

# Fonctions en R

Sophie Baillargeon, Université Laval

2021-03-24

## Table des matières

<b>1</b>	<b>Syntaxe générale d'une fonction R</b>	<b>2</b>
1.1	Composantes d'une fonction R . . . . .	3
1.2	Fonction anonyme . . . . .	4
<b>2</b>	<b>Arguments en entrée</b>	<b>5</b>
2.1	Valeurs par défaut des arguments . . . . .	5
2.1.1	Arguments acceptant seulement quelques chaînes de caractères spécifiques . . . . .	8
2.2	Appel d'une fonction . . . . .	9
2.3	Argument . . . . .	10
2.3.1	Utilité 1 : recevoir en entrée un nombre indéterminé de valeurs à traiter . . . . .	10
2.3.2	Utilité 2 : passer des arguments à une autre fonction . . . . .	12
<b>3</b>	<b>Résultat produit par une fonction</b>	<b>13</b>
3.1	Sortie d'une fonction . . . . .	13
3.2	Retour de plusieurs éléments dans une liste . . . . .	14
3.3	Fonction <code>match.call</code> . . . . .	15
3.4	Effets de bord d'une fonction . . . . .	15
<b>4</b>	<b>Évaluation d'un appel à une fonction et environnements associés</b>	<b>17</b>
4.1	Passage d'arguments par valeur . . . . .	18
4.2	Portée lexicale . . . . .	18
4.3	Chemin de recherche complet . . . . .	20
4.4	Bonnes pratiques dans la création d'une fonction . . . . .	20
<b>5</b>	<b>Exemple de création d'une fonction R</b>	<b>21</b>
5.1	Étapes de développement conseillées . . . . .	21
<b>6</b>	<b>Programmation fonctionnelle</b>	<b>23</b>
<b>7</b>	<b>Résumé</b>	<b>25</b>
	<b>Références</b>	<b>27</b>

---

*Note préliminaire : Lors de leur dernière mise à jour, ces notes ont été révisées en utilisant R version 4.0.3.*

---

Lorsqu'un bout de code R est susceptible d'être utilisé à répétition (p. ex. pour faire un même calcul sur des données différentes), il est préférable d'en faire une fonction R. Les fonctions permettent de :

- rédiger du code plus clair et plus court, donc plus facile à comprendre et à partager ;
- diminuer les risques de faire des erreurs ;

- sauver du temps à long terme.

Bref, faire des fonctions est une bonne pratique de programmation en R.

Comme il a déjà été mentionné dans les [notes sur les concepts de base en R](#), une fonction R est un bout de code qui produit un certain résultat, lorsqu'exécuté. Pour exécuter le code composant une fonction, celle-ci doit être *appelée*. Lorsqu'elle est appelée, la fonction prend des valeurs en entrée, qui sont assignées à des *arguments*. Le code de la fonction réfère à ces arguments de façon à ce que l'appel de la fonction provoque un traitement des valeurs fournies en entrée. En fin de compte, la fonction génère un résultat, qui est la plupart du temps *retourné* dans un objet en sortie. Ce résultat peut aussi être un *effet de bord* (p. ex. la production d'un graphique, l'écriture dans un fichier externe, etc.). Voici une représentation schématique de l'exécution d'un appel à une fonction R



Voyons maintenant comment écrire nos propres fonctions en R.

## 1 Syntaxe générale d'une fonction R

Pour créer une fonction en R, il faut utiliser le mot-clé `function` en respectant la syntaxe suivante :

```
<nom_fonction> <- function(<arg_1>, <arg_2>, <arg_3>) {
  <instructions> # formant le corps de la fonction
}
```

Dans ce gabarit de code, les éléments encadrés des signes `<` et `>` sont des bouts à remplacer par ce qui convient pour la tâche à accomplir. Les éléments `<arg_1>`, `<arg_2>` et `<arg_3>` représentent les arguments de la fonction (aussi appelés paramètres dans d'autres langages informatiques), soit les objets qui peuvent être fournis en entrée à la fonction (qui ne sont pas nécessairement au nombre de trois).

Voici une fonction qui reprend un exemple présenté dans les [notes sur les structures de contrôle en R](#). Elle calcule des statistiques descriptives simples selon le type des éléments du vecteur donné en entrée.

```
stats_desc <- function(x) {
  if (is.numeric(x)) {
    stats <- c(min = min(x), moy = mean(x), max = max(x))
  } else if (is.character(x) || is.factor(x)) {
    stats <- table(x)
  } else {
    stats <- NA
  }
  return(stats)
}
```

Après avoir soumis le code de création de cette fonction dans la console, la fonction se retrouve dans l'environnement de travail. Il est alors possible de l'appeler.

```
stats_desc(x = iris$Species)
```

```
## x
##      setosa versicolor  virginica
##      50          50          50
```

Nous pourrions ajouter un argument à cette fonction. Par exemple, nous pourrions offrir l'option d'une sortie présentée sous la forme d'une matrice plutôt que d'un vecteur.

```
stats_desc <- function(x, sortie_matrice) {
  # Calcul
  if (is.numeric(x)) {
    stats <- c(min = min(x), moy = mean(x), max = max(x))
  } else if (is.character(x) || is.factor(x)) {
    stats <- table(x, dnn = NULL)
  } else {
    stats <- NA
  }
  # Production de la sortie
  if (sortie_matrice) {
    stats <- as.matrix(stats)
    colnames(stats) <- if (is.character(x) || is.factor(x)) "frequence" else "stat"
  }
  return(stats)
}
```

L'argument `dnn = NULL` a aussi été ajouté dans l'appel à la fonction `table` afin de retirer le nom de la dimension (ici `x`) dans la sortie produite par la fonction. Nous pouvons maintenant appeler la fonction comme suit.

```
stats_desc(x = iris$Species, sortie_matrice = TRUE)
```

```
##           frequence
## setosa           50
## versicolor       50
## virginica        50
```

## 1.1 Composantes d'une fonction R

Les composantes d'une fonction R sont :

- la liste de ses arguments, possiblement avec des valeurs par défaut (nous allons y revenir) ;

```
args(stats_desc)
```

```
## function (x, sortie_matrice)
## NULL
```

```
# ou
formals(stats_desc)
```

```
## $x
##
##
## $sortie_matrice
```

- le corps de la fonction, soit les instructions qui la constituent.

```
body(stats_desc)
```

```
## {
##   if (is.numeric(x)) {
##     stats <- c(min = min(x), moy = mean(x), max = max(x))
##   }
##   else if (is.character(x) || is.factor(x)) {
##     stats <- table(x, dnn = NULL)
##   }
## }
```

```
##     else {
##         stats <- NA
##     }
##     if (sortie_matrice) {
##         stats <- as.matrix(stats)
##         colnames(stats) <- if (is.character(x) || is.factor(x))
##             "frequence"
##         else "stat"
##     }
##     return(stats)
## }
```

- l'environnement englobant de la fonction (défini plus loin).

```
environment(stats_desc)
```

```
## <environment: 0x0000017c65a6bed8>
```

## 1.2 Fonction anonyme

Notons qu'une fonction n'a même pas besoin de porter de nom. La grande majorité du temps, une fonction est conçue pour être appelée à plusieurs reprises et il est alors nécessaire qu'elle ait un nom. Cependant, certaines fonctions sont parfois à usage unique.

Par exemple, il est parfois utile de se créer une fonction pour personnaliser le calcul effectué par une fonction de la famille des `apply`. Si cette fonction est très courte et a peu de chance d'être réutilisée, il n'est pas nécessaire de la nommer.

Voici un exemple. Si nous voulions calculer le minimum, la moyenne et le maximum (comme le fait notre fonction `stats_desc`) de toutes les variables numériques du jeu de données `iris`, mais selon le niveau de la variable `Species`, nous pourrions utiliser trois appels à la fonction `aggregate` comme suit.

```
aggregate(formula = . ~ Species, data = iris, FUN = min)
```

```
##      Species Sepal.Length Sepal.Width Petal.Length Petal.Width
## 1      setosa          4.3         2.3          1.0         0.1
## 2 versicolor          4.9         2.0          3.0         1.0
## 3  virginica          4.9         2.2          4.5         1.4
```

```
aggregate(formula = . ~ Species, data = iris, FUN = mean)
```

```
##      Species Sepal.Length Sepal.Width Petal.Length Petal.Width
## 1      setosa       5.006       3.428       1.462       0.246
## 2 versicolor       5.936       2.770       4.260       1.326
## 3  virginica       6.588       2.974       5.552       2.026
```

```
aggregate(formula = . ~ Species, data = iris, FUN = max)
```

```
##      Species Sepal.Length Sepal.Width Petal.Length Petal.Width
## 1      setosa          5.8         4.4          1.9         0.6
## 2 versicolor          7.0         3.4          5.1         1.8
## 3  virginica          7.9         3.8          6.9         2.5
```

Nous pourrions aussi créer une fonction qui calcule les trois statistiques et donner cette fonction en entrée à `aggregate` comme valeur à l'argument `FUN`.

```
aggregate(
  formula = . ~ Species,
  data = iris,
```

```

FUN = function(x) c(min = min(x), moy = mean(x), max = max(x))
)

##      Species Sepal.Length.min Sepal.Length.moy Sepal.Length.max Sepal.Width.min
## 1      setosa           4.300           5.006           5.800           2.300
## 2 versicolor           4.900           5.936           7.000           2.000
## 3 virginica            4.900           6.588           7.900           2.200
##      Sepal.Width.moy Sepal.Width.max Petal.Length.min Petal.Length.moy Petal.Length.max
## 1           3.428           4.400           1.000           1.462           1.900
## 2           2.770           3.400           3.000           4.260           5.100
## 3           2.974           3.800           4.500           5.552           6.900
##      Petal.Width.min Petal.Width.moy Petal.Width.max
## 1           0.100           0.246           0.600
## 2           1.000           1.326           1.800
## 3           1.400           2.026           2.500

```

Nous n'avons jamais donné de nom à la fonction et cela n'a causé aucun problème. Nous n'avons même pas utilisé d'accolades pour encadrer le corps de la fonction. Ce n'est pas nécessaire lorsque celui-ci est composé d'une seule instruction.

## 2 Arguments en entrée

Les arguments d'une fonction sont définis en énumérant leurs noms entre les parenthèses après le mot-clé `function`.

```

<nom_fonction> <- function(<arg_1>, <arg_2>, <arg_3>) {
  <instructions> # formant le corps de la fonction
}

```

Il n'y a pas de restrictions quant au nombre d'arguments que peut posséder une fonction. Exceptionnellement, une fonction peut même ne posséder aucun argument. C'est le cas par exemple de la fonction `getwd` et de la fonction suivante.

```

HelloWorld <- function() cat("Hello World !")

```

Comme nous le savons déjà, pour appeler une fonction sans fournir d'arguments, il faut tout de même utiliser les parenthèses.

```

HelloWorld()

```

```

## Hello World !

```

Omettre les parenthèses retourne le code source de la fonction.

```

HelloWorld

```

```

## function() cat("Hello World !")
## <environment: 0x0000017c65a6bed8>

```

### 2.1 Valeurs par défaut des arguments

Afin de définir une valeur par défaut pour un argument, il faut accompagner son nom dans l'énumération des arguments d'un opérateur `=` et d'une expression R retournant la valeur par défaut. Par exemple, dans la fonction `stats_desc`, il serait préférable de définir un format par défaut pour la sortie.

```

stats_desc <- function(x, sortie_matrice = FALSE) {
  # Calcul
  if (is.numeric(x)) {

```

```

  stats <- c(min = min(x), moy = mean(x), max = max(x))
} else if (is.character(x) || is.factor(x)) {
  stats <- table(x, dnn = NULL)
} else {
  stats <- NA
}
# Production de la sortie
if (sortie_matrice) {
  stats <- as.matrix(stats)
  colnames(stats) <- if (is.character(x) || is.factor(x)) "frequence" else "stat"
}
return(stats)
}

```

La liste des arguments de la fonction devient la suivante.

```
formals(stats_desc)
```

```

## $x
##
##
## $sortie_matrice
## [1] FALSE

```

Les arguments qui ne possèdent pas de valeur par défaut sont obligatoires. Si une fonction est appelée sans donner de valeur en entrée à un argument obligatoire, une erreur est produite.

```
stats_desc(sortie_matrice = FALSE)
```

```
## Error in stats_desc(sortie_matrice = FALSE): argument "x" is missing, with no default
```

Les arguments ayant une valeur par défaut peuvent, pour leur part, ne pas être fournis en entrée, auquel cas leur valeur par défaut est utilisée.

```
stats_desc(x = iris$Sepal.Length)
```

```

##      min      moy      max
## 4.300000 5.843333 7.900000

```

### Valeurs par défaut non mentionnées dans la liste des arguments

Parfois, même si un argument ne possède pas de valeur par défaut dans la liste des arguments, une valeur lui est attribuée dans le corps de la fonction si l'utilisateur appelle la fonction sans assigner de valeur à l'argument. Prenons l'exemple de la [fonction sample](#) vue dans les [notes sur les calculs statistiques en R](#) pour illustrer cette possibilité.

```
args(sample)
```

```

## function (x, size, replace = FALSE, prob = NULL)
## NULL

```

Dans la liste officielle des arguments de cette fonction, l'argument `size` ne possède pas de valeur par défaut. Pourtant, si la fonction `sample` est appelée sans fournir en entrée de valeur à cet argument, aucune erreur n'est générée et une sortie est produite.

```
sample(1:5)
```

```
## [1] 4 1 3 2 5
```

Allons voir le corps de la fonction pour comprendre ce qui se passe.

```
body(sample)
```

```
## {
##   if (length(x) == 1L && is.numeric(x) && is.finite(x) && x >=
##       1) {
##       if (missing(size))
##         size <- x
##       sample.int(x, size, replace, prob)
##   }
##   else {
##     if (missing(size))
##       size <- length(x)
##     x[sample.int(length(x), size, replace, prob)]
##   }
## }
```

La fonction `sample` fait appel à la [fonction `missing`](#) pour tester si une valeur a été fournie en entrée à `size`. Si ce n'est pas le cas, une valeur est assignée à `size`.

Cette façon de faire n'est pas idéale. Il est plus clair pour l'utilisateur de voir les valeurs par défaut dans la liste des arguments. Le comportement par défaut de la fonction `sample` si l'argument `size` n'est pas fourni est tout de même expliqué dans la section intitulée « *Details* » de la [fiche d'aide de la fonction `sample`](#).

### Valeurs par défaut qui dépendent d'autres arguments

Une valeur par défaut peut utiliser les valeurs prises par d'autres arguments. Prenons par exemple la [fonction `sample.int`](#).

```
args(sample.int)
```

```
## function (n, size = n, replace = FALSE, prob = NULL, useHash = (!replace &&
##   is.null(prob) && size <= n/2 && n > 1e+07))
## NULL
```

La valeur par défaut de l'argument `size` est `n`, qui est le nom du premier argument de la fonction. Ainsi, par défaut, l'argument `size` de `sample.int` prend la même valeur que ce qui a été fourni à l'argument `n`.

Nous pouvons également constater que la valeur par défaut de l'argument `useHash` est `(!replace && is.null(prob) && size <= n/2 && n > 1e+07)`, soit une expression faisant intervenir les valeurs prises par tous les autres arguments de la fonction.

Ainsi, la fonction `sample` aurait pu être programmée différemment. La liste de ses arguments aurait pu contenir la valeur par défaut suivante pour l'argument `size` : `if (length(x) == 1L && is.numeric(x) && is.finite(x) && x >= 1) x else length(x)`. Peut-être que les auteurs de la fonction ont trouvé cette expression trop lourde pour la mettre dans la liste des arguments.

### Valeurs par défaut NULL

Pour éviter une valeur par défaut dont l'expression est trop complexe, certains auteurs de fonctions R utilisent NULL comme valeur par défaut dans la liste des arguments, puis définissent la vraie valeur par défaut dans le corps de la fonction. Par exemple, la [fonction `cut`](#) vue dans les [notes sur le prétraitement de données en R](#) implémente cette pratique pour son argument `labels`.

```
args(cut.default)
```

```
## function (x, breaks, labels = NULL, include.lowest = FALSE, right = TRUE,
##   dig.lab = 3L, ordered_result = FALSE, ...)
## NULL
```

Si un utilisateur appelle la fonction `cut` sans fournir de valeur à l'argument `labels`, alors celui-ci prend d'abord la valeur `NULL`. Cependant, la valeur de l'objet `labels` est ensuite modifiée dans le corps de la fonction par des instructions ayant l'allure suivante.

```
if (is.null(labels)) {  
  <instructions> # comportant une nouvelle assignation de valeur à labels  
}
```

Ainsi la fonction `is.null` est utilisée pour vérifier si une valeur a été fournie en entrée lors de l'appel de la fonction. Cette façon de faire est une solution de rechange à l'utilisation de `missing` en omettant une valeur par défaut dans la liste des arguments. Elle comporte cependant l'avantage d'indiquer plus clairement dans la liste des arguments que l'argument n'est pas obligatoire.

### 2.1.1 Arguments acceptant seulement quelques chaînes de caractères spécifiques

Attardons-nous maintenant à un cas particulier de valeur par défaut en R. Supposons qu'une fonction possède un argument qui prend en entrée une chaîne de caractères et que seulement un petit nombre de chaînes de caractères distinctes sont acceptées pour cet argument. C'est le cas par exemple de l'argument `useNA` de la fonction `table`. La fonction accepte seulement les valeurs `"no"`, `"ifany"` ou `"always"` pour cet argument. Donner une valeur autre à l'argument produit une erreur.

```
table(iris$Species, useNA = "test")
```

```
## Error in match.arg(useNA): 'arg' should be one of "no", "ifany", "always"
```

Une pratique courante en R pour un argument de ce type est de lui donner comme valeur dans l'énumération des arguments le vecteur de toutes ses valeurs possibles. C'est ce qui est fait dans la fonction `table`.

```
args(table)
```

```
## function (... , exclude = if (useNA == "no") c(NA, NaN), useNA = c("no",  
##   "ifany", "always"), dnn = list.names(...), deparse.level = 1)  
## NULL
```

La valeur par défaut de l'argument n'est pas, dans ce cas, le vecteur complet `c("no", "ifany", "always")`, mais plutôt le premier élément de ce vecteur, soit `"no"`. Il en est ainsi, car le corps de la fonction contient l'instruction suivante.

```
useNA <- match.arg(useNA)
```

La fonction `match.arg` vérifie que la valeur donnée en entrée à un argument est bien une valeur acceptée ou retourne le premier élément du vecteur de valeurs possibles si aucune valeur n'a été donnée en entrée à l'argument.

Nous devrions reproduire cette façon de faire dans nos propres fonctions qui possèdent un argument du même type que l'argument `useNA` de la fonction `table`. Par exemple, remplaçons l'argument `sortie_matrice` de notre fonction `stats_desc` par l'argument `format_sortie` comme suit.

```
stats_desc <- function(x, format_sortie = c("vecteur", "matrice", "liste")) {  
  # Calcul  
  if (is.numeric(x)) {  
    stats <- c(min = min(x), moy = mean(x), max = max(x))  
  } else if (is.character(x) || is.factor(x)) {  
    stats <- table(x, dnn = NULL)  
  } else {  
    stats <- NA  
  }  
  # Production de la sortie  
  format_sortie <- match.arg(format_sortie)
```



```

if (format_sortie == "matrice") {
  stats <- as.matrix(stats)
  colnames(stats) <- if (is.character(x) || is.factor(x)) "frequence" else "stat"
} else if (format_sortie == "liste") {
  stats <- as.list(stats)
}
return(stats)
}

```

La valeur par défaut de l'argument `format_sortie` est bel et bien `"vecteur"`.

```
stats_desc(x = iris$Sepal.Length)
```

```
##      min      moy      max
## 4.300000 5.843333 7.900000
```

```
stats_desc(x = iris$Sepal.Length, format_sortie = "vecteur")
```

```
##      min      moy      max
## 4.300000 5.843333 7.900000
```

La présence du vecteur des chaînes de caractères possibles dans la définition des arguments est informative, car elle indique à l'utilisateur quelles valeurs sont acceptées par l'argument.

## 2.2 Appel d'une fonction

Les appels à nos propres fonctions respectent les mêmes règles que les [appels à n'importe quelle fonction R](#). En plus du fonctionnement des valeurs par défaut décrit ci-dessus, rappelons que des valeurs peuvent être fournies aux arguments d'une fonction R par nom complet, par nom partiel ou par position. L'assignation de valeurs aux arguments se fait en respectant les règles de préséances suivantes :

1. d'abord les arguments fournis avec un nom exact se voient assigner une valeur (assignation explicite),
2. puis les arguments fournis avec un nom partiel (encore assignation explicite),
3. et finalement les arguments non nommés, selon leurs positions (assignation implicite).

Voici quelques exemples.

```

test_appel <- function(x, option, param, parametre) {
  cat("l'argument x prend la valeur", x, "\n")
  cat("l'argument option prend la valeur", option, "\n")
  cat("l'argument param prend la valeur", param, "\n")
  cat("l'argument parametre prend la valeur", parametre, "\n")
}

```

```
test_appel(1, 2, 3, 4)
```

```
## l'argument x prend la valeur 1
## l'argument option prend la valeur 2
## l'argument param prend la valeur 3
## l'argument parametre prend la valeur 4
```

```
test_appel(1, 2, param = 3, opt = 4)
```

```
## l'argument x prend la valeur 1
## l'argument option prend la valeur 4
## l'argument param prend la valeur 3
## l'argument parametre prend la valeur 2
```

Le nom partiel `opt` a été reconnu comme représentant l'argument `option`. Un nom partiel pouvant représenter plus d'un argument génère cependant une erreur. C'est le cas du nom partiel `par` dans l'exemple suivant, qui pourrait référer à l'argument `param` ou encore à l'argument `parametre`.

```
test_appel(1, par = 2, option = 3, 4)
```

```
## Error in test_appel(1, par = 2, option = 3, 4): argument 2 matches multiple formal arguments
```

Une bonne pratique de programmation en R est d'utiliser l'assignation de valeurs aux arguments par positionnement seulement pour les premiers arguments, ceux les plus souvent utilisés. Les arguments moins communs devraient être nommés, en utilisant leurs noms complets, afin de conserver un code facile à comprendre.

## 2.3 Argument ...

Les deux utilités de l'argument `...` ont été mentionnées lors du premier cours. Nous pouvons recourir à cet argument dans nos propres fonctions, en exploitant l'une ou l'autre de ses utilités.

### 2.3.1 Utilité 1 : recevoir en entrée un nombre indéterminé de valeurs à traiter

L'argument `...` peut permettre de prendre en entrée un nombre indéterminé de valeurs à traiter. Voici un exemple de fonction qui exploite cette utilité de l'argument `...`. Il s'agit d'une extension de la fonction `stats_desc` qui permet de calculer des statistiques descriptives à partir d'observations d'un nombre quelconque de variables fournies sous forme de vecteurs ou facteurs.

```
stats_desc_multi <- function(...) {  
  args <- list(...)  
  lapply(X = args, FUN = stats_desc)  
}
```

Voici un exemple d'appel à cette fonction.

```
stats_desc_multi(iris$Sepal.Length, iris$Petal.Width, iris$Species)
```

```
## [[1]]  
##      min      moy      max  
## 4.300000 5.843333 7.900000  
##  
## [[2]]  
##      min      moy      max  
## 0.100000 1.199333 2.500000  
##  
## [[3]]  
##      setosa versicolor virginica  
##          50          50          50
```

Il est possible d'attribuer des noms aux valeurs attrapées par le `...`. Pour la fonction `stats_desc_multi`, ces noms deviennent les noms des éléments de la liste retournée en sortie.

```
stats_desc_multi(  
  Sepal.Length = iris$Sepal.Length,  
  Petal.Width = iris$Petal.Width,  
  Species = iris$Species  
)
```

```
## $Sepal.Length  
##      min      moy      max  
## 4.300000 5.843333 7.900000  
##
```

```
## $Petal.Width
##      min      moy      max
## 0.100000 1.199333 2.500000
##
## $Species
##      setosa versicolor virginica
##          50          50          50
```

Le corps d'une fonction possédant l'argument ... dans le but de recevoir autant de valeurs à traiter que désiré doit contenir une instruction pour récupérer les objets. Dans le corps de la fonction `stats_desc_multi`, ils sont récupérés par l'instruction `list(...)`. Les valeurs passées en entrée via ... peuvent aussi être référées dans le corps de la fonction par les noms `..1`, `..2`, `..3` et ainsi de suite.

```
test_trois_points <- function(...) {
  cat("le premier argument prend la valeur", ..1, "\n")
  cat("le deuxième argument prend la valeur", ..2, "\n")
}
```

```
test_trois_points("a", 1, TRUE)
```

```
## le premier argument prend la valeur a
## le deuxième argument prend la valeur 1
```

Il vaut cependant mieux s'assurer que suffisamment d'objets ont été passés en entrée avant d'utiliser un de ces noms. Par exemple, si la fonction `test_trois_points` est appelée en lui fournissant un seul objet en entrée, elle génère une erreur.

```
test_trois_points("a")
```

```
## Error in cat("le deuxième argument prend la valeur", ..2, "\n") :
## the ... list contains fewer than 2 elements
```

L'instruction `...length()` retourne le nombre d'objets attrapés par l'argument .... Utilisons cette instruction pour améliorer la fonction `test_trois_points` comme suit.

```
test_trois_points <- function(...) {
  if (...length() >= 1) {
    cat("le premier argument prend la valeur", ..1, "\n")
  }
  if (...length() >= 2) {
    cat("le deuxième argument prend la valeur", ..2, "\n")
  }
}
```

```
test_trois_points("a")
```

```
## le premier argument prend la valeur a
```

Nous pouvons faire encore mieux en bouclant sur les objets fournis en entrée, peu importe leur nombre. Dans ce cas, plutôt que d'utiliser les noms `..1`, `..2`, `..3` et ainsi de suite, nous utilisons la fonction `...elt`, qui prend en entrée un entier spécifiant une position et retourne l'objet attrapé par ... à cette position.

```
test_trois_points <- function(...) {
  for (i in seq_len(...length())) {
    cat("l'argument", i, "prend la valeur", ...elt(i), "\n")
  }
}
```

```
test_trois_points("a", 1, TRUE)
```

```
## l'argument 1 prend la valeur a
## l'argument 2 prend la valeur 1
## l'argument 3 prend la valeur TRUE
```

En résumé, lorsque l'argument ... sert à recevoir en entrée un nombre indéterminé de valeurs à traiter, les [instructions suivantes peuvent être utiles pour manipuler, dans le corps de la fonction, les valeurs attrapées par ...](#) :

- `list(...)`;
- `..1`, `..2`, `..3` et ainsi de suite;
- `...elt(n)` où `n` est un entier;
- `...length()`.

### 2.3.2 Utilité 2 : passer des arguments à une autre fonction

L'argument ... permet également de passer des arguments à une autre fonction appelée dans le corps de la fonction. Comme pour l'utilité 1, le nombre d'éléments attrapés par ... peuvent être en quantité quelconque. Cependant, ceux-ci ne vont cette fois pas être traités directement par la fonction principale. Ils vont seulement être transférés à une autre fonction, appelée par la fonction principale.

Par exemple, l'argument ... serait utile à notre fonction `stats_desc` pour contrôler la façon dont les appels aux fonctions `min`, `mean` et `max` traitent les valeurs manquantes. Dans le corps de la fonction `stats_desc`, les appels aux fonctions auxquelles nous souhaitons permettre le passage d'arguments (ici `min`, `mean` et `max`) doivent contenir l'argument ..., comme dans l'exemple suivant.

```
stats_desc <- function(x, format_sortie = c("vecteur", "matrice", "liste"), ...) {
  # Calcul
  if (is.numeric(x)) {
    stats <- c(min = min(x, ...), moy = mean(x, ...), max = max(x, ...)) # ... ici
  } else if (is.character(x) || is.factor(x)) {
    stats <- table(x, dnn = NULL)
  } else {
    stats <- NA
  }
  # Production de la sortie
  format_sortie <- match.arg(format_sortie)
  if (format_sortie == "matrice") {
    stats <- as.matrix(stats)
    colnames(stats) <- if (is.character(x) || is.factor(x)) "frequence" else "stat"
  } else if (format_sortie == "liste") {
    stats <- as.list(stats)
  }
  return(stats)
}
```

```
stats_desc(x = c(iris$Sepal.Length, NA))
```

```
## min moy max
## NA NA NA
```

```
stats_desc(x = c(iris$Sepal.Length, NA), na.rm = TRUE)
```

```
##      min      moy      max
## 4.300000 5.843333 7.900000
```

Unicité de l'argument ...

Tous les arguments d'une fonction doivent avoir des noms distincts. Par conséquent, une fonction ne peut pas contenir plus d'un argument ..., ce qui l'empêche d'exploiter simultanément les deux utilités de cet argument.

Il serait intéressant par exemple que la fonction `stats_desc_multi` puisse à la fois traiter un nombre indéterminé de variables fournies en entrée et passer un nombre indéterminé d'arguments à la fonction `stats_desc`, qu'elle appelle. Cependant, nous ne pourrions pas utiliser deux arguments ... pour permettre cela.

Une fonction mentionnée dans les [notes sur le prétraitement de données](#), la fonction `ave`, souffre de cette limite. L'argument ... que cette fonction possède permet à l'utilisateur de spécifier un nombre quelconque de variables de regroupement. Il n'est donc pas possible de passer des arguments à la fonction fournie à l'argument `FUN`. Par exemple, l'instruction suivante ne permet pas de calculer des moyennes tronquées de 20% des observations le plus grandes et de 20% des observations les plus petites.

```
ave(x = Orange$circumference, Orange$Tree, FUN = mean, trim = 0.2)
```

```
## [1] 99.57143 99.57143 99.57143 99.57143 99.57143 99.57143 99.57143 135.28571
## [9] 135.28571 135.28571 135.28571 135.28571 135.28571 135.28571 94.00000 94.00000
## [17] 94.00000 94.00000 94.00000 94.00000 94.00000 139.28571 139.28571 139.28571
## [25] 139.28571 139.28571 139.28571 139.28571 111.14286 111.14286 111.14286 111.14286
## [33] 111.14286 111.14286 111.14286
```

Même si l'instruction ne génère pas d'erreur, elle a traité `trim` comme une variable de regroupement (pour laquelle toutes les observations sont associées à la même modalité). Pour que l'argument `trim` de la fonction `mean` soit pris en compte, il faut créer une nouvelle fonction, comme suit.

```
ave(x = Orange$circumference, Orange$Tree, FUN = function(x) mean(x, trim = 0.2))
```

```
## [1] 104.4 104.4 104.4 104.4 104.4 104.4 104.4 142.2 142.2 142.2 142.2 142.2 142.2 142.2
## [15] 97.6 97.6 97.6 97.6 97.6 97.6 97.6 145.8 145.8 145.8 145.8 145.8 145.8 145.8
## [29] 114.2 114.2 114.2 114.2 114.2 114.2 114.2
```

## 3 Résultat produit par une fonction

Une fonction R peut retourner un objet en sortie et/ou produire un effet de bord, telles une impression, la production d'un graphique, l'écriture dans un fichier externe, la modification d'un paramètre de session, etc.

### 3.1 Sortie d'une fonction

Nous pourrions définir la sortie d'une fonction par ce qui est attrapé si l'appel à la fonction est accompagné d'une assignation à un nom. L'objet ainsi créé est la sortie de la fonction. Les expressions « valeur retournée » et « sortie produite » par un appel à une fonction sont synonymes. Dans la fiche d'aide d'une fonction R, la sortie produite est documentée dans la section intitulée « *Value* ».

Une fonction R retourne en sortie :

- ce que retourne l'expression donnée en argument à la fonction `return` dans le corps de la fonction (retour explicite) ;
- ou, en l'absence d'appel à la fonction `return`, ce que retourne la dernière expression évaluée dans le corps de la fonction (retour implicite).

Si cette dernière expression ne retourne rien, alors la fonction ne produit aucune sortie. Dans les exemples précédents, la fonction `HelloWorld` et les fonctions dont le nom débute par `test` ne contiennent aucun appel à la fonction `return` et possèdent un appel à la fonction `cat` comme dernière expression évaluée. La fonction `cat` imprime (en d'autres mots affiche) un résultat, ce qui constitue un type d'effet de bord, mais ne retourne rien.

```
sortie_cat <- cat("ceci est un test\n")
```

```
## ceci est un test
```

```
sortie_cat
```

```
## NULL
```

Nous constatons que l'objet `sortie_cat` prend la valeur `NULL`. Ne rien retourner en sortie signifie en réalité, pour une fonction R, retourner l'objet spécial `NULL`.

Les fonctions `test_appel` et `test_trois_points` ne produisent aucune sortie puisqu'elles retournent implicitement ce qu'un appel à la fonction `cat` retourne, c'est à dire rien. Elles provoquent cependant une impression comme effet de bord.

```
sortie_test <- test_trois_points("oui")
```

```
## l'argument 1 prend la valeur oui
```

```
sortie_test
```

```
## NULL
```

La fonction `stats_desc` utilise la fonction `return` à la dernière ligne du corps de la fonction pour retourner explicitement l'objet `stats`. Cependant, le corps de la fonction `stats_desc_multi` ne contient aucun appel à la fonction `return`. Celle-ci retourne donc le résultat de `lapply(X = args, FUN = stats_desc)`, car il s'agit de la dernière expression évaluée dans le corps de la fonction. Modifions cette dernière fonction pour expérimenter l'utilisation de la fonction `return`.

```
stats_desc_multi <- function(...) {  
  args <- list(...)  
  return(args)  
  cat("Est-ce que cette instruction est évaluée ?")  
  lapply(X = args, FUN = stats_desc)  
}
```

```
stats_desc_multi(rating = attitude$rating, complaints = attitude$complaints)
```

```
## $rating
```

```
## [1] 43 63 71 61 81 43 58 71 72 67 64 67 69 68 77 81 74 65 65 50 50 64 53 40 63 66 78 48 85 82  
##
```

```
## $complaints
```

```
## [1] 51 64 70 63 78 55 67 75 82 61 53 60 62 83 77 90 85 60 70 58 40 61 66 37 54 77 75 57 85 82
```

Cette version de la fonction `stats_desc_multi` retourne la liste des arguments fournis en entrée plutôt que le résultat de l'appel à `lapply` à cause de la présence de l'instruction `return(args)`. Les instructions suivant l'appel à la fonction `return` n'ont même pas été évaluées puisque l'impression demandée par l'appel à la fonction `cat` ajouté au corps de la fonction n'a pas été produite.

## 3.2 Retour de plusieurs éléments dans une liste

La sortie d'une fonction est toujours composée d'un seul objet. Pour retourner plusieurs objets, il faut les combiner en un seul objet (typiquement dans une liste), comme dans l'exemple suivant.

```
stats_desc_multi <- function(...) {  
  call <- match.call()  
  args <- list(...)  
  stats <- lapply(X = args, FUN = stats_desc)  
  return(list(stats = stats, call = call))  
}
```

```
stats_desc_multi(rating = attitude$rating, complaints = attitude$complaints)
```

```
## $stats
## $stats$rating
##      min      moy      max
## 40.00000 64.63333 85.00000
##
## $stats$complaints
##   min  moy  max
## 37.0 66.6 90.0
##
##
## $call
## stats_desc_multi(rating = attitude$rating, complaints = attitude$complaints)
```

Pour faciliter la réutilisation des résultats, il est souhaitable de toujours nommer les éléments d'une liste retournée en sortie.

### 3.3 Fonction `match.call`

L'exemple précédent fait intervenir la fonction `match.call`. Il est commun pour des fonctions d'ajustement de modèle telles que `lm` de retourner dans la sortie une copie de l'appel de la fonction.

```
exemple <- lm(rating ~ raises, data = attitude)
exemple$call
```

```
## lm(formula = rating ~ raises, data = attitude)
```

C'est la fonction `match.call` qui permet de créer cet élément de la sortie.

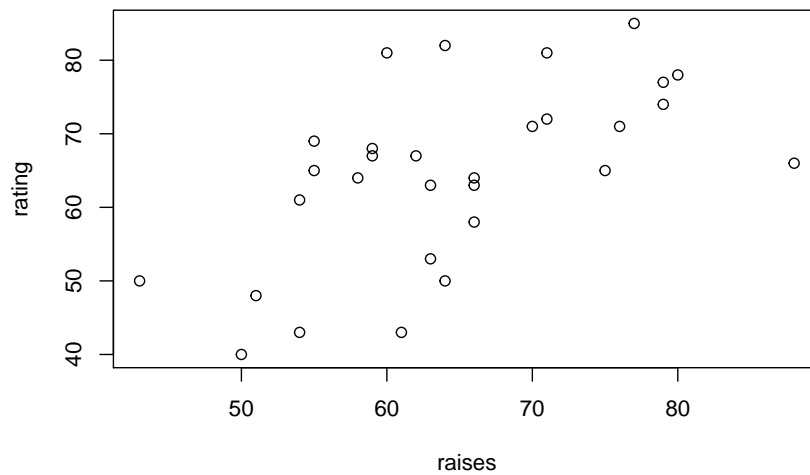
Les fonctions `match.call` et `return` sont des exemples de fonctions seulement utiles dans le corps d'une fonction. Les appeler directement dans la console retourne une erreur ou une sortie sans intérêt.

### 3.4 Effets de bord d'une fonction

Comme pour les fonctions `Hello World`, `test_appel` et `test_trois_points`, l'évaluation de l'appel à une fonction peut produire des « effets de bord » (en anglais *side effects*). Ces effets de bords peuvent être en réalité le but principal de la fonction.

En plus de la **production d'une impression** avec une fonction telle que `print` ou `cat`, un exemple courant d'effet de bord est la **production d'un graphique**.

```
sortie_plot <- plot(rating ~ raises, data = attitude)
```



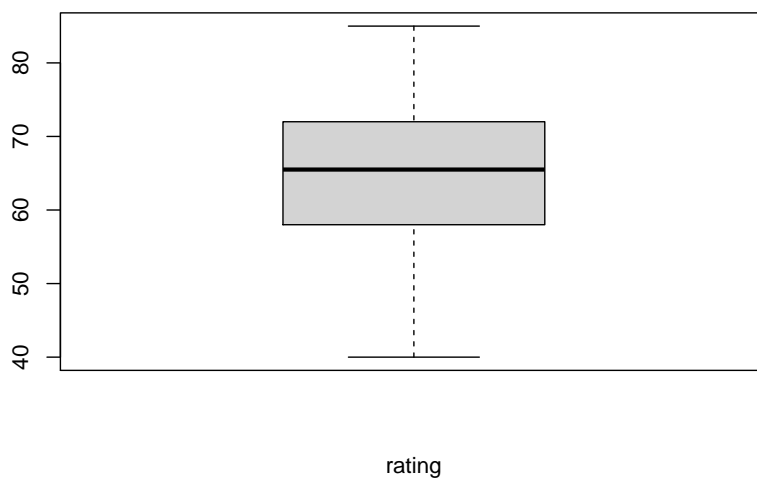
Certaines fonctions graphiques ne retournent pas d'objet.

```
sortie_plot
```

```
## NULL
```

Nous savons déjà cependant que d'autres fonctions graphiques retournent un objet en plus de produire un graphique. Il est donc possible pour une fonction R de produire à la fois un effet de bord et une sortie.

```
sortie_boxplot <- boxplot(x = attitude$rating, xlab = "rating")
```



```
sortie_boxplot
```

```
## $stats
##      [,1]
## [1,] 40.0
## [2,] 58.0
```



```
## [3,] 65.5
## [4,] 72.0
## [5,] 85.0
##
## $n
## [1] 30
##
## $conf
##          [,1]
## [1,] 61.46146
## [2,] 69.53854
##
## $out
## numeric(0)
##
## $group
## numeric(0)
##
## $names
## [1] ""
```

Un autre exemple d'effet de bord est **l'écriture dans un fichier externe**. Par exemple, la fonction `write.table` ne retourne rien dans l'environnement de travail de la session R, mais enregistre des données dans un fichier externe, sur le disque de l'ordinateur.

Finalement, toute **interaction avec l'environnement de travail ou la session R** peut être considérée comme un effet de bord. Les fonctions suivantes sont toutes des exemples de fonctions ayant des effets de bord :

- `library` : charge un package, ce qui modifie le chemin de recherche de R ;
- `setwd` : modifie le répertoire de travail ;
- `options` : modifie les options de la session R ;
- `par` : modifie les paramètres graphiques ;
- etc.

## 4 Évaluation d'un appel à une fonction et environnements associés

Lorsqu'une fonction R est appelée, un environnement est créé spécifiquement pour l'évaluation du corps de la fonction, puis **détruit lorsque l'exécution est terminée**. Rappelons que l'évaluation est simplement la façon dont R s'y prend pour comprendre ce qu'une instruction R signifie. Attardons-nous à comprendre comment R fait pour trouver la valeur d'un objet lorsqu'il évalue les instructions dans le corps d'une fonction.

Au départ, l'environnement créé lors de l'appel d'une fonction contient seulement des *promesses d'évaluation*, car R utilise une évaluation d'arguments dite  *paresseuse*. Il évalue un argument seulement lorsqu'une instruction du corps de la fonction le fait intervenir pour une première fois. Ainsi, au fur et à mesure que les instructions constituant le corps de la fonction sont évaluées, les arguments de la fonction deviennent des objets dans l'environnement temporaire créé spécifiquement pour l'évaluation de la fonction.

La valeur d'un argument auquel une expression a été assignée (explicitement ou implicitement) lors de l'appel de la fonction est obtenue en évaluant cette expression. La valeur d'un argument auquel aucune expression n'a été assignée est pour sa part obtenue en évaluant l'expression fournie comme valeur par défaut dans la définition de la fonction.

Les instructions formant le corps de la fonction créent parfois de nouveaux objets. Ceux-ci sont créés dans l'environnement d'évaluation de la fonction. En informatique, ces objets sont appelés **variables locales**.

## 4.1 Passage d'arguments par valeur

Lors de l'appel de la plupart des fonctions R, lorsqu'un objet est assigné à un argument, une copie de cet objet est créée et l'évaluation des instructions du corps de la fonction affecte cette copie et non l'objet d'origine. Ce mécanisme est appelé en informatique le « [passage d'arguments par valeur](#) ».

Un autre mécanisme possible de passage d'arguments est le « [passage par référence](#) ». Avec ce type de passage, les objets passés ne sont pas recopiés et les instructions du corps d'une fonction peuvent modifier l'objet d'origine. En R, ce type de passage est très rare. Une exception notable à cette observation est le package [data.table](#), qui utilise le passage par référence pour certaines de ses fonctions (notamment l'opérateur `:=`, les fonctions `setorder` et `setcolorder`), comme nous l'avons mentionné dans les [notes sur le prétraitement de données en R](#).

Illustrons ici le passage d'argument de loin le plus usuel en R : le passage d'arguments par valeur. Supposons que notre environnement de travail comporte un objet nommé `x` contenant le nombre 5.

```
x <- 5
x
```

```
## [1] 5
```

Créons une simple fonction R qui ajoute une unité à des nombres.

```
ajoute_1 <- function(x) x + 1
```

Maintenant, appelons cette fonction en lui donnant en entrée l'objet `x` de notre environnement de travail.

```
ajoute_1(x = x)
```

```
## [1] 6
```

La fonction retourne le résultat de `x + 1`, soit 6. Mais est-ce que l'objet `x` a pour autant changé ?

```
x
```

```
## [1] 5
```

Non. Il contient toujours la valeur 5.

Remarquez qu'ici le nom `x` a été utilisé pour deux entités distinctes :

- un objet dans notre environnement de travail,
- un argument de la fonction `ajoute_1`.

Dans l'instruction `ajoute_1(x = x)`, nous avons assigné la valeur contenue dans l'objet `x` à l'argument portant le même nom.

Comment pourrions-nous modifier l'objet `x` de notre environnement de travail à l'aide de la fonction `ajoute_1` ? Il faudrait assigner le résultat de l'instruction `ajoute_1(x = x)` au nom `x` comme suit.

```
x <- ajoute_1(x = x)
```

En fait, cette commande écrase l'ancien objet `x` par un nouveau, contenant la valeur retournée par `ajoute_1(x = x)`.

```
x
```

```
## [1] 6
```

## 4.2 Portée lexicale

Trouver la valeur des arguments et des variables locales en cours d'évaluation de l'appel à une fonction est simple pour R. Ces objets se trouvent directement dans l'environnement d'évaluation de la fonction. On dit en informatique qu'ils ont une **portée locale**.

Mais comment R trouve-t-il la valeur des objets appelés à l'intérieur d'une fonction, qui ne sont ni des arguments ni des variables locales ?

Chaque langage de programmation suit une certaine règle pour résoudre ce problème. Les deux règles les plus courantes sont l'utilisation d'une **portée lexicale** (en anglais *lexical scoping*) ou encore d'une **portée dynamique** (en anglais *dynamic scoping*).

Avec une portée lexicale, si un objet appelé n'est pas trouvé dans l'environnement d'évaluation de l'appel à une fonction, le programme va le chercher dans l'environnement d'où la fonction a été **créée**, nommé **environnement englobant** (en anglais *enclosing environment*). Avec une portée dynamique, le programme va plutôt le chercher dans l'environnement d'où la fonction a été **appelée**, nommé **environnement d'appel** (en anglais *calling environment*).

**R utilise la portée lexicale.**

Voici un petit exemple pour illustrer la portée lexicale.

```
a <- 1
b <- 2
f <- function(x) {
  a*x + b
}
```

Quelle valeur sera retournée par `f(x = 5)` ? Est-ce  $1*5 + 2 = 7$  ? Oui !

```
f(x = 5)
```

```
## [1] 7
```

Les objets nommés `a` et `b` ne se retrouvaient pas dans l'environnement d'exécution de la fonction. Alors R a cherché leurs valeurs dans l'environnement englobant de la fonction `f`, qui est ici l'environnement de travail.

```
environment(f)
```

```
## <environment: 0x0000017c65a6bed8>
```

Il a trouvé `a = 1` et `b = 2`. La fonction `environment` retourne l'environnement englobant d'une fonction.

Modifions maintenant l'exemple comme suit.

```
g <- function(x) {
  a <- 3
  b <- 4
  f(x = x)
}
```

Quelle valeur sera retournée par `g(x = 5)` ? Est-ce  $3*5 + 4 = 19$  ? Non !

```
g(x = 5)
```

```
## [1] 7
```

La fonction `g` est appelée dans l'environnement de travail. Elle appelle elle-même `f`. L'environnement d'appel de `f` est donc l'environnement d'exécution de `g`. Par contre, l'environnement englobant de `f` n'a pas changé. Il est encore l'environnement de travail, car c'est dans cet environnement que la fonction a été définie.

```
environment(f)
```

```
## <environment: 0x0000017c65a6bed8>
```

La portée lexicale permet de s'assurer que le fonctionnement de l'évaluation de l'appel à une fonction ne dépende pas du contexte dans lequel la fonction est appelée. Il dépend seulement de l'environnement d'où la fonction a été créée.

Si la portée en R était par défaut dynamique, `g(x = 5)` aurait retourné la valeur 19.

Et si `f` était créée à l'intérieur de la fonction `g` ?

```
g <- function(x) {  
  f <- function(x) {  
    a*x + b  
  }  
  a <- 3  
  b <- 4  
  f(x = x)  
}
```

Que retourne `g(x = 5)` maintenant ?

```
g(x = 5)
```

```
## [1] 19
```

L'environnement englobant de `f` est maintenant l'environnement d'exécution de `g`, car `f` a été défini dans le corps de la fonction `g`.

Notons que l'environnement englobant des fonctions disponibles en R autres que celles que nous avons créées en cours de session est l'espace de noms du package d'où provient la fonction. Par exemple, l'environnement englobant de la fonction `mean` est l'espace de noms du package `base`. Nous verrons ce qu'est un espace de noms dans les [notes sur l'utilisation de packages R](#).

```
environment(mean)
```

```
## <environment: namespace:base>
```

### 4.3 Chemin de recherche complet

Le chemin de recherche des valeurs des objets lors de l'évaluation de l'appel à une fonction en R ne s'arrête pas à l'environnement d'exécution de la fonction suivi de l'environnement englobant de la fonction. Si l'environnement englobant est l'environnement d'exécution d'une autre fonction, alors la recherche se poursuit dans l'environnement englobant de cette fonction. Plusieurs environnements englobants peuvent ainsi être imbriqués. Ensuite, l'environnement de travail est toujours ajouté au chemin de recherche s'il n'est pas directement l'environnement englobant (potentiellement à la base d'une hiérarchie d'environnements englobants). Et de là, le chemin de recherche se poursuit par les environnements de tous les packages chargés, tel que vu dans les [notes sur des informations techniques concernant R](#). Nous pouvons donc utiliser, dans les fonctions que nous créons, des fonctions provenant d'autres packages. Il faut seulement s'assurer que ces packages soient chargés pour que nos fonctions roulent sans erreur.

Dans le dernier exemple, le chemin de recherche complet lors de l'évaluation de l'appel à la fonction `f`, dans le corps de la fonction `g`, alors que `f` a aussi été créé dans le corps de `g`, est le suivant.

### 4.4 Bonnes pratiques dans la création d'une fonction

Il est recommandé d'utiliser dans une fonction uniquement des objets que nous sommes certains de pouvoir atteindre. L'idéal est de se limiter aux arguments de la fonction, aux objets créés dans la fonction (variables locales) ainsi qu'aux objets se trouvant dans des packages *chargés*.

Ceux qui comprennent bien le concept de portée lexical peuvent aussi s'amuser à utiliser des objets dans l'environnement englobant d'une fonction.

Cependant, il est risqué d'utiliser les objets de l'environnement de travail, même si cet environnement se retrouve toujours dans le chemin de recherche de valeurs des objets lors de l'évaluation de l'appel à une fonction. Le contenu de l'environnement de travail est constamment modifié au fil de nos sessions. Aussi, si

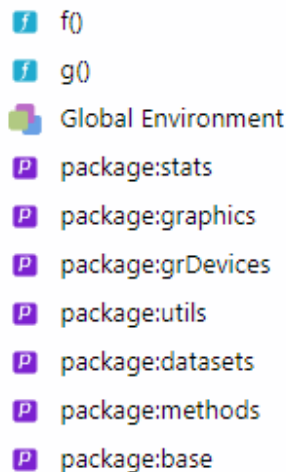


FIGURE 1 – Exemple de chemin de recherche lors de l'évaluation de l'appel à une fonction R

nous partageons nos fonctions avec une autre personne, nous ne contrôlons pas le contenu de l'environnement de travail pendant la session R de cette personne.

Ces recommandations s'appliquent au code dans le corps d'une fonction, mais aussi aux instructions définissant les valeurs par défaut des arguments. Nous avons appris que ces instructions sont évaluées dans le corps de la fonction. Elles peuvent donc contenir sans problème d'autres arguments de la fonction. Cependant, nous devrions éviter d'utiliser des objets provenant de l'environnement de travail dans ces instructions.

## 5 Exemple de création d'une fonction R

Nous allons créer ensemble une fonction qui compte combien de nombres entiers impairs contient un vecteur numérique (cet exemple est tiré de [1]).

### 5.1 Étapes de développement conseillées

1. **Planifier** le travail (pas de programmation encore)
  - définir clairement la tâche à accomplir par la fonction et la sortie qu'elle doit produire ;
  - prévoir les étapes à suivre afin d'effectuer cette tâche ;
  - identifier les arguments devant être fournis en entrée à la fonction.
2. **Développer le corps de la fonction**
  - écrire le programme par étapes, d'abord sans former la fonction, en commentant bien le code et en travaillant sur des mini-données test ;
  - pour chaque petite étape ou sous-tâche, tester interactivement si le programme produit le résultat escompté (tester souvent en cours de travail, ainsi il y a moins de débogage à faire).
3. **Créer la fonction** à partir du programme développé.
4. **Documenter** la fonction : les informations minimales à fournir sont
  - ce que la fonction fait,
  - quels arguments la fonction accepte en entrée (utilités, formats acceptés, valeurs par défaut),
  - ce que la fonction produit comme résultat (sortie retournée et/ou effet de bord).

D'autres étapes de développement seront abordées au prochain cours.

#### 1. Planifier le travail :

- entrée = un vecteur de nombres (= 1 seul argument)
- sortie = le dénombrement (une seule valeur)

- utiliser l'opérateur modulo pour tester si un nombre est impair
- nous pourrions travailler de façon vectorielle ou encore utiliser une boucle sur les éléments du vecteur

## 2. Développer le corps de la fonction :

Création de mini-données test

```
x <- c(6, 3, 5.5, 1, 0, -5)
```

Ce vecteur contient 3 nombres entiers impairs. C'est le résultat que nous visons obtenir.

Code le plus simple qui me vient en tête :

```
sum(x %% 2 == 1)
```

```
## [1] 3
```

Nous obtenons bien 3. Ça marche pour les mini-données test.

Ce code est équivalent à la boucle suivante :

```
k <- 0
for (n in x) {
  if (n %% 2 == 1) {
    k <- k + 1
  }
}
k
```

```
## [1] 3
```

## 3. Créer la fonction à partir du programme développé :

```
compte_impair_vectoriel <- function(x) {
  sum(x %% 2 == 1)
}
```

```
compte_impair_boucle <- function(x) {
  k <- 0
  for (n in x) {
    if (n %% 2 == 1) {
      k <- k + 1
    }
  }
  k
}
```

## 4. Documenter la fonction :

**Option 1** : Documentation en commentaire dans le corps de la fonction.

```
compte_impair_vectoriel <- function(x) {
  # Fonction qui compte combien de nombres entiers impairs contient un vecteur numérique
  # Argument en entrée : x = vecteur numérique
  # Sortie : le nombre de nombres entiers impairs dans x
  sum(x %% 2 == 1)
}
```

**Option 2** : Documentation en commentaire avant la définition de la fonction.

```

# Fonction qui compte combien de nombres entiers impairs contient un vecteur numérique
# Argument en entrée : x = vecteur numérique
# Sortie : le nombre de nombres entiers impairs dans x
compte_impair_boucle <- function(x) {
  k <- 0
  for (n in x) {
    if (n %% 2 == 1) {
      k <- k + 1
    }
  }
  k
}

```

**Options supplémentaires :** Nous verrons d'autres options dans le cours sur les packages.

### Comparaison des 2 fonctions :

Nous avons créé 2 fonctions qui, à première vue, retournent toutes les deux le résultat escompté. Nous devrions par contre les tester sur plus de données pour en être certains. Tenons pour acquis que ces fonctions accomplissent correctement leur tâche.

Dans ce cas, laquelle des 2 fonctions devrions-nous utiliser ?

Réponse : la plus rapide.

Créons un vecteur très grand pour comparer le temps d'exécution des deux fonctions.

```
x <- round(runif(1000000, -10, 10))
```

Utilisons la fonction `system.time` pour évaluer les temps d'exécution.

```
system.time(compte_impair_vectoriel(x))
```

```
##      user  system elapsed
##    0.03    0.00    0.03
```

```
system.time(compte_impair_boucle(x))
```

```
##      user  system elapsed
##    0.78    0.00    0.78
```

L'écart dans les temps d'exécution des deux fonctions se creuse encore plus si nous augmentons la longueur du vecteur `x`.

Nous devrions donc choisir d'utiliser `compte_impair_vectoriel` plutôt que `compte_impair_boucle`.

Nous allons revenir plus tard sur l'optimisation des temps d'exécution de nos fonctions.

## 6 Programmation fonctionnelle

Maintenant que vous savez écrire des fonctions en R, vous pouvez exploiter tout le potentiel de la [programmation fonctionnelle](#), paradigme de programmation exploité par R. En fait, nous avons déjà parlé de ce paradigme dans ce cours. L'[utilisation de fonctions de la famille des `apply`](#) est une forme de programmation fonctionnelle. Contentons nous ici de parler de cet aspect de la programmation fonctionnelle : les fonctions de haut niveau qui prennent d'autres fonctions en entrée, comme les fonctions de la famille des `apply`.

Nous avons déjà donné dans les [notes sur les structures de contrôle](#) un exemple de boucle `for` remplacé par un appel à une fonction de la famille des `apply`. En fait, pratiquement n'importe quelle boucle `for` en R peut être remplacée par un appel à une fonction de la famille des `apply` une fois que nous savons comment

créer de nouvelles fonctions. Par exemple, reprenons l'exemple de boucle suivant, aussi tiré des [notes sur les structures de contrôle](#).

```
modeles <- vector(length = ncol(attitude) - 1, mode = "list")
names(modeles) <- setdiff(names(attitude), "rating")

for (variable in names(modeles)) {
  modeles[[variable]] <- lm(rating ~ ., data = attitude[, c("rating", variable)])
}
```

Il s'agit d'une boucle ajustant plusieurs modèles de régression linéaire simple avec les variables du jeu de données `attitude`. Modifions un peu cet exemple pour conserver des modèles uniquement les coefficients de détermination, non ajustés et ajustés.

```
R2 <- matrix(NA, nrow = 2, ncol = ncol(attitude) - 1)
colnames(R2) <- setdiff(names(attitude), "rating")
rownames(R2) <- c("r.squared", "adj.r.squared")

for (variable in colnames(R2)) {
  reg <- lm(rating ~ ., data = attitude[, c("rating", variable)])
  R2["r.squared", variable] <- summary(reg)$r.squared
  R2["adj.r.squared", variable] <- summary(reg)$adj.r.squared
}
```

R2

```
##               complaints privileges learning raises critical advance
## r.squared      0.6813142  0.1815756 0.3889745 0.3482640 0.02447321 0.02405175
## adj.r.squared  0.6699325  0.1523461 0.3671521 0.3249877 -0.01036703 -0.01080355
```

À l'aide de la bonne fonction, nous allons obtenir exactement le même résultat avec un code plus court, qui n'utilise pas de boucle, mais qui utilise la fonction `sapply`. Voici une fonction qui extrait les deux statistiques à conserver pour un seul modèle de régression.

```
R2_reg_rating_vs_var <- function(variable) {
  reg <- lm(rating ~ ., data = attitude[, c("rating", variable)])
  summary(reg)[c("r.squared", "adj.r.squared")]
}
```

Cette fonction prend en entrée le nom d'une variable provenant de `attitude`, mais autre que la variable réponse `rating`. Nous pouvons itérer sur tous les noms de variables possibles comme suit :

```
sapply(
  X = setdiff(names(attitude), "rating"),
  FUN = R2_reg_rating_vs_var
)
```

```
##               complaints privileges learning raises critical advance
## r.squared      0.6813142  0.1815756 0.3889745 0.348264 0.02447321 0.02405175
## adj.r.squared  0.6699325  0.1523461 0.3671521 0.3249877 -0.01036703 -0.01080355
```

Un des avantages de l'utilisation de fonctions de la famille des `apply` en remplacement de boucles est que nous n'avons pas à initialiser préalablement un objet pour stocker les résultats. La fonction de haut niveau gère cet aspect pour nous. Les adeptes de la programmation fonctionnelle sont également d'avis que l'utilisation de fonction de la famille des `apply` produit un code plus clair qu'une boucle.

Il existe en R d'autres fonctions de haut niveau, qui appliquent itérativement une fonction sur les éléments d'un objet. Pour les intéressés, la fiche d'aide ouverte par la commande `help(funprog)` présente ces fonctions (`Map`, `Filter`, `Reduce`, etc.). Un package R se spécialise aussi dans ce genre de fonctions, il s'agit du [package](#)



purrr.

---

## 7 Résumé

### Syntaxe générale d'une fonction

```
<nom_fonction> <- function(<arg_1>, <arg_2>, <arg_3>) {  
  <instructions> # formant le corps de la fonction  
}
```

Les composantes d'une fonction R sont :

- la liste de ses arguments, possiblement avec des valeurs par défaut ;
- le corps de la fonction, soit le code qui la constitue ;
- l'environnement englobant de la fonction.

**Note** : Il n'est pas obligatoire pour une fonction de porter un nom, ni de posséder des arguments.

### Valeurs par défaut des arguments

Les valeurs par défaut sont **définies dans la liste des arguments**, en accompagnant le nom d'un argument d'un **opérateur =** et d'une **expression R retournant la valeur par défaut**.

**Cas particulier** : argument qui prend en entrée **une seule chaîne de caractères** et que seulement un **petit nombre de chaînes de caractères distinctes sont acceptées** comme valeur de cet argument :

- valeur à droite de l'opérateur = dans la liste des arguments : vecteur de toutes les valeurs acceptées,
- valeur par défaut : élément en position 1 dans le vecteur de toutes les valeurs acceptées,
- dans le corps de la fonction : `arg <- match.arg(arg)`.

### Appel d'une fonction

L'assignation de valeurs aux arguments se fait en respectant les règles de préséances suivantes :

1. d'abord les arguments fournis avec un nom exact se voient assigner une valeur (assignation explicite),
2. puis les arguments fournis avec un nom partiel (encore assignation explicite),
3. et finalement les arguments non nommés, selon leurs positions (assignation implicite).

### Bonne pratique :

- utiliser l'association par positionnement seulement pour les premiers arguments,
- arguments moins communs nommés (code plus facile à comprendre).

### Argument ...

Nous pouvons insérer l'argument `...` dans la liste des arguments des fonctions que nous créons. Dans le corps de la fonction, le traitement de cet argument dépend de son utilité.

- Pour **prendre en entrée un nombre indéterminé de valeurs à traiter** :
  - Le corps de la fonction doit contenir des instructions pour récupérer les éléments attrapés :
    - \* `list(...)` ou ;
    - \* `..1, ..2, ..3` et ainsi de suite ou ;
    - \* `...elt(n)` où `n` est un entier.
  - (`...length()` retourne le nombre d'éléments attrapés par `...`)
- Pour **permettre le passage d'arguments** à une autre fonction :
  - Dans le corps de la fonction, les appels à la ou aux fonctions auxquelles nous souhaitons permettre le passage d'arguments doivent contenir l'argument `...`

## Résultat produit

Une fonction **retourne en sortie** :

- ce que retourne l'expression donnée en argument à la fonction **return** dans le corps de la fonction (retour explicite) ;
- ou, en l'absence d'appel à la fonction **return**, ce que retourne la dernière expression évaluée dans le corps de la fonction (retour implicite).

La sortie d'une fonction est toujours composée d'**un seul objet**. Pour retourner plusieurs objets, il faut les combiner dans un seul objet (typiquement dans une liste).

La fonction `match.call` permet d'obtenir une copie de l'appel de la fonction.

Une fonction peut également produire un ou des **effets de bord** : une impression, la production d'un graphique, l'écriture dans un fichier externe, une interaction avec l'environnement de travail ou la session R, etc.

## Évaluation d'un appel à une fonction

Lorsqu'une fonction R est appelée, un environnement est créé spécifiquement pour l'évaluation du corps de la fonction, puis détruit lorsque l'exécution est terminée.

→ **environnement d'exécution** (ou d'évaluation) = temporaire

Cet environnement contient :

- des objets pour les arguments fournis dans l'appel ;
- des objets pour les arguments non fournis dans l'appel, qui prennent leur valeur par défaut ;
- des objets créés dans les instructions du corps de la fonction (variables locales).

**Évaluation paresseuse** : Les objets associés aux arguments sont créés uniquement lorsqu'une instruction du corps de la fonction les faisant intervenir doit être évaluée. À sa création, l'environnement d'exécution contient uniquement des promesses d'évaluation.

## Passage des arguments par valeur

La grande majorité du temps en R, les objets assignés à des arguments dans un appel à une fonction R sont recopiés et l'évaluation de la fonction affecte ces copies. Elle n'affecte pas les objets d'origine.

## Portée lexicale

Comment R trouve-t-il la valeur des objets appelés dans les instructions du corps d'une fonction qui ne sont ni des arguments ni des variables locales ?

Il les cherche dans l'**environnement englobant** de la fonction = *environnement dans lequel la fonction a été créée*. R utilise donc une **portée lexicale**.

À ne pas confondre : R ne cherche pas dans l'**environnement d'appel** = *environnement dans lequel la fonction est appelée* (à moins que l'environnement englobant soit le même que l'environnement d'appel).

## Chemin de recherche complet lors de l'évaluation de l'appel à une fonction

- environnement d'exécution,
- environnement englobant ou hiérarchie d'environnements englobants,
- environnement de travail,
- environnements des packages chargés.

**Bonne pratique** : Utiliser dans une fonction uniquement des objets que nous sommes certains de pouvoir atteindre, soit

- les arguments de la fonction,

- les objets créés dans la fonction (variables locales),
- les objets se trouvant dans des packages chargés,
- les objets dans l'environnement englobant (si nous comprenons bien le concept de portée lexicale).

**Ne pas utiliser les objets de l'environnement de travail**, car le contenu de cet environnement est constamment modifié.

---

## Références

### Référence citée dans le texte :

[1] Matloff, N. (2011). *The Art of R Programming : A Tour of Statistical Software Design*. No Starch Press. Sections 1.3 et 7.4

### Références supplémentaires :

- Wickham, H. (2019). *Advanced R*. 2e édition. Chapman and Hall/CRC.
  - Fonctions : Chapitre 6 <https://adv-r.hadley.nz/functions.html>
  - Environnements : Chapitre 7 <https://adv-r.hadley.nz/environments.html>
- Fanara, C. (2019). Tutoriel web intitulé « *A Tutorial on Using Functions in R!* ». <https://www.datacamp.com/community/tutorials/functions-in-r-a-tutorial>
- Passage d'arguments par valeur versus par référence : <http://www.mathwarehouse.com/programming/passing-by-value-vs-by-reference-visual-explanation.php>
- Pour en apprendre davantage concernant la programmation fonctionnelle :
  - <https://adv-r.hadley.nz/fp.html>
  - <https://r4ds.had.co.nz/iteration.html#for-loops-vs.functionals>
  - <https://purrr.tidyverse.org/>
  - <https://speakerdeck.com/hadley/the-joy-of-functional-programming>