

Introduction à R

Julien Grosjean

M1 GIL

Année 2016-2017 S2

Généralités

- *Freeware* : <https://www.cran.r-project.org/>
- Pour Linux (aptitude `r-base-dev`), MacOS X et Windows (32 et 64 bits)
- R est un interpréteur (exceptions) et un « vrai » langage :
 - Types de données
 - Branchements conditionnels
 - Boucles
 - Fonctions
 - Modules,
 - Etc.



Pourquoi ? Pour qui ?

- Orienté statistiques (tests stats, fouille de données, apprentissage, graphiques, mathématiques avancées, ...)
- Nombreux « déçus » par les langages/frameworks classiques (Matlab, Excel, etc.)
- Nouvelles disciplines (Big Data, OLAP) : bioinformatique, informatique médicale, *Business analytics*, etc.
- Adaptation du langage S (fin '70 et années '80)
- « Patchwork » de plusieurs bibliothèques existantes : C, C++, Fortran, Java, ... => *all-in-one*

Les bases

- Le type « vecteur »
- Tableaux, matrices => ensemble de vecteurs
- Librairies (base + additionnelle)

```
library()  
library(LIB)  
library(help = LIB)
```

- Affectation de valeur

```
x <- 3 ou  
x = 3  
v <- c(1, 23, 42)
```

- Opérateurs : **<**, **<=**, **>**, **>=**, **!=**, **==**
- Lignes de commande (console) + scripts

Les bases : exemples

```
v <- 1:4; #suite arithmétique de raison 1  
class(v); #renvoie numeric
```

```
v <- seq(from=1, to=2, by=0.2); #sequence  
v[3] # = 1.4 ici (accès aux indices du vecteur à partir de 1)
```

```
b <- c(T, F, F, T, T); # vecteur de booléens
```

```
x <- c(v, b); #concaténation de vecteurs
```

Les bases : fonctions de base

- `length()`
- `sum()`
- `mean()`
- `max()`
- ...
- `names()` : attribuer/récupérer noms des cases du vecteur
- De l'aide ? `Help(fonction/librairie)` ou `?fonction/librairie`

Valeurs particulières

- Valeur infinie (`-Inf` ou `Inf`)
- Valeur indéfinie (`NaN`) : exemple `0/0`
- Donnée manquante (`NA`)
- Le pointeur NULL (`is.null(x)`)

Les bases : data.frame

- Tableau de n vecteurs

Soit x et y des vecteurs de même taille,

```
t<-data.frame(x, y)
```

En colonne par défaut mais aussi possible en ligne

- À partir d'un fichier

```
t<-read.table(FICHIER)
```


Structures algorithmiques

- Conditions

```
if( CONDITION ) {  
    #TODO  
}else{  
    #TODO  
}
```

- Boucles

```
for( INDICE in SEQUENCE) {  
    #TODO  
    #break possible  
}  
while( CONDITION) {  
    #TODO  
}
```

Les bases : fonctions

```
myFunc = function(p1,p2=99) {  
  #TODO  
  return c(p1,p2);  
}
```

```
myFunc(10,22); # appel
```

Et bien d'autres choses...

- Graphiques : plot, hist, barplot, dotchart, pie, boxplot, etc.
- Gestion des chaînes de caractères
- Fonctions mathématiques avancées : cos, sin, tan, quantile, round, union, mod, etc.
- Entrées/sorties : read.csv, read.table, save, print, etc.
- Les bibliothèques dédiées

Organisation du travail

- *Le workspace* (environnement R avec opérations et objects en mémoire) : sauvegarde possible
- Généralement dans un répertoire prédéfini

Références

- <http://www.duclert.org/>
- <http://moo.nac.uci.edu/~hjm/AnRCheatsheet.html>
- <https://cran.r-project.org/doc/contrib/Short-refcard.pdf>