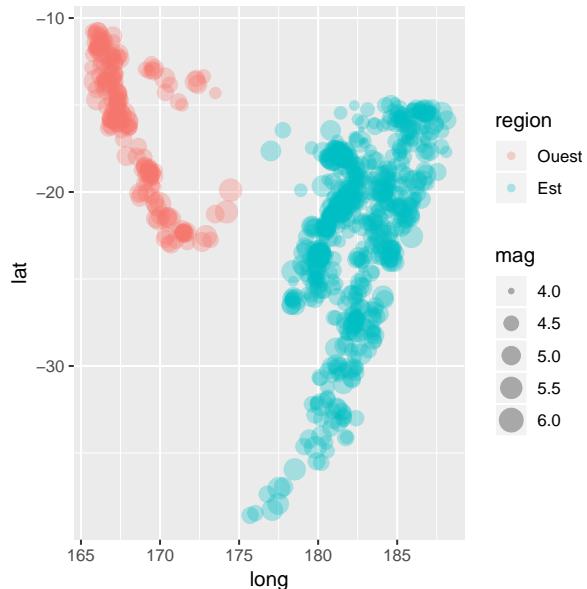
### Superposition avec aspect distinctif selon la variable catégorique

```
ggplot(data = quakes) +
  geom_point(mapping = aes(x = mag, y = stations, colour = region, shape = region))
```



### Représentation de trois variables numériques et d'une variable catégorique

```
ggplot(data = quakes) +
  geom_point(mapping = aes(x = long, y = lat, colour = region, size = mag),
             alpha = 0.3) +
  coord_quickmap()
# coord_quickmap() permet d'avoir des axes sur la même échelle,
# ce qui est préférable ici, car les variables sont des coordonnées géographiques
```



### Enregistrement d'un graphique - fonction ggsave

La fonction `ggsave` enregistre le dernier graphique `ggplot2` produit, en conservant les dimensions de la fenêtre graphique.

```
ggsave(file = "ExempleGraphique_ggplot2.pdf")
```

### Comparaison entre ggplot2 et le système graphique R de base

Aspects plus faciles avec ggplot2 :

- intégration des légendes (souvent automatisée),
- représentations multivariables,
- ajouts d'éléments qui requièrent un calcul statistique (p. ex. barres erreurs).

Avantages du système graphiques R de base :

- les configurations graphiques sont souvent plus simples.

Références :

- <http://seananderson.ca/ggplot2-FISH554/>
  - <http://flowingdata.com/2016/03/22/comparing-ggplot2-and-r-base-graphics/>
- 

## Autres possibilités graphiques en R (seulement mentionnées)

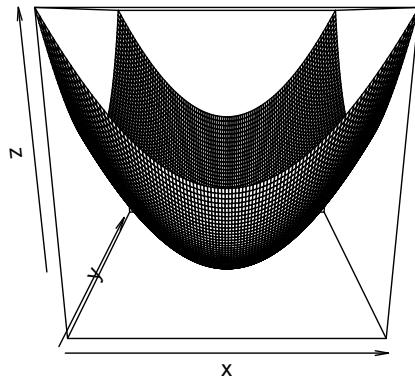
- Graphiques 3D orientables :
  - fonction `persp` du système de base,
  - fonctions `cloud` et `wireframe` du package `lattice`,
  - package `rgl` : fonction `plot3d` et autres.
- Cartes géographiques :
  - packages `maps` et `mapdata`,
  - plusieurs packages sont offerts.
- Graphiques interactifs :
  - package `plotly` pour créer des graphiques `Plotly`,
  - plusieurs packages sont offerts.
- [Shiny App](#)
- et bien d'autres : <https://CRAN.R-project.org/view=Graphics>

## Graphiques 3D

### Fonction `persp` du système de base

Graphique 3D produit avec la fonction `persp` du système graphique de base :

```
x <- y <- seq(from = -1, to = 1, length = 100)
grille <- expand.grid(x = x, y = y)
z <- grille$x^2 + grille$y^2
persp(x = x, y = y, z = matrix(z, ncol = length(x)), zlab = "z")
```

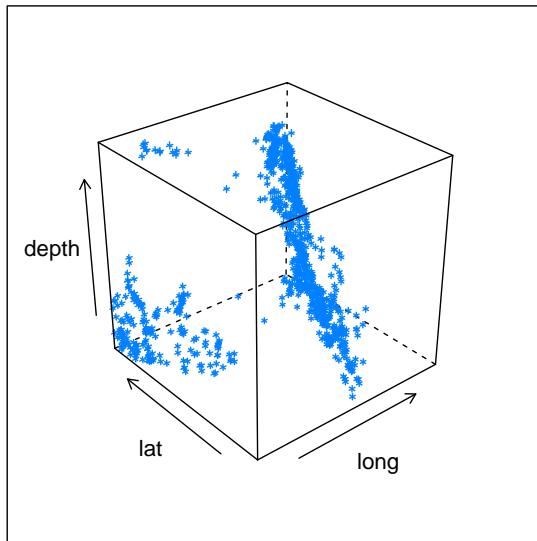


Cette fonction ne permet pas de représenter un nuage de points en 3 dimensions. Elle permet plutôt de tracer une fonction mathématique de 2 variables. Il faut fournir à l'argument `z` la valeur de la fonction mathématique pour toutes les combinaisons des valeurs en `x` et `y`.

#### Fonction `cloud` du package `lattice`

Graphique 3D produit avec la fonction `cloud` du package `lattice` :

```
cloud(depth ~ long + lat, data = quakes)
```



#### Graphiques 3D orientables - package `rgl`

Le package `rgl` permet de produire des graphiques 3D. Ces graphiques sont affichés dans une fenêtre interactive dans laquelle il est possible de tourner le graphique. Il n'est cependant pas possible de visualiser le résultat dans un document HTML ou PDF. Voici un exemple de commande pour créer un graphique 3D avec la fonction `plot3d` du package `rgl`.

```

library(rgl)
plot3d(x = quakes$long, y = quakes$lat, z = quakes$depth,
       col = ifelse(quakes$region == "Ouest", rgb(1, 0, 0, 0.3), rgb(0, 0, 1, 0.3)),
       aspect = TRUE, type = "s", size = 1)
# Le graphique produit n'apparaît pas dans le fichier PDF, car il est affiché dans
# une fenêtre graphique interactive qui ne peut pas être intégrée à un PDF.

```

## Cartes géographiques

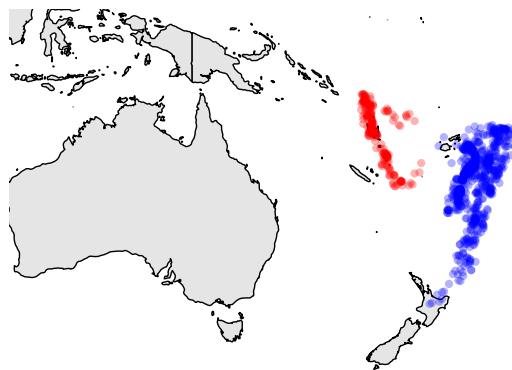
### Packages `maps` et `mapdata`

Carte tirée de banques de données accessibles par l'intermédiaire du package `maps`.

```

library(maps)
map(database = "world", xlim = c(115, 190), ylim = c(-50, 0),
     col = "gray90", fill = TRUE)
points(x = quakes$long, y = quakes$lat, pch = 20,
       col = ifelse(quakes$region == "Ouest", rgb(1, 0, 0, 0.3), rgb(0, 0, 1, 0.3)))

```



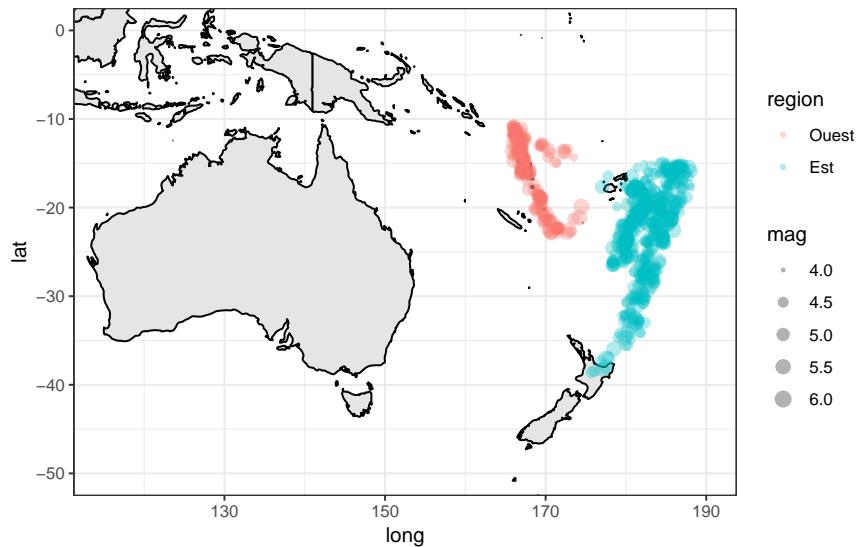
Au besoin, encore plus de cartes sont accessibles par l'intermédiaire du package `mapdata`.

Voici un exemple d'utilisation d'une carte avec `ggplot2` :

```

monde <- map_data("world") # va chercher des données provenant du package maps ou mapdata
ggplot() +
  geom_polygon(data = monde,
               aes(x=long, y = lat, group = group), fill = "gray90", col = "black") +
  coord_quickmap(xlim = c(115, 190), ylim = c(-50, 0)) +
  geom_point(data = quakes,
             aes(x = long, y = lat, colour = region, size = mag),
             shape = 20, alpha = .3) +
  theme_bw()

```



Références :

- <http://www.michaeljgrogan.com/creating-maps-ggplot2-libraries/>
- <http://eriqande.github.io/rep-res-web/lectures/making-maps-with-R.html>

### Autres packages pour l'intégration de cartes géographiques dans un graphique R

Il existe plusieurs autres packages pour la création de cartes en R :

- CRAN task view : <https://CRAN.R-project.org/view=Spatial>
- ressources web sur le sujet :
  - <https://statnmap.com/fr/>
  - <https://geocompr.robinlovelace.net/adv-map.html>
- package **ggmap** : permet d'intégrer une carte Google Maps à un graphique produit avec **ggplot2** (requiert un compte sur <https://cloud.google.com/maps-platform/> pour avoir accès aux cartes)
- package **tmap** : permet notamment de facilement créer des **cartes choroplèthes**
- package **leaflet** : permet de créer des cartes interactives avec la librairie JavaScript du même nom <https://leafletjs.com/>
- etc.

## Graphiques interactifs

Un graphique interactif est un graphique qui n'est pas statique (une image fixe), mais qui est plutôt capable d'interagir avec la personne qui l'observe. La forme la plus simple d'interaction est de fournir de l'information supplémentaire lorsque la souris est placée sur certains éléments du graphique. Une forme plus avancée d'interaction est de modifier certains éléments du graphique en fonction de choix fait par l'observateur, par exemple via des boutons ou un menu cliquable.

Ce genre de graphique se visualise typiquement dans un navigateur internet. Il s'intègre facilement à un document HTML (comme une page web), mais pas à un document PDF.

### Graphiques interactifs Plotly - package **plotly**

L'application web Plotly (<https://plot.ly/>) est devenue très populaire pour produire des graphiques interactifs. Le package R **plotly** permet de créer facilement des graphiques Plotly en R. Voici un exemple affiché seulement dans la version HTML des notes, pas dans le PDF.

```

library(plotly)

plot_ly(quakes, x = ~mag, y = ~stations, color = ~depth,
        size = ~depth, text = ~paste("Region: ", region))

```

## Autres packages pour la création de graphiques interactifs en R

Les tutoriels réalisés par les étudiants de ce cours en 2018 portaient sur 4 packages R pour la création de graphiques interactifs :

- package [plotly](#) mentionné ci-dessus :  
[https://stt4230.rbind.io/tutoriels\\_etudiants/hiver\\_2018/plotly/](https://stt4230.rbind.io/tutoriels_etudiants/hiver_2018/plotly/)
- package [googleVis](#) pour créer des Google charts :  
[https://stt4230.rbind.io/tutoriels\\_etudiants/hiver\\_2018/googlevis/](https://stt4230.rbind.io/tutoriels_etudiants/hiver_2018/googlevis/)
- package [highcharter](#) basé sur la librairie JavaScript Highcharts :  
[https://stt4230.rbind.io/tutoriels\\_etudiants/hiver\\_2018/highcharter/](https://stt4230.rbind.io/tutoriels_etudiants/hiver_2018/highcharter/)
- package [ggiraph](#) pour rendre des graphiques ggplot2 interactifs :  
[https://stt4230.rbind.io/tutoriels\\_etudiants/hiver\\_2018/ggiraph/](https://stt4230.rbind.io/tutoriels_etudiants/hiver_2018/ggiraph/)

## Conclusion

### Quelques conseils :

- Trouver de l'information :
  - Système graphique de base :  
lire les fiches d'aide, nous y apprenons toutes les possibilités des fonctions ;
  - Autres systèmes (en particulier ggplot2) et packages graphiques :  
faire des recherches sur internet.
- Stratégie 1 - Procéder par essai/erreur jusqu'à obtenir le graphique souhaité.
- Stratégie 2 - Partir d'un modèle :
  - un graphique que nous avons déjà fait ou
  - un graphique trouvé dans un livre ou sur internet (voir les références).

### R = un excellent outil pour les graphiques

- R produit des graphiques d'assez bonne qualité pour des publications et facilement reproductibles à partir de leur code source.
  - La variété des graphiques possibles de créer en R est impressionnante.
  - De nouvelles fonctions graphiques sont fréquemment ajoutées.
-

# Synthèse

## Packages graphiques en R

- `graphics` et `grDevices` : intégrés au R de base => chargés par défaut lors de l'ouverture d'une session R

## Systèmes graphiques alternatifs :

- `lattice` : emphase sur les représentations multivariables
- `ggplot2` : conçu en ayant comme objectif la simplicité d'utilisation et la qualité des graphiques produits

## Packages offrant d'autres types de graphiques, notamment :

- graphiques 3D orientables : `rgl`
- cartes géographiques : `maps`, `mapdata`, etc.
- graphiques interactifs : `plotly`, etc.
- et bien d'autres : <https://cran.r-project.org/web/views/Graphics.html>

## Procédure de création d'un graphique de base

1. La configuration des paramètres graphiques (au besoin) :
  - énoncé `par` ou `layout`.
2. L'initialisation d'un graphique :
  - fonction de base : `plot`,
  - ou fonction pour un type spécifique graphiques : `pairs`, `matplot`, `pie`, `barplot`, `mosaicplot`, `hist`, `boxplot`, `qqnorm`, `qqplot`, `curve`, etc.
3. L'ajout séquentiel d'éléments au graphique (au besoin) :
  - fonctions d'ajouts à un graphique déjà initialisé :
    - `points`, `matpoints`, `lines`, `matlines`, `abline`, `segments`, `arrows`, `rect`, `polygon`, `legend`, `text`, `mttext`, `title`, `axis`, `box`, `qpline`, etc. ;
    - `matplot`, `barplot`, `hist`, `boxplot`, `curve` avec l'argument `add = TRUE`.

Arguments et paramètres graphiques ⇒ mise en forme et annotations

## Fonction `plot`

- Fonction générique qui possède plusieurs méthodes :
  - Le comportement de `plot` dépend des types d'objets fournis en entrée comme deux premiers arguments.
- Choix de `plot` = bon type de graphique à produire selon le nombre et la nature des variables.
- Observations numériques de deux variables en entrée ⇒ **diagramme de dispersion**.

## Autres fonctions de création d'un graphique

Représentations d'une variable (observations stockées dans `x`)

Type de graphique	Appel de fonction	Type de <code>x</code>
diagramme en secteurs ( <i>pie chart</i> )	<code>pie(table(x), ...)</code>	facteur
diagramme en bâtons ( <i>bar plot</i> )	<code>barplot(table(x), ...)</code>	facteur
histogramme	<code>hist(x, ...)</code>	vecteur numérique
courbe de densité à noyau ( <i>kernel density plot</i> )	<code>plot(density(x), ...) → méthode plot.density</code>	vecteur numérique
diagramme en boîte ( <i>boxplot</i> )	<code>boxplot(x, ...)</code>	vecteur numérique
diagramme quantile-quantile théorique normal	<code>qqnorm(x, ...)</code>	vecteur numérique

Représentation d'une **expression mathématique** : `curve(expr, ...)`

Représentations de deux variables (observations stockées dans `x` et `y`)

Type de graphique	Appel de fonction	Type de <code>x</code>	Type de <code>y</code>
diagramme en bâtons empilés ou groupés	<code>barplot(table(x, y), ...)</code> avec <code>beside = TRUE</code> pour bâtons groupés	facteur	facteur
diagramme en mosaïque	<code>mosaicplot(tablex, y, ...)</code>	facteur	facteur
diagrammes en boîte juxtaposés	<code>boxplot(y ~ x, ...)</code>	facteur	vecteur
diagramme de dispersion ( <i>scatterplot</i> ) ou en lignes ( <i>line chart</i> )	<code>plot(x, y, ...)</code> → méthode <code>plot.default</code>	vecteur	vecteur
diagramme quantile-quantile empirique	<code>qqplot(x, y, ...)</code>	vecteur	vecteur
		numérique	numérique

Représentation de plus de trois variables numériques :

- matrice de diagrammes de dispersion : `pairs(~ x + y + z, ...)`
- diagrammes de dispersion superposés : `matplot(matriceX, matriceY, ...)`

## Arguments et paramètres graphiques

- Arguments communs à plusieurs fonctions graphiques :
  - titre et noms d'axes : `main, sub, xlab, ylab`;
  - étendue numérique couverte : `xlim, ylim`;
  - type de représentation : `type`.
- Paramètres graphiques :
  - autour de la zone graphique : `ann, bty`;
  - style des éléments : `lty, lwd, pch`;
  - police de caractères : `font, font.main, font.sub, font.axis, font.lab, family`;
  - grosseur des éléments : `cex, cex.main, cex.sub, cex.axis, cex.lab`;
  - couleur des éléments : `col, col.main, col.sub, col.axis, col.lab, bg`;
  - argument de la fonction `par` uniquement :
    - \* disposition des graphiques : `mfrow` ou `mfcol, new`;
    - \* largeur des marges : `mar, oma`;
    - \* etc.

## Ajout d'éléments à un graphique

Fonction(s) R	Élément(s) ajouté(s)
<code>points</code> et <code>matpoints</code>	points selon des coordonnées
<code>lines</code> et <code>matlines</code>	segments de droites reliant des points
<code>abline</code>	droites traversant toute la zone graphique
<code>segments</code>	segments de droites entre des paires de coordonnées
<code>arrows</code>	flèches entre des paires de coordonnées
<code>rect</code>	rectangles
<code>polygon</code>	polygones quelconques
<code>legend</code>	légende
<code>text</code>	texte dans la zone graphique
<code>mtext</code>	texte dans la marge
<code>title</code>	titre
<code>axis</code>	axe
<code>box</code>	boîte autour de la zone graphique
<code>qqline</code>	ligne dans un graphique quantile-quantile théorique

## Exemple de code avec le système graphique de base

```
# Configuration de paramètres graphiques
par.default <- par(bty="n")
# Initialisation du graphique
plot(x = quakes$mag, y = quakes$stations,
      xlab = "magnitude du séisme",
      ylab = "nombre de stations rapportant le séisme",
      main = "1000 séismes près de Fiji")
# Ajout d'une droite de régression
lmout <- lm(stations ~ mag, data = quakes)
abline(lmout)
# Réattribution des valeurs par défaut
# aux paramètres graphiques
par(par.default)
```

## Possibilités graphiques dans le R de base

Annotations mathématiques :

- Utilisation de la fonction `expression` avec la syntaxe détaillée dans la [fiche d'aide `plotmath`](#) pour créer les valeurs attribuées aux arguments d'éléments textuels dans un graphique (ex. `main`, `xlab`, `ylab`, `labels`, `legend`, `text`, etc.).

## Plusieurs graphiques dans une même fenêtre :

division de la fenêtre en sous-fenêtres

- Sous-fenêtres de tailles égales :
  - arguments `mfrow` ou `mfcol` de la fonction `par`,
- Sous-fenêtres de tailles quelconques :
  - fonction `layout`.

## Aspects techniques des graphiques en R

- **Générer un graphique** : soumettre le programme de création
- **Fenêtre graphique** :
  - tous les graphiques sont créés dans la même fenêtre,
  - RStudio conserve un historique des graphiques produits,
  - une nouvelle fenêtre graphique s'ouvre avec la commande `windows()` (Windows), `quartz()` (Mac) ou `X11()` (Linux).
- **Enregistrer un graphique** :
  - par le menu de la fenêtre graphique ou
  - avec une fonction  `bmp`,  `postscript`,  `pdf`,  `png`,  `tiff`,  `svg` ou  `jpeg`, selon le format désiré.

Pour fermer une connexion avec une fenêtre graphique ou un fichier  
→  `dev.off()`.

## Package `ggplot2`

Principes de base de la grammaire graphique :

Graphique statistique = représentation de **données**, dans un **système de coordonnées** spécifique, divisée en éléments de base :

- **éléments géométriques** (*geoms*) : points, lignes, barres, etc. ;
- **propriétés visuelles** (*aesthetics*) des éléments géométriques : axes, couleurs, formes, tailles, etc.
- **transformations statistiques**, si désiré : courbe de régression ou de lissage, région d'erreur, etc.

Un graphique est spécifié en **associant des variables**, provenant des données, à **des propriétés visuelles** des éléments géométriques du graphique.

## Survol des principales fonctions de `ggplot2`

- `ggplot` : initialisation d'un objet de classe  `ggplot`
- `qplot` : initialisation rapide (*q* pour *quick*) d'un objet  `ggplot` (utilisation similaire à  `plot`, moins de possibilités que  `ggplot`)
- `+` : opérateur pour l'ajout de couches ou la modification de configurations dans un objet  `ggplot`
- fonctions de type  `geom_` :
  - spécification de couches à ajouter à un graphique
- `aes` : création d'un  `mapping`, soit une association entre des propriétés visuelles et des variables
- `ggsave` : enregistrement d'un graphique

**Produire un graphique = afficher un objet  `ggplot`**

## Quelques fonctions de type geom\_

Fonction	Type de graphique	Élément(s) ajouté(s)
<code>geom_point</code>	diagramme de dispersion	points selon des coordonnées
<code>geom_line</code>	diagramme en lignes	segments de droites reliant des points
<code>geom_bar</code>	diagramme en bâtons	barres disjointes, de hauteurs spécifiées ou calculées = fréquences des niveaux d'un facteur
<code>geom_histogram</code>	histogramme	barres collées, de hauteurs calculées = fréquences d'observations d'une variable numérique tombant dans des intervalles joints ( <i>bin</i> )
<code>geom_boxplot</code>	diagramme en boîte	<i>boxplots</i>
<code>geom_density</code>	courbe de densité à noyau	courbe de la densité estimée par noyau ( <i>kernel density</i> )
<code>geom_qq</code>	diagramme quantile-quantile théorique	points pour les couples de quantiles empiriques et théoriques

La plupart de ces fonctions cachent des transformations statistiques (p. ex. `geom_bar` et `geom_histogram` calculent des fréquences, `geom_boxplot` calcule des quantiles, `geom_density` estime une densité, etc.)

## Quelques autres fonctions de ggplot2

Pour ajouter des couches ou modifier des configurations dans un objet `ggplot` initialisé :

- fonctions `ggtitle`, `xlab`, `ylab` :  
ajouter un titre, modifier les noms d'axes ;
- fonctions de type `coord_` :  
modifier des configurations reliées au système de coordonnées ;
- fonctions de type `theme_` :  
modifier des configurations reliées à l'apparence du graphique ;
- fonctions de type `scale_` :  
modifier les échelles de certaines propriétés visuelles (p. ex. couleurs, formes, tailles, etc.)
- fonctions de type `facet_` :  
créer des panneaux (*panel*) de graphiques côte-à-côte
  - chacun des sous-graphique est conditionnel à la valeur de facteurs(s).

Autres : <http://ggplot2.tidyverse.org/reference/>

## Exemple de code avec ggplot2

```
# Chargement du package
library(ggplot2)
# Initialisation d'un objet ggplot
graph <- ggplot(data = quakes,
                 mapping = aes(x = mag, y = stations)) +
  # Ajout de points
  geom_point() +
  # Ajout d'un titre
  ggtitle(label = "1000 séismes près de Fiji") +
  # Ajout d'une droite de régression
  geom_smooth(method='lm')
# Affichage du graphique
graph
```

# Références

## Livres :

- Murrell, P. (2011). R Graphics, Second Edition. CRC Press. URL pour le code R : <https://www.stats.auckland.ac.nz/~paul/RG2e/>
- Chang, W. (2012). R Graphics Cookbook : Practical Recipes for Visualizing Data. O'Reilly Media, Inc. URL pour le code R : <http://www.cookbook-r.com/Graphs/>
- Wickham, H. (2016). ggplot2 : Elegant Graphics for Data Analysis, Second Edition. Springer. URL pour le code R : <https://github.com/hadley/ggplot2-book>
- Wickham, H. et Grolemund, G. (2016). R for Data Science. O'Reilly Media, Inc. Chapitre 3. URL <http://r4ds.had.co.nz/>

(les trois derniers traitent surtout ou exclusivement de ggplot2) :

## Ressources web :

### Système de base :

- <http://www.statmethods.net/graphs/index.html>
- <http://www.statmethods.net/advgraphs/index.html>

### Package lattice :

- <https://CRAN.R-project.org/package=lattice>
- <http://lattice.r-forge.r-project.org/>

### Package ggplot2 :

- <https://CRAN.R-project.org/package=ggplot2>
- <http://ggplot2.tidyverse.org/>
- Tutoriels, rédigés par d'anciens étudiants du cours :
  - Boxplot : [https://stt4230.rbind.io/tutoriels\\_etudiants/hiver\\_2015/boxplot\\_ggplot2/](https://stt4230.rbind.io/tutoriels_etudiants/hiver_2015/boxplot_ggplot2/)
  - Graphique de série temporelle :  
[https://stt4230.rbind.io/tutoriels\\_etudiants/hiver\\_2015/graphique\\_temporel\\_ggplot2/](https://stt4230.rbind.io/tutoriels_etudiants/hiver_2015/graphique_temporel_ggplot2/)
  - Histogramme et courbe de densité de Kernel :  
[https://stt4230.rbind.io/tutoriels\\_etudiants/hiver\\_2015/histogramme\\_ggplot2/](https://stt4230.rbind.io/tutoriels_etudiants/hiver_2015/histogramme_ggplot2/)
  - Diagramme en bâtons : [https://stt4230.rbind.io/tutoriels\\_etudiants/hiver\\_2015/diagramme\\_batons\\_ggplot2/](https://stt4230.rbind.io/tutoriels_etudiants/hiver_2015/diagramme_batons_ggplot2/)
- Tutoriel créé par des employés de la BBC : <https://bbc.github.io/rcookbook/>

## Exemples de graphiques

- <http://www.r-graph-gallery.com/> : exemples de graphiques avec leur code source permettant de les reproduire (tous systèmes, mais beaucoup de ggplot2)

## Autres

- Conseils pour améliorer nos graphiques, illustrés avec ggplot2
  - <http://www.thinkingondata.com/6-tips-to-make-your-visualizations-look-professional/>
- Référence concernant les couleurs en R :
  - <http://research.stowers.org/mcm/efg/R/Color/Chart/>
- Feuilles de triche (cheat sheets) :
  - Système de base
    - \* <http://www.joyce-robbins.com/wp-content/uploads/2016/04/BaseGraphicsCheatsheet.pdf>
    - \* Paramètres graphiques : <http://gastonsanchez.com/visually-enforced/resources/2015/09/22/R-cheat-sheet-graphical-parameters/>

- `ggplot2` :
    - \* <https://www.rstudio.com/wp-content/uploads/2016/11/ggplot2-cheatsheet-2.1.pdf>
    - \* traduction française de ce document :  
<http://thinkr.fr/pdf/ggplot2-french-cheatsheet.pdf>
  - Vidéos :
    - vidéos de Roger Peng concernant `ggplot2` :
      - \* <https://www.youtube.com/watch?v=HeqHMM4ziXA>
      - \* <https://www.youtube.com/watch?v=n8kYa9vu1l8>
  - Revue des packages R ajoutant de nouvelles possibilités graphiques à R :
    - <https://CRAN.R-project.org/view=Graphics>
- 

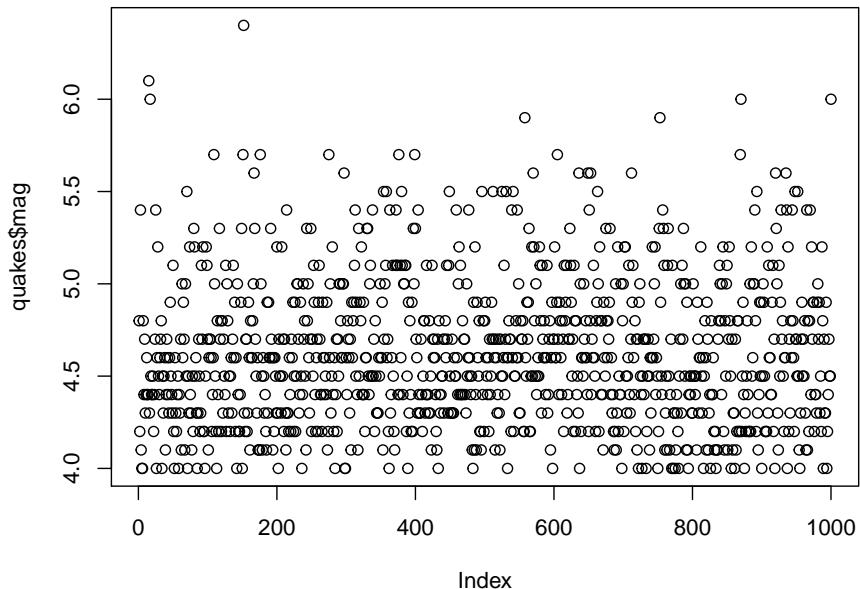
## Annexe

### Graphiques produits par `plot` selon les types d'objets donnés en entrée aux arguments `x` et `y`

Voici quelques essais pour mieux comprendre la fonction `plot`. Appelons la fonction en lui fournissant en entrée des arguments `x` et `y` de différentes classes et observons le résultat. Les données `quakes` sont utilisées ici.

#### Un seul vecteur numérique

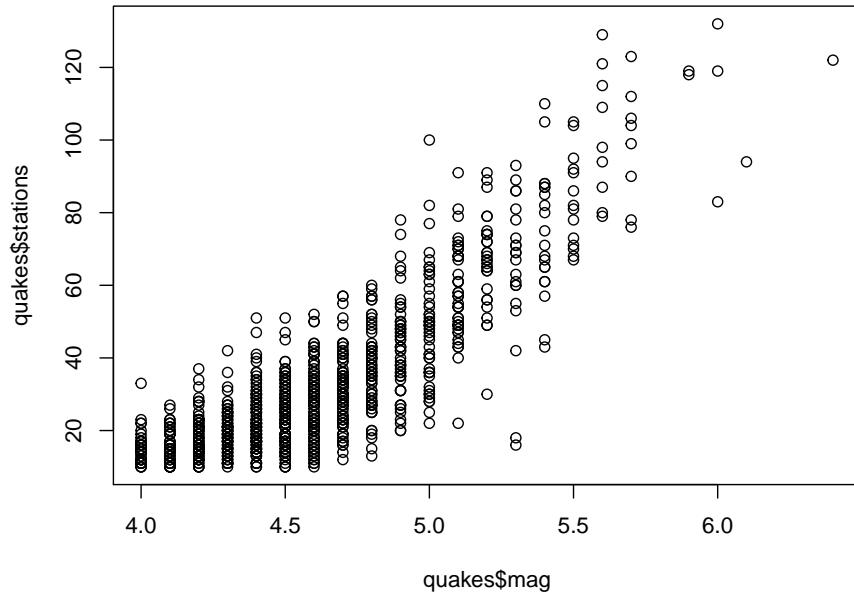
```
plot(x = quakes$mag)
```



Si seulement un vecteur est fourni en entrée, en argument `x`, et qu'aucun argument `y` n'est fourni, alors les observations contenues dans le vecteur sont placées sur l'axe des `y` (ordonnées), et les entiers 1 à `length(x)` sont placés sur l'axe des `x` (abscisses). C'est la méthode `plot.default` qui a tracé ce graphique.

#### Deux vecteurs numériques

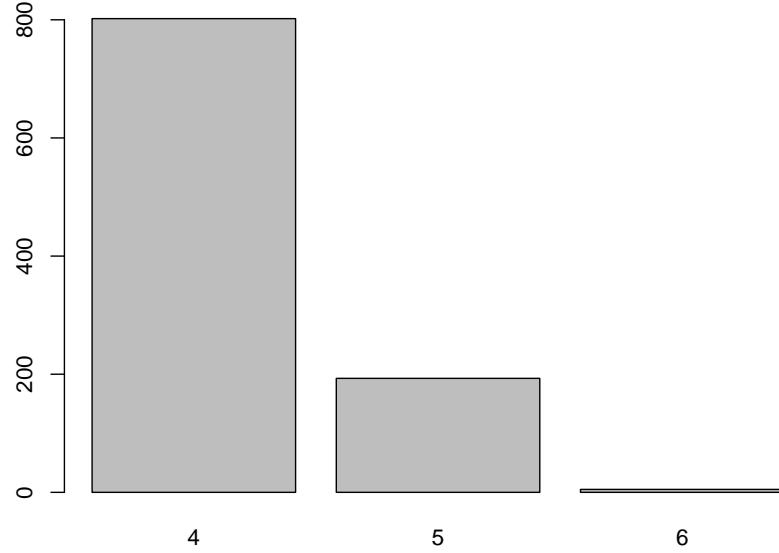
```
plot(x = quakes$mag, y = quakes$stations)
```



Un diagramme de dispersion est produit, encore par la méthode `plot.default`.

### Un seul facteur

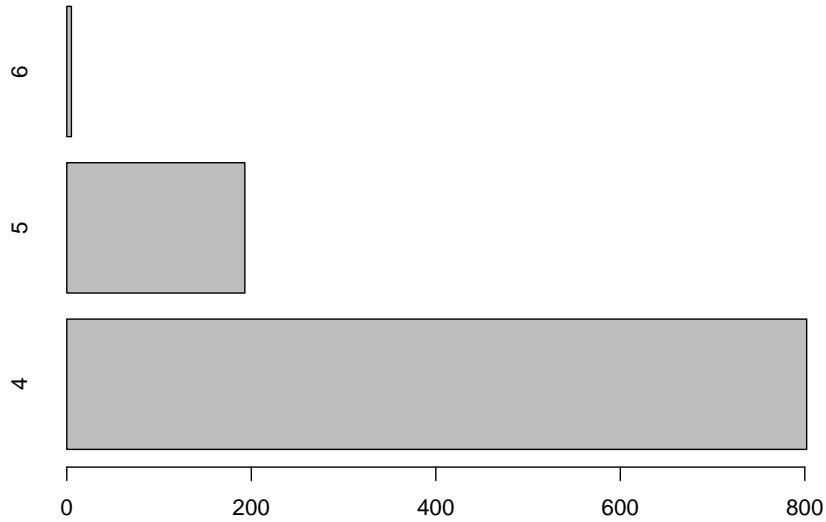
```
plot(x = quakes$magFactor)
```



Si seulement un facteur est fourni en entrée, un diagramme en bâtons est produit. La fonction générique `plot` a commencé par envoyer les arguments qu'elle a reçus à la méthode `plot.factor` parce que la classe de l'argument `x` fourni en entrée était `factor`. Ensuite, la méthode `plot.factor` a choisi d'appeler la fonction `barplot`, parce qu'aucun argument `y` n'a été fourni. En fait, ce sont les fréquences des niveaux du facteur (`table(x)`) qui sont données comme premier argument à `barplot`.

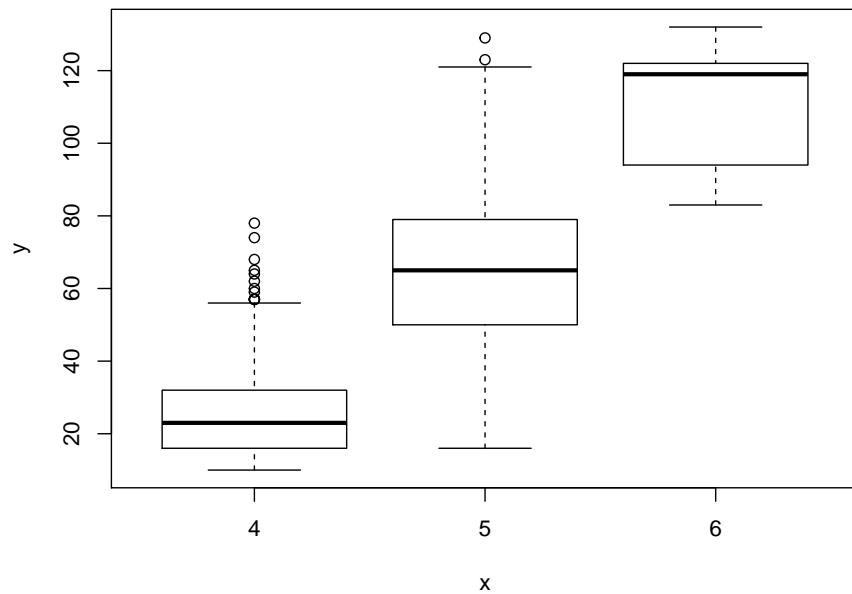
Ainsi, les arguments acceptés par la fonction générique `plot` lorsqu'elle reçoit en entrée comme premier argument (`x`) un facteur sans recevoir d'argument `y` sont les arguments acceptés par la fonction `barplot`. Par exemple, nous pourrions produire un diagramme à bâtons horizontaux comme suit.

```
plot(x = quakes$magFactor, horiz = TRUE)
```



### Un facteur et un vecteur numérique

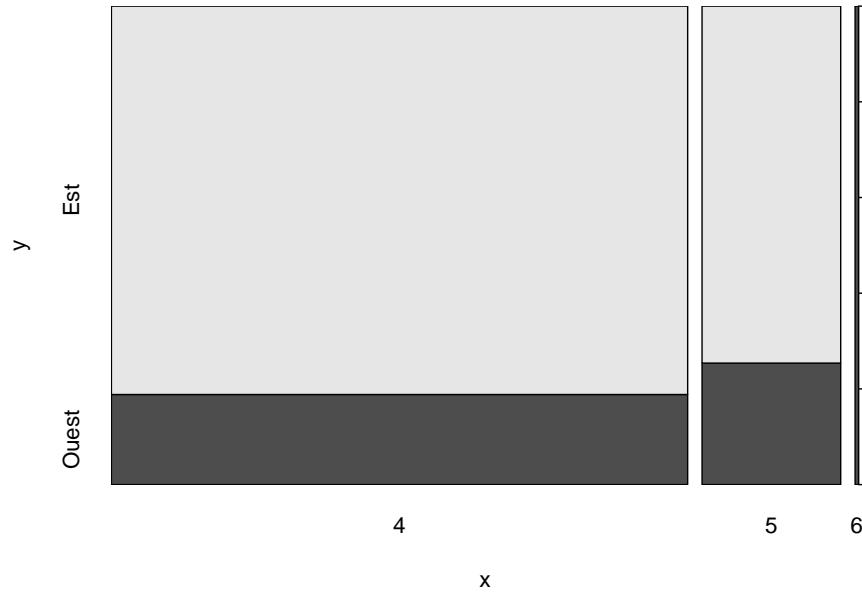
```
plot(x = quakes$magFactor, y = quakes$stations)
```



Si un vecteur numérique est fourni en argument y, avec le facteur fourni en x, la méthode `plot.factor` choisit d'appeler la fonction `boxplot` pour produire des boxplots de la variable y pour chacun des niveaux de la variable x.

### Deux facteurs

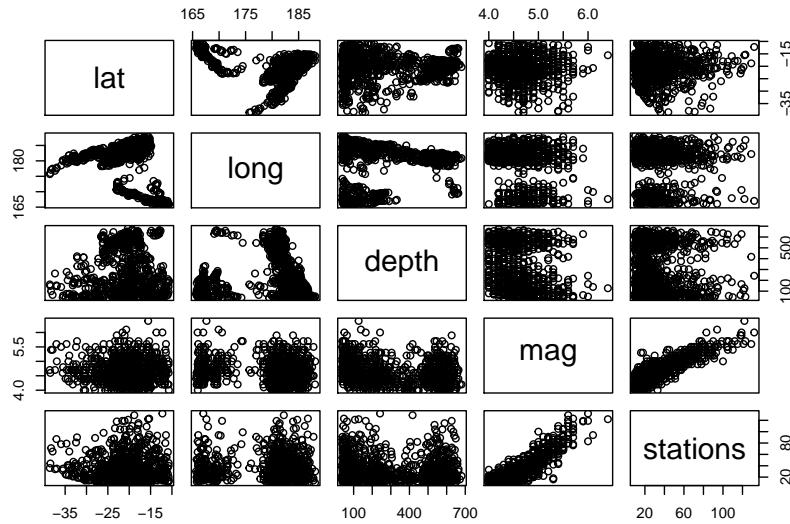
```
plot(x = quakes$magFactor, y = quakes$region)
```



Si deux facteurs sont fournis en entrée, la méthode `plot.factor` choisit d'appeler la fonction `spineplot`. Elle produit un type particulier de diagramme en mosaïque, nommé en anglais *spineplot*.

**Data frame à plus de 2 variables contenant des valeurs numériques ou transformables en valeurs numériques**

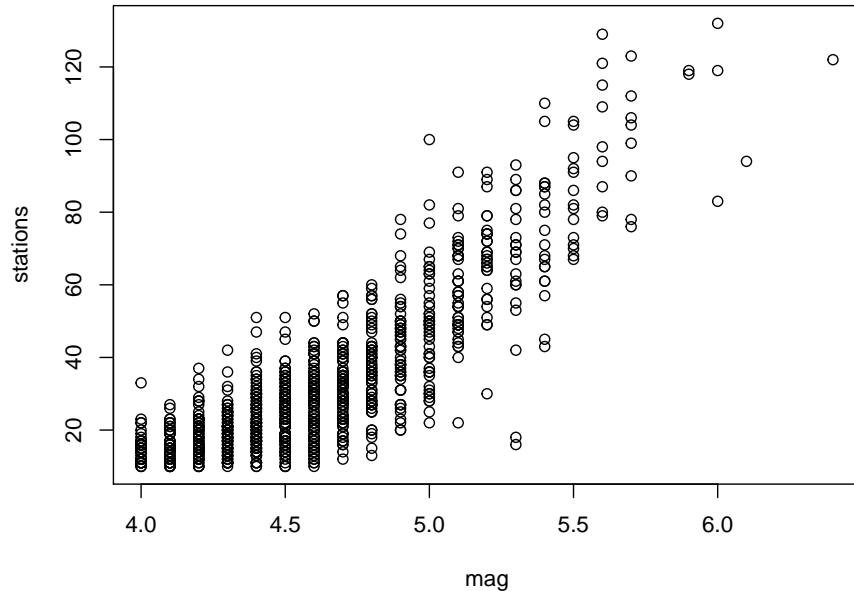
```
# Utilisation du data frame quakes original, sans les deux facteurs ajoutés
plot(datasets::quakes)
```



Si un data frame est fourni en entrée à la fonction générique `plot`, elle délègue son travail à la méthode `plot.data.frame`. Si ce data frame possède plus de 2 variables contenant des valeurs numériques (ou transformables en valeurs numériques), alors la méthode `plot.data.frame` choisit d'appeler la fonction `pairs`. Celle-ci produit une matrice de diagramme de dispersion.

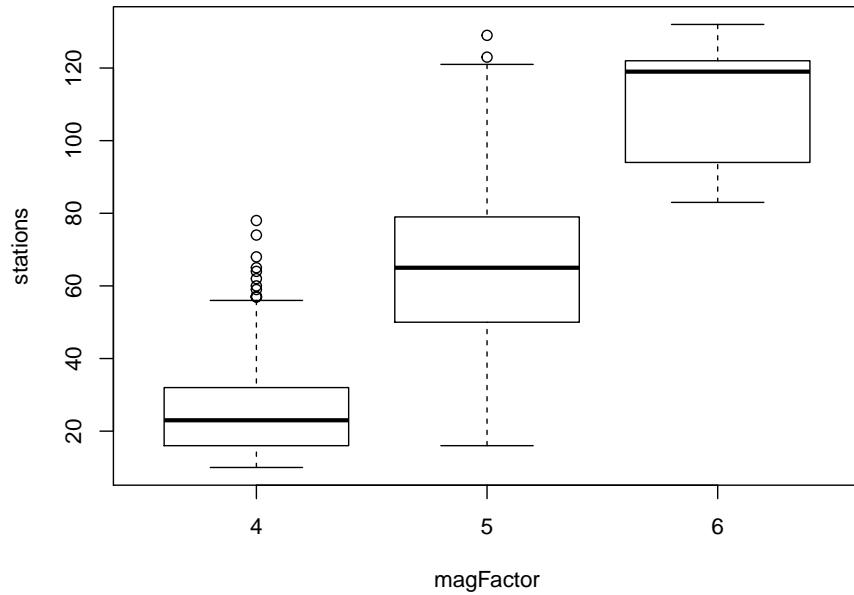
**Data frame à 2 variables**

```
plot(quakes[, c("mag", "stations")])
```



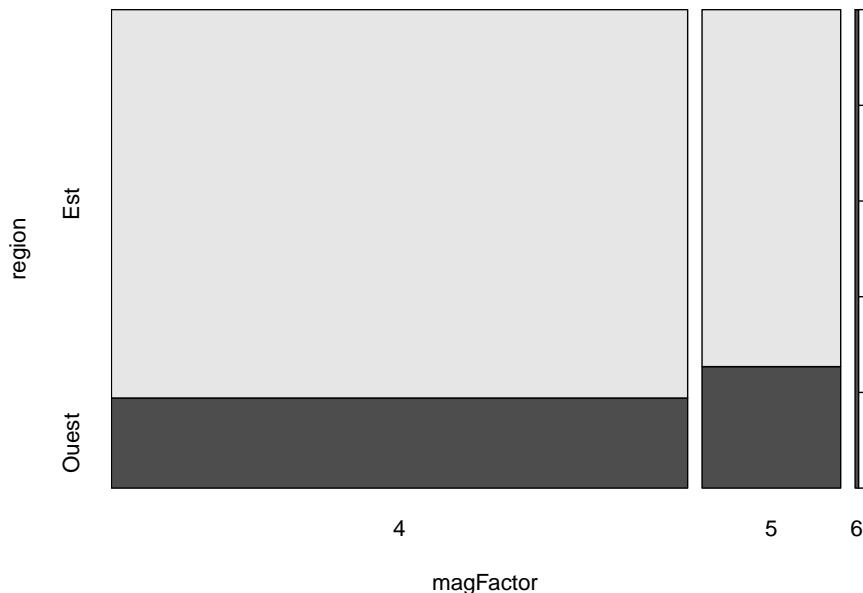
Si un data frame contenant seulement 2 variables est fourni en entré à `plot`, la méthode `plot.data.frame` évalue le type des données dans ces variables. Si les deux variables contiennent des valeurs numériques, la méthode `plot.default` est appelée pour produire un seul diagramme de dispersion. Les observations dans la première colonne sont placées sur l'axe des x et celle de la deuxième colonne sur l'axe des y.

```
plot(quakes[, c("magFactor", "stations")])
```



Si la première colonne est un facteur, des boxplots des observations dans la deuxième colonne, par niveaux du facteur dans la première colonne, sont produits avec la fonction `boxplot`.

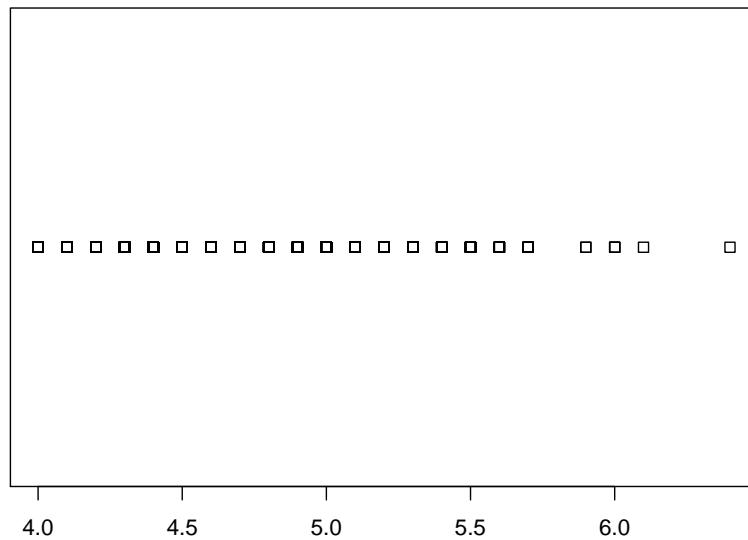
```
plot(quakes[, c("magFactor", "region")])
```



Si les deux variables sont des facteurs, `plot.data.frame` se comporte alors comme `plot.factor` lorsqu'il reçoit deux facteurs.

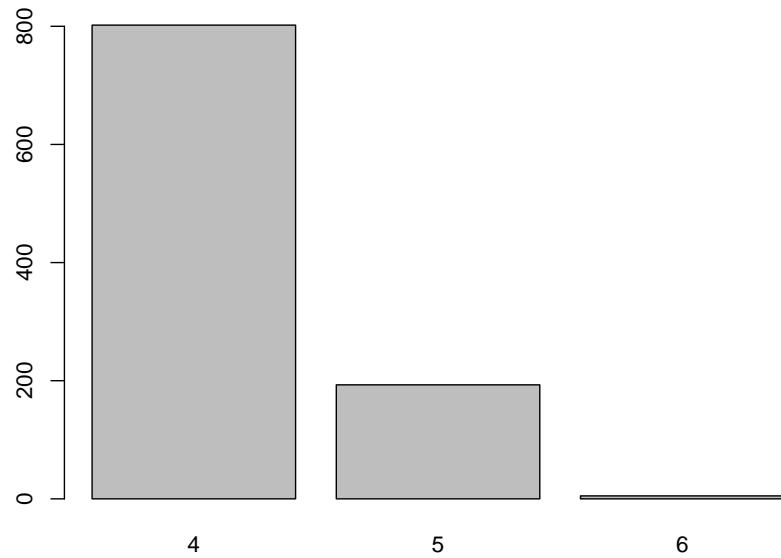
#### Data frame à une colonne

```
plot(quakes[, "mag", drop = FALSE])
```



Si un data frame à seulement une colonne est fourni à `plot` et que cette colonne contient des observations numériques, un diagramme de dispersion à une dimension est produit, en appelant la fonction `stripchart`.

```
plot(quakes[, "magFactor", drop = FALSE])
```



Si la colonne contient plutôt un facteur, c'est un diagramme en bâtons qui est produit par la fonction `barplot`.

**Remarque :** Sans l'argument `drop = FALSE` dans l'extraction d'une colonne de `quakes`, le résultat de l'extraction aurait été un vecteur ou un facteur (un objet à une dimension) plutôt qu'un data frame à une colonne.

```
str(quakes[, "magFactor"])

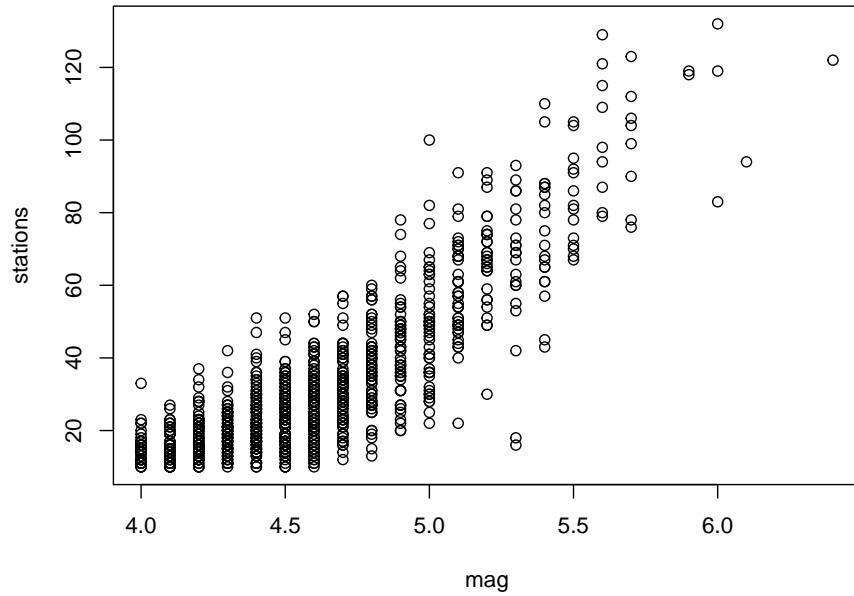
##  Factor w/ 3 levels "4","5","6": 1 1 2 1 1 1 1 1 1 1 ...
str(quakes[, "magFactor", drop = FALSE])

## 'data.frame':   1000 obs. of  1 variable:
## $ magFactor: Factor w/ 3 levels "4","5","6": 1 1 2 1 1 1 1 1 1 ...
```

### Formule en entrée

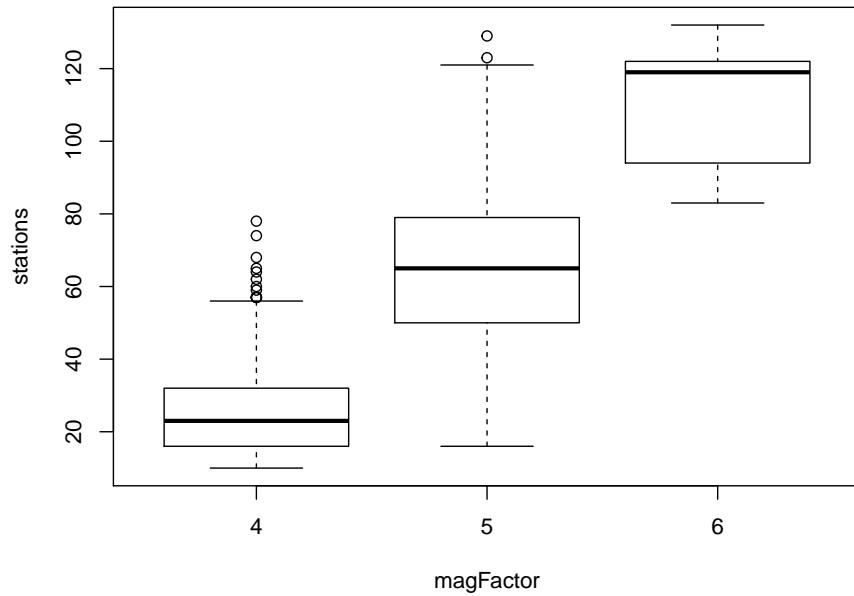
La méthode `plot` accepte aussi une formule en entrée, auquel cas la méthode `plot.formula` est appelée. La méthode `plot.formula` est très versatile. La variable que nous lui fournissons dans la partie de gauche de la formule (à gauche de `~`) est celle positionnée sur l'axe des y. La variable que nous lui fournissons dans la partie de droite de la formule est celle positionnée sur l'axe des x.

```
plot(stations ~ mag, data = quakes)
```



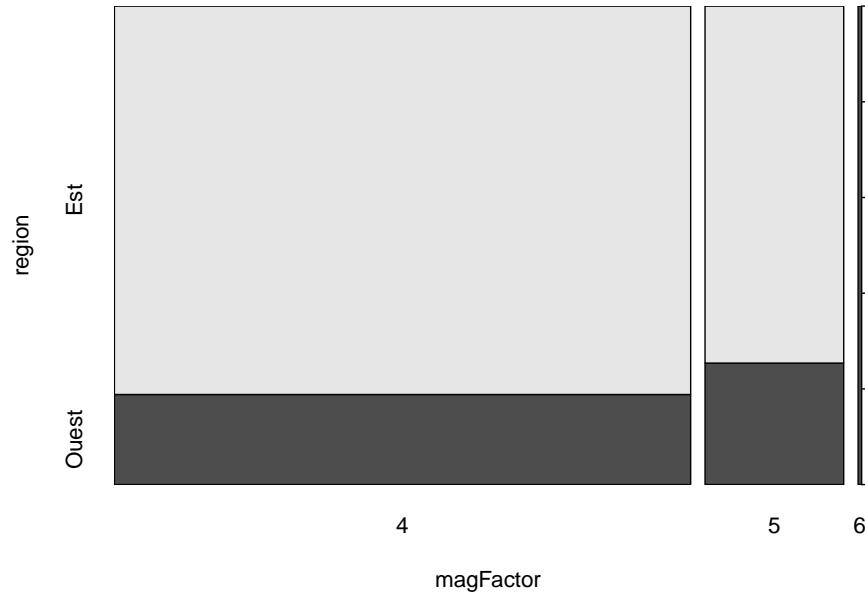
Si les deux variables sont numériques, un diagramme de dispersion est produit. Mais si la variable à droite est un facteur, des boxplots de la variable dans la partie de gauche de la formule, par niveau de ce facteur, sont produits.

```
plot(stations ~ magFactor, data = quakes)
```



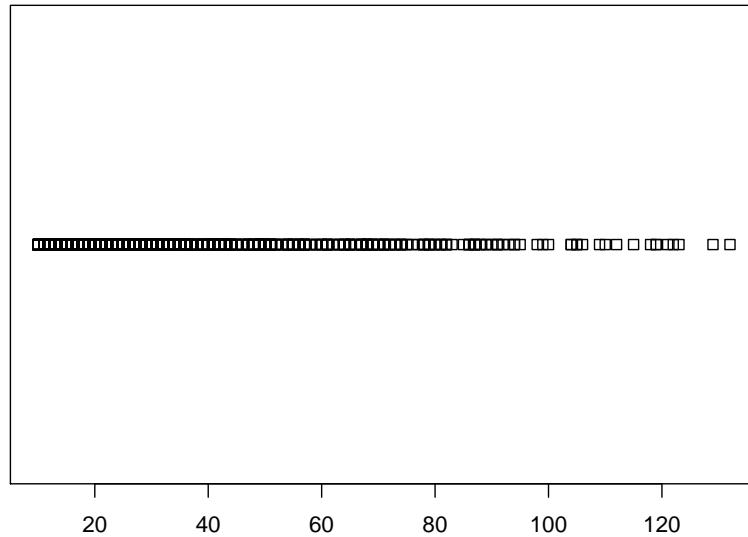
Si les deux variables sont des facteurs, un diagramme en mosaïque est produit.

```
plot(region ~ magFactor, data = quakes)
```



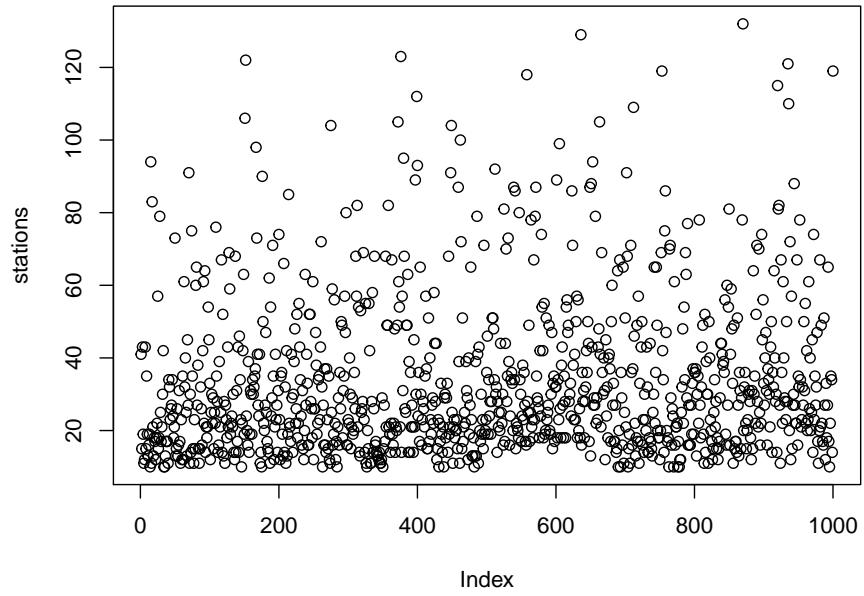
Nous pouvons aussi spécifier seulement une variable dans la formule. Par exemple, avec une seule variable dans la partie de droite de la formule (rien dans la partie de gauche de la formule), nous obtenons le même résultat qu'en fournissant un data frame à une colonne à la fonction générique `plot`.

```
plot(~ stations, data = quakes)
```



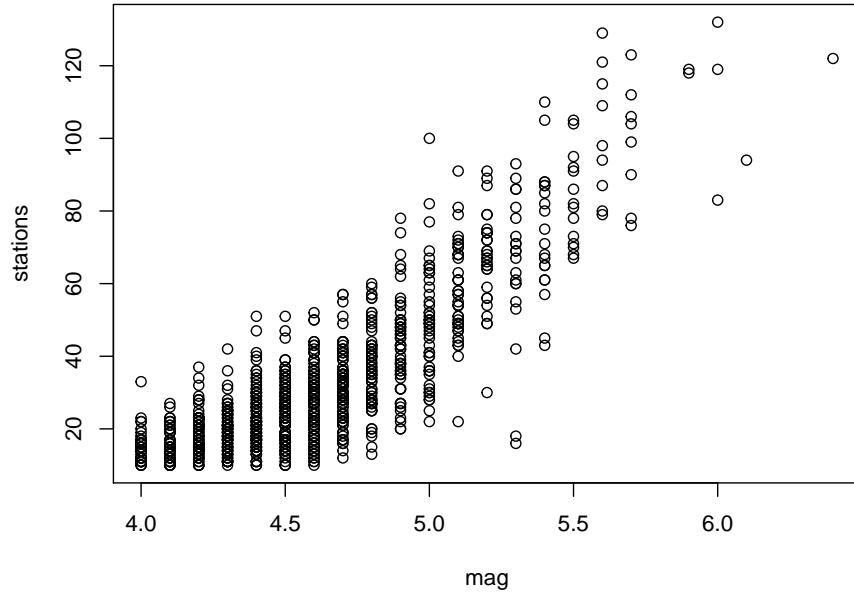
En plaçant plutôt l'unique variable dans la partie de gauche de la formule, avec la valeur 1 dans la partie de droite, nous obtenons le même résultat qu'en fournissant un seul vecteur ou facteur à la fonction générique `plot`.

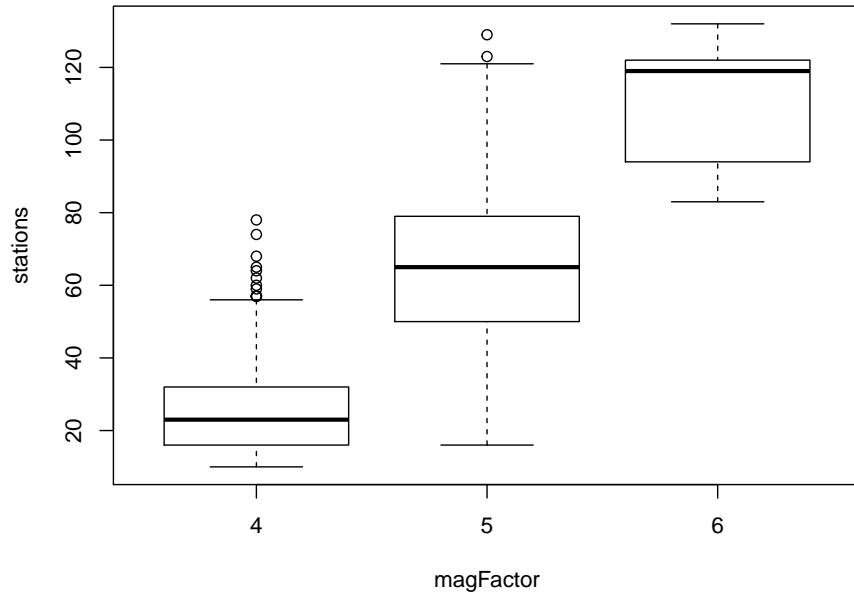
```
plot(stations ~ 1, data = quakes)
```



Si la formule contient une variable dans la partie de gauche et plus d'une variable dans la partie de droite, plus d'un graphique est produit.

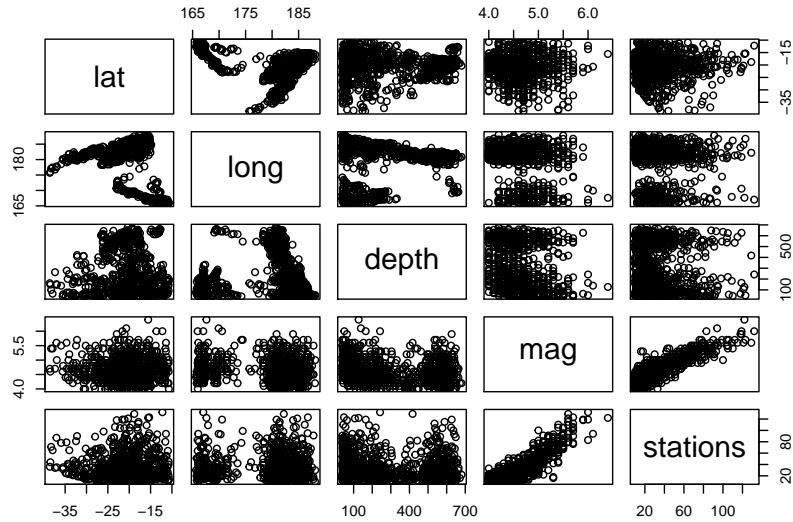
```
plot(stations ~ mag + magFactor, data = quakes)
```





Mais si la formule ne contient pas de variable dans la partie de gauche et plus de deux variables dans la partie de droite, une matrice de diagrammes de dispersion est produite.

```
plot(~ lat + long + depth + mag + stations, data = quakes)
```

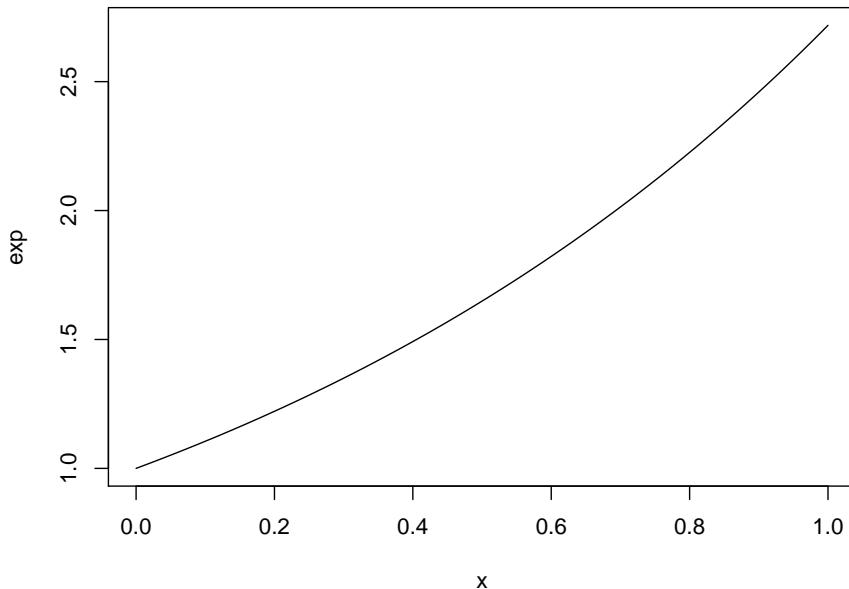


Bref, la méthode `plot.formula` fait appel à une des autres méthodes vues jusqu'à maintenant, en fonction de la position et de la nature (vecteur numérique ou facteur) des variables dans la formule.

### Fonction en entrée

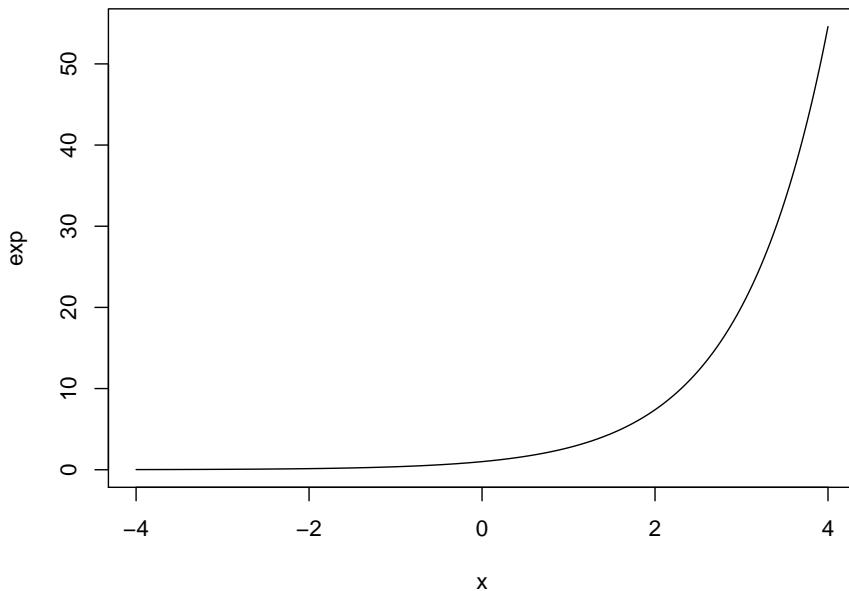
Il y a en R plusieurs fonctions prenant comme premier argument un vecteur numérique et retournant un vecteur de valeurs numériques, par exemple les fonctions R de type mathématique `abs`, `sqrt`, `exp`, `log`, `sin`, `cos`, etc. Ce sont en quelque sorte des représentations de fonctions mathématiques d'une variable. La fonction générique `plot` peut tracer ces fonctions, grâce à sa méthode `plot.function`, après avoir évalué leurs valeurs en certains points.

```
plot(exp)
```



Par défaut, elle fait une évaluation de la valeur de la fonction en 101 points également répartis entre 0 et 1. Il est possible de modifier ces points grâce aux arguments `from`, `to` et `n` (une séquence de points est créée avec la fonction `seq`).

```
plot(exp, from = -4, to = 4, n = 501)
```



En fait, la méthode `plot.function` appelle la fonction `curve`.