

Initiation à R

É.Hirchaud
F.Simonnet

26 Septembre 2013

Table des matières

1	L'environnement RStudio	1
1.1	Éléments de RStudio	1
2	Premiers Pas	1
2.1	R est une calculette	1
2.2	Créer des objets et les utiliser	2
2.2.1	Vecteurs	2
2.2.2	Matrices et data frame	3
3	Erreurs, importer exporter des données	4
3.1	Importer et exporter des données	4
4	Graphiques	4
5	Écrire et utiliser une fonction maison	5
5.0.1	Préparation	5
6	Programation les bases	6
6.0.2	Les boucles	6
6.1	Les conditions	6

1 L'environnement RStudio

1.1 Éléments de RStudio

Présentation des différentes fenêtres de RStudio et personnalisation.

2 Premiers Pas

2.1 R est une calculette

```
1 # Une opération simple
2 10 + 3
3 # Une Opération plus complexe
4 10 / (3+8) * 78
5 # les différents opérateurs sont :
6 # multiplication *
7 # addition +
8 # division /
9 # soustraction -
```

EXERCICE :

Calculer le temps passé dans votre dernière structure professionnelle (université, labo, entreprise ...) puis diviser par la différence entre l'année en cours et votre année de naissance. Multiplier par 100 pour obtenir un pourcentage du temps passé dans cette structure.

```
1 # Stockage d'une valeur dans une variable
2 nombreX <- 50.8
3
4 # Accès à la valeur stockée dans la variable nombreX
5 # Utiliser l'auto complétion. Commencer par écrire nomb puis appuyer sur la
   touche de tabulation.
6 nombreX
7
8 # stockage de la variable nombreY (le symbole <- est identique à =)
9 nombreY <- 7.4
10
11 # Utiliser l'auto complétion, si plusieurs solutions existent vous pouvez
12 # continuer de taper le nom de la variable pour que le choix soit plus
13 # restreint puis utiliser les flèches du clavier pour choisir la bonne
14 # variable. Une fois sélectionnée, appuyer soit sur entrée.
15 nombreX + nombreY
16
17 # Stockage du résultat de l'opération dans la variable sommeXY
18 # Conseil : utiliser l'historique en appuyant sur la 'flèche haut'
19 # on peut retrouver des commandes déjà écrites.
20 sommeXY <- nombreX + nombreY
```

EXERCICE :

Répéter l'exercice précédant mais en utilisant des variables.

2.2 Créer des objets et les utiliser

2.2.1 Vecteurs

```
1 # Création d'un vecteur numérique
2 monVecteur1 <- c(20.7676, -45, 78, 12)
3
4 # Remarque sur les symbole > et + au début de la ligne de commande
5 #
6 c(20, 45, 78,
7
8 # Séquence de nombres de 1 à 30
9 maSuite <- seq(from = 1, to = 30)
10
11 # Une séquence de nombres paire
12 maSuitePaire <- seq(from = 2, to = 20, by = 2)
13
14
15 # Nous avons utiliser la fonction seq()
16 # Pour trouver à quoi servent les arguments taper la commande
17 ?seq
18
19 # La documentation de cette fonction s'affiche sur la panneau en bas à
   droite de RStudio
```

EXERCICE En utilisant la fonction `seq()`, créer un vecteur de nombres impaires allant de 7 à 77.

```
1 # La fonction rnorm() permet de créer des vecteurs avec des nombres alé
   atoirs.
2 monVecteurAlea <- rnorm(45)
```

```

3
4 # Création d'un vecteur avec chaîne de caractère.
5 monVecteurA <- c("Mut_1", "Mut_2", "Mut_3")
6
7 # La fonction rep() permet de répéter une chaîne de caractères un certain
  nombre de fois.
8 rep("Mut", 4)
9
10 # La fonction paste() permet de coller des chaînes de caractères en les sé
    parant par un caractère.
11 paste("Mut", 1, sep = "_")

```

EXERCICE En utilisant les fonctions *rep()* et *paste()* et *seq()* créer un vecteur identique à *monVecteurA*, mais allant de 1 à 20. Dans un premier temps utiliser des étapes intermédiaires.

```

1 # Accès à l'élément d'indice 1
2 monVecteur1[1]
3
4 # Accéder aux éléments d'indice 3 , 4 puis 5.
5
6 # Accès aux éléments d'indice 1 à 3
7 monVecteur1[1:3]
8
9 # Accéder aux éléments d'indice 1 3 4
10 monVecteur1[c(1, 3, 4)]
11
12 # Pour connaître la longueur d'un vecteur on utilise la fonction length()
13 length(monVecteur1)

```

EXERCICE Trouver une manière d'accéder au dernier élément d'un vecteur.

```

1 # Opérations sur tous les éléments d'un vecteur
2 monVecteur1 + 5.5
3 log(monVecteur1)
4 log10(monVecteur1)
5 log2(monVecteur1)
6 log(monVecteur1, 2)
7 abs(monVecteur1)
8 round(monVecteur1)
9 round(monVecteur1, 2)
10
11 # Opérations qui travaillent sur tout un vecteur
12 sum(monVecteur1)
13 mean(monVecteur1)
14 median(monVecteur1)
15 sd(monVecteur1)
16 var(monVecteur1)
17 max(monVecteur1)
18 min(monVecteur1)
19 range(monVecteur1)
20
21 # Addition de deux vecteurs
22 monVecteur2 <- c(10, 100, 5, 2)
23 monVecteur1 + monVecteur2

```

2.2.2 Matrices et data frame

```

1 # Matrice avec des nombres aléatoires
2 maMatrice <- matrix(rnorm(100), ncol = 10)

```

```

3
4 # S'informer sur la matrice
5 # voir les n premières lignes d'une matrice
6 head(maMatrice)

```

EXERCICE :

- Combien de lignes la fonction `head()` affiche - t'elle par défaut
- Comment afficher plus ou moins de ligne ? Utiliser la documentation de la fonction `head` pour trouver.
- Comment voir la matrice à partir des dernières lignes ?

```

1 #Dimension de la matrice : nombre de lignes et nombre de colonnes
2 dim(maMatrice)

```

EXERCICE :

- Quel type d'objet renvoie la fonction `dim()`
- La fonction `ncol()` permet de nous renseigner sur le nombre de colonnes. Trouver une fonction similaire pour trouver le nombre de lignes
- Avec la fonction `dim()` afficher uniquement le nombre de lignes

```

1 # Une matrice est un ensemble de vecteur.
2 # Chaque colonne est un vecteur ainsi que chaque ligne.
3 # Pour récupérer la 9ème colonne.
4 col9 <- maMatrice[, 9]

```

EXERCICE : Recupérer les colonnes 1, 7 et 3 de la matrice *maMatrice*

```

1 # La fonction colnames() permet d'obtenir le nom des colonnes d'une matrice
  ou d'un data.frame
2 colnames(maMatrice)
3
4 #Donner un nom aux colonnes
5 colnames(maMatrice) <- LETTERS[1:ncol(maMatrice)]
6 colnames(maMatrice)
7
8 #Type et classe d'objet#
9 typeof(maMatrice)
10 class(maMatrice)
11
12 #Transformer une matrice en data.frame
13 monDataFrame <- as.data.frame(maMatrice)
14
15 #Autre moyen d'accéder aux colonnes avec un data.frame
16 colonneC <- monDataFrame$C

```

3 Erreurs, importer exporter des données

3.1 Importer et exporter des données

```

1 # Importation de données
2
3 #La fonction essentiel read.delim
4 #Sans aucun argument
5 mesEchantillons <- read.delim("samples.txt")
6
7 #Avec des arguments
8 mesEchantillons <- read.delim("samples.txt", row.names = 1, dec = ",",
  stringsAsFactor = FALSE, header = TRUE)
9

```

```

10 #Observer le nom des colones
11
12
13 mesEchantillons <- read.delim("samples2.txt", row.names = 1, dec = ",",
    stringsAsFactor = FALSE, header = TRUE)
14
15 #Possible d'importer des données avec R studio, mais tous les arguments ne
    sont pas disponibles.
16
17 #Sauvegarde dans un fichier
18 #La fonction essentiel write.table
19 write.table(x = mesEchantillons, file = "mesEchantillons.txt", col.names =
    NA , row.names = TRUE, quote = FALSE, sep = "\t")

```

Exercice : retrouver les cinq erreurs de cette ligne de code.

```

1 #Ligne de commande erronée.
2 matriceBug <- read.delin("sample.txt" , rown.names = 1 head = true)

```

4 Graphiques

```

1 #la fonction plot()
2 x<-1:20
3 plot(x, x^2)
4
5 #Il existe plusieurs fonction de base pour les graphiques.
6 # plot()
7 # hist()
8 # boxplot()
9 # Utilisation basique
10 boxplot(maMatrice)
11
12
13 #Arguments communs à toutes les fonctions
14 plot(x, x^2, xlim=c(0, 30), ylim=c(-100, 500), xlab="Variable x", ylab="
    Variable x au carré", main="Carré des valeurs de 1 à 20", cex.axis=1.5,
    cex.lab=1.5, cex.main=2, bty="l", pch=16)
15
16
17 # Sauvegarde dans un fichier image
18 # Dans l'onglet Plots : Export-'Save Plot As Image'
19 # File name : boxPlot
20 # La même chose avec la commande :
21 png("boxPlot.png")
22 plot(x, x^2, xlim=c(0, 30), ylim=c(-100, 500), xlab="Variable x", ylab="
    Variable x au carré", main="Carré des valeurs de 1 à 20", cex.axis=1.5,
    cex.lab=1.5, cex.main=2, bty="l", pch=16, col = 2)
23
24 #Pour ajouter une légende
25 legend("topright", legend = "X", pch = 16, col = 2)
26
27 # Ferme la fenêtre graphique et enregistre le fichier.
28 dev.off()
29
30 # Il existe 71 paramètres pour affiner les graphiques
31 # Liste des 71 paramètres de la fonction par()
32 par()

```

5 Écrire et utiliser une fonction maison

5.0.1 Préparation

Il est plus facile d'écrire des fonctions dans un ou plusieurs fichiers. Créer un nouveau fichier et le nommer `mesFonctions.R`. La syntaxe pour écrire une fonction est la suivante :

```
1 maFonction <- function(argument1, argument2 = valeurParDefaut){
2
3   #On utilise le nom des argument comme variable pour faire des calcul,
   appeler des fonctions.
4   resultatTemporaire <- argument1 + argument2
5
6   #On continue différent traitement avec d'autre variable créer dans la
   fonction
7   resultatTraiter <- uneAutreFonction(resultatTemporaire)
8
9   #On retourne le resultats avec la fonction return()
10  return(resultatTraiter)
11
12
13 }
```

Enoncé : Créer la fonction *ingredientsPateAPizza* qui prend comme argument le nombre de pizza, par défaut 1. Et retourne un vecteur avec les quantités pour X pizzas. :

```
1 ingredientsPateAPizza(nbPizza = 1)
2 Farine Eau Levure
3 [1] 500 250 20
```

Les ingrédients pour une pizza

- farine : 500
- eau : 250
- levure : 20

Tester sa fonction : La fonction est écrite dans le fichier `fonction.R`, mais elle est encore inconnue dans le *workspace*. Pour la mettre en mémoire il faut exécuter la fonction `source()`.

```
1 source("fonction.R")
```

Si il n'y pas d'erreurs de syntaxe, la fonction sera reconnue par R ce que l'on remarque sur la fenêtre *Workspace*. On peut maintenant utiliser notre fonction de la même façon que n'importe quelles autres.

Exercice : Créer la fonction *repIngred* qui prend comme arguments

- *vecIngred* Un vecteur de nombre dont chaque indice à le nom d'un ingrédients.
- *imageName* Une chaîne de caractère qui correspond au nom du fichier image.
- *recipiesName* Une chaîne de caractère qui correspond au nom de la recette.

6 Programation les bases

6.0.2 Les boucles

Le mot *for* permet de répéter une instruction un certain de nombre de fois.

```
1 x <- 1:20
2
3 for(i in 1:length(x)){
4
5   x + 100
6
7 }
```

Exercice : Dans un nouveau fichier de votre choix. Écrire une boucle *for* exécutant 10 fois la fonction *ingredientsPateAPizza*

6.1 Les conditions

les instructions *if* et *else* permettent d'exécuter des parties de codes sous certaines conditions.

```
1 if ( x > 0){
2     print("x est positif")
3
4 }else{
5     print("x est négatif")
6 }
7
8
9 if ( x > 0){
10     print("x est positif")
11
12 }else{
13     if(x != 0){
14         print("x est négatif")
15     }else{
16         print("x est nul")
17     }
18 }
```

Exercice : Dans la fonction *ingredientsPateAPizza* ajouter une instruction *if else* qui vérifie que l'arguments *nbPizza* soit bien positif et non nul. Auquel cas un message du style "Veuillez saisir un nombre de pizza positif" devra s'afficher.