

Introduction to R and Statistical Data Analysis

PART I

Petr Nazarov

petr.nazarov@crp-sante.lu

17-10-2011

◆ Information package (1)

◆ Installation (2)

◆ R interface (3)

- ◆ typing commands, calling functions, embedded help and demo

◆ Variables and basic operations (4)

- ◆ variables, types of data, scalar data, vectors, matrixes, data frames, lists.

◆ Data import and export (5)

- ◆ work folders, use scan, read/write tables, load/save data

◆ Control workflow and custom functions (6)

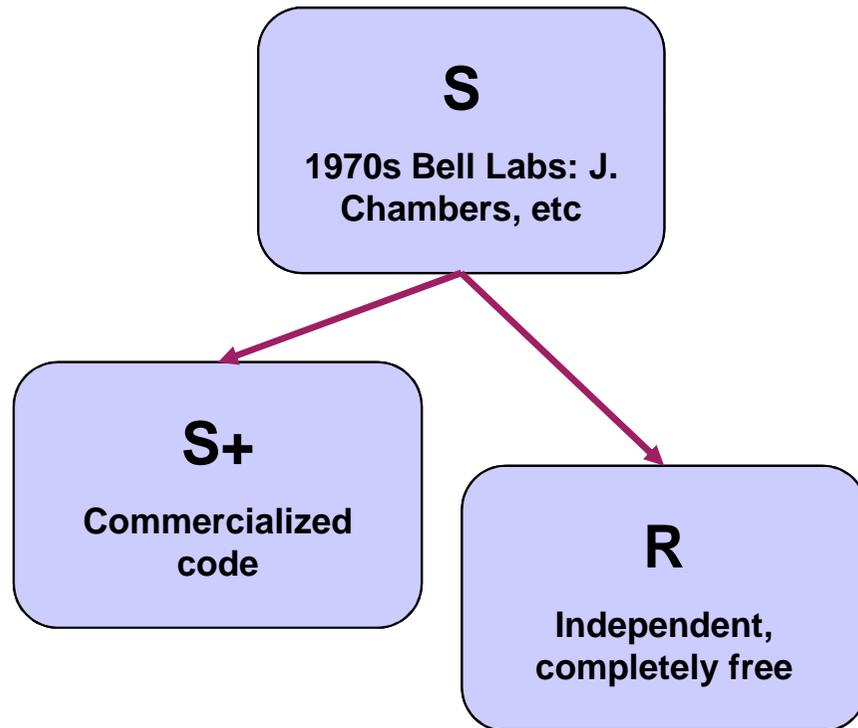
- ◆ if, while, repeat, next, break, custom functions, use external scripts

◆ Data visualization (7)

- ◆ variables, types of data, scalar data, vectors, matrixes, data frames, lists.

Look for corresponding scripts at
<http://edu.sablab.net/r2011/scripts>

History



R was created by Ross Ihaka and Robert Gentleman at the University of Auckland, New Zealand, and is now developed by the **R Development Core Team**.

Positive Features

- ◆ Scripting language of high level
- ◆ Interactive work with data (similar to MATLAB)
- ◆ Works under Windows, Linux, Mac OS
- ◆ Free code + verified algorithms
- ◆ Fast developing: new version each 2 month
- ◆ Extremely wide application: from biology to theoretical physics and computer sciences
- ◆ Very good information support
- ◆ Fast installation (with some exclusions)

Negative

- ◆ Not memory efficient (comparing to C)
- ◆ Slower then C/C++
- ◆ No advanced built-in GUI development tools

1. INFORMATION PACKAGE

Main Web-page:

cran.r-project.org

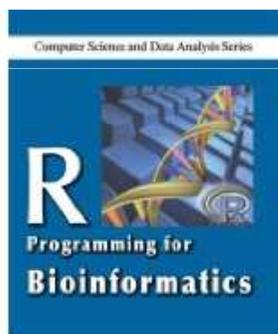
cran.r-project.org/manuals.html
cran.r-project.org/web/packages/
cran.r-project.org/other-docs.html

R-Project Seek Engine:

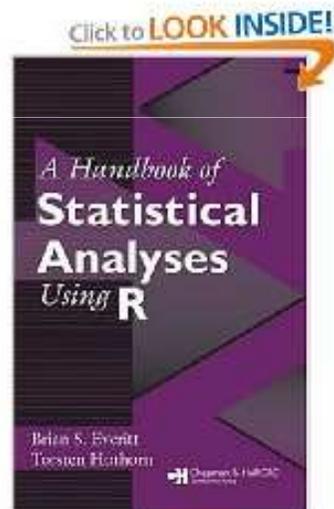
www.rseek.org

R/Bioconductor

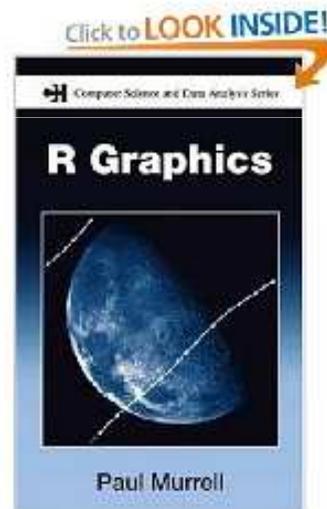
www.bioconductor.org



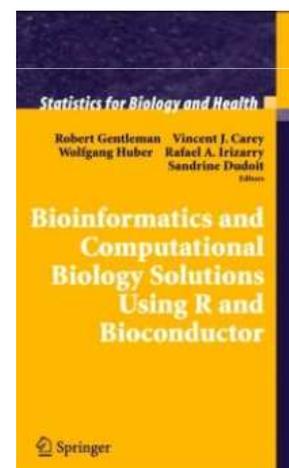
Robert Gentleman



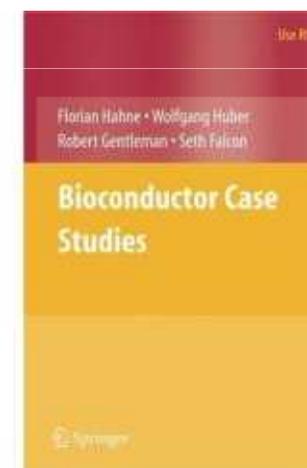
John Fox, Teunis Hothorn



Paul Murrell



Springer



Springer

All the materials are available here:

<http://edu.sablab.net/r2011>

Data can be downloaded from:

<http://edu.sablab.net/data/txt>

1. Download Binaries

<http://cran.r-project.org/bin/>

<http://cran.r-project.org/bin/windows/base/> (for Windows)

2. Install R (basic packages are automatically installed)

3. Run R and install additional packages (need Internet)

```
install.packages(package_name)
```

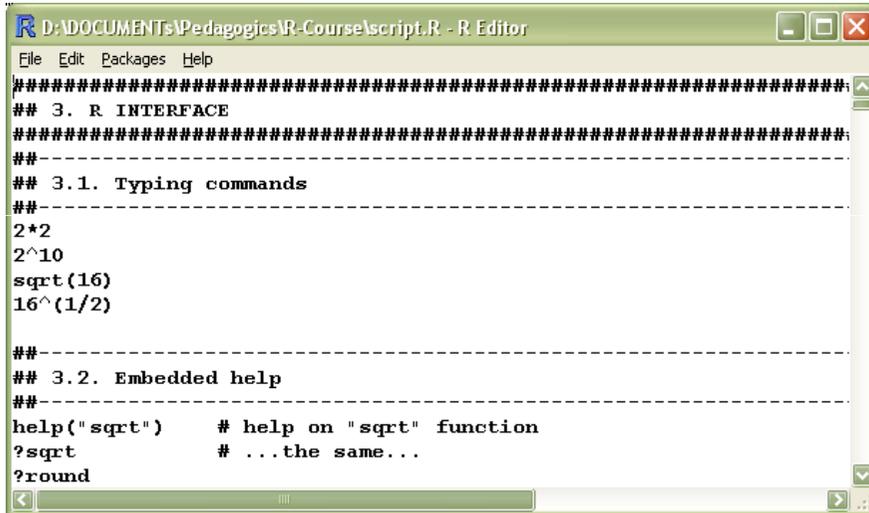
```
install.packages("rgl")
```

4. Another method: using Bioconductor tools

<http://www.bioconductor.org/install/>

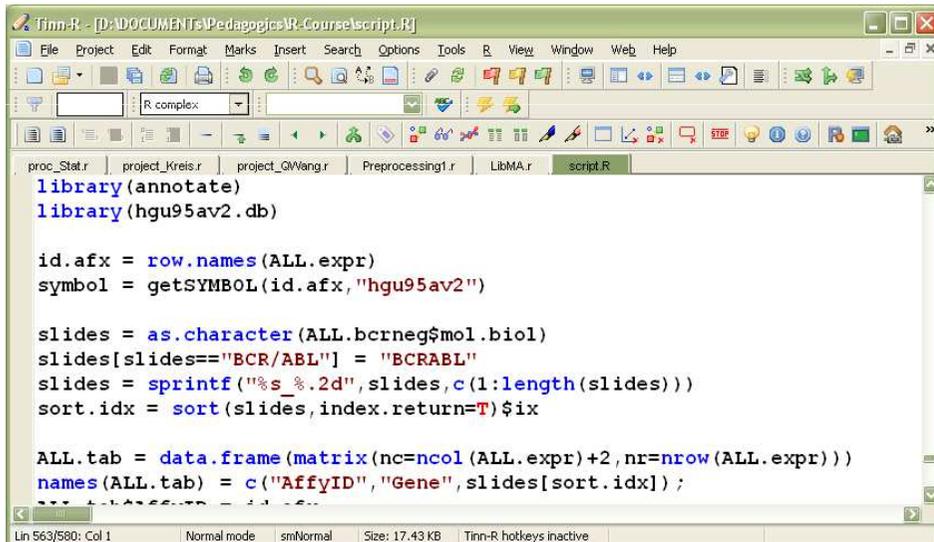
See more packages at <http://cran.r-project.org/web/packages/>

Built-in Script Editor



```
R D:\DOCUMENTS\PedagogicsR-Course\script.R - R Editor
File Edit Packages Help
#####
## 3. R INTERFACE
#####
##
## 3.1. Typing commands
##-----
2*2
2^10
sqrt(16)
16^(1/2)
##-----
##
## 3.2. Embedded help
##-----
help("sqrt") # help on "sqrt" function
?sqrt       # ...the same...
?round
```

Alternative: Tinn-R
(highly recommended for win-users)



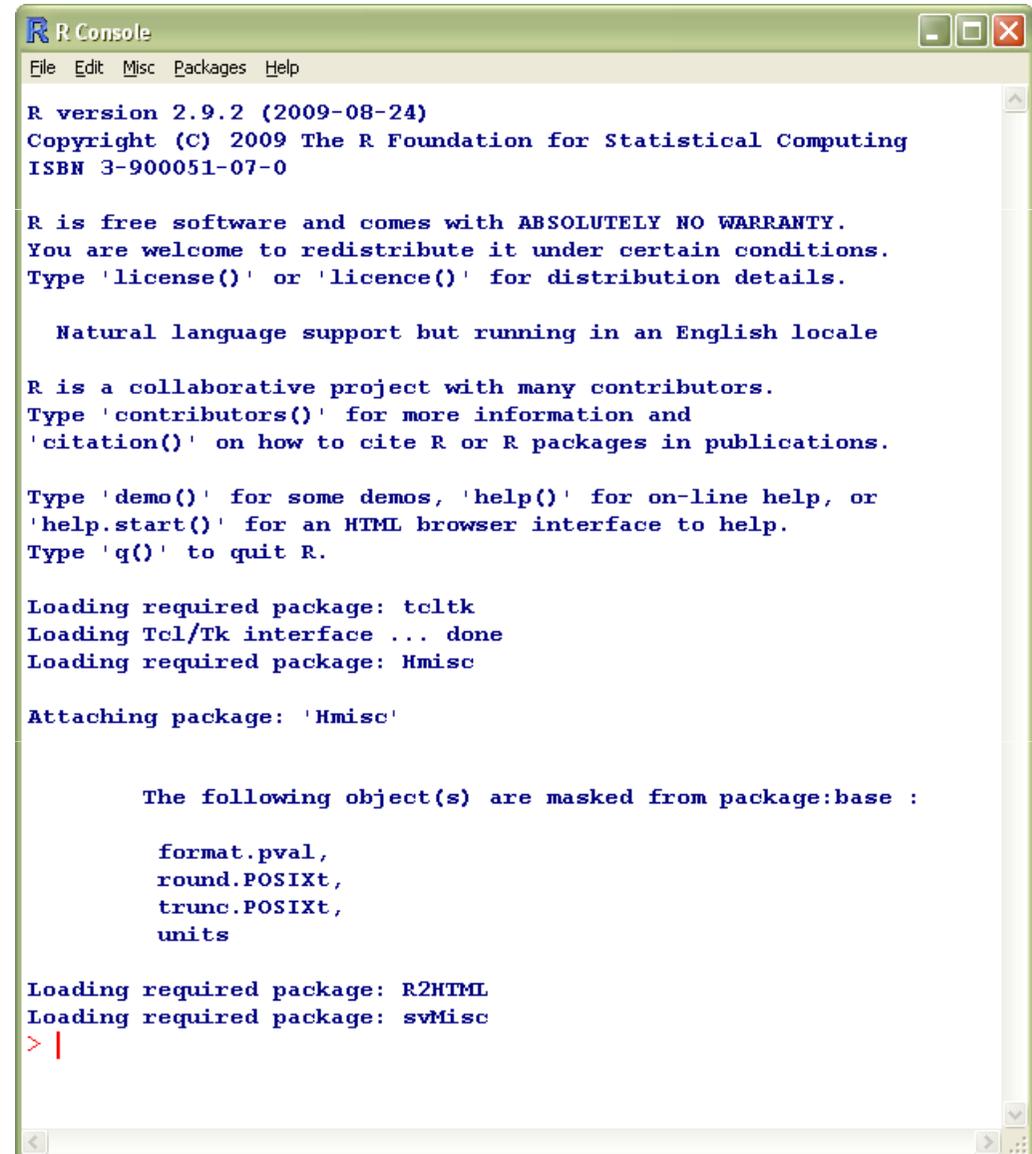
```
Tinn-R - [D:\DOCUMENTS\PedagogicsR-Course\script.R]
File Project Edit Format Marks Insert Search Options Tools R View Window Web Help
R complex
proc_Stat.r | project_kreis.r | project_QWang.r | Preprocessing1.r | LibMA.r | script.R
library(annotate)
library(hgu95av2.db)

id.afx = row.names(ALL.expr)
symbol = getSYMBOL(id.afx, "hgu95av2")

slides = as.character(ALL.bcrneg$mol1.biol)
slides[slides=="BCR/ABL"] = "BCRABL"
slides = sprintf("%s %.2d", slides, c(1:length(slides)))
sort.idx = sort(slides, index.return=T)$ix

ALL.tab = data.frame(matrix(nc=ncol(ALL.expr)+2, nr=nrow(ALL.expr)))
names(ALL.tab) = c("AffyID", "Gene", slides[sort.idx])
```

Console



```
R Console
File Edit Misc Packages Help

R version 2.9.2 (2009-08-24)
Copyright (C) 2009 The R Foundation for Statistical Computing
ISBN 3-900051-07-0

R is free software and comes with ABSOLUTELY NO WARRANTY.
You are welcome to redistribute it under certain conditions.
Type 'license()' or 'licence()' for distribution details.

Natural language support but running in an English locale

R is a collaborative project with many contributors.
Type 'contributors()' for more information and
'citation()' on how to cite R or R packages in publications.

Type 'demo()' for some demos, 'help()' for on-line help, or
'help.start()' for an HTML browser interface to help.
Type 'q()' to quit R.

Loading required package: tcltk
Loading Tcl/Tk interface ... done
Loading required package: Hmisc

Attaching package: 'Hmisc'

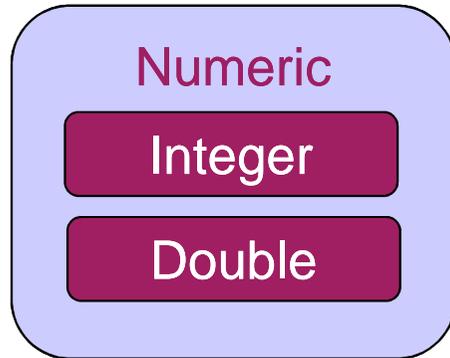
The following object(s) are masked from package:base :

format.pval,
round.POSIXt,
trunc.POSIXt,
units

Loading required package: R2HTML
Loading required package: svMisc
> |
```

```
#####  
## 2. INSTALL R PACKAGES  
#####  
install.packages("rgl")  
#####  
## 3. R INTERFACE  
#####  
##-----  
## 3.1. Typing commands  
##-----  
2*2  
2^10  
sqrt(16)  
16^(1/2)  
##-----  
## 3.2. Calling functions  
##-----  
log(100)  
log(100, base=10)  
log(100, b=10)  
log(100, 10)  
##-----  
## 3.3. Embedded help  
##-----  
help("sqrt")    # help on "sqrt" function  
?sqrt          # ...the same...  
?round  
??round        # fuzzy search for "round" in all help topics  
apropos("plot") # propose commands with the word "plot" inside the name  
## Demos  
demo()         # show available demos  
demo("image")  # start demo "image"  
demo(persp)  
demo(plotmath)
```

Scalar Data



```
1
3.141593
```



```
TRUE
FALSE
```



```
"Hello, world!"
```



```
> answer=factor(c("yes", "no"))
> answer
[1] yes no
Levels: no yes
```

has a sense to use only in vectors or data frames

Data Containers



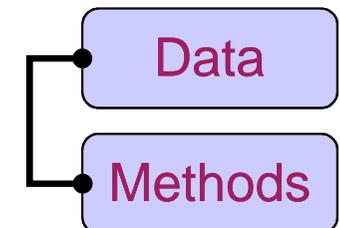
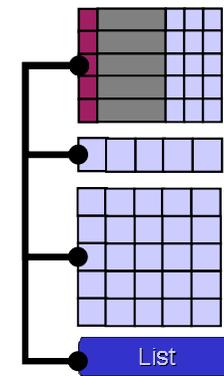
```
> x
[1] 1 2 3 4 5
```



```
> A
  [,1] [,2] [,3]
[1,]  1   3   5
[2,]  2   4   1
```



| | name | marks |
|---|-------|-------|
| 1 | Alex | 10 |
| 2 | Jean | 8 |
| 3 | David | 7 |



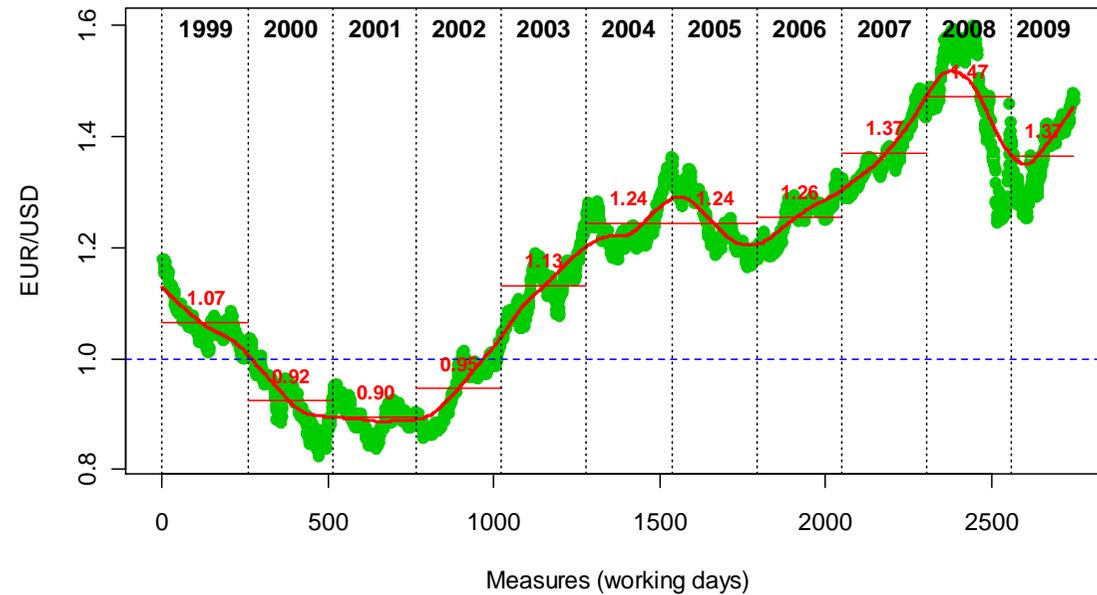
currency.txt

R Data Editor

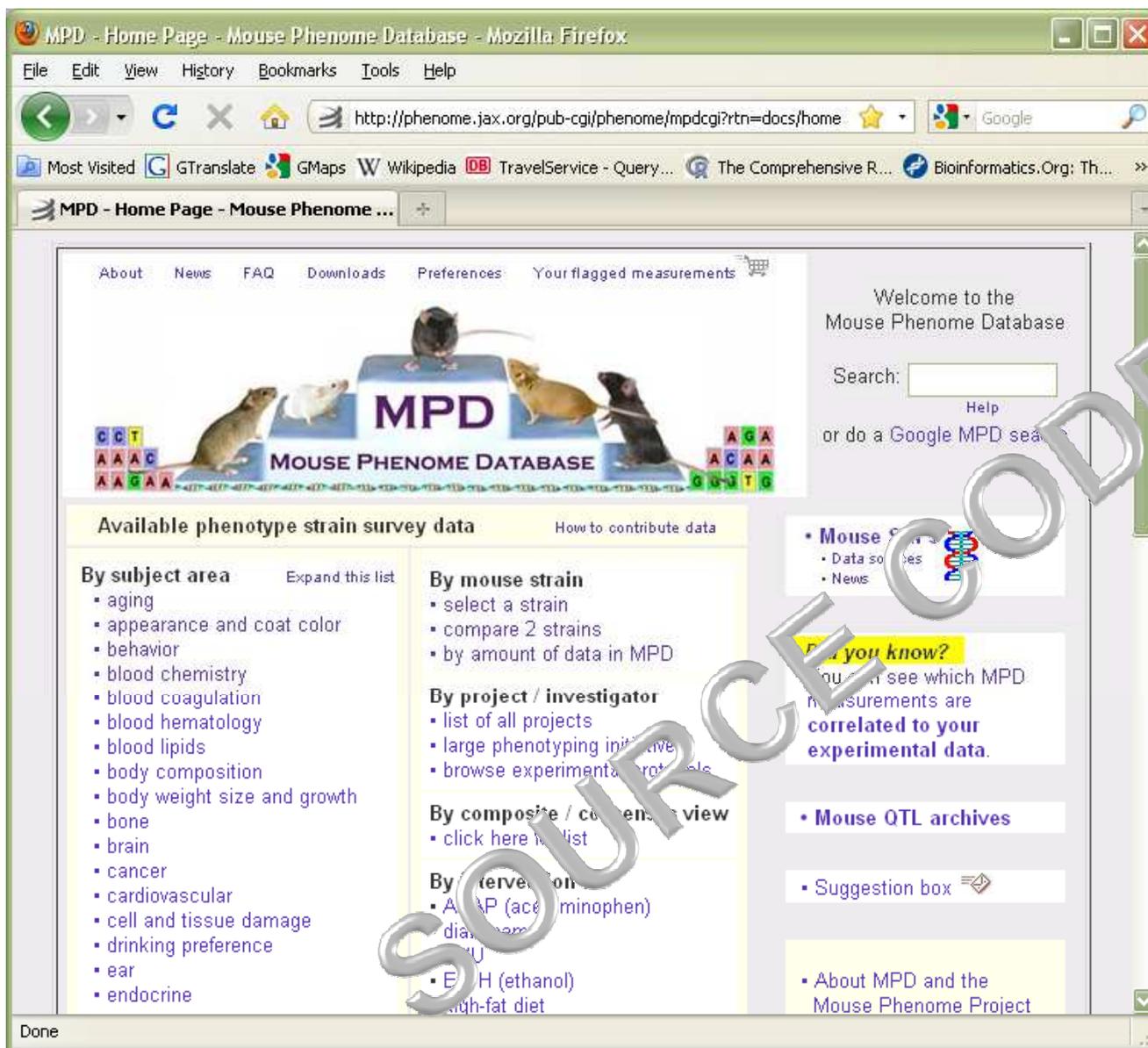
File Edit Help

| | Date | EUR |
|----|------------|--------|
| 1 | 1999-01-04 | 1.1789 |
| 2 | 1999-01-05 | 1.179 |
| 3 | 1999-01-06 | 1.1743 |
| 4 | 1999-01-07 | 1.1632 |
| 5 | 1999-01-08 | 1.1659 |
| 6 | 1999-01-11 | 1.1569 |
| 7 | 1999-01-12 | 1.152 |
| 8 | 1999-01-13 | 1.1744 |
| 9 | 1999-01-14 | 1.1653 |
| 10 | 1999-01-15 | 1.1626 |
| 11 | 1999-01-18 | 1.1612 |
| 12 | 1999-01-19 | 1.1616 |
| 13 | 1999-01-20 | 1.1575 |
| 14 | 1999-01-21 | 1.1572 |
| 15 | 1999-01-22 | 1.1567 |
| 16 | 1999-01-25 | 1.1584 |
| 17 | 1999-01-26 | 1.1582 |
| 18 | 1999-01-27 | 1.1529 |
| 19 | 1999-01-28 | 1.141 |
| 20 | 1999-01-29 | 1.1384 |

EUR/USD ratio for 11 years



SOURCE CODE



Tordoff MG, Bachmanov AA
 Survey of calcium & sodium intake and metabolism with bone and body composition data
 Project symbol: **Tordoff3**
 Accession number: **MPD:103**

mice.txt

**Thank you for your
attention**

to be continued...