

# Découverte de

**David Causeur**

*Laboratoire de Mathématiques Appliquées*

*Agrocampus Ouest*

*IRMAR CNRS UMR 6625*

*<http://www.agrocampus-ouest.fr/math/causeur/>*

# Plan du cours

- 1 R, Logiciel libre
- 2 Présentation
- 3 Importation
- 4 Manipulation
- 5 Description graphique
- 6 Traitements automatiques

# Le système GNU

Système GNU (extrait de <http://www.gnu.org/home.fr.html>)

## Qu'est-ce que GNU ?

Le projet GNU a été lancé en 1984 pour développer un système d'exploitation complet de style Unix, qui soit un [logiciel libre](#) : le système GNU.

# Le système GNU

Système GNU (extrait de <http://www.gnu.org/home.fr.html>)

Une alternative aux propositions commerciales

Logiciel libre (extrait de <http://www.gnu.org/home.fr.html>)

## Qu'est-ce qu'un logiciel libre ?



Un « [logiciel libre](#) » se réfère à la liberté, pas au prix. Pour comprendre le concept, vous devez penser à « libre » comme dans « liberté d'expression », pas comme dans « bière gratuite ». (NdT : en anglais, le mot « free » veut dire libre, mais aussi gratuit d'où la confusion possible).

Les logiciels libres donnent aux utilisateurs **la liberté d'exécuter, de copier, de distribuer, d'étudier, de modifier et d'améliorer** les logiciels. Plus précisément, cela se réfère à quatre types de liberté pour les utilisateurs de logiciels :

# R, environnement libre d'analyse statistique

Téléchargement à partir de <http://www.r-project.org>



## About R

[What is R?](#)  
[Contributors](#)  
[Screenshots](#)  
[What's new?](#)

## Download, Packages

[CRAN](#)

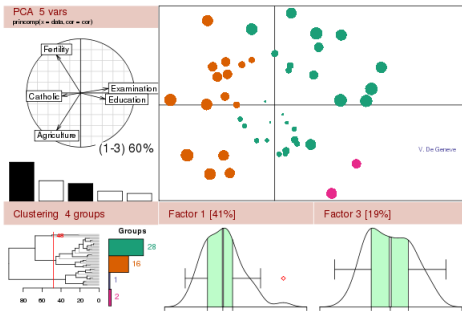
## R Project

[Foundation](#)  
[Members & Donors](#)  
[Mailing Lists](#)  
[Bug Tracking](#)  
[Developer Page](#)  
[Conferences](#)  
[Search](#)

## Documentation

[Manuals](#)  
[FAQs](#)  
[The R Journal](#)

## The R Project for Statistical Computing



## Getting started:

- R is a free software environment for statistical computing and graphics. It compiles and runs on a wide variety of UNIX platforms, Windows and MacOS. To [download R](#), please choose your preferred [CRAN mirror](#).

# R, environnement libre d'analyse statistique

Téléchargement à partir de <http://www.r-project.org>  
Le réseau CRAN



## CRAN

[Mirrors](#)  
[What's new?](#)  
[Task Views](#)  
[Search](#)

## About R

[R Homepage](#)  
[The R Journal](#)

## Software

[R Sources](#)  
[R Binaries](#)  
[Packages](#)  
[Other](#)

## Documentation

[Manuals](#)  
[FAQs](#)  
[Contributed](#)

## The Comprehensive R Archive Network

### Frequently used pages

#### Download and Install R

Precompiled binary distributions of the base system and contributed packages, **Windows and Mac** users most likely want one of these versions of R:

- [Linux](#)
- [MacOS X](#)
- [Windows](#)

#### Source Code for all Platforms


Windows and Mac users most likely want the precompiled binaries listed in the upper box, not the source code. The sources have to be compiled before you can use them. If you do not know what this means, you probably do not want to do it!


- **The latest release** (2010-04-22): [R-2.11.0.tar.gz](#) (read [what's new](#) in the latest version).
- Sources of [R alpha and beta releases](#) (daily snapshots, created only in time periods before a planned release).
- Daily snapshots of current patched and development versions are [available here](#). Please read about [new features and bug fixes](#) before filing corresponding feature requests or bug reports.
- Source code of older versions of R is [available here](#).
- Contributed extension [packages](#)

# Plan du cours

- 1 R, Logiciel libre
- 2 Présentation**
- 3 Importation
- 4 Manipulation
- 5 Description graphique
- 6 Traitements automatiques

# Environnement R

Démarrage de 

 R Console


```
R version 2.10.0 (2009-10-26)
Copyright (C) 2009 The R Foundation for Statistical Computing
ISBN 3-900051-07-0
```

```
R est un logiciel libre livré sans AUCUNE GARANTIE.
Vous pouvez le redistribuer sous certaines conditions.
Tapez 'license()' ou 'licence()' pour plus de détails.
```

```
R est un projet collaboratif avec de nombreux contributeurs.
Tapez 'contributors()' pour plus d'information et
'citation()' pour la façon de le citer dans les publications.
```

```
Tapez 'demo()' pour des démonstrations, 'help()' pour l'aide
en ligne ou 'help.start()' pour obtenir l'aide au format HTML.
Tapez 'q()' pour quitter R.
```


# Environnement R

Démarrage de 

Session de travail : opérations effectuées entre le démarrage et la sortie de R

- Entrées/sorties de données
- Production de tableaux de résultats
- Production de graphiques

# Environnement R

Démarrage de 

Session de travail : opérations effectuées entre le démarrage et la sortie de R

- Entrées/sorties de données
- Production de tableaux de résultats
- Production de graphiques

Cadre minimal de travail

- Console : cœur de la session de travail
- Script(s) : espace de rédaction des commandes

# Langage R

## Principe général

> sortie = opération(entrée)

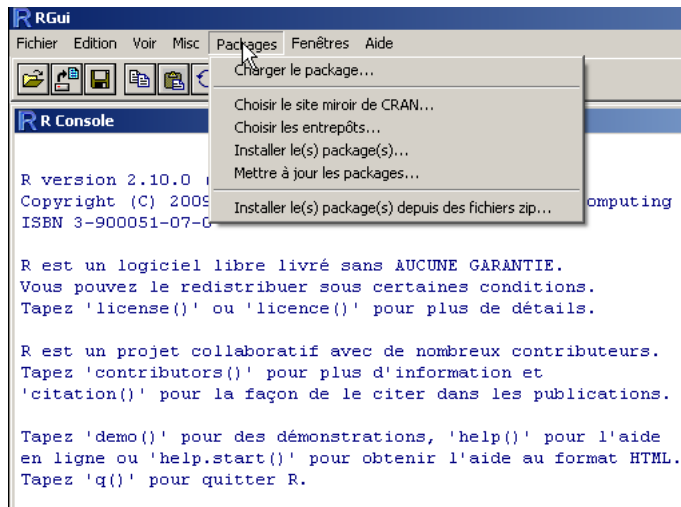
## Exemples

- `help.start()`
- `x=2`
- `y=sqrt(x)`
- `y`
- `objects()`

**Exercice** : Créer un script `graphsqrtr.R` qui produit le tracé de la fonction  $f(x) = \sqrt{x}$  pour  $0 \leq x \leq 20$

# Modulation des fonctionnalités de R

## Modules de fonctions : Packages



The screenshot shows the RGui application window. The title bar reads 'RGui'. The menu bar includes 'Fichier', 'Edition', 'Voir', 'Misc', 'Packages', 'Fenêtres', and 'Aide'. The 'Packages' menu is open, displaying the following options:

- Charger le package...
- Choisir le site miroir de CRAN...
- Choisir les entrepôts...
- Installer le(s) package(s)...
- Mettre à jour les packages...
- Installer le(s) package(s) depuis des fichiers zip...

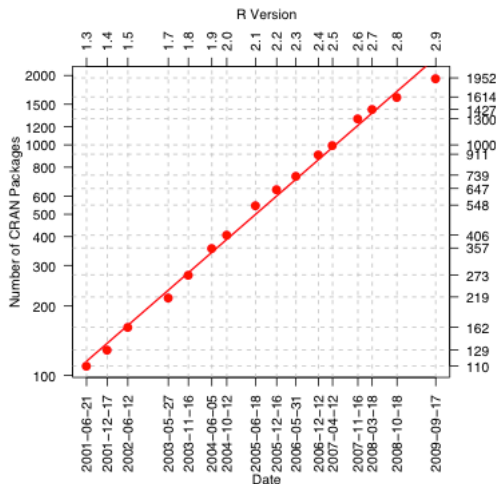
The 'R Console' pane is visible below the menu, displaying the following text:

```
R version 2.10.0  
Copyright (C) 2009  
ISBN 3-900051-07-0  
  
R est un logiciel libre livré sans AUCUNE GARANTIE.  
Vous pouvez le redistribuer sous certaines conditions.  
Tapez 'license()' ou 'licence()' pour plus de détails.  
  
R est un projet collaboratif avec de nombreux contributeurs.  
Tapez 'contributors()' pour plus d'information et  
'citation()' pour la façon de le citer dans les publications.  
  
Tapez 'demo()' pour des démonstrations, 'help()' pour l'aide  
en ligne ou 'help.start()' pour obtenir l'aide au format HTML.  
Tapez 'q()' pour quitter R.
```

# Modulation des fonctionnalités de R

Modules de fonctions : Packages

Croissance exponentielle



# Le projet Bioconductor

Projet interne à R (extrait de <http://www.bioconductor.org>)



**BIOCONDUCTOR**  
open source software for bioinformatics

Bioconductor is an open source and open development software project for the analysis and comprehension of genomic data.

accueil getting started overview téléchargements documentation publications workshops cabig

**project news**

- ▶ [2010-04-23](#)  
Bioconductor 2.6, consisting of 389 packages and designed to work with R version 2.11, was released today.
- ▶ [2009-10-26](#)  
BioC 2.5, consisting of 352 packages and designed to work with R 2.10.z, was released today.

[Voir aussi](#)

# Le projet Bioconductor

Projet interne à R (extrait de <http://www.bioconductor.org>)

Installation de Bioconductor (version allégée)

```
> source("http://www.bioconductor.org/biocLite.R")
```

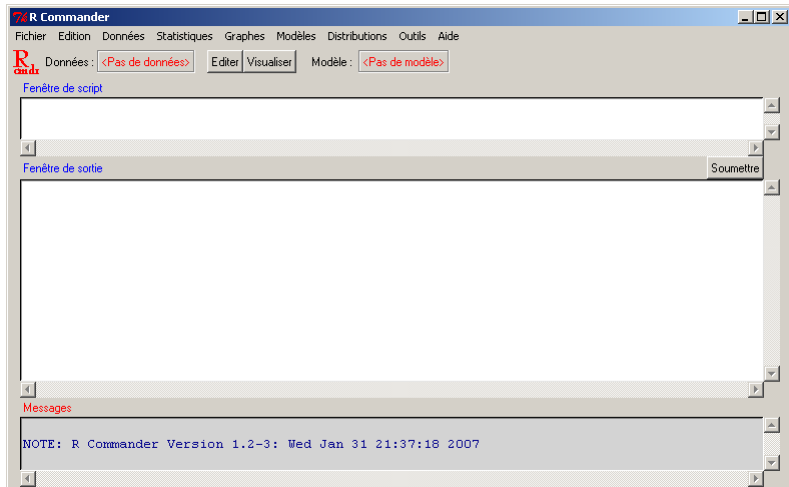
```
> biocLite()
```

# Plan du cours

- 1 R, Logiciel libre
- 2 Présentation
- 3 Importation**
- 4 Manipulation
- 5 Description graphique
- 6 Traitements automatiques

# Le package Rcmdr

## Environnement Rcmdr

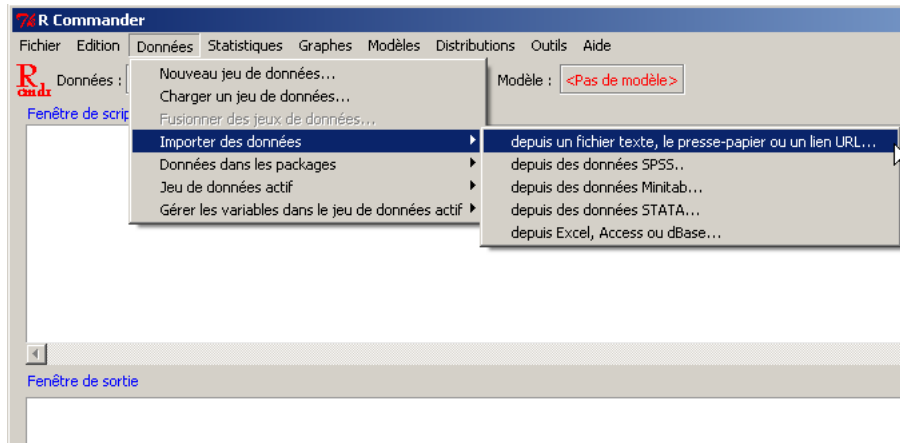


# Le package Rcmdr

Environnement Rcmdr

Importation des données `poulets`

<http://www.agrocampus-ouest.fr/math/causeur/Donnees/poulets.txt>

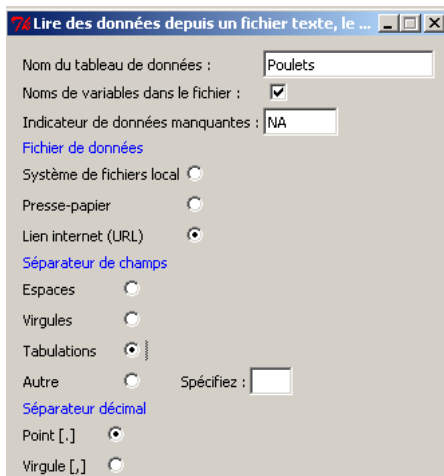


# Le package Rcmdr

## Environnement Rcmdr

## Importation des données `poulets`

<http://www.agrocampus-ouest.fr/math/causeur/Donnees/poulets.txt>

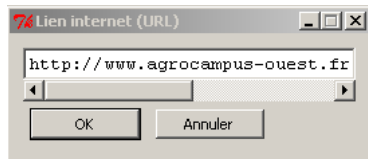


# Le package Rcmdr

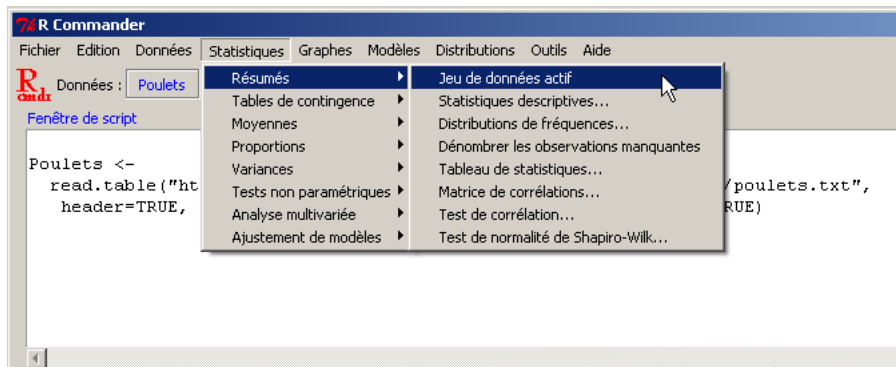
Environnement Rcmdr

Importation des données `poulets`

<http://www.agrocampus-ouest.fr/math/causeur/Donnees/poulets.txt>



# Statistiques élémentaires



The screenshot shows the R Commander application window. The title bar reads "74 R Commander". The menu bar includes "Fichier", "Edition", "Données", "Statistiques", "Graphes", "Modèles", "Distributions", "Outils", and "Aide". The "Statistiques" menu is open, displaying a list of options: "Résumés", "Tables de contingence", "Moyennes", "Proportions", "Variances", "Tests non paramétriques", "Analyse multivariée", and "Ajustement de modèles". A sub-menu is visible under "Résumés", listing: "Jeu de données actif", "Statistiques descriptives...", "Distributions de fréquences...", "Dénombrer les observations manquantes", "Tableau de statistiques...", "Matrice de corrélations...", "Test de corrélation...", and "Test de normalité de Shapiro-Wilk...". A mouse cursor is positioned over the "Jeu de données actif" option. In the background, the "Données" pane shows "Poulets" and the "Fenêtre de script" pane contains the R code: 

```
Poulets <-  
  read.table("ht  
  header=TRUE,
```

# Statistiques élémentaires

Fenêtre de sortie										Soumettre
3rd Qu.: 427.06	3rd Qu.:2865.5	3rd Qu.:333.58	3rd Qu.:205.94	3rd Qu.:334.65	3rd Qu.:113.85	3rd Qu.:352.35				
Max. :1649.74	Max. :5001.6	Max. :567.83	Max. :359.69	Max. :473.29	Max. :145.48	Max. :608.90				
		NA's : 1.00								
G6.8U.1.ddrt	IGFBP5.metab	C3.9D.1.ddrt	G3.6A.1.ddrt	PLTP.metab	C5.7E.3.ddrt	G4.5B.2.ddrt				
Min. : 62.67	Min. : 62.33	Min. : 628.7	Min. : 72.7	Min. : 72.84	Min. : 68.28	Min. : 58.74				
1st Qu.: 91.00	1st Qu.:101.47	1st Qu.:1345.5	1st Qu.:112.3	1st Qu.:103.88	1st Qu.:110.86	1st Qu.: 93.51				
Median :157.90	Median :141.39	Median :2090.6	Median :176.1	Median :149.95	Median :167.89	Median :139.46				
Mean :188.23	Mean :192.99	Mean :2618.4	Mean :220.0	Mean :190.08	Mean :209.21	Mean :189.61				
3rd Qu.:275.27	3rd Qu.:267.34	3rd Qu.:3915.1	3rd Qu.:316.2	3rd Qu.:261.19	3rd Qu.:294.68	3rd Qu.:264.95				
Max. :393.97	Max. :386.32	Max. :5643.7	Max. :481.4	Max. :424.12	Max. :440.11	Max. :444.71				
FBP.metab	C5.13B.1.ddrt	G5.14A.3.ddrt	IGF1.metab	C8.9A.1.ddrt	G6.8V.1.ddrt	regime	genotype	groupe		
Min. : 165.3	Min. : 60.71	Min. : 63.24	Min. : 55.34	Min. : 35.34	Min. : 64.22	A:14	G:13	AG:7		
1st Qu.: 282.9	1st Qu.: 98.05	1st Qu.:102.37	1st Qu.: 96.74	1st Qu.: 60.11	1st Qu.:117.26	N:13	M:14	AM:7		
Median : 471.4	Median :138.45	Median :168.74	Median :136.11	Median : 83.95	Median :162.98			NG:6		
Mean : 558.6	Mean :194.16	Mean :204.52	Mean :179.04	Mean :118.75	Mean :250.53			NM:7		
3rd Qu.: 815.2	3rd Qu.:259.81	3rd Qu.:295.24	3rd Qu.:246.10	3rd Qu.:154.46	3rd Qu.:392.05					
Max. :1333.9	Max. :554.29	Max. :534.82	Max. :465.93	Max. :335.53	Max. :812.64					

# Plan du cours

- 1 R, Logiciel libre
- 2 Présentation
- 3 Importation
- 4 Manipulation**
- 5 Description graphique
- 6 Traitements automatiques

# Extraction de données


## Sélection d'individus (lignes)

- Sélection d'un individu  
> Poulets[1,]
- Sélection de plusieurs individus  
> Poulets[1:10,]  
> Poulets[Poulets\$genotype=="G",]

## Sélection de variables (colonnes)

- Sélection d'une variable  
> Poulets[,1]  
> Poulets[,"G1.6E.3.ddrt"]  
> Poulets\$G1.6E.3.ddrt
- Sélection de plusieurs variables (colonne)  
> Poulets[,1:2]  
> Poulets[,c("G1.6E.3.ddrt","G6.8V.3.ddrt")]

## Transformation de variables

**Exercice** : créer un tableau de données (`lpoulets`) de même structure que `poulets` dans lequel les données d'expression sont obtenues par passage au  $\log_2$  (fonction `log2` de ) des données de `poulets`.

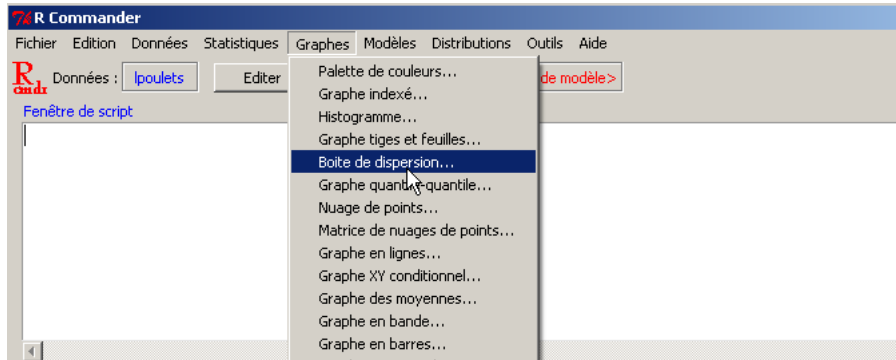
# Plan du cours

- 1 R, Logiciel libre
- 2 Présentation
- 3 Importation
- 4 Manipulation
- 5 Description graphique**
- 6 Traitements automatiques

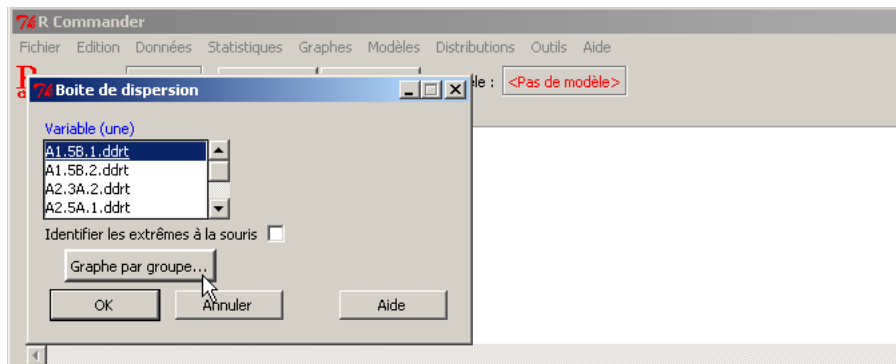
# Boîtes de dispersion



# Boîtes de dispersion

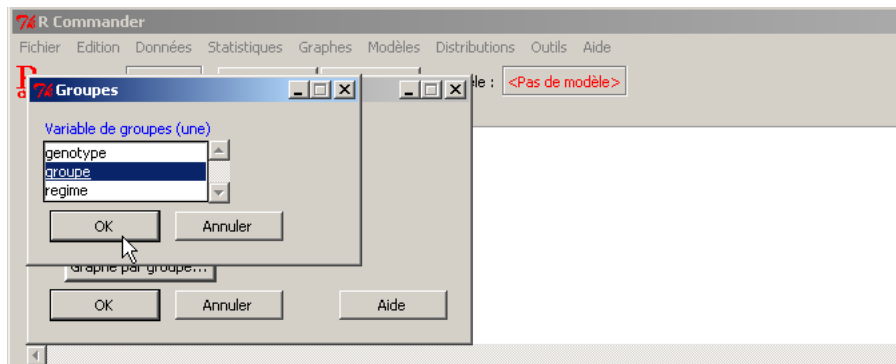


# Boîtes de dispersion

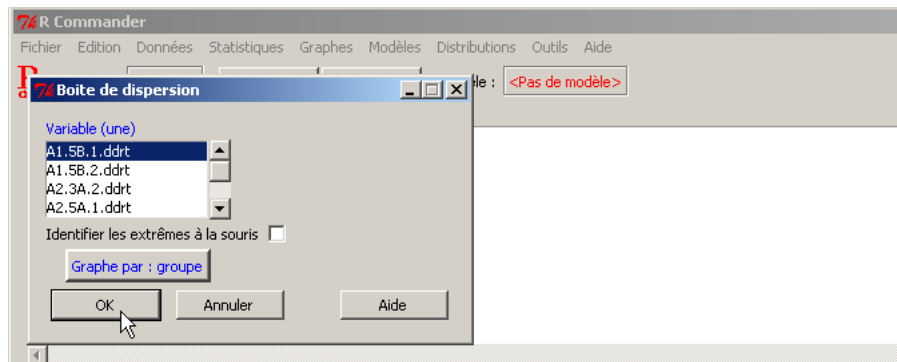


The screenshot shows the R Commander software interface. The main window has a menu bar with the following items: Fichier, Edition, Données, Statistiques, Graphes, Modèles, Distributions, Outils, Aide. A toolbar is visible below the menu bar. A dialog box titled "Boîte de dispersion" is open in the foreground. The dialog box has a title bar with standard window controls (minimize, maximize, close) and a file name field containing "<Pas de modèle>". Inside the dialog, there is a section labeled "Variable (une)" with a list box containing the following items: A1.5B.1.ddrt (selected), A1.5B.2.ddrt, A2.3A.2.ddrt, and A2.5A.1.ddrt. Below the list box is a checkbox labeled "Identifier les extrêmes à la souris" which is currently unchecked. There is a button labeled "Graphe par groupe..." with a mouse cursor pointing to it. At the bottom of the dialog are three buttons: "OK", "Annuler", and "Aide".

# Boîtes de dispersion



# Boîtes de dispersion



## Boîtes de dispersion multiples

**Exercice 1** : créer un graphique juxtaposant les boîtes de dispersion des expressions de chacun des 20 premiers gènes du tableau de données.

**Exercice 2** : créer un graphique juxtaposant les boîtes de dispersion des expressions de chaque biopuce.

# Plan du cours

- 1 R, Logiciel libre
- 2 Présentation
- 3 Importation
- 4 Manipulation
- 5 Description graphique
- 6 Traitements automatiques**

## Automatisation d'un traitement statistique

Fonctions d'automatisation :

- `apply(X,MARGIN,FUN)` : applique la fonction `FUN` à une des marges (ligne si `MARGIN=1`, colonne sinon) du tableau `X`

**Exercice** : créer le vecteur `moyennes` des expressions moyennes par puces.

- `sweep(x,MARGIN,STATS,FUN="-")` : standardise un tableau de données par application de la fonction `FUN` (par défaut, `FUN` est une soustraction) à une des marges (ligne si `MARGIN=1`, colonne sinon) du tableau `x`. La fonction `FUN` définit une opération impliquant une statistique `STATS`

**Exercice** : créer un tableau de données (`lcpoulets`) de même structure que `poulets` dans lequel les données d'expression sont centrées par biopuce.

## Création d'une fonction

```
NomFonction = fonction(a1,a2,...) # Nom et arguments
{                                # Début
...                              # Corps
return(sortie)                  # sortie est le résultat
}
```

**Exercice 1** : créer la fonction qui calcule  $z = 1/(x + \sqrt{y})$ .

**Exercice 2** : calculer les probabilités critiques des tests de comparaison des expressions moyennes par génotype pour chaque gène et donner la liste de ceux qui sont significatifs pour un seuil de 0.05.