# Développement de packages R

# Sophie Baillargeon, Université Laval

# 2020-03-24

# Table des matières

Etape 1. Ecrire les fonctions	2
Étape 2. Créer la structure de fichiers du package  Répertoire principal portant le nom du package  Sous-répertoire nommé R  Sous-répertoire nommé man  Fichier nommé DESCRIPTION  Fichier nommé NAMESPACE  Autres fichiers et répertoires	3 3 3 3 4 5
Étape 3. Écrire la documentation des fonctions et du package  Comment écrire de la documentation avec roxygen2?	5 6 10
Lancement des commandes de construction et de vérification en RStudio	11 12 12 13 13 14
Étape 5. Si désiré, partager le package	14
Étape 6. Au besoin, mettre à jour le package	14
Résumé	16
Références	18
Annexe Résolution de problèmes déjà rencontrés	18 18

Pour créer un package R, il faut d'abord développer une structure de fichiers similaires à ce qui se retrouve dans n'importe quel *package source*. Pour voir de quoi à l'air une telle structure de fichier, il suffit de :

<sup>•</sup> télécharger un package source à partir du CRAN (par exemple ggplot2\_3.1.0.tar.gz sur http://CRAN.R-project.org/package=ggplot2),

<sup>•</sup> décompresser le fichier (il y a 2 étapes de décompression à effectuer).

L'exemple utilisé ici, soit le package ggplot2, est un gros package. Seuls certains de ses fichiers et répertoires sont nécessaires dans tout package R.

Après avoir développé les fichiers sources, qui incluent du code et de la documentation, il faut construire le package. Nous verrons ici comment réaliser cette construction avec RStudio.

# Étape 1. Écrire les fonctions

Cette étape fait appel à ce que nous avons appris dans toutes les notes de cours précédentes, en particulier celles sur les fonctions en R.

Lorsque nous créons des fonctions dans le but de les inclure dans un package, il faut garder en tête le chemin de recherche complet qu'utilisera R lors de l'exécution de ces fonctions. Ce chemin est le suivant :

- 1. l'environnement local d'exécution de la fonction,
- 2. l'environnement englobant de la fonction, soit l'espace de noms du package,
- 3. l'environnement des objets importés par le package,
- 4. l'environnement du package R de base,
- 5. l'environnement de travail,
- 6. les environnements de tous les packages chargés.

Nous savons déjà que, dans ce chemin de recherche, nous ne pouvons pas nous fier au contenu de l'environnement de travail, car il varie constamment en cours de session. Nous ne pouvons pas non plus nous fier aux packages chargés. En effet, la liste des packages chargés varie aussi en cours de session et d'un utilisateur à l'autre. Le bon fonctionnement des fonctions d'un package devrait dépendre d'un seul appel à la fonction library, celui servant à charger le package en question.

Ainsi, dans le corps des fonctions d'un package, nous pouvons toujours faire appel

- aux arguments et variables locales,
- à tout objet, public ou privé, contenu dans le package,
- à des objets du package R de base, qui est toujours inclus dans le chemin de recherche.

Si nous souhaitons faire appel à des objets autres, par exemple des fonctions provenant d'autres packages, il faudra faire le nécessaire pour inclure ces fonctions dans *l'environnement des objets importés par le package*. Aussi, dans le corps de nos fonctions, il est recommandé de faire appel à ces objets en utilisant l'opérateur : : pour indiquer clairement de quel package ils proviennent. Par exemple, si nous voulons utiliser la fonction dist du package stats, il est bien de l'appeler comme suit : stats::dist(...).

Aussi, si notre but est éventuellement de rendre notre package public sur le CRAN, il faut respecter les politiques du CRAN : <a href="http://cran.r-project.org/web/packages/policies.html">http://cran.r-project.org/web/packages/policies.html</a>. La plus importante de ces politiques est que notre package doit passer le R CMD check --as-cran sans erreurs ni avertissements (<a href="https://cran.r-project.org/doc/manuals/r-release/R-exts.html#Checking-packages">https://cran.r-project.org/doc/manuals/r-release/R-exts.html#Checking-packages</a>, nous en reparlons plus loin). Pour passer cette vérification sans problèmes, notre code ne devrait pas, entre autres :

- contenir des symboles non-ASCII (donc pas d'accents) : afin de s'assurer que le code est fonctionnel sur n'importe quelle plateforme informatique ;
- utiliser d'association partielle d'argument (donc nous devons utiliser le nom complet des arguments dans les appels de fonction) :

  afin de s'assurer que le code demeure fonctionnel si des arguments sont ajoutés aux définitions des
  - afin de s'assurer que le code demeure fonctionnel si des arguments sont ajoutés aux définitions des fonctions appelées dans le code (ces nouveaux arguments pourraient porter des noms qui entrent en conflit avec l'association partielle);
- toujours utiliser comme valeurs logiques TRUE ou FALSE (donc pas de T ou F):

  parce qu'un objet R nommé T ou F peut être défini, écrasant du coup les définitions T <- TRUE et F <FALSE, alors qu'il est impossible de créer un nouvel objet R nommé TRUE ou FALSE (mots-clés protégés).

# Étape 2. Créer la structure de fichiers du package

Un package, dans sa version source, est simplement un répertoire de fichiers compressés. Cependant, ce répertoire doit comprendre des fichiers et des sous-répertoires portant des noms précis.

Nous pouvons créer cette structure de fichiers à l'aide de :

- la fonction package.skeleton du package utils ou
- la fonction create\_package du package usethis ou
- les fonctionnalités de création de projets RStudio.

Cependant, nous pouvons aussi créer manuellement les répertoires et fichiers. Peu importe la procédure utilisée, il faut d'abord avoir un répertoire principal portant le nom du package. Le contenu de base d'un répertoire de package est le suivant :

- sous-répertoire nommé R,
- sous-répertoire nommé man,
- fichier nommé DESCRIPTION,
- fichier nommé NAMESPACE.

# Répertoire principal portant le nom du package

Premièrement, un nom doit être choisi pour le package. Ce nom peut contenir uniquement les caractères ASCII suivants : lettres, chiffres et . (point). Il ne peut donc pas contenir de lettres accentuées, ni le caractère \_ (tiret bas ou *underscore* en anglais). Il doit être composé d'au moins 2 caractères, débuter par une lettre et ne pas terminer par un point<sup>1</sup>.

Ensuite, il faut créer un répertoire portant ce nom. Par exemple, si nous voulions créer un package nommé manhattan, il faudrait d'abord créer un répertoire nommé manhattan sur notre ordinateur. Ce pourrait être par exemple le répertoire "C:/coursR/manhattan".

# Sous-répertoire nommé R

Le sous-répertoire R doit comprendre des fichiers de scripts R (portant l'extension .R) contenant le code de création des fonctions ainsi que les commentaires roxygen2, si c'est l'outil que nous utilisons pour générer la documentation (nous y reviendrons plus loin). Le développeur peut choisir les noms qu'il veut pour ces fichiers et il peut y répartir le code source comme il le veut. Évidemment, une bonne pratique est de répartir le code source de façon à ce que ce soit facile de s'y retrouver. Certains préconisent la stratégie « un fichier par fonction, portant le nom de la fonction », mais ce n'est pas obligatoire. La stratégie opposée, « un seul fichier contenant toutes les fonctions », représente rarement une bonne répartition du code, à moins que les fonctions soient très peu nombreuses.

# Sous-répertoire nommé man

Le sous-répertoire man doit comprendre des fichiers sources de fiches d'aide R (portant l'extension .Rd). Nous allons voir à la prochaine étape comment générer ces fichiers à l'aide du package roxygen2.

### Fichier nommé DESCRIPTION

L fichier 'DESCRIPTION est très important pour la construction du package. C'est un fichier court, comportant de l'information de base sur le package. Ce fichier doit respecter une syntaxe précise.

## Exemple:

<sup>&</sup>lt;sup>1</sup>https://cran.r-project.org/doc/manuals/r-release/R-exts.html#The-DESCRIPTION-file

Package: manhattan Version: 1.0.0 Date: 2019-03-21 License: GPL-3

Title: Distance de Manhattan

Description: Calcul de la distance de Manhattan entre deux points.

Author: Sophie Baillargeon [aut, cre],

Autre Auteur [aut]

Maintainer: Sophie Baillargeon <sophie.baillargeon@mat.ulaval.ca>

Imports: stats
LazyData: true

Ce fichier contient des champs nommés. Un champ débute par son nom (première lettre toujours majuscule), immédiatement suivi d'un deux-points et d'un espace. Vient ensuite la valeur fournie à ce champ. Les valeurs données aux champs peuvent s'étendre sur plus d'une ligne.

Les champs obligatoires sont les suivants : Package, Version, License, Title, Description, Author, et Maintainer.

Les champs Author, et Maintainer peuvent être remplacés par un champ Authors@R. Les codes de rôles acceptés dans ces champs sont énumérés dans la fiche d'aide ouverte pas la commande help(person). Par exemple, le champ Authors@R suivant pourrait remplacer les champs Author, et Maintainer dans l'exemple précédent.

Voici quelques informations spécifiques à certains champs :

- Package : le nom du package fourni dans ce champ doit correspondre parfaitement au nom du répertoire contenant les fichiers sources du package ;
- Version : il s'agit d'une séquence d'au minimum 2 (souvent 3) nombres entiers non négatifs séparés par un seul caractère . (point) ou (tiret) ;
- Maintainer : ce champ doit contenir un seul nom, celui de la personne à contacter pour toute question ou tout problème à rapporter concernant le package, suivi d'une adresse courriel valide placée entre les caractères < et > ;
- Imports : ce champ contient la liste des packages à partir desquels des objets sont importés (nous laisserons roxygen2 écrire ce champ pour nous);
- LazyData : si la valeur true est fournie dans ce champ, les jeux de données contenus dans le package seront directement accessibles dans l'environnement du package, sans que l'utilisateur ait besoin d'utiliser la commande data (nous avions discuté des différentes façons d'accéder à des jeux de données dans un package dans les notes sur l'utilisation de packages R).

Le fichier DESCRIPTION doit contenir uniquement des caractères ASCII, sauf s'il contient un champ Encoding.

L'information complète et officielle concernant ce fichier peut être trouvée sur la page web suivante :  $\frac{\text{http:}}{\text{cran.r-project.org/doc/manuals/r-release/R-exts.html}\#\text{The-DESCRIPTION-file}$ 

# Fichier nommé NAMESPACE

Le fichier NAMESPACE permet de définir quels objets sont accessibles dans un package, c'est-à-dire exportés de l'espace de noms du package. Ce fichier permet aussi d'identifier quels objets provenant d'autres packages sont utilisés dans le package, donc de définir le contenu de l'environnement des objets importés par le package. Nous allons voir comment utiliser roxygen2 pour générer ce fichier.

# Autres fichiers et répertoires

Un package peut contenir plusieurs autres fichiers et sous-répertoires. Voici quelques autres répertoires parfois nécessaires :

- Sous-répertoire nommé data : Si le package contient des jeux de données, ils se trouvent habituellement dans ce répertoire. Ceux-ci sont typiquement dans un format de données R (souvent .rda ou .rdata).
- Sous-répertoire nommé **src** : Si le package contient du code C, C++ ou Fortran, ce code doit se trouver dans ce répertoire.
- Sous-répertoire nommé vignettes : Si le package contient de la documentation autre que les fiches d'aide (par exemple un guide d'utilisateur), il devrait idéalement se trouver dans ce répertoire.
- $\bullet \ \ etc.: http://cran.r-project.org/doc/manuals/r-release/R-exts.html \# Package-subdirectories \\$

#### Autres fichiers utiles:

- Fichier nommé NEWS : Ce fichier décrit les modifications apportées à un package lors d'une mise à jour. Sa mise en forme n'est pas vraiment importante.
- etc.: http://cran.r-project.org/doc/manuals/r-release/R-exts.html#Package-structure

# Étape 3. Écrire la documentation des fonctions et du package

Documenter ses fonctions est une étape intégrée au développement des fonctions. Cependant, il faut maintenant écrire la documentation dans un format qui produira correctement les fiches d'aide qui doivent être incluses dans le package.

Ces fiches d'aide proviennent en fait de fichiers portant l'extension .Rd. Cependant, nous ne verrons pas comment éditer directement ces fichiers. Nous allons plutôt apprendre à utiliser le package roxygen2, qui permet de générer des fichiers .Rd, ainsi que le fichier NAMESAPCE, de façon automatique, à partir de commentaires intégrés au code.

L'utilisation de roxygen2 comporte les avantages suivants :

- la documentation se situe dans le même fichier que le code, ce qui aide à se rappeler que la documentation doit être mise à jour si le code est modifié,
- la syntaxe roxygen2 est un peu plus simple que la syntaxe des fichiers .Rd.

Rappelons que dans un package il faut obligatoirement documenter :

# Les fonctions publiques

Il est essentiel de faire des fiches d'aide pour ces fonctions, afin que tout utilisateur comprenne comment appeler correctement la fonction. Si un utilisateur a mal compris la documentation et fournit par erreur une valeur d'argument invalide en entrée, il est aussi souhaitable que la fonction retourne une erreur informative. Le code de ces fonctions publiques comporte donc typiquement de la validation des arguments fournis en entrée.

À l'opposé, les **fonctions privées ou internes** ne sont pas conçues pour être appelées par n'importe quel utilisateur. Ces fonctions ne sont pas exportées de l'espace de noms et elles ne sont pas documentées officiellement. Il est tout de même bon de documenter minimalement ces fonctions pour nous-mêmes, mais nous n'avons pas à produire de fiches d'aide pour elles. De plus, pour ne pas alourdir ces fonctions, elles comportent typiquement peu ou pas de validation d'arguments.

### Les jeux de données dans le répertoire data

Chaque jeu de données dans le répertoire data d'un package doit être décrit dans une fiche d'aide pour expliquer son contenu.

## Les classes et méthodes S4 ou RC publiques

Les packages exploitant le système de programmation orientée objet  $S4^2$  ou le système  $RC^3$  doivent fournir des fiches d'aide pour les classes et méthodes publiques.

#### Il est aussi recommandé de documenter :

• Les méthodes S3 pour des fonctions génériques :

Nous pouvons leur faire une fiche d'aide indépendante ou encore les documenter dans la même fiche d'aide que la fonction qui crée les objets de la classe en question.

• Le package lui-même :

Il est utile de créer une fiche d'aide présentant le package.

# Comment écrire de la documentation avec roxygen2?

Il suffit d'insérer des « commentaires » roxygen2 dans le code source des fonctions, donc dans les scripts situés dans le sous-répertoire R. Un commentaire roxygen2 débute par #', ce qui le distingue d'un commentaire ordinaire, qui débute par #.

### Exemples:

D'abord, voici des exemples de scripts R contenant des commentaires roxygen2. Des explications se retrouvent après les exemples. Ces exemples poursuivent l'exemple de création d'une fonction qui calcule la distance de Manhattan entre deux points provenant des notes Tests et exceptions en R.

#### Fichier C:/coursR/manhattan/R/distman.R:

```
#' Distance de Manhattan
#'
#' Calcule la distance de Manhattan entre deux points
#'
#' @param point1 Un vecteur numerique des coordonnees du premier point.
#' Oparam point2 Un vecteur numerique des coordonnees du deuxieme point.
#' Oreturn une seule valeur : la distance de Manhattan entre
#'
                               \code{point1} et \code{point2}
#' @author Sophie Baillargeon
#' @export
#' @examples
\#' distman(point1 = c(0,-5), point2 = c(0,-15))
distman <- function(point1, point2) {</pre>
  # Validation des arguments
  if (length(point1) != length(point2))
    stop("'point1' and 'point2' must have the same length")
  if (!is.null(dim(point1)) | !is.null(dim(point2)))
    warning("'point1' and 'point2' are treated as dimension 1 vectors")
  # Calculs
  out <- list(dist = sum(abs(point1 - point2)))</pre>
  # Sortie
  class(out) <- "distman"</pre>
  return(out)
}
```

 $<sup>^2</sup> https://cran.r-project.org/doc/manuals/r-release/R-exts.html \# Documenting-S4-classes-and-methods for the control of the$ 

<sup>&</sup>lt;sup>3</sup>https://r-pkgs.org/man.html#man-classes

# Fichier C:/coursR/manhattan/R/distman2.R:

```
#' Distance de Manhattan
#' Calcule la distance de Manhattan entre deux points
#' Utilise la fonction \code{\link[stats]{dist}} du package \pkg{stats}.
#' Oparam point1 Un vecteur numerique des coordonnees du premier point.
#' @param point2 Un vecteur numerique des coordonnees du deuxieme point.
#' @return \item{dist}{ la distance de Manhattan entre \code{point1} et \code{point2} }
#' @return \item{call}{ une copie de l'appel de la fonction }
#' @author Sophie Baillargeon
#' @export
#' @importFrom stats dist
#' @examples
\#' distman2(point1 = c(0,-5), point2 = c(0,-15))
distman2 <- function(point1, point2) {</pre>
 call <- match.call()</pre>
 dist <- stats::dist(rbind(point1, point2), method = "manhattan")</pre>
 out <- list(dist = as.vector(dist), call = call)</pre>
 class(out) <- "distman"</pre>
 return(out)
}
```

### Fichier C:/coursR/manhattan/R/print.distman.R:

```
#' @describeIn distman Affiche un objet de classe \code{"distman"}
#' @param x Un objet produit par la fonction \code{distman}, a afficher.
#' @param \dots D'autres arguments passes a d'autres methodes.
#' @export
print.distman <- function(x, ...) {
   cat("Manhattan distance between 'point1' and 'point2' :", x$dist, "\n")
   invisible(x)
}</pre>
```

# Fichier C:/coursR/manhattan/R/manhattan.R:

```
#' @details
#' \tabular{ll}{
#' Package: \tab manhattan\cr
#' Type: \tab Package\cr
#' Version: \tab 1.0.0\cr
#' Date: \tab 2019-04-02\cr
#' License: \tab GPL-3\cr
#' }
"_PACKAGE"

#' Points aleatoires
#'
#' Coordonnes en deux dimensions de 10 points aleatoires.
#'
```

```
#' @format Une matrice contenant 10 points designes par les

#' coordonnees suivantes.

#' \describe{

#' \item{\code{X}}{\coordonnee en X}

#' \item{\code{Y}}{\coordonnee en Y}

#' Becamples

#' distman(point1 = points[4, ], point2 = points[8, ])

#'

#' * Note : Ces donnees ont ete crees par les instructions suivantes

#' points <- matrix(sample(1:10, size = 20, replace = TRUE),

#' nrow = 10, ncol = 2)

#' colnames(points) <- c("X", "Y")

"points"</pre>
```

## **Explications:**

Pour les fonctions et les méthodes associées à des fonctions génériques, il suffit de mettre en entête au code source les commentaires roxygen2 qui généreront la documentation. Pour le package et les jeux de données, il faut ajouter un fichier d'extension .R dans le répertoire R. Ce fichier doit contenir les commentaires roxygen2 pour documenter globalement le package, suivi de l'instruction "\_PACKAGE", et les commentaires pour documenter les jeux de données, chaque bloc suivi du nom du jeu de données sous forme de chaîne de caractères (par exemple "points") .

La première phrase de ces commentaires doit être le **titre** de la fiche d'aide. Ce titre peut optionnellement être précédé du tag **@title**.

Cette phrase doit être suivie d'une ligne de commentaire roxygen2 vide, puis du texte à mettre dans la section **Description**. Ce paragraphe peut optionnellement être précédé du tag @description.

Si nous souhaitons avoir une section **Details**, il faut mettre à la suite de ce texte une ligne de commentaire roxygen2 vide. Tout le texte après cette ligne vide, mais avant les lignes débutant par un tag roxygen2 formera la section **Details**. Ces paragraphes peuvent optionnellement être précédés du tag **@details**.

Ainsi, les deux documentations suivantes sont équivalentes :

```
#' Distance de Manhattan
#'
#' Calcule la distance de Manhattan entre deux points
#'
#' Utilise la fonction \code{\link[stats]{dist}} du package \pkg{stats}.
#'
#' @param point1 Un vecteur numerique des coordonnees du premier point.
#' Oparam point2 Un vecteur numerique des coordonnees du deuxieme point.
#' @return \item{dist}{ la distance de Manhattan entre \code{point1} et \code{point2} }
#' Oreturn \item{call}{ une copie de l'appel de la fonction }
#' @author Sophie Baillargeon
#' @export
#' @importFrom stats dist
#' @examples
\#' distman2(point1 = c(0,-5), point2 = c(0,-15))
distman2 <- function(point1, point2) {</pre>
  ... (code omis ici)
```

et

```
#' @title Distance de Manhattan
#' @description Calcule la distance de Manhattan entre deux points
#' @details Utilise la fonction \code{\link[stats]{dist}} du package \pkg{stats}.
#' @param point1 Un vecteur numerique des coordonnees du premier point.
#' @param point2 Un vecteur numerique des coordonnees du deuxieme point.
#' @return \item{dist}{ la distance de Manhattan entre \code{point1} et \code{point2} }
#' @return \item{call}{ une copie de l'appel de la fonction }
#' @author Sophie Baillargeon
#' @author Sophie Baillargeon
#' @importFrom stats dist
#' @examples
#' distman2(point1 = c(0,-5), point2 = c(0,-15))
distman2 <- function(point1, point2) {
    ... (code omis ici)
}</pre>
```

Les sections qui suivent le titre et la description, qui sont obligatoires, et les informations détaillées (cette section est optionnelle), débutent toutes par des tags. Il faut obligatoirement décrire, si la fonction en possède :

- les arguments en entrée avec le tag <code>Oparam</code> : il faut une description par argument, de la forme <code>Oparam</code> nomArgument description, où la description peut s'étendre sur plusieurs lignes;
- la sortie avec le tag @return : si la fonction retourne une liste, il est recommandé d'avoir une description par élément de la liste, de la forme @return \item{nomElementDeLaListe}{description}, où la description peut encore une fois s'étendre sur plusieurs lignes.

Il est recommandé de toujours mettre des exemples dans une fiche d'aide. Dans la documentation roxygen2, ceux-ci doivent être ajoutés à la fin du bloc de documentation, après le tag @examples.

Nous pouvons aussi ajouter les informations suivantes :

- les noms des auteurs avec le tag @author,
- des références avec le tag @references,
- des liens vers les fiches d'aide de fonction en lien avec la fonction documentée avec le tag @seealso,
- etc. (Il ne semble pas encore exister de liste formelle de tous les tags roxygens, mais ces tags sont présentés via les vignettes du package roxygen2 : https://cran.r-project.org/web/packages/roxygen2/vignettes/rd.html.)

#### Documenter plusieurs fonctions dans la même fiche

Les tags @rdname et @describeIn servent à présenter plusieurs fonctions dans la même fiche d'aide. Dans l'exemple, la fonction distman et la méthode print.distman sont documentées dans la même fiche d'aide grâce au tag @describeIn. Ce tag doit être suivi du nom de la fiche dans laquelle de l'information doit être ajouté, puis d'une courte description à propos de la fonction ou méthode documentée.

### Documenter des jeux de données

Pour documenter des jeux de données, deux tags supplémentaires sont disponibles :

- Cformat pour décrire la structure R qui contient les données,
- Osource pour fournir une référence concernant la provenance des données.

# Nom des fiches d'aide

Nous pouvons contrôler le nom des fiches d'aide avec le tag @name. Sans ce tag, la fiche d'aide porte le nom de la fonction ou le nom du jeu de données qui suit le bloc de documentation roxygen2. Pour une

documentation de package, l'instruction "\_PACKAGE" qui suit le bloc de documentation indique à roxygen2 d'utiliser le nom du package suivi de -package comme nom de fiche d'aide. Dans l'exemple, le nom de la fiche d'aide du package est donc manhattan-package. Cependant, roxygen2 crée aussi un alias pour le nom de la fiche d'aide qui est le nom du package.

Pour ouvrir une fiche d'aide avec la fonction help ou l'opérateur ?, il faut lui donner en entrée le nom de la fiche ou un alias du nom de la fiche. Il est donc important de nommer intelligemment nos fiches d'aide. La norme est de donner à une fiche d'aide le nom de l'objet principal qu'elle documente et d'ajouter des alias pour les noms des autres objets documentés dans la fiche (c'est ce que fait roxygen2 de façon automatique lorsque les tags @rdname et @describeIn sont utilisés).

#### Texte formaté dans des fiche d'aide

Dans les commentaires roxygen2, il est possible de formater du texte en utilisant les tags de mises en forme acceptés dans les fichiers .Rd. Ces tags, par exemple \code{}, \pkg{}, \item{}, \link{}, \dots, etc., sont documentés sur la page web suivante : https://cran.r-project.org/web/packages/roxygen2/vignettes/formatting.html.

# Tags pour définir le NAMESPACE

Il est aussi essentiel de mettre les tags pour l'écriture du  ${\tt NAMESPACE}$  :

- le tag @export exporte une fonction du NAMESPACE (donc la rend publique),
- le tag @importFrom assure l'importation des fonctions provenant d'autres packages qui sont utilisées dans le code de notre package.

# Comment générer les fichiers .Rd et le fichier NAMESPACE?

Pour générer les fichiers .Rd et le fichier NAMESPACE, il suffit de lancer la commande roxygenize du package roxygen2. La commande peut être lancée directement dans la console. Par exemple

```
roxygenize("C:/coursR/manhattan")
```

Elle peut aussi être lancée par l'intermédiaire d'un menu de RStudio (voir prochaine section).

Dans l'exemple présenté précédemment, les commentaires roxygen2 des fichiers "distman.R", "distman2.R", "print.distman.R" et "manhattan.R", du sous-répertoire "C:/coursR/manhattan/R", ont produit les fichiers suivants :

- NAMESPACE dans le répertoire C:/coursR/manhattan/;
- distman.Rd, distman2.Rd, manhattan.Rd et points.Rd dans le répertoire C:/coursR/manhattan/man/.

Le fichier C:/coursR/manhattan/NAMESPACE est le suivant :

```
# Generated by roxygen2: do not edit by hand

S3method(print,distman)
export(distman)
export(distman2)
importFrom(stats,dist)
```

Voici de quoi a l'air un des fichiers d'extension .Rd généré.

Fichier "C:/coursR/manhattan/man/distman.Rd":

```
% Generated by roxygen2: do not edit by hand
% Please edit documentation in R/distman.R, R/print.distman.R
\name{distman}
```

```
\alias{distman}
\alias{print.distman}
\title{Distance de Manhattan}
\usage{
distman(point1, point2)
\method{print}{distman}(x, ...)
\arguments{
\item{point1}{Un vecteur numerique des coordonnees du premier point.}
\item{point2}{Un vecteur numerique des coordonnees du deuxieme point.}
\item{x}{Un objet produit par la fonction \code{distman}, a afficher.}
\item{\dots}{D'autres arguments passes a d'autres methodes.}
}
\value{
une seule valeur : la distance de Manhattan entre
                           \code{point1} et \code{point2}
}
\description{
Calcule la distance de Manhattan entre deux points
\section{Methods (by generic)}{
\itemize{
\item \code{print}: Affiche un objet de classe \code{"distman"}
}}
\examples{
distman(point1 = c(0,-5), point2 = c(0,-15))
\author{
Sophie Baillargeon
}
```

Ces fichiers utilisent une syntaxe inspirée de la syntaxe de LaTeX. Lorsque nous utilisons roxygen2, nous n'avons pas besoin d'éditer directement ces fichiers, donc nous n'avons pas besoin de comprendre leur syntaxe. La syntaxe de roxygen2 est plus simple.

# Étape 4. Construire et vérifier le package

Une fois les étapes 1 à 3 complétées, nous sommes prêts à construire le fichier compressé du package à partir du répertoire contenant les fichiers source. En fait, nous allons aussi à cette étape vérifier le répertoire afin de s'assurer qu'il produira un package conforme.

Ces tâches se réalisent avec les utilitaires :

- R CMD check,
- R CMD build,
- R CMD INSTALL.

Ces utilitaires s'exploitent par des commandes soumises dans le terminal sous Unix / Linux ou Mac OS X

/ OS X / macOS et dans une fenêtre invite de commandes sous Windows. Dans ces notes, nous verrons comment utiliser un outil qui soumet ces commandes à notre place, sans que nous ayons besoin d'ouvrir le terminal ou la fenêtre d'invite de commandes.

Les utilitaires R CMD check, build et INSTALL sont inclus dans l'installation de base de R (dans le package utils). Cependant, ils nécessitent des outils supplémentaires pour fonctionner : les outils de développement de logiciel GNU, incluant un compilateur C/C++. Voici comment s'assurer que ces outils sont installés sur notre ordinateur, selon le système d'exploitation (plus de détails sur la page https://support.rstudio.com/hc/en-us/articles/200486498-Package-Development-Prerequisites) :

- Windows : il faut installer les « Rtools », téléchargeable sur la page web https://cran.r-project.org/bin/windows/Rtools/ (plus de détails dans le Guide d'installation ou de mise à jour de R et RStudio);
- Mac OS X / OS X / macOS : il faut installer les « Apple Xcode developer tools », disponibles gratuitement sur le « App Store », s'ils ne sont pas déjà installés (souvent installés par défaut) ;
- $\bullet \quad \text{Unix / Linux : il faut s'assurer d'avoir install\'e R accompagn\'e des ses « development tools » (\textbf{r-base-dev}).}$

Le package pkgbuild comporte une fonction pour tester si tout le nécessaire au développement de package est installé et fonctionne correctement.

```
library(pkgbuild)
has_build_tools()
```

Si cet appel à la fonction has\_build\_tools retourne TRUE, alors tout est fonctionnel.

De plus, une des commandes R pour le développement de packages, soit R CMD check, a besoin d'une installation de LaTeX pour tester la création de la documentation des packages en format PDF. Pour développer des packages, vous avez donc le choix entre :

- installer LaTeX sur votre ordinateur :
  - une version gratuite pour Windows est MiKTeX (https://miktex.org/download),
  - une version gratuite pour Mac OS X / OS X / macOS est MacTeX (http://www.tug.org/mactex/),
  - les systèmes Unix / Linux viennent habituellement par défaut avec une distribution de LaTeX;
- omettre la création de la documentation PDF lors de la soumission de la commande R CMD check grâce à l'option --no-manual (plus d'informations à venir en temps opportun).

# Lancement des commandes de construction et de vérification en RStudio

RStudio peut lancer pour nous les commandes de construction et de vérification de packages. Le logiciel rend le processus vraiment plus simple que de lancer les commandes manuellement dans le terminal ou l'invite de commandes. Nous apprendrons donc seulement comment construire et vérifier un package avec RStudio.

#### Sous-étape a) Créer un projet RStudio avec le répertoire principal du package

Il faut tout d'abord créer un projet RStudio avec le répertoire principal de notre package. Pour ce faire, nous pouvons procéder comme suit :

- ouvrir le menu « File » et sélectionner « New Project... » (il y a aussi un bouton dans la barre de RStudio en haut à droite, nommé à l'origine « Project : (None) » qui ouvre un menu contenant aussi l'élément « New Project... »),
- sélectionner « Existing Directory »,
- sélectionner le répertoire principal du package, c'est-à-dire le répertoire portant le nom du package et contenant les fichiers sources,
- cliquer sur « Create Project ».

Le projet sera créé et ouvert. Ça ajoute des fichiers dans le répertoire de notre package. Nous ne nous préoccuperons pas de ces fichiers. Ils sont automatiquement ignorés lors de la construction du package avec RStudio.

**Note** : Si nous sélectionnons « New Directory » plutôt que « Existing Directory », puis « R package », RStudio crée un squelette de répertoire de fichiers source de package.

# Sous-étape b) Configurer les options de RStudio

Voici quelques configurations de RStudio que je conseille d'utiliser.

#### Options globales:

(à modifier une seule fois)

- par le menu « Tools > Global Options... »,
- dans « Packages », décocher « Cleanup output after successful R CMD check ».

Les répertoires générés par la commande R CMD check ne seront ainsi pas effacés et nous pourrons, par exemple, aller y récupérer la version PDF de la documentation du package.

# Options du projet :

(à modifier pour chaque nouveau projet)

- par le menu « Tools > Project Options... »;
- dans « Build Tools » :
  - décocher « Use devtools package functions if available » (seulement nécessaire si nous souhaitons que la commande R CMD check génère la version PDF de la documentation du package),
  - cocher « Generate documentation with Roxygen »,
  - si le menu de configuration ne s'ouvre pas automatiquement, cliquez sur « Configure... », puis assurez-vous que les options suivantes soient cochées :
    - \* « Use roxygen to generate » : « Rd files » et « NAMESPACE »,
    - \* « Automatically roxygenize when running » : tout cocher.

Avec ces dernières configurations, la majorité des commandes de construction et de vérification de package lancées par le menu « Build » (voir ci-dessous) vont d'abord soumettre la commande roxygenize sur le répertoire du package avant de faire leur travail. Ainsi, les fichiers .Rd de documentation et le fichier NAMESPACE seront mis à jour à chaque lancement d'une de ces commandes

Nous pouvons aussi soumettre la commande roxygenize par le menu « Build > Document ».

Note: Si vous n'avez pas de compilateur LaTeX sur votre ordinateur, ajoutez l'option suivante à R CMD check: --no-manual. Cette fonction indique à R CMD check de ne pas vérifier la compilation de la version PDF de la documentation du package.

#### Sous-étape c) Construire à partir du menu « Build » de RStudio

Il est préférable de toujours d'abord s'assurer que le package passe sans erreur ou avertissements problématiques la vérification faite par la commande R CMD check. Ensuite, nous pouvons construire le package, soit dans sa version source, soit dans sa version binaire. Voici comment faire tout ça facilement en RStudio.

- Pour vérifier le package :
   menu « Build > Check Package » (lance en fait la commande R CMD check).
- Pour créer le package source (qui est aussi la version Unix / Linux) :
  menu « Build > Build Source Package » (lance en fait la commande R CMD build)

  → un fichier nommé "nomPackage\_numeroVersion.tar.gz" (dans notre exemple
  manhattan\_1.0.0.tar.gz) sera créé dans le répertoire au-dessus du répertoire principal du package.
- Pour créer le package binaire (si nous travaillons sous Windows ou Mac OS X / OS X / macOS):
   menu « Build > Build Binary Package » (lance en fait la commande R CMD INSTALL --build)
   → un fichier nommé nomPackage\_numeroVersion.zip (sous Windows) ou nomPackage\_numeroVersion.tgz

(sous Mac OS X / OS X / mac OS) sera créé dans le répertoire au-dessus du répertoire principal du package.

La commande « Install and Restart » est pratique en cours de travail. Elle permet de construire le package dans le bon format pour notre système d'exploitation, l'installer (donc remplacer l'ancienne installation par la nouvelle) et charger de nouveau le package (avec la commande library).

Vous trouverez en annexe des solutions à certains problèmes techniques déjà rencontrés lors du lancement d'une commande R CMD.

#### Lancement des commandes de construction et de vérification avec devtools

Nous ne couvrirons pas cette option ici, mais les commandes de construction et de vérification d'un package peuvent aussi être soumises à l'aide de fonctions du package devtools (https://github.com/hadley/devtools). Comme les utilitaires de RStudio, ces fonctions permettent de soumettre des commandes R CMD sans passer nous-mêmes par le terminal ou l'invite de commandes. Avec devtools, tout se réalise via des commandes soumises dans la console R.

# Étape 5. Si désiré, partager le package

Notre package est maintenant prêt à être utilisé. Si nous souhaitons le partager avec le grand public, nous pouvons le rendre disponible sur le CRAN. Pour ce faire, il faut d'abord s'assurer de respecter les politiques du CRAN: http://cran.r-project.org/web/packages/policies.html. Comme mentionné précédemment, il faut donc que notre package passe le R CMD check --as-cran sans erreurs ni avertissements.

Une fois s'être assuré de respecter les politiques du CRAN, nous pouvons soumettre notre package au CRAN en ligne par l'intermédiaire de l'interface web suivante : https://cran.r-project.org/submit.html

Il suffit de suivre les instructions.

# Étape 6. Au besoin, mettre à jour le package

Mettre à jour un package signifie de le modifier pour ajouter des fonctionnalités et/ou corriger des bogues. Lors d'une mise à jour, il faut suivre les étapes suivantes.

# Sous-étape a) Incrémenter le numéro de version :

Cette incrémentation doit être effectuée dans le fichier DESCRIPTION, ainsi que dans tout commentaire roxygen2 mentionnant la date de construction du package, par exemple dans la fiche d'aide du package.

Rappelons qu'un numéro de version est une séquence d'au minimum 2 (souvent 3) nombres entiers non négatifs séparés par un seul caractère caractère . (point) ou – (tiret). Ces nombres ne sont pas contraints à être compris entre 0 et 9.

Voici les règles que plusieurs développeurs de packages R suivent dans la numérotation des versions de leurs packages. Ils forment les numéros de version de 3 nombres entiers séparés par un point. En cours de développement, soit avant d'avoir une version qu'ils considèrent suffisamment testée, ils utilisent un 0 comme premier nombre dans le numéro de version (par exemple 0.9.12). Lorsqu'ils jugent leur package assez fiable pour être rendu disponible à plus grande échelle qu'à l'interne, ils changent le premier nombre dans le numéro de version pour 1, ce qui fait retomber à zéro les nombres suivants. La première version officielle porte donc le numéro de version 1.0.0.

Ensuite, ils font évoluer les numéros de version comme suit.

• Lors d'une mise à jour majeure (beaucoup de nouvelles fonctionnalités) : le premier nombre de la numérotation est incrémenté de 1, les nombres subséquents retombent à 0.

- Lors d'une mise à jour mineure (seulement quelques fonctionnalités pas trop importantes ont été ajoutées ou d'importants bogues ont été réglés) : le premier nombre est inchangé, par contre le deuxième est incrémenté de 1 et le dernier retombe à 0.
- Si seulement quelques bogues ont été corrigés, sans changer du tout les fonctionnalités : le troisième nombre de la numérotation est incrémenté de 1, sans modifier les deux premiers nombres.

La page Wikipédia https://en.wikipedia.org/wiki/Software\_versioning traite de ce sujet et propose d'autres règles.

# Sous-étape b) Faire les mises à jour dans le code :

L'ajout de fonctionnalités ou la correction de bogues impliquent des modifications à apporter au code.

## Sous-étape c) Documenter les modifications :

Les commentaires roxygen2 produisant les fiches d'aide doivent être mis à jour de façon à refléter les modifications apportées au code. Une correction d'un bogue ne nécessite pas toujours de mise à jour des fiches d'aide, mais un ajout de fonctionnalités en nécessite toujours.

Je conseille vivement aussi de documenter les mises à jour dans un fichier NEWS. Ce fichier est un point de repère pour un utilisateur d'un package qui souhaite identifier ce qui a changé lors d'une mise à jour, donc ce qui pourrait affecter son utilisation du package.

Il n'y a pas de consensus en R à propos de comment rédiger le fichier NEWS, ni de l'endroit où le placer dans les fichiers source. Nous pouvons observer, notamment, les pratiques suivantes :

- fichier NEWS en format texte simple (dont le nom de fichier ne porte pas d'extension), placé dans le répertoire principal (au même niveau que les fichiers DESCRIPTION et NAMESPACE);
- fichier NEWS.md en format Markdown, placé dans le répertoire principal;
- fichier NEWS en format texte simple (dont le nom de fichier ne porte pas d'extension), placé dans le sous-répertoire inst;
- fichier NEWS.Rd en format .Rd, placé dans le sous-répertoire inst.

La première des pratiques est probablement la plus répandue.

Voici à quoi pourrait ressembler un fichier NEWS en format texte simple pour l'exemple du package manhattan, après l'avoir mis à jour.

Fichier "C:/coursR/manhattan/NEWS":

```
Changements dans manhattan version 1.1.0 (2019-04-02)

* modifications des fonctions distman et distman2 afin qu'elles puissent calculer la distance entre plusieurs points

* ajout d'une methode plot pour un objet de classe "distman"

Changements dans manhattan version 1.0.0 (2019-03-21)

* première version du package manhattan
```

Si nous installions la version 1.1.0 du package et que nous chargions le package en R avec la commande library, nous pourrions afficher son fichier NEWS dans la console avec la instructions suivantes :

```
nouvelles <- news(package = "manhattan")
print(nouvelles, doBrowse = FALSE)</pre>
```

Nous pouvons aussi voir le fichier NEWS des packages distribués sur le CRAN directement sur leur page web du CRAN.

## Sous-étape d) Reconstruire et revérifier le package

Il faut finalement construire et vérifier de nouveau le package. Un nouveau fichier compressé sera produit, portant le nom du package accompagné du nouveau numéro de version.

# Résumé

Étapes de création d'un package R

#### 1. Écrire les fonctions

Objets utilisables dans le corps des fonctions d'un package :

- arguments et variables locales,
- tout objet, public ou privé, contenu dans le package,
- objets provenant d'autres packages, à la condition d'inclure ces objets dans l'environnement des objets importés par le package,
- objets du package R de base.

Si but = mettre sur le CRAN : respecter les politiques http://cran.r-project.org/web/packages/policies.html

# 2. Créer la structure de fichiers du package

Répertoire principal, portant le nom du package, contenant :

- sous-répertoire nommé R : répertoire des fichiers contenant le code source R des fonctions,
- sous-répertoire nommé man : répertoire des fichiers sources des fiches d'aide,
- fichier nommé DESCRIPTION : informations descriptives générales du package,
- fichier nommé NAMESPACE : fichier permettant de définir quels objets sont publics (exportés) et d'identifier les fonctions d'autres packages (excluant le package base) utilisées dans le package (objets importés).
- autres éléments, selon les besoins : fichier NEWS, sous-répertoire data, sous-répertoire src, etc.

# 3. Écrire la documentation des fonctions et du package

Éléments à documenter :

- les fonctions publiques,
- les jeux de données,
- les classes et méthodes S4 ou RC publiques,
- les méthodes S3 pour des fonctions génériques (si pertinent),
- le package lui-même (recommandé).

# Documentation d'un package avec roxygen2:

Commentaires dans le code source R pour générer les fiches d'aide (contenu du sous-répertoire man) et le fichier NAMESPACE

- symbole de commentaire roxygen2 : #';
- commentaires en entête des fonctions ou
  - avant "\_PACKAGE" pour la fiche d'aide globale du package,
  - avant une chaîne de caractères contenant le nom d'un objet contenant des données;
- 1<sup>ère</sup> ligne => titre de la fiche, suivi d'une ligne vide;
- $1^{er}$  paragraphe => champ Descripion de la fiche, suivi d'une ligne vide;
- autres paragraphes (optionnels) => champ Details;

- tags:
  - Cparam : description des paramètres => champ Arguments;
  - @return : description de la sortie => champ Value;
  - pour le fichier NAMESPACE : @export, @importFrom, etc.;
  - autres sections : @examples, @author, @references, @format, etc.;
  - pour jumeler des fiches : @rdname, @describeIn;
  - métadonnées : @name. @aliases. etc.

## 4. Construire et vérifier le package

Outils nécessaires :

- R.
- outils de développement de logiciel GNU,
- compilateur LaTeX (uniquement pour générer la documentation au format PDF).

#### R CMD check

Effectue diverses vérifications du package :

- option --as-cran avant de soumettre an CRAN,
- option --no-manual si aucun compilateur LaTeX n'est installé sur notre ordinateur.

#### R CMD build

Construit le package source (.tar.gz, = version Unix / Linux).

#### R CMD INSTALL --build

Construit le package binaire (.zip sous Windows, .tgz sous Mac OS X / OS X / macOS )

### Construire et vérifier le package avec RStudio :

- Sous-étape a) Créer un projet RStudio avec notre package
- Sous-étape b) Configurer les options de RStudio
- Sous-étape c) Compiler à partir du menu « Build » de RStudio
  - vérifier le package : menu « Build > Check Package »,
  - créer le package source : menu « Build > Build Source Package »,
  - créer le package binaire : menu « Build > Build Binary Package ».

La compilation « Install and Restart » est pratique en cours de travail. Elle permet de

- compiler le package dans le bon format pour notre système d'exploitation,
- l'installer (donc remplacer l'ancienne installation par la nouvelle) et
- charger de nouveau le package (avec la commande library).

#### 5. Si désiré, partager le package

Il n'est pas compliqué de soumettre un package au CRAN: https://cran.r-project.org/submit.html

# 6. Au besoin, mettre à jour le package

Mise à jour = modification pour ajouter des fonctionnalités et/ou corriger des bogues :

- Sous-étape a) Incrémenter le numéro de version
  - dans le fichier DESCRIPTION
  - dans la fiche d'aide du package
- Sous-étape b) Faire les mises à jour dans le code
- Sous-étape c) Documenter les modifications :
  - $-\,$ mettre à jour les commentaires  ${\tt roxygen2}$  produisant les fiches d'aide
  - décrire brièvement les changements dans fichier NEWS (non obligatoire, mais c'est une bonne pratique)

• Sous-étape d) Reconstruire et revérifier le package

Références

- R Core Team (2019). Writing R Extensions. R Foundation for Statistical Computing. Chapitre 4. URL https://cran.r-project.org/doc/manuals/r-release/R-exts.html
- Wickham, H. (2015). R packages. O'Reilly Media, Inc.
  - URL première édition (quelques informations dans cette version sont déjà obsolètes) : http://r-pkgs.had.co.nz/
  - URL deuxième édition (en développement) : https://r-pkgs.org
- Hadley Wickham, Peter Danenberg and Manuel Eugster (2018). roxygen2: In-Line Documentation for R. R package version 6.1.1.
  - URL https://CRAN.R-project.org/package=roxygen2
    - \* les vignettes sont informatives : https://cran.r-project.org/web/packages/roxygen2/vignettes/roxygen2.html
  - documentation avec roxygen2 : http://r-pkgs.had.co.nz/man.html

Pour le développement avec RStudio :

• https://support.rstudio.com/hc/en-us/articles/200486488-Developing-Packages-with-RStudio

Pour le développement en utilisant le package devtools :

- https://github.com/r-lib/devtools
- https://rawgit.com/rstudio/cheatsheets/master/package-development.pdf

Annexe

# Résolution de problèmes déjà rencontrés

Voici des problèmes déjà rencontrés par des étudiants du cours lors de la construction ou la vérification d'un package et leurs solutions.

# Erreur « Installation failed »

PROBLÈME : Une erreur de ce type est produite par l'outil « Check Package » :

\* checking whether package 'manhattan' can be installed ... ERROR Installation failed.

See 'C:/coursR/NomRépertoireAvecAccents/manhattan.Rcheck/00install.out' for details.

CAUSE : Le nom complet du répertoire du package contient des accents ou des espaces. (Cela ne cause pas toujours une erreur, ça dépend de la version de R utilisée et de votre système d'exploitation.)

SOLUTION: Utiliser un répertoire dont le nom complet ne comprend aucun accent, ni aucun espace.

# Erreur de permission d'écriture dans un répertoire d'installation

PROBLÈME : Une erreur de ce type est produite par l'outil « Build Binary Package », « Install and Restart » ou « Clean and Rebuild » :

ERREUR : pas de permission pour installer dans le répertoire  $\mbox{'C:/Program Files/R/R-3.5.2/library'}$ 

CAUSE: Le compte utilisé n'a pas la permission d'écrire dans le répertoire où R cherche à installer le package. Si vous travaillez à partir d'un compte administrateur, vous ne devriez pas rencontrer ce problème.

SOLUTION : Modifier les options des outils de construction pour demander l'installation dans un autre répertoire, un pour lequel le compte utilisateur à une permission en écriture.

Étape 1 : trouver le bon répertoire à utiliser

Il faut utiliser un des répertoires dans lesquels R recherche des installations de packages. Un vecteur contenant tous ces répertoires est retourné par la commande suivante dans la console R:

#### .libPaths()

- ## [1] "C:/Users/Sophie/Documents/R/win-library/3.6"
- ## [2] "C:/Program Files/R/R-3.6.2/library"

Dans mon cas, le premier répertoire énuméré, soit "C:/Users/Sophie/Documents/R/win-library/3.5", est localisé dans les fichiers du compte utilisateur. J'y ai donc assurément une permission en écriture. C'est lui que je vais utiliser.

Étape 2 : ajouter une option aux outils de construction

- a) ouvrir la fenêtre de configuration des outils de construction
- b) ajouter l'option --library="C:/Users/Sophie/Documents/R/win-library/3.5" (à adapter selon le nom du répertoire que vous avez choisi d'utiliser) à deux endroits :
- la configuration de l'outil « **Install and Restart** » et « **Clean and Rebuild** » se fait dans le champ, nommé « Build and Reload R CMD INSTALL additionnal options : »
- la configuration de l'outil « **Build Binary Package** » se fait dans le dernier champ, nommé « Build Binary Package R CMD INSTALL additionnal options : »

## Vérification qui gèle à l'étape « checking PDF version of manual »

PROBLÈME : Outil « **Check Package** » qui n'arrive pas à terminer son check, qui gèle à l'étape « checking PDF version of manual ».

 ${\it CAUSE}$  : Je ne sais pas exactement... Ça ressemble à un bogue.

SOLUTION : Modifier les options de l'outil « Check Package »

- a) ouvrir la fenêtre de configuration des outils de construction
- b) ajouter dans le champ nommé « Check Package R CMD check additionnal options : » l'option suivante : --no-manual