

R Highlight!

# CRAN & Bioconductor

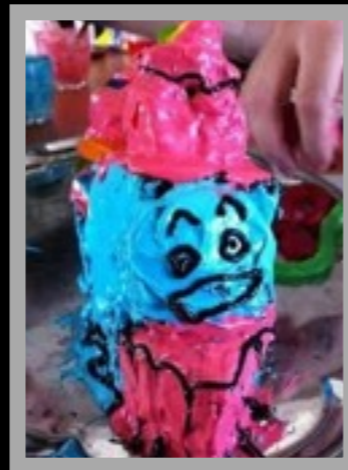
Major repositories for **R packages**  
that extend R functionality

# **CRAN**: Comprehensive R Archive Network

- CRAN is a network of mirrored servers around the world that administer and distribute R itself, R documentation and **R packages** (basically add on functionality!)
- There are currently ~9,000 packages on CRAN in the areas of finance, bioinformatics, machine learning, high performance computing, multivariate statistics, natural language processing, *etc. etc.*

<https://cran.r-project.org/>

# Side-note: R packages come in all shapes and sizes



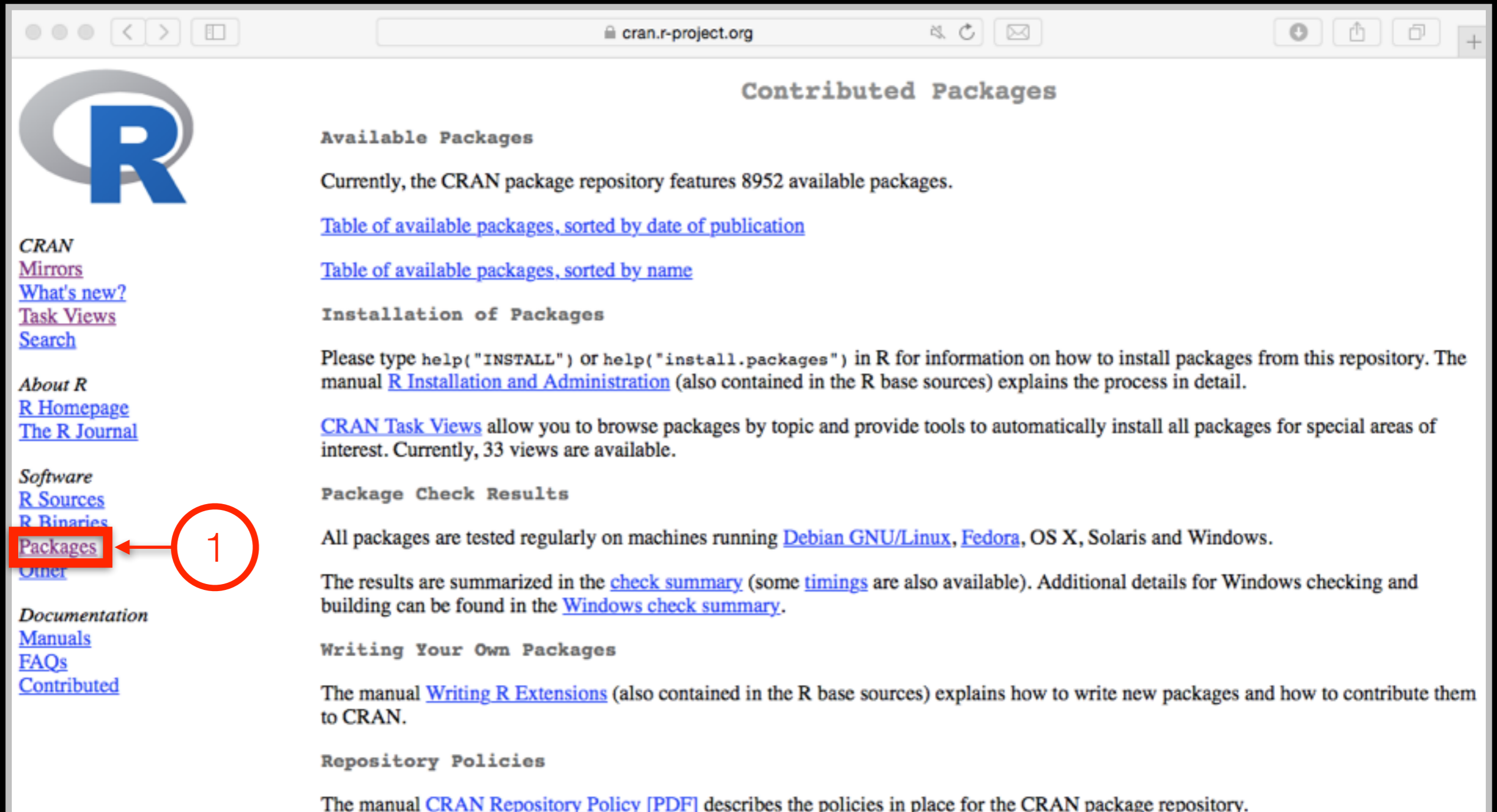
R packages can be of variable quality and often there are multiple packages with overlapping functionality.

**Refer to relevant publications, package citations, update/maintenance history, documentation quality and your own tests!**

“ The journal has sufficient experience with CRAN and Bioconductor resources to endorse their use by authors. We do not yet provide any endorsement for the suitability or usefulness of other solutions. ”

From: “Credit for Code”. *Nature Genetics* (2014), 46:1

# <https://cran.r-project.org>



The screenshot shows the CRAN website interface. The browser address bar displays `cran.r-project.org`. The main heading is **Contributed Packages**. The left sidebar contains a navigation menu with the following items: **CRAN**, [Mirrors](#), [What's new?](#), [Task Views](#), [Search](#), **About R**, [R Homepage](#), [The R Journal](#), **Software**, [R Sources](#), [R Binaries](#), [Packages](#) (highlighted with a red box), [Other](#), **Documentation**, [Manuals](#), [FAQs](#), and [Contributed](#). The main content area is titled **Contributed Packages** and includes sections for **Available Packages** (8952 packages), **Installation of Packages** (with instructions on using `help("INSTALL")` or `help("install.packages")`), **Package Check Results** (tested on Debian GNU/Linux, Fedora, OS X, Solaris, and Windows), and **Writing Your Own Packages** (with instructions on using `Writing R Extensions`).

# Installing a package

RStudio > Tools > Install Packages

```
> install.packages("bio3d")
```

```
> library("bio3d")
```

# Pick a package to explore and install

## Rmarkdown

- Reports, websites, documenting etc.: Promoting reproducibility.

## ggplot2

- Popular graphics package: We have already explored this.

## bio3d

- Widely used and highly cited structural bioinformatics package.

# Bioconductor

R packages and utilities for working with  
high-throughput genomic data

<http://bioconductor.org>





**More pragmatic:**

Bioconductor is a **software repository of R packages** with **some rules and guiding principles.**

**Version 3.3 had 1211 software packages.**

Bioconductor has  
emphasized

**Reproducible Research**

since its start, and has been  
an early adapter and driver  
of tools to do this.

“Bioconductor: open software development for computational biology and bioinformatics”

Gentleman et al

Genome Biology 2004, 5:R80

“Orchestrating high-throughput genomic analysis with Bioconductor”

Huber et al

Nature Methods 2015, 12:115-121

# Installing a bioconductor package

```
> source("https://bioconductor.org/biocLite.R")  
> biocLite()  
> biocLite("GenomicFeatures")
```

See: <http://www.bioconductor.org/install/>

# Summary

- R is a powerful data programming language and environment for statistical computing, data analysis and graphics.
- Introduced R syntax and major R data structures (called vectors, matrices data.frames and lists).
- Demonstrated using R for exploratory data analysis and graphics.
- Introduced CRAN and Bioconductor package repositories.

# Learning Resources

- **TryR**. An excellent interactive online R tutorial for beginners.  
< <http://tryr.codeschool.com/> >
- **RStudio**. A well designed reference card for RStudio.  
< <https://help.github.com/categories/bootcamp/> >
- **DataCamp**. Online tutorials using R in your browser.  
< <https://www.datacamp.com/> >
- **R for Data Science**. A new O'Reilly book that will teach you how to do data science with R, by Garrett Grolemund and Hadley Wickham.  
< <http://r4ds.had.co.nz/> >