

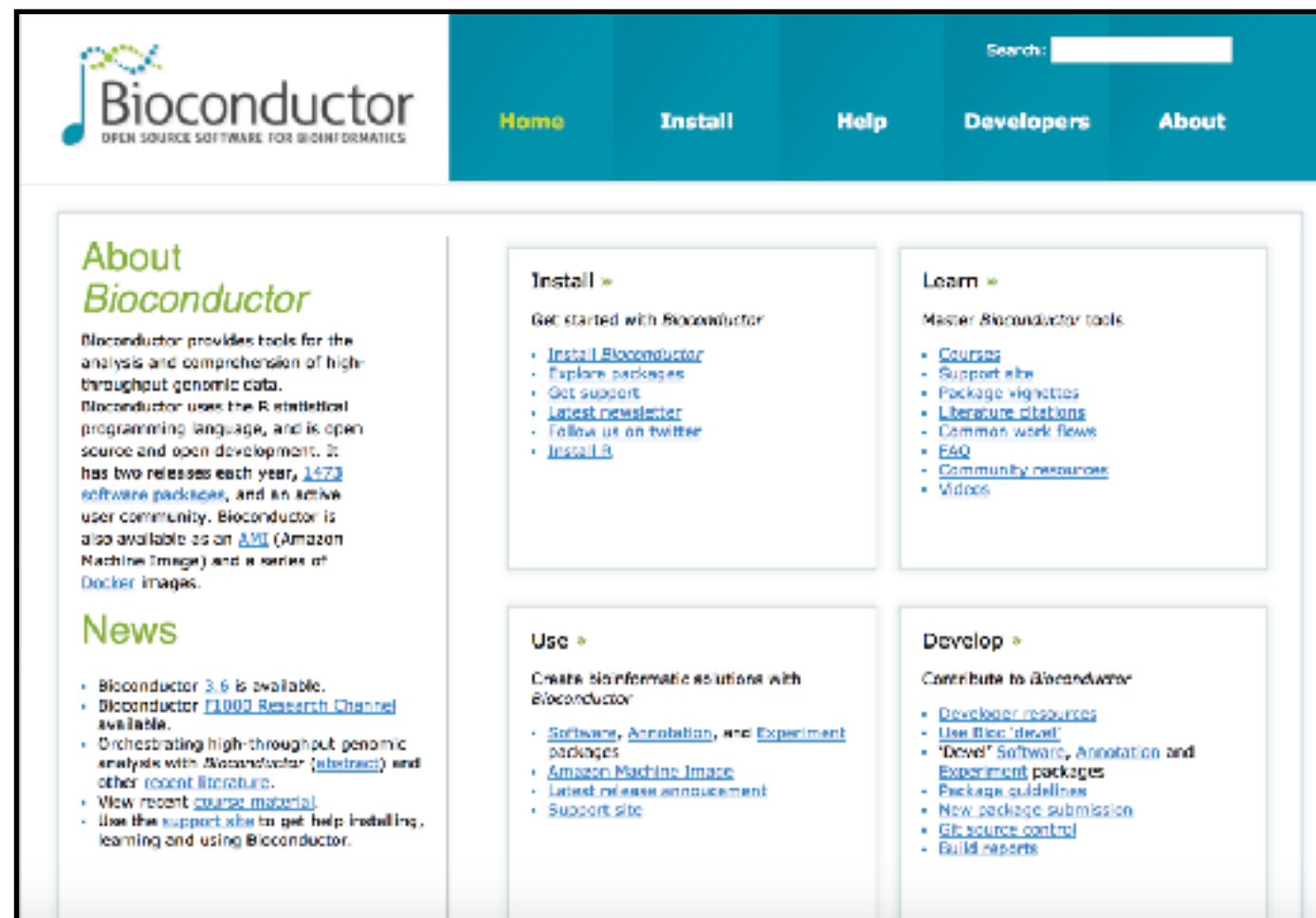
# Using Bioconductor Packages

R for Beginners 2



# What is Bioconductor?

- <https://www.bioconductor.org/>
- Provides tools for the analysis and comprehension of high-throughput genomic data
  - Sequencing: RNASeq, ChIPSeq, variants, copy number, ...
  - Microarrays: expression, SNP, ...
  - Flow cytometry, proteomics, images, ...



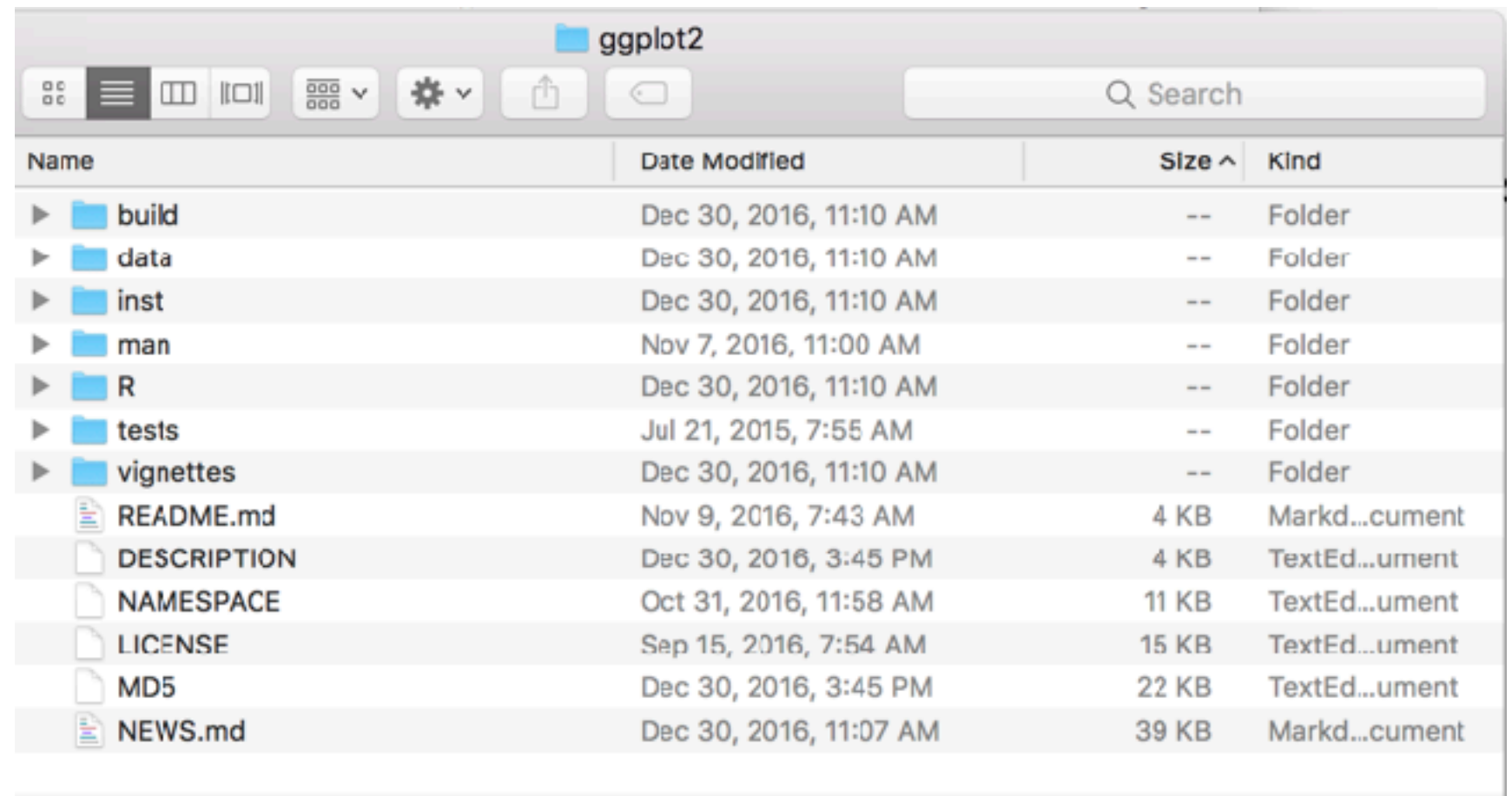
# What is Bioconductor (cont.)?

- Distributes as R Packages Open Source and open development
- Started in 2001
- Has 2 releases each year (every 6 months)
- As of Jan 2018 Bioconductor version 3.6 (Release)
  - **1473** software packages
  - **908** annotation packages
  - **325** experiment packages

# What is an R Package?

- Packages are the fundamental units of reproducible R code
  - Improve existing base R functionalities, or by adding new ones.
  - Include R functions, documentation that describes how to use them and sample data
    - R Functions
    - Documentation
    - Sample data

## ***Package Folder***



The screenshot shows a file explorer window for a folder named 'ggplot2'. The window displays a list of files and folders with columns for Name, Date Modified, Size, and Kind. The files and folders are as follows:

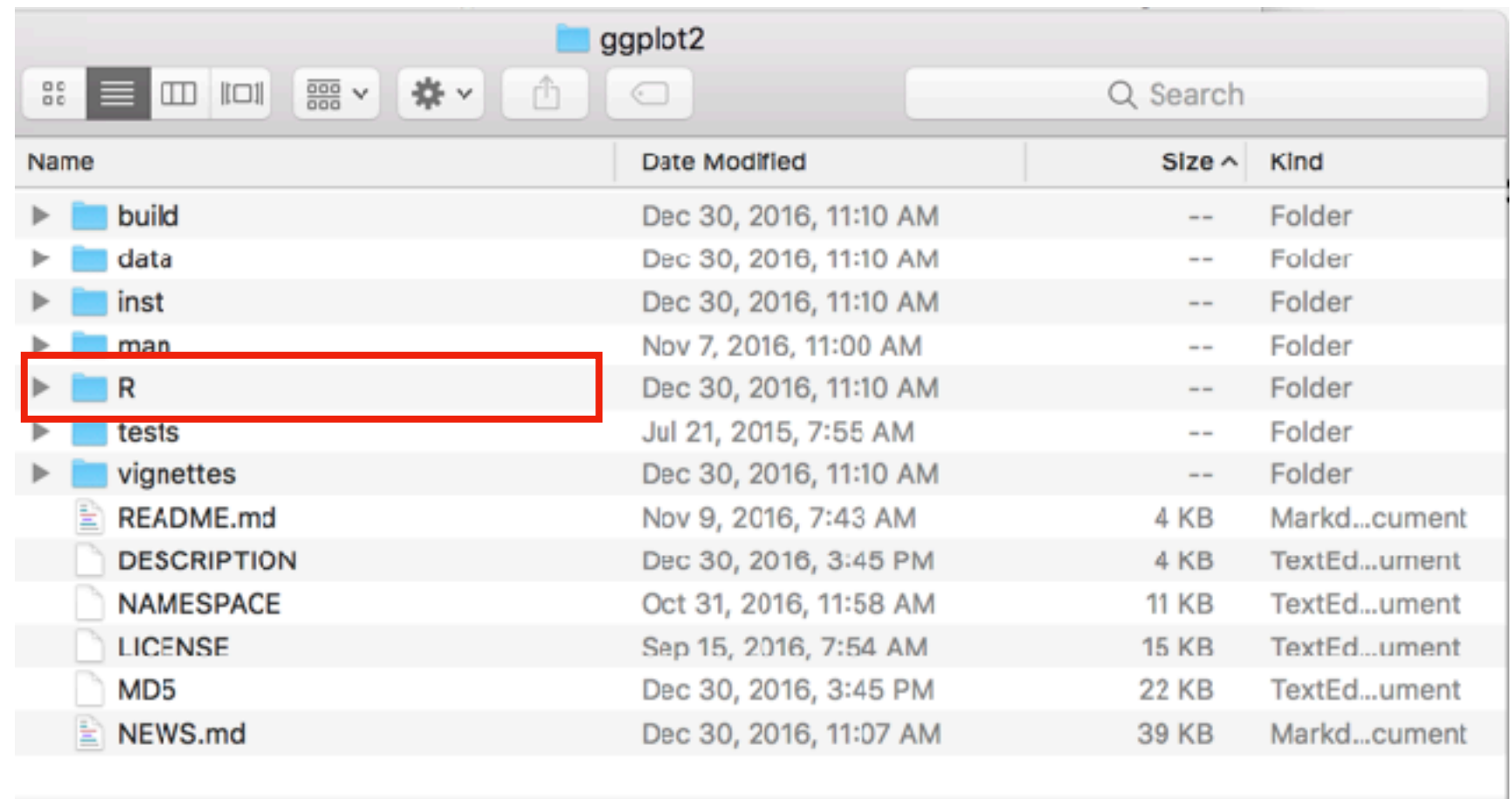
Name	Date Modified	Size	Kind
▶ build	Dec 30, 2016, 11:10 AM	--	Folder
▶ data	Dec 30, 2016, 11:10 AM	--	Folder
▶ inst	Dec 30, 2016, 11:10 AM	--	Folder
▶ man	Nov 7, 2016, 11:00 AM	--	Folder
▶ R	Dec 30, 2016, 11:10 AM	--	Folder
▶ tests	Jul 21, 2015, 7:55 AM	--	Folder
▶ vignettes	Dec 30, 2016, 11:10 AM	--	Folder
README.md	Nov 9, 2016, 7:43 AM	4 KB	Markd...cument
DESCRIPTION	Dec 30, 2016, 3:45 PM	4 KB	TextEd...ument
NAMESPACE	Oct 31, 2016, 11:58 AM	11 KB	TextEd...ument
LICENSE	Sep 15, 2016, 7:54 AM	15 KB	TextEd...ument
MD5	Dec 30, 2016, 3:45 PM	22 KB	TextEd...ument
NEWS.md	Dec 30, 2016, 11:07 AM	39 KB	Markd...cument

# What is an R Package?

- Packages are the fundamental units of reproducible R code
  - Improve existing base R functionalities, or by adding new ones.
  - Include R functions, documentation that describes how to use them and sample data
    - **R Functions**
    - Documentation
    - Sample data

## ***Package Folder***

- ***R Folder***



Name	Date Modified	Size ^	Kind
▶ build	Dec 30, 2016, 11:10 AM	--	Folder
▶ data	Dec 30, 2016, 11:10 AM	--	Folder
▶ inst	Dec 30, 2016, 11:10 AM	--	Folder
▶ man	Nov 7, 2016, 11:00 AM	--	Folder
▶ <b>R</b>	Dec 30, 2016, 11:10 AM	--	Folder
▶ tests	Jul 21, 2015, 7:55 AM	--	Folder
▶ vignettes	Dec 30, 2016, 11:10 AM	--	Folder
README.md	Nov 9, 2016, 7:43 AM	4 KB	Markd...cument
DESCRIPTION	Dec 30, 2016, 3:45 PM	4 KB	TextEd...ument
NAMESPACE	Oct 31, 2016, 11:58 AM	11 KB	TextEd...ument
LICENSE	Sep 15, 2016, 7:54 AM	15 KB	TextEd...ument
MD5	Dec 30, 2016, 3:45 PM	22 KB	TextEd...ument
NEWS.md	Dec 30, 2016, 11:07 AM	39 KB	Markd...cument

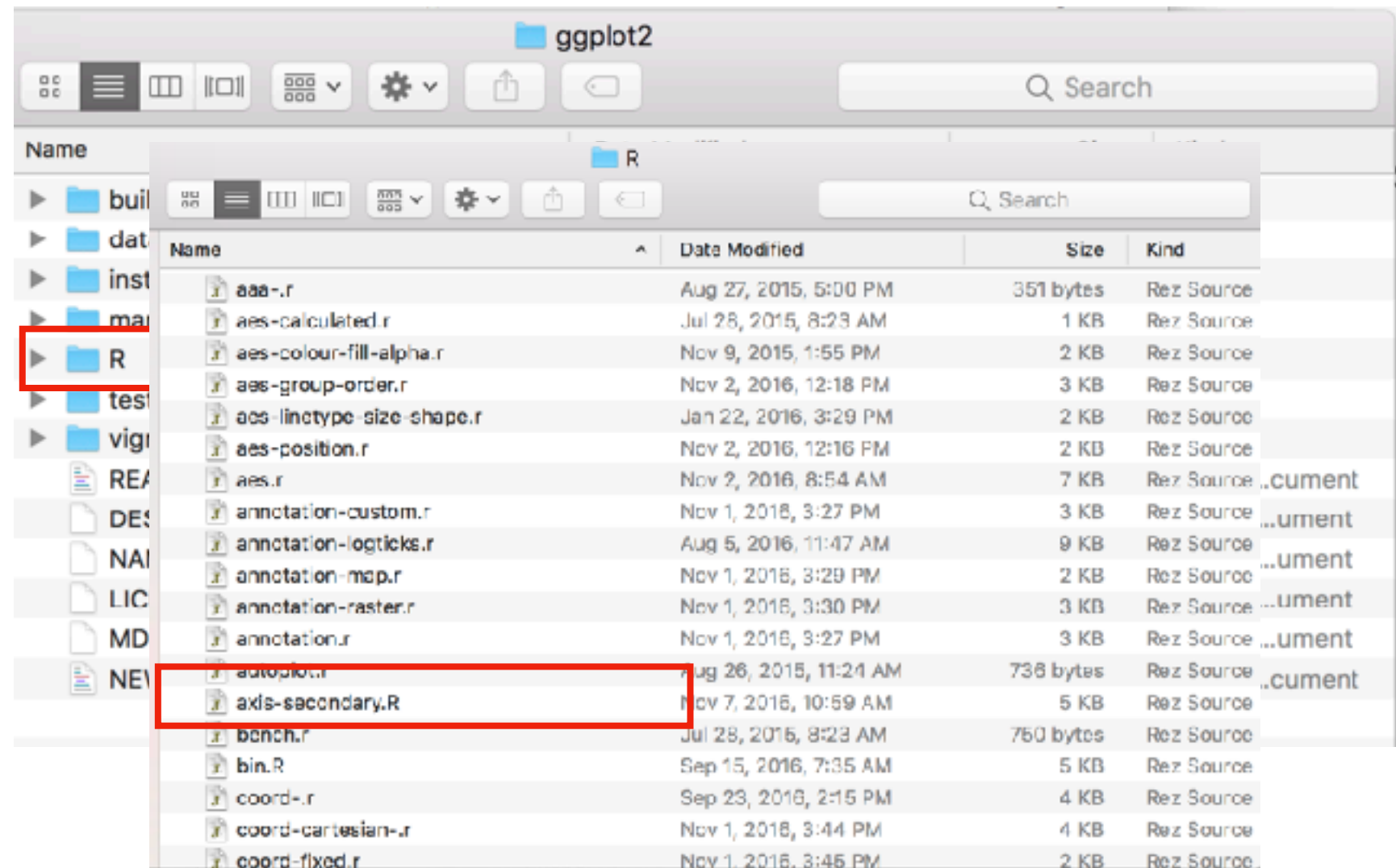
# What is an R Package?

- Packages are the fundamental units of reproducible R code
  - Improve existing base R functionalities, or by adding new ones.
  - Include R functions, documentation that describes how to use them and sample data

- **R Functions**
- Documentation
- Sample data

## ***Package Folder***

- ***R Folder***
- ***.R Files***





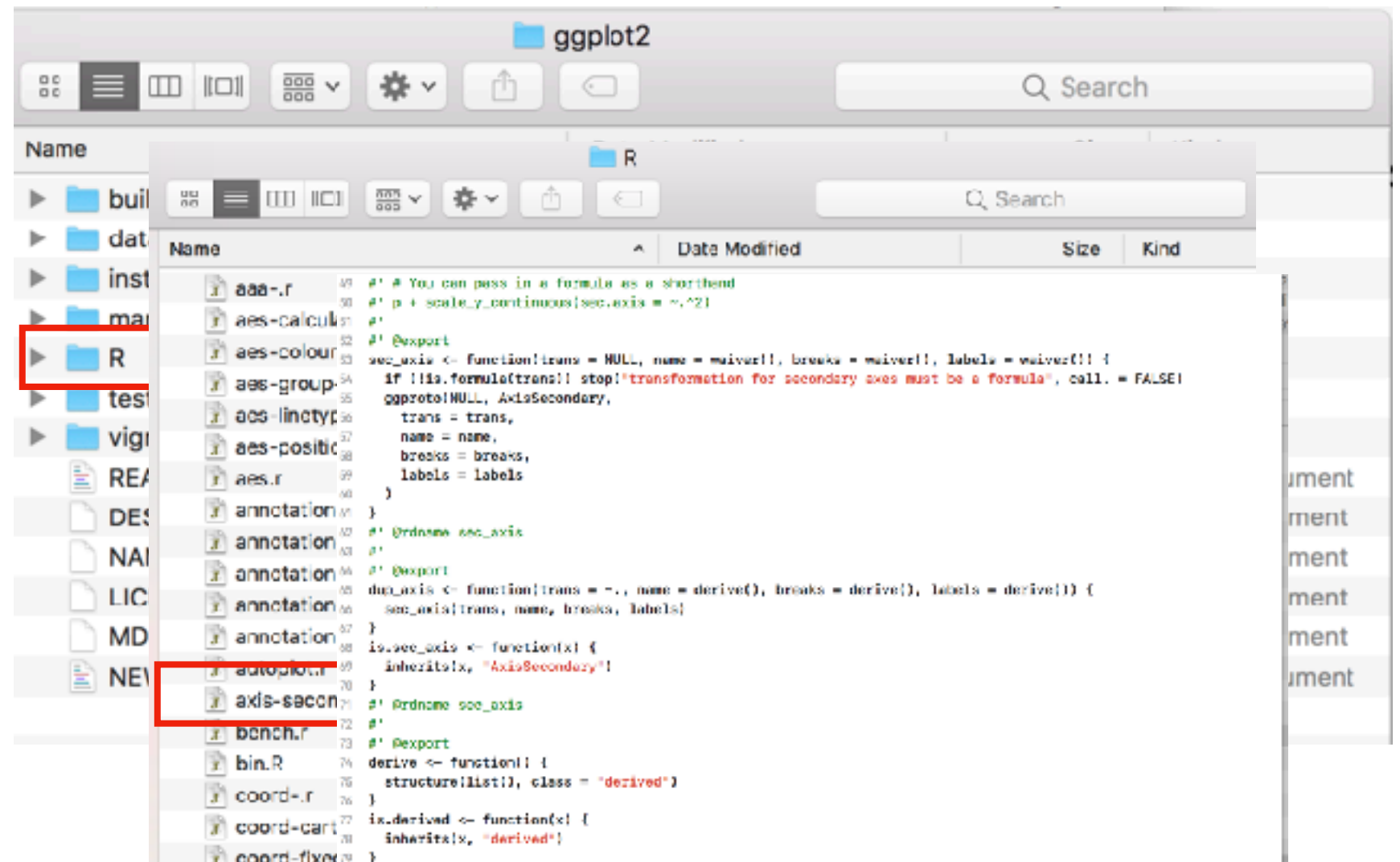
# What is an R Package?

- Packages are the fundamental units of reproducible R code
  - Improve existing base R functionalities, or by adding new ones.
  - Include R functions, documentation that describes how to use them and sample data

- **R Functions**
- Documentation
- Sample data

## ***Package Folder***

- ***R Folder***
  - ***.R Files***
  - ***Functions written in R code***



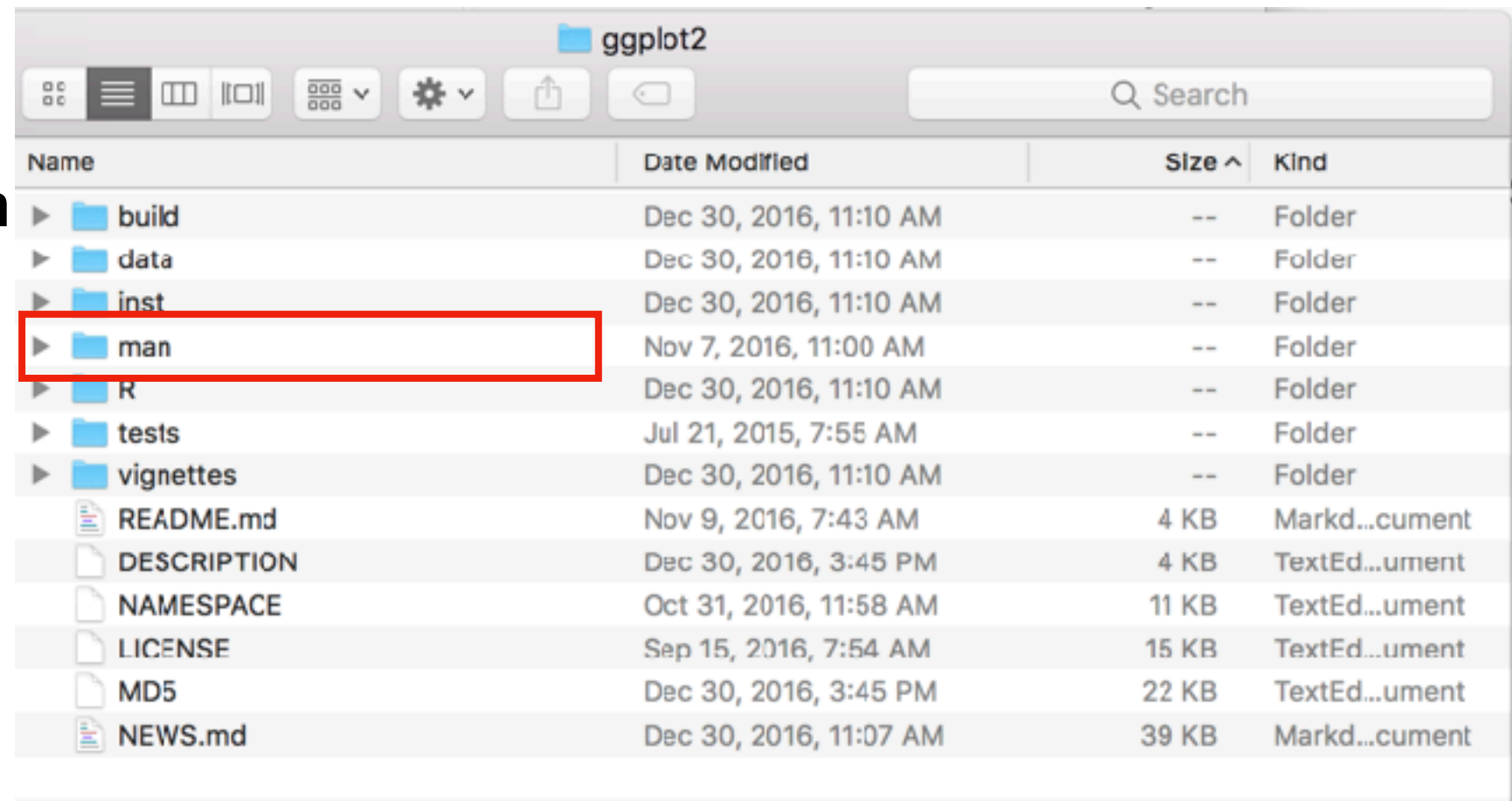
# What is an R Package?

- Packages are the fundamental units of reproducible R code
  - Improve existing base R functionalities, or by adding new ones.
  - Include R functions, documentation that describes how to use them and sample data

- R Functions
- **Documentation**
- Sample data

## ***Package Folder***

- ***man Folder***



Name	Date Modified	Size ^	Kind
▶ build	Dec 30, 2016, 11:10 AM	--	Folder
▶ data	Dec 30, 2016, 11:10 AM	--	Folder
▶ inst	Dec 30, 2016, 11:10 AM	--	Folder
▶ <b>man</b>	Nov 7, 2016, 11:00 AM	--	Folder
▶ R	Dec 30, 2016, 11:10 AM	--	Folder
▶ tests	Jul 21, 2015, 7:55 AM	--	Folder
▶ vignettes	Dec 30, 2016, 11:10 AM	--	Folder
README.md	Nov 9, 2016, 7:43 AM	4 KB	Markd...cument
DESCRIPTION	Dec 30, 2016, 3:45 PM	4 KB	TextEd...ument
NAMESPACE	Oct 31, 2016, 11:58 AM	11 KB	TextEd...ument
LICENSE	Sep 15, 2016, 7:54 AM	15 KB	TextEd...ument
MD5	Dec 30, 2016, 3:45 PM	22 KB	TextEd...ument
NEWS.md	Dec 30, 2016, 11:07 AM	39 KB	Markd...cument



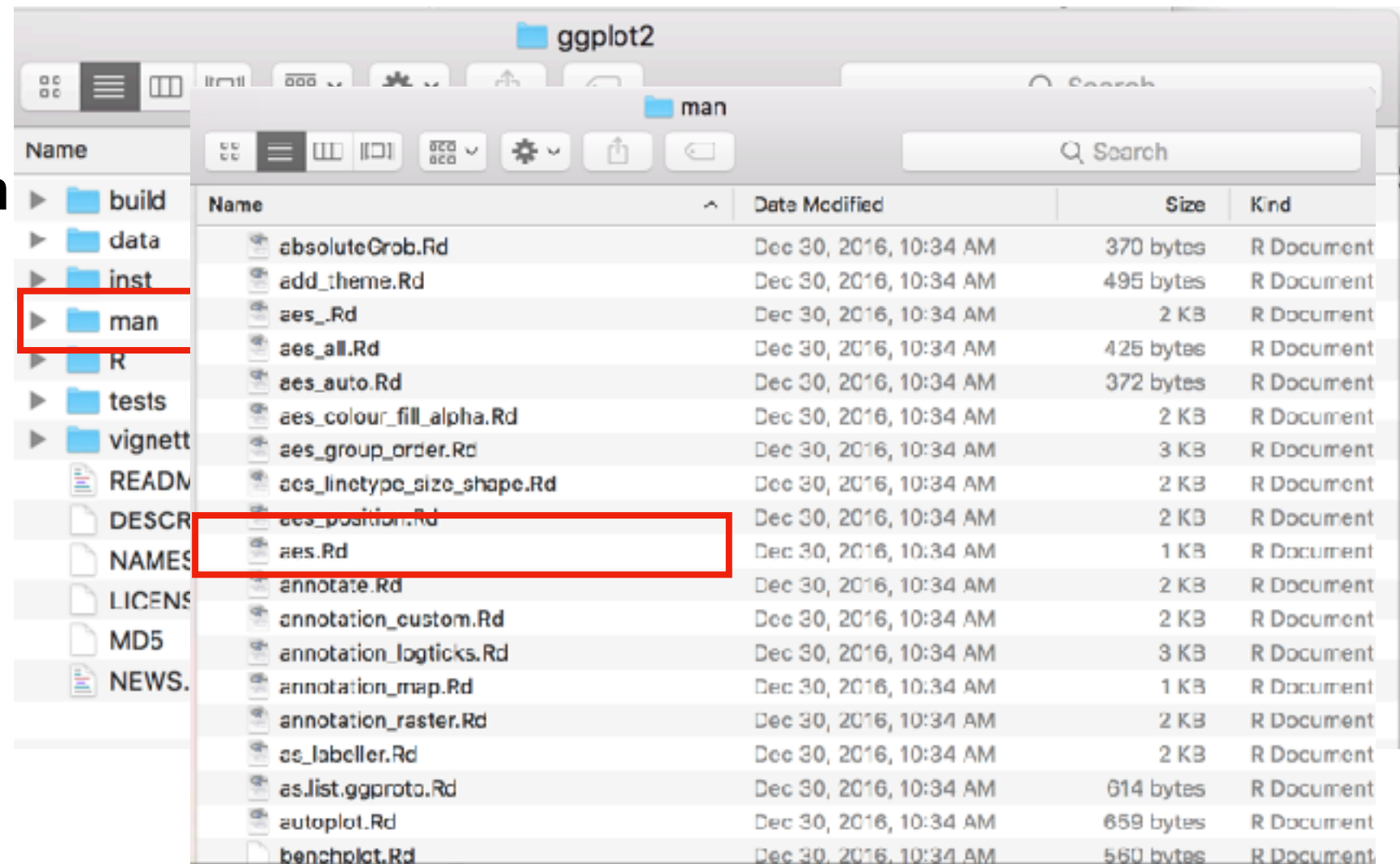
# What is an R Package?

- Packages are the fundamental units of reproducible R code
  - Improve existing base R functionalities, or by adding new ones.
  - Include R functions, documentation that describes how to use them and sample data

- R Functions
- **Documentation**
- Sample data

## ***Package Folder***

- ***man Folder***
  - ***.Rd files***



# What is an R Package?

- Packages are the fundamental units of reproducible R code
  - Improve existing base R functionalities, or by adding new ones.
  - Include R functions, documentation that describes how to use them and sample data

- R Functions

- **Documentation**

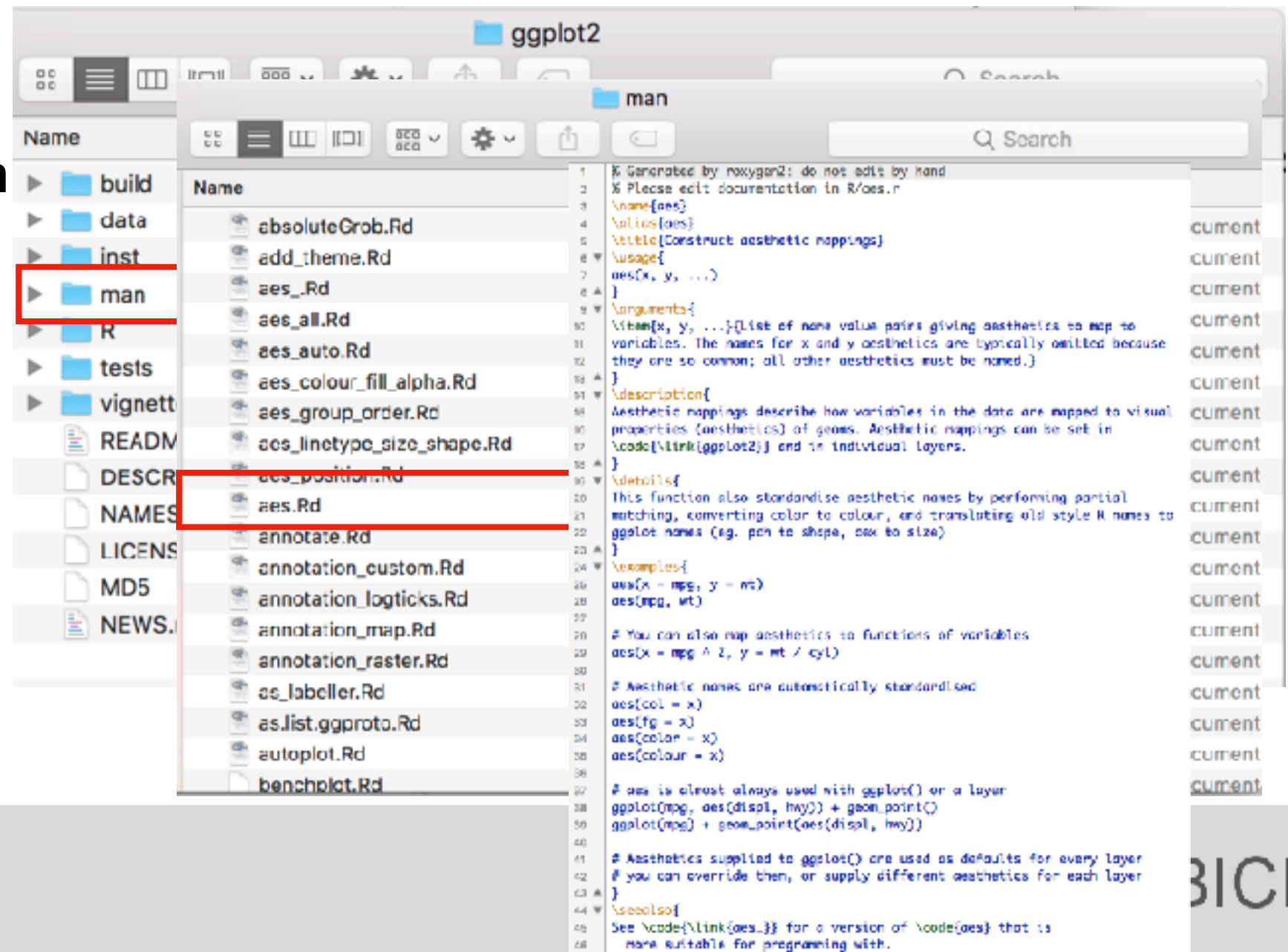
- Sample data

**Package Folder**

- **man Folder**

- **.Rd files**

- **Code used to generate standard manual**

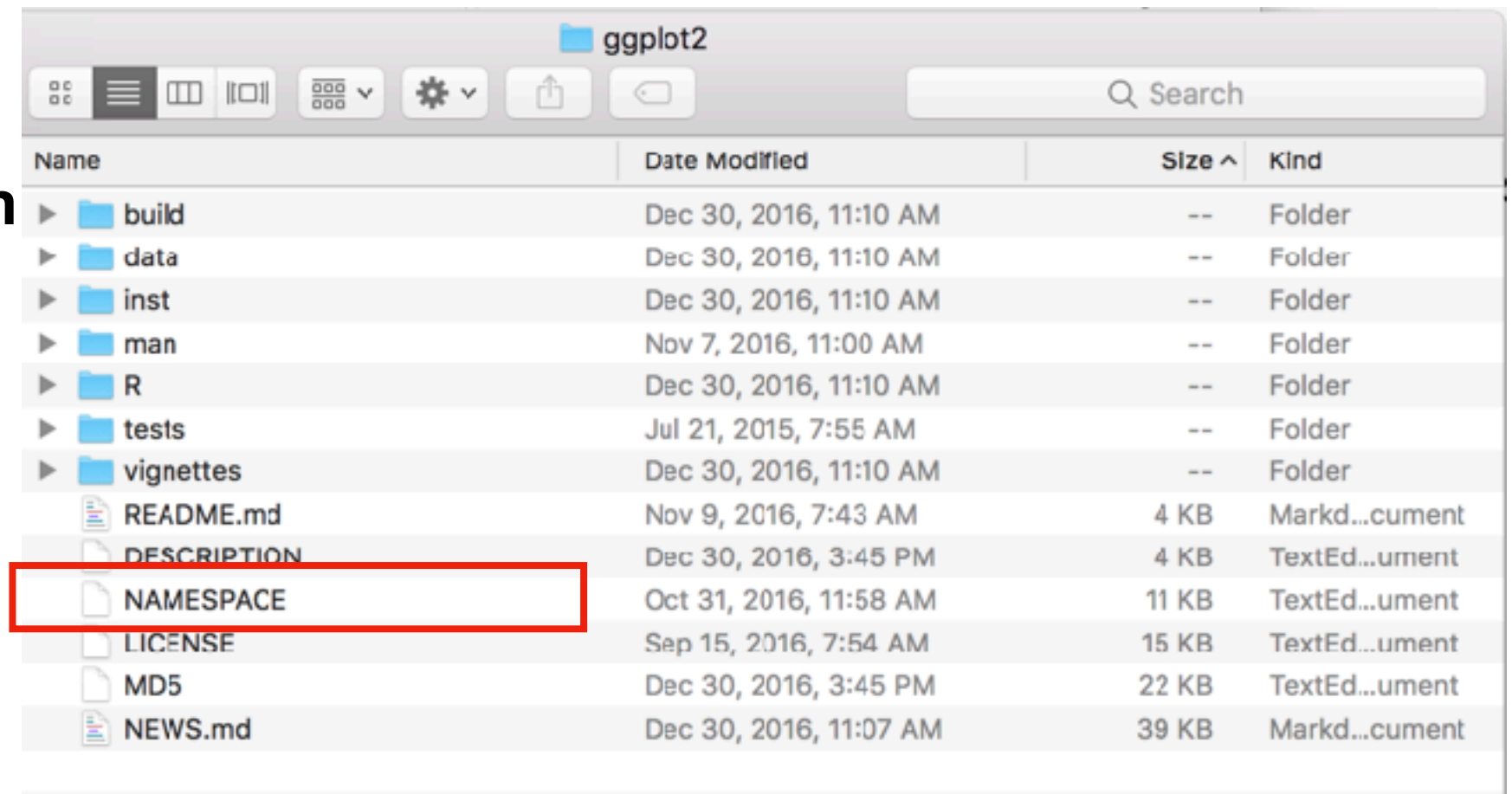


# What is an R Package?

- Packages are the fundamental units of reproducible R code
  - Improve existing base R functionalities, or by adding new ones.
  - Include R functions, documentation that describes how to use them and sample data
    - R Functions
    - **Documentation**
    - Sample data

## *Package Folder*

- **NAMESPACE**



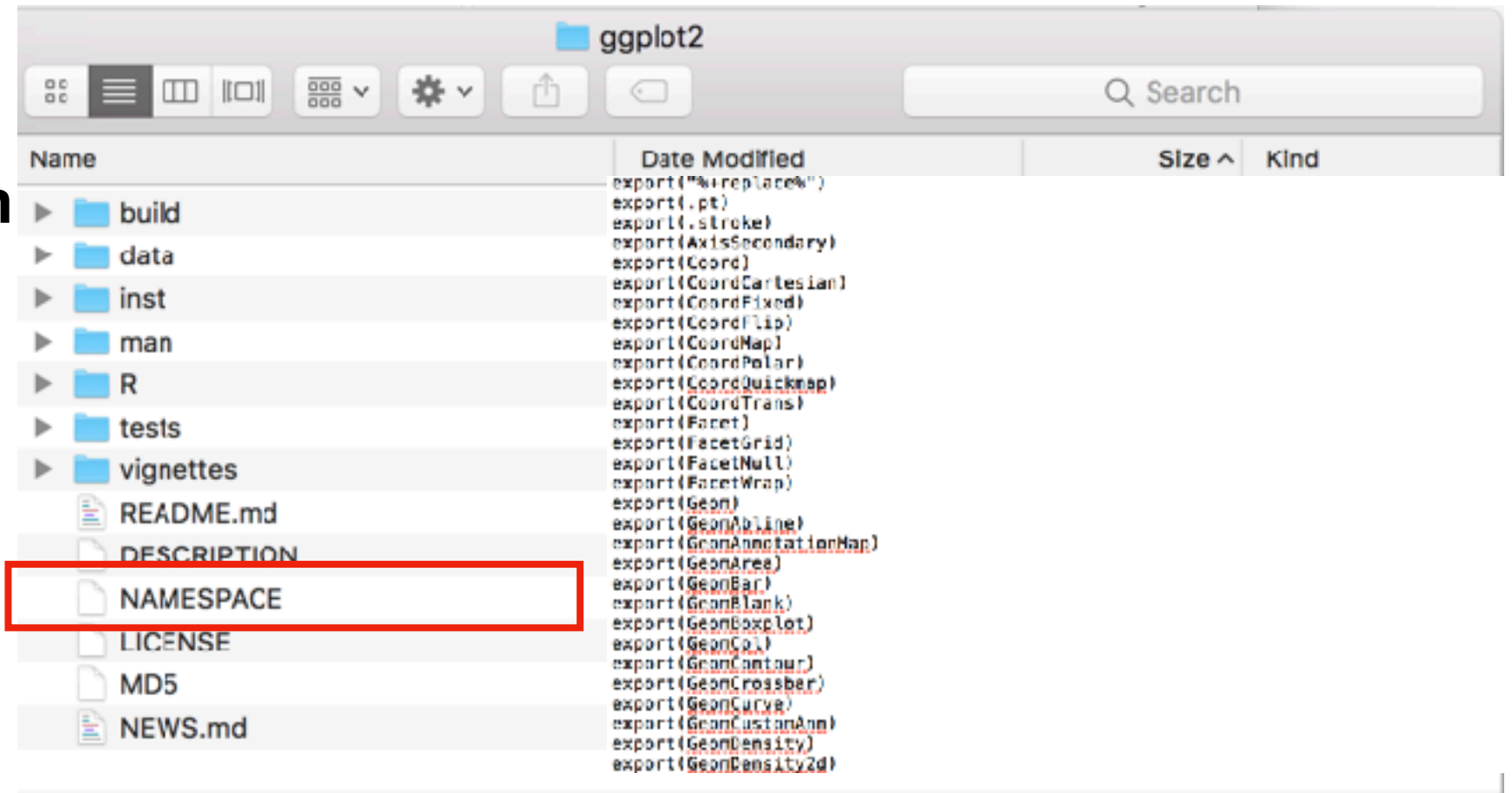
Name	Date Modified	Size ^	Kind
▶ build	Dec 30, 2016, 11:10 AM	--	Folder
▶ data	Dec 30, 2016, 11:10 AM	--	Folder
▶ inst	Dec 30, 2016, 11:10 AM	--	Folder
▶ man	Nov 7, 2016, 11:00 AM	--	Folder
▶ R	Dec 30, 2016, 11:10 AM	--	Folder
▶ tests	Jul 21, 2015, 7:55 AM	--	Folder
▶ vignettes	Dec 30, 2016, 11:10 AM	--	Folder
README.md	Nov 9, 2016, 7:43 AM	4 KB	Markd...cument
DESCRIPTION	Dec 30, 2016, 3:45 PM	4 KB	TextEd...ument
<b>NAMESPACE</b>	Oct 31, 2016, 11:58 AM	11 KB	TextEd...ument
LICENSE	Sep 15, 2016, 7:54 AM	15 KB	TextEd...ument
MD5	Dec 30, 2016, 3:45 PM	22 KB	TextEd...ument
NEWS.md	Dec 30, 2016, 11:07 AM	39 KB	Markd...cument

# What is an R Package?

- Packages are the fundamental units of reproducible R code
  - Improve existing base R functionalities, or by adding new ones.
  - Include R functions, documentation that describes how to use them and sample data
    - R Functions
    - **Documentation**
    - Sample data

## *Package Folder*

- **NAMESPACE**
- *Specifies what functions and objects are exposed*



Name	Date Modified	Size ^	Kind
▶ build	export("%replace%")		
▶ data	export(.pt)		
▶ inst	export(.stroke)		
▶ man	export(AxisSecondary)		
▶ R	export(Coord)		
▶ tests	export(CoordCartesian)		
▶ vignettes	export(CoordFixed)		
README.md	export(CoordFlip)		
DESCRIPTION	export(CoordMap)		
<b>NAMESPACE</b>	export(CoordPolar)		
LICENSE	export(CoordQuickmap)		
MD5	export(CoordTrans)		
NEWS.md	export(Facet)		
	export(FacetGrid)		
	export(FacetNull)		
	export(FacetWrap)		
	export(Geom)		
	export(GeomAbline)		
	export(GeomAnnotationMap)		
	export(GeomArea)		
	export(GeomBar)		
	export(GeomBlank)		
	export(GeomBoxplot)		
	export(GeomCo)		
	export(GeomContour)		
	export(GeomCrossbar)		
	export(GeomCurve)		
	export(GeomCustomAnn)		
	export(GeomDensity)		
	export(GeomDensity2d)		



# What is an R Package?

- Packages are the fundamental units of reproducible R code
  - Improve existing base R functionalities, or by adding new ones.
  - Include R functions, documentation that describes how to use them and sample data

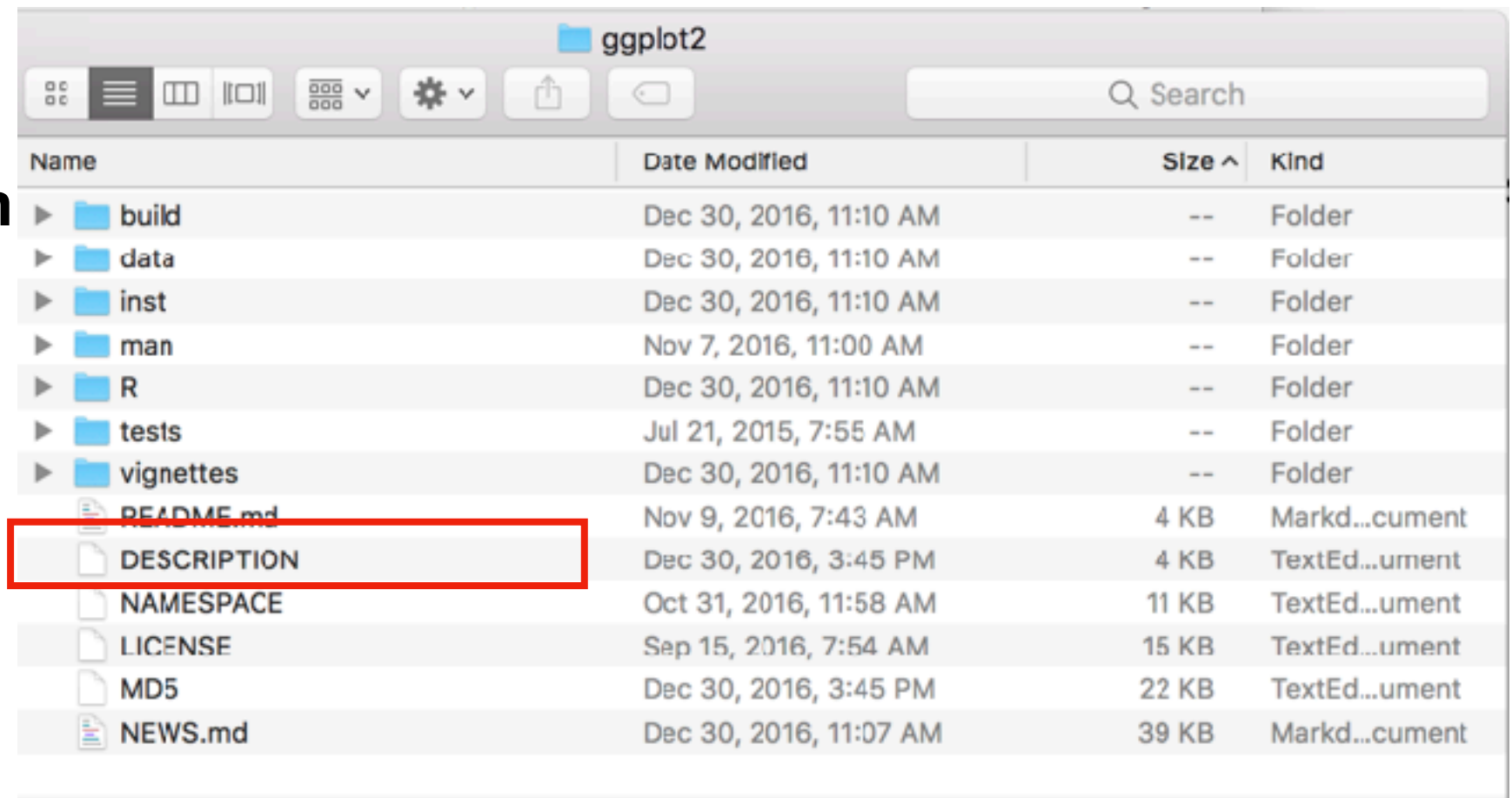
- R Functions

- **Documentation**

- Sample data

## ***Package Folder***

- ***DESCRIPTION***



Name	Date Modified	Size ^	Kind
▶ build	Dec 30, 2016, 11:10 AM	--	Folder
▶ data	Dec 30, 2016, 11:10 AM	--	Folder
▶ inst	Dec 30, 2016, 11:10 AM	--	Folder
▶ man	Nov 7, 2016, 11:00 AM	--	Folder
▶ R	Dec 30, 2016, 11:10 AM	--	Folder
▶ tests	Jul 21, 2015, 7:55 AM	--	Folder
▶ vignettes	Dec 30, 2016, 11:10 AM	--	Folder
▶ README.md	Nov 9, 2016, 7:43 AM	4 KB	Markd...cument
DESCRIPTION	Dec 30, 2016, 3:45 PM	4 KB	TextEd...ument
NAMESPACE	Oct 31, 2016, 11:58 AM	11 KB	TextEd...ument
LICENSE	Sep 15, 2016, 7:54 AM	15 KB	TextEd...ument
MD5	Dec 30, 2016, 3:45 PM	22 KB	TextEd...ument
NEWS.md	Dec 30, 2016, 11:07 AM	39 KB	Markd...cument

# What is an R Package?

- Packages are the fundamental units of reproducible R code
  - Improve existing base R functionalities, or by adding new ones.
  - Include R functions, documentation that describes how to use them and sample data

- R Functions

- **Documentation**

- Sample data

## ***Package Folder***

- ***DESCRIPTION***

- ***Dependencies***

- ***Package Imports***

- ***System Requirements***

- ***Suggested Packages***





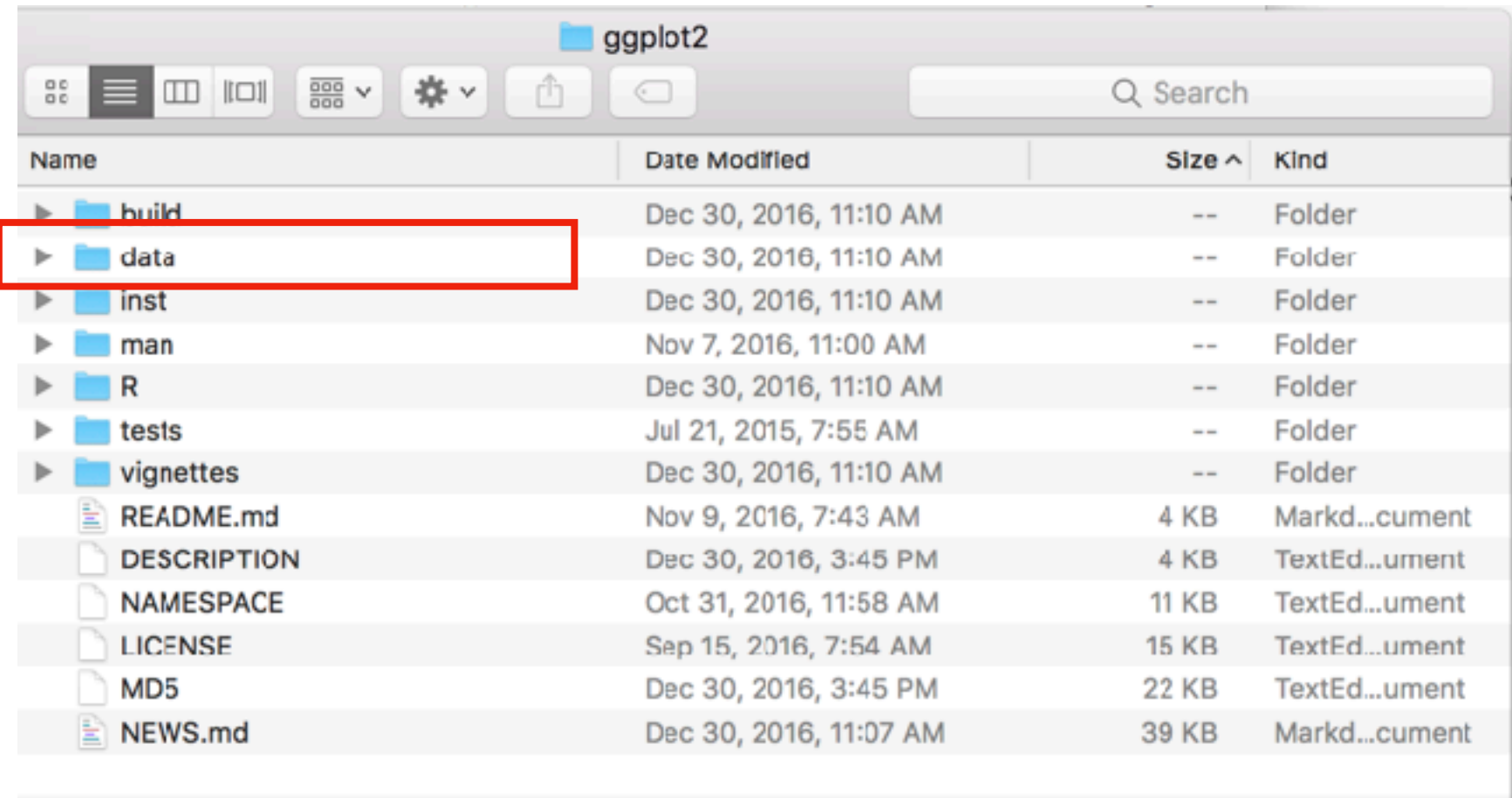
# What is an R Package?

- Packages are the fundamental units of reproducible R code
  - Improve existing base R functionalities, or by adding new ones.
  - Include R functions, documentation that describes how to use them and sample data

- R Functions
- Documentation
- **Sample data**

## ***Package Folder***

- ***data Folder***



Name	Date Modified	Size ^	Kind
▶ build	Dec 30, 2016, 11:10 AM	--	Folder
▶ data	Dec 30, 2016, 11:10 AM	--	Folder
▶ inst	Dec 30, 2016, 11:10 AM	--	Folder
▶ man	Nov 7, 2016, 11:00 AM	--	Folder
▶ R	Dec 30, 2016, 11:10 AM	--	Folder
▶ tests	Jul 21, 2015, 7:55 AM	--	Folder
▶ vignettes	Dec 30, 2016, 11:10 AM	--	Folder
README.md	Nov 9, 2016, 7:43 AM	4 KB	Markd...cument
DESCRIPTION	Dec 30, 2016, 3:45 PM	4 KB	TextEd...ument
NAMESPACE	Oct 31, 2016, 11:58 AM	11 KB	TextEd...ument
LICENSE	Sep 15, 2016, 7:54 AM	15 KB	TextEd...ument
MD5	Dec 30, 2016, 3:45 PM	22 KB	TextEd...ument
NEWS.md	Dec 30, 2016, 11:07 AM	39 KB	Markd...cument

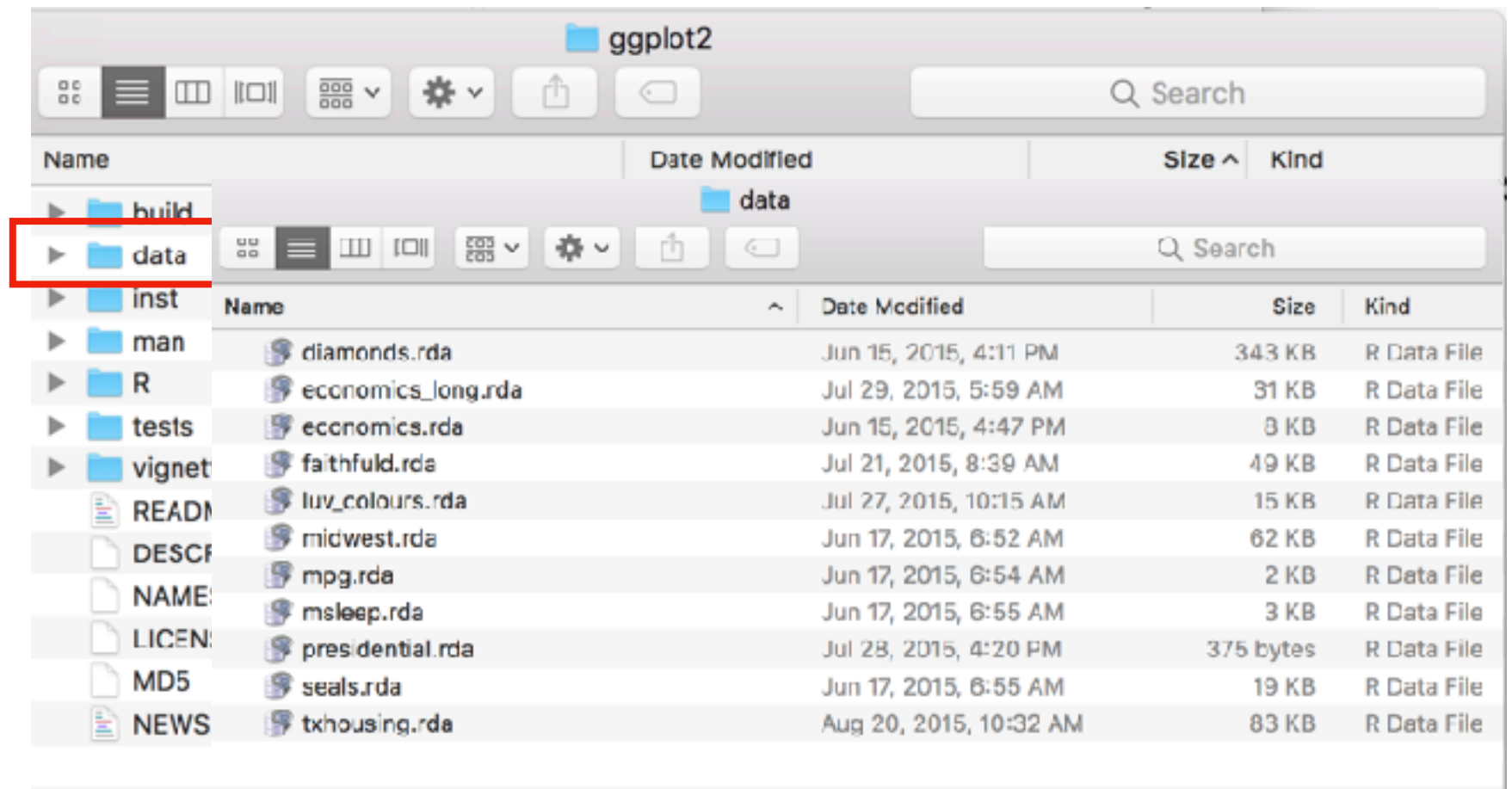
# What is an R Package?

- Packages are the fundamental units of reproducible R code
  - Improve existing base R functionalities, or by adding new ones.
  - Include R functions, documentation that describes how to use them and sample data

- R Functions
- Documentation
- **Sample data**

## ***Package Folder***

- ***data Folder***
- ***.Rda files***
  - ***Intrinsic to function utility***
  - ***Run example successfully***



# What is a Bioconductor Package?

- Packages that meet requirements for R packages and aligned with goals set by Bioconductor project
  - To provide widespread access to a broad range of **powerful statistical** and **graphical methods** for the analysis of **genomic data**.
  - To facilitate the inclusion of **biological metadata** in the analysis of genomic data, e.g. literature data from PubMed, annotation data from Entrez genes.
  - To provide a common **software platform** that enables the rapid development and deployment of extensible, scalable, and interoperable software.
  - To further scientific understanding by producing high-quality **documentation and reproducible research**.

# Types of Bioconductor Packages

- **Software Packages:** perform analytic calculations
- **Annotation Packages:** database-like packages that provide information linking identifies (e.g. Entrez Gene name) to other information (e.g. chromosomal location)
- **Experiment data Packages:** provide data sets that are used by software packs to illustrate particular analysis. Contain curated data from an experiment, teaching course or publication.

# Bioconductor Landing Page

The screenshot shows the Bioconductor landing page. The header includes the Bioconductor logo, the text 'OPEN SOURCE SOFTWARE FOR BIOINFORMATICS', a search bar, and navigation links: Home, Install, Help, Developers, and About. The main content area is divided into four columns. The first column contains 'About Bioconductor' and 'News'. The second column contains 'Install' and 'Use'. The third column contains 'Learn' and 'Develop'. A red arrow points from the 'News' section to the 'Software, Annotation, and Experiment packages' link in the 'Use' section.

**Bioconductor**  
OPEN SOURCE SOFTWARE FOR BIOINFORMATICS

Search:

[Home](#) [Install](#) [Help](#) [Developers](#) [About](#)

## About Bioconductor

Bioconductor provides tools for the analysis and comprehension of high-throughput genomic data. Bioconductor uses the R statistical programming language, and is open source and open development. It has two releases each year, [1473 software packages](#), and an active user community. Bioconductor is also available as an [AMI](#) (Amazon Machine Image) and a series of [Docker](#) images.

## News

- Bioconductor [3.6](#) is available.
- Bioconductor [F1000 Research Channel](#) available.
- Orchestrating high-throughput genomic analysis with *Bioconductor* ([abstract](#)) and other [recent literature](#).
- View recent [course material](#).
- Use the [support site](#) to get help installing, learning and using Bioconductor.

## Install »

Get started with *Bioconductor*

- [Install Bioconductor](#)
- [Explore packages](#)
- [Get support](#)
- [Latest newsletter](#)
- [Follow us on twitter](#)
- [Install R](#)

## Learn »

Master *Bioconductor* tools

- [Courses](#)
- [Support site](#)
- [Package vignettes](#)
- [Literature citations](#)
- [Common work flows](#)
- [FAQ](#)
- [Community resources](#)
- [Videos](#)

## Use »

Create bioinformatic solutions with *Bioconductor*

- [Software, Annotation, and Experiment packages](#)
- [Amazon Machine Image](#)
- [Latest release announcement](#)
- [Support site](#)

## Develop »

Contribute to *Bioconductor*

- [Developer resources](#)
- [Use Bioc 'devel'](#)
- 'Devel' [Software, Annotation and Experiment packages](#)
- [Package guidelines](#)
- [New package submission](#)
- [Git source control](#)
- [Build reports](#)



# Bioconductor Search/Discover Packages (biocView)

The screenshot shows the Bioconductor website's 'All Packages' page. At the top, there is a teal navigation bar with links for Home, Install, Help, Developers, and About. A search bar is located in the top right corner. Below the navigation bar, the page title 'All Packages' is displayed in green. On the left side, there is a sidebar titled 'Bioconductor version 3.6 (Release)' with an autocomplete search field. The sidebar lists various categories with their respective package counts: Software (1476), AssayDomain (573), BiologicalQuestion (559), Infrastructure (323), ResearchField (413), StatisticalMethod (489), Technology (933), WorkflowStep (774), AnnotationData (908), and ExperimentData (325). The main content area is titled 'Packages found under Software:' and features a table with columns for Package, Maintainer, and Title. The table lists several packages related to Affymetrix array analysis, all maintained by Tobias Verbeke, Willem Ligtenberg. A red arrow points to the top navigation bar, and another red arrow points to the search bar.

Package	Maintainer	Title
<a href="#">a4</a>	Tobias Verbeke, Willem Ligtenberg	Automated Affymetrix Array Analysis Umbrella Package
<a href="#">a4Base</a>	Tobias Verbeke, Willem Ligtenberg	Automated Affymetrix Array Analysis Base Package
<a href="#">a4Classif</a>	Tobias Verbeke, Willem Ligtenberg	Automated Affymetrix Array Analysis Classification Package
<a href="#">a4Core</a>	Tobias Verbeke, Willem Ligtenberg	Automated Affymetrix Array Analysis Core Package
<a href="#">a4Preproc</a>	Tobias Verbeke, Willem Ligtenberg	Automated Affymetrix Array Analysis Preprocessing Package
<a href="#">a4Reporting</a>	Tobias Verbeke, Willem Ligtenberg	Automated Affymetrix Array Analysis Reporting Package
<a href="#">ASAPEnrichment</a>	Steffen Gratz	Gene expression enrichment in human brain



# Bioconductor Landing Page

- Title
- author / maintainer
- short description
- citation
- installation instructions
- shields
- documentation
- download statistics

# Bioconductor Landing Page

**Bioconductor**  
OPEN SOURCE SOFTWARE FOR BIOINFORMATICS

Home Install Help Developers About

Search:

Home » Bioconductor 3.6 » Software Packages » DiffBind

## DiffBind

platforms all downloads top 5% posts 53 / 1 / 1 / 8 in Bloc 6 years  
build ok

DOI: [10.18129/B9.bioc.DiffBind](https://doi.org/10.18129/B9.bioc.DiffBind) [f](#) [t](#)

### Differential Binding Analysis of ChIP-Seq Peak Data

Bioconductor version: Release (3.6)

Compute differentially bound sites from multiple ChIP-seq experiments using affinity (quantitative) data. Also enables occupancy (overlap) analysis and plotting functions.

Author: Rory Stark<[rory.stark@cruk.cam.ac.uk](mailto:rory.stark@cruk.cam.ac.uk)>, Gord Brown <[gdbzork@gmail.com](mailto:gdbzork@gmail.com)>  
Maintainer: Rory Stark<[rory.stark@cruk.cam.ac.uk](mailto:rory.stark@cruk.cam.ac.uk)>  
Citation (from within R, enter `citation("DiffBind")`):  
Stark R and Brown G (2011). *DiffBind: differential binding analysis of ChIP-Seq peak data*.  
<http://bioconductor.org/packages/release/bioc/vignettes/DiffBind/inst/doc/DiffBind.pdf>.  
Ross-Innes CS, Stark R, Teschendorff AE, Holmes KA, Ali HR, Dunning MJ, Brown GD, Gojis O, Ellis IO, Green AR, Ali S, Chin S, Palmieri C, Caldas C and Carroll JS (2012). "Differential oestrogen receptor binding is associated with clinical outcome in breast cancer." *Nature*, **481**, pp. -4. <http://www.nature.com/nature/journal/v481/n7381/full/nature10730.html>.

#### Documentation »

*Bioconductor*

- Package [vignettes](#) and manuals.
- [Workflows](#) for learning and use.
- [Course and conference](#) material.
- [Videos](#).
- Community [resources](#) and [tutorials](#).

*R / CRAN* packages and [documentation](#)

#### Support »

Please read the [posting guide](#). Post questions about Bioconductor to one of the following locations:

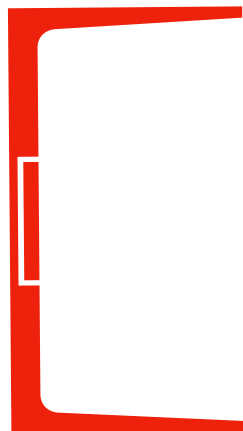
- [Support site](#) - for questions about Bioconductor packages
- [Bioc-devel](#) mailing list - for package developers

Shields  
Package health  
And stats

Package  
Information  
(DESCRIPTION)

# Bioconductor Landing Page (cont)

Documentation  
and Help



## Installation

To install this package, start R and enter:

```
## try http:// if https:// URLs are not supported
source("https://bioconductor.org/biocLite.R")
biocLite("DiffBind")
```

## Documentation

To view documentation for the version of this package installed in your system, start R and enter:

```
browseVignettes("DiffBind")
```

[PDF](#)

[R Script](#)

DiffBind: Differential binding analysis of ChIP-Seq peak data

[PDF](#)

Reference Manual

[Text](#)

NEWS

# Bioconductor Packages Page (cont)

  
**Package  
Information  
(DESCRIPTION)**

## Details

biocViews	<a href="#">ChIPSeq</a> , <a href="#">DifferentialPeakCalling</a> , <a href="#">Sequencing</a> , <a href="#">Software</a>
Version	2.6.5
In Bioconductor since	BioC 2.9 (R-2.14) (6 years)
License	Artistic-2.0
Depends	R ( $\geq 3.4$ ), <a href="#">GenomicRanges</a> , <a href="#">SummarizedExperiment</a>
Imports	<a href="#">RColorBrewer</a> , <a href="#">amap</a> , <a href="#">edgeR</a> , <a href="#">gplots</a> , <a href="#">grDevices</a> , <a href="#">limma</a> , <a href="#">GenomicAlignments</a> , <a href="#">locfit</a> , <a href="#">stats</a> , <a href="#">utils</a> , <a href="#">IRanges</a> , <a href="#">zlibbioc</a> , <a href="#">lattice</a> , <a href="#">systemPipeR</a> , <a href="#">tools</a> , <a href="#">Rcpp</a> , <a href="#">dplyr</a> , <a href="#">ggplot2</a> , <a href="#">BiocParallel</a> , <a href="#">parallel</a> , <a href="#">S4Vectors</a> , <a href="#">Rsamtools</a> , <a href="#">DESeq2</a> , <a href="#">methods</a> , <a href="#">graphics</a> , <a href="#">ggrepel</a>
LinkingTo	<a href="#">Rsamtools</a> ( $\geq 1.19.38$ ), <a href="#">Rcpp</a>
Suggests	<a href="#">DESeq</a> , <a href="#">BiocStyle</a> , <a href="#">testthat</a>
SystemRequirements	
Enhances	<a href="#">rgl</a> , <a href="#">XLConnect</a>
URL	
Depends On Me	<a href="#">ChIPQC</a> , <a href="#">vulcan</a>
Imports Me	
Suggests Me	
	<a href="#">Build Report</a>

# Bioconductor Landing Documentation

- Manual
- Vignettes

## Installation

To install this package, start R and enter:

```
## try http:// if https:// URLs are not supported
source("https://bioconductor.org/biocLite.R")
biocLite("DiffBind")
```

## Documentation

To view documentation for the version of this package installed in your system, start R and enter:

```
browseVignettes("DiffBind")
```

[PDF](#)

[R Script](#)

DiffBind: Differential binding analysis of ChIP-Seq peak data


[PDF](#)

Reference Manual

[Text](#)

NEWS

# Bioconductor Packages Manual



R topics documented:	
DiffBind-package	2
dba	3
DBA object methods	6
DBA tamoxifen resistance dataset	7
dba.analyze	8
dba.contrast	10
dba.count	12
1	
2	
DiffBind-package	
dba.load	15
dba.mask	16
dba.overlap	18
dba.peakset	21
dba.plotBox	25
dba.plotHeatmap	28
dba.plotMA	31
dba.plotPCA	33
dba.plotVenn	36
dba.plotVolcano	38
dba.report	40
dba.size	43

Contains entries for each function contained in the package



# Bioconductor Packages Manual

## R topics documented:

[DiffBind package](#) .....  
[dba](#) .....  
[DBA object methods](#) .....  
[DBA tamoxifen resistance dataset](#) .....  
[dba.analyze](#) .....  
[dba.contrast](#) .....  
[dba.count](#) .....

dba

3

### Author(s)

Rory Stark <rory.stark @at@ cruk.cam.ac.uk> and Gord Brown <gdbzork @at@ gmail.com>

dba

*Construct a DBA object*

### Description

Constructs a new DBA object from a sample sheet, or based on an existing DBA object

### Usage

```
dba(DBA,mask, minOverlap=2,  
    sampleSheet="dba_samples.csv",  
    config=data.frame(RunParallel=TRUE, reportInit="DBA", DataType=DBA_DATA_GRANGES,  
                      AnalysisMethod=DBA_DESEQ2, minQct=15, fragmentSize=125,  
                      bCorPlot=FALSE, th=0.05, bUsePval=FALSE),  
    peakCaller="raw", peakFormat, scoreCol, bLowerScoreBetter,  
    filter, skipLines=0,  
    bAddCallerConsensus=FALSE,  
    bRemoveM=TRUE, bRemoveRandom=TRUE,  
    bSummarizedExperiment=FALSE,  
    bCorPlot, attributes)
```

### Arguments

DBA	existing DBA object – if present, will return a fully-constructed DBA object based on the passed one, using criteria specified in the mask and/or minOverlap parameters. If missing, will create a new DBA object based on the sampleSheet.
mask	logical or numerical vector indicating which peaksets to include in the resulting model if basing DBA object on an existing one. See <a href="#">dba.mask</a> .
minOverlap	only include peaks in at least this many peaksets in the main binding matrix if basing DBA object on an existing one. If minOverlap is between zero and one, peak will be included from at least this proportion of peaksets.
sampleSheet	data frame containing sample sheet, or file name of sample sheet to load (ignored if DBA is specified). Columns names in sample sheet may include:

2

[dba.load](#) .....  
[dba.mask](#) .....  
[dba.overlap](#) .....  
[dba.peakset](#) .....  
[dba.plotBox](#) .....  
[dba.plotHeatmap](#) .....  
[dba.plotMA](#) .....  
[dba.plotPCA](#) .....  
[dba.plotVenn](#) .....  
[dba.plotVolcano](#) .....  
[dba.report](#) .....  
[dba.size](#) .....

# Bioconductor Packages Manual

## R topics documented:

[DiffBind package](#) .....  
[dba](#) .....  
[DBA object methods](#) .....  
[DBA tamoxifen resistance dataset](#) .....  
[dba.analyze](#) .....  
[dba.contrast](#) .....  
[dba.count](#) .....

2

[dba.load](#) .....  
[dba.mask](#) .....  
[dba.overlap](#) .....  
[dba.peakset](#) .....  
[dba.plotBox](#) .....  
[dba.plotHeatmap](#) .....  
[dba.plotMA](#) .....  
[dba.plotPCA](#) .....  
[dba.plotVenn](#) .....  
[dba.plotVolcano](#) .....  
[dba.report](#) .....  
[dba.save](#) .....

dba

3

## Author(s)

Rory Stark <rory.stark @at@ cruk.cam.ac.uk> and Gord Brown <gdbzcrk @at@ gmail.com>

dba

*Construct a DBA object*

## Description

Constructs a new DBA object from a sample sheet, or based on an existing DBA object

## Usage

```
dba(DBA,mask, minOverlap=2,  
    sampleSheet="dba_samples.csv",  
    config=data.frame(RunParallel=TRUE, reportInit="DBA", DataType=DBA_DATA_GRANGES,  
                      AnalysisMethod=DBA_DESEQ2, minQct=15, fragmentSize=125,  
                      bCorPlot=FALSE, th=0.05, bUsePval=FALSE),  
    peakCaller="raw", peakFormat, scoreCol, bLowerScoreBetter,  
    filter, skipLines=0,  
    bAddCallerConsensus=FALSE,  
    bRemoveM=TRUE, bRemoveRandom=TRUE,  
    bSummarizedExperiment=FALSE,  
    bCorPlot, attributes)
```

## Arguments

DBA	existing DBA object – if present, will return a fully-constructed DBA object based on the passed one, using criteria specified in the mask and/or minOverlap parameters. If missing, will create a new DBA object based on the sampleSheet.
mask	logical or numerical vector indicating which peaksets to include in the resulting model if basing DBA object on an existing one. See <a href="#">dba.mask</a> .
minOverlap	only include peaks in at least this many peaksets in the main binding matrix if basing DBA object on an existing one. If minOverlap is between zero and one, peak will be included from at least this proportion of peaksets.
sampleSheet	data frame containing sample sheet, or file name of sample sheet to load (ignored if DBA is specified). Columns names in sample sheet may include:

# Bioconductor Packages Vignettes

- A vignette is an executable document consisting of a collection of **code chunks** and **documentation text chunks**.
- Vignettes provide **dynamic**, **integrated**, and **reproducible** documents that can be automatically updated if either data or analyses are changed.
- Provides a task - oriented description of package functionality
- Comes in two forms
  - R Script
  - PDF

## Documentation

To view documentation for the version of this package installed in your system, start R and enter:

```
browseVignettes("DiffBind")
```

[PDF](#)[R Script](#)

DiffBind: Differential binding analysis of ChIP-Seq peak data

[PDF](#)

Reference Manual

[Text](#)

NEWS

# Bioconductor Packages Vignettes PDF

## 3 Example: Obtaining differentially bound sites

Description  
Of what  
Code will do



This section offers a quick example of how to use *DiffBind* to identify significantly differentially bound sites using affinity (read count) data.

The dataset for this example consists of ChIPs against the transcription factor ERα using five breast cancer cell lines ([2]). Three of these cell lines are responsive to tamoxifen treatment, while two others are resistant to tamoxifen. There are at least two replicates for each of the cell lines, with one cell line having three replicates, for a total of eleven sequenced libraries. Of the five cell lines, two are based on MCF7 cells: the regular tamoxifen responsive line, as well as MCF7 cells specially treated with tamoxifen until a tamoxifen resistant cell line is obtained. For each sample, we have one peakset originally derived using the MACS peak caller ([1]), for a total of eleven peaksets. Note that to save space in the package, only data for chromosome 18 is used for the vignette. The metadata and peak data are available in the *extra* subdirectory of the *DiffBind* package directory; you can make this your working directory by entering:

```
> library(DiffBind)
> setwd(system.file("extra", package="DiffBind"))
```

Obtaining the sites significantly differentially bound (DB) between the samples that respond to tamoxifen and those that are resistant can be done in a five-step script:

```
> print(savewd)
> tamoxifen <- dba(sampleSheet="tamoxifen.csv")
> tamoxifen <- dba.count(tamoxifen)
> tamoxifen <- dba.contrast(tamoxifen)
> tamoxifen <- dba.analyze(tamoxifen)
> tamoxifen.DB <- dba.report(tamoxifen)
```

Code



The following subsections describe these steps in more detail



# Bioconductor Packages Vignettes

## R Script

```
### R code from vignette source 'DiffBind.Rnw'

#####
### code chunk number 1: style
#####
BiocStyle::latex()
savedwd <- getwd()

#####
### code chunk number 2: DiffBind.Rnw:136-140
#####
tmp <- tempfile(as.character(Sys.getpid()))
pdf(tmp)
savewarn <- options("warn")
options(warn=-1)

#####
### code chunk number 3: DiffBind.Rnw:144-146
#####
library(DiffBind)
setwd(system.file("extra", package="DiffBind"))

#####
### code chunk number 4: DiffBind.Rnw:151-157 (eval = FALSE)
#####
## print(savedwd)
## tamoxifen <- dba(sampleSheet="tamoxifen.csv")
## tamoxifen <- dba.count(tamoxifen)
## tamoxifen <- dba.contrast(tamoxifen)
## tamoxifen <- dba.analyze(tamoxifen)
## tamoxifen.DB <- dba.report(tamoxifen)

#####
### code chunk number 5: sampSheet
#####
samples <- read.csv(file.path(system.file("extra", package="DiffBind"),
                                     "tamoxifen.csv"))

names(samples)
samples

#####
### code chunk number 6: dbaConstruct
#####
tamoxifen <- dba(sampleSheet="tamoxifen.csv")
```

Code



# Install Bioconductor Packages

Use the biocLite.R script to install Bioconductor, CRAN and github packages.  
To install core packages, type the following in an R command window:

```
## try http:// if https:// URLs are not supported  
source("https://bioconductor.org/biocLite.R")  
biocLite()
```

Install specific packages using:

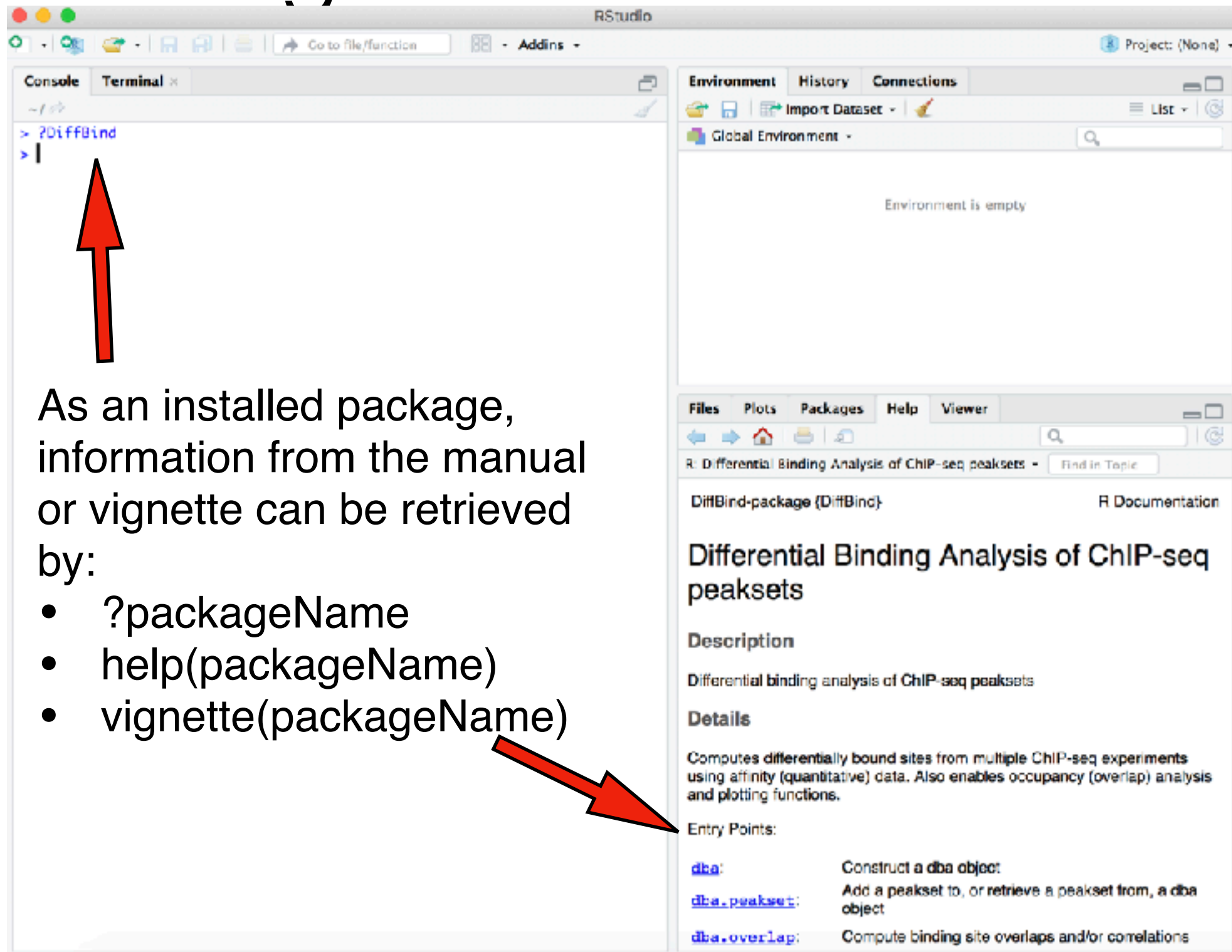
```
biocLite("DiffBind")
```

Once installed, the package can be load into an R session

```
library("DiffBind")
```



# Accessing Manual From RStudio



The screenshot shows the RStudio interface. In the console on the left, the command `?DiffBind` has been entered, and a red arrow points to it. On the right, the 'Help' pane displays the documentation for the 'DiffBind' package, titled 'Differential Binding Analysis of ChIP-seq peaksets'. A red arrow points from the text '`vignette(packageName)`' in the list below to the 'Details' section of the help page.

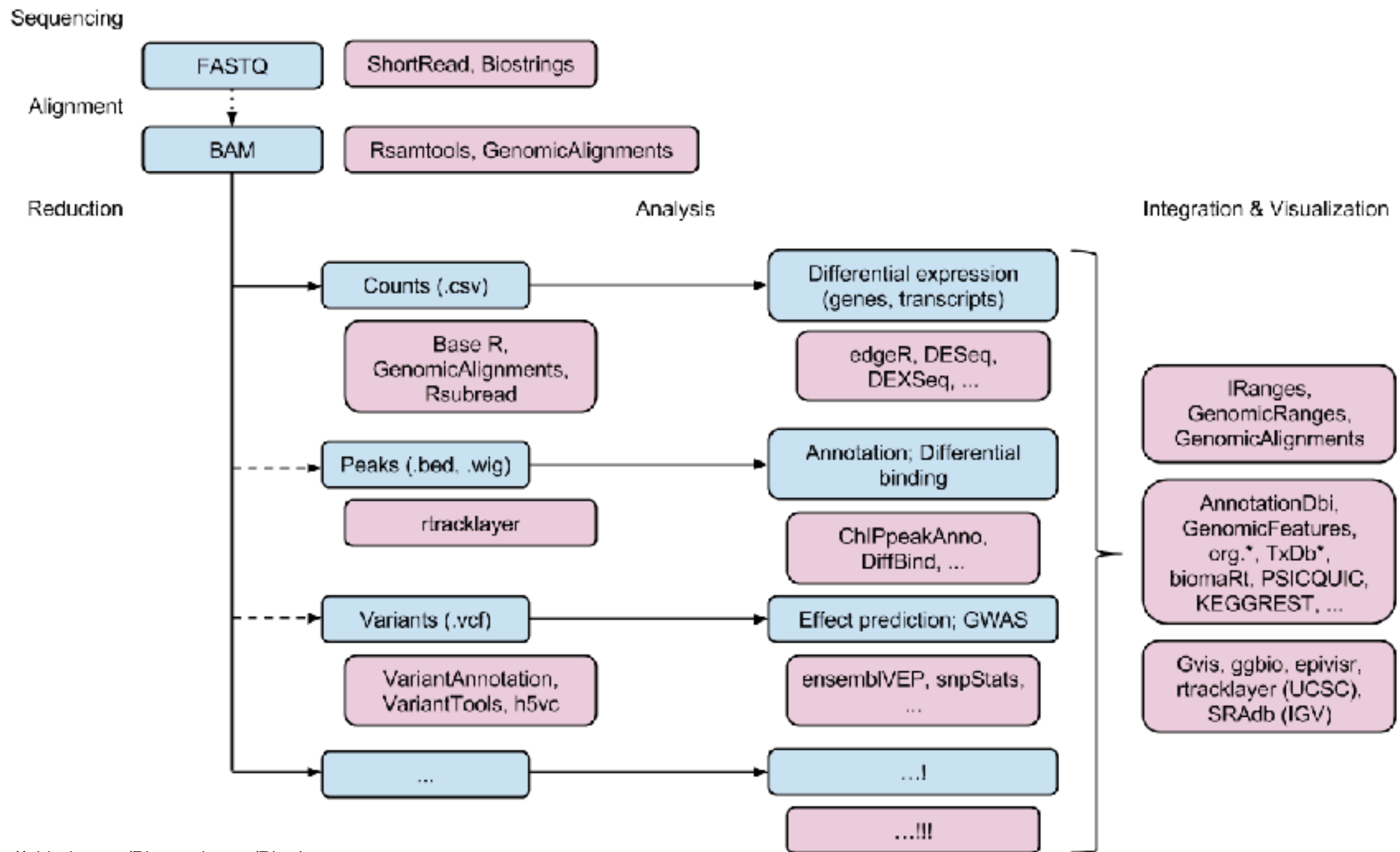
As an installed package, information from the manual or vignette can be retrieved by:

- `?packageName`
- `help(packageName)`
- `vignette(packageName)`

The help page for 'DiffBind' includes the following sections:

- Description**: Differential binding analysis of ChIP-seq peaksets
- Details**: Computes differentially bound sites from multiple ChIP-seq experiments using affinity (quantitative) data. Also enables occupancy (overlap) analysis and plotting functions.
- Entry Points**:
  - `dba`: Construct a dba object
  - `dba.peakset`: Add a peakset to, or retrieve a peakset from, a dba object
  - `dba.overlap`: Compute binding site overlaps and/or correlations

# Bioconductor NGS Workflow



<https://github.com/Bioconductor/BiocIntro>

# Bioconductor Infrastructure

- Biological Sequences: Biostrings
- Genomic Coordinates: GenomicRanges
- Aligned Reads: GenomicAlignment
- Multiple Assay and Experiments: SummarizedExperiment
- Annotation Resources

# Biological Sequences: Biostrings

## Manipulation of Biological Sequences

- DNA
- RNA
- Protein

```
> library(Biostrings)
Loading required package: XVector

Attaching package: 'Biostrings'

The following object is masked from 'package:DelayedArray':

    type

The following object is masked from 'package:base':

    strsplit

> dna <- DNAStringSet(c("AACAT", "GGCGCCT"))
> reverseComplement(dna)
A DNAStringSet instance of length 2
width seq
[1] 5 ATGTT
[2] 7 AGGCGCC
> |
```

# Genomic Coordinates: GenomicRanges

GRanges: Genomic coordinates to represent:

- annotations (exons, genes, regulatory marks, ...)
- data (called peaks, variants, aligned reads)

GRanges

- length()
- seqnames()
- start()
- end()
- width()
- strand()

```
> exons(TxDb.Hsapiens.UCSC.hg38.knownGene)
GRanges object with 581036 ranges and 1 metadata column:
```

	seqnames <Rle>	ranges <IRanges>	strand <Rle>	exon_id <integer>
[1]	chr1	[29554, 30039]	+	1
[2]	chr1	[30267, 30667]	+	2
[3]	chr1	[30366, 30503]	+	3
[4]	chr1	[30564, 30667]	+	4
[5]	chr1	[30976, 31097]	+	5
...	...	...	...	...
[581032]	chrUn_KI270750v1	[148668, 148843]	+	581032
[581033]	chrUn_KI270752v1	[ 144, 268]	+	581033
[581034]	chrUn_KI270752v1	[ 21813, 21944]	+	581034
[581035]	chrUn_KI270752v1	[ 3497, 3623]	-	581035
[581036]	chrUn_KI270752v1	[ 9943, 10067]	-	581036

```
-----
seqinfo: 455 sequences (1 circular) from hg38 genome
```

DataFrame

- mcols()
- gr\$exon\_id



# Genomic Coordinates: GenomicRanges

GRangesList: genomic coordinates grouped into list elements (e.g., paired-end reads; exons grouped by transcript)

## GRangesList

- length()
- range()
- shift()

```
Console Terminal x
~/
> exonsBy(TxDb.Hsapiens.UCSC.hg38.knownGene, 'tx')
GRangesList object of length 197782:
$1
GRanges object with 3 ranges and 3 metadata columns:
      seqnames      ranges strand | exon_id exon_name
      <Rle>        <IRanges> <Rle> | <integer> <character>
[1] chr1 [29554, 30039]      + |      1      <NA>
[2] chr1 [30564, 30667]      + |      4      <NA>
[3] chr1 [30976, 31097]      + |      5      <NA>
      exon_rank
      <integer>
[1] 1
[2] 2
[3] 3
$2
GRanges object with 2 ranges and 3 metadata columns:
      seqnames      ranges strand | exon_id exon_name exon_rank
      <Rle>        <IRanges> <Rle> | <integer> <character> <integer>
[1] chr1 [30267, 30667]      + |      2      <NA>      1
[2] chr1 [30976, 31109]      + |      6      <NA>      2
$3
GRanges object with 1 range and 3 metadata columns:
      seqnames      ranges strand | exon_id exon_name exon_rank
      <Rle>        <IRanges> <Rle> | <integer> <character> <integer>
[1] chr1 [30366, 30503]      + |      3      <NA>      1
...
<197779 more elements>
-----
seqinfo: 455 sequences (1 circular) from hg38 genome
>
```

## GRanges

- grl[[2]]
- grl\$2

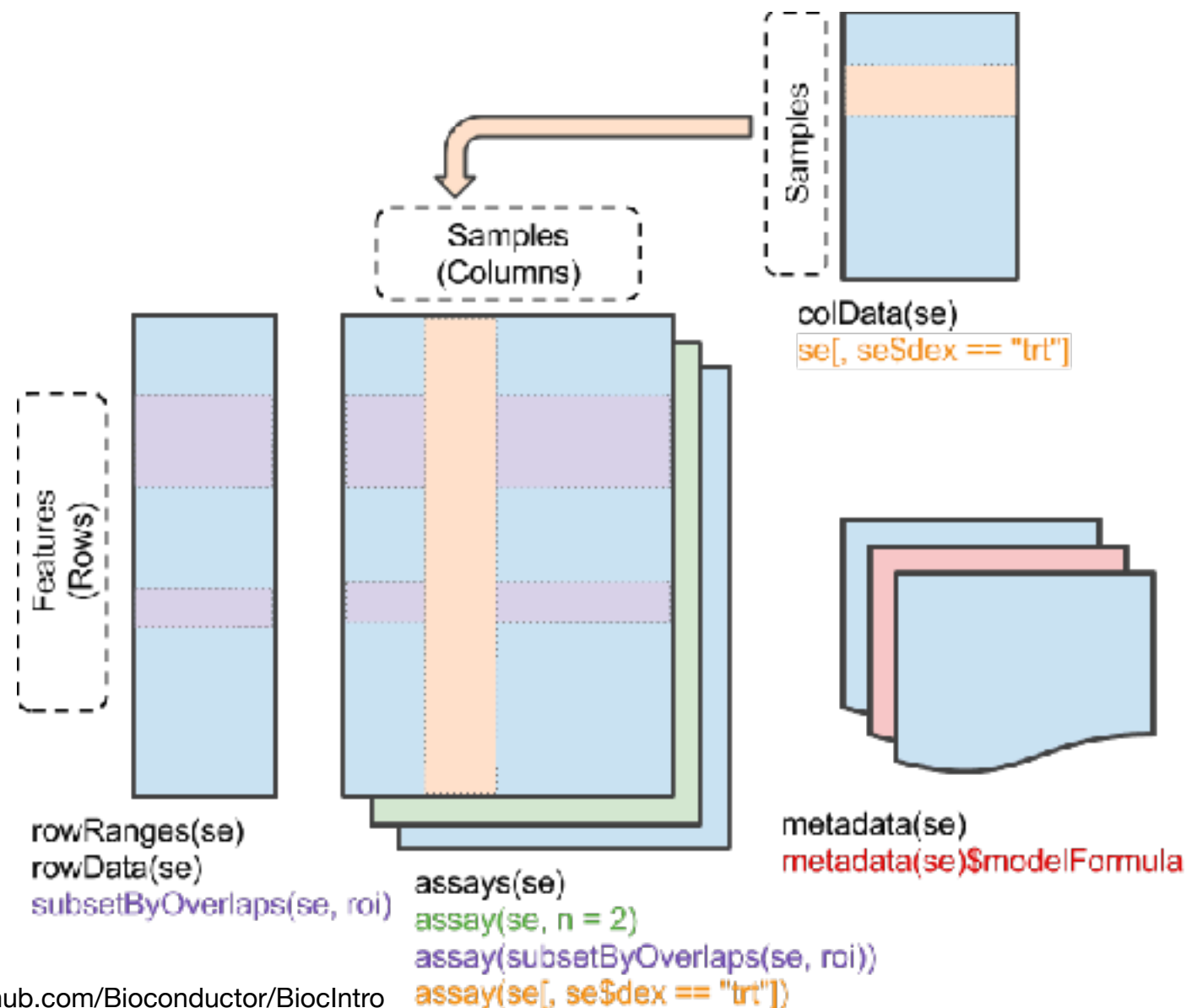
# Aligned Reads: GenomicAlignment

## Representation of aligned reads

```
> aln1_file <- system.file("extdata", "ex1.bam", package="Rsamtools")
> aln1 <- readGAlignments(aln1_file)
> aln1
GAlignments object with 3271 alignments and 0 metadata columns:
      seqnames strand      cigar    qwidth    start      end
      <Rle>   <Rle> <character> <integer> <integer> <integer>
 [1]    seq1      +      36M        36         1        36
 [2]    seq1      +      35M        35         3        37
 [3]    seq1      +      35M        35         5        39
 [4]    seq1      +      36M        36         6        41
 [5]    seq1      +      35M        35         9        43
 ...
[3267]   seq2      +      35M        35      1524      1558
[3268]   seq2      +      35M        35      1524      1558
[3269]   seq2      -      35M        35      1528      1562
[3270]   seq2      -      35M        35      1532      1566
[3271]   seq2      -      35M        35      1533      1567
      width    njunc
      <integer> <integer>
 [1]         36         0
 [2]         35         0
 [3]         35         0
 [4]         36         0
 [5]         35         0
 ...
[3267]         35         0
[3268]         35         0
[3269]         35         0
[3270]         35         0
[3271]         35         0
-----
seqinfo: 2 sequences from an unspecified genome
```

# Multiple Assay and Experiments: SummarizedExperiment

The SummarizedExperiment container contains one or more assays, each represented by a matrix-like object of numeric or other mode. The rows typically represent genomic ranges of interest and the columns represent samples.



- Coordinate feature x sample 'assays' with row (feature) and column (sample) descriptions.
- `colData()` data frame for description of samples
- `rowRanges()` *GRanges* / *GRangeList* or data frame for description of features
- `exptData()` to describe the entire object
- `assays()` can be any matrix-like object

# Annotation Resources

- *Bioconductor* provides extensive access to 'annotation' resources (see the AnnotationData biocViews hierarchy)
  - Packages for querying on-line resources; each of these have informative vignettes.
    - biomaRT
    - KEGGREST
- AnnotationDbi is a standard package:
  - **org** packages (e.g. org.Hs.eg.db) contain maps between different gene identifiers (e.g. ENTREZ and SYMBOL).
  - **TxDb** packages (e.g. TxDb.Hsapiens.UCSC.hg38.knownGene) contain gene models (e.g., exon coordinates, exon/transcript relationships) driven from common sources.
  - **BSgenome** packages (e.g. BSgenome.Hsapiens.UCSC.hg38) contain whole genome of organism
- VariantAnnotation and ensemblVEP provide access to sequence annotation facilities, e.g., to identify coding variants.

# Other Resources

- Bioconductor provided training Resources:
  - <https://bioconductor.org/help/course-materials/>
- Support Site:
  - <https://support.bioconductor.org/>
- Basic Introduction to analyze data:
  - <https://bioconductor.org/help/workflows/>