

Practical R: Packages

Abhijit Dasgupta

BIOF 339

What are packages in R?

Packages

Packages are collections of functions, and sometimes data, that are usually unified for a common purpose

*If **functions** are recipes, then **packages** are recipe books*

If you want to cook from a recipe, you first have to grab the recipe book from your shelf

*Similarly, if you want to use a function from a package, you first have to grab or activate the package in **your current R session***

This is done using the `library` function

For example,

```
library(tidyverse)
library(janitor)
```

Packages

There is another way to access functions from packages, if you're really only going to use one function from it.

The general form for this is `<package>::<function>` (note the **two** colons)

For example, if you just want to use the `clean_names` function from the **janitor** package, you can do so by

```
janitor::clean_names(dataset)
```

where `dataset` is the name of the data.frame whose column names you want to clean.

Important operational notes

Install packages once per computer

Activate a package once per R session

Never install packages inside a R Markdown file

The **pacman** package and the `pacman::p_load` function saves you a bunch of trouble by installing a package only if it doesn't exist on your computer and then activating the package. This one function removes a lot of the operational issues in installing and loading packages in R.

Where are the packages?

CRAN

CRAN is the Comprehensive R Archive Network, a network of mirrored repositories containing R packages.

Today, it really doesn't matter which of the repositories you use.

In RStudio, the default repository is **Global (CDN) - RStudio** which is a version in the cloud that typically works the fastest.

CRAN

You can install packages from CRAN using the following means:

```
install.packages("<package name>")
```

Or, if you want to be explicit, or are not using RStudio,

```
install.packages("<package name>", repos  
= "<repository URL>")
```

.pull-right[Using the RStudio *Packages* panel

(see next slide)

You can find packages using CRAN [Task Views](#)

Go to file/function

Addins

_output.yml

projects.Rmd

packages.Rmd*

02-DataStructures.Rmd

factors.Rmd

Knit

Insert

Run

74

75 Today, it really doesn't matter which of the repositories you use.

76

77 In RStudio, the default repository is ****Global (CDN) - RStudio**** which is a version in the cloud that typically works the fastest.

78

79 `![:scale 50%](../img/pkg1.png)`

80

81 ---

82

83 ## CRAN

84

85 You can install packages from CRAN using the follow

86

87 `.pull-left[`

88 ``install.packages("<package name>")``

89

90 Or, if you want to be explicit, or are not using RS

91

92 ``install.packages("<package name>", repos = "<repos`

93

94 `]`

95 `.pull-right[`

96 Using the RStudio `_Packages_` panel

97

98

98:1 # CRAN

Install Packages

Install from: [Configuring Repositories](#)

Repository (CRAN)

Packages (separate multiple with space or comma):

tidy

tidybayes

tidyBF

tidyboot

tidycat

tidycensus

tidycode

tidycomm

tidydice

tidyfast

tidyfst

tidyft

tidygapminder

tidygate

tidygenomics

tidygeocoder

tidygeoRSS

tidygraph

tidyHeatmap

tidyhydat

tidyjson

tidylo

Install

Cancel

Console

Terminal

~/ARAASTAT/Teaching/BIOF339/

output file: packages.knit.md

/Applications/RStudio.app/Contents/MacOS/pandoc/pandoc +RTS -K512m -RTS packages.utf8.md --to html4 --from markdown+autolink_bare_uris+tex_math_single_backslash --output package s.html --email-obfuscation none -V 'mathjax-url=https://mathjax.rstudio.com/latest/MathJa x.js?config=TeX-MML-AM_CHTML' -V 'title-slide-class=center, middle, inverse, title-slide' --standalone --section-divs --template /Library/Frameworks/R.framework/Versions/3.6/Resou rces/library/xaringan/rmarkdown/templates/xaringan/resources/default.html --no-highlight - -css ../robot.css --css ../robot-fonts.css --css ../sfah.css --include-in-header /var/fol ders/k4/xvcmx4yx76xdb141zy3hq8rc0000gn/T//RtmpLUdoPY/rmarkdown-strc7f13f79dae8.html --lua- filter /Library/Frameworks/R.framework/Versions/3.6/Resources/library/rmarkdown/rmd/Lua/pa gebreak.lua --lua-filter /Library/Frameworks/R.framework/Versions/3.6/Resources/library/rm arkdwn/rmd/Lua/latex-div.lua --include-before-body /var/folders/k4/xvcmx4yx76xdb141zy3hq8 rc0000gn/T//RtmpLUdoPY/xaringanc7f13217bfbb.md --include-after-body /var/folders/k4/xvcmx 4yx76xdb141zy3hq8rc0000gn/T//RtmpLUdoPY/xaringanc7f16da22c80.js --variable title-slide=tr ue --variable math=true

Output created: packages.html

> |

Packages

Help

Viewer

Install

Update

Search

System Library

	Name	Description	Version	
<input type="checkbox"/>	abhiR	What the Package Does (Title Case)	0.1.0	
<input type="checkbox"/>	abind	Combine Multidimensional Arrays	1.4-5	
<input type="checkbox"/>	acepack	ACE and AVAS for Selecting Multiple Regression Transformations	1.4.1	
<input type="checkbox"/>	acs	Download, Manipulate, and Present American Community Survey and Decennial Data from the US Census	2.1.4	
<input type="checkbox"/>	addinslist	Discover and Install Useful RStudio Addins	0.3	
<input type="checkbox"/>	ade4	Analysis of Ecological Data: Exploratory and Euclidean Methods in Environmental Sciences	1.7-15	
<input type="checkbox"/>	afex	Analysis of Factorial Experiments	0.27-2	
<input type="checkbox"/>	airway	RangedSummarizedExperiment for RNA-Seq in airway smooth muscle cells, by Himes et al PLoS One 2014	1.4.0	
<input type="checkbox"/>	ALL	A data package	1.26.0	
<input type="checkbox"/>	annotate	Annotation for microarrays	1.62.0	
<input type="checkbox"/>	AnnotationDbi	Manipulation of SQLite-based annotations in Bioconductor	1.46.1	
<input type="checkbox"/>	AnnotationFilter	Facilities for Filtering Bioconductor Annotation Resources	1.8.0	
<input type="checkbox"/>	AnnotationHub	Client to access AnnotationHub resources	2.16.1	

Environment

History

Connections

Build

Git

Tutorial

Build All

More

cp make_subdirs slides/lectures/week1/Makefile; make -C slides/lectures/week1

Makefile:8: *** missing separator. Stop.

make: *** [slides/lectures/week1] Error 2

Exited with status 2.

GitHub

GitHub is where many R packages reside during development.

To install a package directly from GitHub, you need the **remotes** package, and then you can use

```
remotes::install_github("<owner>/<repo>")
```

For example, if you want to install the development version of **dplyr**:

```
remotes::install_github("tidyverse/dplyr")
```

Bioconductor

The [Bioconductor](#) is a R organization dedicated to bioinformatics. It has its own repository of over 1900 packages

To install Bioconductor packages, you first need to install the **BiocManager** package from CRAN (note the upper and lower case letters). Then you can install packages by

```
BiocManager::install('<package name>')
```

For example, if you want to install the **DESeq2** package that computes differential gene expressions:

```
BiocManager::install('DESeq2')
```

Installing packages, a summary

From CRAN

```
install.packages("tidyverse")
```

From Bioconductor

```
install.packages("BiocManager") # do once  
BiocManager::install('limma')
```

From GitHub

```
install.packages('remotes') # do once  
remotes::install_github("rstudio/rmarkdown")  
# usual format is username/package name
```

GitHub often hosts development version of packages published on CRAN or Bioconductor

Both CRAN and Bioconductor have stringent checks to make sure packages can run properly, with no obvious program flaws. There are typically no guarantees about analytic or theoretical correctness, but most packages have been crowd-validated and there are several reliable developer groups including RStudio

Packages commonly used

An incomplete listing

Data ingestion

Package	Description
readr	Read Rectangular Text Data
readxl	Read Excel Files
haven	Import and Export 'SPSS', 'Stata' and 'SAS' Files
DBI	R Database Interface
rvest	Easily Harvest (Scrape) Web Pages
jsonlite	A Simple and Robust JSON Parser and Generator for R

Data munging

Package	Description
tidyr	Tidy Messy Data
dplyr	A Grammar of Data Manipulation
stringr	Simple, Consistent Wrappers for Common String Operations
lubridate	Make Dealing with Dates a Little Easier
forcats	Tools for Working with Categorical Variables (Factors)
purrr	Functional Programming Tools
janitor	Simple Tools for Examining and Cleaning Dirty Data

Data visualization

Package	Description
ggplot2	Create Elegant Data Visualisations Using the Grammar of Graphics
lattice	Trellis Graphics for R
visdat	Preliminary Visualisation of Data
naniar	Data Structures, Summaries, and Visualisations for Missing Data
htmlwidgets	HTML Widgets for R
leaflet	Create Interactive Web Maps with the JavaScript 'Leaflet' Library
highcharter	A Wrapper for the 'Highcharts' Library
plotly	Create Interactive Web Graphics via 'plotly.js'

There is an entire package ecosystem around `ggplot2` that can be seen [here](#). These include specialized plots, different themes and colors, animations, etc.

Statistics

Data description

Package	Description
tableone	Create 'Table 1' to Describe Baseline Characteristics with or without Propensity Score Weights
table1	Tables of Descriptive Statistics in HTML
stargazer	Well-Formatted Regression and Summary Statistics Tables
arsenal	An Arsenal of 'R' Functions for Large-Scale Statistical Summaries
gtsummary	Presentation-Ready Data Summary and Analytic Result Tables
flextable	Functions for Tabular Reporting
Hmisc	Harrell Miscellaneous

Statistics

Analysis

Package	Description
stats	The R Stats Package
survival	Survival Analysis
infer	Tidy Statistical Inference
rsample	General Resampling Infrastructure
broom	Convert Statistical Objects into Tidy Tibbles
finalfit	Quickly Create Elegant Regression Results Tables and Plots when Modelling

Statistical modeling

Package	Description
stats	The R Stats Package
survival	Survival Analysis
recipes	Preprocessing Tools to Create Design Matrices
rms	Regression Modeling Strategies
broom	Convert Statistical Objects into Tidy Tibbles
rsample	General Resampling Infrastructure

Machine Learning

Package	Description
caret	Classification and Regression Training
parsnip	A Common API to Modeling and Analysis Functions
yardstick	Tidy Characterizations of Model Performance
rpart	Recursive Partitioning and Regression Trees
party	A Laboratory for Recursive Partytioning
randomForest	Breiman and Cutler's Random Forests for Classification and Regression
baguette	Efficient Model Functions for Bagging
kernlab	Kernel-Based Machine Learning Lab
earth	Multivariate Adaptive Regression Splines

Reporting

```
There is a binary version available but the source version is later:
  binary source needs_compilation
articles 0.16 0.17 FALSE

The downloaded binary packages are in
  /var/folders/k4/xvcmx4yx76xdbl41zy3hq8rc0000gn/T//RtmpcobB10/downloaded_packages

The downloaded binary packages are in
  /var/folders/k4/xvcmx4yx76xdbl41zy3hq8rc0000gn/T//RtmpcobB10/downloaded_packages
```

Package	Description
rmarkdown	Dynamic Documents for R
knitr	A General-Purpose Package for Dynamic Report Generation in R
bookdown	Authoring Books and Technical Documents with R Markdown
distill	'R Markdown' Format for Scientific and Technical Writing
rticles	Article Formats for R Markdown
blogdown	Create Blogs and Websites with R Markdown
flexdashboard	R Markdown Format for Flexible Dashboards
shiny	Web Application Framework for R