

TD - 2.0

Retour sur R

Vincent PAYET

« Jusqu'ici, tout va bien »
— Hubert Koundé



Objectifs : cette fiche propose un rappel sur les bases de , quelques éléments du langage et des bonnes pratiques d'utilisation du logiciel.


Table des matières

1	Le logiciel R	1
2	Les bases	2
2.1	R en 12 lignes	2
2.2	Des explications	3
3	Bonnes pratiques	4
4	Documentation	4


1 Le logiciel R

 [4] est un logiciel dédié à l'analyse de données, à la statistique et à la représentation graphique de données. C'est aussi un langage de programmation. Cela implique un certain apprentissage du langage. Parmi les avantages de ce logiciel citons en deux :

- c'est un langage de programmation, ce qui permet de conserver une trace exhaustive de la démarche d'une analyse.
- c'est un logiciel libre, il est légalement téléchargeable et utilisable (et modifiable, même si cela vous concerne moins).

Pour installer  : <http://cran.r-project.org>

2 Les bases

Il faut du temps pour apprendre un langage et il est probablement impossible de connaître  intégralement avec ses milliers de modules complémentaires (nommés des *packages*, avec l'accent de *Mr Bean*). Toutefois, pour commencer à discuter de statistique avec votre ordinateur, il ne faut que quelques éléments de langage, le reste viendra avec la pratique. Voici le verbatim d'un échange de quelques commandes accompagnées de commentaires. Les entrées sont précédées du symbole `>`, le prompt de la console, les sorties commencent par un numéro entre crochets qui permet de connaître le rang de la valeur en début de ligne. Les commentaires (parties non évaluées du programme) sont introduits par le symbole dièse `#`.

Pour ces exemples, copiez le fichier de données `exemple.txt` sur le bureau (ou dans votre dossier de travail si vous savez déjà le modifier).

2.1 R en 12 lignes

```
> 40+2                                # R comme une calculatrice.
[1] 42

> x <- 6*7                            # On peut stocker un résultat dans un objet
> z <- x+2                            # et manipuler cet objet
> z                                    # Pour voir l'objet, équivalent à print(z).
[1] 44

> y <- c(1,4,6,7,10)                  # On peut stocker plusieurs valeurs dans un vecteur
> y*2                                # et manipuler ce vecteur...
[1]  2  8 12 14 20

> mean(y)                            # ... éventuellement avec des fonctions statistiques.
[1] 5,6

> y[2] ; y[1:3]; y[-1]                # On peut accéder par les index à une partie d'un objet.
[1] 4
[1] 1 4 6
[1]  4  6  7 10

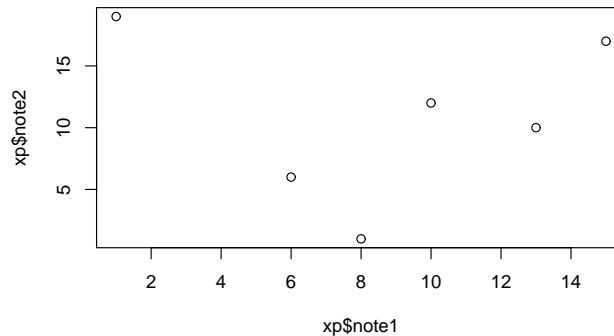
> # La fonction read.table sert à importer des données.
> xp <- read.table("exemple.txt",header=TRUE)
> xp[1,2] ; xp[,3:4]                  # On peut accéder aux lignes et/ou aux colonnes par les indices
[1] 10

  note2 groupe
1    12      1
2    17      1
3     1      2
4     6      2
5    10      3
6    19      3

> xp$noms; xp$Noms                    # on peut aussi accéder aux colonnes avec le symbole $
NULL
```

```
[1] alf joe foo bar tim hal
Levels: alf bar foo hal joe tim

> # Notez que Noms et noms ne sont pas équivalents. La casse compte.
> plot(xp$note1,xp$note2) # la fonction plot est une fonction graphique
```



2.2 Des explications

Pour essayer ces commandes, il n'est pas nécessaire de recopier les commentaires (par exemple `# R comme une calculatrice`).

Le symbole `<-` permet de nommer un objet. On dit qu'on **assigne** le résultat de l'opération `6*7` à `x`. Les noms suivis de parenthèses sont des **fonctions**. Entre les parenthèses on précise les **paramètres** des fonctions, séparés par des virgules. On a vu ici quatre fonctions :

- `c()` : permet de créer un vecteur.
- `mean()` : calcule la moyenne.
- `read.table()` : lit un fichier et permet d'importer des données. Cette fonction construit un objet particulier, un `data.frame`, on peut le voir comme un jeu de données. Le paramètre `header=TRUE` permet de préciser que la première ligne du fichier contient des noms de colonnes, des entêtes. Le mot `TRUE` (en majuscule) est un mot réservé du langage pour donner la valeur d'un objet dit booléen. un booléen est un objet de type *logical* qui vaut vrai ou faux (`TRUE` ou `FALSE`).
- `plot()` : fonction générique pour réaliser des graphiques. Ici on fournit en paramètres deux vecteurs et la fonction retourne un nuage de points de coordonnées `note1` en abscisse et `note2` en ordonnée.


Les crochets permettent de sélectionner un sous-élément en précisant leur index : `y[2]` renvoie la deuxième valeur de `y`. Pour les objets composés de lignes et de colonnes on peut accéder à une valeur particulière : `xp[1,2]` renvoie la valeur sur la 1^{er} ligne et la 2^e colonne. On peut aussi accéder aux lignes ou aux colonnes :

- `xp[1,]` renvoie la 1^{er} ligne et toutes les colonnes
- `xp[,3]` renvoie la 3^e colonne et toutes les lignes

On peut retenir que les indices sont toujours à lire ainsi : [Lignes, Colonnes].


Le symbole `$` permet d'accéder aux colonnes par leur nom. Enfin on peut mettre plusieurs commandes sur la même ligne en les séparant par un `;` comme `y[2]` ; `y[1:3]`.

3 Bonnes pratiques

Au delà du langage, pour utiliser efficacement le logiciel il y a quelques bonnes pratiques à adopter rapidement. Elles sont explicitées dans le TD 1 [3]. La liste est la suivante :

- Définir un répertoire de travail (menu fichier, changer le répertoire courant)
- Utiliser des scripts pour conserver une trace de vos analyses, et de vos apprentissages
- Consulter l'aide pour les nouvelles fonctions avec `help(fonction)` ou le symbole `? : ?mean`. Par défaut l'aide s'ouvre sous forme de page web.
- Constituer un glossaire des fonctions de base vues pendant les TD.

4 Documentation

La documentation sur  est très riche, livres, manuels téléchargeables, sites internet... Vous pouvez notamment consulter les manuels *An introduction to R* en anglais [5], *R pour les débutants* en français [2], et le livre *Statistique avec R* [1].

Références

- [1] P-A. Cornillon and A. Guyader. *Statistiques avec R*. PUR, 2012.
- [2] E. Paradis. *R pour les débutants*. <http://cran.r-project.org/manuals.html>, 2005.
- [3] V. Payet. *TDR1, Premiers pas*. ISARA, 2012.
- [4] R Core Team. *R : A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2012. ISBN 3-900051-07-0.
- [5] D. M. Venables, W. N. Smith and the R Core Team. *An Introduction to R. Notes on R : A Programming Environment for Data Analysis and Graphics*. <http://cran.r-project.org/manuals.html>, 2012.