

Exploration des données en R

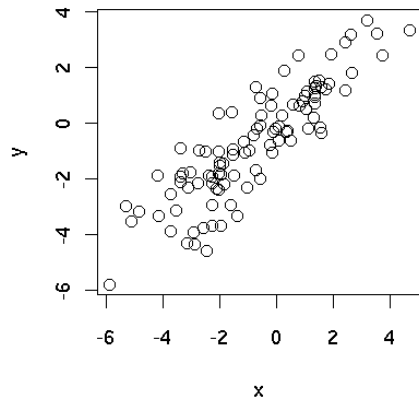
Christophe Pallier et Christophe Lalanne

`christophe.pallier@m4x.org / christophe.lalanne@gmx.net`

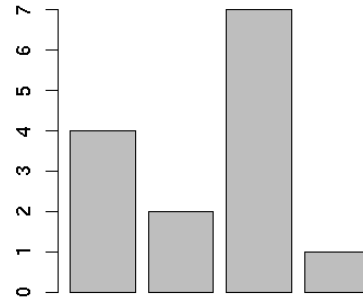
Mastère de Sciences Cognitives, EHESS – Paris 5 – ENS

fonctions graphiques

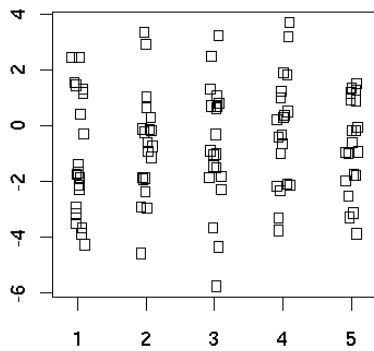
plot(x,y)



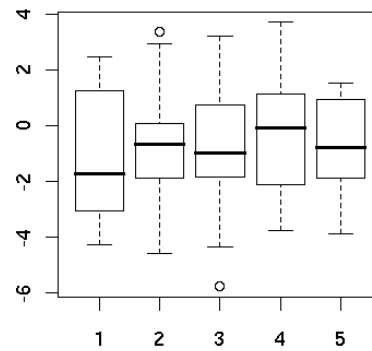
barplot(c(4,2,7,1))



stripchart(y~b)



boxplot(y~b)





fonctions graphiques

1. Observer les résultats des lignes suivantes:

```
a=rnorm(100,mean=100, sd=10)
```

```
plot(a)
```

```
boxplot(a)
```

```
hist(a)
```

```
rug(a)
```

```
hist(a,breaks=20)
```

```
plot(density(a))
```

```
qqnorm(a)
```

2. Jouer avec la taille de l'échantillon aléatoire (10, 20, 1000).
3. Essayer d'autres lois: runif, rbinom, rcauchy
4. Afficher, côte à côte, 20 boxplots d'échantillons normaux ($N(0,1)$) de 10 éléments, puis de 100 éléments, puis de 1000 éléments
5. Créer un vecteur "mixture" (50 échantillons $N(0,1)$ et 50 échantillons $N(2,3)$) et afficher son histogramme et sa densité estimée.



Exploration de données

```
data(iris)
help(iris)
head(iris)
summary(iris)
attach(iris)
plot(iris)
plot(iris, col=as.numeric(iris$Species))
boxplot(Petal.Length)
boxplot(iris[,1:4])
boxplot(Petal.Width~Species)
```



Exploration de données

```
sujet=gl(10,10)
condition=gl(5,2,100)
y=rnorm(100,mean=10)+rep(rnorm(10,sd=2),rep(10,10))+ \
    rep(rep(rnorm(5),2),10)

d<-data.frame(sujet,condition,y)
d
table(a,b)
par(mfrow=c(2,2))
plot(y~sujet)
barplot(tapply(y,sujet,mean))
plot(y~condition)
barplot(tapply(y,condition,mean))
x11()
interaction.plot(condition,sujet,y)
l=aov(y~condition+Error(sujet/condition))
summary(l)
```



Résumé des commandes graphiques de base

- univarié:
 - continue: `hist(x)`, `rug(x)`, `plot(density(x))`, `boxplot(x)`
 - discrete: `barplot(table(x))`
- bivarié (2 variables)
 - continues: `plot(x, y)`, `qqplot(x, y)`
 - discrete: `mosaicplot(x, y)`
 - 1 continue, 1 discrete: `boxplot(x ~ a)`
- multivariées: `matplot`, `pairs`, `ggobi`