

Iago Mosqueira

Introduction to R (and computing)

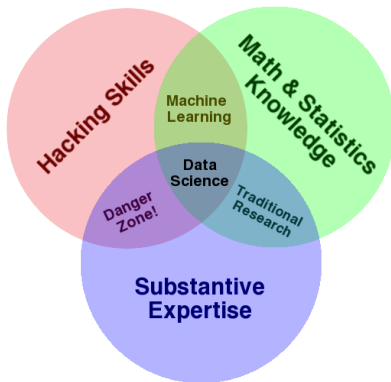
July 8, 2012

## 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)

# Data analyst

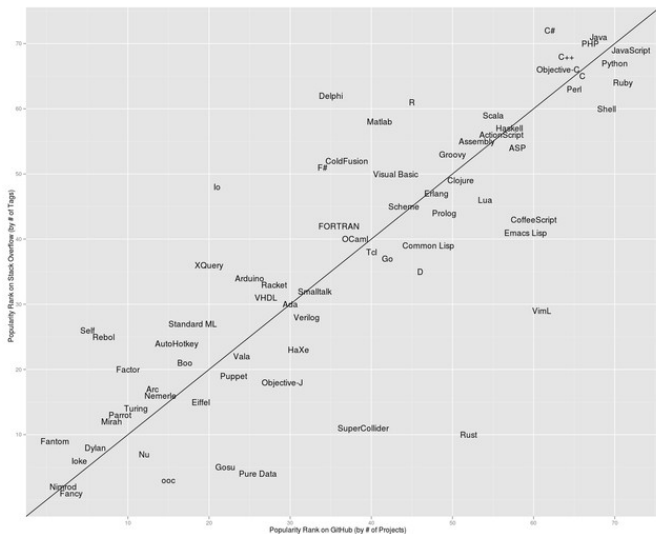


# What is R



- 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

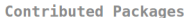
### GNU General Public License



**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>



## 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.

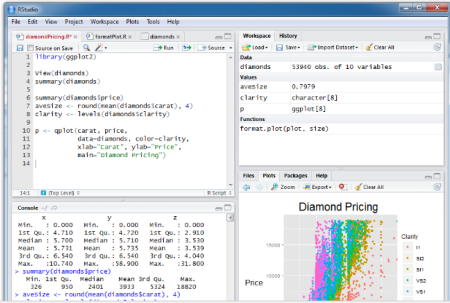
# RStudio

[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 with the following components:

- Source Editor:** Contains R code for loading ggplot2, viewing the 'diamonds' dataset, and creating a scatter plot of Price vs. Clarity.
- Environment/History:** Shows the 'diamonds' dataset with 53940 observations and 10 variables.
- Console:** Displays the output of the R code, including summary statistics for the 'diamonds' dataset.
- Plots:** A scatter plot titled 'Diamond Pricing' showing Price (Y-axis) versus Clarity (X-axis).



This is a smaller screenshot showing a different view of the RStudio interface, likely a video frame from a screencast.

 **Screencast**  
RStudio in 2 minutes



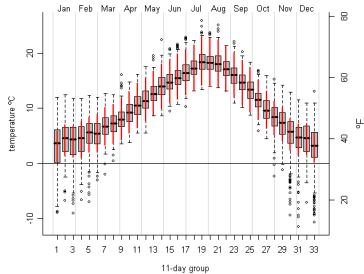
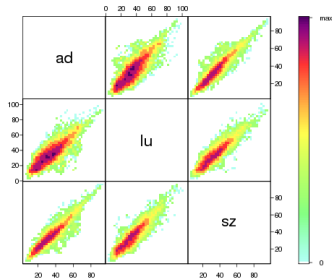
## Basic features

- 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`

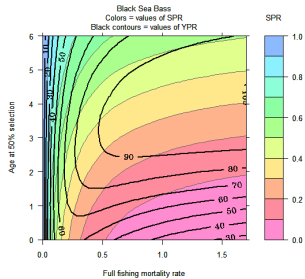
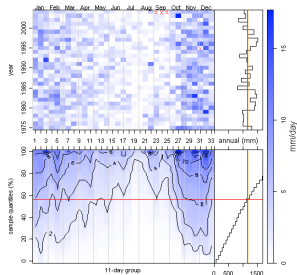
## What else can it do?

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

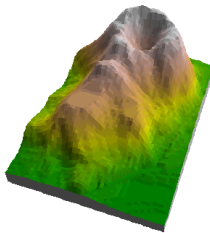
# Eye candy



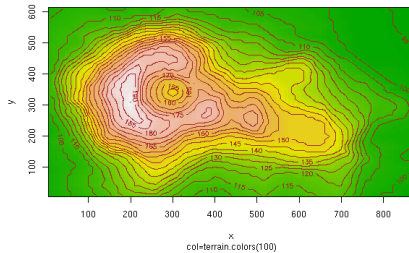
# Eye candy



# Eye candy



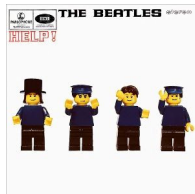
*Maunga Whau Volcano*



## 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>