

An introduction to 

# Background Material and Homework Exercises


Ronald B. Geskus  
Academic Medical Center  
Amsterdam  
the Netherlands

email: `statistics@inter.nl.net`


November 2008


R is extremely powerful, yet not easy to master in the beginning. The main disadvantage compared to other statistical programs is the very basic user interface. In this short course, I want to explain the basic concepts of R. Of course, learning to use R in three hours is not a realistic goal. However, my hope is that once you have obtained some basic knowledge of R, you can learn more by yourself if you know where to find the relevant information. Therefore, I will also explain where extra information can be found and how your specific questions may be answered.


## Preliminary Work

Before the start of this short course, I ask you to have  installed on your computer and to have obtained some very basic familiarity with the program. This will take about one hour. For this, you need to have an internet connection. I assume you use MS Windows.


**Installation** Go to the website <http://www.r-project.org>. In the column on the left hand side, click on the CRAN link under the subheading **Download**. The right hand side changes to a page with “CRAN Mirrors”. This is a list of computers from which the R software and extensions can be downloaded. Click on a link (e.g. the one in the Netherlands). Now both the left hand side and the right hand side of the page change. (This page can be saved as a bookmark/favorite.) On the right hand side, click on Windows under the subheading **Download and Install R**. Then click on base and download/install R-2.8.0-win32.exe. You can use the default installation options.

**First Steps** Start the program. One window inside the main  window is opened. Read the text in the **R Console** window. In the console window, write the text `demo(graphics)` after the “ $\geq$ ”-sign and hit the *Return/Enter* key. A graphics window is opened. Hitting *Return/Enter* changes the graph that is displayed. After a few more graphs, the demo is finished. Next write the text “`help.start()`” in the console window. A page is opened in your web browser with links to documentation, the installed packages, a search engine and the FAQ.

**Install Extra Packages** We will use some extra packages that need to be installed from the internet as well. The easiest way is through the menu in . Click on the **Packages** menu and select **Select CRAN mirror...** Again a download computer location needs to be chosen. Next click on **Install package(s)...** in the same **Packages** menu. A long list of available additions (called *packages*) is shown. Install the packages `Rcmdr`,

`RcmdrPlugin.Export`, `RcmdrPlugin.HH` and `RcmdrPlugin.TeachingDemos`. (They can be selected together via the standard MS Windows procedure using the **Shift** or **Ctrl** key.) This will take a couple of minutes, so in the mean time have a further look at the  web site (see next topic below). As soon as installation is finished, go to **Open the Rcmdr GUI** below.

**A Further Look at the Website** Go back to the CRAN mirror. On the left hand side, several links are of interest. The most interesting one is the link `Contributed` under **Documentation**. Here you can find weeks of freely downloadable reading material. Also have a look at the FAQs one day you start using R seriously. Other interesting links are `Packages` under **Software** and `Task Views and Search` under **CRAN**.

**Open the Rcmdr GUI** In the R console window, write `library(Rcmdr)`. You may be asked to install some extra packages. Click “yes”. This will again take several minutes. In the mean time, we can have a further look at the  web site. After this, an extra window called “R Commander” is opened. Make this a full screen window. Open a data set of your own via the menu **Data** → **Import Data**. Play with the data, e.g. via “View data set” or some of the options under the **Statistics** or **Graphs** menu.

**End Exercise** Close the program, either via the menu **File Exit** under the “R commander” window, or by typing “q()” in the R console window.

## Exercises

### Veni

We will do some simple analyses on the `Arrests` data, which comes with the `effects` package. If you installed the `Rcmdr` package, the `effects` package has been installed as well. But we do need to make the package available via `library(effects)`.

1. Investigate the meaning of the variables via `help(Arrests)`.
2. Have a look at the first ten rows of the data set.
3. Make a summary of the data (use `summary`).
4. Make a simple 2 by 6 frequency table of the number of people released/brought to police station per year (hint: use `table`).

5. Test whether blacks and whites have the same age distribution in the data; perform a t-test as well as a Mann-Whitney test (use `help.search` if you don't know which functions to use)
6. Perform a chi-squared test for independence between race and release status.
7. A data frame is a list with a class "data.frame" assigned. Select the records from individuals aged 60 years or older and store it in an object named `tmp.x`. Create a new object `tmp` that only contains the columns `released` and `colour` of the data set `tmp.x`. Compare `print(tmp)`, `tmp` and `print.default(tmp)`. Can you explain the differences and similarities?


## Vidi





We again use the Arrests data.

- Plot age as a function of year using the standard `plot` function, and save the graph as a pdf file.
- Check whether "year" is a factor. If not, make it into a factor.
- Again plot age as a function of year. Can you explain the difference in type of graph?
- Make a histogram of the age distribution of the blacks and the whites using the `lattice` package.
- Fit a logistic regression model that predicts whether someone will be released based on individual characteristics. Use a model with main effects for "employed", "citizen" and "checks", and in which the effects of age and year are allowed to depend on colour. (Hint: use the `glm` function with argument `family` set to `binomial`.)
- Test whether the overall effect of age is significant.

## Vici

No homework exercises, but here are some suggestions for further reading:

- A basic introduction to , somewhat in line with my course, can be found at <http://www.math.ilstu.edu/dhkim/Rstuff/Rtutor.html>

- A further introduction to  is found on <http://www.statmethods.net/index.html>.  
This page shows the commands to be used for many aspects of a statistical analysis.
- There is a book called “R for SAS and SPSS users”. Its website is <http://rforsasandspssusers.com/>.  
This web site also has a freely downloadable early version of the book. It also has a comparison table that lists for many SAS and SPSS procedures the corresponding R packages.
- Further options are [http://zoonek2.free.fr/UNIX/48\\_R/all.html](http://zoonek2.free.fr/UNIX/48_R/all.html),  
and the information on the R web site:  
<http://cran.r-project.org/other-docs.html>,  
<http://www.r-project.org/doc/bib/R-publications.html>,  
the R newsletter (under Documentation)  
and the manuals <http://cran.r-project.org/manuals.html>.
- If you want to experiment with a GUI for , have a look at <http://www.sciviews.org/SciViews-R>,  
<http://www-sre.wu-wien.ac.at/SimpleR/>  
or <http://stats.math.uni-augsburg.de/JGR/>.  
There also exists a commercial GUI for , called R-Plus,  
<http://www.experience-rplus.com/>;  
a free trial can be downloaded.
- Editors for  are  
Tinn-R (<http://www.sciviews.org/Tinn-R/index.html>),  
Relax (<http://www.wiwi.uni-bielefeld.de/~wolf/software/relax/relax.html>)  
or WinEDT (<http://www.winedt.com/>) together with the RWinEDt  
plug-in, see  
<http://cran.r-project.org/web/packages/RWinEdt/vignettes/RWinEdt.pdf>.
- A web site with many beautiful graphs and the corresponding R code is <http://addictedtor.free.fr/graphiques/>.
- There exist programs that allow to run R from Excel (<http://rcom.univie.ac.at/>) and from SPSS (see the SPSS web site, search on R plug-in or look at <http://www.spss.com/devcentral/>).