

# DIY R PACKAGE: SIMPLIFY LIFE WITH R

---

Christina Nieuwoudt

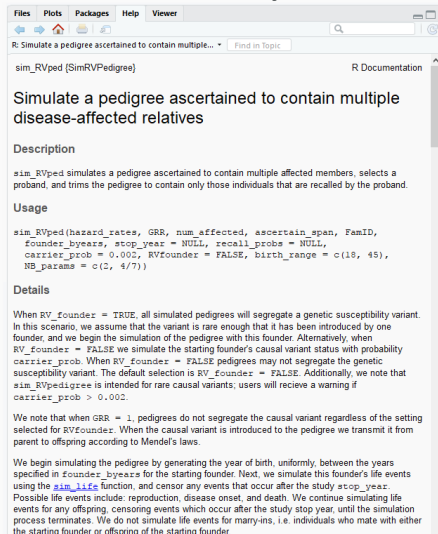
Supervisor: Jinko Graham

May 27, 2016

Simon Fraser University

Department of Statistics

No more digging through old notes to figure out what you did months ago!  
Store information where you use it: in R.



The screenshot shows the R Documentation page for the `sim_RVped` function. The page title is "Simulate a pedigree ascertained to contain multiple disease-affected relatives". The description states that `sim_RVped` simulates a pedigree ascertained to contain multiple affected members, selects a proband, and trims the pedigree to contain only those individuals that are recalled by the proband. The usage section shows the function signature: `sim_RVped(hazard_rates, GRR, num_affected, ascertain_span, FamID, founder_years, stop_year = NULL, recall_probs = NULL, carrier_prob = 0.002, RVfounder = FALSE, birth_range = c(18, 45), NB_params = c(2, 4/7))`. The details section explains that when `RV_founder = TRUE`, all simulated pedigrees will segregate a genetic susceptibility variant. In this scenario, we assume that the variant is rare enough that it has been introduced by one founder, and we begin the simulation of the pedigree with this founder. Alternatively, when `RV_founder = FALSE`, we simulate the starting founder's causal variant status with probability `carrier_prob`. When `RV_founder = FALSE`, pedigrees may not segregate the genetic susceptibility variant. The default selection is `RV_founder = FALSE`. Additionally, we note that `sim_RVped` is intended for rare causal variants; users will receive a warning if `carrier_prob > 0.002`. We note that when `GRR = 1`, pedigrees do not segregate the causal variant regardless of the setting selected for `RVfounder`. When the causal variant is introduced to the pedigree we transmit it from parent to offspring according to Mendel's laws. We begin simulating the pedigree by generating the year of birth, uniformly, between the years specified in `founder_years` for the starting founder. Next, we simulate this founder's life events using the `sim_life` function, and censor any events that occur after the study `stop_year`. Possible life events include: reproduction, disease onset, and death. We continue simulating life events for any offspring, censoring events which occur after the study stop year, until the simulation process terminates. We do not simulate life events for marry-ins, i.e. individuals who mate with either the starting founder or offspring of the starting founder.

No more tedious, time-consuming data import!

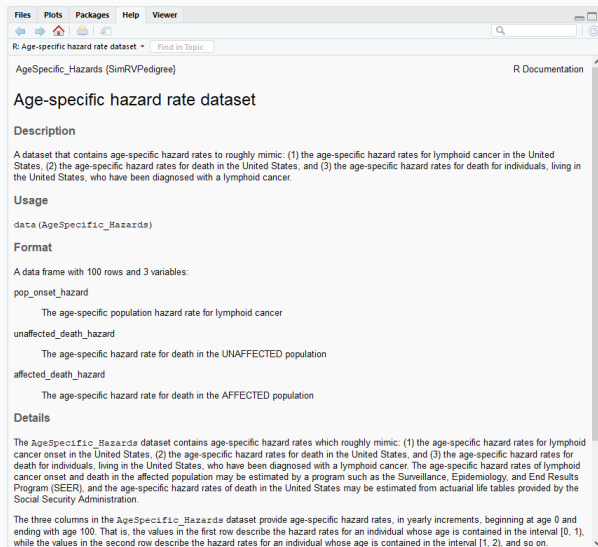
```
1 getwd()  
2 setwd("C:/Users/cnieuwoudt/Google Drive/LymphomaStudy  
  /PedSim/PedSimData")  
3 LC_SIMhazards <- read.csv("LC_SIMhazards.csv",  
4                           stringsAsFactors=FALSE)  
5 LC_SIMpartition <- read.csv("LC_SIMpartition.csv",  
6                             stringsAsFactors=FALSE)
```

Simply load the package and then the data.

```
library(SimRVPedigree)  
data("AgeSpecific_Hazards")
```



BONUS: You can clearly document your datasets, which may save you valuable time in the future.



The screenshot shows the R documentation interface for the 'Age-specific hazard rate dataset'. The window title is 'R: Age-specific hazard rate dataset'. The top menu bar includes 'Files', 'Plots', 'Packages', 'Help', and 'Viewer'. Below the menu is a search bar with the text 'Find in Topic'. The main content area is titled 'Age-specific hazard rate dataset' and includes a 'Description' section, a 'Usage' section with the command `data(AgeSpecific_Hazards)`, a 'Format' section, and a 'Details' section. The 'Description' section states: 'A dataset that contains age-specific hazard rates to roughly mimic: (1) the age-specific hazard rates for lymphoid cancer in the United States, (2) the age-specific hazard rates for death in the United States, and (3) the age-specific hazard rates for death for individuals, living in the United States, who have been diagnosed with a lymphoid cancer.' The 'Usage' section shows the command `data(AgeSpecific_Hazards)`. The 'Format' section states: 'A data frame with 100 rows and 3 variables:'. The 'Details' section states: 'The AgeSpecific\_Hazards dataset contains age-specific hazard rates which roughly mimic: (1) the age-specific hazard rates for lymphoid cancer onset in the United States, (2) the age-specific hazard rates for death in the United States, and (3) the age-specific hazard rates for death for individuals, living in the United States, who have been diagnosed with a lymphoid cancer. The age-specific hazard rates of lymphoid cancer onset and death in the affected population may be estimated by a program such as the Surveillance, Epidemiology, and End Results Program (SEER), and the age-specific hazard rates of death in the United States may be estimated from actuarial life tables provided by the Social Security Administration.' The 'Details' section also states: 'The three columns in the AgeSpecific\_Hazards dataset provide age-specific hazard rates, in yearly increments, beginning at age 0 and ending with age 100. That is, the values in the first row describe the hazard rates for an individual whose age is contained in the interval [0, 1), while the values in the second row describe the hazard rates for an individual whose age is contained in the interval [1, 2), and so on.'

Files Plots Packages Help Viewer

R: Age-specific hazard rate dataset Find in Topic

AgeSpecific\_Hazards (SimRVPedigree) R Documentation

## Age-specific hazard rate dataset

### Description

A dataset that contains age-specific hazard rates to roughly mimic: (1) the age-specific hazard rates for lymphoid cancer in the United States, (2) the age-specific hazard rates for death in the United States, and (3) the age-specific hazard rates for death for individuals, living in the United States, who have been diagnosed with a lymphoid cancer.

### Usage

```
data(AgeSpecific_Hazards)
```

### Format

A data frame with 100 rows and 3 variables:

pop\_onset\_hazard

The age-specific population hazard rate for lymphoid cancer

unaffected\_death\_hazard

The age-specific hazard rate for death in the UNAFFECTED population

affected\_death\_hazard

The age-specific hazard rate for death in the AFFECTED population

### Details

The AgeSpecific\_Hazards dataset contains age-specific hazard rates which roughly mimic: (1) the age-specific hazard rates for lymphoid cancer onset in the United States, (2) the age-specific hazard rates for death in the United States, and (3) the age-specific hazard rates for death for individuals, living in the United States, who have been diagnosed with a lymphoid cancer. The age-specific hazard rates of lymphoid cancer onset and death in the affected population may be estimated by a program such as the Surveillance, Epidemiology, and End Results Program (SEER), and the age-specific hazard rates of death in the United States may be estimated from actuarial life tables provided by the Social Security Administration.

The three columns in the AgeSpecific\_Hazards dataset provide age-specific hazard rates, in yearly increments, beginning at age 0 and ending with age 100. That is, the values in the first row describe the hazard rates for an individual whose age is contained in the interval [0, 1), while the values in the second row describe the hazard rates for an individual whose age is contained in the interval [1, 2), and so on.

## Track and Revert Changes

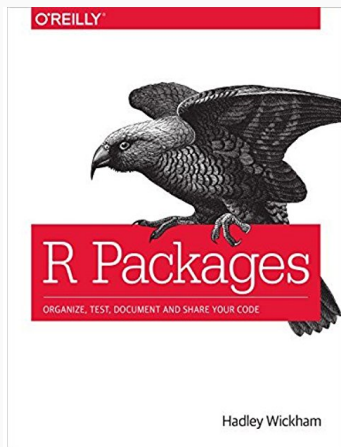
- ▶ Git
- ▶ Subversion

## Useful links:

- ▶ <https://support.rstudio.com/hc/en-us/articles/200532077-Version-Control-with-Git-and-SVN>
- ▶ <https://www.r-bloggers.com/version-control-with-git/>

OMG, THANK YOU HADLEY WICKHAM!

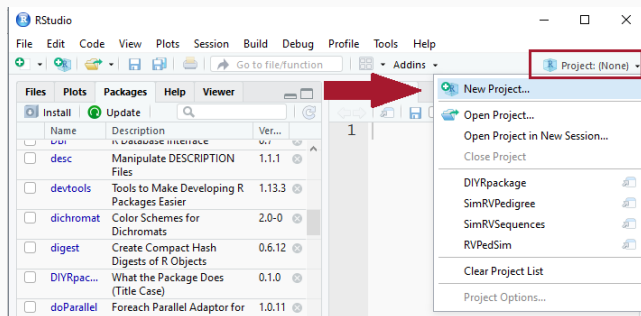
Wickham, Hadley: *R Packages: Organize, Test, Document, and Share Your Code*. 1st Edition, O'Reilly Media (2015).



Available online <http://r-pkgs.had.co.nz/>

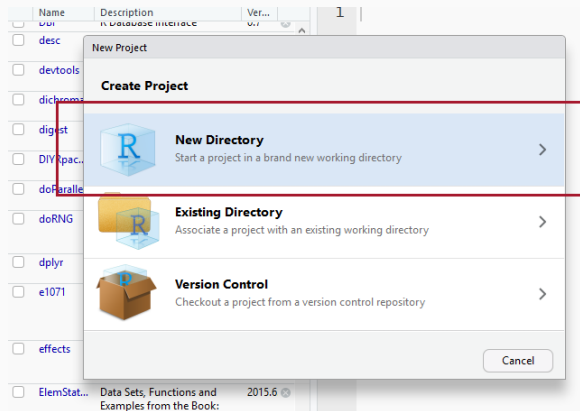
- ▶ Developer Tools (NOT R PACKAGES)
  - ▶ **Windows:**
    - ▶ Rtools: <https://cran.r-project.org/bin/windows/Rtools/>
  - ▶ **Mac:**
    - ▶ XCode: Available for free in App Store
    - ▶ Command Line Tools for Xcode:  
<http://developer.apple.com/downloads>
  - ▶ **Linux:** See HW's book.
- ▶ R packages:
  - ▶ devtools
  - ▶ roxygen2
  - ▶ knitr
  - ▶ testthat

# CREATE R PACKAGE



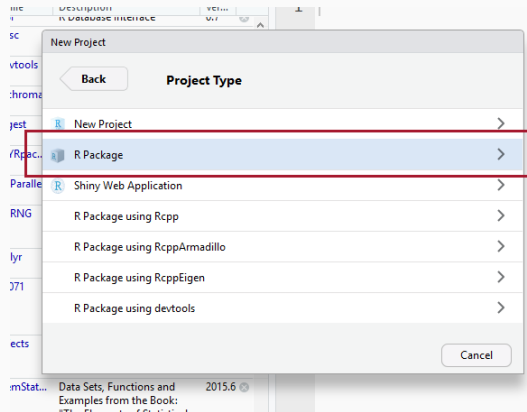
Click on the *Project* drop down menu and select *New Project*.





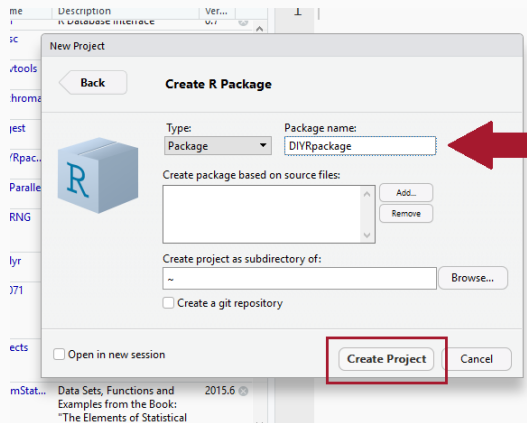
Select *New Directory* from the *New Project* options.

# CREATE R PACKAGE



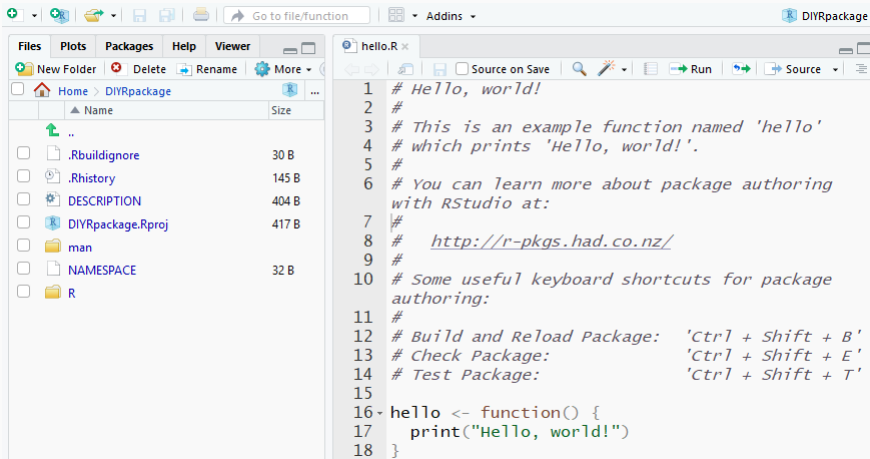
Choose *R Package* from the *Project Type* options.

# CREATE R PACKAGE

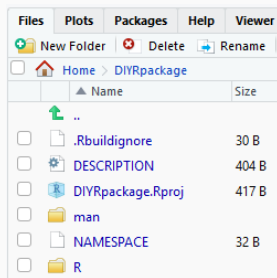


Name the package and click *Create Project*.

# CREATE R PACKAGE



R generates a new package with basic package files and a hello world function.



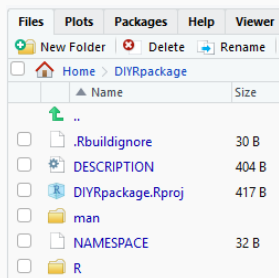
R package directory files:

- ▶ .Rproj file
- ▶ .Rbuildignore
- ▶ DESCRIPTION
- ▶ NAMESPACE
- ▶ \R
- ▶ \man

**.Rproj file:** a text file that stores project preferences.

**.Rbuildignore:** Important for CRAN packages, for more info check out Chapter 2 of HW's book.

Don't worry about them. You don't need to edit them by hand.



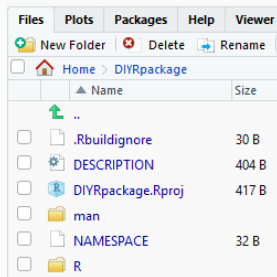
R package directory files:

- ▶ .Rproj file
- ▶ .Rbuildignore
- ▶ DESCRIPTION
- ▶ NAMESPACE
- ▶ \R
- ▶ \man

## NAMESPACE

Important for CRAN packages.

Don't worry about it; roxygen2 manages this for you.

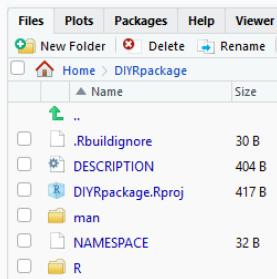


R package directory files:

- ▶ .Rproj file
- ▶ .Rbuildignore
- ▶ DESCRIPTION
- ▶ NAMESPACE
- ▶ \R
- ▶ \man

## DESCRIPTION

Contains package metadata, more on this soon.



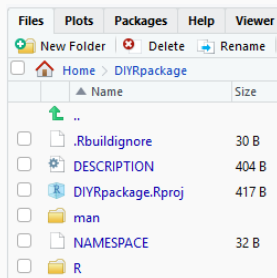
R package directory files:

- ▶ .Rproj file
- ▶ .Rbuildignore
- ▶ DESCRIPTION
- ▶ NAMESPACE
- ▶ \R
- ▶ \man

**\R**

All R code is stored here.





R package directory files:

- ▶ .Rproj file
- ▶ .Rbuildignore
- ▶ DESCRIPTION
- ▶ NAMESPACE
- ▶ \R
- ▶ \man

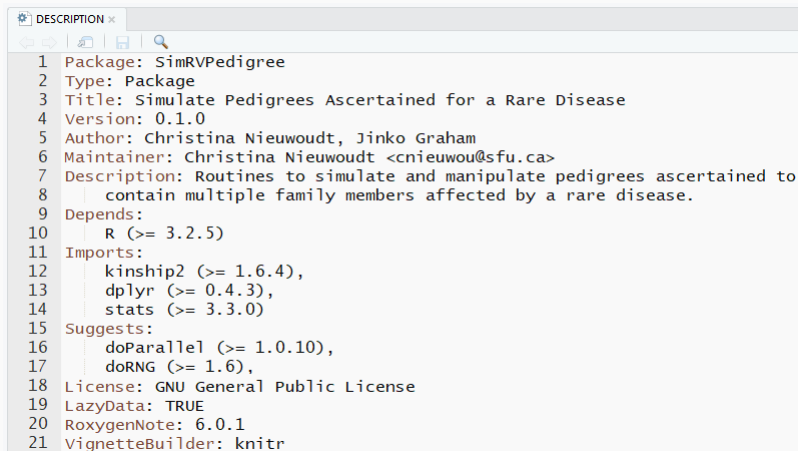
man

This is where description files are stored.

NEVER edit these by hand; roxygen2 does all of the work for you.

Additionally, you may find the following files to be useful additions to your R package directory:

- ▶ `\data`
- ▶ `\vignettes`
- ▶ `\tests`

A screenshot of a text editor window titled "DESCRIPTION x". The window contains a DESCRIPTION file for an R package. The file lists package metadata such as name, type, title, version, author, maintainer, description, dependencies, imports, suggests, license, lazy data, roxygen note, and vignette builder. The text is as follows:

```
1 Package: SimRVPedigree
2 Type: Package
3 Title: Simulate Pedigrees Ascertained for a Rare Disease
4 Version: 0.1.0
5 Author: Christina Nieuwoudt, Jinko Graham
6 Maintainer: Christina Nieuwoudt <cnieuwou@sfu.ca>
7 Description: Routines to simulate and manipulate pedigrees ascertained to
8   contain multiple family members affected by a rare disease.
9 Depends:
10   R (>= 3.2.5)
11 Imports:
12   kinship2 (>= 1.6.4),
13   dplyr (>= 0.4.3),
14   stats (>= 3.3.0)
15 Suggests:
16   doParallel (>= 1.0.10),
17   doRNG (>= 1.6),
18 License: GNU General Public License
19 LazyData: TRUE
20 RoxygenNote: 6.0.1
21 VignetteBuilder: knitr
```

---

For more information, refer to Chapter 4 of “*R Packages*”

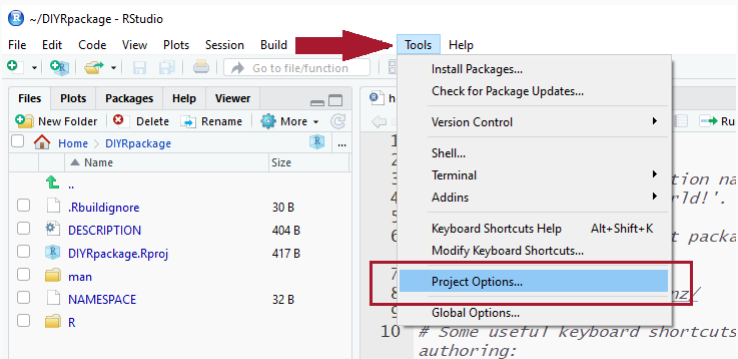
In conjunction with `devtools`, `roxygen2` is a really nice R package that will simplify development.

The `roxygen2` package:

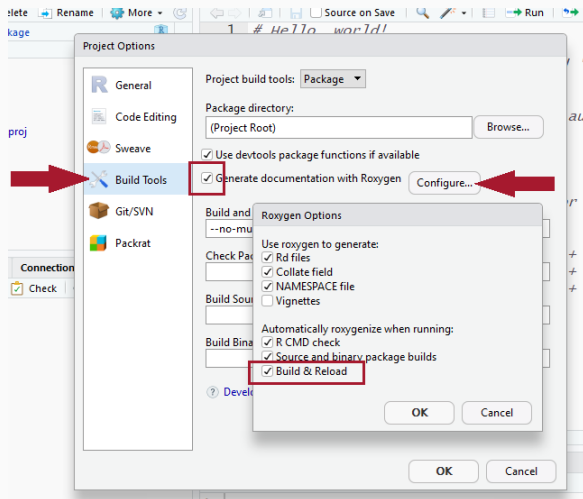
- ▶ manages the `NAMESPACE` file,
- ▶ handles text formatting in documentation, and
- ▶ generates `.Rd` files in the `\man` directory for documented functions.

All this can be yours... but first, you must configure the project build tools to use `roxygen2`.

# CONFIGURE ROXYGEN2

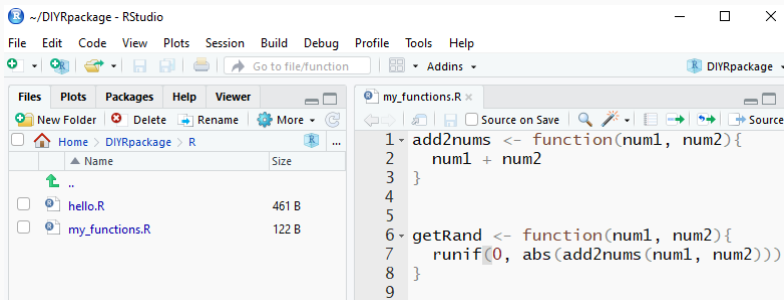


To configure, click on the *Tools* menu and select *Project Options...*

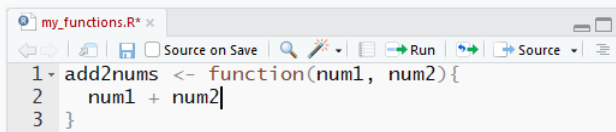


Under *Build Tools*, configure Roxygen as displayed above.

- ▶ R functions are stored in .R files in the \R directory.
- ▶ You may store and document multiple functions in the same script.



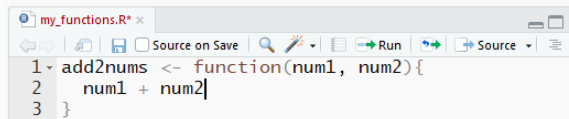
# CREATE R FUNCTION AND DOCUMENTATION



```
my_functions.R* x
Source on Save
1 add2nums <- function(num1, num2){
2   num1 + num2
3 }
```

Let's start by documenting this simple function, `add2num`, which returns the sum of `num1` and `num2`.



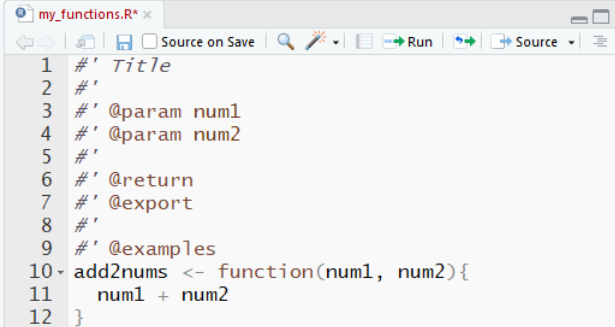


The screenshot shows an RStudio editor window with a single file named 'my\_functions.R'. The editor contains the following R code:

```
1 add2nums <- function(num1, num2){  
2   num1 + num2  
3 }
```

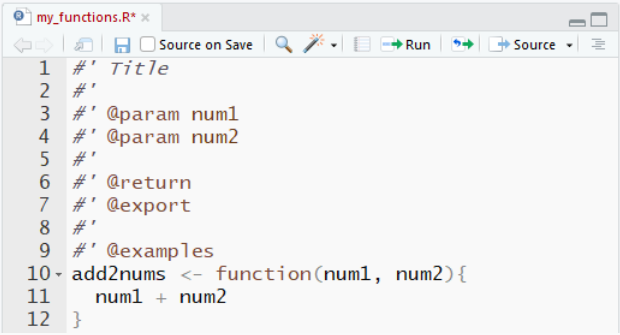
The code is displayed with line numbers 1, 2, and 3 on the left margin. The function is named 'add2nums' and takes two arguments, 'num1' and 'num2'. The body of the function consists of a single line 'num1 + num2' followed by a closing curly brace '}'.

- To generate a description skeleton with roxygen2:
  1. place cursor inside function definition
  2. hold: **Shift + Ctrl/Cmd + Alt + R**



```
my_functions.R* x
Source on Save
Run
Source
1 #' Title
2 #'
3 #' @param num1
4 #' @param num2
5 #'
6 #' @return
7 #' @export
8 #'
9 #' @examples
10 add2nums <- function(num1, num2){
11   num1 + num2
12 }
```

Now we have a description skeleton generated by roxygen2.



The screenshot shows an RStudio editor window titled "my\_functions.R\* x". The toolbar includes icons for navigation, saving, and running code. The code in the editor is as follows:

```
1 #' Title
2 #'
3 #' @param num1
4 #' @param num2
5 #'
6 #' @return
7 #' @export
8 #'
9 #' @examples
10 add2nums <- function(num1, num2){
11   num1 + num2
12 }
```

## Quick Observations:

- ▶ all roxygen2 comments start with `#'` , and
- ▶ roxygen2 keywords start with `@`.

`@param` documents function arguments.

`@return` documents what the function returns to the user.  
This can be a single object or a list of objects.

`@export` Tells R to make the function available to anyone using your package. For now, always include `@export`.

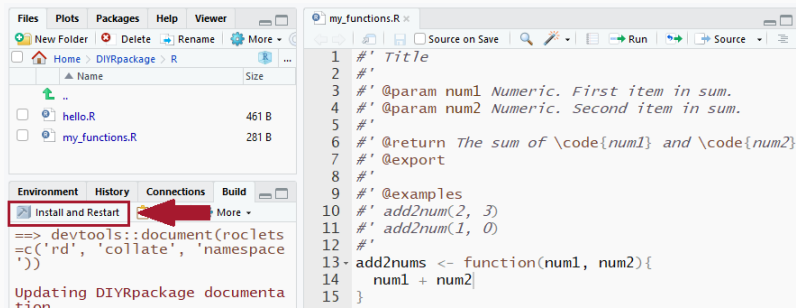
`@examples` documents example R code.

Other useful keywords:

`@inheritParams` inherit parameters of another function.

`@seealso` provide links to other functions or resources.

To convert roxygen2 comments to .Rd files:  
click on *Install and Restart*, under the *Build* tab.

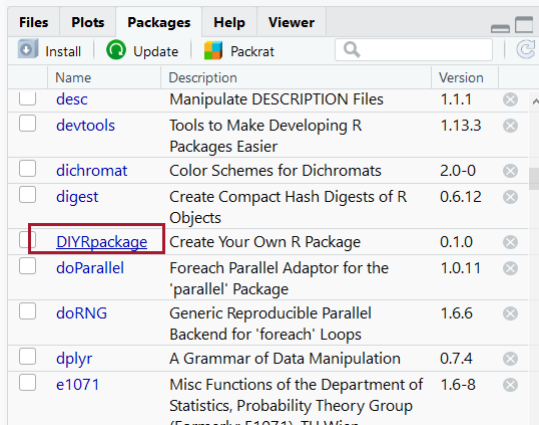


The screenshot shows the RStudio interface. On the left, the 'Files' pane displays a directory structure with 'hello.R' (461 B) and 'my\_functions.R' (281 B). Below it, the 'Environment' pane shows the command `devtools::document(roclerts = c('rd', 'collate', 'namespace'))` and the message 'Updating DIYRpackage documentation'. The 'Build' tab is selected, and the 'Install and Restart' button is highlighted with a red box and a red arrow. On the right, the 'my\_functions.R' file is open in the editor, showing roxygen2 comments and a function definition:

```

1 #' Title
2 #'
3 #' @param num1 Numeric. First item in sum.
4 #' @param num2 Numeric. Second item in sum.
5 #'
6 #' @return The sum of \code{num1} and \code{num2}
7 #' @export
8 #'
9 #' @examples
10 #' add2num(2, 3)
11 #' add2num(1, 0)
12 #'
13 add2nums <- function(num1, num2){
14   num1 + num2
15 }
```

Find your package in the list of installed packages.



The screenshot shows the RStudio interface with the 'Packages' tab selected. The 'Update' button is highlighted with a green circle. A search bar is visible at the top right of the package list. The list of packages includes:

	Name	Description	Version	
<input type="checkbox"/>	desc	Manipulate DESCRIPTION Files	1.1.1	ⓧ ^
<input type="checkbox"/>	devtools	Tools to Make Developing R Packages Easier	1.13.3	ⓧ
<input type="checkbox"/>	dichromat	Color Schemes for Dichromats	2.0-0	ⓧ
<input type="checkbox"/>	digest	Create Compact Hash Digests of R Objects	0.6.12	ⓧ
<input type="checkbox"/>	<b>DIYRpackage</b>	Create Your Own R Package	0.1.0	ⓧ
<input type="checkbox"/>	doParallel	Foreach Parallel Adaptor for the 'parallel' Package	1.0.11	ⓧ
<input type="checkbox"/>	doRNG	Generic Reproducible Parallel Backend for 'foreach' Loops	1.6.6	ⓧ
<input type="checkbox"/>	dplyr	A Grammar of Data Manipulation	0.7.4	ⓧ
<input type="checkbox"/>	e1071	Misc Functions of the Department of Statistics, Probability Theory Group (Formerly E1071), TU Wien	1.6-8	ⓧ

Navigate to the documented function.

## Create Your Own R Package



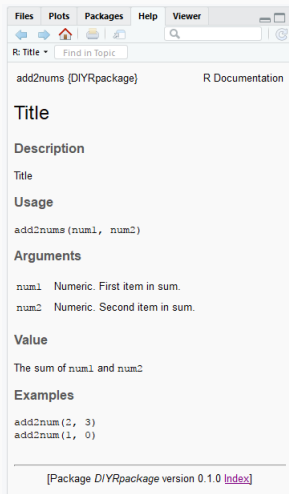
Documentation for package  
'DIYRpackage' version 0.1.0

- [DESCRIPTION file](#).

## Help Pages

<a href="#">add2nums</a>	Title
<a href="#">hello</a>	Hello, World!

# UPDATE AND NAVIGATE TO DOCUMENTATION



The screenshot shows the RStudio interface with the 'Viewer' pane displaying the documentation for the `add2nums` function from the `DIYRpackage`. The documentation includes sections for Title, Description, Usage, Arguments, Value, and Examples.

**add2nums** (DIYRpackage) R Documentation

**Title**

**Description**

**Usage**

```
add2nums(num1, num2)
```

**Arguments**

`num1` Numeric. First item in sum.  
`num2` Numeric. Second item in sum.

**Value**

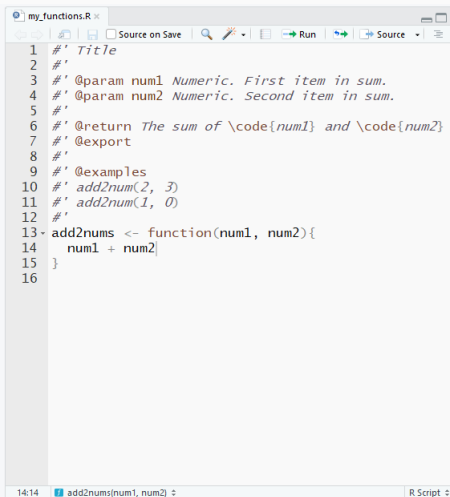
The sum of `num1` and `num2`

**Examples**

```
add2num(2, 3)  
add2num(1, 0)
```

---

[Package *DIYRpackage* version 0.1.0 [index](#)]



The screenshot shows the RStudio interface with the 'Source' pane displaying the source code for the `my_functions.R` file. The code defines the `add2num` and `add2nums` functions.

```
1 #' Title  
2 #'  
3 #' @param num1 Numeric. First item in sum.  
4 #' @param num2 Numeric. Second item in sum.  
5 #'  
6 #' @return The sum of \code{num1} and \code{num2}  
7 #' @export  
8 #'  
9 #' @examples  
10 #' add2num(2, 3)  
11 #' add2num(1, 0)  
12 #'  
13 add2nums <- function(num1, num2){  
14   num1 + num2  
15 }  
16
```



If “*Description*” is not specified, “*Title*” is recycled and used again for “*Description*.”

The screenshot displays the RStudio interface with two panes. The left pane shows the documentation for the `add2nums` function from the `DIYRpackage`. The right pane shows the source code of `my_functions.R`. Red boxes and arrows illustrate how the `Title` field in the documentation is recycled to serve as the `Description` when no explicit description is provided in the code comments.

**Left Pane (Documentation):**

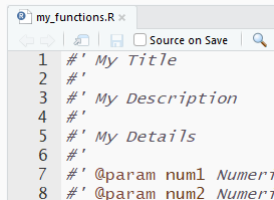
- Title:** add2nums (DIYRpackage)
- Description:** Title
- Usage:** add2nums(num1, num2)
- Arguments:**
  - num1 Numeric. First item in sum.
  - num2 Numeric. Second item in sum.
- Value:** The sum of num1 and num2
- Examples:**

```
add2num(2, 3)
add2num(1, 0)
```

**Right Pane (Source Code):**

```
1 #' Title
2 #'
3 #' @param num1 Numeric. First item in sum.
4 #' @param num2 Numeric. Second item in sum.
5 #'
6 #' @return The sum of \code{num1} and \code{num2}
7 #' @export
8 #'
9 #' @examples
10 #' add2num(2, 3)
11 #' add2num(1, 0)
12 #'
13 add2nums <- function(num1, num2){
14   num1 + num2
15 }
16
```

NOTE: There are no keyword specifiers for *"Title"*, *"Description"*, or *"Details"*. To specify, simply separate text for each section by a blank line.



```
1 #' My Title
2 #'
3 #' My Description
4 #'
5 #' My Details
6 #'
7 #' @param num1 Numerical
8 #' @param num2 Numerical
```

roxygen2 always treats the first chunk as *"Title"*, the second chunk as *"Description"*, and the third chunk as *"Details"*.

# CREATE DESCRIPTION AND DETAILS

The image shows the relationship between R package documentation sections and their corresponding source code comments. Red boxes highlight specific sections in the documentation and their corresponding comment lines in the source code, with red arrows indicating the mapping.

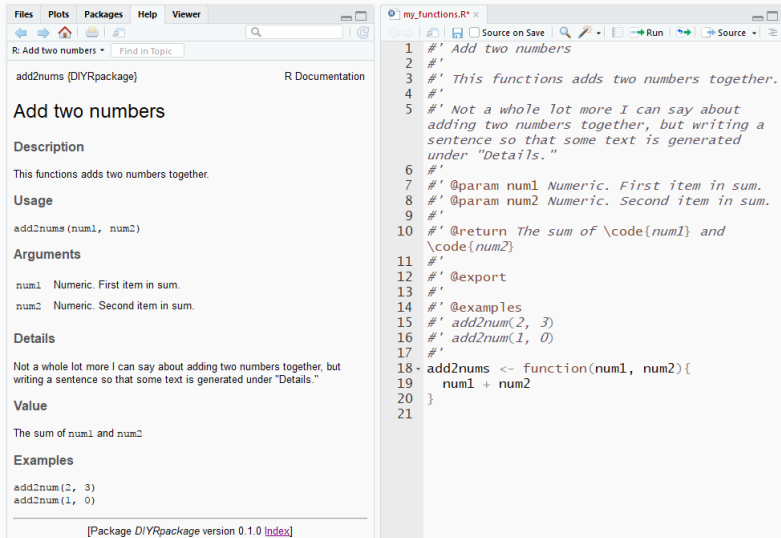
**Documentation Sections (Left Panel):**

- My Title** (linked to line 1: `#' My Title`)
- Description** (linked to line 3: `#' My Description`)
- Usage** (linked to line 5: `#' My Details`)
- Arguments** (linked to line 7: `#' @param num1 Numeric. First item in sum.`)
- Details** (linked to line 13: `#' @examples`)
- Value** (linked to line 17: `add2nums <- function(num1, num2){`)
- Examples** (linked to line 18: `num1 + num2`)

**Source Code (Right Panel):**

```
1 #' My Title
2 #'
3 #' My Description
4 #'
5 #' My Details
6 #'
7 #' @param num1 Numeric. First item in sum.
8 #' @param num2 Numeric. Second item in sum.
9 #'
10 #' @return The sum of \code{num1} and \code{num2}
11 #' @export
12 #'
13 #' @examples
14 #' add2num(2, 3)
15 #' add2num(1, 0)
16 #'
17 add2nums <- function(num1, num2){
18   num1 + num2
19 }
20
```

[Package *DIYRpackage* version 0.1.0 [Index](#)]



The screenshot displays the RStudio interface with two panes. The left pane shows the documentation for the `add2nums` function from the `DIYRpackage`. The right pane shows the source code for `my_functions.R`.

**R: Add two numbers** Find in Topic

`add2nums` (DIYRpackage) R Documentation

## Add two numbers

### Description

This functions adds two numbers together.

### Usage

```
add2nums(num1, num2)
```

### Arguments

`num1` Numeric. First item in sum.  
`num2` Numeric. Second item in sum.

### Details

Not a whole lot more I can say about adding two numbers together, but writing a sentence so that some text is generated under "Details."

### Value

The sum of `num1` and `num2`

### Examples

```
add2num(2, 3)  
add2num(1, 0)
```

[Package *DIYRpackage* version 0.1.0 [Index](#)]

**my\_functions.R**

```
1 #' Add two numbers  
2 #'  
3 #' This functions adds two numbers together.  
4 #'  
5 #' Not a whole lot more I can say about  
6   adding two numbers together, but writing a  
7   sentence so that some text is generated  
8   under "Details."  
9 #'  
10 #' @param num1 Numeric. First item in sum.  
11 #' @param num2 Numeric. Second item in sum.  
12 #'  
13 #' @return The sum of \code{num1} and  
14   \code{num2}  
15 #'  
16 #' @export  
17 #'  
18 #' @examples  
19 #' add2num(2, 3)  
20 #' add2num(1, 0)  
21 #'  
22 add2nums <- function(num1, num2){  
23   num1 + num2  
24 }  
25
```

Let's use roxygen2 to simplify documentation of getRand.

```
21 #' Title
22 #'
23 #' @param num1
24 #' @param num2
25 #'
26 #' @return
27 #' @export
28 #'
29 #' @examples
30 getRand <- function(num1, num2){
31   runif(0, abs(add2nums(num1, num2)))|
32 }
33
```


To avoid re-documenting `num1` and `num2`, we replace `@param` with `@inheritParams`.

```
21 #' Title
22 #'
23 #' @inheritParams add2nums
24 #' @importFrom stats runif
25 #'
26 #' @return
27 #' @export
28 #'
29 #' @examples
30 #' getRand(2, 3)
31 getRand <- function(num1, num2){
32   runif(0, abs(add2nums(num1, num2)))
33 }
```


To import functions from other packages:

- ▶ Use @importFrom <package> <function>
- ▶ Do not use: library(package)

```
#' Title
#'  
#' @inheritParams add2nums  
#' @importFrom stats runif  
#'  
#' @return  
#' @export  
#'  
#' @examples  
#' getRand(2, 3)  
getRand <- function(num1, num2){  
  runif(0, abs(add2nums(num1, num2)))  
}
```

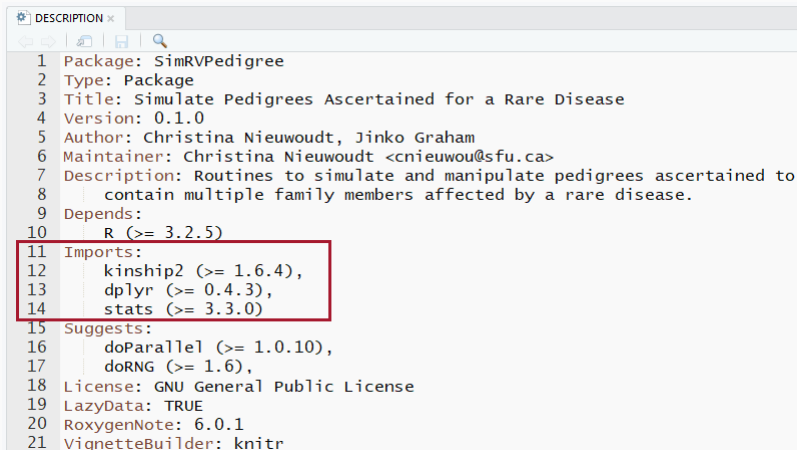


```
#' Title
#'  
#' @inheritParams add2nums  
#'  
#' @return  
#' @export  
#'  
#' @examples  
#' getRand(2, 3)  
library(stats)  
getRand <- function(num1, num2){  
  runif(0, abs(add2nums(num1, num2)))  
}
```



## INCLUDE PACKAGES IN DESCRIPTION

Include any required packages in the DESCRIPTION file, under Imports, along with the version number of the package.



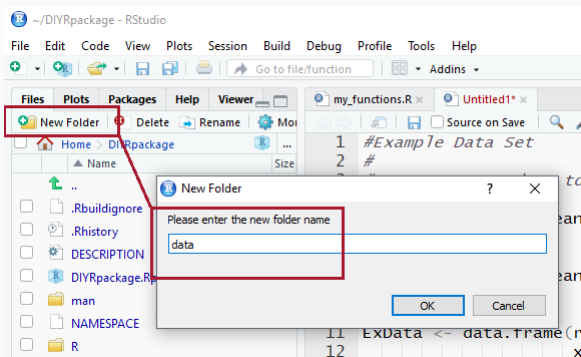
```
DESCRIPTION x
1 Package: SimRVPedigree
2 Type: Package
3 Title: Simulate Pedigrees Ascertained for a Rare Disease
4 Version: 0.1.0
5 Author: Christina Nieuwoudt, Jinko Graham
6 Maintainer: Christina Nieuwoudt <cnieuwou@sfu.ca>
7 Description: Routines to simulate and manipulate pedigrees ascertained to
8 | contain multiple family members affected by a rare disease.
9 Depends:
10 | R (>= 3.2.5)
11 Imports:
12 | kinship2 (>= 1.6.4),
13 | dplyr (>= 0.4.3),
14 | stats (>= 3.3.0)
15 Suggests:
16 | doParallel (>= 1.0.10),
17 | doRNG (>= 1.6),
18 License: GNU General Public License
19 LazyData: TRUE
20 RoxygenNote: 6.0.1
21 VignetteBuilder: knitr
```



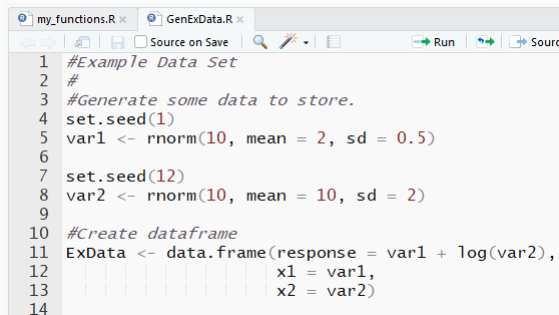
# CREATE AND STORE DATA SET

Data sets are stored in `\data`. We will need to create this directory.

1. Click on *New Folder* under *Files*
2. Name and create the data directory.

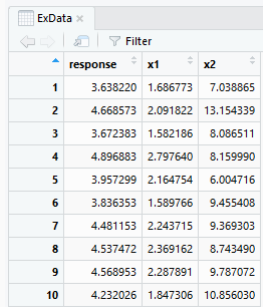


If you are following along, execute GenExData.R.



```
1 #Example Data Set
2 #
3 #Generate some data to store.
4 set.seed(1)
5 var1 <- rnorm(10, mean = 2, sd = 0.5)
6
7 set.seed(12)
8 var2 <- rnorm(10, mean = 10, sd = 2)
9
10 #Create dataframe
11 ExData <- data.frame(response = var1 + log(var2),
12                       x1 = var1,
13                       x2 = var2)
14
```

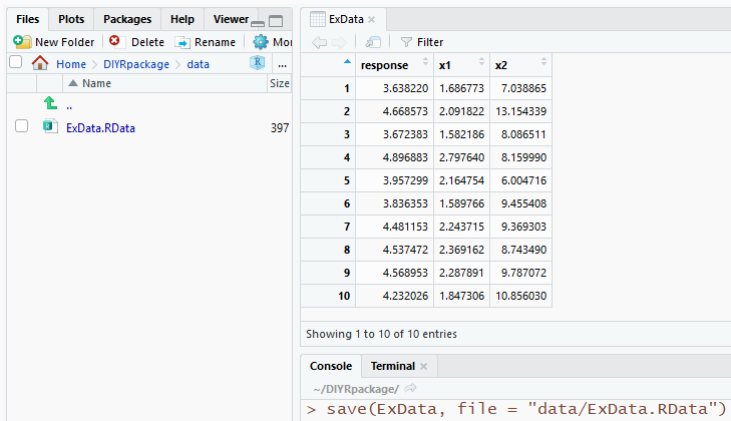
Goal: store ExData in the \data directory.



	response	x1	x2
1	3.638220	1.686773	7.038865
2	4.668573	2.091822	13.154339
3	3.672383	1.582186	8.086511
4	4.896883	2.797640	8.159990
5	3.957299	2.164754	6.004716
6	3.836353	1.589766	9.455408
7	4.481153	2.243715	9.369303
8	4.537472	2.369162	8.743490
9	4.568953	2.287891	9.787072
10	4.232026	1.847306	10.856030

In the console execute the command:  
`save(ExData, file = "data/ExData.RData")`

# CREATE AND STORE DATA SET



The screenshot displays the RStudio interface. On the left, the 'Files' pane shows the project structure: Home > DIYRpackage > data. A file named 'ExData.RData' is listed with a size of 397 bytes. The main editor pane on the right shows the 'ExData' dataset as a data frame with 10 rows and 4 columns: 'response', 'x1', and 'x2'. Below the data frame, it indicates 'Showing 1 to 10 of 10 entries'. At the bottom, the 'Console' pane shows the command: `> save(ExData, file = "data/ExData.RData")`.

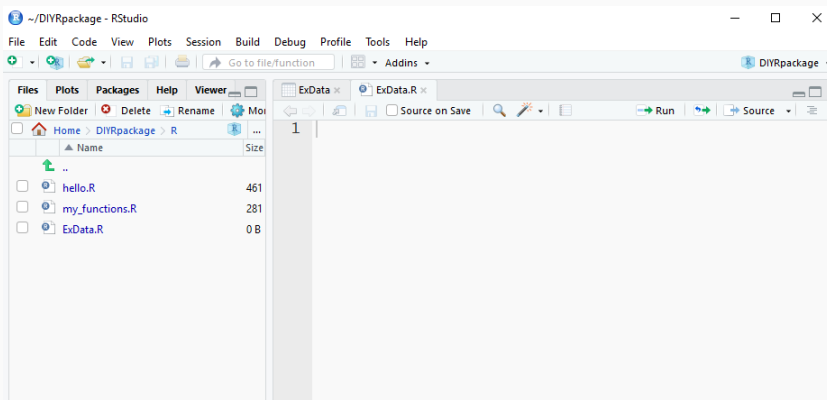
	response	x1	x2
1	3.638220	1.686773	7.038865
2	4.668573	2.091822	13.154339
3	3.672383	1.582186	8.086511
4	4.896883	2.797640	8.159990
5	3.957299	2.164754	6.004716
6	3.836353	1.589766	9.455408
7	4.481153	2.243715	9.369303
8	4.537472	2.369162	8.743490
9	4.568953	2.287891	9.787072
10	4.232026	1.847306	10.856030

```
> save(ExData, file = "data/ExData.RData")
```

- ▶ Documentation for a data set is created in an .R script, stored in the \R directory, with the same name as data set stored in \data.
- ▶ Each data set needs a separate documentation file.

# CREATE DATA DOCUMENTATION

1. Hold **Shift + Ctrl/Cmd + N** to create a new script.
2. Save as `ExData.R` in `\R`.



Files Plots Packages Help Viewer

R: ExData ▾ Find in Topic

ExData (DIYRpackage) R Documentation

## ExData

### Description

ExData Description

### Usage

ExData

### Format

A data set with 10 rows and 3 variables:

response

sum of var1 and log(var2)

var1

normal, mean = 2, sd = 0.5

var2

normal, mean = 10, sd = 2

### Details

ExData Details

### Source

[http://www.web\\_site.com](http://www.web_site.com)

---

[Package DIYRpackage version 0.1.0 [Index](#)]

ExData x ExData.R x DatasetDocumentationSkeleton.R x Untitled

Source on Save Run Source

```
1 #' ExData
2 #'
3 #' ExData Description
4 #'
5 #' ExData Details
6 #'
7 #' @format A data set with 10 rows and 3
  variables:
8 #' \describe{
9 #'   \item{response}{sum of var1 and log(var2)}
10 #'   \item{var1}{normal, mean = 2, sd = 0.5}
11 #'   \item{var2}{normal, mean = 10, sd = 2}
12 #' }
13 #'
14 #' @source \url{http://www.web_site.com}
15 #'
16 "ExData"
17
18
```

14:1 (Top Level) ↕ R Script ↕

As with function documentation, `roxygen2` treats the first chunk as “*Title*”, the second chunk as “*Description*”, and the third chunk as “*Details*”.

The screenshot shows the RStudio interface with the documentation for the `ExData` function. The left pane displays the rendered documentation, and the right pane shows the source code with roxygen2 comments. Red boxes and arrows highlight the mapping between the rendered sections and the source code lines.

**Rendered Documentation (Left Pane):**

- ExData**
- Description**
- Usage**
- Format**
- Details**

**Source Code (Right Pane):**

```

1 #' ExData
2 #'
3 #' ExData Description
4 #'
5 #' ExData Details
6 #'
7 #' @format A data set with 10 rows and 3
  variables:
8 #' \describe{
9 #'   \item{response}{sum of var1 and log(var2)}
10 #'   \item{var1}{normal, mean = 2, sd = 0.5}
11 #'   \item{var2}{normal, mean = 10, sd = 2}
12 #' }
13 #'
14 #' @source \url{http://www.web_site.com}
15 #'
16 "ExData"
17
18
  
```

**Mapping:**

- ExData** (Section) → Line 1: `1 #' ExData`
- ExData Description** (Section) → Line 3: `3 #' ExData Description`
- ExData Details** (Section) → Line 5: `5 #' ExData Details`



# CREATE DATA DOCUMENTATION

The screenshot shows the RStudio interface with the documentation for the 'ExData' package. The left pane displays the package documentation, and the right pane shows the corresponding R source code. Red boxes highlight the '@format' and '@source' tags in the code, with red arrows pointing to their respective sections in the documentation.

**Documentation (Left Pane):**

- ExData (DIYRpackage) R Documentation
- ExData
- Description
- ExData Description
- Usage
- ExData
- Format**
- A data set with 10 rows and 3 variables:
- response
- sum of var1 and log(var2)
- var1
- normal, mean = 2, sd = 0.5
- var2
- normal, mean = 10, sd = 2
- Details
- ExData Details
- Source**
- [http://www.web\\_site.com](http://www.web_site.com)
- [Package DIYRpackage version 0.1.0 [Index](#)]

**Source Code (Right Pane):**

```
1 #' ExData
2 #'
3 #' ExData Description
4 #'
5 #' ExData Details
6 #'
7 #' @format A data set with 10 rows and 3
8 #' variables:
9 #' \describe{
10 #'   \item{response}{sum of var1 and log(var2)}
11 #'   \item{var1}{normal, mean = 2, sd = 0.5}
12 #'   \item{var2}{normal, mean = 10, sd = 2}
13 #' }
14 #' @source \url{http://www.web_site.com}
15
16 "ExData"
17
18
```

- ▶ @format is used to give an overview of the data
- ▶ @source is used to provide the data source. For multiple sources, use @source before each source.

The last line is always the dataset name in quotation marks.

Note: this is not a roxygen2 comment.

The screenshot shows the RStudio interface with two panes. The left pane displays the rendered documentation for a dataset named 'ExData'. The right pane shows the source R code, which uses roxygen2-style comments to generate the documentation. The last line of the code, line 16, is highlighted with a red box.

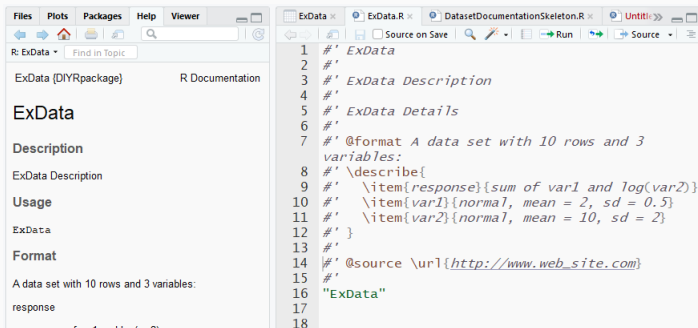
**Left Pane (Rendered Documentation):**

- Files: ExData (DIYRpackage) R Documentation
- ExData
- Description
  - ExData Description
- Usage
  - ExData
- Format
  - A data set with 10 rows and 3 variables:
  - response
    - sum of var1 and log(var2)
  - var1
    - normal, mean = 2, sd = 0.5
  - var2
    - normal, mean = 10, sd = 2
- Details
  - ExData Details
- Source
  - [http://www.web\\_site.com](http://www.web_site.com)

**Right Pane (R Code):**

```

1 #' ExData
2 #'
3 #' ExData Description
4 #'
5 #' ExData Details
6 #'
7 #' @format A data set with 10 rows and 3
  variables:
8 #' \describe{
9 #'   \item{response}{sum of var1 and log(var2)}
10 #'   \item{var1}{normal, mean = 2, sd = 0.5}
11 #'   \item{var2}{normal, mean = 10, sd = 2}
12 #' }
13 #'
14 #' @source \url{http://www.web_site.com}
15 #
16 "ExData"
17
18
  
```

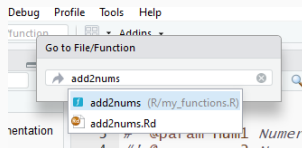


The screenshot displays the RStudio interface with two panes. The left pane shows the 'ExData' package documentation, including sections for Description, Usage, and Format. The right pane shows the source code for 'ExData' documentation, which includes comments for the package name, description, details, format, and source, as well as the package name 'ExData'.

```
1 #' ExData
2 #'
3 #' ExData Description
4 #'
5 #' ExData Details
6 #'
7 #' @format A data set with 10 rows and 3
  variables:
8 #' \describe{
9 #'   \item{response}{sum of var1 and log(var2)}
10 #'   \item{var1}{normal, mean = 2, sd = 0.5}
11 #'   \item{var2}{normal, mean = 10, sd = 2}
12 #' }
13 #'
14 #' @source \url{http://www.web_site.com}
15 #'
16 "ExData"
17
18
```

Never include @export in data documentation,  
@export is used for functions only.

- ▶ To generate roxygen2 description skeleton:
  1. place cursor inside function definition
  2. hold: **Shift + Ctrl/Cmd + Alt + R**
- ▶ To jump to function definition:
  1. hold: **Ctrl/Cmd + .**
  2. type function name and choose from drop down list.



After clicking *Install and Reload*, if you receive the error message:

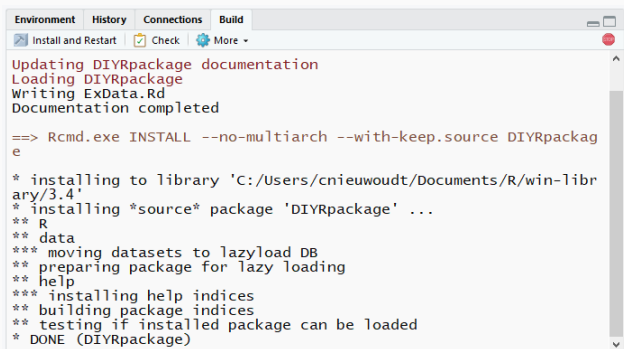
```
Warning: The existing 'NAMESPACE' file was not generated  
by roxygen2, and will not be overwritten.
```

Try the following fix:

- ▶ Backup the NAMESPACE file (just in case)
- ▶ Delete the NAMESPACE file from the package directory
- ▶ In the console execute the command:  
`devtools::document()`.

Roxygen should create a new NAMESPACE file. If it works, you should not have to repeat this fix, again.

Occasionally, the package locks after clicking *Install and Reload*.



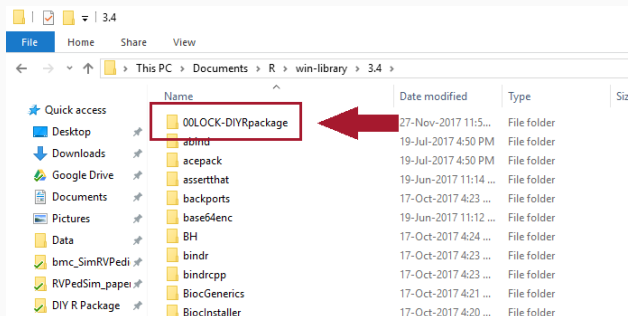
```
Environment History Connections Build
Install and Restart Check More v

Updating DIYRpackage documentation
Loading DIYRpackage
Writing ExData.Rd
Documentation completed

==> Rcmd.exe INSTALL --no-multiarch --with-keep.source DIYRpackag
e

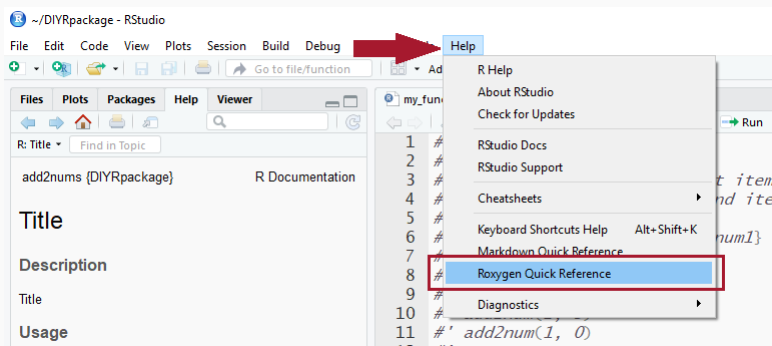
* installing to library 'C:/Users/cnieuwoudt/Documents/R/win-libr
ary/3.4'
* installing *source* package 'DIYRpackage' ...
** R
** data
*** moving datasets to lazyload DB
** preparing package for lazy loading
** help
*** installing help indices
** building package indices
** testing if installed package can be loaded
* DONE (DIYRpackage)
```

You'll notice that after you hit *Install and Reload*, the **STOP** symbol does not go away and the library doesn't reload.

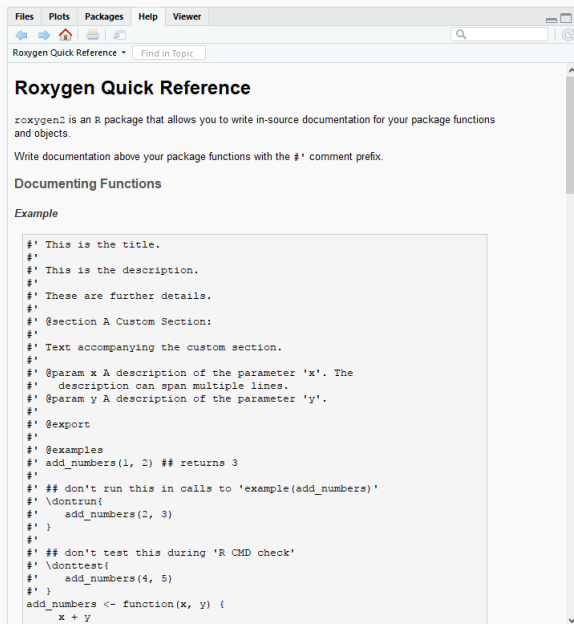


- ▶ Navigate to the library with the same version number as your current installation of R.
- ▶ Locate the file entitled “00LOCK-package\_name”.
- ▶ Delete it.

In Rstudio, additional information for `roxygen2` may be found by navigating to *Roxygen Quick Reference* in the *Help* menu







The screenshot shows the RStudio interface with the 'Help' tab selected. The 'Roxygen Quick Reference' page is displayed, featuring a search bar and a table of contents. The main content area shows the 'Documenting Functions' section, which includes an 'Example' of R code with roxygen2 comments. The code defines a function 'add\_numbers' and includes various roxygen2 tags like '@section', '@param', '@export', '@examples', and '\donttest'.

Files Plots Packages Help Viewer

Roxygen Quick Reference Find in Topic

## Roxygen Quick Reference

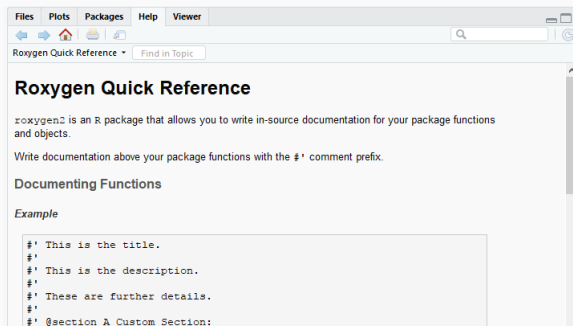
roxygen2 is an R package that allows you to write in-source documentation for your package functions and objects.

Write documentation above your package functions with the `#'` comment prefix.

### Documenting Functions

*Example*

```
#' This is the title.
#
#' This is the description.
#
#' These are further details.
#
#' @section A Custom Section:
#
#' Text accompanying the custom section.
#
#' @param x A description of the parameter 'x'. The
#'   description can span multiple lines.
#' @param y A description of the parameter 'y'.
#
#' @export
#
#' @examples
#' add_numbers(1, 2) ## returns 3
#
#' ## don't run this in calls to 'example(add_numbers)'
#' \donttest{
#'   add_numbers(2, 3)
#' }
#
#' ## don't test this during 'R CMD check'
#' \donttest{
#'   add_numbers(4, 5)
#' }
add_numbers <- function(x, y) {
  x + y
}
```



## Additional references:

- ▶ <https://cran.r-project.org/web/packages/roxygen2/vignettes/roxygen2.html>
- ▶ Hadley Wickham's book "R Packages", Chapter 5
- ▶ ... the all-mighty and powerful Google.

- ▶ Function Documentation: Chapter 5, pp. 49
- ▶ roxygen2 introduction: Chapter 5
- ▶ Data Documentation: Chapter 9.1.1, pp. 107
- ▶ .Rbuildignore: Chapter 2
- ▶ DESCRIPTION file: Chapter 4