An Introduction to Scientific Image Data Analysis using R

M. Austenfeld

- Introduction
- First steps with the R environment and Language
- Statistics and Plots in R
- Data Transfer and Analysis with ImageJ and R
- Useful Packages and Resources

- R is a language and environment for statistical computing and graphics

- OpenSource (GPL, GPL2)

- R is an implementation of the S programming language combined with lexical scoping semantics inspired by Scheme

- First binary 1993 (source avail. 1995) by Ross Ihaka and Robert Gentleman

- First production release 2000 (1.0)

- Current Release 2.15.2

Why Use R?

- Open Source and available for all popular OS environments
- Huge community
- Well documented (Many specialized books)
- Statistical routines not yet available in other packages
- Can be extended with over 4000 Packages (at the moment)!
- Packages can be written in different languages (C, C++, Java, Fortran)
- Can be called from Java (libraries available)
- Many different free GUI-Applications available
- Many Plot types available

Installation

Windows: “base” installation

MacOSX: pkg installation

Linux: sudo yum install R-core.i386 or R-core.x86_64 (Fedora)

sudo apt-get install r-base (Ubuntu)

Start Linux:

e.g. /usr/sbin/R (Shell)
or:
/usr/sbin/R -g TK & (with Window)

Packages Linux:

e.g.: /usr/lib/R/site-library

Data Editor Matrices and Dataframes:

edit(object)

R Workspace

The so-called workspace is the current **R** working environment. It contains all defined objects.

This workspace can be saved or reloaded with the following commands:

Save Workspace:

save.image(„yourFile.RData")

Save objects to a file:

save(object list,file="myfile.RData")

Load Workspace:

load(„yourFile.RData")

Basic Data Structures

Vectors
Values have the same data type (Character or numeric.. etc.)

Factors
A Factor is an ordered collection of items (categorial values). The different values are called levels. They are stored internally as integers

Matrices
A two-dimensional structure which can contain one datatype. Values are in rows and columns. Matrices are 2-d arrays. For more dimensions arrays are needed!

Data frames
For tabular data. Contains vectors with same length. Each column can have a different datatype.

Lists
A list is a collection of objects. A list can contain e.g. a list (2d) of arbitrary length. Different objects in one list are possible.

Arrays
The **array** is similar to a matrix but can have more dimensions

Conversion

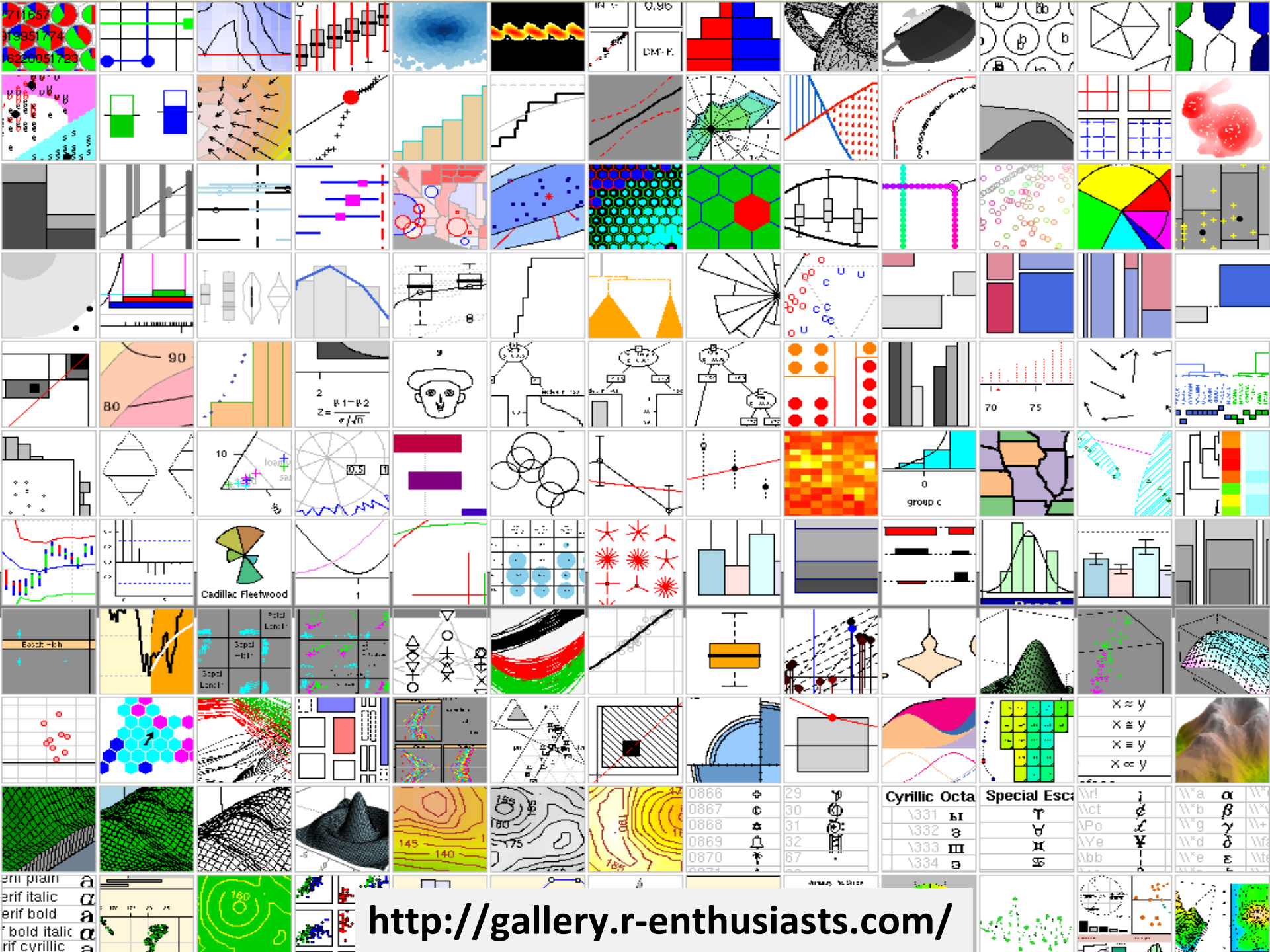
•matrix	as.matrix	is.matrix
•vector	as.vector	is.vector
•array	as.array	is.array

Information

•class(x)
•dim(x)
•length(x)

Basic Data Types

•integer	as.integer	is.integer
•double	as.double	…
•numeric	as.numeric	is.numeric
•complex	as.complex	…
•character	as.character	…
•logical	as.logical	…
•raw	as.raw	…


http://gallery.r-enthusiasts.com/

First Steps with the R Environment and Language

Cran Task Views

Bayesian	Bayesian Inference
ChemPhys	Chemometrics and Computational Physics
ClinicalTrials	Clinical Trial Design, Monitoring, and Analysis
Cluster	Cluster Analysis & Finite Mixture Models
DifferentialEquations	Differential Equations
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
HighPerformanceComputing	High-Performance and Parallel Computing with R
MachineLearning	Machine Learning & Statistical Learning
MedicalImaging	Medical Image Analysis
Multivariate	Multivariate Statistics
NaturalLanguageProcessing	Natural Language Processing
OfficialStatistics	Official Statistics & Survey Methodology
Optimization	Optimization and Mathematical Programming
Pharmacokinetics	Analysis of Pharmacokinetic Data
Phylogenetics	Phylogenetics, Especially Comparative Methods
Psychometrics	Psychometric Models and Methods
ReproducibleResearch	Reproducible Research
Robust	Robust Statistical Methods
SocialSciences	Statistics for the Social Sciences
Spatial	Analysis of Spatial Data
Survival	Survival Analysis
TimeSeries	Time Series Analysis
gR	gRaphical Models in R

Image Packages

CRAN packages Medical Images:

adimpro (core)
AnalyzeFMRI (core)
arf3DS4 (core)
cudaBayesreg
DATforDCEMRI (core)
dcemriS4 (core)
dpmixsim (core)
dti (core)
fmri (core)
mmand (core)
mritc (core)
neuRosim (core)
occ (core)
oro.dicom (core)
oro.nifti (core)
PET (core)
PTAk
Rniftilib (core)
RNiftyReg (core)
tractor.base (core)

General Image Packages

EBImage
EBImage is an R package which provides general purpose functionality for the reading, writing, processing and analysis of images.

biOps
Image processing and analysis

rimage?
Image Processing Module for R

RImageJ?
R bindings for ImageJ

Spatial Packages (Raster data)

rgdal
Bindings to Frank Warmerdam's Geospatial Data Abstraction Library (GDAL)

raster
Reading, writing, manipulating, analyzing and modeling of gridded spatial data

spatstat
A package for analysing spatial data, mainly Spatial Point Patterns

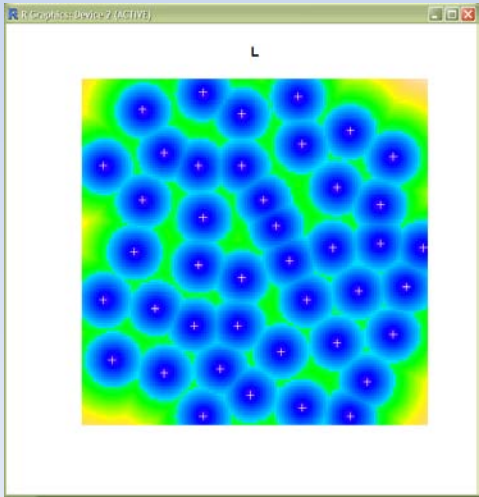
Spatstat

im create a pixel image
as.im convert other data to a pixel image
pixellate convert other data to a pixel image
as.matrix.im convert pixel image to matrix
as.data.frame.im convert pixel image to data frame
plot.im plot a pixel image on screen as a digital image
contour.im draw contours of a pixel image
persp.im draw perspective plot of a pixel image
rgbim create colour-valued pixel image
hsvim create colour-valued pixel image
[.im extract a subset of a pixel image
[<-.im replace a subset of a pixel image
shift.im apply vector shift to pixel image
X print very basic information about image X
summary(X) summary of image X
hist.im histogram of image
mean.im mean pixel value of image
integral.im integral of pixel values
quantile.im quantiles of image
cut.im convert numeric image to factor image
is.im test whether an object is a pixel image
interp.im interpolate a pixel image
blur apply Gaussian blur to image
connected find connected components
compatible.im test whether two images have comp. dim.
harmonise.im make images compatible
commonGrid find a common pixel grid for images
eval.im evaluate any expression involving images
scaletointerval rescale pixel values
zapsmall.im set very small pixel values to zero
levelset level set of an image
solutionset region where an expression is true
imcov spatial covariance function of image
convolve.im spatial convolution of images
transect.im line transect of image

The package supports
- creation, manipulation and plotting of point patterns
- exploratory data analysis
- simulation of point process models
- parametric model-fitting
- hypothesis tests and model diagnostics

e.g.:

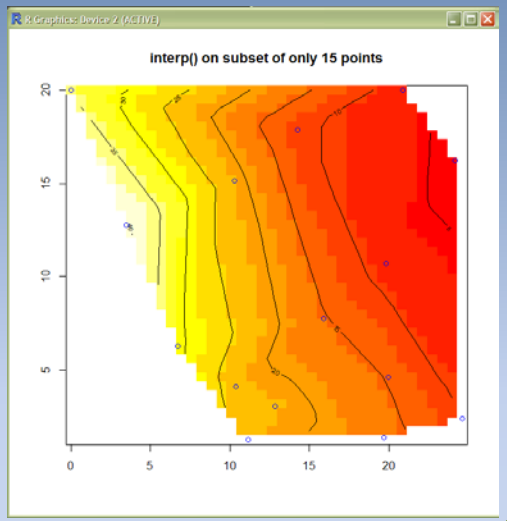
Summary statistics for a point pattern
Summary statistics for a multitype point pattern
Line segment patterns
Tessellations
Three-dimensional point patterns
……..



Interpolation, Spatial Modelling

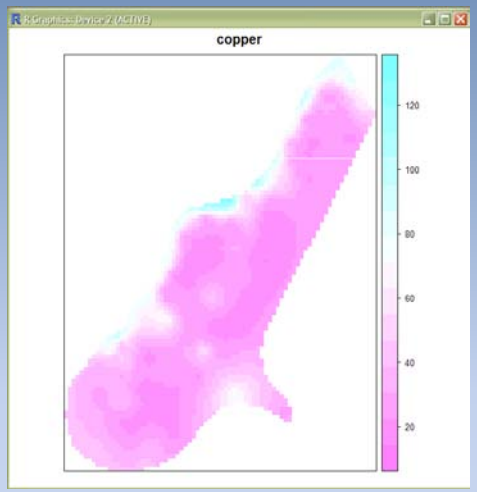
akima

Linear or cubic spline interpolation for irregular gridded data



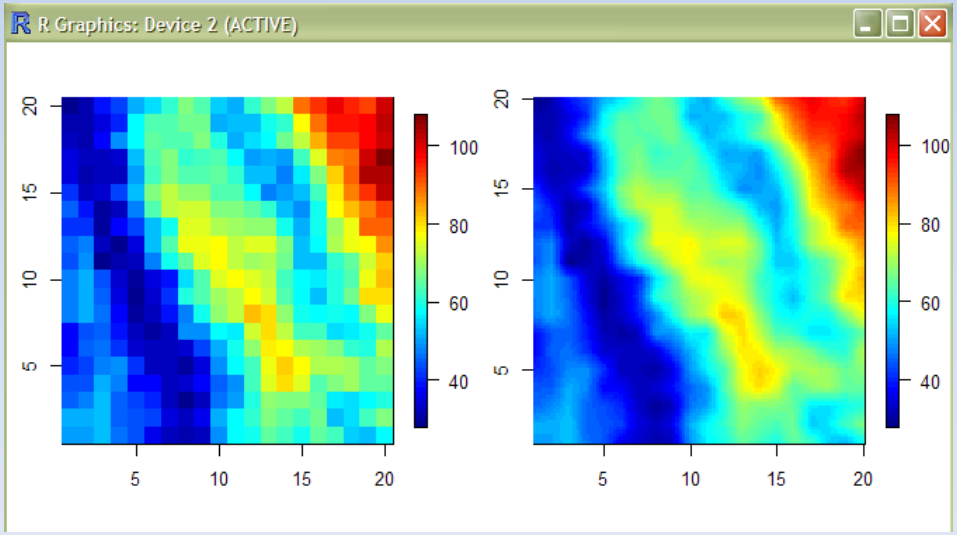
gstat

spatial and spatio-temporal geostatistical modelling, prediction and simulation



fields

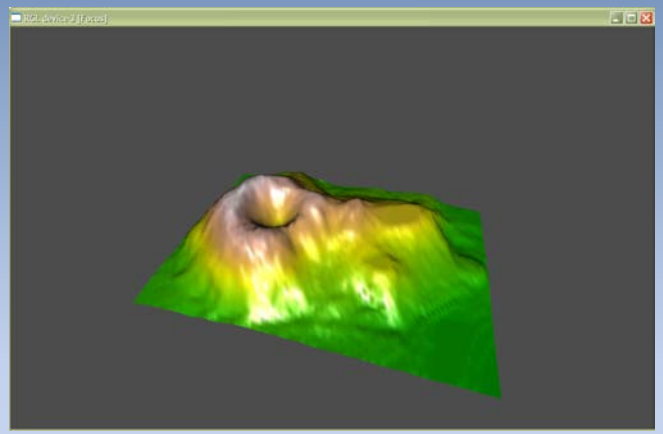
Fields is for curve, surface and function fitting with an emphasis on splines, spatial data and spatial statistics



Plot Packages

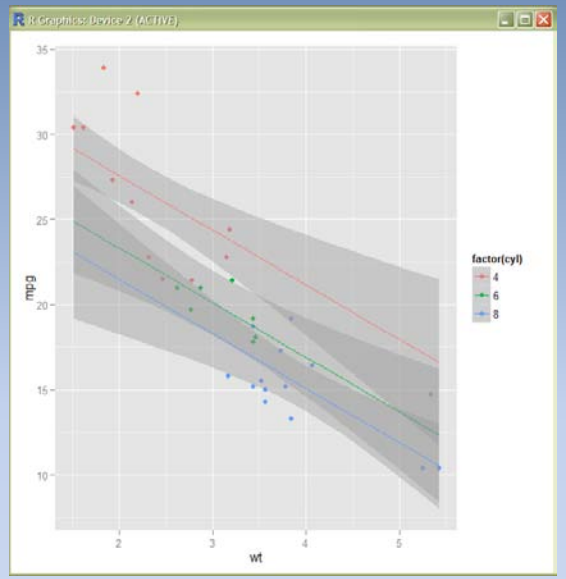
rgl

http://rgl.neoscientists.org/about.shtml



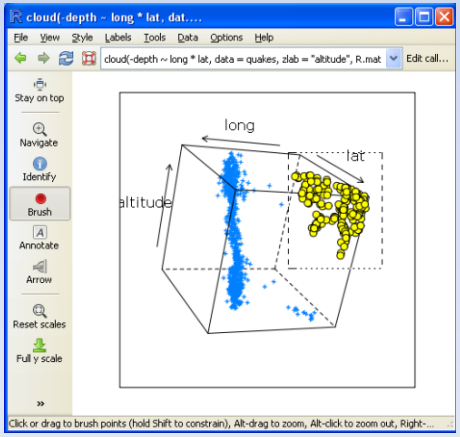
ggplot2

http://ggplot2.org/



playwith

http://code.google.com/p/playwith/



And:

animation
scatterplot3d
plotrix
……..

Reporting in R

Sweave

"Sweave is a tool that allows to embed the R code for complete data analyses in latex documents. The purpose is to create dynamic reports, which can be updated automatically if data or analysis change."

Source: http://www.statistik.lmu.de/~leisch/Sweave/

Knitr

"The **knitr** package was designed to be a transparent engine for dynamic report generation with R, solve some long-standing problems in Sweave, and combine features in other add-on packages into one package."

Source: http://yihui.name/knitr/

Knitr example


```
library(ggplot2)
pie <- ggplot(diamonds, aes(x = factor(1), fill = cut)) +
    xlab("cut") + geom_bar(width = 1)
pie + coord_polar(theta = "y")  # a pie chart
pie + coord_polar()  # the bullseye chart
```

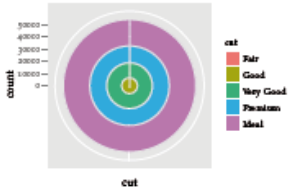


Figure 6: Two plots were produced in this chunk, but only the last one is kept. This can be useful when we experiment with many plots, but only want the last result. (Adapted from the ggplot2 website)

When multiple plots are produced by a code chunk, we may want to show them as an animation with the option `fig.show='animate'`. Figure 7 shows a simple clock animation; you may compare the code to Figure 5 to understand that high-level plots are always recorded, regardless of where they appeared.

```
par(mar = rep(3, 4))
for (i in seq(pi/2, -4/3 * pi, length = 12)) {
    plot(0, 0, pch = 20, ann = FALSE, axes = FALSE)
    arrows(0, 0, cos(i), sin(i))
    axis(1, 0, "VI"); axis(2, 0, "IX")
    axis(3, 0, "XII"); axis(4, 0, "III"); box()
}
```

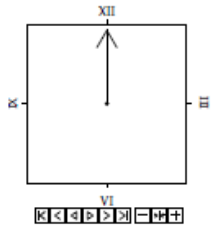
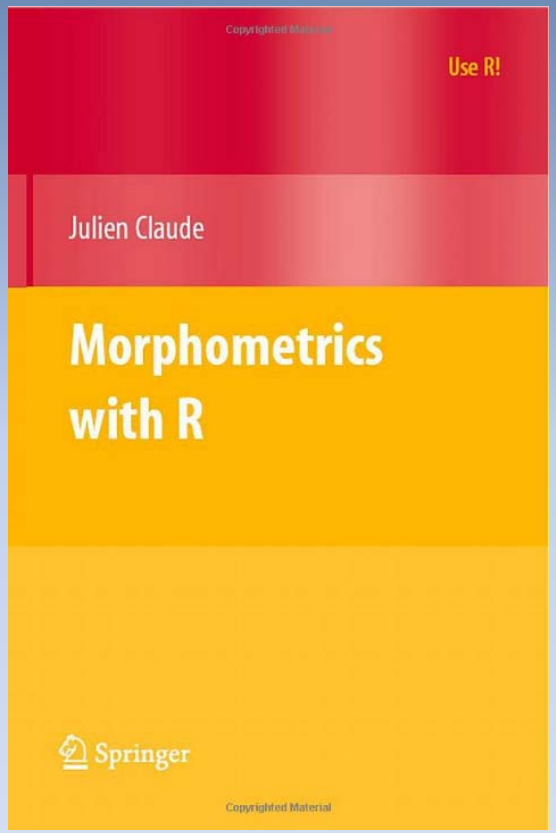


Figure 7: A clock animation. You have to view it in Adobe Reader: click to play/pause; there are also buttons to speed up or slow down the animation.

We can also set the alignment of plots easily with the `fig.align` option; this document uses `fig.align='center'` as a global option, and we can also set plots to be left/right-aligned. Figure 8 is an example

Source: http://yihui.name/knitr/

Morphometrics



Packages:

Momocs

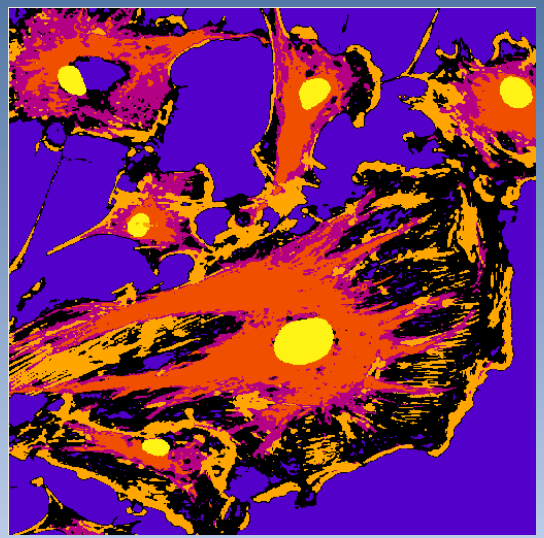
shapes

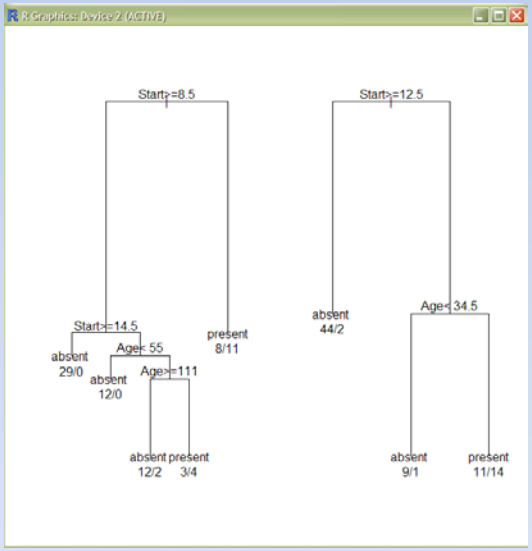
Source: http://www.isem.cnrs.fr/spip.php?article835

Maschine Learning / Supervised or Unsupervised Learning

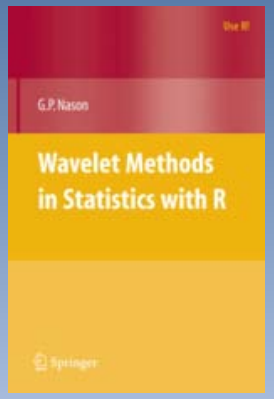
CRAN packages:

- ahaz
- arules
- BayesTree
- Boruta
- BPHO
- bst
- C50
- caret
- COREIearn
- CoxBoost
- Cubist
- e1071 (core)
- earth
- elasticnet
- ElemStatLearn
- evtree
- gafit
- GAMBoost
- gamboostLSS
- gbm (core)
- glmnet
- glmpath
- GMMBoost
- grplasso
- grpreg
- hda
- ipred
- kernlab (core)
- klaR
- lars
- lasso2
- LiblineaR
- LogicForest
- LogicReg
- longRPart
- mboost (core)
- mvpart
- ncvreg
- nnet (core)
- oblique.tree
- pamr
- party
- partykit
- penalized
- penalizedSVM
- predbayescor
- quantregForest
- randomForest (core)
- randomSurvivalForest
- rattle
- rda
- rdetools
- REEMtree
- relaxo
- rgenoud
- rgp
- rminer
- ROCR
- rpart (core)
- rpartOrdinal
- RPMM
- RSNNS
- RWeka
- sda
- SDDA
- svmpath
- tgp
- tree
- TWIX
- varSelRF

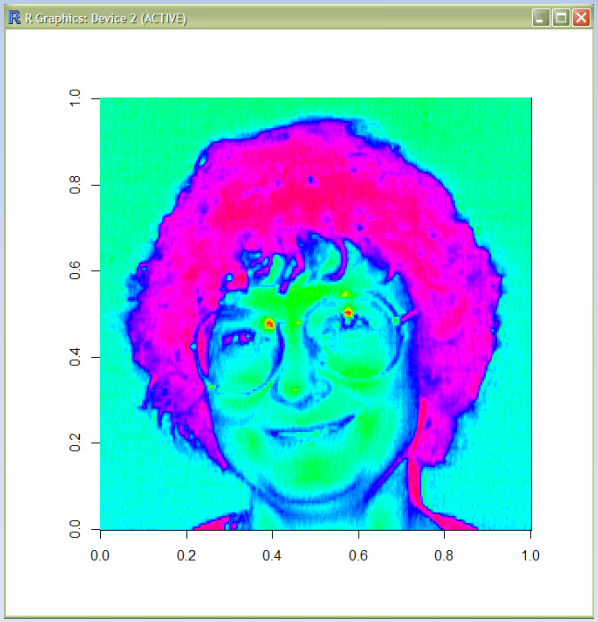


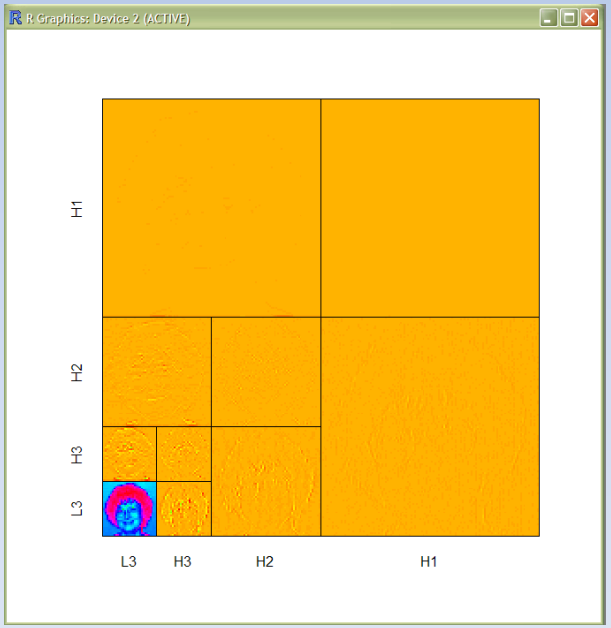


Wavelett Packages



- wavethresh
- wavelets
- waveslim





Source: Package “waveslim”, Author: Brandon Whitcher

Using R in Java
Libraries and Projects

rJava
http://www.rforge.net/rJava/

Rserve
http://rforge.net/Rserve/

RCaller
http://www.mhsatman.com/rcaller.php

JVM-based Interpreter for the R Language
renjin
http://code.google.com/p/renjin/

R GUI's

R Commander
http://socserv.mcmaster.ca/jfox/Misc/Rcmdr/

JGR
http://rforge.net/JGR/

Deducer
http://www.deducer.org/

RStudio
http://rstudio.org/

RKWard
http://rkward.sourceforge.net/

Tinn-R
http://www.sciviews.org/Tinn-R/

StatET
http://www.walware.de/goto/statet

Elastic-R

http://www.elastic-r.net/

Bio7
http://bio7.org

Thank You!

Numeric Precision

http://cran.r-project.org/doc/FAQ/R-FAQ.html#Why-doesn_0027t-R-think-these-numbers-are-equal_003f

http://rwiki.sciviews.org/doku.php?id=misc:r_accuracy:high_precision_arithmetic

Batch Execution of R

R CMD BATCH [options] infile [outfile]

Run **R** non-interactively with input from
infile and send output (stdout/stderr) to another file.