

Prise en main du Logiciel R → 2 mots sur R

R est un logiciel de statistique disponible librement sur *internet*.

Vous pouvez le télécharger à <http://www.r-project.org/>

Il est déjà installé dans ces salles

The screenshot shows the R Project for Statistical Computing website as it appeared in the early 2000s, viewed through a Windows Internet Explorer browser. The browser's address bar shows the URL <http://www.r-project.org/>. The website features the R logo on the left and a navigation menu with links such as "About R", "What is R?", "Contributors", "Screenshots", "What's new?", "Download CRAN", "R Project Foundation", "Members & Donors", "Mailing Lists", "Bug Tracking", "Developer Page", "Conferences", "Search", and "Documentation". The main content area is titled "The R Project for Statistical Computing" and displays several statistical plots: a PCA plot titled "PCA 5 vars" with a biplot showing variables like Fertility, Catholic, Agriculture, Examination, and Education; a bar chart showing the distribution of a variable; a dendrogram for clustering; and two histograms showing the distribution of data across groups. The bottom section, titled "Getting Started:", contains a list of bullet points providing information about R as a free software environment and how to download it from the CRAN mirror.

The R Project for Statistical Computing

PCA 5 vars
princomp(x = data[, cor = cor])

Fertility
Catholic
Agriculture
Examination
Education
(1-3) 60%

Clustering 4 groups

Factor 1 [41%]
Factor 3 [19%]

Groups
28
16
1
2

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](#).
- If you have questions about R like how to download and install the software, or what the license terms are, please read our [answers to frequently asked](#)

Prise en main du Logiciel R ➡ 2 mots sur R

R est constitué d'un ensemble de **librairies** utilisables dans des domaines précis de la statistique. Nous commencerons à l'utiliser pour des fonctions de calculs simples mais il deviendra un véritable outil de travail nous simplifiant la pratique de la Biostatistique.

R est également un langage de programmation.

Le **calcul vectoriel** et le traitement de **fichiers** feront partie de son utilisation.

Nous allons commencer par nous servir du logiciel dans le cadre **d'exercices de difficulté croissante**. Ne restez pas en plan et demandez de l'aide si nécessaire **DES LES PREMIERES ETAPES!!!**

Plutôt que du Bla Bla inutile à ce stade, entrons en piste maintenant!...

Lisez attentivement les instructions contenues dans les **dias** suivantes et essayer dès que possible de voler de vos propres ailes

objectif : ***les mains dans le cambouis***

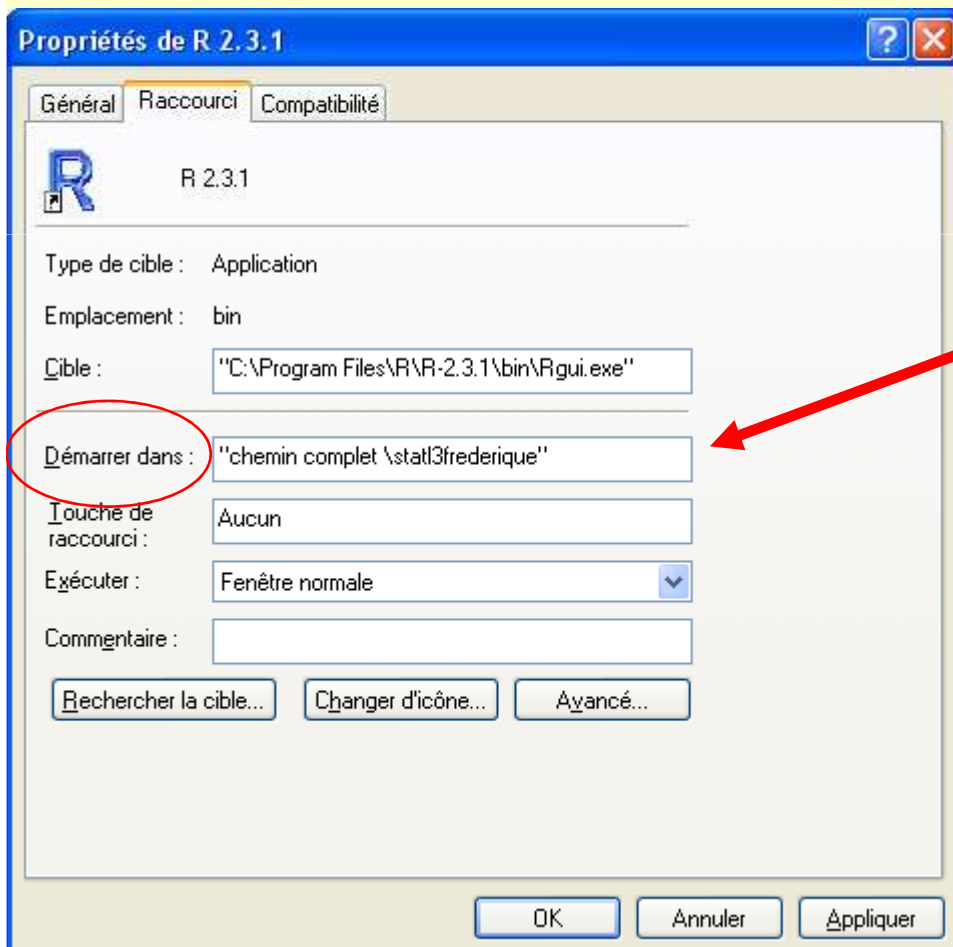
Prise en main du Logiciel R ➔ Raccourci

⇒ **étape 1** : Bien définir le raccourci pour la suite du travail

a/ Créer un répertoire StatL3_prénom **sans accent ni blanc** (ex: statl3_frederique)

b/ Copier raccourci R dans ce répertoire

c/ Clic droit souris sur ce raccourci et indiquer le chemin complet du répertoire que vous venez de créer (cliquer ensuite sur Appliquer et OK)



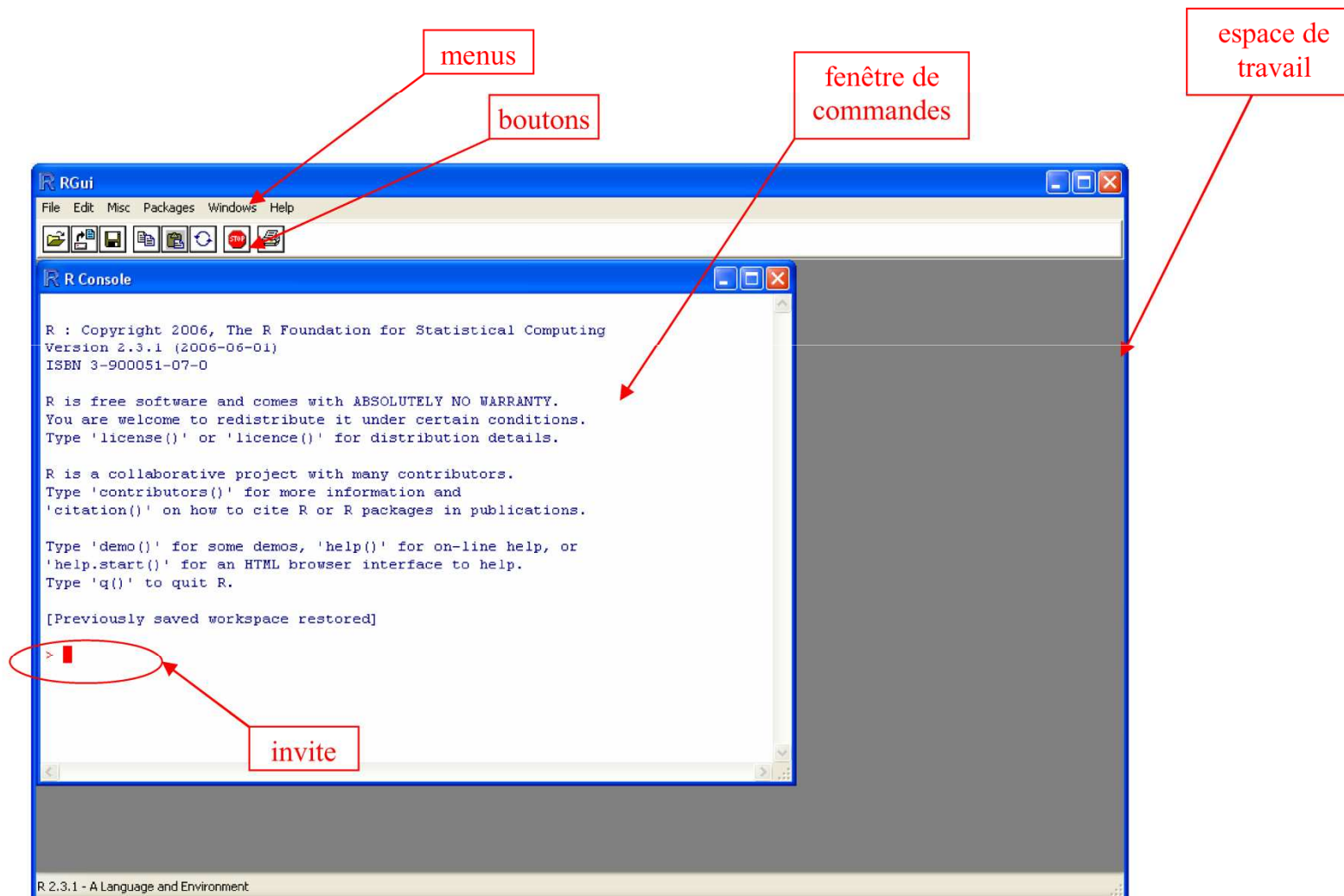
ICI

⇒ **Attention** :

à la fin de chaque séance, il vous faudra copier votre environnement de travail sur une clé USB (données effacée du disque périodiquement)

Prise en main du Logiciel R ➡ l'interfaçage

GUI : Graphical User Interface



Prise en main du Logiciel R ➡ Conseil

- ⇒ Il vous est loisible de travailler sous word, d'y prendre à la volée vos notes, d'y ajouter d'éventuelles remarques et de copier/coller tout ce que bon vous semble des lignes de commandes aux graphes réalisés en salle.
- ⇒ Utiliser la touche « Impression système »
ou la combinaison CTRL+ Impression système

Prise en main du Logiciel R → Calculs simples -1-

⇒ étape 2 : calcul de la concentration d'une solution saline

4 variables simples seront utilisées (destinées à recevoir des valeurs réelles)

conc_mere, **vol_prel**, **vol_final** et **conc_exp**

Où conc_mere contient la concentration de la solution mère, vol_prel : le volume prélevé de cette solution, vol_final : le volume final de la dilution et conc_exp : la concentration expérimentale (à calculer)

Vous connaissez l'équation : $\text{conc_mere} \times \text{vol_prel} = \text{vol_final} \times \text{conc_exp}$

Unités choisies : concentrations en mM et volumes en ml

Pour une variable recevoir un contenu correspond à une opération d'affectation notée du sigle < suivit du signe –

Ce qui donne par exemple : **conc_mere <- 100**

Dans une fenêtre R taper les instructions suivantes :

```
> vol_prel <- 25
> conc_mere <- 100
> vol_final <- 250
```

Prise en main du Logiciel R ➡ Calculs simples -1-

⇒ étape 2 (suite) : calcul de la concentration d'une solution saline

L'opérateur de multiplication est noté *
celui de la division est noté /
celui de l'addition est noté +
et l'opérateur de la soustraction est noté -
Ce qui donne dans notre exemple :

```
> vol_prel <- 25  
> conc_mere <- 100  
> vol_final <- 250  
> conc_exp <- vol_prel*conc_mere/vol_final  
> conc_exp  
[1] 10
```

Pour obtenir le contenu d'une variable affectée ou calculée, il suffit d'indiquer son nom et taper la touche ENTREE. C'est ce qui a été fait sur la dernière instruction. Le résultat figure en bleu , le nombre entre crochet indique le nombre de lignes contenues dans la variable (ici une seule valeur, donc une seule ligne) il est suivi du contenu de la variable, en l'occurrence la concentration recherchée : 10 (mM)

Prise en main du Logiciel R → Calculs simples -2-

⇒ étape 3 : travail sur les notes d'un groupe d'étudiants

calcul de sommes pondérées

et d'une moyenne globale sur un ensemble de notes d'étudiants en *Biostat*

Nous allons utiliser ici des vecteurs

- Définition de 2 variables (vecteurs) :

NoteCc (notes du contrôle continu) et **NoteExam** (notes obtenues à l'examen)

Utilisation de **c** : combine values into a **vector** or list

c est une fonction qui regroupe des valeurs dans un vecteur ou une liste

> **help("c")** pour invoquer l'aide sur cette fonction

remarque : comparer avec `note=scan()`

```
> NoteCc <- c(12,20,0,15,17.5,5,10,16,14,12)
```

```
> NoteCc
```

```
[1] 12.0 20.0 0.0 15.0 17.5 5.0 10.0 16.0 14.0 12.0
```


Prise en main du Logiciel R → Calculs simples -2-

⇒ étape 3 (suite) : travail sur les notes d'un groupe d'étudiants

- Calculs sur ces vecteurs

mean pour calculer une moyenne (des valeurs contenues dans une liste)

var et **sd** pour calculer respectivement la variance et l'écart type d'une distribution

length pour connaître l'effectif des valeurs contenues dans un vecteur

```
> NoteExam <- c(18,7.5,10.5,6,16,9,14,9,13,15.5)
```

```
> mean(NoteExam)
```

```
[1] 11.85
```

```
> sd(NoteExam)
```

```
[1] 4.02
```

⇒ Un peu d'anglais...

- moyenne : mean

- écart-type : standard deviation (sd)

- résumé : summary

Prise en main du Logiciel R ➔ Calculs simples -2-

⇒ étape 3 (suite) : travail sur les notes d'un groupe d'étudiants

summary, une fonction incontournable

```
> summary(NoteCc)

  Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
 0.00  10.50   13.00   12.15   15.75   20.00
```

- Calculs sur ces vecteurs ⇒ définition d'une nouvelle variable **NoteEcrit** (vecteur) en tenant compte des coefficients (0.6 et 0.4) des notes
- Calcul de la **moyenne** du groupe (une nouvelle variable est sûrement impliquée)

```
> NoteEcrit <- 0.6*NoteExam+0.4*NoteCc
> NoteEcrit
 [1] 15.6 12.5 6.3 9.6 16.6 7.4 12.4 11.8 13.4 14.1
> MoyGroupe <- mean(NoteEcrit)
> MoyGroupe
 [1] 11.97
```

Prise en main du Logiciel R → Calculs sur fichiers

⇒ étape 4 : suivre la croissance bactérienne

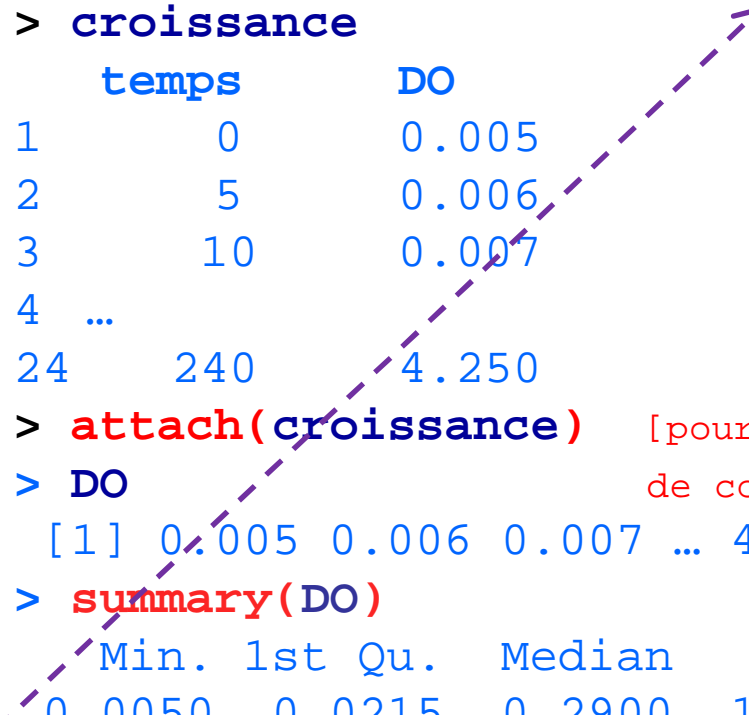
R est un logiciel permettant de manipuler des tableaux de données volumineux

temps	DO
0	0.005
5	0.006
10	0.007
15	0.008
20	0.011
30	0.017
40	0.023
45	0.033
60	0.080
75	0.085
90	0.120
105	0.250
120	0.330
135	0.600
150	1.020
165	1.950
180	3.340
200	4.050
210	4.250
220	4.170
230	4.280
240	4.250

bacterial1.txt

- ⇒ créer le fichier texte des données de façon cohérente
- ⇒ 2 colonnes (les titres de colonnes deviendront 2 variables sous R)
- ⇒ importer les données dans R avec la fonction **read.table**

```
> croissance <-  
  read.table("./data/bacterial1.txt", h=TRUE)  
> croissance  
   temps    DO  
1      0 0.005  
2      5 0.006  
3     10 0.007  
4    ...  
24   240 4.250  
> attach(croissance) [pour pouvoir utiliser les entête  
> DO                de colonnes comme noms de variables]  
[1] 0.005 0.006 0.007 ... 4.280 4.250  
> summary(DO)  
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.  
0.0050  0.0215  0.2900  1.4110  2.8300  4.2800
```



Prise en main du Logiciel R ➡ Calculs sur fichiers

⇒ **étape 4 (suite)** : Quelques manipulations de données
recherche du minimum, de la taille du tableau, ordonner les valeur avec **sort**
afficher les données correspondant à un critère

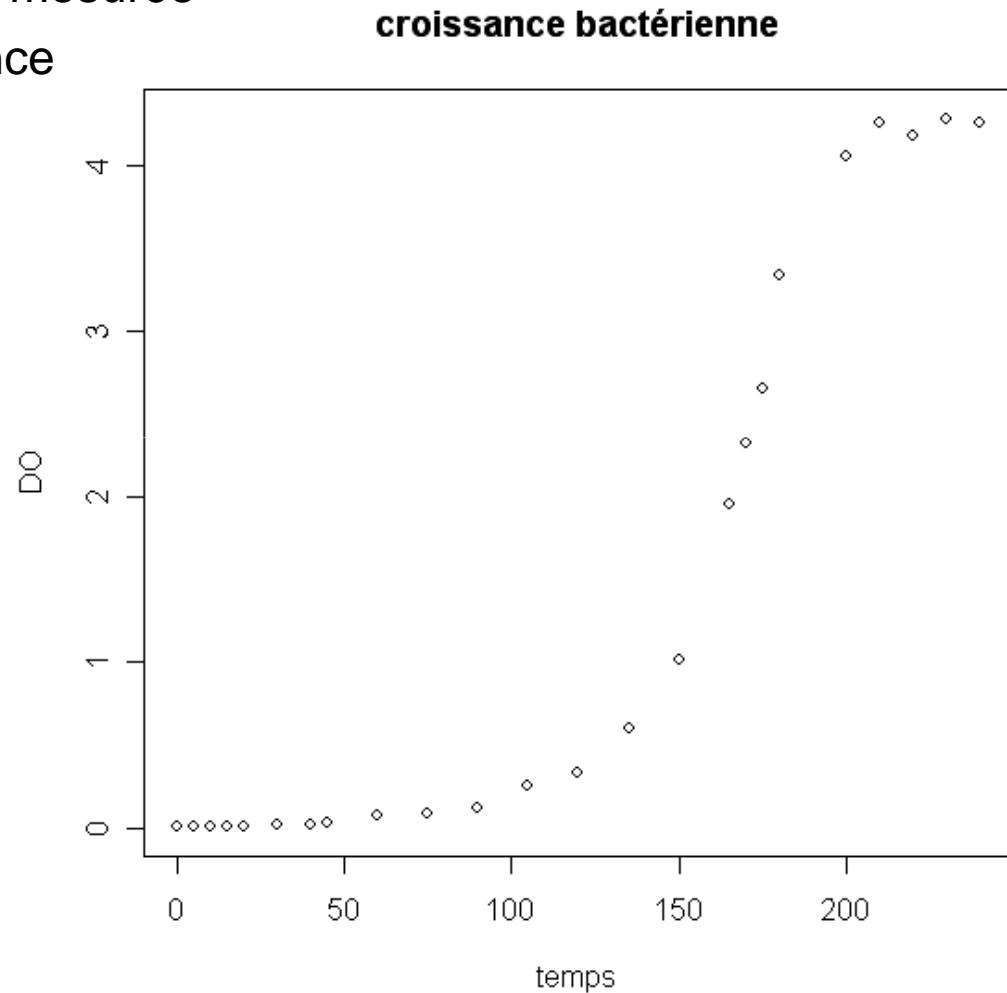
```
> croissance <- read.table("bacterial.txt",h=TRUE)
> attach(croissance) [associe les entêtes colonnes à des variables vecteurs]
> min(DO)
[1] 0.005
> length(DO)
[1] 24
> sum(DO)/length(DO)
[1] 1.411042
> mean(DO)
[1] 1.411042
> sort(DO)
[1] 0.005 0.006 0.007 ... 4.170 4.250 4.250 4.280
> croissance[DO==4.25,]
      temps    DO
21      210 4.25
22      240 4.25
> croissance[DO==4.25 & temps>210,]
```

Prise en main du Logiciel R ➡ Les premiers graphes

⇒ étape 4 : suivre la croissance bactérienne <côté plot>

⇒ Réaliser le **graphe** de la croissance mesurée
et importée dans le tableau croissance

⇒ Graphe simple



```
> plot(x=temps,y=DO,main="croissance bactérienne")
```

Prise en main du Logiciel R ➡ Les premiers graphes

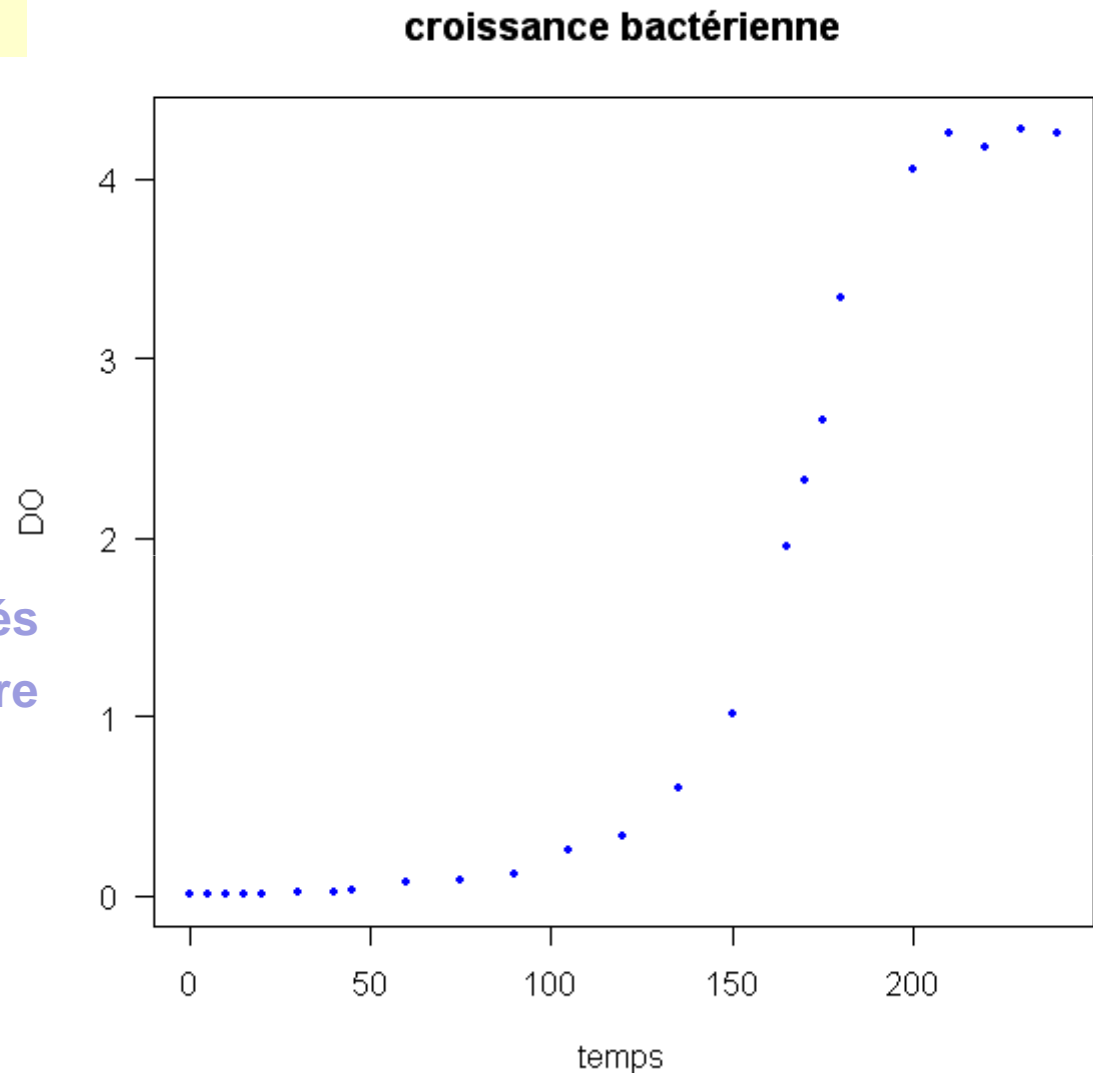
⇒ étape 4 : suivre la croissance bactérienne <côté plot>

```
> plot(x=temps,y=DO,main="croissance bactérienne")
```

Prise en main du Logiciel R ➡ Les premiers graphes

⇒ étape 4 : *plot ... plus élaboré*

A retenir : les graphes sont tracés
« à la carte », chaque paramètre
précisant l'option choisie



```
> plot(x=temps,y=DO,main="croissance bactérienne",las=1,pch=20,  
      col="blue")
```

Prise en main du Logiciel R ➡ Les premiers graphes

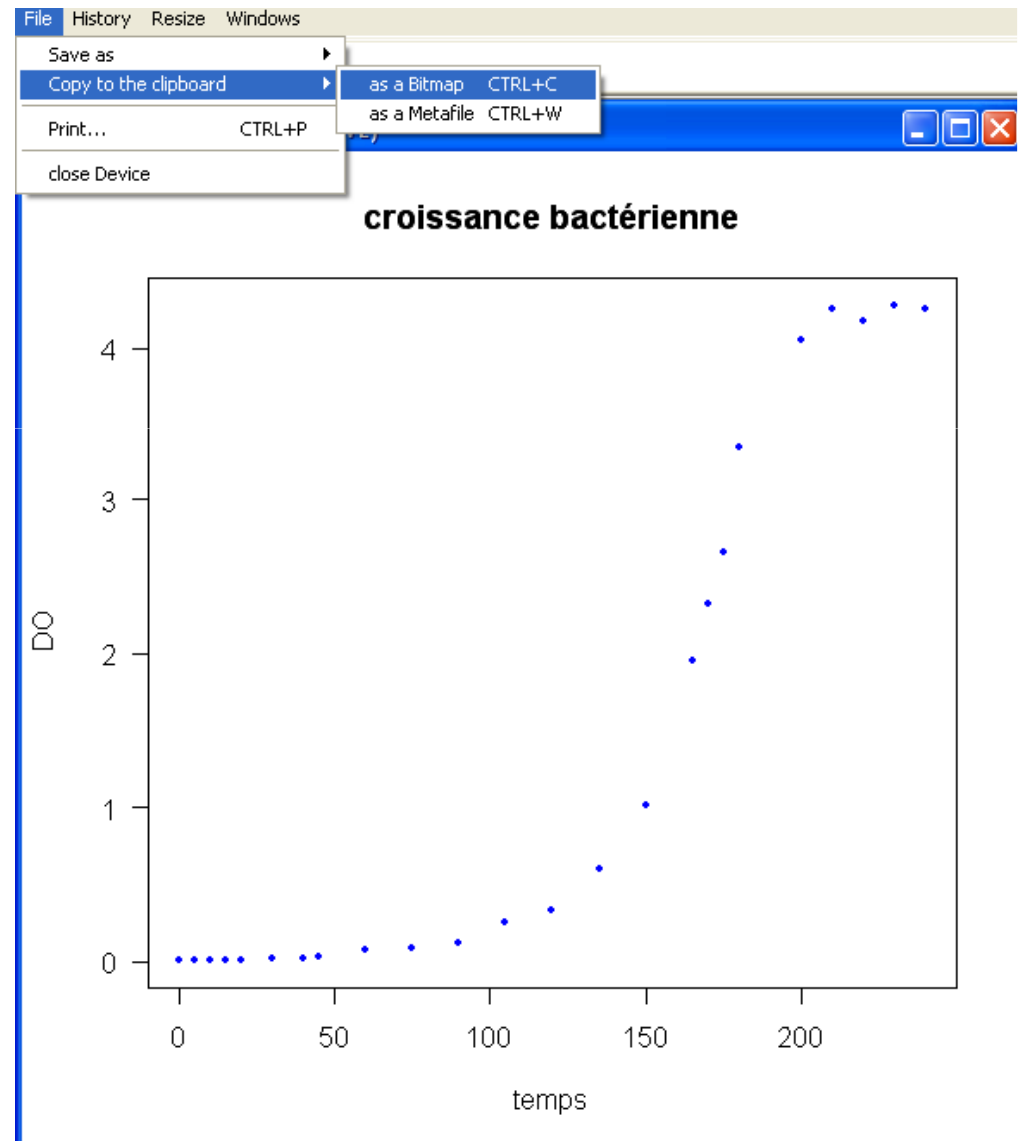
⇒ étape 4 : *plot ... plus élaboré*

```
> plot(x=temps,y=DO,main="croissance bactérienne",las=1,pch=20,  
      col="blue")
```


Prise en main du Logiciel R ➡ Les premiers graphes

⇒ Sauvegarder un graphe en dessin pour l'insérer dans un rapport

Utiliser ensuite « insertion image » dans le menu correspondant de Word, EXCEL ou Powerpoint.



Prise en main du Logiciel R ➡ Résumé

objectif : *les stats par la pratique*

- ⇒ Première expérience avec le logiciel R
- ⇒ Affectation de valeurs à des variables et calculs simples
- ⇒ Fonctions appliqués à des vecteurs et calculs sur vecteurs
- ⇒ Premières fonctions de statistique descriptive
- ⇒ Premiers graphes
- ⇒ Importation de fichiers de données

Investissement : reprenez ce travail et entraînez vous

**Cette étape simple est considérée comme définitivement acquise
(pensez-y pour l'examen)**

Site web : vous y retrouverez tout ce dont vous avez besoin