# R software: advantages and opportunities

Antonio Gasparrini
on behalf of the R.LSHTM group

Centre for Statistical Methodology (CSM) Forum

LSHTM 14 October 2011

## The R software

R is a freely-available programming language and software environment for statistical computing and graphics

Originally created by Ross Ihaka and Robert Gentleman at the University of Auckland in the early 90's

Currently maintained and extended by the *R Development Core Team* and a vast R community

# Objects and environments

Define two *objects*:
```
> x <- sort(runif(50,10,90))
> y <- 10 + 0.2*x + rnorm(50,0,2)
```

Run a regression model and extract the coefficients:
```
> model <- lm(y~x)
> coef(model)
(Intercept)            x
 10.2740494    0.1957665
```

Multiple objects of different classes defined in an *environment*

Motto: 'everything is an object'

# Programming in R

*Functional* and *object-oriented* programming style

Structure of a function:
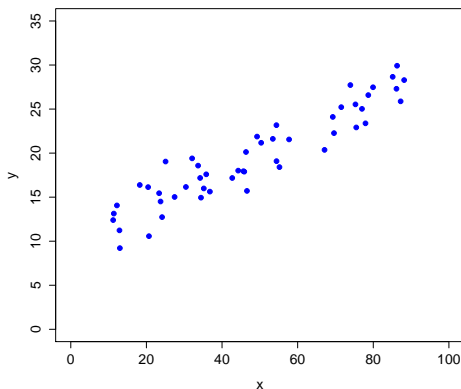```
> fun <- function(arg1,arg2,...) {
+    expression1
+    expression2
+    ...
+    return(output)
+ }
```

Example. A function for the mean:
```
> mymean <- function(object) sum(object)/length(object)
> mymean(x)
[1] 53.3065
```
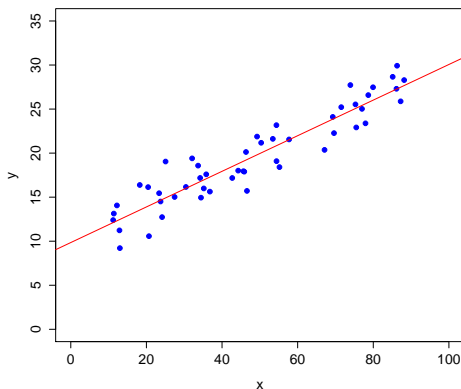
# Graphics in R (I)

```
> plot(x,y,pch=19,col="blue",ylim=c(0,35),xlim=c(0,100))
```
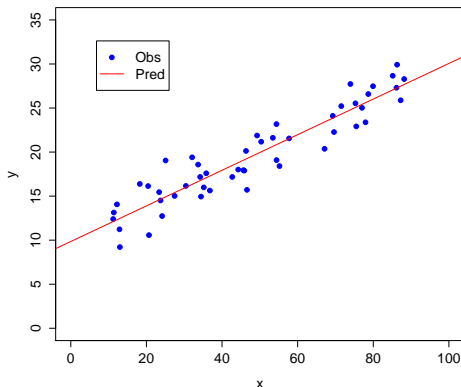
# Graphics in R (II)
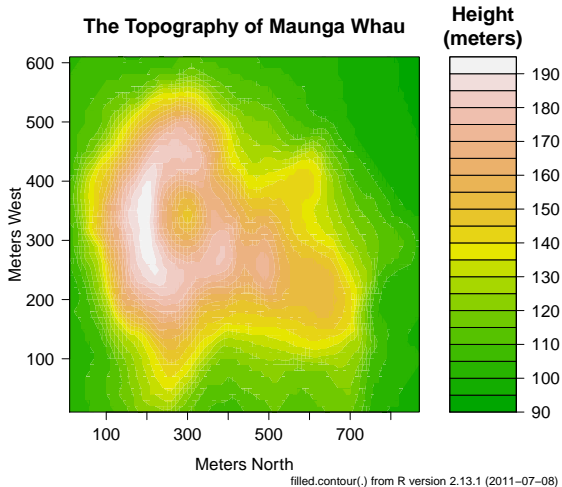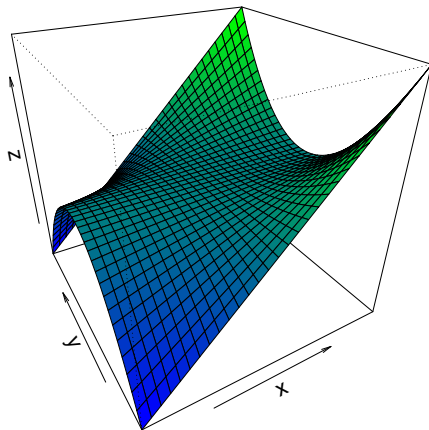
```
> abline(model,col=2)
```

# Graphics in R (III)

```
> legend("topleft",c("Obs","Pred"),col=c(4,2),pch=c(19,NA),
+  lty=c(NA,1),inset=0.1)
```

# Advanced graphics (1)



**The Topography of Maunga Whau**

**Height (meters)**

Meters West

Meters North

filled.contour(.) from R version 2.13.1 (2011–07–08)

# Advanced graphics (2)

## Packages

R computing structure is organized in *packages*: collections of code + data + documentation

Actually, something more: a *coherent functional environment* nested in the global environment, used for specific applications

A set of core packages are directly maintained and developed by the R Development Core Team

In addition, the capabilities of R are extended through user-created packages: on 13 Oct 2011, 3,350 deposited on CRAN

## Open-source software

R is a free software released under the *GNU General Public License*

This means:

- freely available without a fee
- free to see how it is written
- free to reuse the code under the same terms
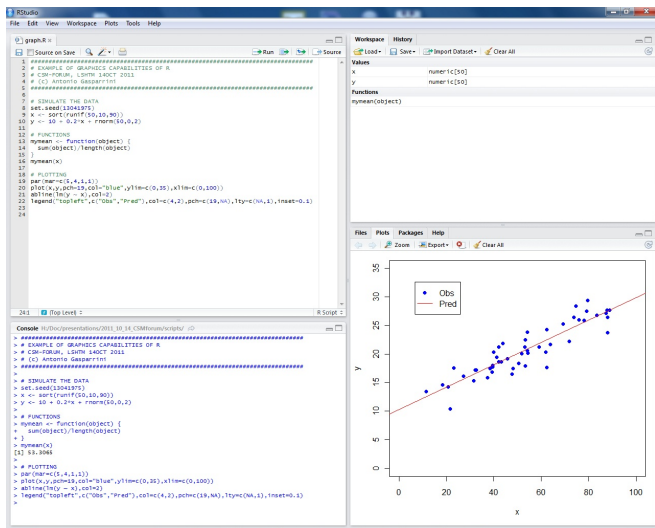
The R community

# Help and documentation

- Help pages
- Package vignettes
- Manuals (mostly free), books
- R mailing list, task views
- Search engines (RSeek, R-Search)
- Journals (R Journal, Journal of Statistical Software)

# R console
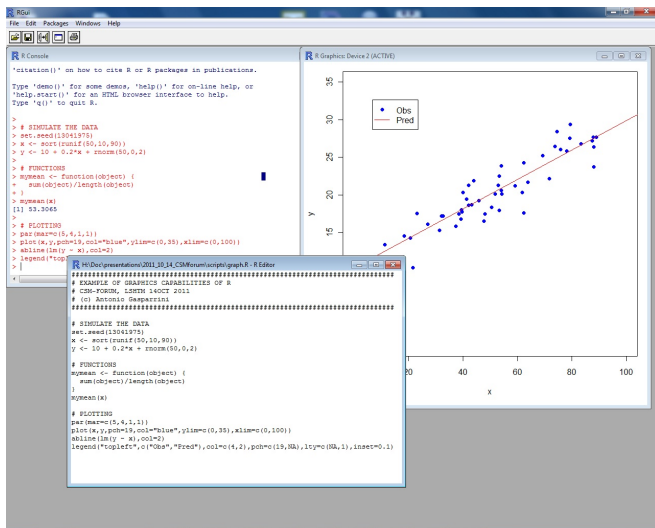
# GUI: RStudio

## So why using R?

- Freely available, open-source
- Performance comparable with (or superior to) other commercial software
- Flexible and easy programming methods
- High-level graphical capabilities
- Vast documentation, wide community of users

And:

- Optimal choice for LSHTM students from developing countries
- Standard software in other UK universities: Warwick, Lancaster, Oxford , Cambridge, UCL, Edinburgh, Glasgow

# The R.LSHTM group

- Antonio Gasparrini
- Edmond Ng
- Jennifer Rogers
- Thomas Chu
- Richard Silverwood
- Jonathan Bartlett

- Tieble Traore
- Liz Turner
- David Prieto
- Rosalba Radice
- Andrew Cox

# Activities within LSHTM

- Dedicated web-pages with info and links
- A mailing list (r@lists.lshtm.ac.uk)
- Planned R user meetings, with seminars and help desk
- Planned R courses, possibly starting early next year

# Useful links

The R web-page:
http://www.r-project.org/

CSM R web-page:
http://csm.lshtm.ac.uk/themes/statistical-computing/r/

LSHTM R wiki:
http://wiki.lshtm.ac.uk/fsg/index.php5/R:RIndex

LSHTM R mailing list:
https://lists.lshtm.ac.uk/sympa/info/r
r@lists.lshtm.ac.uk