

Programmation sous R

Chapitre 2: Graphiques - Statistiques

Mohamed Essaied Hamrita

mhamrita@gmail.com

github.com/Hamrita

Université de Sousse - Tunisie

2023-2024

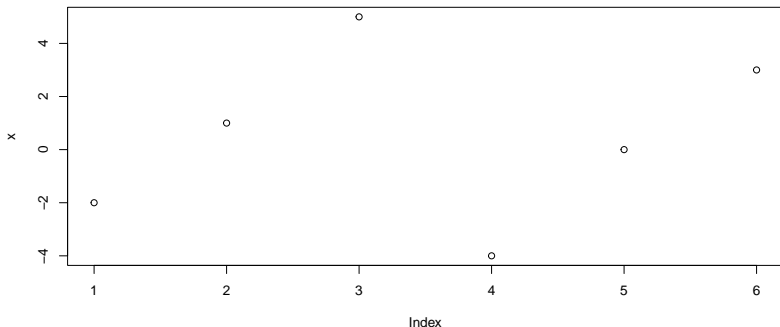
- 1 Graphiques
- 2 Statistique univariée
- 3 Variable qualitative
- 4 Variable quantitative
- 5 Analyse statistique

Graphiques

Graphiques

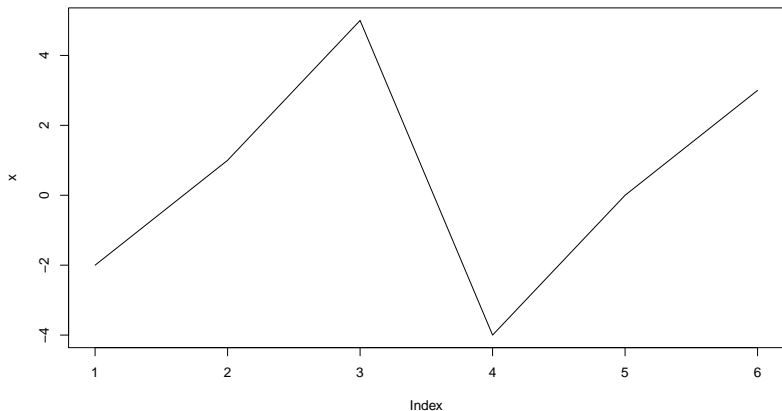
Cette section explique comment créer des types de graphique de base. La commande la plus simple à utiliser pour représenter graphiquement un ensemble de points est la commande `plot(x,y)`. La commande `plot` a plusieurs arguments. Par défaut, cette commande trace l'ensemble des points en points.

```
x=c(-2,1,5,-4,0,3); plot(x)
```



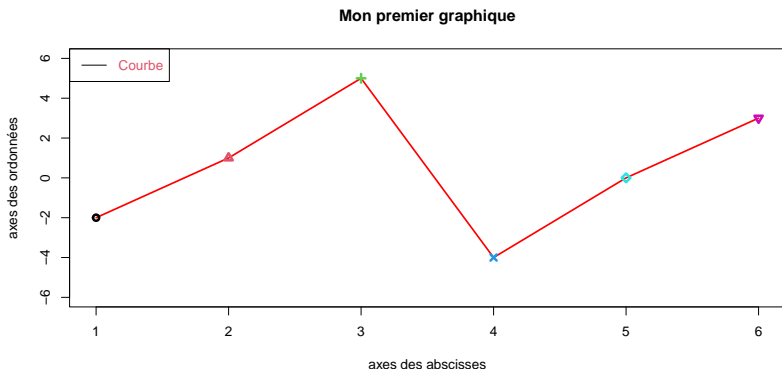
Pour tracer une ligne, on doit ajouter l'argument `type="l"`.

```
plot(x, type="l")
```



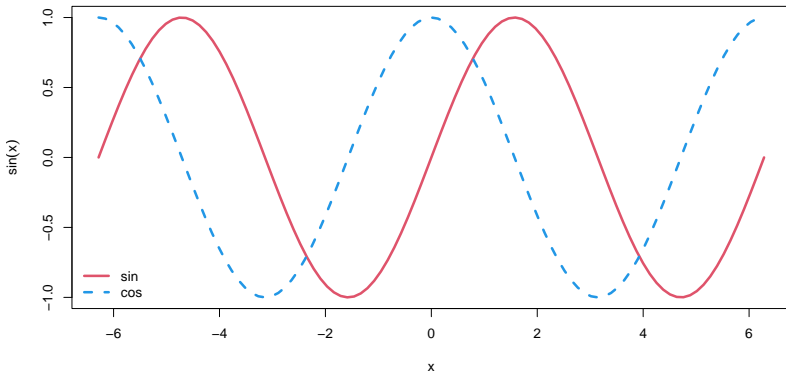
Bien sûr, le logiciel R présente plusieurs arguments pour la fonction `plot`, tels que le paramétrage des couleurs, largeur du trait de la courbe, les étiquettes (labels) des axes, etc ...

```
plot(x,type="l", col="red",lwd=2, xlab="axes des abscisses",
     ylab="axes des ordonnées", main="Mon premier graphique",
     ylim=c(-6,6))
points(x,col=1:6,pch=1:6,lwd=3)
legend("topleft", "Courbe", text.col=2, lty=1)
```



La représentation des courbes des fonctions peut être faite de deux manières; soit par la fonction `plot`, soit par la fonction `curve`.

```
xx=seq(-2*pi, 2*pi, len=100); yy= sin(xx)
plot(xx,yy,type="l",xlab="x", ylab=expression(sin(x)),
      col=2, lwd=3)
curve(cos, -2*pi, 2*pi, col=4, lwd=3, lty=2, add=T)
legend("bottomleft",c("sin", "cos"), lty=c(1,2), col=c(2,4),
      lwd=3,bty="n")
```



Les symboles graphiques

La figure ci-dessous montre les différents types de points:

0 □	1 ○	2 △	3 +	4 ×
5 ◇	6 ▽	7 ⊠	8 ✱	9 ⬠
10 ⊕	11 ⊗	12 ⊞	13 ⊠	14 ⊠
15 ■	16 ●	17 ▲	18 ◆	19 ●

Les types des traits

Le type de traits peut être spécifier en utilisant le paramètre graphique `lty`. Les types de traits disponibles dans R sont :

6. **'twodash'**

- - - - -

5. **'longdash'**

— — — — —

4. **'dotdash'**

. - . - . - . - . - . - . -

3. **'dotted'**

.

2. **'dashed'**

- - - - -

1. **'solid'**

—————

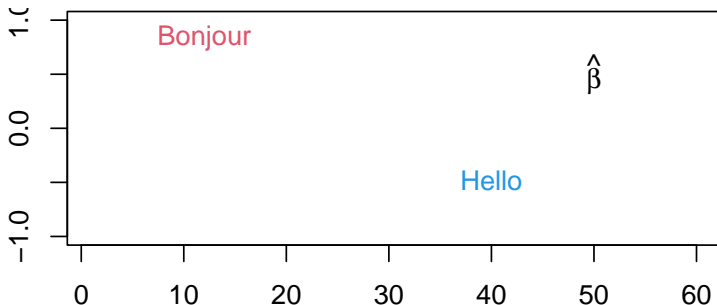
Ajouter un texte

Pour ajouter du texte à un graphique avec le logiciel statistique R, les fonctions `text()` et `mtext()` peuvent être utilisées.

```
text(x,y,label)
```

`x` et `y` sont les coordonnées du texte à ajouter et `label` est le texte à écrire sur le graphique.

```
x1=cos(seq(0,pi,len=60)); plot(x1,type="n", xlab="", ylab="")
text(12, x1[12], "Bonjour", col=2); text(40, x1[40], "Hello", col=4)
text(50,0.5, expression(hat(beta)))
```



ggplot

Une autre manière pour faire la représentation graphique est l'utilisation de la fonction `ggplot` du package `ggplot2` qui doit être installer par la commande `install.package("ggplot2")`. Après l'installation, on fait appel au package à l'aide `library("ggplot2")`.

```
install.packages("ggplot2")  
library("ggplot2")
```

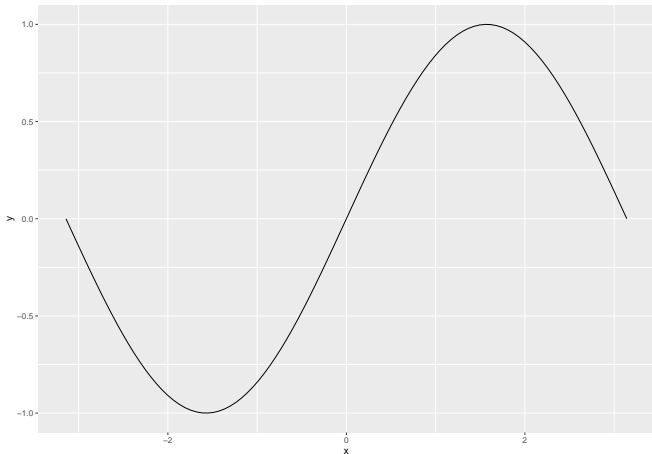
ggplot

Une autre manière pour faire la représentation graphique est l'utilisation de la fonction `ggplot` du package `ggplot2` qui doit être installer par la commande `install.package("ggplot2")`. Après l'installation, on fait appel au package à l'aide `library("ggplot2")`.

```
install.packages("ggplot2")  
library("ggplot2")
```

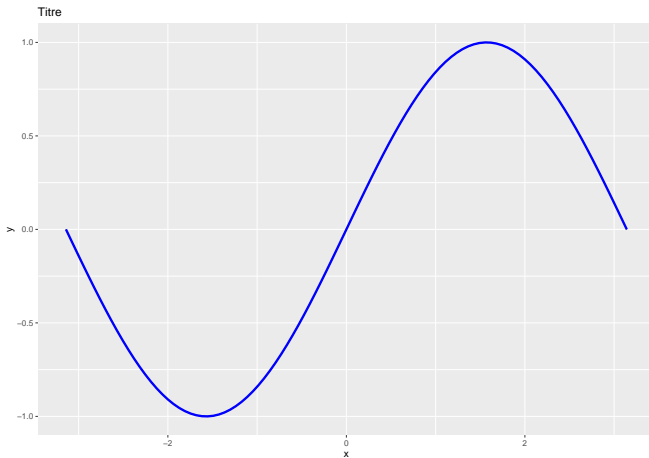
La fonction `ggplot` prend comme un premier argument une `data.frame` qui contient les données à représenter. Un deuxième argument `aes(x,y)` spécifie les valeurs des abscisses et les ordonnées.

```
x=seq(-pi, pi, len=100)
y=sin(x); dd=data.frame(x,y)
p=ggplot(dd,aes(x,y))+geom_line()
p
```



Les paramètres de la largeur et la couleur de la courbe doivent être spécifiés dans `geom_line()`. L'ajout d'un titre se fait par l'ajout de `ggtitle()`.

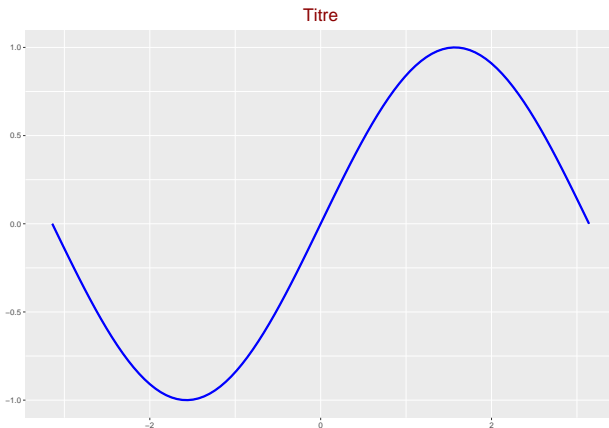
```
p=p+geom_line(linewidth=1.2, colour="blue")+ ggtitle("Titre")  
p
```



Si on veut centrer le titre ou le mettre en couleur ou encore le mettre en gras, on ajoutera `theme(plot.title = element_text(hjust = 0.5, size=20, color="darkred"))`.

```
p=p+theme(plot.title = element_text(hjust = 0.5, size=20,  
                                     color="darkred"))+labs(x="",y="")
```

p



Section 2

Statistique univariée

Statistique univariée

On entend par **statistique univariée** l'étude d'une seule variable, que celle-ci soit **qualitative** ou **quantitative**. La statistique univariée fait partie de la statistique descriptive.

Statistique univariée

On entend par **statistique univariée** l'étude d'une seule variable, que celle-ci soit **qualitative** ou **quantitative**. La statistique univariée fait partie de la statistique descriptive.

Une **variable qualitative** (aussi appelée variable catégorique) réfère à une caractéristique qui n'est pas quantifiable. Une variable catégorique peut être nominale ou ordinale.

Statistique univariée

On entend par **statistique univariée** l'étude d'une seule variable, que celle-ci soit **qualitative** ou **quantitative**. La statistique univariée fait partie de la statistique descriptive.

Une **variable qualitative** (aussi appelée variable catégorique) réfère à une caractéristique qui n'est pas quantifiable. Une variable catégorique peut être nominale ou ordinale.

- Une variable **nominale** décrit un nom, une étiquette ou une catégorie sans ordre naturel. Le sexe est un exemple.
- Une variable **ordinale** est une variable dont les valeurs sont définies par une relation d'ordre entre les catégories possibles. La variable mention est une variable ordinale parce que la catégorie "très bien" est meilleure que la catégorie "bien" qui est meilleure de la catégorie "passable".

Statistique univariée

On entend par **statistique univariée** l'étude d'une seule variable, que celle-ci soit **qualitative** ou **quantitative**. La statistique univariée fait partie de la statistique descriptive.

Une **variable qualitative** (aussi appelée variable catégorique) réfère à une caractéristique qui n'est pas quantifiable. Une variable catégorique peut être nominale ou ordinale.

- Une variable **nominale** décrit un nom, une étiquette ou une catégorie sans ordre naturel. Le sexe est un exemple.
- Une variable **ordinale** est une variable dont les valeurs sont définies par une relation d'ordre entre les catégories possibles. La variable mention est une variable ordinale parce que la catégorie "très bien" est meilleure que la catégorie "bien" qui est meilleure de la catégorie "passable".

Une **variable quantitative** est une caractéristique quantifiable dont les valeurs sont des nombres. Les variables numériques peuvent être **continues** ou **discrètes**.

Statistique univariée

On entend par **statistique univariée** l'étude d'une seule variable, que celle-ci soit **qualitative** ou **quantitative**. La statistique univariée fait partie de la statistique descriptive.

Une **variable qualitative** (aussi appelée variable catégorique) réfère à une caractéristique qui n'est pas quantifiable. Une variable catégorique peut être nominale ou ordinale.

- Une variable **nominale** décrit un nom, une étiquette ou une catégorie sans ordre naturel. Le sexe est un exemple.
- Une variable **ordinale** est une variable dont les valeurs sont définies par une relation d'ordre entre les catégories possibles. La variable mention est une variable ordinale parce que la catégorie "très bien" est meilleure que la catégorie "bien" qui est meilleure de la catégorie "passable".

Une **variable quantitative** est une caractéristique quantifiable dont les valeurs sont des nombres. Les variables numériques peuvent être **continues** ou **discrètes**.

- Variables continues: On dit qu'une variable est **continue** si elle prend un nombre infini de valeurs réelles possibles à l'intérieur d'un intervalle donné. Prenons la taille d'un élève par exemple.

Statistique univariée

On entend par **statistique univariée** l'étude d'une seule variable, que celle-ci soit **qualitative** ou **quantitative**. La statistique univariée fait partie de la statistique descriptive.

Une **variable qualitative** (aussi appelée variable catégorique) réfère à une caractéristique qui n'est pas quantifiable. Une variable catégorique peut être nominale ou ordinale.

- Une variable **nominale** décrit un nom, une étiquette ou une catégorie sans ordre naturel. Le sexe est un exemple.
- Une variable **ordinale** est une variable dont les valeurs sont définies par une relation d'ordre entre les catégories possibles. La variable mention est une variable ordinale parce que la catégorie "très bien" est meilleure que la catégorie "bien" qui est meilleure de la catégorie "passable".

Une **variable quantitative** est une caractéristique quantifiable dont les valeurs sont des nombres. Les variables numériques peuvent être **continues** ou **discrètes**.

- Variables continues: On dit qu'une variable est **continue** si elle prend un nombre infini de valeurs réelles possibles à l'intérieur d'un intervalle donné. Prenons la taille d'un élève par exemple.
- Variables discrètes: Contrairement à une variable continue, une variable **discrète** ne peut prendre qu'un nombre fini de valeurs réelles possibles à l'intérieur d'un intervalle donné. Le nombre d'enfants dans un ménage est un exemple.

Section 3

Variable qualitative

Variable qualitative

Une variable qualitative peut être représentée, soit par un diagramme à barres, soit par un diagramme en secteurs.

Variable qualitative

Une variable qualitative peut être représentée, soit par un diagramme à barres, soit par un diagramme en secteurs.

Exemple:

En 2005, les recettes du budget de l'État se présentaient de la façon suivante (en milliards) :

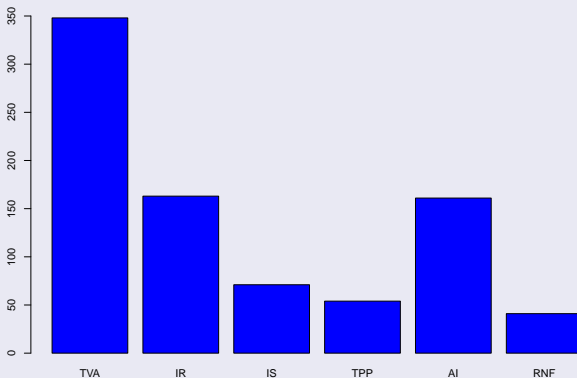
Source	TVA	IR	IS	TPP	AI	RNF
RF	348	163	71	54	161	41

Le graphique à barres est parfois appelé graphique à bandes ou graphique à bâtons. Il peut être horizontal ou vertical.

Graphique à barres

Diagramme à barres vertical

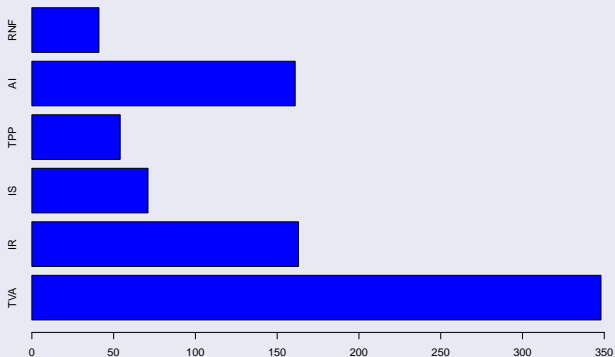
```
RF=c(348,163,71,54,161,41)
names=c("TVA","IR","IS","TPP","AI","RNF")
barplot(RF, names.arg=names, col="blue", ylim = c(0,350))
```



Graphique à barres

Diagramme à barres horizontal

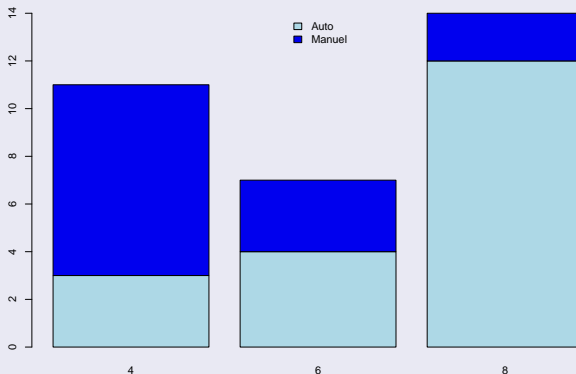
```
barplot(RF, names.arg=names, col="blue", horiz = T, xlim = c(0,350) )
```



Graphique à barres

Diagramme à barres groupées

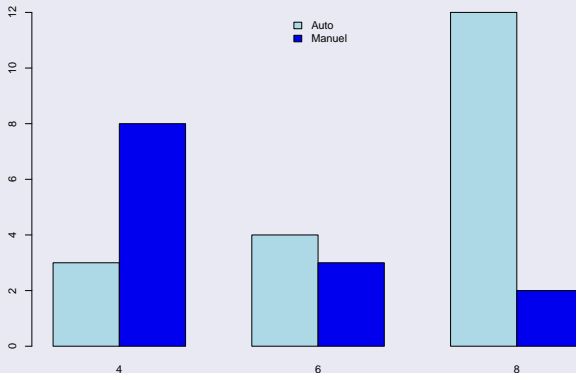
```
cyl=factor(mtcars$cyl)
am=factor(mtcars$am, labels = c("Auto", "Manuel"))
tab=table(am, cyl);barplot(tab, col=c("lightblue","blue2"))
legend("top", c("Auto","Manuel"), fill=c("lightblue","blue2"), box.lty = 0)
```



Graphique à barres

Diagramme à barres groupées

```
barplot(tab, col=c("lightblue","blue2"), beside=T)  
legend("top", c("Auto","Manuel"), fill=c("lightblue","blue2"), box.lty = 0)
```

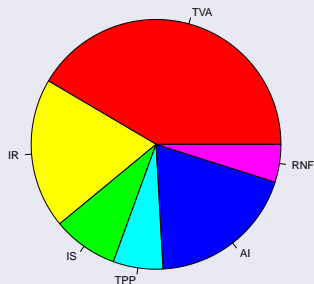


Graphique circulaire

Ce type de graphique est formé d'un cercle divisé en secteurs. Chaque secteur représente une catégorie particulière.

Diagramme simple

```
pie(RF, names, col=rainbow(length(RF)))
```

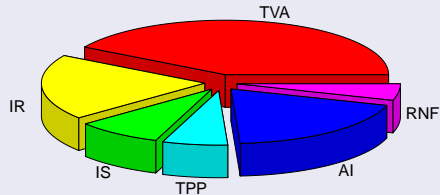


Graphique circulaire

Ce type de graphique est formé d'un cercle divisé en secteurs. Chaque secteur représente une catégorie particulière.

Diagramme 3D

```
plotrix::pie3D(RF, labels=names, explode=0.1)
```



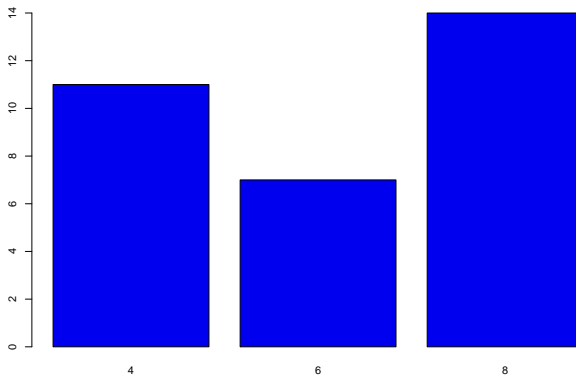
Section 4

Variable quantitative

Variable quantitative discrète

Une variable quantitative discrète peut être représentée graphiquement par un graphique en bâton. Ce graphique s'obtient avec la fonction `barplot()`.

```
barplot(table(mtcars$cyl), col="blue2")
```



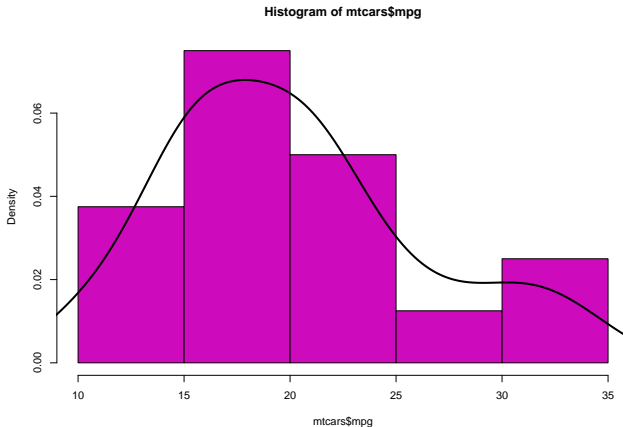
Variable quantitative continue

Une variable quantitative continue peut être représentée par un histogramme. Il est souvent employé pour montrer les caractéristiques principales de la distribution des données de façon pratique. Ce type de graphique s'obtient sous R à l'aide de la fonction `boxplot`.

Variable quantitative continue

Une variable quantitative continue peut être représentée par un histogramme. Il est souvent employé pour montrer les caractéristiques principales de la distribution des données de façon pratique. Ce type de graphique s'obtient sous R à l'aide de la fonction `boxplot`.

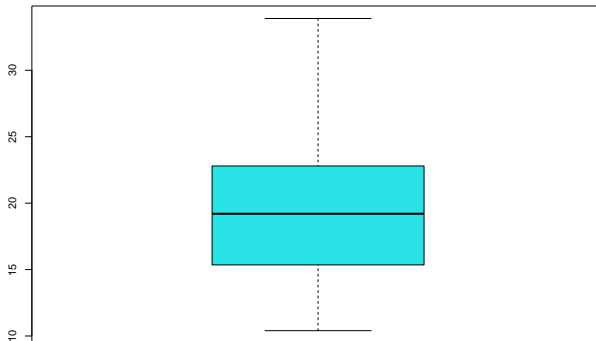
```
hist(mtcars$mpg, col=6, probability = T )
lines(density(mtcars$mpg), lwd=3)
```



Variable quantitative continue

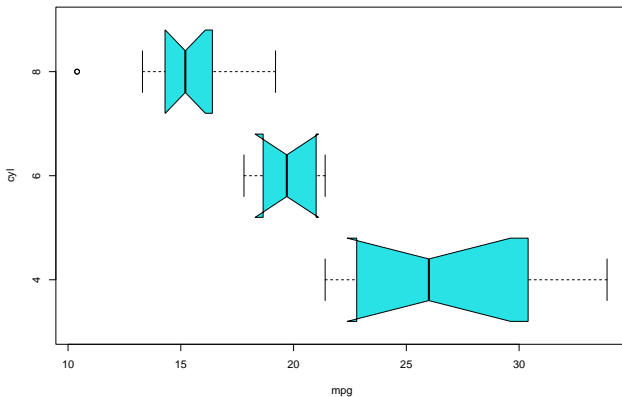
On peut aussi résumer une variable statistique quantitative par un diagramme appelé boîte à moustache (boxplot).

```
boxplot(mtcars$mpg, col=5)
```



Variable quantitative continue

```
boxplot(mpg ~ cyl , data=mtcars, col=5, notch=T, horizontal = T)
```



Section 5

Analyse statistique

Variable qualitative

Pour une variable qualitative, on peut déterminer un tableau des fréquences (absolues ou relatives).

```
(t1=table(mtcars$cyl)) # absolues
```

```
 4  6  8
11  7 14
```

```
t1/sum(t1) # relatives
```

```
      4      6      8
0.34375 0.21875 0.43750
```

```
(t2=table(mtcars$cyl, mtcars$gear )); t2/sum(t2)
```

```
      3  4  5
4  1  8  2
6  2  4  1
8 12  0  2
```

```
      3      4      5
4 0.03125 0.25000 0.06250
6 0.06250 0.12500 0.03125
8 0.37500 0.00000 0.06250
```

Variable quantitative

Une première analyse statistique pour une variable quantitative est l'analyse descriptive.

```
summary(mtcars$mpg)
```

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
10.40	15.43	19.20	20.09	22.80	33.90

```
pastecs::stat.desc(mtcars$mpg, norm=T, basic=F )
```

median	mean	SE.mean	CI.mean.0.95	var	std.dev
19.2000000	20.0906250	1.0654240	2.1729465	36.3241028	6.0269481
coef.var	skewness	skew.2SE	kurtosis	kurt.2SE	normtest.W
0.2999881	0.6106550	0.7366922	-0.3727660	-0.2302812	0.9475647
normtest.p					
0.1228814					