

# Une interface graphique pour analyser des données distantes sous R

Raphaël Coudret<sup>(1)</sup>, Gilles Durrieu<sup>(2)</sup>, Jérôme Saracco<sup>(1)</sup>

<sup>(1)</sup>Équipe CQFD, INRIA Bordeaux - Sud-Ouest  
et IMB, Université de Bordeaux

<sup>(2)</sup>LMBA, Université de Bretagne Sud

1<sup>ères</sup> Rencontres R, 2 Juillet 2012, Bordeaux

# Sommaire

## Accéder aux données

- Les fonctionnalités de MySQL

- Utiliser des commandes SQL sous R

## Des méthodes d'analyse dans une interface graphique

- Les fenêtres et les boutons de Gtk+ ...

- ... disponibles sous R

## Application à des mesures biologiques

- Contexte

- Un estimateur de densité particulier

Accéder aux données

# Les fonctionnalités de MySQL

- ▶ Mise en ordre de grands jeux de données  
→ Tables, bases, index
- ▶ Connexion sécurisée à la base de données  
→ Droits de lecture, d'écriture, de création de tables
- ▶ Interopérabilité  
→ Systèmes d'exploitation, langages de programmation

Exploitation des données, travail en groupe,  
diffusion de résultats

## Les fonctionnalités de MySQL

Commande	mysqldump	mysqlimport	mysql
Emploi	Faire une copie de la base de données	Importer des données dans une table	Se connecter au serveur MySQL
			Envoyer des commandes SQL

Option	-h	-u	-p	-e
Emploi	Adresse IP du serveur	Nom d'utilisateur	Mot de passe	Commande à envoyer

Exemple :

```
mysql -h 164.12.89.261 -u root -p
-e "DROP TABLE nomDb.nomTable"
```

## Utiliser des commandes SQL sous R

### *Package RMySQL :*

- ▶ Connexion à une base de données : `dbConnect`.  
Produit un objet que l'on nomme `conObj`.
- ▶ Création du script SQL :  
chaîne de caractère stockée par exemple dans `script`.
- ▶ Envoi de la commande :  
`retour <- dbSendQuery(conObj, script)`
- ▶ Récupération éventuelle des résultats de la commande :  
`resultat <- fetch(retour, n=-1)`
- ▶ Déconnexion : `dbDisconnect`

# Des méthodes d'analyse dans une interface graphique

## Les fenêtres et les boutons de Gtk+ ...

### Gtk+

- ▶ Bibliothèque permettant de créer des objets graphiques.
- ▶ Est à la base de GNOME et GIMP.

### Son utilisation

- ▶ Créer ses objets graphiques :  
Glade génère un fichier .XML.
- ▶ Lier ces objets à des programmes (C, Python, R, etc...).



## ... disponibles sous R

### *Package RGtk2*

- ▶ Initialiser les différents types d'objets : `gtkWindowGetType()`
- ▶ Charger le fichier .XML : `gtkBuilder()`
- ▶ Appeler une fonction lors d'une interaction avec un objet :  
`gSignalConnect()`
- ▶ Accéder aux propriétés d'un objet : `fenetre$destroy()`
- ▶ S'aider de la documentation :  
<http://developer.gnome.org/gtk/stable/>

# Application à des mesures biologiques

## Contexte

- ▶ Données : écartements entre les deux parties de coquilles d'huîtres au cours du temps.
- ▶ Particularité : plus de 50 000 mesures par jour et par animal.
- ▶ Objectif : déterminer la santé d'un animal pour fabriquer un indicateur de la qualité de l'eau.
- ▶ Observation : les cycles d'ouverture et de fermeture des huîtres varient avec leur état de santé.
- ▶ Problème : déterminer quand l'huître est ouverte ou fermée.

<http://molluscan-eye.epoc.u-bordeaux1.fr/>

## Un estimateur de densité particulier

- ▶ Nous souhaitons estimer la densité des écartements des huîtres pendant une journée.
- ▶ Nous utilisons un estimateur à noyau gaussien (Parzen, 1962).
- ▶ Nous supposons que la densité à estimer a deux modes.
- ▶ Nous construisons un estimateur qui possède également deux modes (Coudret et al.).
- ▶ La position de l'antimode de la densité estimée détermine les valeurs d'écartements correspondant respectivement aux ouvertures et aux fermetures.

Perspectives et limites

## Perspectives et limites

### Des utilisations possibles variées

- ▶ Données protégées.
- ▶ Codes diffusables (produit fini, enseignement).

### Quelques problèmes...

- ▶ *Packages* difficiles à installer (en particulier sous Windows).
- ▶ Compatibilité des interfaces graphiques entre les systèmes d'exploitation.

## Références



R. Coudret, G. Durrieu, and J. Saracco.

Comparison of kernel density estimators with assumption on number of modes.

*Submitted paper.*



M. Lawrence and D. Temple Lang.

RGtk2 : a graphical user interface toolkit for R.

*Journal of Statistical Software*, 37(8) :1–52, 2010.



E. Parzen.

On estimation of probability density function and mode.

*Ann. Math. Statist.*, 33(3) :1065–1076, 1962.



M. Sow, G. Durrieu, L. Briollais, P. Ciret, and J.-C. Massabuau.

Water quality assessment by means of HFNI valvometry and high-frequency data modeling.

*Environmental Monitoring and Assessment*, 182(1-4) :155–170, 2011.

## Exemple avec RMySQL

```
1 conObj <- dbConnect(MySQL(), host="147.210.161.10",
2   user="*****", password="*****", dbname="dbEels")
3 script <- "SHOW COLUMNS FROM dbEels.site"
4 retour <- dbSendQuery(conObj, script)
5 resultats <- fetch(retour, n=-1)
6 print(resultats)
7 dbDisconnect(conObj)
```

**Code** : accès à la liste des colonnes d'une table SQL.



## Exemple avec RGtk2

```
1  gtkWindowGetType()
2  gtkButtonGetType()
3  builder <- gtkBuilder()
4  builder$addFromFile("ex.xml")
5  builder$connectSignals(NULL)
6  bouton <- builder$getObject("bouton")
7  fenetre <- builder$getObject("fenetre")
8  fenetre$show()
9
10 gSignalConnect(bouton, "clicked", function(btn){
11     fenetre2 <- gtkWindow()
12 })
```

Code : création d'une multitude de fenêtres.