Quantitative fisheries science in R and FLR

Iago Mosqueira
Systems Modeling
iago.mosqueira@cefas.co.uk

Cefas

http://r-project.org

Cefas

What is R

R is a scripting language for statistical data manipulation and analysis. It was inspired by, and is mostly compatible with, the statistical language S developed by AT&T. The name S, obviously standing for statistics, was an allusion to another programming language developed at AT&T with a one-letter name, C.

Cefas

Why R

- yauh peng, yauh leng - “both inexpensive and beautiful.”
- Open source (GPL) and free
- Cross-platform
- object-oriented and functional programming structure
- Interactive and batch modes
- Huge user/developer base

Cefas

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 a widely used free software license, originally written by Richard Stallman for the GNU project. The GPL is the most popular and well-known example of the type of strong copyleft license that requires derived works to be available under the same copyleft. 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.

The GNU Lesser General Public License (LGPL) is a modified, more permissive, version of the GPL, originally intended for some software libraries. There is also a GNU Free Documentation License, which was originally intended for use with documentation for GNU software, but has also been adopted for other uses, such as the Wikipedia project.

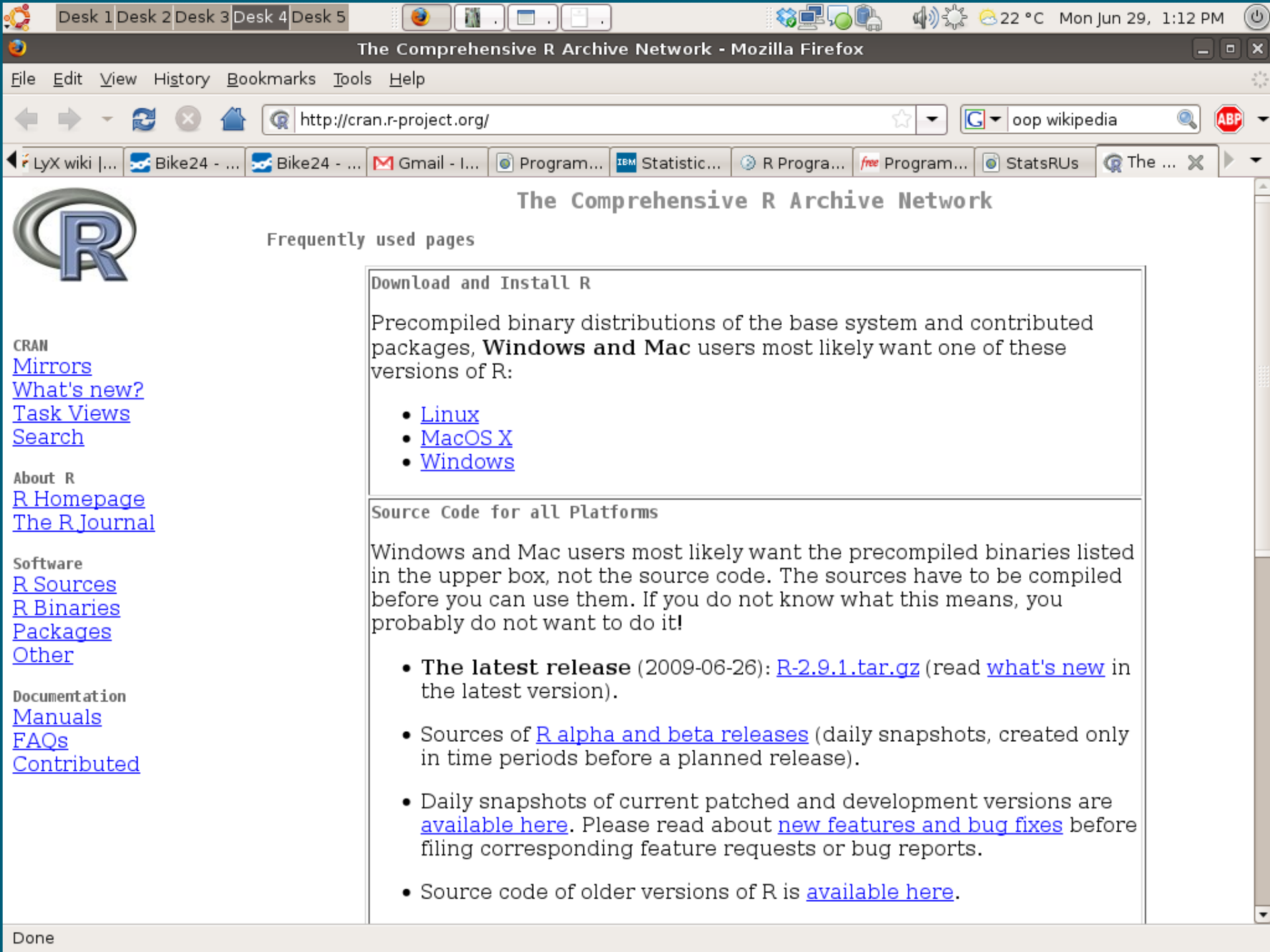
The Affero General Public License (GNU AGPL) is a similar license with a focus on networking server software. The GNU AGPL is similar to the GNU General Public License, except that it additionally covers the use of the software over a computer network, requiring that the complete source code be made available to any network user of the AGPLed work, for example a web application. The Free Software Foundation recommends that this license is considered for any software that will commonly be run over the network.

Contents [hide]

GNU General Public License

GNU GPLv3 Logo

Author	Free Software Foundation
Version	3
Publisher	Free Software Foundation, Inc.
Published	June 29, 2007
DFSG compatible	Yes
Free software	Yes
OSI approved	Yes
Copyleft	Yes
Linking from code with a different license	No


Desk 1 Desk 2 Desk 3 Desk 4 Desk 5
22 °C Mon Jun 29, 1:12 PM
The Comprehensive R Archive Network - Mozilla Firefox
File Edit View History Bookmarks Tools Help
http://cran.r-project.org/
oop wikipedia
LyX wiki |...
Bike24 - ...
Bike24 - ...
Gmail - I...
Program...
Statistic...
R Progra...
Program...
StatsRUs
The ...
The Comprehensive R Archive Network
Frequently used pages
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
Download and Install R
Precompiled binary distributions of the base system and contributed packages, Windows and Mac users most likely want one of these versions of R:
Linux
MacOS X
Windows
Source Code for all Platforms
Windows and Mac users most likely want the precompiled binaries listed in the upper box, not the source code. The sources have to be compiled before you can use them. If you do not know what this means, you probably do not want to do it!
The latest release (2009-06-26): R-2.9.1.tar.gz (read what's new in the latest version).
Sources of R alpha and beta releases (daily snapshots, created only in time periods before a planned release).
Daily snapshots of current patched and development versions are available here. Please read about new features and bug fixes before filing corresponding feature requests or bug reports.
Source code of older versions of R is available here.
Done

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

Contributed Packages

Installation of Packages

Please type `help("INSTALL")` or `help("install.packages")` in R for information on how to install packages from this directory. The manual R Installation and Administration (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, 24 views are available.

Daily Package Check Results

All packages are tested regularly on machines running Debian GNU/Linux. Packages are also checked under MacOS X and Windows, but only at 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.

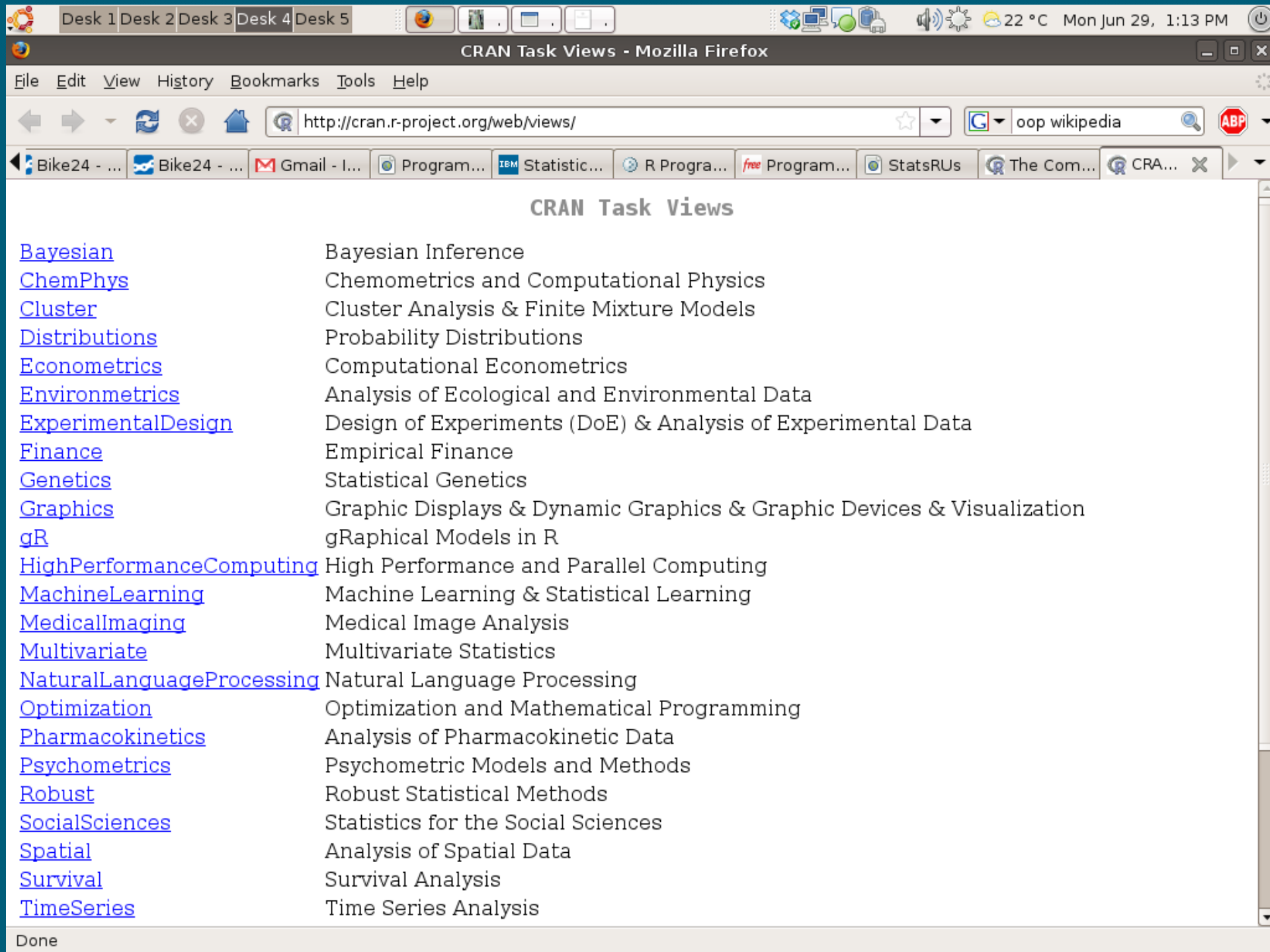
Writing Your Own Packages

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

Available Bundles and Packages

Currently, the CRAN package repository features 1850 objects including 1843 packages and 7 bundles containing 26 packages, for a total of 1869 available packages.

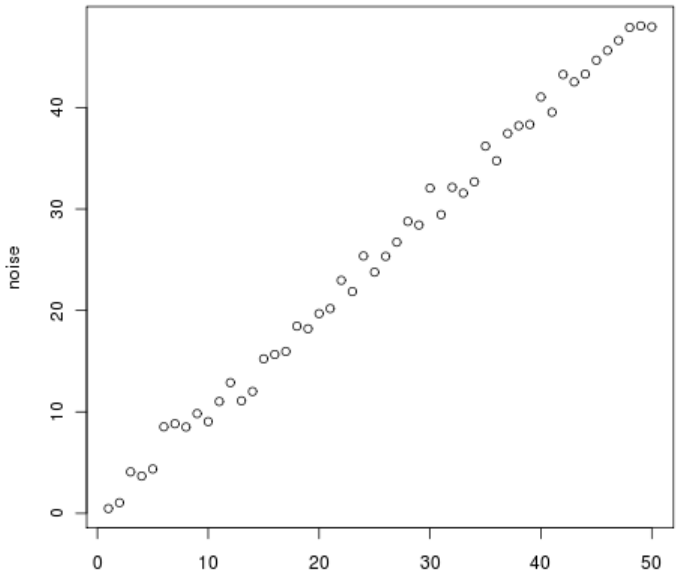
A B C D E F G H I J K L M N O P Q R S T U V W X Y Z

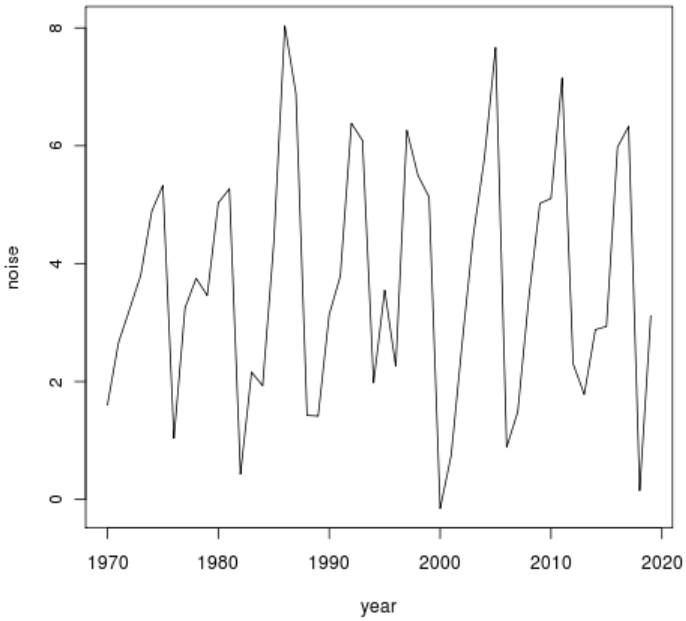

Desk 1 Desk 2 Desk 3 Desk 4 Desk 5
22 °C Mon Jun 29, 1:13 PM
CRAN Task Views - Mozilla Firefox
File Edit View History Bookmarks Tools Help
http://cran.r-project.org/web/views/
oop wikipedia
Bike24 - ... Bike24 - ... Gmail - I... Program... Statistic... R Progra... Program... StatsRUs The Com... CRA...
CRAN Task Views
Bayesian Bayesian Inference
ChemPhys Chemometrics and Computational Physics
Cluster Cluster Analysis & Finite Mixture Models
Distributions Probability Distributions
Econometrics Computational Econometrics
Environmetrics Analysis of Ecological and Environmental Data
ExperimentalDesign Design of Experiments (DoE) & Analysis of Experimental Data
Finance Empirical Finance
Genetics Statistical Genetics
Graphics Graphic Displays & Dynamic Graphics & Graphic Devices & Visualization
gR gRaphical Models in R
HighPerformanceComputing High Performance and Parallel Computing
MachineLearning Machine Learning & Statistical Learning
MedicalImaging Medical Image Analysis
Multivariate Multivariate Statistics
NaturalLanguageProcessing Natural Language Processing
Optimization Optimization and Mathematical Programming
Pharmacokinetics Analysis of Pharmacokinetic Data
Psychometrics Psychometric Models and Methods
Robust Robust Statistical Methods
SocialSciences Statistics for the Social Sciences
Spatial Analysis of Spatial Data
Survival Survival Analysis
TimeSeries Time Series Analysis
Done

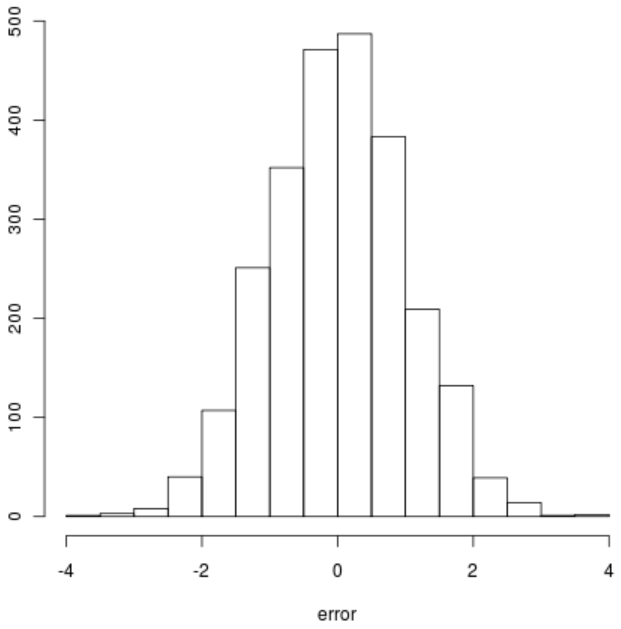
Basic features

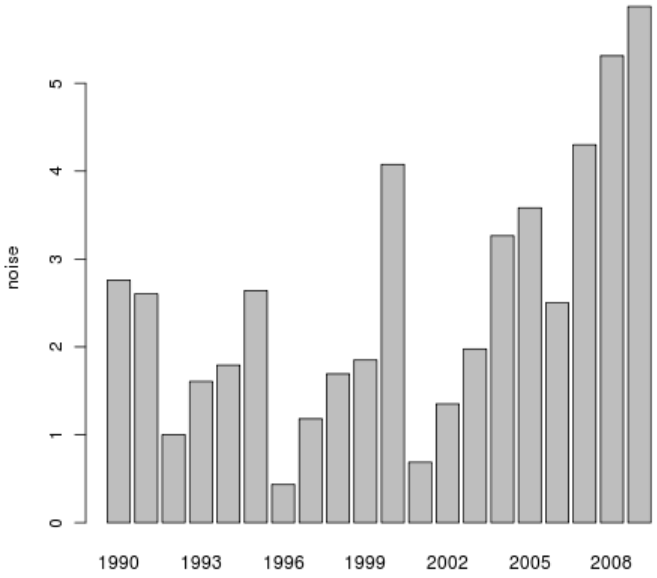
- Numerous procedures (algebra, matrix, stats)
- Named storage (everything is an object)
- Functions
- Classes and methods (S3, S4)
  - Vectors, matrices and arrays
  - Lists and data.frames
- Special values (NA, NaN, Inf, NULL)
- Logical vectors and boolean algebra
- `basic_features.R`

Cefas

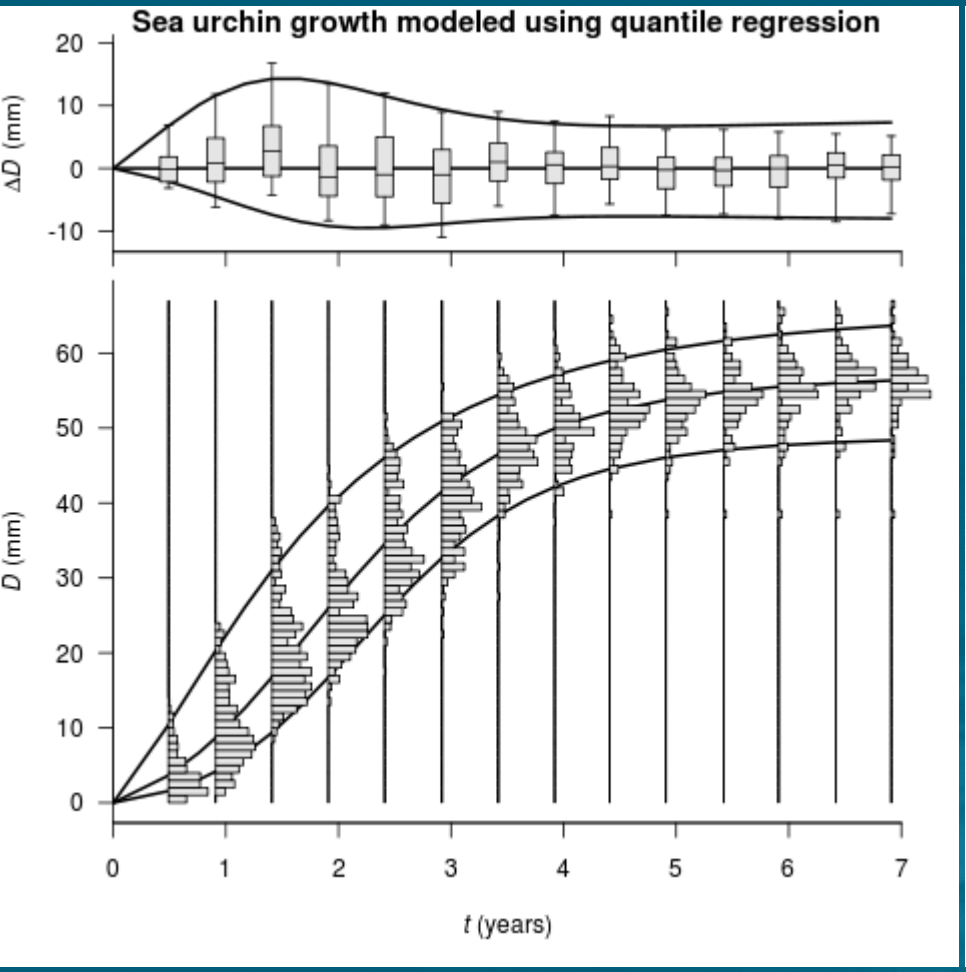

noise
0
10
20
30
40
0
10
20
30
40
50

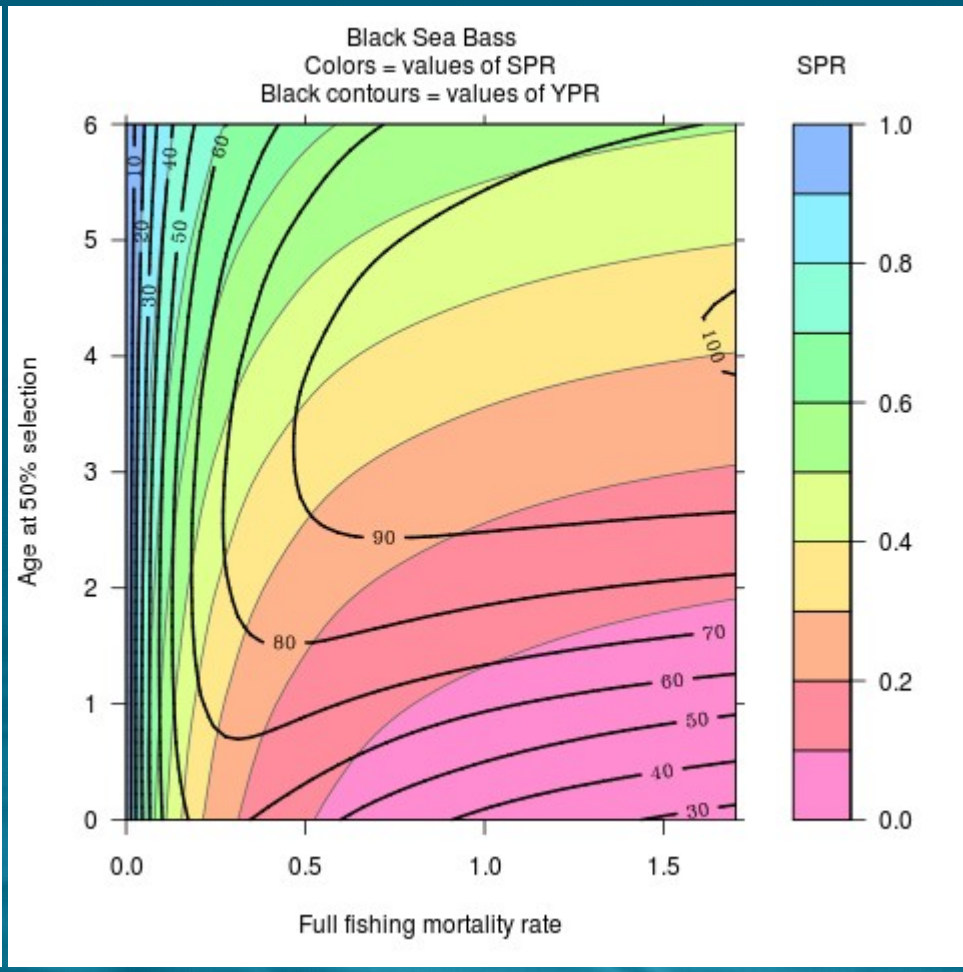

noise
0
2
4
6
8
1970
1980
1990
2000
2010
2020
year

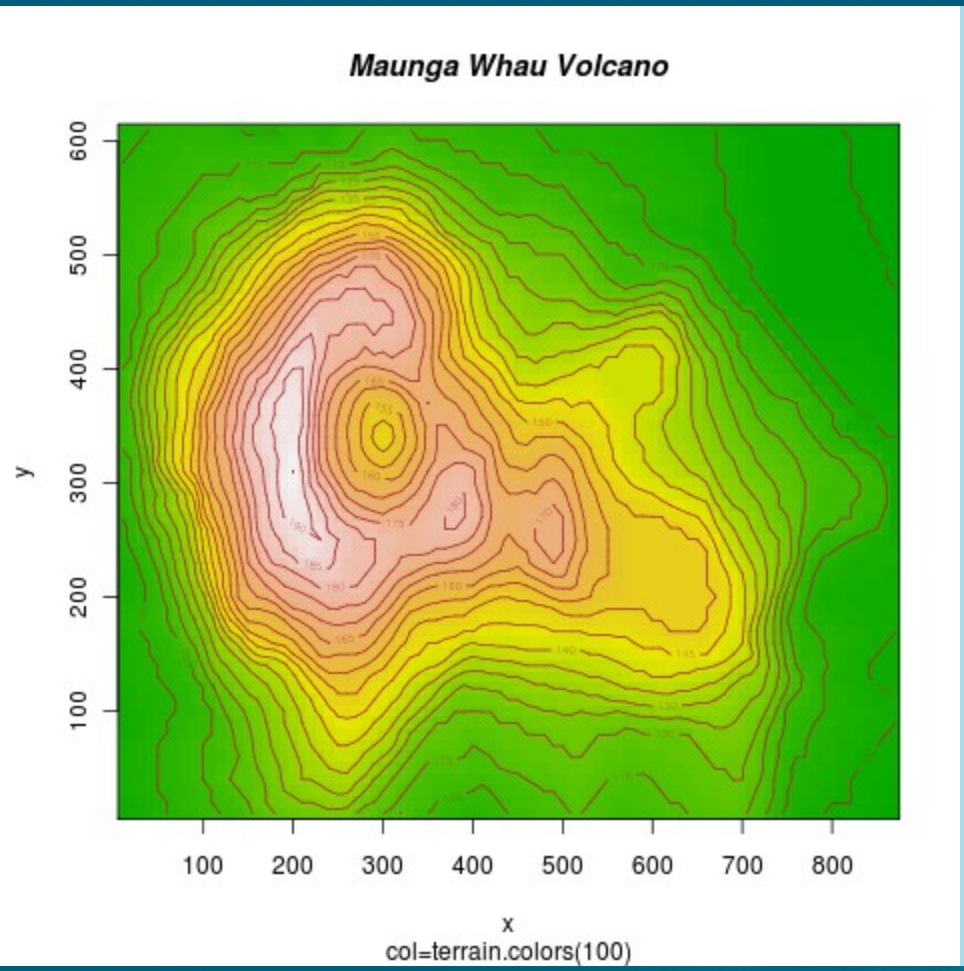

0
100
200
300
400
500
-4
-2
0
2
4
error

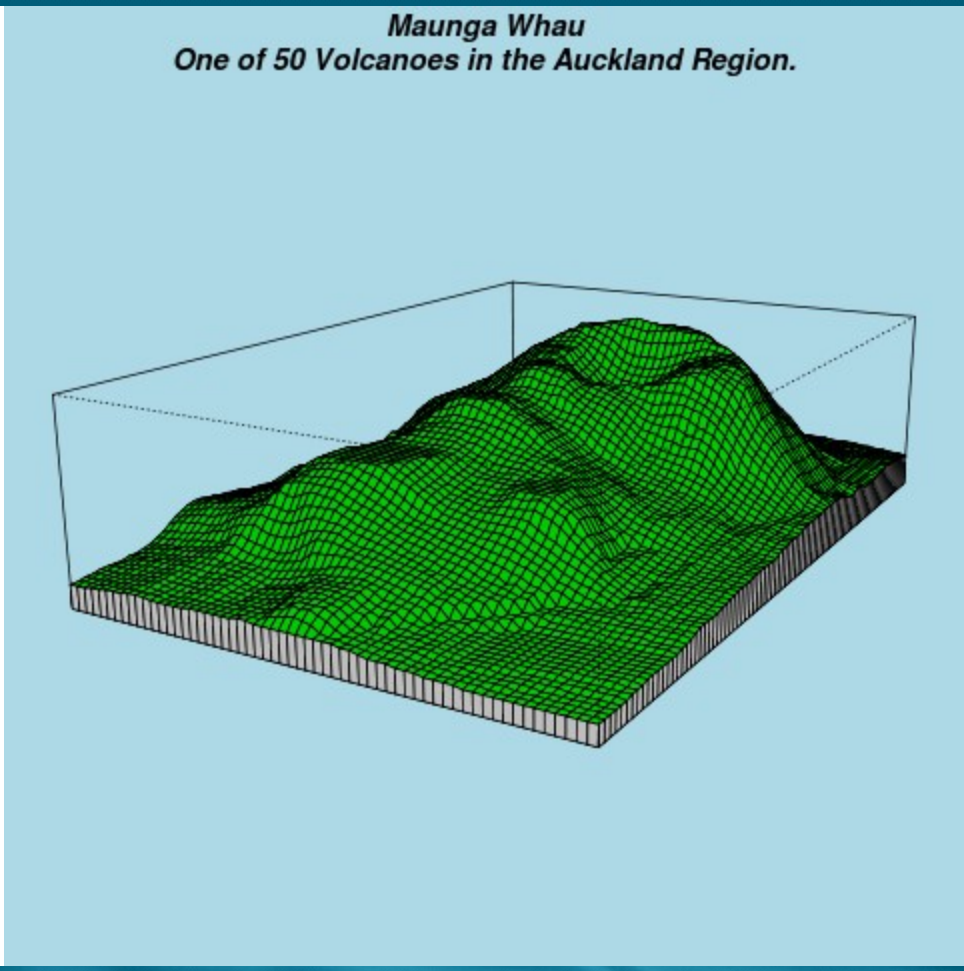

noise
0
1
2
3
4
5
1990
1993
1996
1999
2002
2005
2008

fas

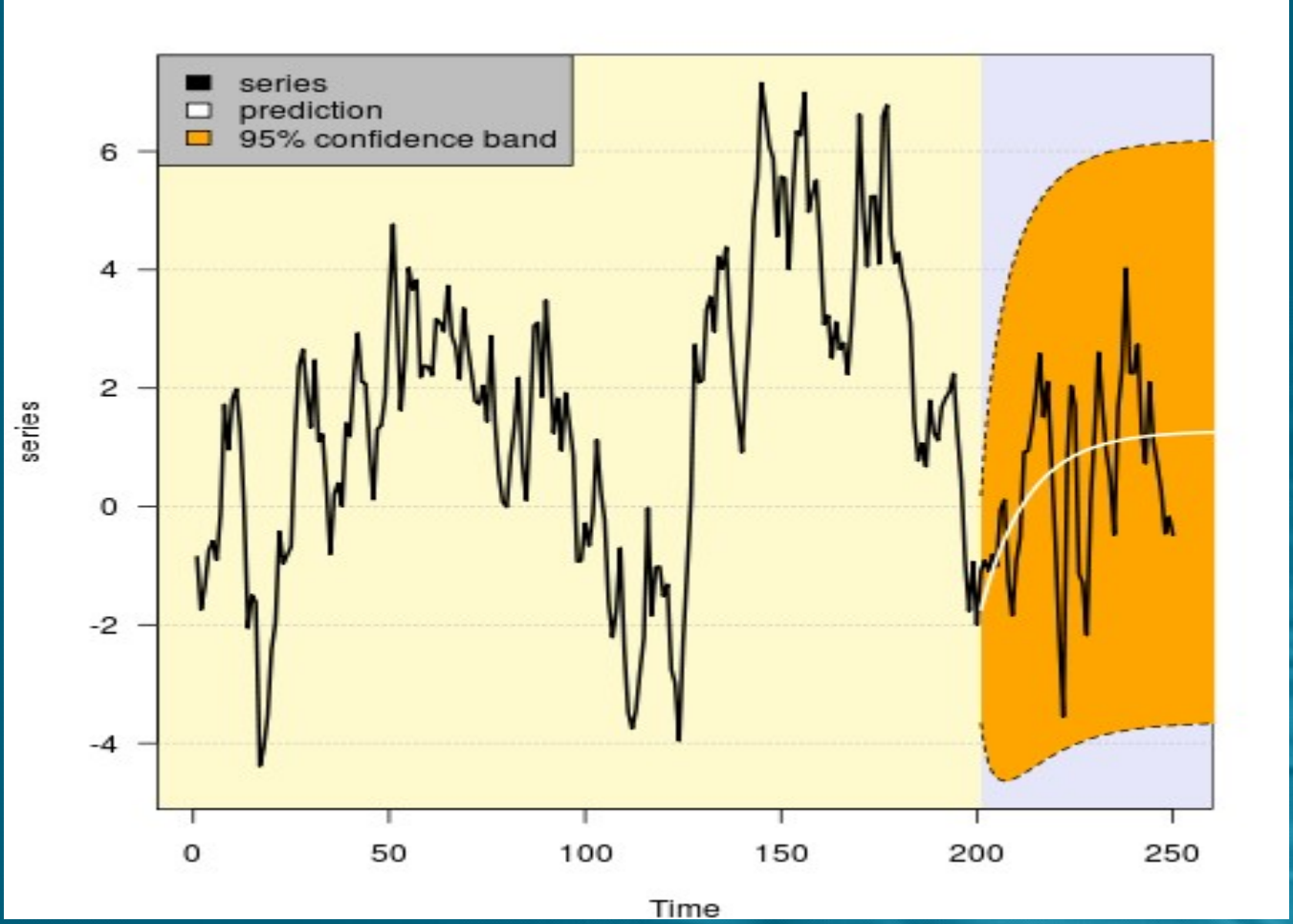

Sea urchin growth modeled using quantile regression
ΔD (mm)
D (mm)
t (years)

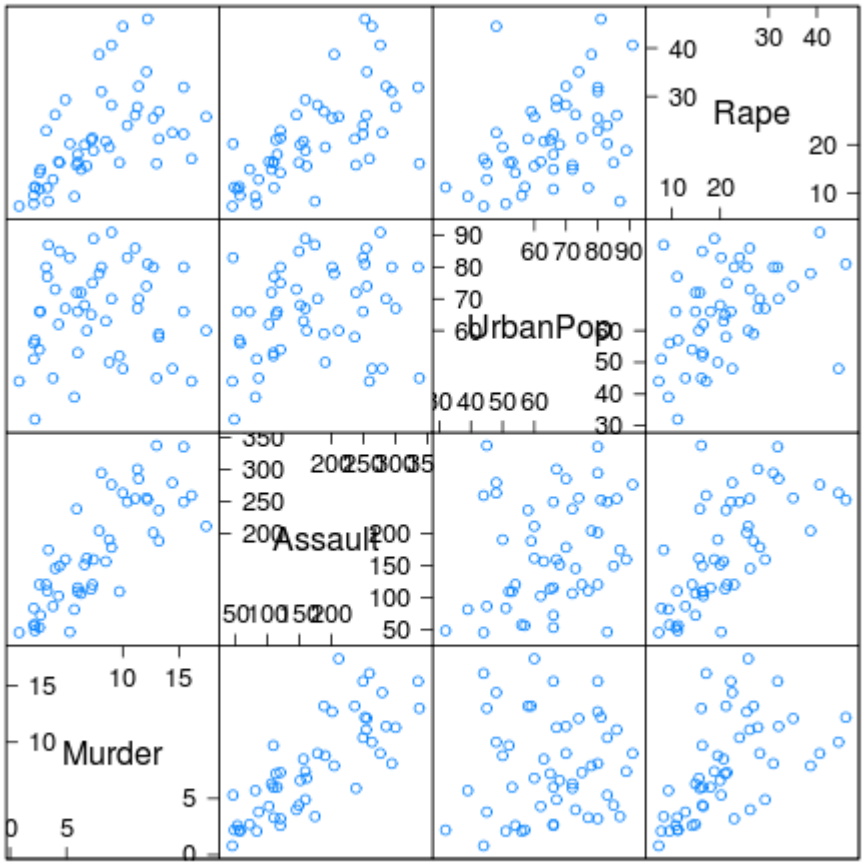

Black Sea Bass
Colors = values of SPR
Black contours = values of YPR
SPR
Age at 50% selection
Full fishing mortality rate


Maunga Whau Volcano
y
100
200
300
400
500
600
100
200
300
400
500
600
700
800
x
col=terrain.colors(100)

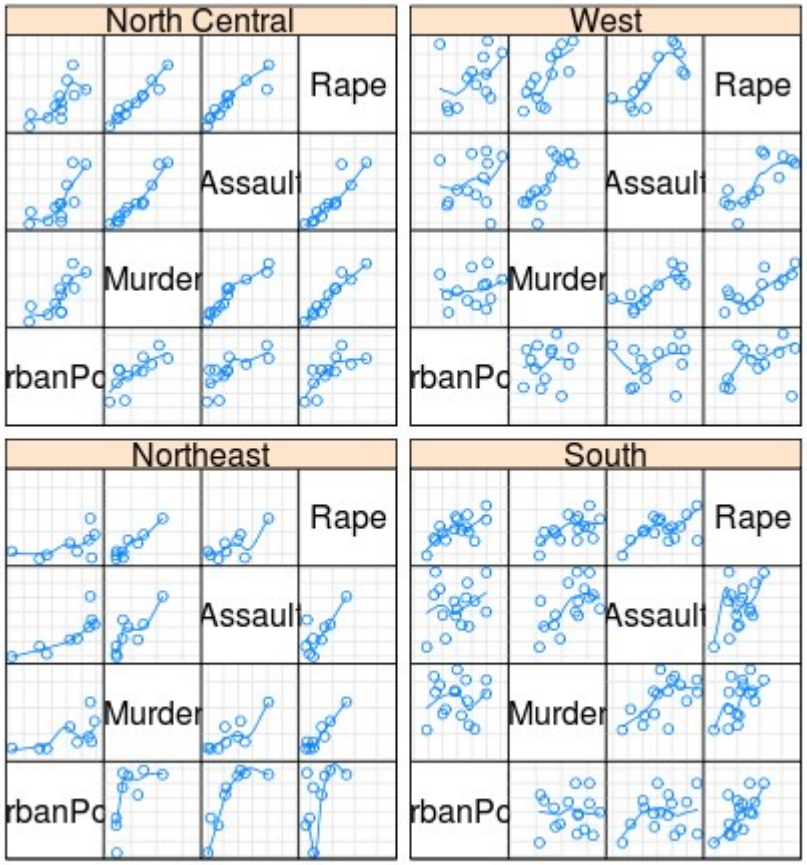

Maunga Whau
One of 50 Volcanoes in the Auckland Region.

Cefas


series
prediction
95% confidence band
6
4
2
0
-2
-4
series
0
50
100
150
200
250
Time

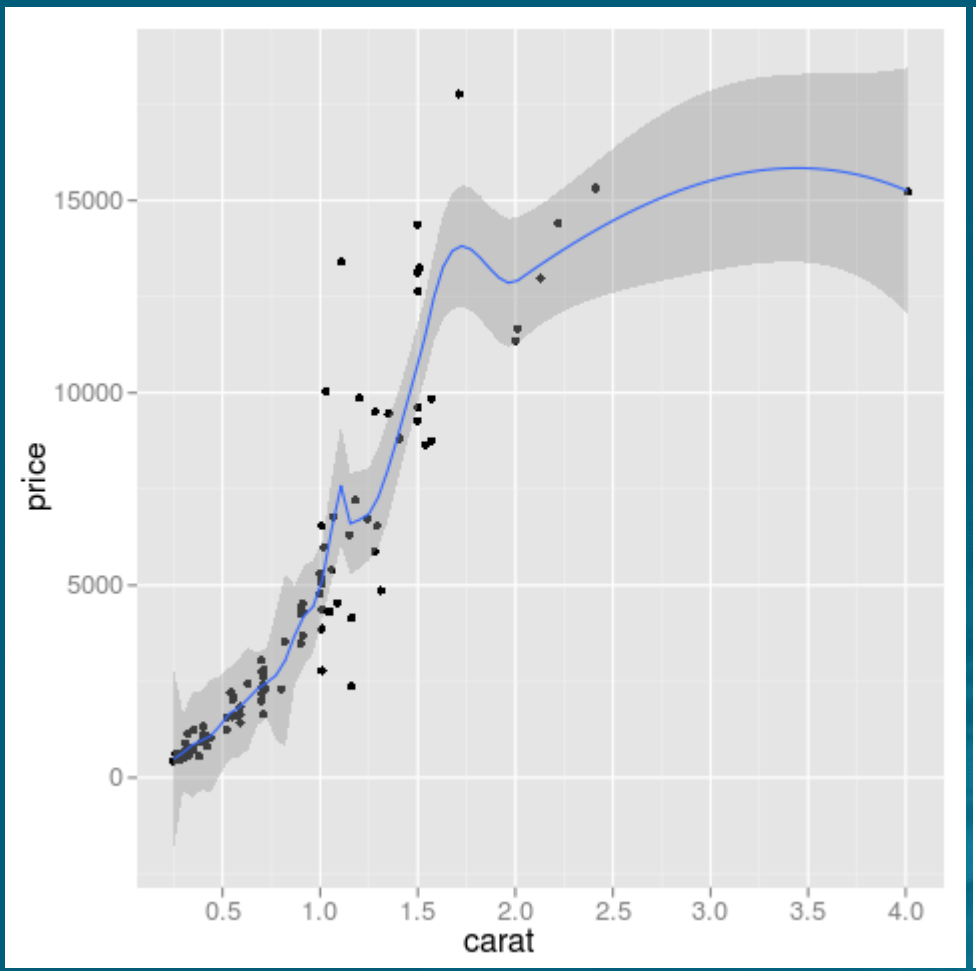


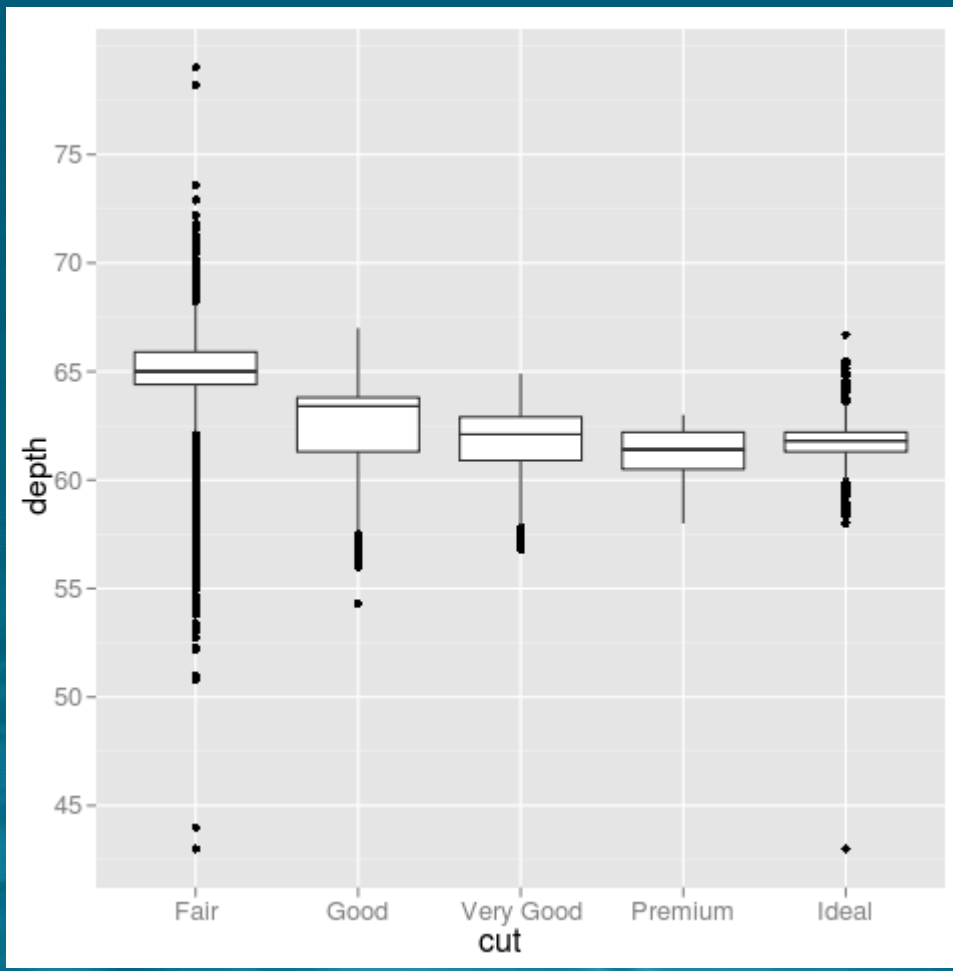
Scatter Plot Matrix



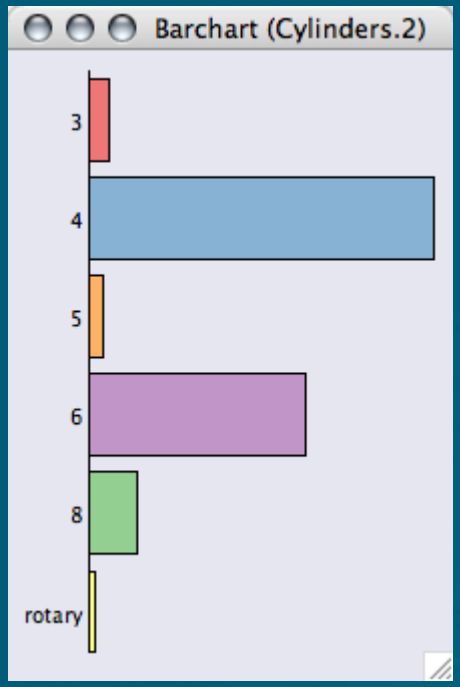
Scatter Plot Matrix

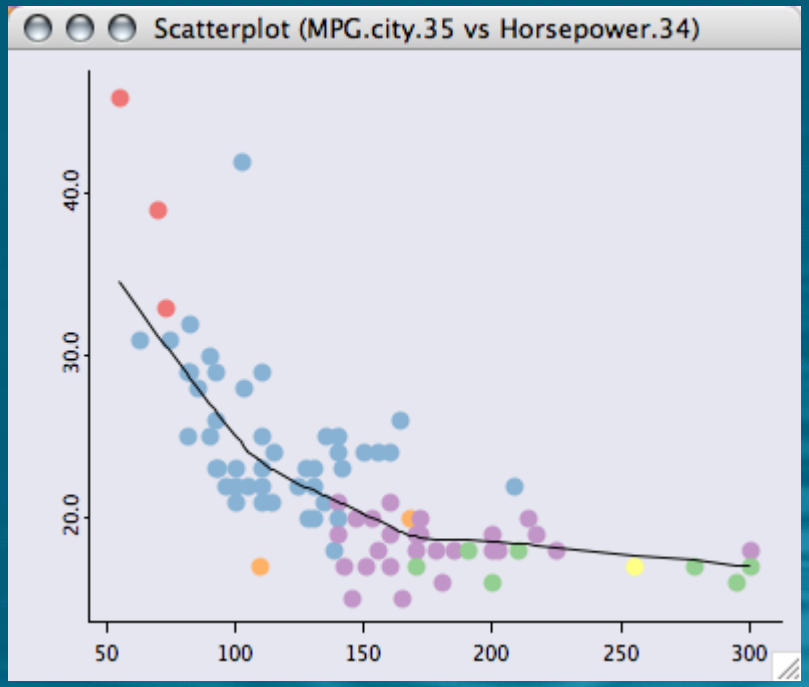
Cefas


price
15000
10000
5000
0
0.5
1.0
1.5
2.0
2.5
3.0
3.5
4.0
carat


depth
75
70
65
60
55
50
45
Fair
Good
Very Good
Premium
Ideal
cut

Cefas


Barchart (Cylinders.2)
3
4
5
6
8
rotary


Scatterplot (MPG.city.35 vs Horsepower.34)
40.0
30.0
20.0
50
100
150
200
250
300

Cefas

help!

- help("function") or ?function
- example("function")
- demos and vignettes
- help.search("keyword")
- http://rseek.org
- R-help and R-devel mailing lists

Cefas

pearson Search

Results **1** - **10** for **pearson**. (**0.35** seconds)

Google™ Custom Search

R: **Pearson**-type goodness-of-fit test

20 Oct 2008 **...** The fourth element of the list, "test" , is a data frame with one row containing the **Pearson**-type goodness-of-fit test statistic stat . **...**

bm2.genes.nig.ac.jp/RGM2/R_current/library/msm/... /**pearson**.msm.html

pearson.dat

65.04851 59.77827 63.25094 63.21404 64.95532 63.34242 65.75250 62.79238 61.13723 64.28113 63.02254 64.24221 65.37053 64.08231 64.72398 63.99574 66.06509 **...**

www.stat.columbia.edu/~gelman/arm/examples/correlation/**pearson**.dat

R: **Pearson's** Chi-squared Test for Contingency Tables

Then, **Pearson's** chi-squared test of the null hypothesis that the joint distribution of the cell counts in a 2-dimensional contingency table is the product **...**

www.rforge.net/doc/packages/Deducer/chi.squared.test.html

Pearson's Chi-square Test for Count Data

Then, **Pearson's** chi-square test of the null that the joint distribution of the cell counts in a 2-dimensional contigency table is the product of the row and **...**

www.math.montana.edu/Rweb/Rhelp/chisq.test.html

R: Plot Precision of Estimate of **Pearson** Correlation Coefficient

Plot Precision of Estimate of **Pearson** Correlation Coefficient. Description. This function plots the precision (margin of error) of the product-moment linear **...**

bm2.genes.nig.ac.jp/RGM2/R_current/library/.../plotCorrPrecision.html

Introductions | Task Views | Support Lists | Functions | Books | Related Tools

R Data Analysis Examples: Multinomial Logistic Regression

Next we see information on the distribution of the **pearson's** residuals which help us assess how well the model fits. There are two sets of these residuals, **...**

www.ats.ucla.edu/stat/r/dae/mlogit.htm

Quick-R: Correlations

The rcorr( ) function in the Hmisc package produces correlations/covariances and significance levels for **pearson** and spearman correlations. **...**

www.statmethods.net/stats/correlations.html

Regression * Model behind the regression + Model and assumptions **...**

n <- 10 > x <- rnorm(n) > y <- rnorm(n) > cor(x,y) [1] -0.4132864 > cor.test(x,y) **Pearson's** product-moment correlation data: x and y t = -1.2837, df = 8, **...**

zoonek2.free.fr/UNIX/48_R/09_Regression.txt

Quick-R: Search Page

Description: **Pearson**, Spearman, Kendall, Polyserial, Polychoric Correlations ... URL: http://www.statmethods.net/stats/correlations.html **...**

www.statmethods.net/search/index.asp?QU=options&Page=3&Action=Search

Estimators and Statistical Tests TODO: give the structure of this **...**

... add an extreme value [1] -0.1949206 + **Pearson** residuals One can sometimes spot outliers with the **Pearsons** residuals: sample density / density according **...**

zoonek2.free.fr/UNIX/48_R/08_Estimators_and_Tests.txt

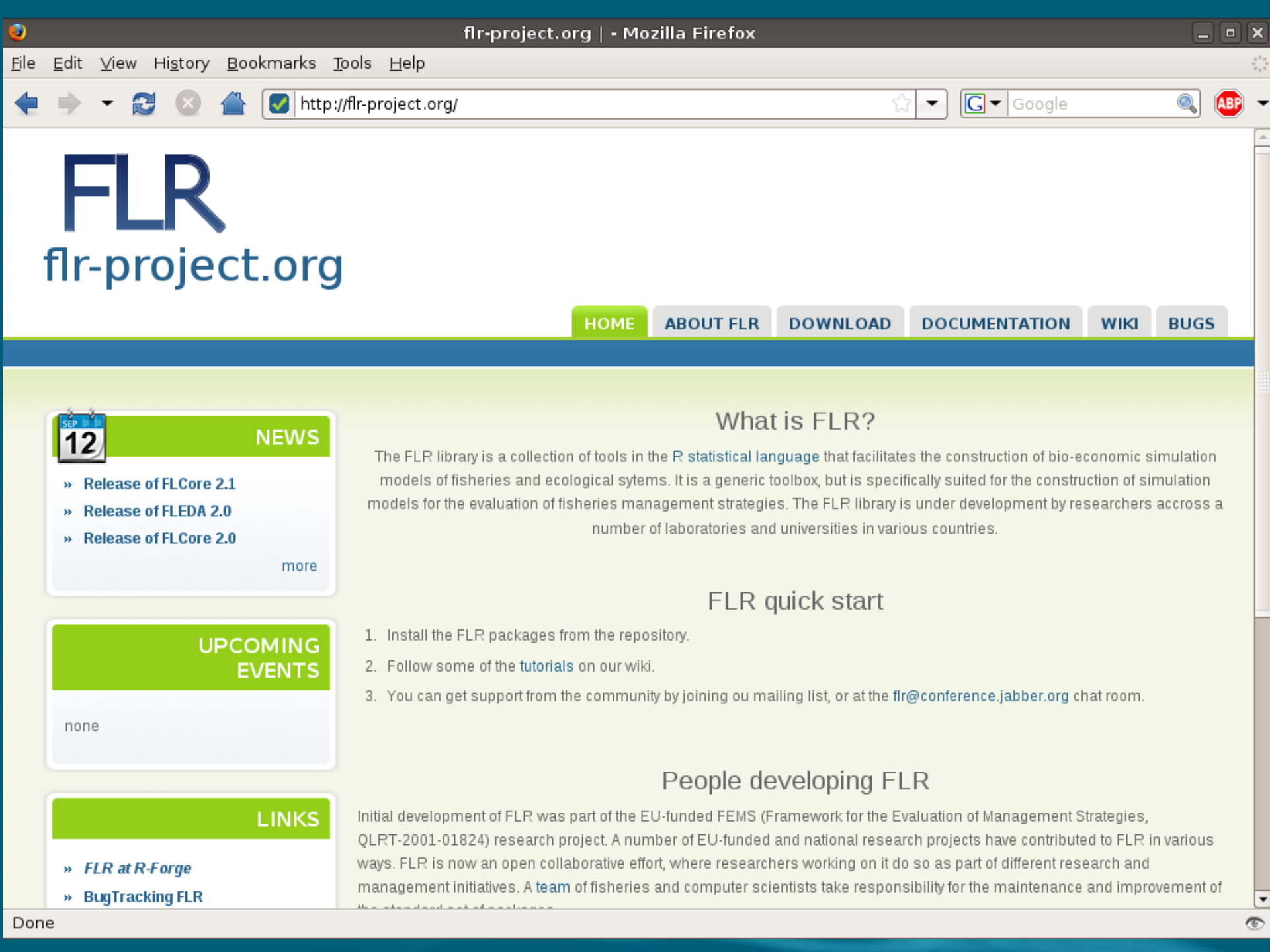
Generalized Linear Models: logistic regression, Poisson regression **...**

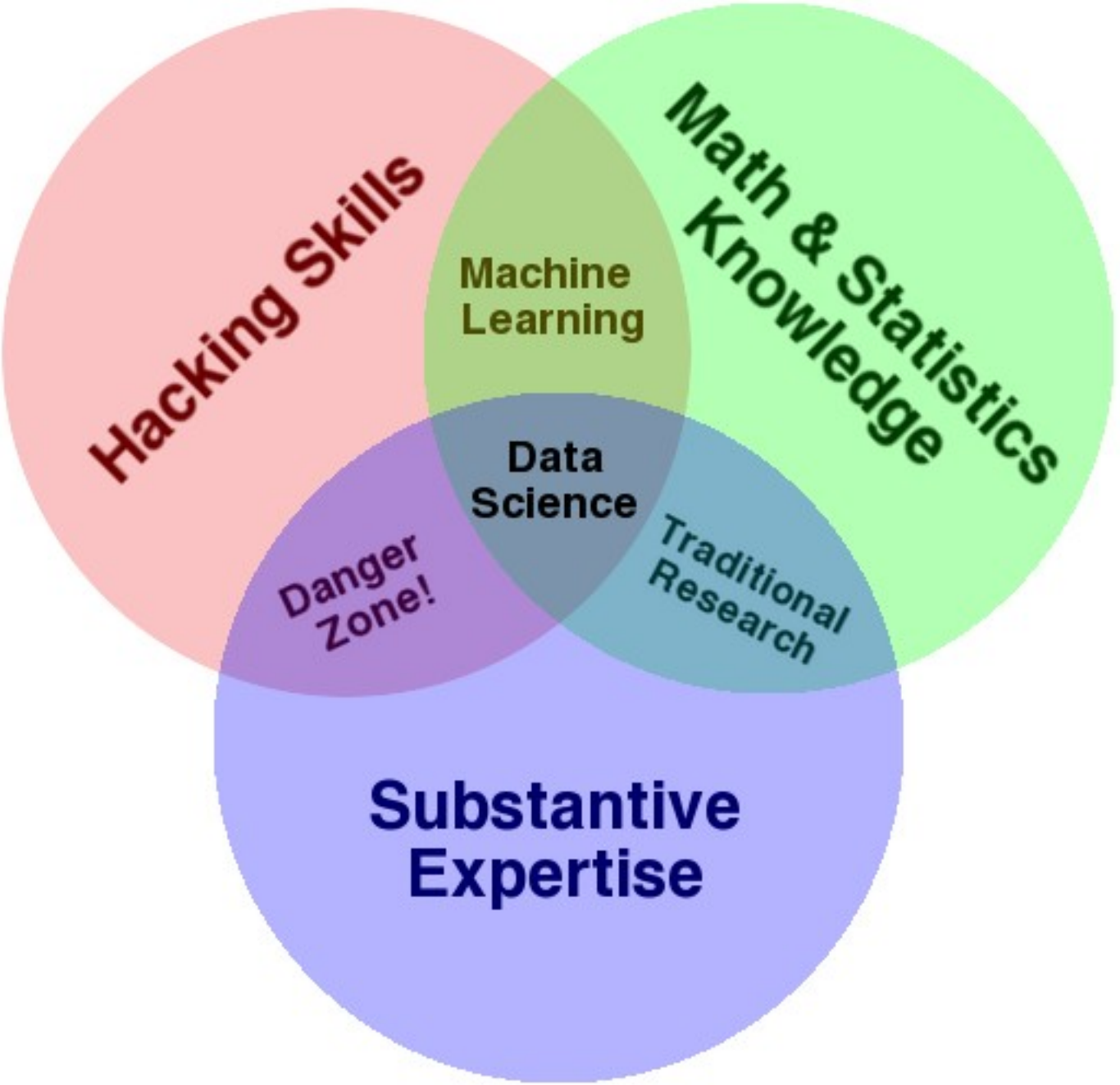
The **Pearson** residuals are defined as y_i - \hat y_i r_i = ---------- s_i where

FLR

- R packages for fisheries science
- Open Source, collaborative work
- Assessment, EDA, MSE, simulation
- Based on S4 classes and methods
- Reuse of legacy code, C, C++, F77, F90
- Currently version 2 in beta
- `http://flr-project.org`

Cefas


flr-project.org | - Mozilla Firefox
File Edit View History Bookmarks Tools Help
http://flr-project.org/
Google
FLR
flr-project.org
HOME
ABOUT FLR
DOWNLOAD
DOCUMENTATION
WIKI
BUGS
SEP 12
NEWS
» Release of FLCore 2.1
» Release of FLEDA 2.0
» Release of FLCore 2.0
more
UPCOMING EVENTS
none
LINKS
» FLR at R-Forge
» BugTracking FLR
What is FLR?
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 sytems. It is a generic toolbox, but is specifically suited for the construction of simulation models for the evaluation of fisheries management strategies. The FLR library is under development by researchers accross a number of laboratories and universities in various countries.
FLR quick start
1. Install the FLR packages from the repository.
2. Follow some of the tutorials on our wiki.
3. You can get support from the community by joining ou mailing list, or at the flr@conference.jabber.org chat room.
People developing FLR
Initial development of FLR was part of the EU-funded FEMS (Framework for the Evaluation of Management Strategies, QLRT-2001-01824) research project. A number of EU-funded and national research projects have contributed to FLR in various ways. FLR is now an open collaborative effort, where researchers working on it do so as part of different research and management initiatives. A team of fisheries and computer scientists take responsibility for the maintenance and improvement of
Done


Hacking Skills
Math & Statistics Knowledge
Substantive Expertise
Machine Learning
Data Science
Danger Zone!
Traditional Research

Cefas

Links

- http://r-project.org
- http://flr-project.org
- http://rseek.org
- http://cran.r-project.org
- http://r-forge.r-project.org
- http://addictedtor.free.fr/

Cefas