

Iago Mosqueira

Introduction to R (and computing)

July 11, 2012

## Why programming?

“Can one be a good data analyst without being a half-good programmer? The short answer to that is, ‘No’. The long answer to that is, ‘No!’.”

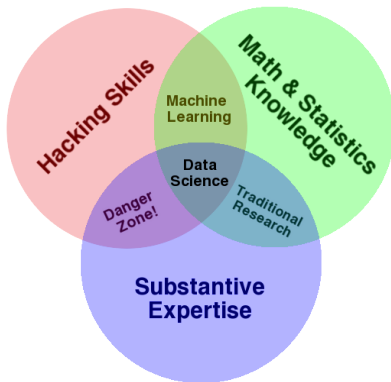
– Frank Harrell, 1999 S-PLUS User Conference, New Orleans (October 1999)

But this should be easy

“Managing fisheries is hard: it’s like managing a forest, in which the trees are invisible and keep moving around”

- Professor John Shepherd

# Data analyst

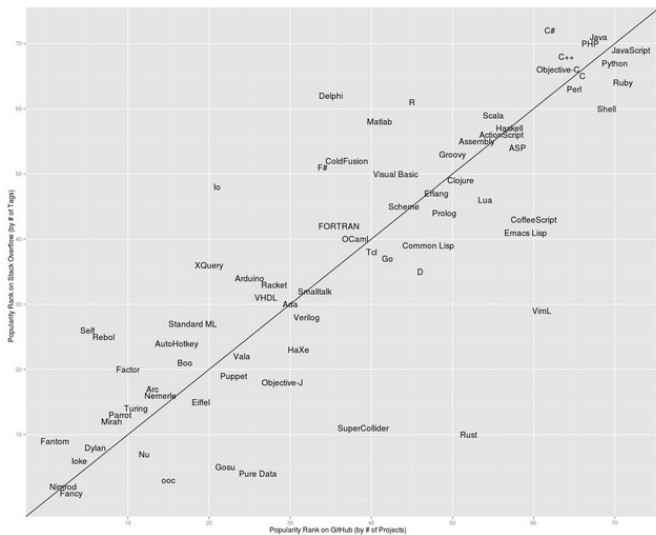


# What is R



- Data analysis and statistics environment
- Interpreted computer language
- Open-source software project
- Active community of developers and practitioners
- Current version: 2.15.1, 2012-06-22, Roasted Marshmallows

# Why R?



# GPL v3.0

## GNU General Public License

From Wikipedia, the free encyclopedia

"*GPL*" *redirects here*. For other uses, see *GPL (disambiguation)*.

The **GNU General Public License** (**GNU GPL** or simply **GPL**) is the most widely used<sup>[5]</sup> *free software license*. It was originally written by *Richard Stallman* for the *GNU Project*.

The GPL is the first *copyleft* license for general use, which means that derived works can only be distributed under the same license terms. Under this philosophy, the GPL grants the recipients of a computer program the rights of the *free software definition* and uses copyleft to ensure the freedoms are preserved, even when the work is changed or added to. This is in distinction to *permissive free software licenses*, of which the *BSD licenses* are the standard examples.

### Contents [hide]

- 1 History
- 2 Versions
  - 2.1 Version 1
  - 2.2 Version 2
  - 2.3 Version 3
- 3 Terms and conditions
  - 3.1 Copyleft
- 4 Licensing and contractual issues
- 5 Copyright holders
- 6 Linking and derived works
  - 6.1 Libraries
  - 6.2 Communicating and bundling with non-GPL programs
- 7 The GPL in court
- 8 Compatibility and multi-licensing
  - 8.1 Multi-licensing
- 9 Adoption
- 10 Use for text and other media

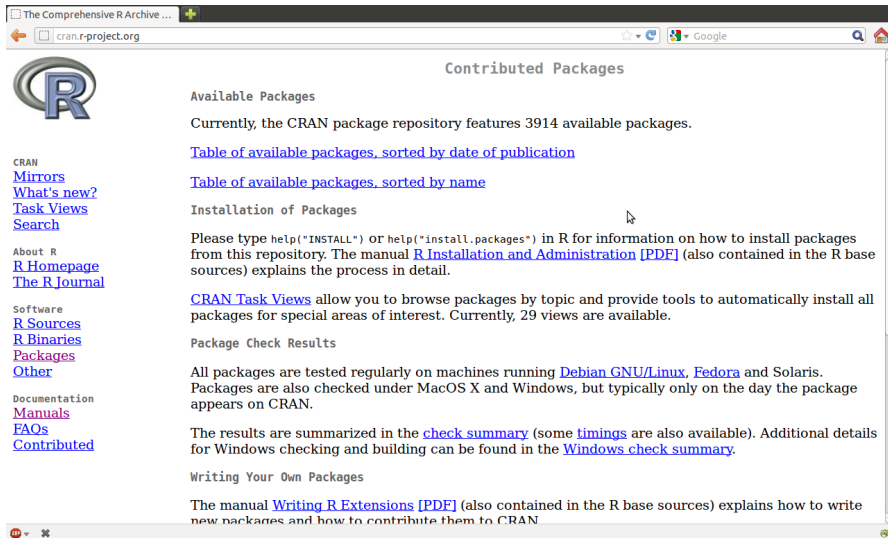
### GNU General Public License



**Free as in Freedom**

GNU GPLv3 Logo


<b>Author</b>	Free Software Foundation
<b>Version</b>	3
<b>Publisher</b>	Free Software Foundation, Inc.
<b>Published</b>	29 June 2007
<b>DFSG compatible</b>	Yes <sup>[1]</sup>
<b>FSF approved</b>	Yes <sup>[2]</sup>
<b>OSI approved</b>	Yes <sup>[3]</sup>
<b>Copyleft</b>	Yes <sup>[2][4]</sup>
<b>Linking from code with a different license</b>	No (except for linking GNU AGPLv3 with GNU GPLv3 – see section)
<b>Website</b>	<a href="http://www.gnu.org/licenses">www.gnu.org/licenses</a>

A screenshot of a web browser displaying the CRAN (Comprehensive R Archive Network) website. The browser's address bar shows 'cran.r-project.org'. The page features the CRAN logo on the left and a sidebar with navigation links. The main content area is titled 'Contributed Packages' and includes sections for 'Available Packages', 'Installation of Packages', 'Package Check Results', and 'Writing Your Own Packages'. Each section contains descriptive text and links to further resources.

The Comprehensive R Archive ...

cran.r-project.org

Google



## Contributed Packages

### Available Packages

Currently, the CRAN package repository features 3914 available packages.

[Table of available packages, sorted by date of publication](#)

[Table of available packages, sorted by name](#)

### Installation of Packages

Please type `help("INSTALL")` or `help("install.packages")` in R for information on how to install packages from this repository. The manual [R Installation and Administration \[PDF\]](#) (also contained in the R base sources) explains the process in detail.

[CRAN Task Views](#) allow you to browse packages by topic and provide tools to automatically install all packages for special areas of interest. Currently, 29 views are available.

### Package Check Results

All packages are tested regularly on machines running [Debian GNU/Linux](#), [Fedora](#) and Solaris. Packages are also checked under MacOS X and Windows, but typically only on the day the package appears on CRAN.

The results are summarized in the [check summary](#) (some [timings](#) are also available). Additional details for Windows checking and building can be found in the [Windows check summary](#).

### Writing Your Own Packages

The manual [Writing R Extensions \[PDF\]](#) (also contained in the R base sources) explains how to write new packages and how to contribute them to CRAN.

CRAN

- [Mirrors](#)
- [What's new?](#)
- [Task Views](#)
- [Search](#)

About R

- [R Homepage](#)
- [The R Journal](#)

Software

- [R Sources](#)
- [R Binaries](#)
- [Packages](#)
- [Other](#)


Documentation

- [Manuals](#)
- [FAQs](#)
- [Contributed](#)



# Task views


Blog with Knitr and Je... Welcome | Travelling... iagomosqueira/ICESSAI Home - TCSAI2012 Oráculo manual y art... The Comprehensive R...  
cran.r-project.org ggplot2



CRAN Task Views

<a href="#">CRAN</a>	<a href="#">Bayesian</a>	Bayesian Inference
<a href="#">Mirrors</a>	<a href="#">ChemPhys</a>	Chemometrics and Computational Physics
<a href="#">What's new?</a>	<a href="#">ClinicalTrials</a>	Clinical Trial Design, Monitoring, and Analysis
<a href="#">Task Views</a>	<a href="#">Cluster</a>	Cluster Analysis & Finite Mixture Models
<a href="#">Search</a>	<a href="#">DifferentialEquations</a>	Differential Equations
	<a href="#">Distributions</a>	Probability Distributions
	<a href="#">Econometrics</a>	Computational Econometrics
	<a href="#">Environmetrics</a>	Analysis of Ecological and Environmental Data
	<a href="#">ExperimentalDesign</a>	Design of Experiments (DoE) & Analysis of Experimental Data
<a href="#">About R</a>	<a href="#">Finance</a>	Empirical Finance
<a href="#">R Homepage</a>	<a href="#">Genetics</a>	Statistical Genetics
<a href="#">The R Journal</a>	<a href="#">Graphics</a>	Graphic Displays & Dynamic Graphics & Graphic Devices & Visualization
	<a href="#">HighPerformanceComputing</a>	High-Performance and Parallel Computing with R
<a href="#">Software</a>	<a href="#">MachineLearning</a>	Machine Learning & Statistical Learning
<a href="#">R Sources</a>	<a href="#">MedicalImaging</a>	Medical Image Analysis
<a href="#">R Binaries</a>	<a href="#">Multivariate</a>	Multivariate Statistics
<a href="#">Packages</a>	<a href="#">NaturalLanguageProcessing</a>	Natural Language Processing
<a href="#">Other</a>	<a href="#">OfficialStatistics</a>	Official Statistics & Survey Methodology
	<a href="#">Optimization</a>	Optimization and Mathematical Programming
<a href="#">Documentation</a>	<a href="#">Pharmacokinetics</a>	Analysis of Pharmacokinetic Data
<a href="#">Manuals</a>	<a href="#">Phylogenetics</a>	Phylogenetics, Especially Comparative Methods
<a href="#">FAQs</a>	<a href="#">Psychometrics</a>	Psychometric Models and Methods
<a href="#">Contributed</a>	<a href="#">ReproducibleResearch</a>	Reproducible Research
	<a href="#">Robust</a>	Robust Statistical Methods
	<a href="#">SocialSciences</a>	Statistics for the Social Sciences
	<a href="#">Spatial</a>	Analysis of Spatial Data
	<a href="#">Survival</a>	Survival Analysis
	<a href="#">TimeSeries</a>	Time Series Analysis


# RStudio

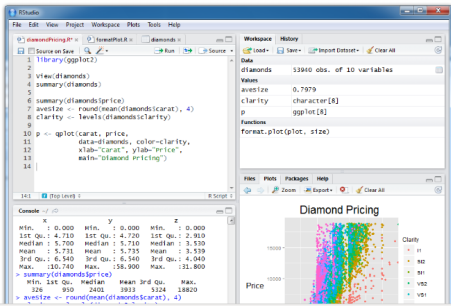


[Home](#) [Screenshots](#) [Download](#) [Docs](#) [Support](#) [Development](#) [Blog](#)


## Welcome to RStudio

RStudio™ is a free and open source integrated development environment (IDE) for [R](#). You can run it on your desktop (Windows, Mac, or Linux) or even over the web using RStudio Server.

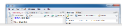
 **Download RStudio**  
for Windows, Mac or Linux



The screenshot shows the RStudio IDE interface. The top menu bar includes File, Edit, View, Project, Workspace, Plots, Tools, and Help. The left pane shows a script editor with R code for loading the 'diamonds' dataset and creating a scatter plot. The bottom-left pane shows the console output, including summary statistics for the 'diamonds' dataset. The bottom-right pane displays a scatter plot titled 'Diamond Pricing' with 'Price' on the y-axis and 'Clarity' on the x-axis. The plot shows a positive correlation between price and clarity, with points colored by clarity levels (I1, SI1, VS2, VS1).



This screenshot shows a close-up of the RStudio script editor, displaying R code for loading the 'diamonds' dataset and creating a scatter plot. The code includes comments and function calls like 'library(ggplot2)', 'data=diamonds', and 'ggplot(diamonds, aes(carat, price))'.



This screenshot shows a close-up of the RStudio console window, displaying the output of the R code executed in the script editor. The output includes summary statistics for the 'diamonds' dataset, such as the number of observations, the range of carat and price, and the distribution of clarity levels.

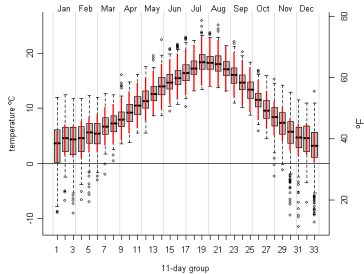
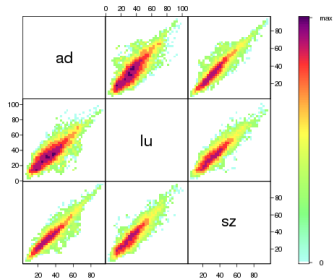
## Basic features

- Numerous procedures (algebra, matrix, stats)
- Named storage (everything is an object)
- Functions
- Classes and methods (S3, S4)
- Special values (NA, NaN, Inf, NULL)
- Logical objects and boolean algebra
- `basic_features.R`

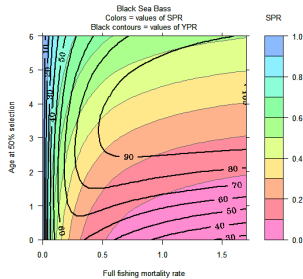
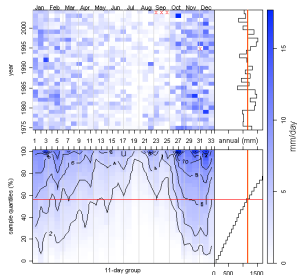
## What else can it do?

- Data handling and storage
- Matrix algebra
- Regular expressions
- Statistics!
- OOP
- Programming
- Graphics

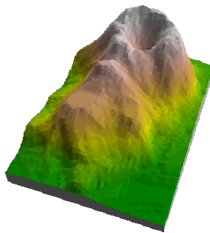
# Eye candy



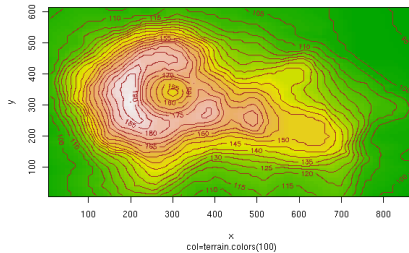
# Eye candy



# Eye candy



*Maunga Whau Volcano*

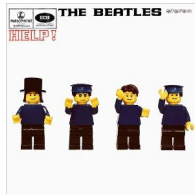


## What doesn't it do

- No DB, but connections (SQL, NoSQL, Spreadsheets)
- No GUI, but IDE & GUI toolsets - CLI
- Slow, but C/C++, HPC
- No commercial support, but community
- Think for you



# Help!



- Help for each function and data type
- ?mean
- ??mean
- ?help
- <http://rseek.org>
- stackoverflow, <http://stackoverflow.com/questions/tagged/r>
- Mailing lists

- Stock assessment and provision of management advice
  - ▶ Well tested, robust methods
  - ▶ Open to detailed inspection
- Data and model validation through simulation
- Risk analysis
- Capacity development & education
- Promote collaboration and openness in quantitative fisheries science
  - ▶ Open source
  - ▶ Community involvement
  - ▶ R as lingua franca
- Support the development of new models and methods
  - ▶ Extensible toolset
  - ▶ Links to other tools (ADMB, BUGS, ...)


start - FLR Project

flr-project.org/doku.php?id=start

Google

# FLR PROJECT

## TABLE OF CONTENTS




### Training Course

#### Intro to R and FLR for Fisheries


3-5 JULY 2012 – Varese, Italy

Registration now open at [FishReg EC JRC](#)



The FLR library is a collection of tools in the [R](#) statistical language that facilitates the construction of bio-economic simulation models of fisheries and ecological systems. It is a generic toolbox, but is specifically suited for the construction of simulation models for evaluations of fisheries management strategies. The FLR library is under development by researchers across a number of laboratories and universities.

*Love it !! Take me to the [Table of Contents](#) !!*




You can install the **stable releases** for [R >= 2.13.0](#) with

```
install.packages(repos="http://flr-project.org/R")
```

or you can install the **development packages** for [R >= 2.14.0](#) with

```
install.packages(repos="http://flr-project.org/Rdevel")
```



You can get help by subscribing our [mailing list](#) and posting your doubt, or reading the **tutorials** on the ["Teach yourself FLR" wiki](#).

# Tools of the trade

- Version Control Systems
  - ▶ CVS
  - ▶ SVN
  - ▶ git
- Editors & IDEs
- Literate Programming
  - ▶ Sweave
  - ▶ knitr
- Validation, Verification and Testing (VV&T)

# Sexy data analysis



# Setting up R & RStudio

- <http://cran.r-project.org>
- <http://rstudio.org>