

B.I.R.O.

Best Information through Regional Outcomes

A Public Health Project funded by the European Commission, DG-SANCO 2005

Statistical routines

Luca Rossi

Rome

Software







- PostgreSQL (Data)
- R (Stastical Analysis)
- Latex (PDF Output)

R (programming language) - Wikipedia, the free encyclopedia - Mozilla Firefox

File Modifica Visualizza Cronologia Segnalibri Strumenti ?

http://en.wikipedia.org/wiki/R-Project

Google

Come iniziare Ultime notizie

Google R-project logo Cerca Cerca nelle news Desktop Rolling Stone Rhapsody.com Galleria pulsanti Gmail Impostazioni

R (programming language)

From Wikipedia, the free encyclopedia

(Redirected from R-Project)

The **R programming language**, sometimes described as **GNU S**, is a programming language and software environment for statistical computing and graphics. It was originally created by Ross Ihaka and Robert Gentleman at the University of Auckland, New Zealand, and is now developed by the *R Development Core Team*. R is considered by its developers to be an implementation of the S programming language, with semantics derived from Scheme. The name R comes partly from the first name of the two original authors, and partly as a word play on the name 'S'.[1]

R is widely used for statistical software development and data analysis, and has become a de-facto standard among statisticians for the development of statistical software.[2] R's source code is freely available under the GNU General Public License, and pre-compiled binary versions are provided for Microsoft Windows, Mac OS X, and several Linux and other Unix-like operating systems. R uses a command line interface, though several graphical user interfaces are available.

Contents [hide]

1 Features
2 Critiques
3 Packages
4 Development
4.1 Milestones
5 Productivity tools
6 CRAN

R



Designed by	Ross Ihaka and Robert Gentleman
Developer	R Development Core Team
Latest release	2.6.1 / November 26, 2007
Latest unstable release	Through SVN
Influenced by	S
OS	Cross-platform
License	GNU General Public License
Website	http://www.r-project.org/

WIKIPEDIA
The Free Encyclopedia

navigation
- Main page
- Contents
- Featured content
- Current events
- Random article

interaction
- About Wikipedia
- Community portal
- Recent changes
- Contact Wikipedia
- Donate to Wikipedia
- Help

search
Go Search

toolbox
- What links here

Esecuzione script parzialmente permessa, 1/2 (wikipedia.org) | <SCRIPT>: 14 | <OBJECT>: 0 Opzioni...

Completato song

start Untitled2 - Ope... Desktop 2 Firefox WP7_V1 1_July... 15.30

LaTeX - Wikipedia, the free encyclopedia - Mozilla Firefox

File Modifica Visualizza Cronologia Segnalibri Strumenti ?

http://en.wikipedia.org/wiki/LaTeX

Come iniziare Ultime notizie

Google R-project logo Cerca Cerca nelle news Desktop Rolling Stone Rhapsody.com Galleria pulsanti Gmail Impostazioni

LaTeX

From Wikipedia, the free encyclopedia

"LATEX" redirects here. For other uses, see latex (disambiguation).
For help editing formulas on Wikipedia, see Help:Displaying a formula.

LaTeX is a document markup language and document preparation system for the TeX typesetting program. Within the typesetting system, its name is styled as LaTeX.



The LaTeX logo, typeset with LaTeX

LaTeX is widely used by mathematicians, scientists, philosophers, engineers, scholars in academia and the commercial world, and other professionals.[1] As a primary or intermediate format (e.g. translating DocBook and other XML-based formats to PDF), LaTeX is used because of the quality of typesetting achievable by TeX. The typesetting system offers programmable desktop publishing features and extensive facilities for automating most aspects of typesetting and desktop publishing, including numbering and cross-referencing, tables and figures, page layout and bibliographies.

LaTeX is intended to provide a high-level language that accesses the power of TeX. LaTeX essentially comprises a collection of TeX macros and a program to process LaTeX documents. Because the TeX formatting commands are very low-level, it is usually much simpler for end-users to use LaTeX.

LaTeX was originally written in the early 1980s by Leslie Lamport at SRI International [2]. It has become the dominant method for using TeX—few people write in plain TeX any more. The current version is LaTeX2e (styled LaTeX 2ε).

LaTeX, like TeX, is free software.

Contents [hide]

1 The typesetting system

WIKIPEDIA
The Free Encyclopedia

navigation
- Main page
- Contents
- Featured content
- Current events
- Random article

interaction
- About Wikipedia
- Community portal
- Recent changes
- Contact Wikipedia
- Donate to Wikipedia
- Help

search
Go Search

toolbox
- What links here

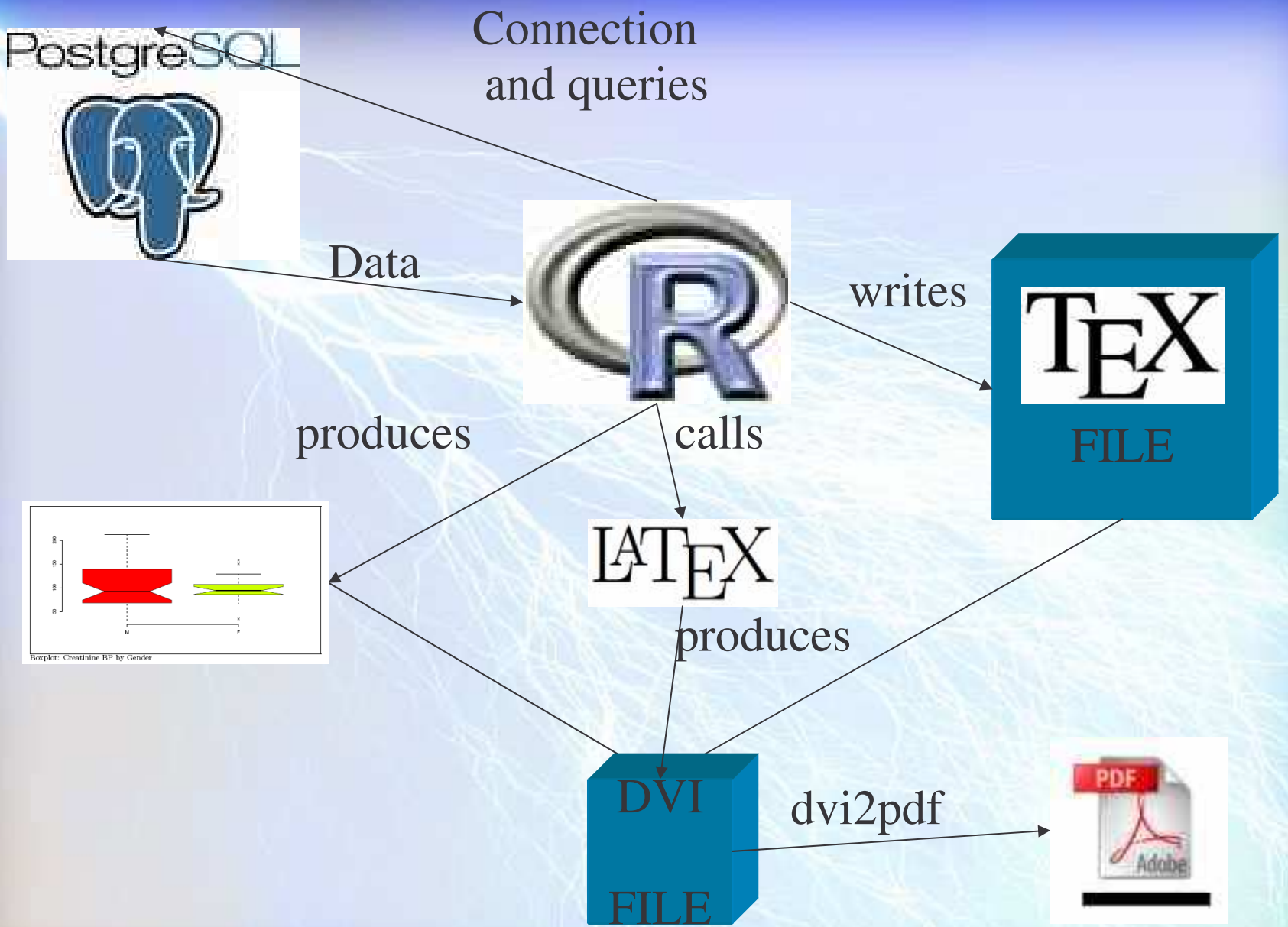
Esecuzione script parzialmente permessa, 1/2 (wikipedia.org) | <SCRIPT>: 14 | <OBJECT>: 0 Opzioni...

Completato song

start Untitled2 - Ope... Desktop 2 Firefox WP7_V1 1_July... 15.33

How does it work?

- R connects himself to the Postgres DB
- R does some queries and gets the data
- R does statistical analysis
- R prints a .tex file and saves .eps files (graphs)
- R calls latex that compiles the .tex file and convert .dvi to .pdf file


PostgreSQL
Connection
and queries
Data
writes
TEX
FILE
produces
calls
LATEX
produces
Boxplot: Creatinine BP by Gender
DVI
FILE
dvi2pdf
PDF
Adobe

Connection to the PostgreSQL DB

Most used R Packages:

- RODBC (for Windows/Linux/Mac)
- Rdbi with Rdbi.PgSQL or DBI with DBI.PgSQL (for Linux/Mac)

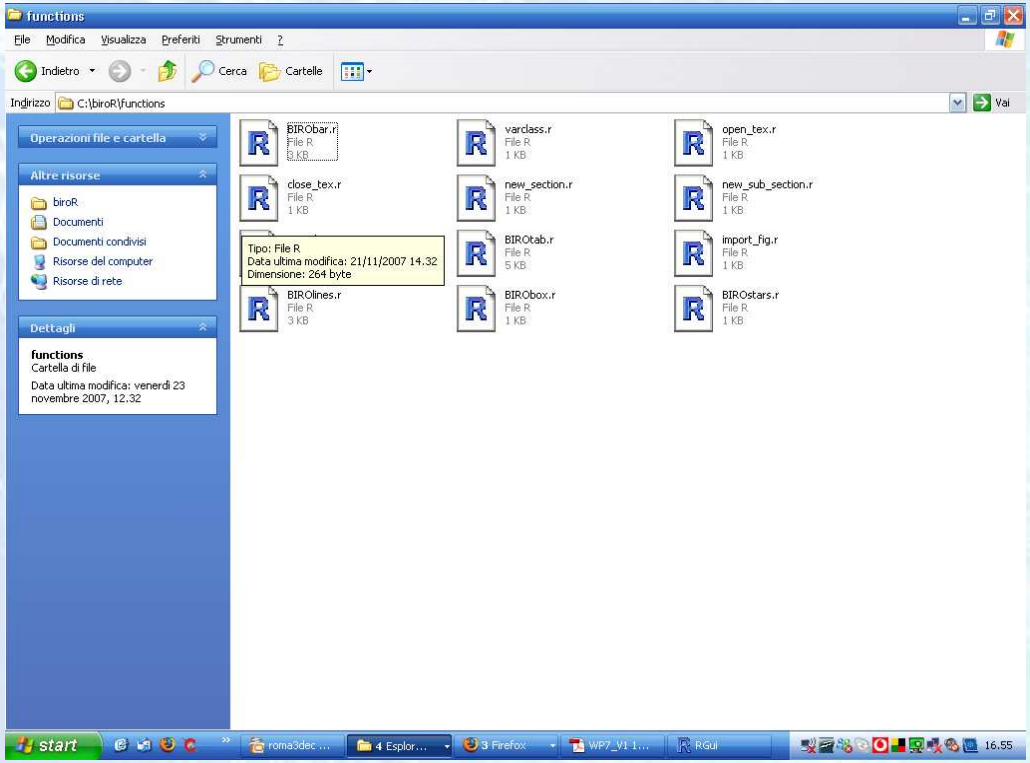
The prototype has been tried on Windows O.S. (XP and Vista)

Statistical Analysis

- This phase concentrates on the creation of input/output functions based on a test sample.
- The prototype application is based on the BIRO report template (deliverable)
- Outputs relate to simple statistics e.g. percentages, absolute frequencies and means.
- Ad hoc written R functions print graphs and tables via Latex

Output functions

- R loads these functions in his workspace:
  - BIROtab
  - BIRObar
  - BIROlines
  - BIRObox
  - BIROstars
  - Other functions to manage the PDF output



Functions

- At the moment the ad hoc written R software needs only one package, RODBC, to connect to the PostgreSQL DB (BIRO Database Engine)

BIROtab()

- This function translates a R table...

```
> a<-rbinom(10000,4,0.5) # A simple way to generate
> b<-rbinom(10000,1,0.5) # two categorical variables
> table(a,b)
   b
a     0    1
  0  311  303
  1 1282 1215
  2 1858 1876
  3 1263 1235
  4  326  331
```

This is a tipical ‘low level’ R table

BIROtab()

- ...into a Latex table...

```
\begin{tabular}{ |c |c |c |c |}
\hline
\multirow{ 1 }{*}{ Var A }   &  b1  &  b2   &\\
\hline
a1   &311  & 303  &  614   \\
a2   &1282  & 1215  &  2497   \\
a3   &1858  & 1876  &  3734   \\
a4   &1263  & 1235  &  2498   \\
a5   &326  & 331  &  657   \\
\hline
& 5040 &4960 &10000 \\
\hline
\end{tabular}
```

BIROtab()

- ... to obtain tables like these (in PDF)

Var A	b1	b2	
a1	311	303	614
a2	1282	1215	2497
a3	1858	1876	3734
a4	1263	1235	2498
a5	326	331	657
	5040	4960	10000

Example

Var A	b1 (%)	b2 (%)	
a1	3.11	3.03	6.14
a2	12.82	12.15	24.97
a3	18.58	18.76	37.34
a4	12.63	12.35	24.98
a5	3.26	3.31	6.57
	50.4	49.6	100

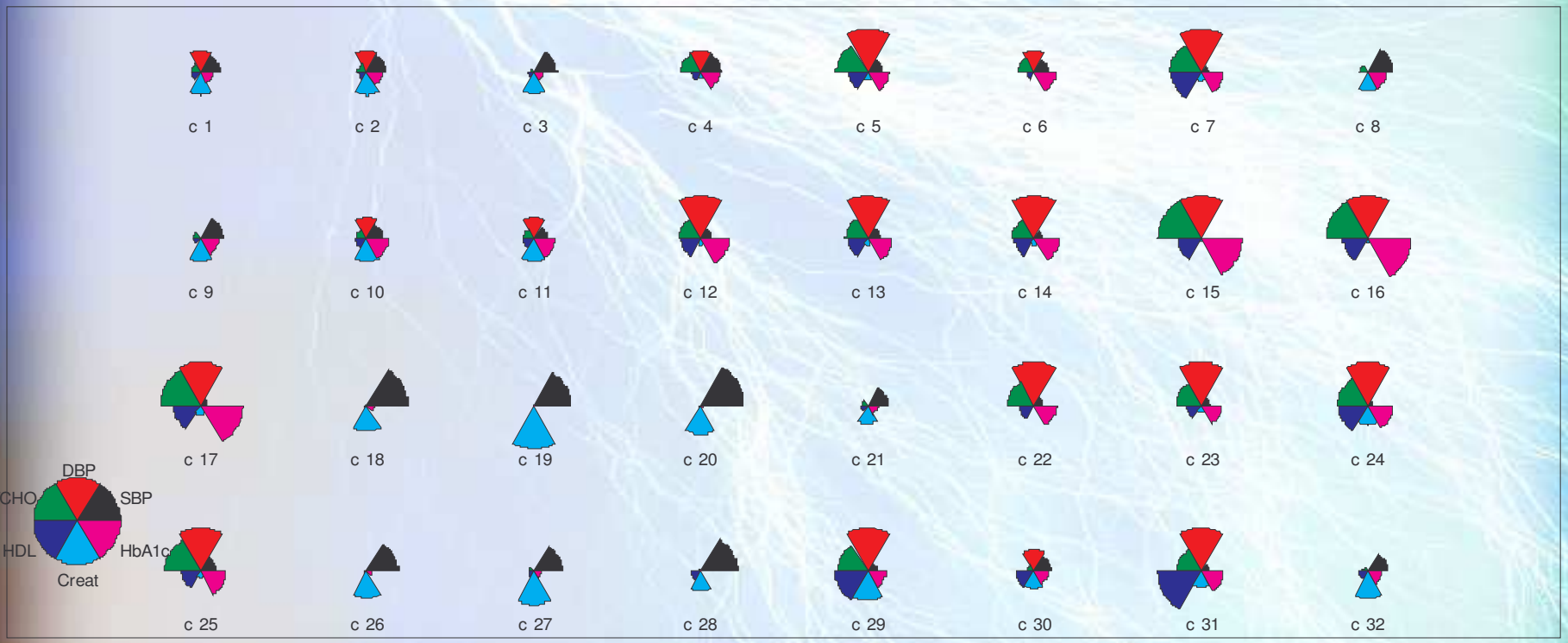
Example

BIROtab

- Through it you can:
  - calculate percentages and absolute frequencies
  - do landscape or portrait oriented tables
  - change width of table (and caption)
  - change variables labels

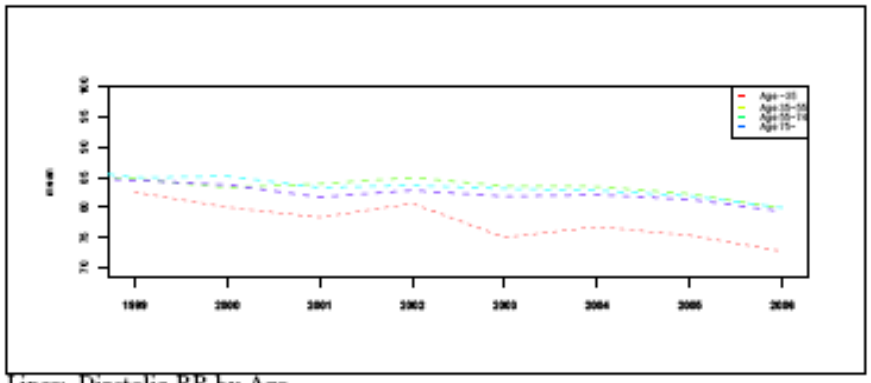
BIROstars()

- For starplots (one for each center)



BIROlines()

- For line plots



Lines: Diastolic BP by Age

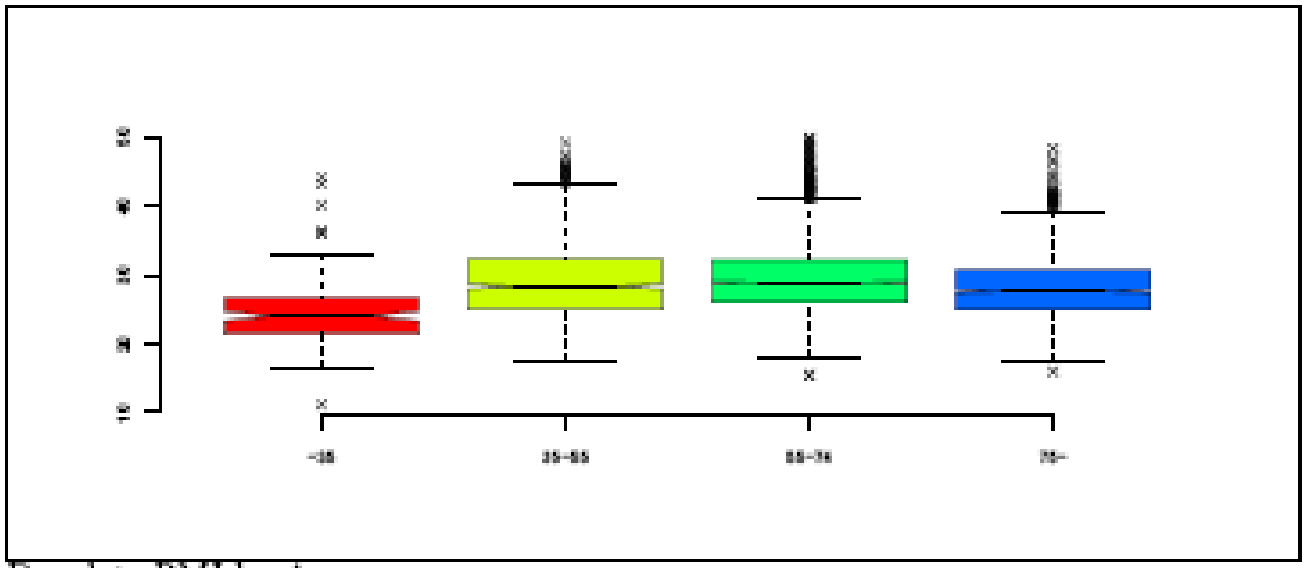
Parameters:

- Y-Variable
- Time Variable
- Stratification Variable

BIRObox()

- For boxplots

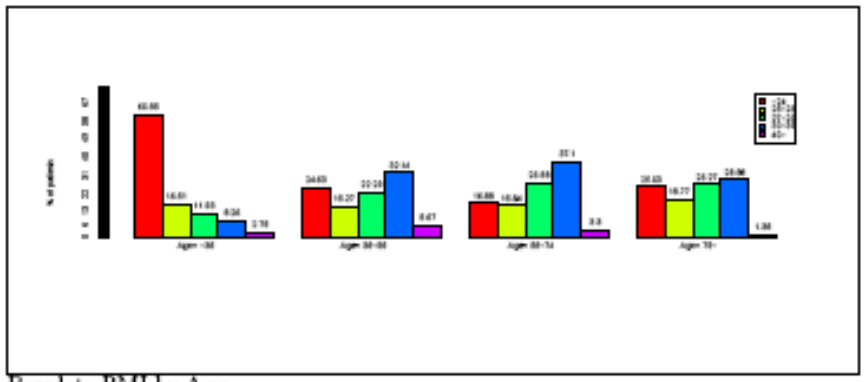
Specific option to print notches



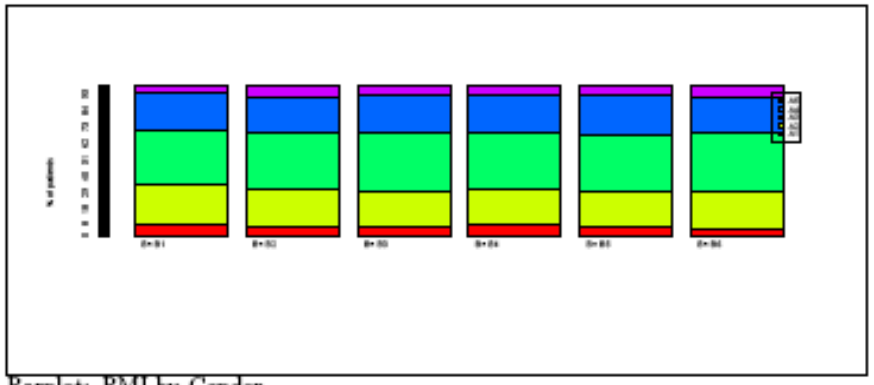
Boxplot: BMI by Age

BIRObar()

- For two different kinds of barplots:



Barplot: BMI by Age



Barplot: BMI by Gender

- Option for percentages/frequencies over each bar

R prints a tex file

- R copies a basic template file called “layout.tex” into a new one called “report.tex”, where new latex commands will be appended.
- Operation is performed using the R function cat repetitively

TEX file

- You can change a TEX file and produce a great report, article, book directly in PDF format.
- If you are a skilled Latex user, you can easily change a tex file and “compile” it to obtain a finely printed article/report that can have a complex structure.

R calls Latex

- This operation is done with the R function shell
- MiKTeX (free latex for Windows) is run automatically within R, by setting an environment variable pointing to the installation path

```
shell (paste("latex ",texfile,sep=""))

shell (paste("dvipdfm -o",pdffile,dvifile,sep=" "))
```

In the next future

- HTML tables (Web portal)
- Graphs in eps/jpeg/gif format
- Latex can manage the bibliography via BibTex...any hints/suggestions?
- Fine tuning of the Statistical Engine
  - Definition of “pre-processed” Statistical Objects
  - Central Engine
  - Multivariate Analysis

Thanks!