

Lecture 28: Nov 28, 2018

R Packages

- *Motivation*
- *Creating an R Package*
 - *RStudio Cloud* and *RStudio Desktop*
- *Structure of R Packages*

James Balamuta
STAT 385 @ UIUC



Announcements

- **Group Project Final Report, Demo Video, and Peer Evaluation** due **Tuesday, December 18th at 10:00 PM**
 - Details: <http://stat385.stat.illinois.edu/group-projects/>
- **hw09** due **Friday, November 30th at 6:00 PM**
- **Quiz 13** covers Week 12 contents @ [CBTF](#).
 - Window: Nov 27th - Nov 29th
 - Sign up: <https://cbtf.engr.illinois.edu/sched>
- Want to review your homework or quiz grades?
Schedule an appointment.

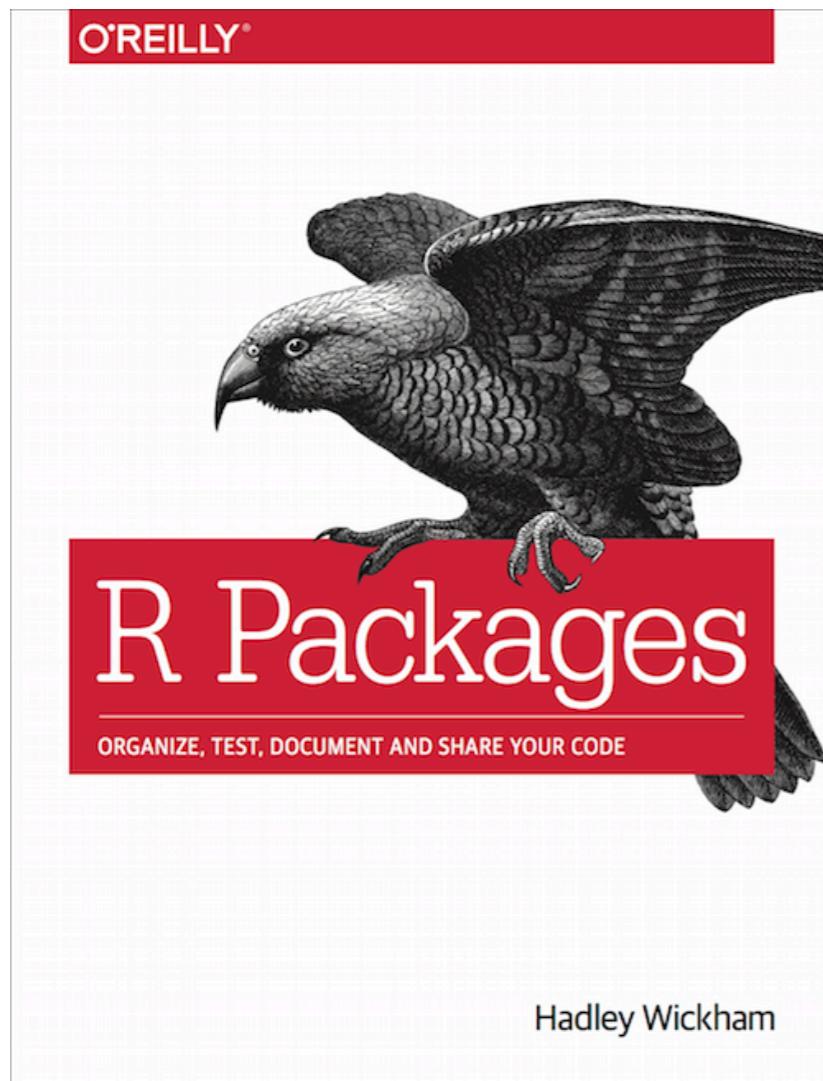
Last Time

- **OOP**
 - Object oriented programming emphasizes objects in a collection with interactions.
 - Focuses on **encapsulation**, **inheritance**, and **polymorphism**.
- **S3 Objects**
 - Each object has an underlying **class** attribute.
 - **Generic functions** are used to dispatch to the appropriate method function for an object.
- **Unpaired (Two-Sample) t-Test**
 - Case study in implementing a method for assessing group difference.

Lecture Objectives

- Explain the **benefits** of *R* packages
- **Structuring** code for an *R* package
- **Developing** unit tests for functions

Motivation



“Packages are the **fundamental units of reproducible** *R* code. They include reusable *R* functions, the documentation that describes how to use them, and sample data.”

– Hadley Wickham in [R packages](#)

R Code vs. *R* Packages

... in the context of shipping containers ...



[Source](#)



[Source](#)



[Source](#)



[Source](#)

13,468 R packages



exist on [Comprehensive R Archival Network \(CRAN\)](#)

R Package Structure

... extending the *R* language ...

Data

trees

Girth	Height	Volume

mtcars

mpg	...	gear	carb

...

Help



...

Functions

mean(x, ...)
sd(x, ...)
sum(x, ...)
prod(x, ...)
plot(x, y, ...)

...

Why an *R* Package?

... why not ???

1. Code is already organized via functions.
2. Nothing already exists
3. Give back to the community: **pay it forward**
4. Portfolio of your coding and development skills.

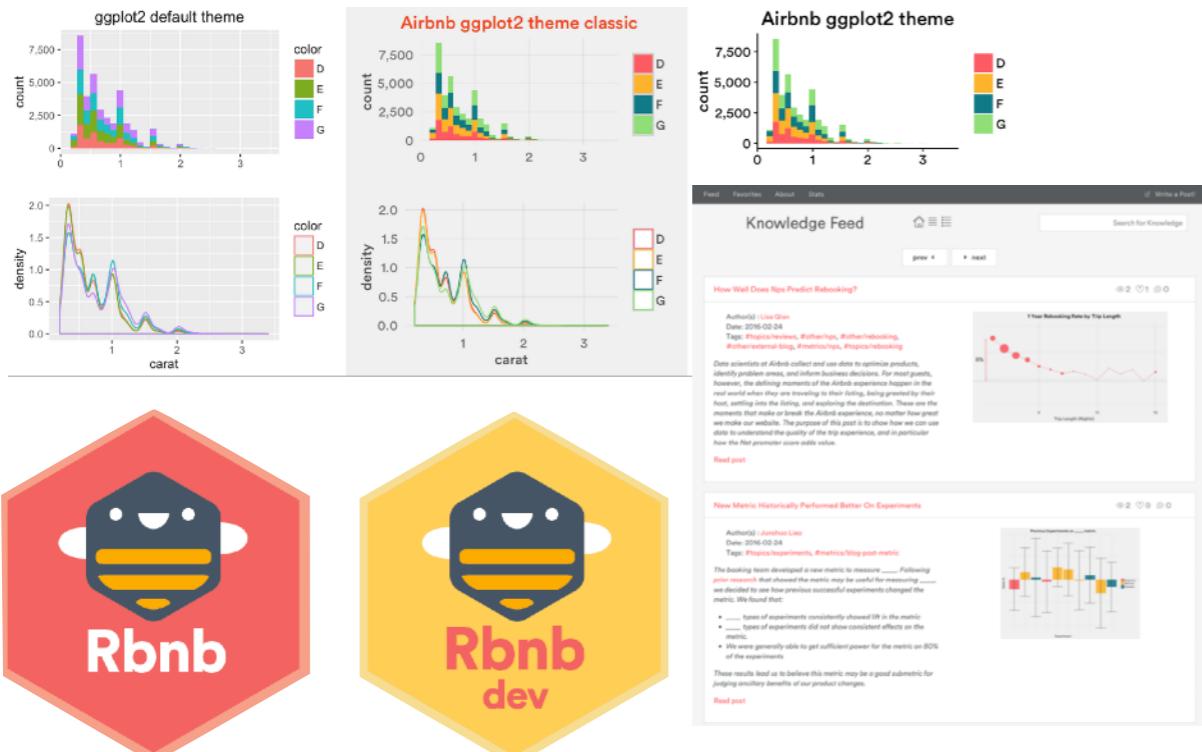
R Script vs. R Package

... the battle for the ages ...

R Script	R Package
Focus on one data analysis project	Generalized to be used on more than one data analysis project.
Creates analysis results	Builds new <i>R</i> objects
Pollutes the Global Environment	Chunked logic through functions

Rbnb

... industry example of Airbnb designing an *R* package ...



- [rstudio::conf 2017 - Airbnb @ 10:47](#)
- [Using R packages and education to scale Data Science at Airbnb](#)
- [Airbnb visualization techniques](#)
- [Airbnb Knowledge Repo](#)
- [PeerJ: How R Helps Airbnb Make the Most of Its Data](#)

Internal R Packages

... more guidelines ...

RStudio Views: Some Ideas for your Internal R Package

The screenshot shows the RStudio interface with a LaTeX source file named `example_beamer_illinois.Rmd` open in the Source Editor. The code includes Beamer configuration, section titles, lists, and a LaTeX block. To the right, three Beamer slides are displayed:

- Slide 1:** Title "Illinois" UIUC Beamer Theme, author John and Mary Doe, department Department of Magic, University of Illinois at Urbana-Champaign, date January 21, 2018.
- Slide 2:** Frame Title, Unordered List (University of Illinois at Urbana-Champaign, Department of Statistics, Illinois Informatics Institute), Ordered List (links to thecoatlessprofessor.com and github.com/coatless).
- Slide 3:** L^AT_EX, Binomial Theorem, equation
$$f(k) = \binom{n}{k} p^k (1-p)^{n-k}$$
.

<https://github.com/coatless/uiucthemes>

Creating an *R* Packages

Tools to Develop Packages

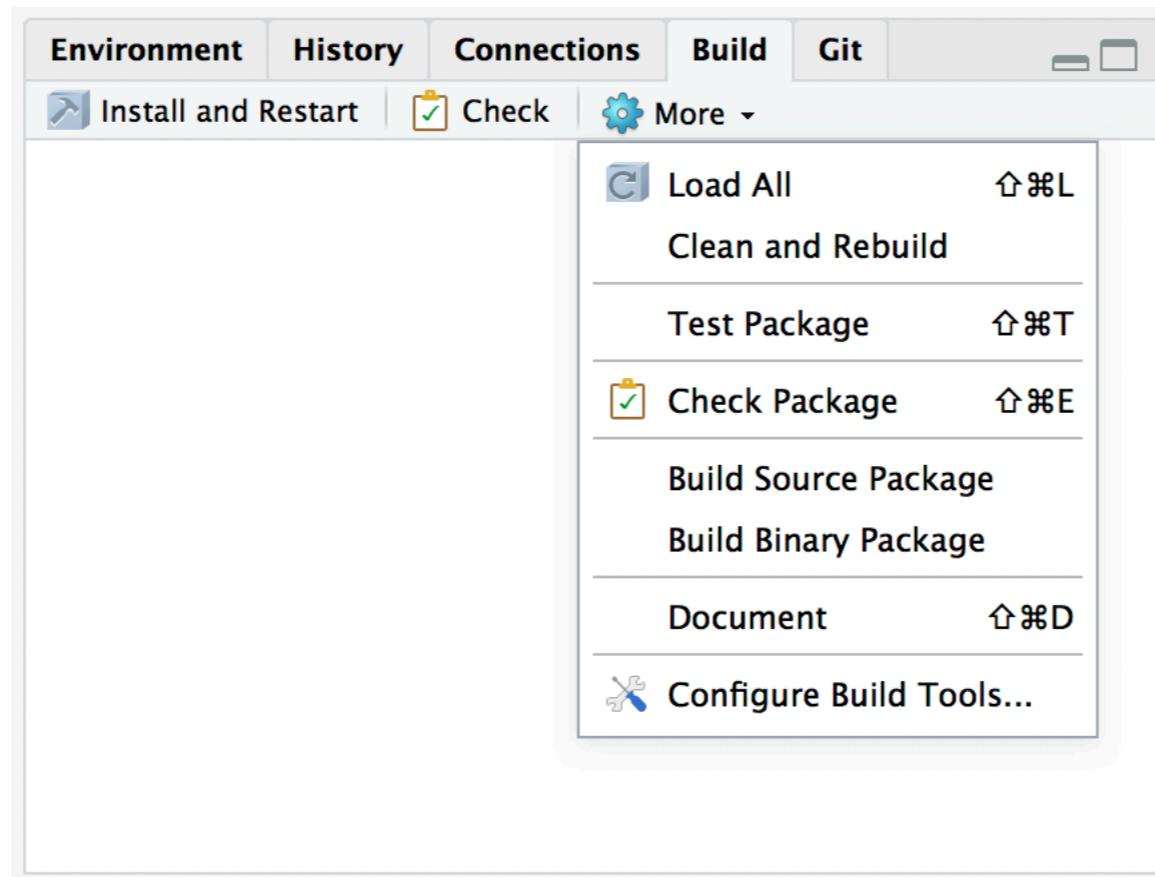
... useful *R* packages for writing *R* packages ...

```
install.packages(c("devtools", "testthat", "roxygen2"))
```



Package Build Tab

... RStudio's built in project support for packages ...



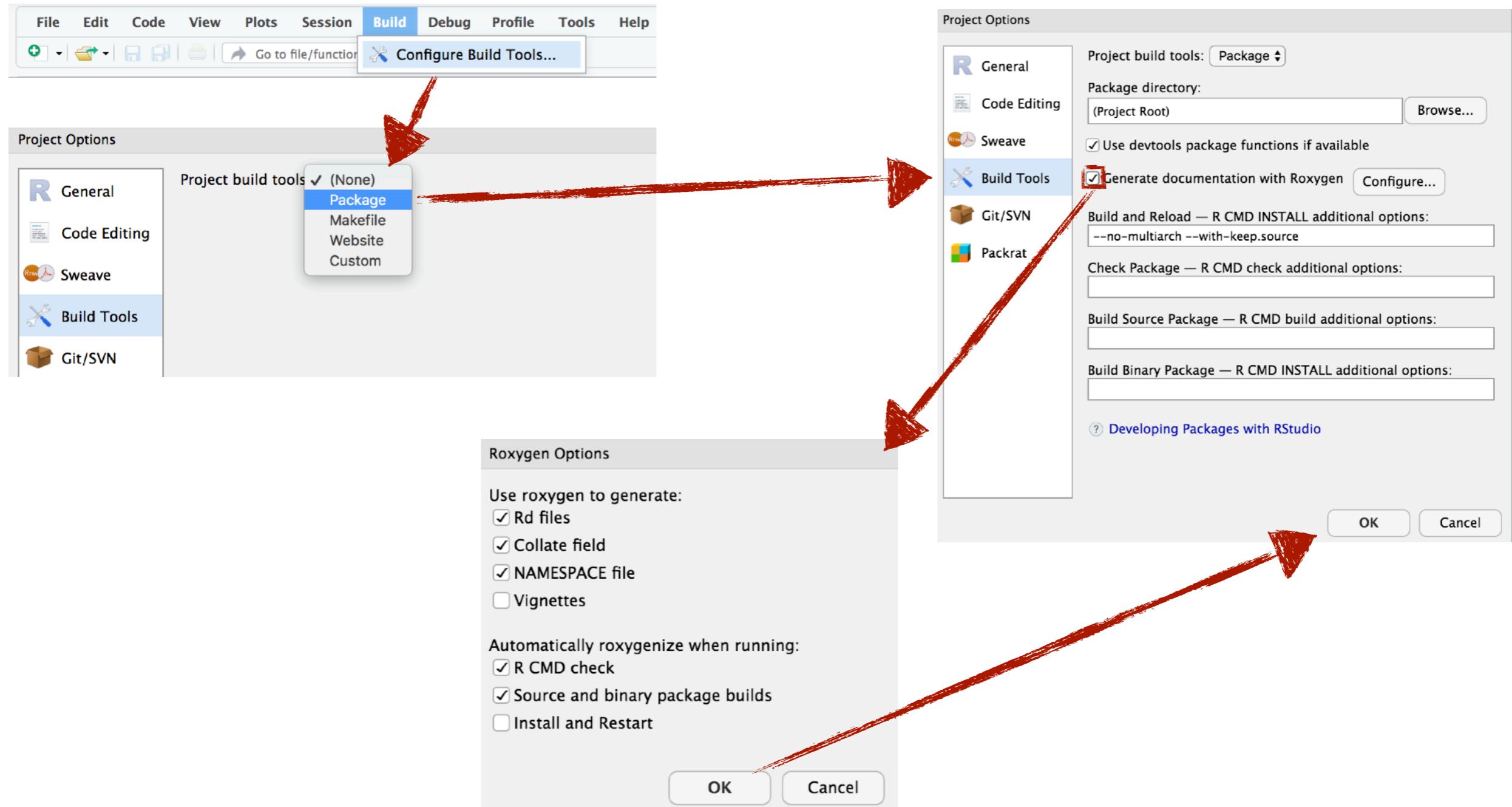
(how to enable the tab is shown next)



Studio Cloud

Creating an R Package

... in RStudio Cloud ...



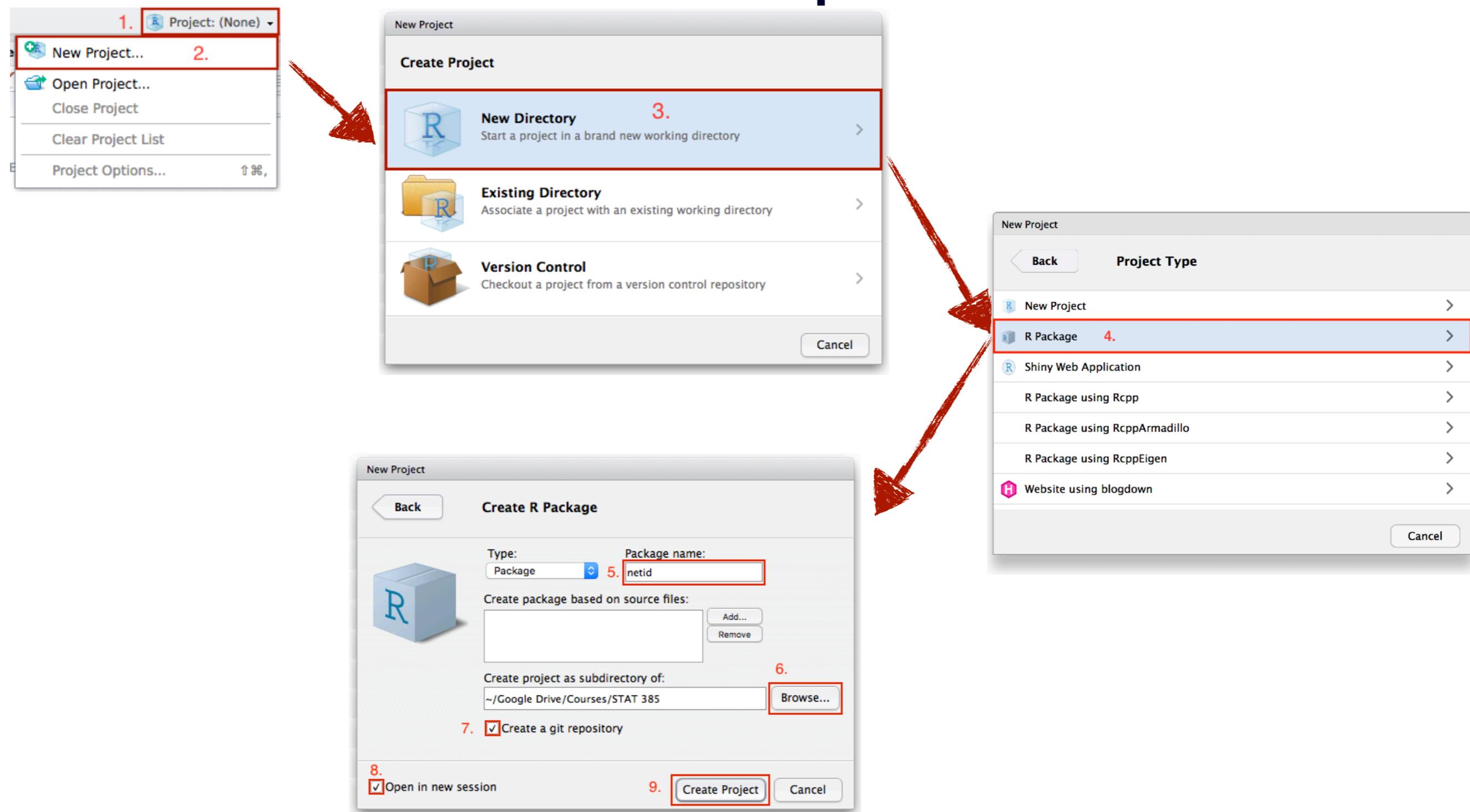
Then run in R
devtools::setup()



Desktop / Server

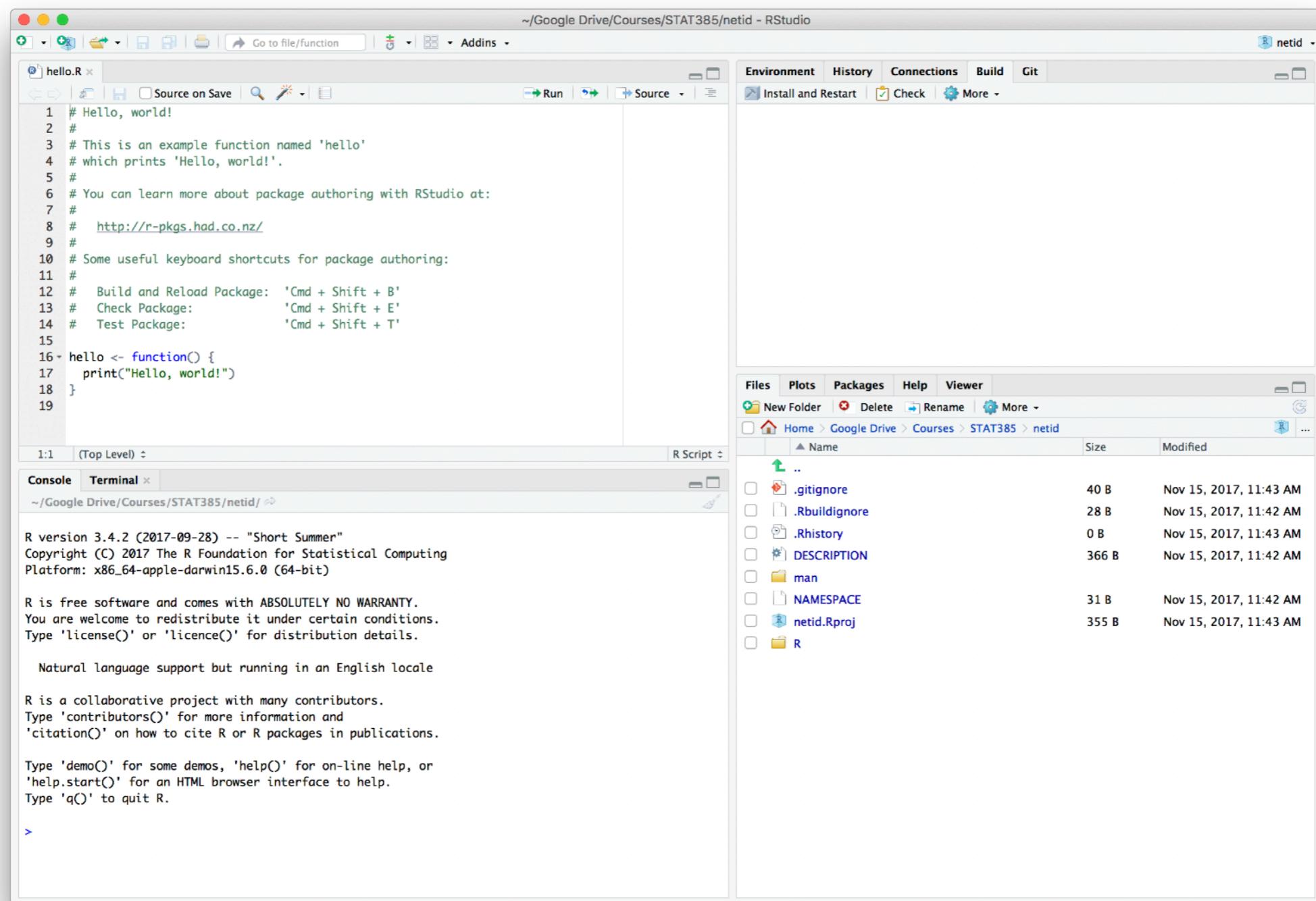
Creating an *R* Package

... in RStudio **Desktop / Server** ...



Creating an *R* Package

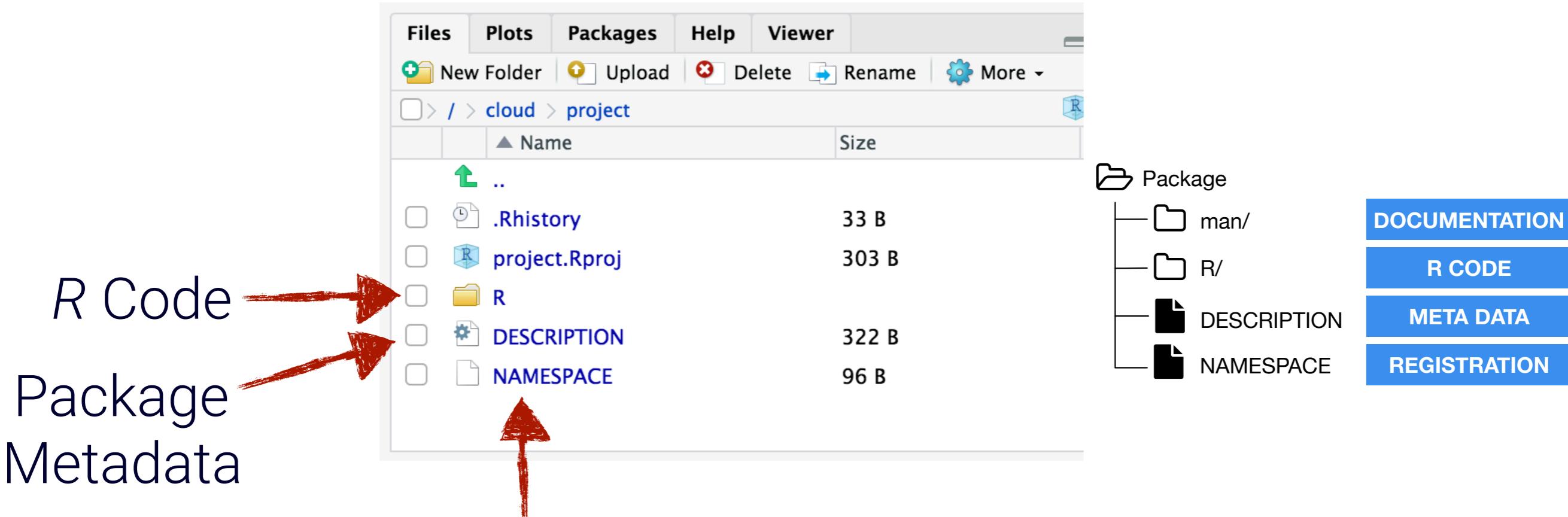
... in RStudio **Desktop / Server** ...



Structure of *R* Packages

Overview

... looking inside an R package ...



Function Exports (others can use your *R* code)
Package Imports (using others *R* code)

Example R Packages

... R & D packages with narrow scope ...

R Package Examples

Repositories 9 People 1 Teams 0 Projects 0 Settings

Pinned repositories

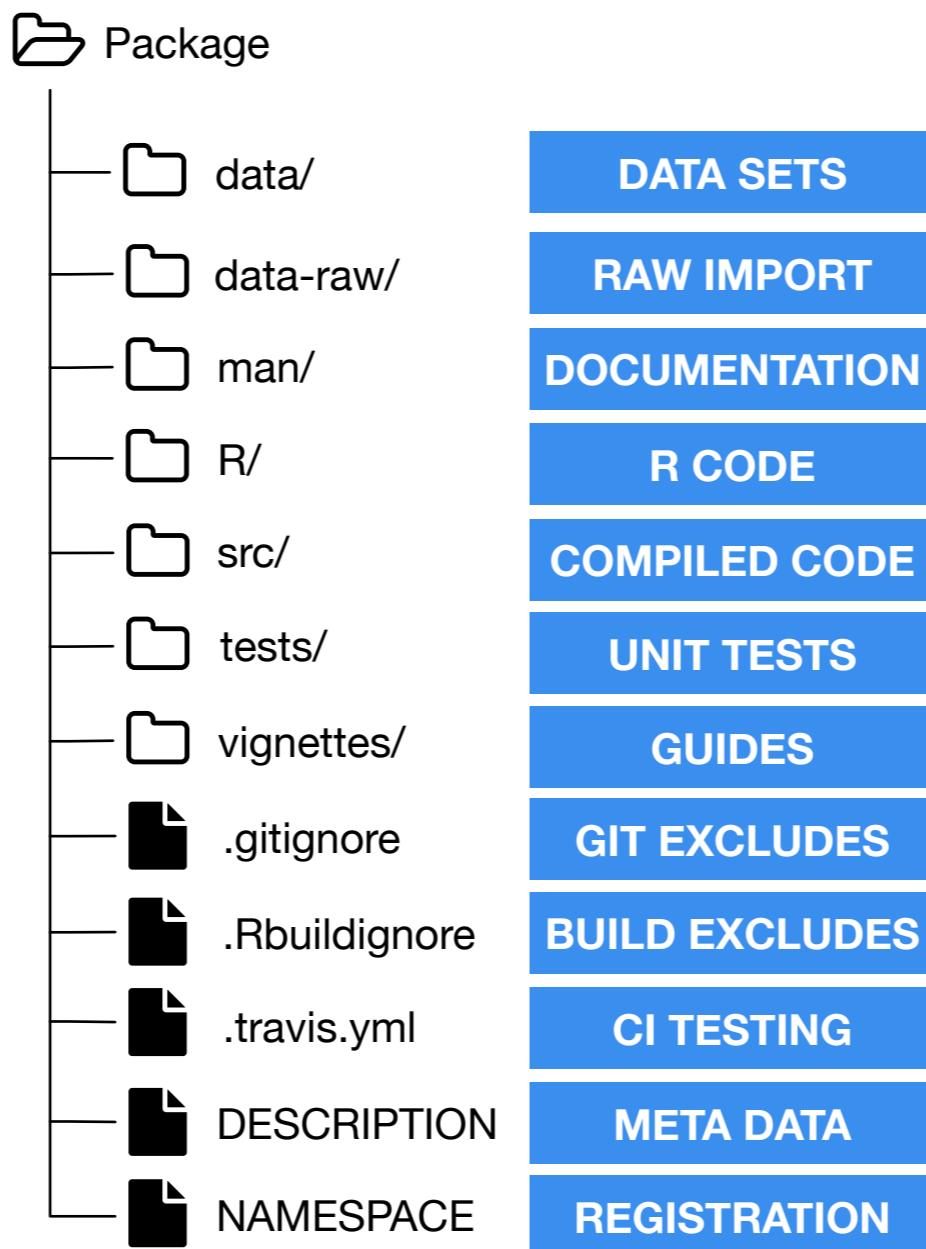
- rcpp-headers-subdirs**
Rcpp: Including C++ code in a Subdirectory within Src
C++ 4 ⚡ 1
- r-c-api-remap-printf**
R's C API Demo: Remapping C's printf() to Rprintf() and capturing console output
C 1 ⚡ 1
- rcpp-cpp11-usage**
Rcpp and Travis-CI: Compiling Code with C++11
C++
- rcpp-headers-src**
Rcpp: Using header to share C++ code within the same directory
C++

Customize pinned repositories

<https://github.com/r-pkg-examples>

All possible options

... there are many more things to add ...



Package Metadata

... editing **DESCRIPTION** ...



Four spaces
if new line
used

Name of Package



Package: **netid**

Title: Sample STAT 385 Package

Version: 0.0.1

Authors@R: c(

```
  person("First Name", "Last Name",  
         email = "netid@illinois.edu",  
         role = c("aut", "cre", "cph")),  
  person("James", "Balamuta",  
         email = "balamut2@illinois.edu",  
         role = c("aut", "cph")))
```

Description: Contained within are our first attempts at learning how to make a package in R.

Depends: R (>= 3.5.0)

License: GPL (>= 2)

Encoding: UTF-8

LazyData: true

Licensing

... choosing the right license ...

James Balamuta is not a lawyer.*

* His advice is not a substitute for consulting with a lawyer when licensing code. Furthermore, using his advice in relation to license is done at your own risk. He cannot be held liable for any code licensing issues that may arise.

Why License?

... provides clarity as to how others can use your code ...

See [Section 1.1.2: Licensing](#) within [Writing R Extensions](#)

View differences between licenses at:

[choosealicense.com](#)

[tldrlegal.com](#)

If you want to make money or do commercial endeavors, make sure to **contact a lawyer** before choosing a license!

Specifying Contributions

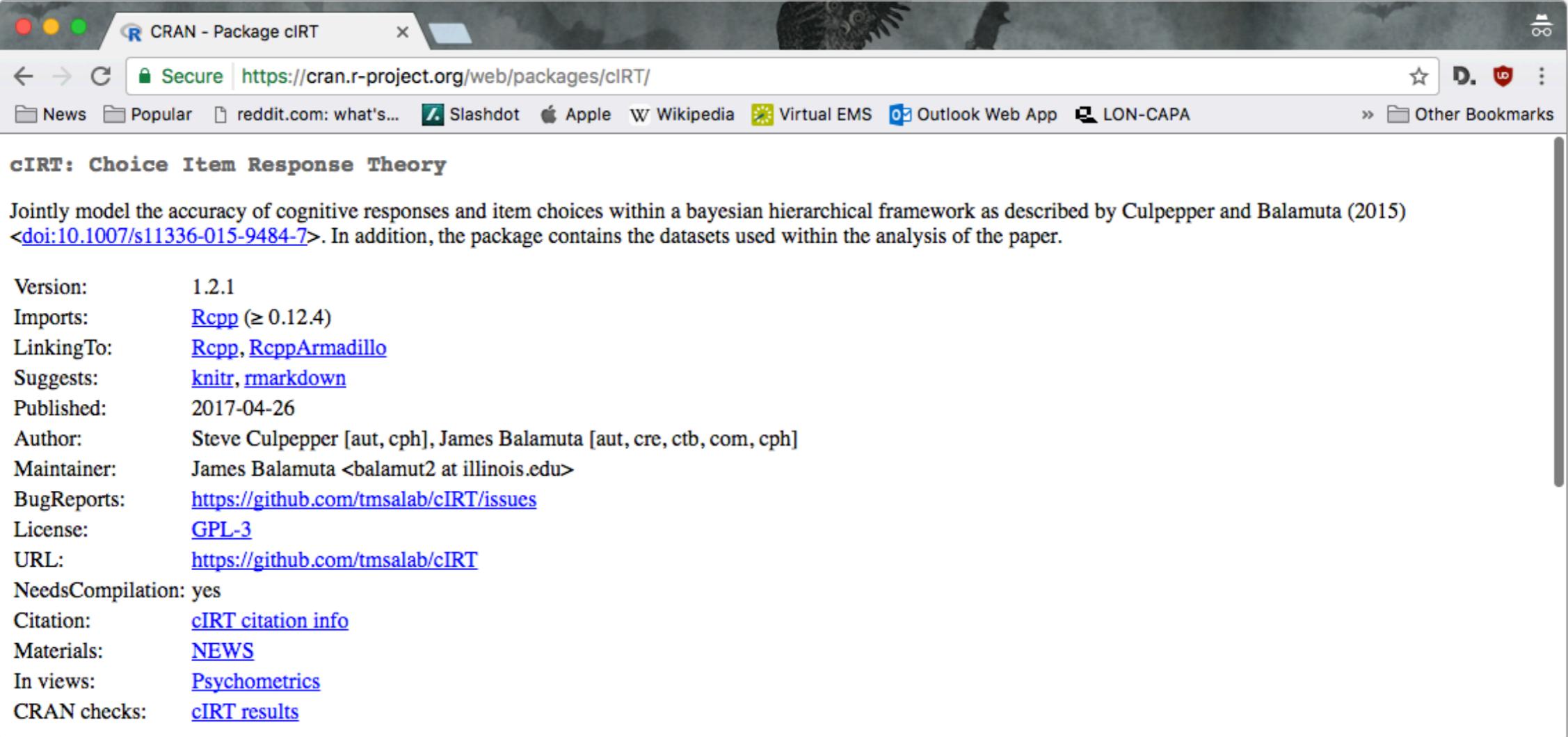
... MARC Code List for Relators ...

Role	Position	Description
aut	Author	substantial contributions to the package
ctb	Contributor	authors who have made smaller contributions
cre	Creator	the package maintainer
cph	Copyright Holder	intellectual property rights holders

For more details, see **?person**

DESCRIPTION on CRAN

... how metadata is viewed ...



The screenshot shows a web browser window with the title "CRAN - Package cIRT". The address bar indicates a secure connection to <https://cran.r-project.org/web/packages/cIRT/>. The page content is the DESCRIPTION file for the cIRT package. It includes the package name, version (1.2.1), imports (Rcpp), linking to Rcpp and RcppArmadillo, suggesting knitr and rmarkdown, and listing authors Steve Culpepper and James Balamuta, along with their contact information and GitHub repository. The page also lists the maintainer, bug reports, license (GPL-3), URL, needs compilation (yes), citation info, news, views, and CRAN checks.

cIRT: Choice Item Response Theory

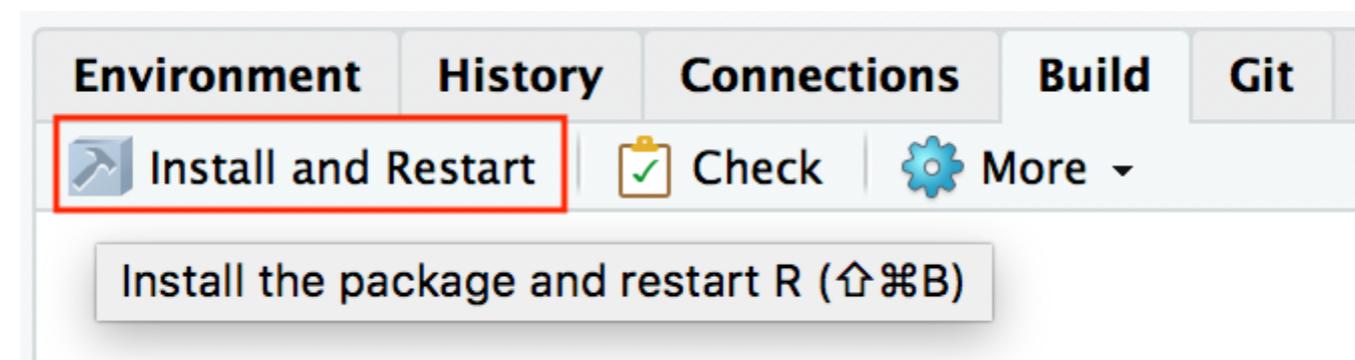
Jointly model the accuracy of cognitive responses and item choices within a bayesian hierarchical framework as described by Culpepper and Balamuta (2015) [doi:10.1007/s11336-015-9484-7](https://doi.org/10.1007/s11336-015-9484-7). In addition, the package contains the datasets used within the analysis of the paper.

Version: 1.2.1
Imports: [Rcpp](#) ($\geq 0.12.4$)
LinkingTo: [Rcpp](#), [RcppArmadillo](#)
Suggests: [knitr](#), [rmarkdown](#)
Published: 2017-04-26
Author: Steve Culpepper [aut, cph], James Balamuta [aut, cre, ctb, com, cph]
Maintainer: James Balamuta <balamut2 at illinois.edu>
BugReports: <https://github.com/tmsalab/cIRT/issues>
License: [GPL-3](#)
URL: <https://github.com/tmsalab/cIRT>
NeedsCompilation: yes
Citation: [cIRT citation info](#)
Materials: [NEWS](#)
In views: [Psychometrics](#)
CRAN checks: [cIRT results](#)

[cIRT CRAN page](#) | [DESCRIPTION file source on GitHub](#)

Build Package

... creates and loads the *R* package ...



[CNTRL / CMD + SHIFT + B]

The screenshot shows the RStudio interface with the Build tab selected. The output console displays the command: `==> R CMD INSTALL --no-multiarch --with-keep.source project`. Below it, the process of installing the "netid" package is shown, including the creation of help indices and building package indices.

```
==> R CMD INSTALL --no-multiarch --with-keep.source project
* installing to library '/home/rstudio-user/R/x86_64-pc-linux-gnu-library/3.4'
No man pages found in package 'netid'
* installing *source* package 'netid' ...
** help
*** installing help indices
** building package indices
** testing if installed package can be loaded
* DONE (netid)
```

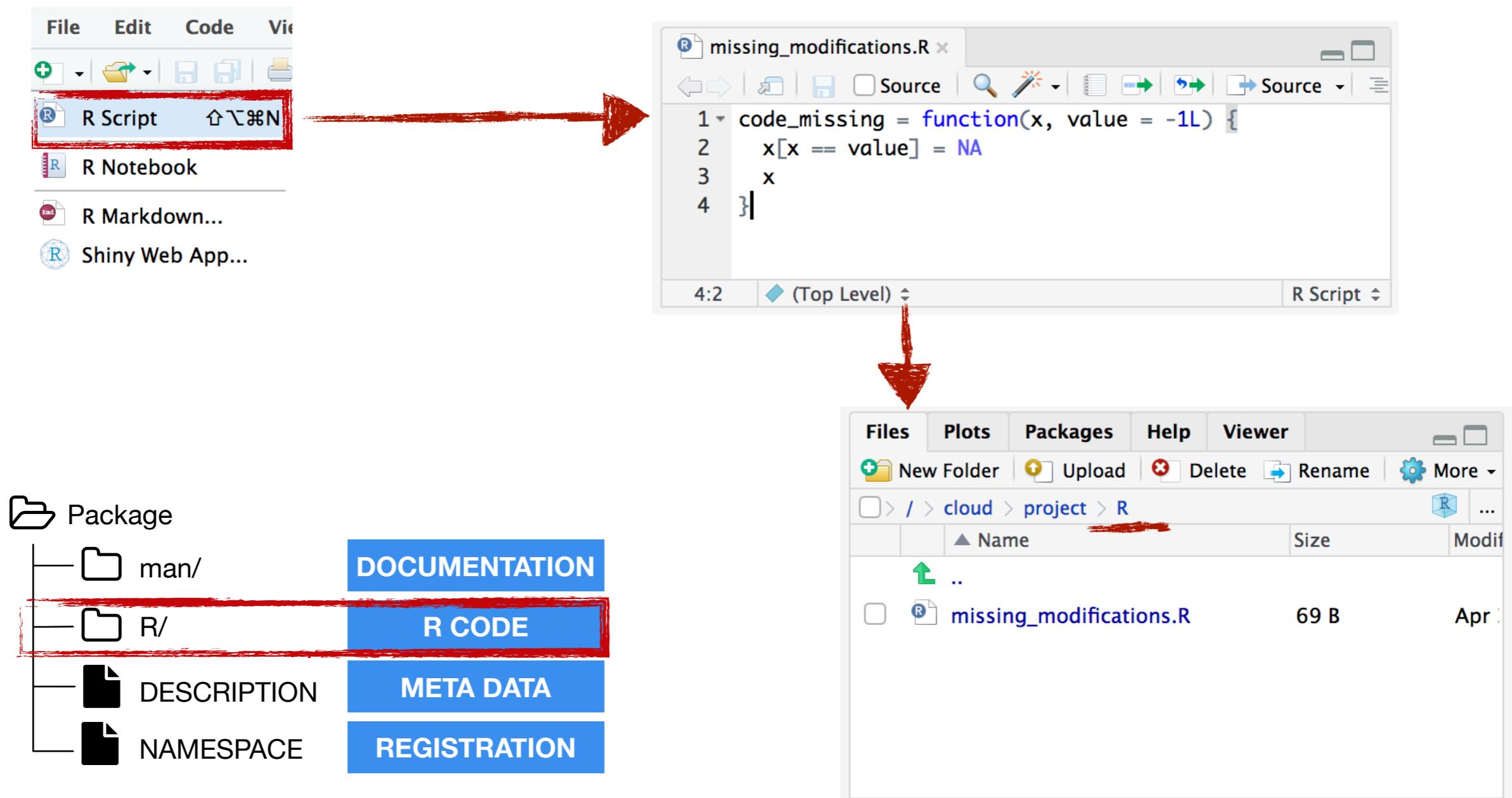
The screenshot shows the RStudio interface with the Packages tab selected. A red arrow points from the "Install and Restart" button in the top-left screenshot to this table. The table lists various R packages installed in the current session:

Name	Description	Version
magrittr	A Forward-Pipe Operator for R	1.5
memoise	Memoisation of Functions	1.1.0
mime	Map Filenames to MIME Types	0.5
netid	Sample STAT 385 Package	0.0.1
openssl	Toolkit for Encryption, Signatures and Certificates Based on OpenSSL	1.0.1
project	What the Package Does (one line, title case)	0.0.0.9000
R6	Classes with Reference Semantics	2.2.2
Rcpp	Seamless R and C++ Integration	0.12.16
roxygen2	In-Line Documentation for R	6.0.1
rprojroot	Finding Files in Project Subdirectories	1.3-2
rstudioapi	Safely Access the RStudio API	0.7
stringi	Character String Processing Facilities	1.1.7
stringr	Simple, Consistent Wrappers for Common	1.3.0

Adding an R Function

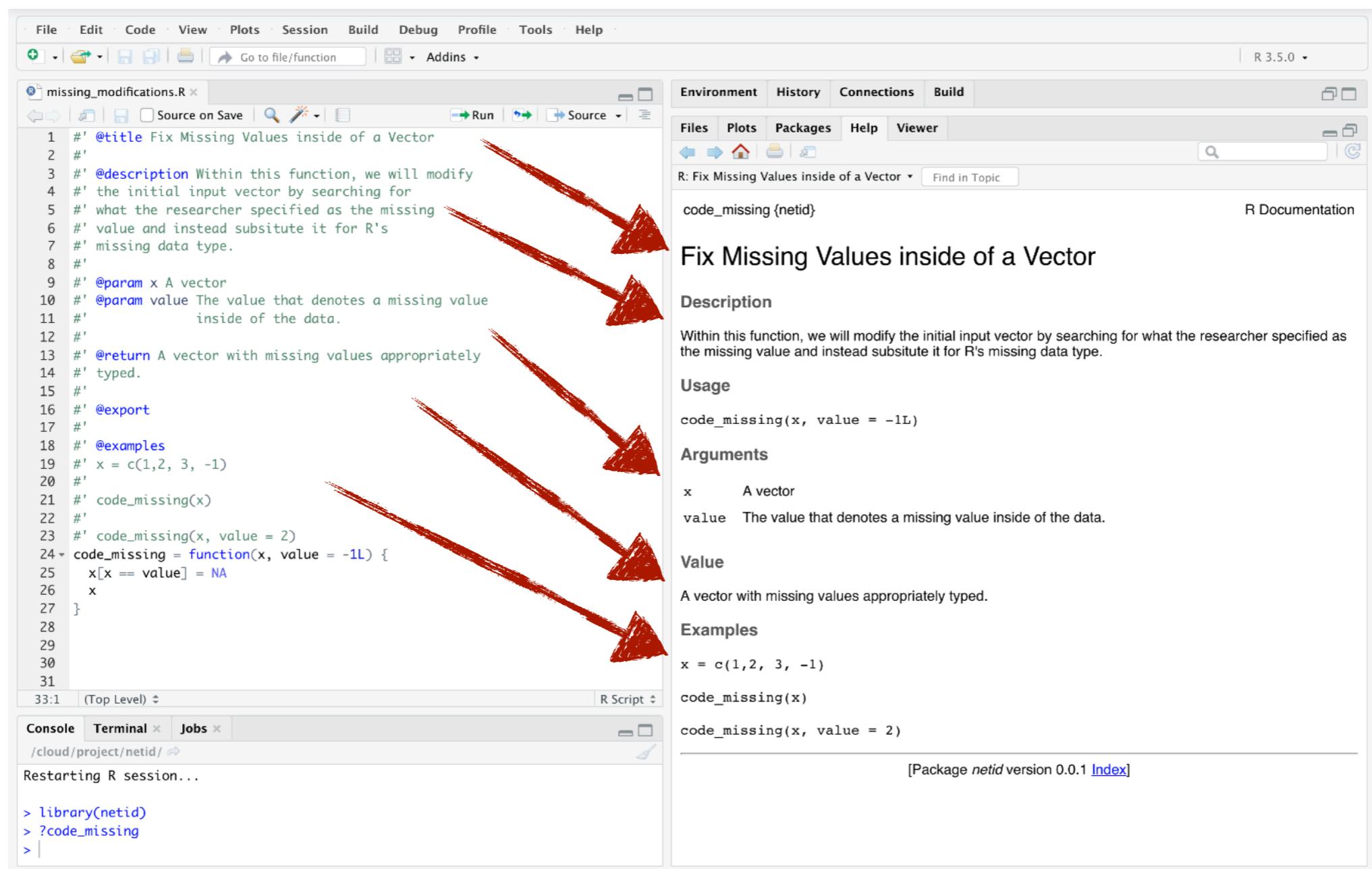
... saving a function to the **R/** directory ...

R/missing_modifications.R



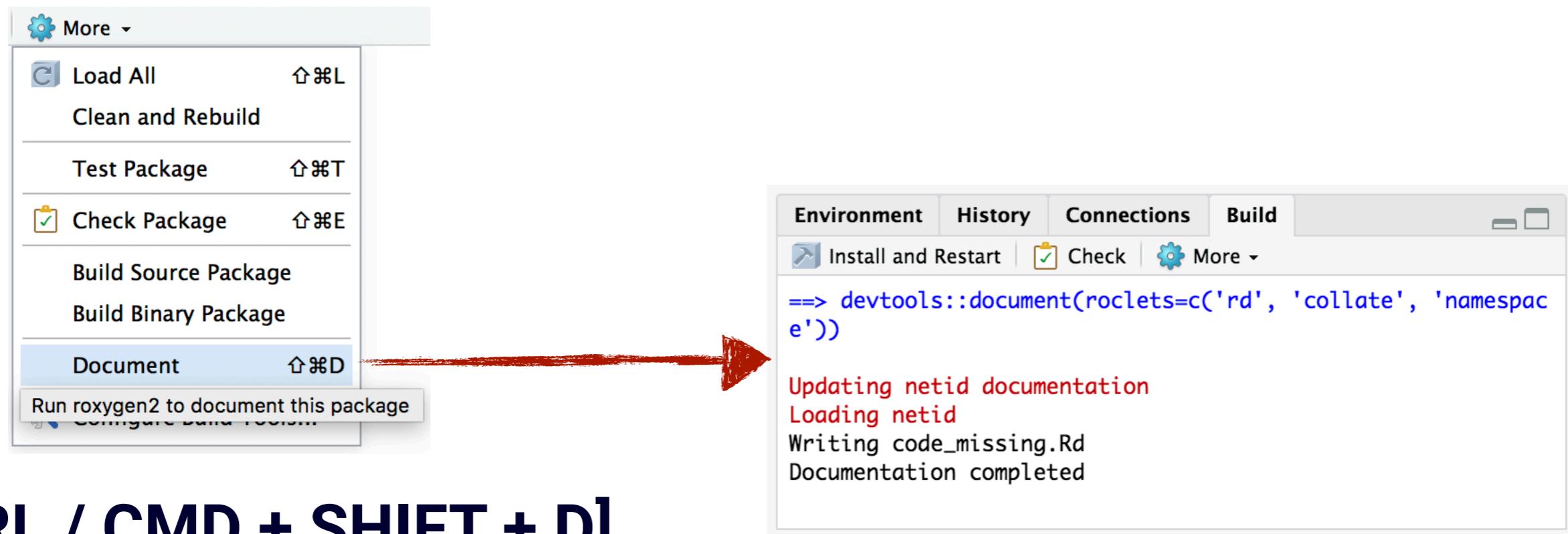
Adding Documentation

... roxygen2's inline documentation feature ...

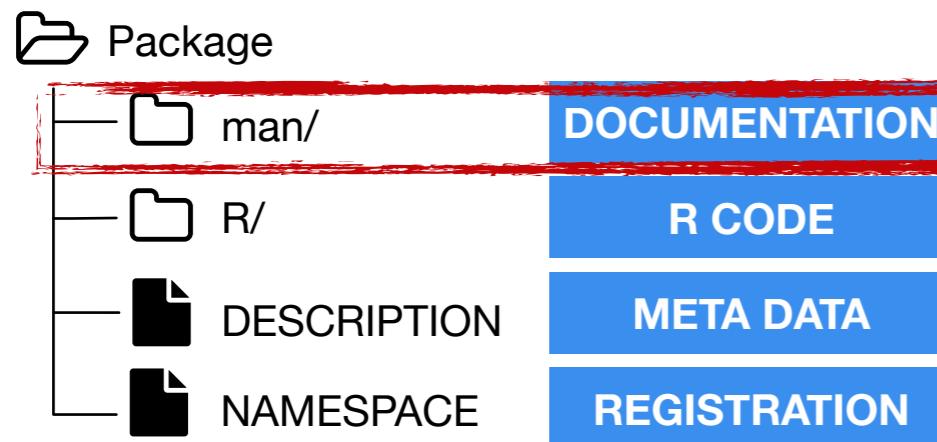


Building Documentation

... translating roxygen2 comments to Rd files ...



[CNTRL / CMD + SHIFT + D]



The screenshot shows the RStudio Cloud file browser. The current path is /cloud/project/man. The table lists files in the directory:

	Name	Size	Modi
	code_missing.Rd	536 B	Apr

Autogenerates Rd Help Files

... roxygen2 saves time because it automatically generates Rd ...

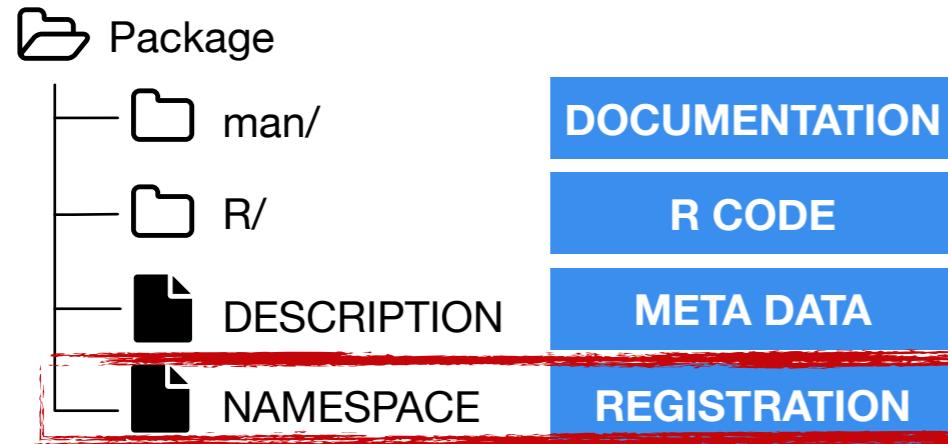
The screenshot shows the RStudio interface with several windows open:

- Source Editor:** A file named `missing_modifications.R` containing R code. The code defines a function `code_missing` that replaces missing values in a vector with a specified value.
- Rd Editor:** A file named `code_missing.Rd` showing the generated R documentation. It includes sections for `title`, `description`, `usage`, `arguments`, `value`, `description`, `examples`, and `value`.
- Environment:** Shows the current R environment with tabs for `Files`, `Plots`, `Packages`, `Help` (selected), and `Viewer`.
- Help Viewer:** A detailed view of the `code_missing` documentation, including the title "Fix Missing Values inside of a Vector", a description of the function's purpose, usage information, arguments, and examples.
- Console:** Shows the command `?code_missing` being run in the R console.

A large watermark-like text "Generated" is overlaid on the bottom left of the Rd Editor window.

Package Namespace

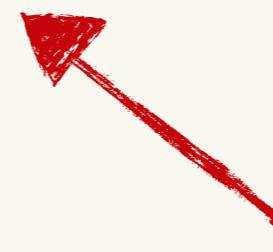
... autogenerated **NAMESPACE** ...



Recognized by **@export**
in function comments

Generated by roxygen2: do not edit by hand

export(code_missing)



export()

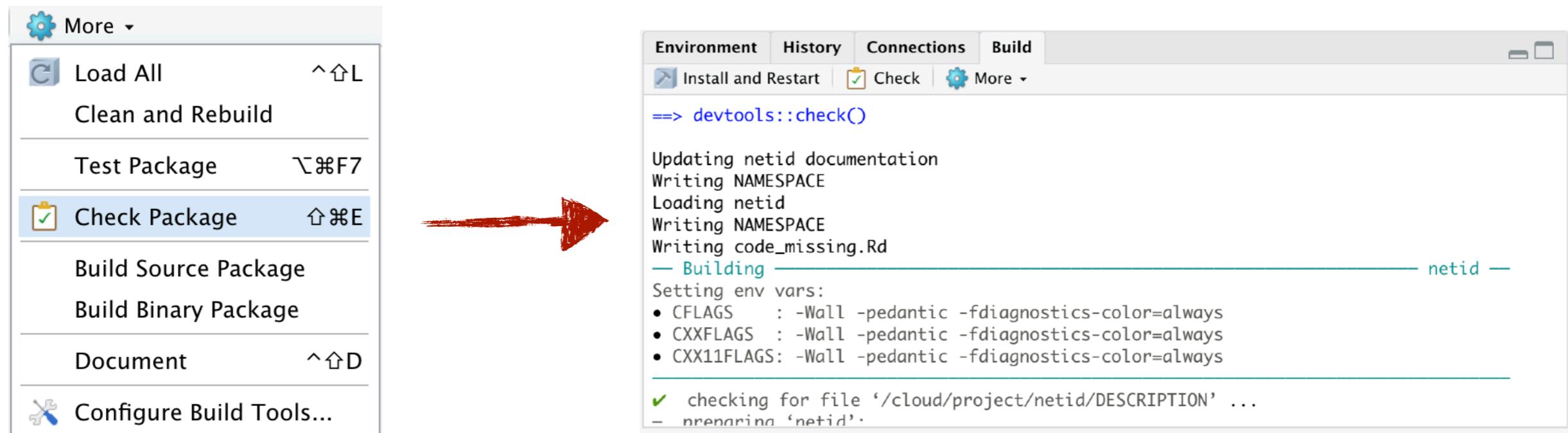
exposes the function

The screenshot shows an RStudio interface with the following code in the "missing_modifications.R" script:

```
1 #' @title Fix Missing Values inside of a Vector
2 #' 
3 #' @description Within this function, we will modify
4 #' the initial input vector by searching for
5 #' what the researcher specified as the missing
6 #' value and instead substitute it for R's
7 #' missing data type.
8 #' 
9 #' @param x A vector
10 #' @param value The value that denotes a missing value
11 #'           inside of the data.
12 #' 
13 #' @return A vector with missing values appropriately
14 #' typed.
15 #' 
16 #' @export
17 #' 
18 #' @examples
19 #' x = c(1,2, 3, -1)
20 #' 
21 #' code_missing(x)
22 #' 
23 #' code_missing(x, value = 2)
24 #' code_missing = function(x, value = -1L) {
25 #'   x[x == value] = NA
26 #'   x
27 #' }
28 #' 
29 #' 
30 #' 
```

Check Package

... verifies R package meets a minimal criterion ...



[CNTRL / CMD + SHIFT + E]

Hope

R CMD check passes...

-- R CMD check results ----- netid 0.0.1
Duration: 9.4s

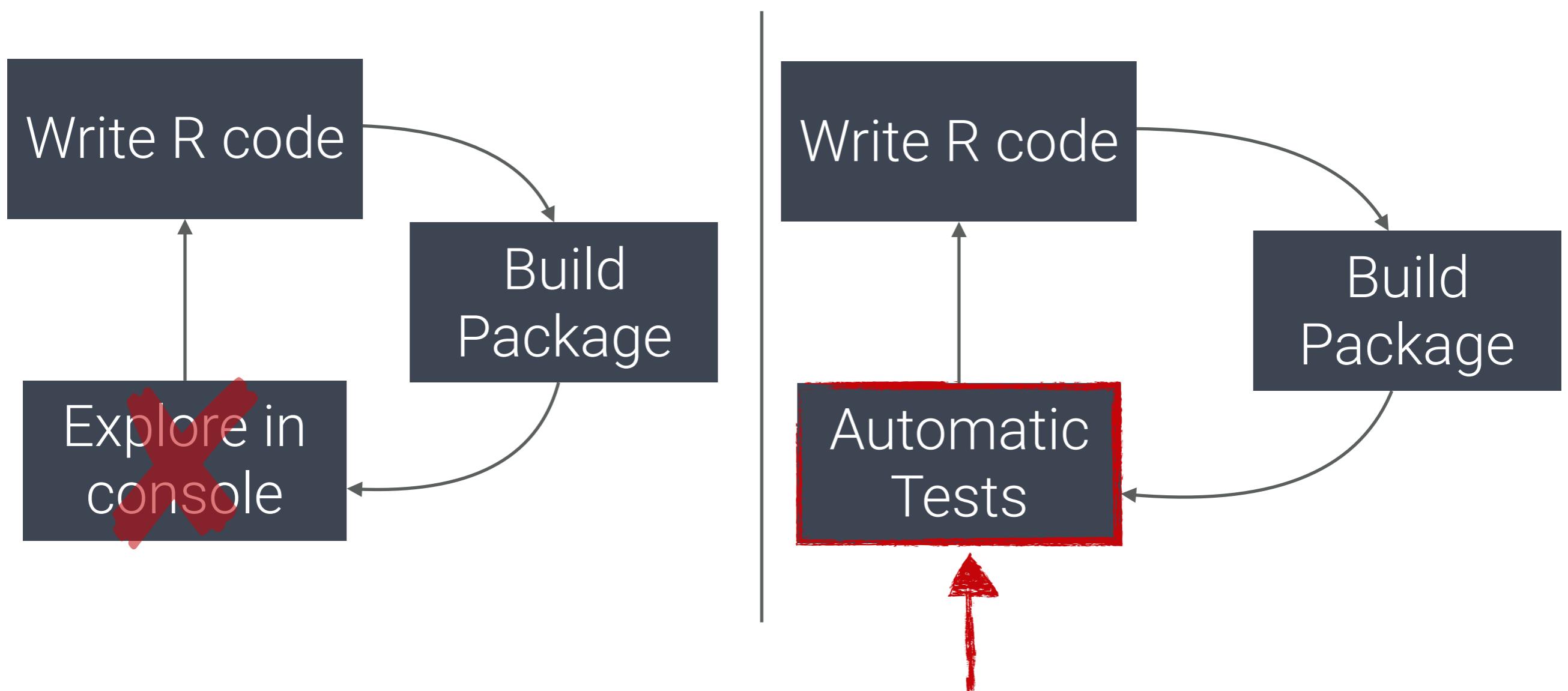
0 errors ✓ | 0 warnings ✓ | 0 notes ✓

R CMD check succeeded

Unit Testing

Testing Functions

... naive vs. unit tests ...



Definition:

Unit Testing refers to verifying the smallest parts of a piece of software, referred to as units, meet expectations set forth.

Test File

Test No NAs

Expectation 1
Expectation 2

Test NAs Present

Expectation 1

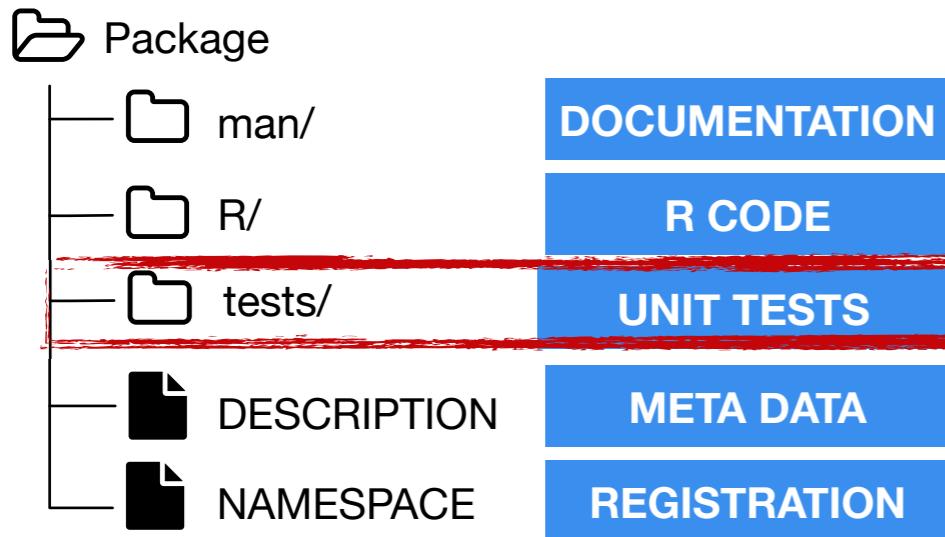


... framework for writing unit tests ...

Function	Description
<code>expect_equal()</code>	Expectation: is the object equal to a
<code>expect_true()</code> <code>expect_false()</code>	Expectation: is the object true/false?
<code>expect_output()</code> <code>expect_error()</code>	Expectation: does code produce output/ message/warning/error?
<code>expect_message()</code> <code>expect_warning()</code>	
<code>test_package()</code> <code>test_dir()</code>	Run all tests in package or directory

<http://testthat.r-lib.org/reference/index.html>

Test Example



```
# Enable testthat  
usethis::use_testthat()
```

```
# Create a test file  
usethis::use_test("missing-mods")
```

Inside:

```
## tests/testthat/  
## test-missing-mods.R
```

```
context("Missing mods")
```

```
test_that("Check coding", {  
  x = c(1, 2, 1, -1, 0)  
  expect_equal(  
    code_missing(x, -1),  
    c(1, 2, 1, NA, 0)  
  )  
})
```

Automatic Test Suites

- GitHub: Enables version control and integrates with...
- **Travis-CI**: Continuous integration
 - e.g. run code every for each commit
- **Codecov**: Coverage of Tests
 - e.g. what parts of the code are tested vs. not?

GitHub + Travis-CI

... instant breakage information ...

The image shows two screenshots illustrating the integration of GitHub and Travis-CI. On the left, the GitHub repository page for 'coatless / searcher' shows a 'Code' tab with a large green 'Passes' badge. A red arrow points from this badge to the word 'Error'd below it. On the right, the Travis-CI build history for 'Build #4' is shown, which failed. The log details a warning about 'browse_url' and ends with an error message: 'Found warnings, treating as errors (as requested)'. A red arrow points from the 'Error'd text to the error log.

coatless / searcher

Unwatch 3

Code Issues 4 Pull requests 0 Projects 0 Wiki Insights Settings

Branch: master

Commits on Jul 14, 2018

Update searcher GIF coatless committed on Jul 14, 2018

Passes

Commits on Jul 1, 2018

Release candidate sent to CRAN coatless committed on Jul 1, 2018

Add R suffix for search engines to ensure R relevant results (close #10...) coatless committed on Jul 1, 2018

Verified

Handle line endings and merge conflicts on NEWS coatless committed on Jul 1, 2018

Error'd

coatless / searcher build passing

Current Branches Build History Pull Requests Build #4 More options

X master Remove Rd listing for browse_url -o #4 failed

-o Commit abbfbf8 ↗ Ran for 4 min 38 sec

↳ Compare 54962dc..abbfbf8 ↗ 11 months ago

↳ Branch master ↗

James J Balamuta

R

Job log View config

X Remove log ↴ Raw log

worker_info system_info R-build R-check 0.54s 0.01s fail log out Top

1 Worker information
6 mode of '/usr/local/clang-5.0.0/bin' changed from 0777 (rwxrwxrwx) to 0775 (rwxrwxr-x)

7 Build system information

1482 Building package

1493 Checking package

1557 checking Rd cross-references ... WARNING

1558 Missing link or links in documentation object 'search_site.Rd':
1559 'browse_url'

1560

1561 See section 'Cross-references' in the 'Writing R Extensions' manual.

1562

1563

1564

1565 R CMD check fail logs

1568 R CMD check log logs

1626 R CMD check out logs

1767 Found warnings, treating as errors (as requested).

GitHub + Code Coverage

... what parts of code have unit tests ...

The image shows a composite screenshot of GitHub and Code Coverage tools. On the left, there's a 'Coverage Sunburst' chart with a large green segment and a smaller orange segment. Below it is a table of 'ALL RECENT COMMITS' by user 'coatless'. To the right is a code editor displaying R code for search functions, with red and green highlights indicating coverage status. A red arrow points from the GitHub commit list to the coverage table below, and another red arrow points from the code editor to the coverage table.

ALL RECENT COMMITS

- Update searcher GIF (coatless, 4 months ago, master, 85f4775, CI Passed)
- Release candidate sent to CRAN (coatless, 5 months ago, master, 3615941, CI Passed)
- Add R suffix for search engines to ensure R relevant results (close #10) (#11) (coatless, 5 months ago, master, 5e49e0e, CI Passed)
- Improve test coverage and remove default switch case due to match.arg call (coatless, 5 months ago, master, 74e0365, CI Passed)
- Handle line endings and merge conflicts on NEWS (coatless, 5 months ago, master, 8bcc19c, CI Passed)
- Add ixquick to README and roll a new point release (coatless, 5 months ago, master, 5a64865, CI Passed)

Coverage

File	Statements	Branches	Conditions	Functions	Coverage (%)
R	61	54	0	7	88.52%
Project Totals (2 files)	61	54	0	7	88.52%

Code Editor (R code)

```
search_site = function(query,
  site = c(
    "google",
    "bing",
    "stackoverflow",
    "so",
    "github",
    "gh",
    "duckduckgo",
    "ddg",
    "bitbucket",
    "bb",
    "ixquick"
  ),
  rlang = TRUE) {
  site = tolower(site)
  site = match.arg(site)
  switch(
    site,
    google = search_google(query, rlang),
    stackoverflow = ,
    # empty case carried below
    so = search_stackoverflow(query, rlang),
    github = ,
    # empty case carried below
    gh = search_github(query, rlang),
    bitbucket = ,
    # empty case carried below
    bb = search_bitbucket(query, rlang),
    bing = search_bing(query, rlang),
    duckduckgo = ,
    # empty case carried below
    ddg = search_ddg(query, rlang),
    ixquick = search_ixquick(query, rlang)
  )
}
```

Red: No test covers code
Green: Unit test evaluates code

Acknowledgements

- Hadley Wickham's book aptly title R Packages
- Writing Extensions for R

Recap

- **Motivation**
 - R packages lower the bare to sharing code
 - Provide means to easily re-use functions in other analyses
- **Creating an R Package**
 - RStudio has built in package templates
 - Possible to use **devtools::create_package()** and **package.skeleton()**
- **Structure of R Packages**
 - Packages require **NAMESPACE** + **DESCRIPTION** files **R/** + **man/** folders
- **Unit Testing**
 - Standardize testing and incorporate it into an automatic framework

This work is licensed under the
Creative Commons
Attribution-NonCommercial-
ShareAlike 4.0 International
License

