

Flip-flop seminar #5

10 December 2009

This seminar

- Getting started with R and Tinn-R
- Editing text and producing tables with R

Choice of stats software (this slide comes from a previous flip-flop seminar)

- Most people use what other people around them are using
- So will we by focusing first on R, SPSS, and STATA
- By and large, R beats STATA and STATA beats SPSS in speed, documentation and how many things you can do with it; and when it comes to graphs, open-source features, debugging codes, and price, nothing beats R (see <http://www.r-project.org/>)
- Yet in some really basic respects (e.g. weighting your data, or working with large data files on your PC) SPSS is by far better than STATA and STATA is better than R. However, you may just as well substitute SPSS with the free alternative PSPP (<http://www.gnu.org/software/pspp/pspp.html>)

But don't just follow the herd

- It is good to know that even among general-purpose stats software there are many-many alternatives to the above (e.g., BMDP, Gauss, MatLab, SAS, Statistica)
- See also <http://www.jstatsoft.org/> (note the searchable archive!)
- Search for e.g. “free statistical software” or “open-source statistics” and be amazed how far you can get for free
- The future anyway belongs to “cloud-computing” so don't imagine that you will stick with one software for ever

Intro-level advice on R

- Free intros under “Manuals” at <http://www.r-project.org/>
- Online help instantly available from the Help menu in R
- Plenty of shorter on-line intros are available too; see, e.g., <http://www.ats.ucla.edu/stat/>

A comprehensive intro that I like

- Everitt, Brian S., and Torsten Hothorn. 2006. *A Handbook of Statistical Analyses Using R*. Boca Raton, FL: Chapman & Hall/CRC
- All the script examples as well as pre-publication versions of the all chapters are included in the HSAUR package for R
- Run this (line by line) from your R prompt to see how it works:

```
install.packages("HSAUR")           # installs HSAUR
library("HSAUR")                    # activates HSAUR
vignette("Ch_introduction_to_R",    package = "HSAUR")
                                     # opens chapter 1
edit (vignette("Ch_introduction_to_R", package =
  "HSAUR"))                          # Opens file w. all codes in ch. 1
vignette(package = "HSAUR")         # lists all chapters
```

Advanced advice on R

- Plenty of books on R
 - e.g. check out the “Use R” series of Springer
- Internet
 - Listserver: <http://www.r-project.org/mail.html>
 - Google for keywords (e.g. “correlated random numbers in R”)
 - <http://addictedtor.free.fr/graphiques/> (vast collection of innovative R graphs where you can pick the layout that you like and look up the script to produce it)

Downloading R

The screenshot shows a web browser window displaying the R Project website. The browser's address bar shows the URL `http://www.r-project.org/`. The website header features the R logo and the text "The R Project for Statistical Computing".

On the left side, there are several navigation links:

- About R
 - [What is R?](#)
 - [Contributors](#)
 - [Screenshots](#)
 - [What's new?](#)
- Download, Packages
 - [CRAN](#)
- R Project
 - [Foundation](#)
 - [Members & Donors](#)
 - [Mailing Lists](#)
 - [Bug Tracking](#)
 - [Developer Page](#)
 - [Conferences](#)
 - [Search](#)
- Documentation
 - [Manuals](#)
 - [FAQs](#)
 - [The R Journal](#)
 - [Wiki](#)
 - [Books](#)
 - [Certification](#)
 - [Other](#)

The main content area displays several statistical plots and charts:

- PCA 5 vars**: A scatter plot showing the first three principal components. The x-axis is labeled "Factor 1 [41%]" and the y-axis is labeled "Factor 3 [19%]". The plot shows two distinct clusters of points, one in orange and one in green. A small pink point is also visible. The text "V. De Geneve" is present in the bottom right of the plot.
- Clustering 4 groups**: A dendrogram showing hierarchical clustering of data points into four groups. A vertical red line is drawn at a distance of 45. Below the dendrogram is a bar chart showing the number of points in each group: Group 1 (green) has 28 points, Group 2 (orange) has 16 points, and Group 3 (pink) has 2 points. The text "(1-3) 60%" is also present.
- Getting started**: A text box containing the following information:

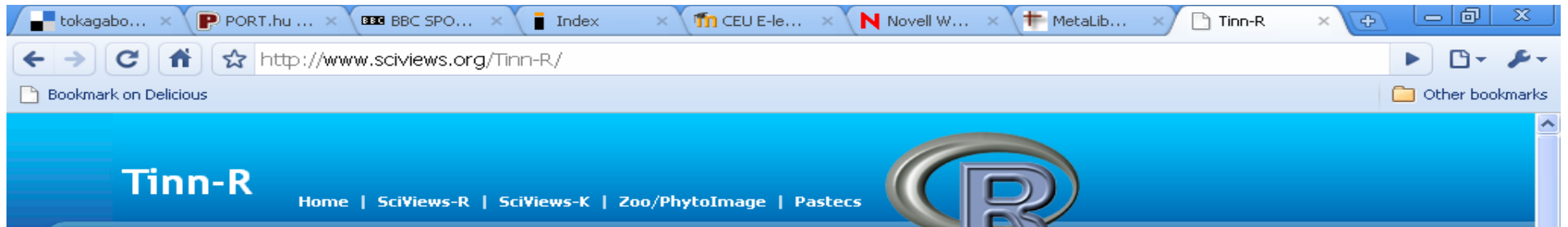
Getting started:

 - R is a free software environment for statistical computing and graphics. It compiles and runs on a wide variety of UNIX platforms, Windows and MacOS. To [download R](#), please choose your preferred [CRAN mirror](#).
 - If you have questions about R like how to download and install the software, or what the license terms are, please read our [answers to frequently asked questions](#) before you send an email.

News :

The browser's status bar at the bottom shows the download of "Tinn-R 1.15.1.7 stable.zip" and a "Show all downloads..." button.

Tinn-R



” Tinn-R is free, simple but efficient replacement for the basic code editor provided by Rgui.

```
MASS_ch01.r
0 # 1.1 A quick overview of S
1
2 2 + 3
3 sqrt(3/4)/(1/3 - 2/pi^2)
4 library(MASS)
5 mean(chem)
6 m <- mean(chem); v <- var(chem)/length(chem)
7 m/sqrt(v)
8
9 std.dev <- function(x) sqrt(var(x))
10 t.test.p <- function(x, mu=0) {
11   n <- length(x)
12   t <- sqrt(n) * (mean(x) - mu) / std.dev(x)
13   2 * (1 - pt(abs(t), n - 1))
14 }
15
16 t.stat <- function(x, mu = 0) {
17   n <- length(x)
18   t <- sqrt(n) * (mean(x) - mu) / std.dev(x)
19 }
20
```

```
R Console
> 2 + 3
[1] 5
> sqrt(3/4)/(1/3 - 2/pi^2)
[1] 6.626513
> library(MASS)
> mean(chem)
[1] 4.280417
> m <- mean(chem); v <- var(chem)/length(chem)
> m/sqrt(v)
[1] 3.958487
> source('C:/Documents and Settings/jcfaris/dados de aplicativos/Tinn-R/temp/MASS_ch01.r')
>
```

Tinn-R on top of R

Tinn-R, edit code and run it in R

Jun 27, 2008

Software

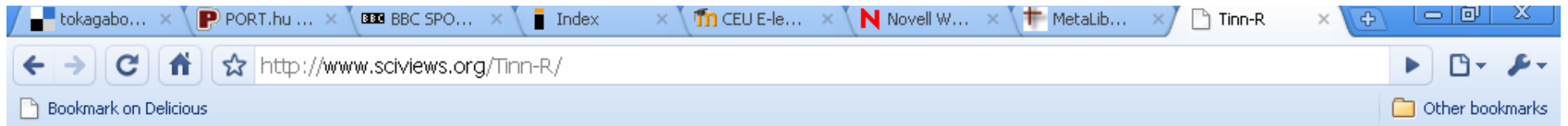
- SciViews-R
- SciViews-K
- Zoo/PhytoImage
- Tinn-R
- Pastecs
- LaboKit
- ShellAxis

Scientific computing

- R translation in French
- R GUI projects
- Benchmark
- Various data analysis software

Links

- Philippe Grosjean's home page
- Subaquatic photography



```
[1] 3.958487  
> source('C:/Documents and Settings/jcfaria/dados de aplicativos/Tinn-R/temp/MASS_ch01.r')  
>
```

Tinn-R on top of R

Tinn-R, edit code and run it in R

Jun 27, 2008

Tinn stands for the recursive acronym 'Tinn is not Notepad'. It is initially written as a small ASCII file editor primarily intended as a better replacement for the default Notepad.exe program distributed with Windows. With versions, it grew to a fairly complete, yet simple and compact code editor.

Tinn is written in Delphi 5 and it runs only under Windows (9X/Me/2000/XP). It is free and distributed under GPL 2.

The specific version (Tinn-R) contains enhancements to allow syntax highlighting of S language (in *.R, *.r, *.Q or *.q files), that is, the language used by the open source statistical software R (<http://www.r-project.org>).

It also pops up additional menu and toolbar when it detects Rgui running on the same computer. These addons interact with the R console and allow to submit code in part or in whole and to control R directly.

Tinn-R is a small, free and simple, yet efficient, replacement for the basic code editor provided by Rgui. Latest Tinn-R version and source is available on [SourceForge](http://sourceforge.net). Note that, starting from version 1.18.x.x, Tinn-R is not compatible anymore with Rgui in MDI mode (only SDI), nor with S-PLUS! Latest compatible version is 1.17.2.4, which is still distributed here.

Download and install Tinn-R for Windows

Oct 25, 2006

Please, visit [SourceForge](http://sourceforge.net) to download the latest version of Tinn-R.

Starting from Tinn-R 1.16.1.4 stable, the installer is required. Tinn-R now comes with two databases (R tips and R cards) that must be installed at the same time as Tinn-R itself. The latest stable version that can be installed manually (version, 1.15.1.7) is still downloadable from here (to run it on any computer from an USB key, for instance, just unzip the file on the key and run Tinn-R.exe from there).

- [Setup for Tinn-R](#), old stable version (1.17.2.4) (.exe, 5.2 Mb) compatible with Rgui in SDI or MDI mode, with SciViews R Console and with S-PLUS.
- [Tinn-R, last runnable version without installer](#) (1.15.1.7) (.zip, 2.9 Mb).

Plenty of alternatives to Tinn-R

- See http://www.sciviews.org/_rgui/, especially the menu points on
 - [Emacs \(ESS\)](#) (a more comprehensive programmers' editor than Tinn-R);
 - [IDE/Script Editors](#) (the direct alternatives to Tinn-R, including the very comprehensive jEdit at www.jedit.org) and
 - [R Commander](#) (an R-package that helps you do the things that you are most likely to need - i.e. to read in statistical data and analyse it with regression etc. - through a clickable pull-down menu like the one in SPSS, STATA, etc.)

An R package for those who are used to SPSS

- Handles user-defined missing values, variable and value labels; SPSS-style recodes and descriptive stats
- Produces pretty journal-style tables
- Helps with simulations and bootstrapping
- To install and see what it does, run this:

```
install.packages("memisc")
```

```
library(memisc)
```

```
?memisc
```

Customizing your R settings and start-up

(This stuff is really not necessary to get started with R but can make the visually-sensitive feel better about the software)

The Rprofile.site file

- This should be in the “\etc” subfolder in your R program directory (under Windows, this is most probably sg. like “c:\Program Files\R\R-2.9.2\etc\”)
- Contains commands that will be automatically executed when you start R; e.g. you can customize here the look of the R prompt ...

```
options(continue="      ", prompt="R> ")
```

- ... or make sure that the packages for some commonly used R textbooks are installed on your computer:

```
library(utils); necessary = c('HSAUR', 'faraway',  
  'DAAG', 'MASS', 'SoDA', 'RGraphics')  
installed = necessary %in% installed.packages()[,  
  'Package']  
if (length(necessary[!installed]) >=1)  
install.packages(necessary[!installed], dep = T)
```

The Rconsole file

- This should also be in the “\etc” subfolder in your R program directory (under Windows, this is most probably sg. like “c:\Program Files\R\R-2.9.2\etc\”)
- Contains your default setting; can change font type, size and style to:

```
font = TT Courier New
```

```
points = 12
```

```
style = bold
```

- You can also change display colors etc. The `rgb.txt` file (in the same folder) lists all available colors. My favorite scheme is:

```
background = Black
```

```
normaltext = Green
```

```
usertext = White
```

```
highlight = DarkRed
```

Editing HLM output in R

- I cleaned up a bit, improved, and uploaded all files used for this demonstration to the \Readings\R folder of the members-only area of POLBERG website
- The ZIP file contains:
 - “EES-Media_HLMoutput_editing.R” – the R code (open it in Tinn_R!)
 - 29 HLM output files (with *.txt extension). These are the input data
 - 10 files called “Table.....doc”. These are the products of the code
 - A PDF of the paper for which this code was written in the first place (building on previous codes for other papers). It shows the ultimate look of Tables 3.a to 5.c (after the minor manual editing described in the comments to the code)

What's next

- We sort of finished going through section 1 of the code at the seminar. Let me know if something remains unclear
- If there is interest in the topic, in January we can continue with reading through the rest of the code. Let me know of any good idea about doing this effectively
- Enjoy the festive season