

Working with R in batch mode: Some notes for beginners

Michael Lundholm*

January 2, 2013
Version 1.2

1 Introduction

The purpose of this note is help beginners to R¹ to use the software in *batch mode*. A majority of beginners (in particular students) are using computers with some Windows operating system (say Windows XP or Windows 7). A minority are using computers with the Mac OS X operating system (say Snow Leopard, Lion or Mountain Lion). For both Windows and Mac OS X, R is shipped with a graphical user interface (GUI), which if started invokes the R console which invites to an interactive use of R; it is there commands and instructions are written by the user and the results are displayed. The user writes commands, gets the output, write new commands etc.²

The standard GUI for the Mac OS X system is the app **R.app** and can be seen in Figure 1 on page 2. The Windows GUI is **Rgui.exe** and can be seen in Figure 2 on page 3.³ Similar GUI for interactive use of R are also supplied by third parties. One example, available for not only Windows and Mac OS X system, but also GNU/Linux systems, is RStudio.⁴

The interactive use of R is not without pitfalls as the following simple example shows:

Example 1 An interactive R session contains the following commands:

*Department of Economics, Stockholm University, michael.lundholm@ne.su.se.

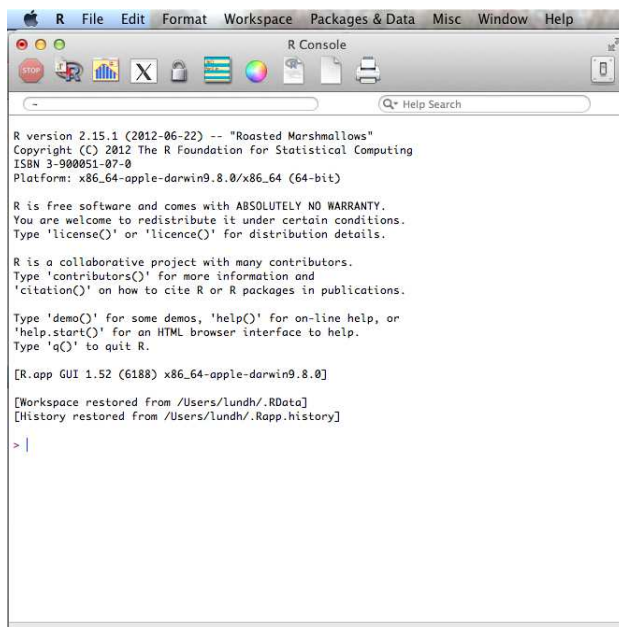
¹R Core Team (2012). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL <http://www.R-project.org/>.

²Users of GNU/Linux systems do probably not need the information in this note, although section 2 can be helpful.

³For both operating systems the GUI exists in 32 and 64 bit versions. The Mac OS X 64 bit version is called **R64.app** but the Windows version is also called **Rgui.exe** but resides in another directory. In terms of the GUI there is no difference between the 32 and 64 bit versions.

⁴<http://rstudio.org>.

Figure 1: The R graphical user interface and its console for Mac OS X.



```
> data(airquality, package = "datasets")
> Ozone <- airquality$Ozone
> M.Ozone <- median(Ozone, na.rm = TRUE)
> Low.Ozone <- Ozone < M.Ozone
> mean(Low.Ozone, na.rm = TRUE)
```

```
[1] 0.5
```

The user then saves the commands as an R script, in a file named `ozone.r`. However, a mistake is made and the second row above defining the variable `Ozone` is missing in the script. The script has the following content:

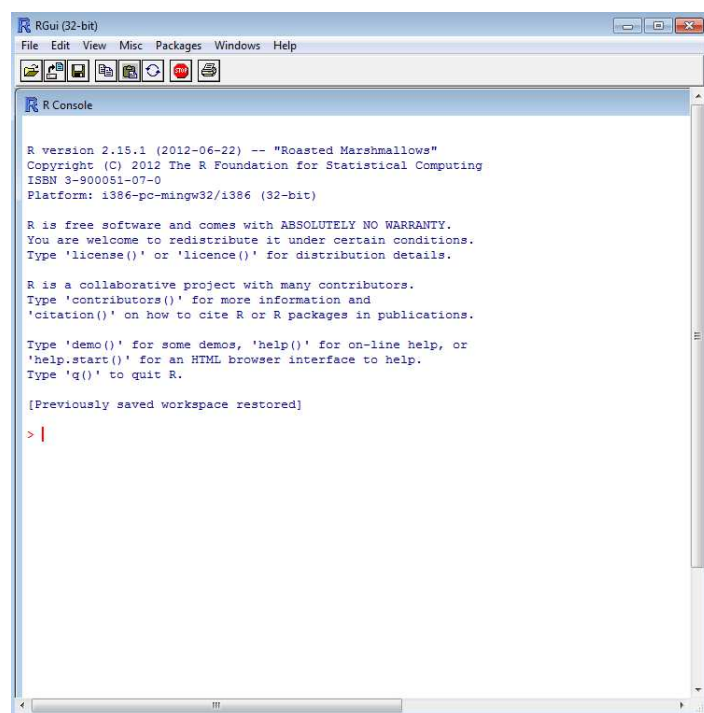
```
data(airquality, package = "datasets")
M.Ozone <- median(Ozone, na.rm = TRUE)
Low.Ozone <- Ozone < M.Ozone
mean(Low.Ozone, na.rm = TRUE)
```

In order to check that the script is working the user adds the following command to the R-session:

```
> source("ozone.r")
```

Since the script is incomplete because one row is missing, there should be an error message. However, no error message appears because the `Ozone`

Figure 2: The R graphical user interface and its console for Windows.



variable is defined in the running R session. ■

In order to get an indication, via an error message, that the script is incomplete we need a way to run R without the information that is stored in R's memory. Basically, that is done by running R in batch mode.

Batch mode is a way to use the software without using the graphical user interface and the console. Generally batch mode or *batch processing*⁵ is when a program is executed on a computer without manual intervention. More specifically for the R-context a *program* is a sequence instructions to R, saved in an R script file, which in batch mode are executed starting with a clean (empty) R-session. By executing the instructions in the script in batch mode we can check the script for internal consistency. A commented script thus checked for internal consistency is an integer part of all good procedures of *reproducible econometrics*.⁶

These notes are organised as follows: In section 2 R's batch mode will be discussed. Section 3 will give practical advice how R is invoked from the command prompt on Windows systems and from the terminal on Mac OS X systems. Finally, section 4 will give some advice regarding the working environment for R users in Windows and Mac OS X that do not want to use the graphical user interfaces.

2 R's batch mode

Although the GUI's are most common among Windows and Mac OS X users the most basic way of invoking R is via the command line.⁷ Disregarding exactly how we do this on various systems (we return to it in the next section) we use the command syntax:

```
R [options] [<infile] [>outfile]
```

where **options** are discussed below, **infile** is the name of the file containing R code and **outfile** is the name of the file that will contain the R output.

A number of options exists but for our purposes only a few need to be mentioned:⁸

--save

--no-save "Control whether data sets should be saved or not at the end of the R session. ... in non-interactive use one of these must be specified or implied by some other option ...".

⁵http://en.wikipedia.org/wiki/Batch_processing.

⁶Koenker R, Zeileis A (2009). On Reproducible Econometric Research, Journal of Applied Econometrics, 24(5), 833-847. doi:10.1002/jae.1083.

⁷See Appendix B.1, pp.81ff in R Core Team (2012).

⁸R Core Team (2012), p.89.

`--restore`

`--no-restore` “Control whether saved images (file `.RData` in the directory where R was started) should be restored at startup or not. The default is to restore.”

`--vanilla` Combines `--no-save` and `--no-restore` (and also prevents R to read a set of other files which else are read by default).

Note that the default behaviour is to read any `.RData` file in the directory from which R is started (i.e., the working directory). Since we probably have saved our work this file may contain objects and data that relate to our script. However, we want R to execute our script with nothing in its memory so we must invoke R without reading this file. The easiest way to do this is to use the `--vanilla` option since it also prevents R to do a lot of other things.⁹ The preferred way to invoke R therefore is:¹⁰

```
R --vanilla <infile >outfile
```

3 Starting R from the command–prompt/terminal

How do we practically invoke R in batch mode using the information from the previous section? Instead of using the GUI we use the so called *command prompt* on Windows systems and *terminal* on Mac OS X systems.¹¹

3.1 Windows and the command prompt

The program called `Command prompt` on Windows is found under at the Start button `All programs - Accessories`. Double-click on its icon so that it starts. We will see a window looking as the one in Figure 3 on page 6.

We can now enter commands to the command prompt.¹² For instance we can try to start R. On Windows systems R comes in several versions; one is the graphical user interface `Rgui.exe` and another is `R.exe`; we will return the various versions below. We try to start the latter by giving the command `R` at the command prompt; see Figure 4 at page 6.

We see that we get an error message:

```
‘R’ is not recognized as an internal or external command,  
operable program or batch file.
```

⁹For the meaning of the term *vanilla*, see http://en.wikipedia.org/wiki/Vanilla_software.

¹⁰The alternative `R CMD BATCH [options] infile outfile` by default invokes R with the options `--save` and `--restore` which is opposite to `--vanilla`. Do not use it if you want to run your script in a clean R session! See R Core Team (2012), p.91 and `?BATCH` in the R console.

¹¹In the following “Windows” is taken to mean *Windows 7*.

¹²For a documentation of the various commands, see e.g., http://en.wikipedia.org/wiki/List_of_MS-DOS_commands.

Figure 3: The Windows command prompt.

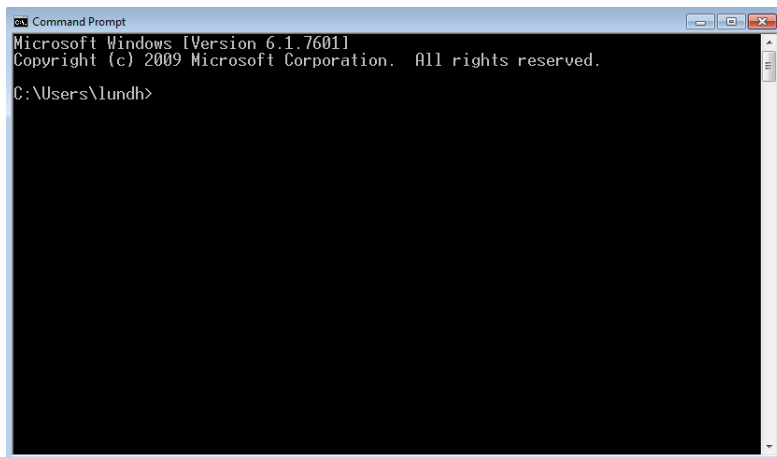
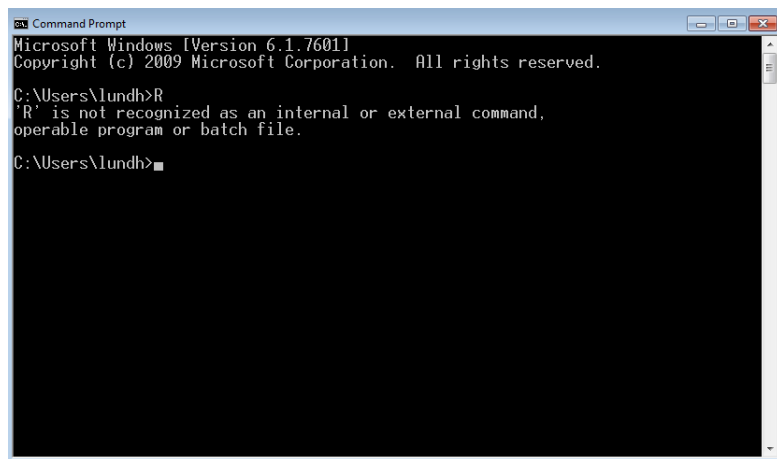


Figure 4: Trying to start R at the command prompt.



In plain words: Although R is installed the Windows system does not find the file **R.exe**. This is because this file is in a directory which is not in the *default search path* of Windows.¹³ The reason is that the default behaviour of the installation script is to install new version of R in different directories. It is up to the user to decide where Windows should look for R (i.e., which version to use).¹⁴

If this error message occurs we have to change the search path so that it includes the path to the directory where **R.exe** is.¹⁵ We do this as follows:

1. Open the Control Panel, choose 'System', click on 'Advanced system settings' in the menu to the left and choose the 'Environment variables' button; see Figure 5.
2. Choose the variable 'Path' and click on the Edit-button. Now we have to edit 'Variable value' so that it also contains the path to the directory where the various R binaries (programs) are installed.
3. The path to this file is found in Windows explorer by looking for the directories where R was installed. See Figure 6. Typically the path is something like

`C:\Program Files\R\R-2.15.1\bin\x64`

and it should be appended to the already existing path in 'Variable value'.

4. Note that the various paths in the search path are separated with a semi-colon so we should add an initial ';' to the string characters that we append to 'Variable value'. The result can be seen in Figure 7.

Now we are ready to start R from the command prompt, but before we do that we need to explore the various versions of R that exist under Windows. If we have a look at Figure 6 we have a long list of files that have type **Application**; all these files have the file extension **exe**. The various versions of R are:

Rgui.exe The standard GUI. Double-clicking on the file will start the graphical user interface.

Rterm.exe An R console. Can be used for interactive use like the GUI, although the user interface is text based.

¹³M. Lundholm, Some notes on file names, version 1.3, 2012-05-31, page 7. <http://people.su.se/~lundh/reproduce/fnames13.pdf>.

¹⁴See FAQ 2.30 in 'R for Windows FAQ. Version for R-2.15.1', url <http://ftp.sunet.se/pub/lang/CRAN/bin/windows/base/rw-FAQ.html>.

¹⁵The following for Windows 7 is based on <http://www.computerhope.com/issues/ch000549.htm> (with some modification), but information about the case Windows XP can be found in the same source.

Figure 5: The Windows Control Panel.

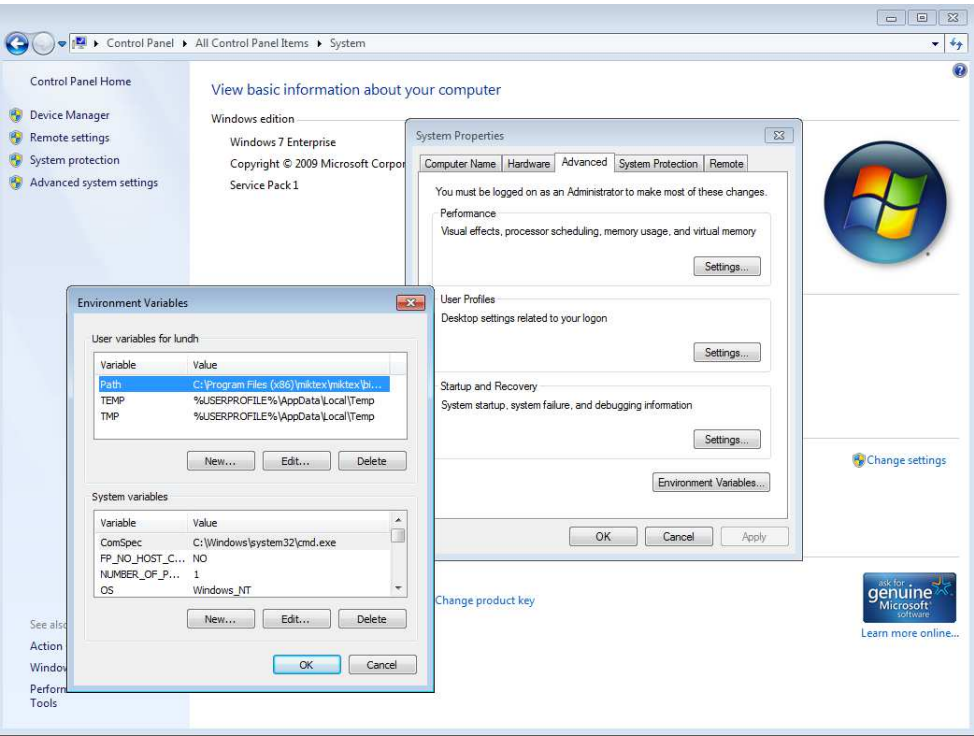


Figure 6: The path to the R binaries (programs) in Windows Explorer.

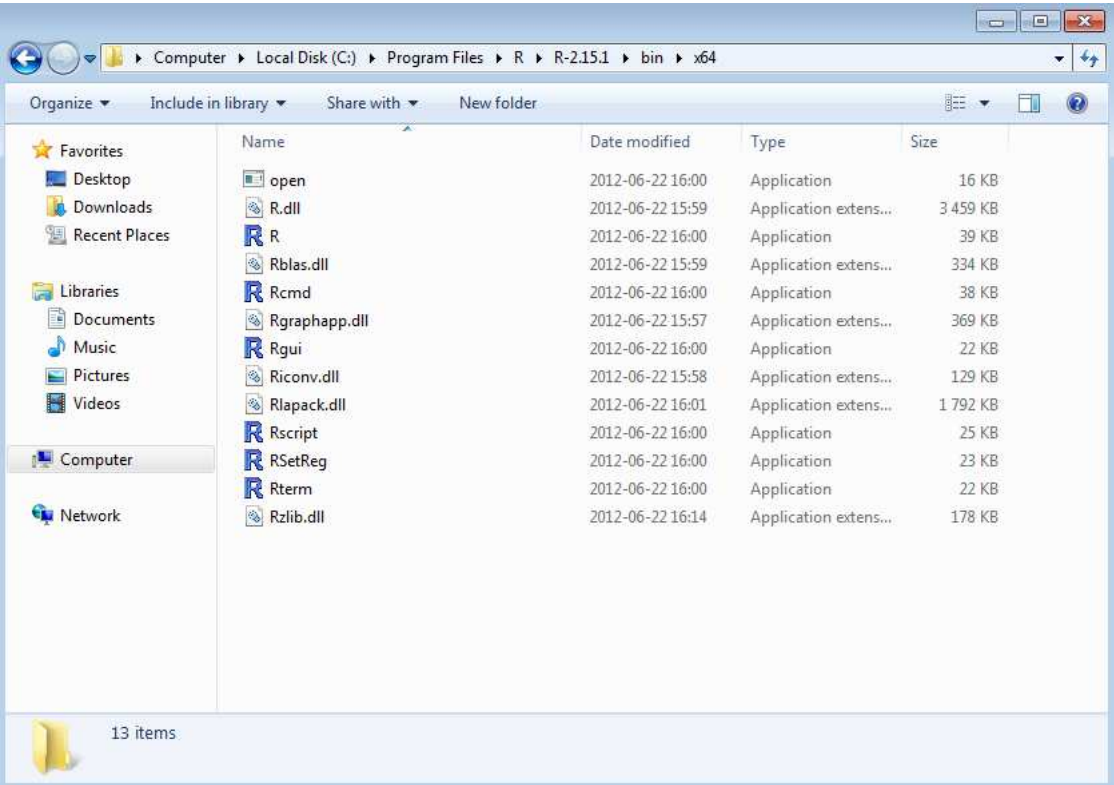
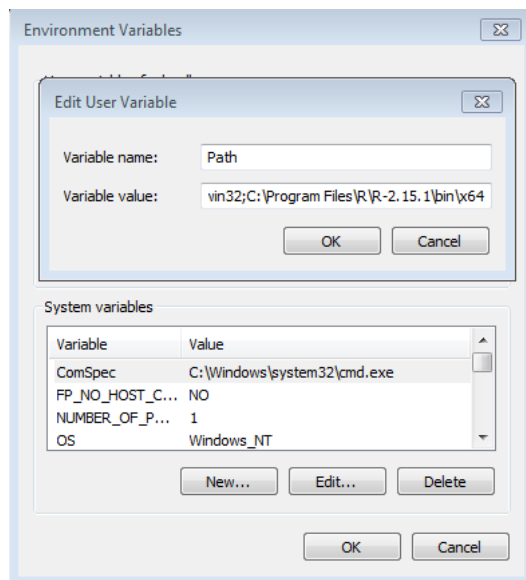


Figure 7: ‘Variable value’ now appended with the path to to `R.exe`.



R.exe At the command prompt the command `R` will open an R console exactly as `Rterm.exe`, but it can also be used to execute R scripts as discussed in the previous section. Of all these binaries it is most similar to R on Mac OS X (and GNU/Linux and other UNIX derivatives) and the one that is most suitable for our purposes.

Rcmd.exe At the command prompt it provides an interface to R corresponding to R `CMD` `command` `args`.

Rscript.exe The version of R that can be used for advanced scripting together with other software. It is, however, not suitable for our purposes.

You may want to investigate the differences between `R.exe`, `Rcmd.exe` and `Rscript.exe` when they are executed from the command prompt to learn about there differences.

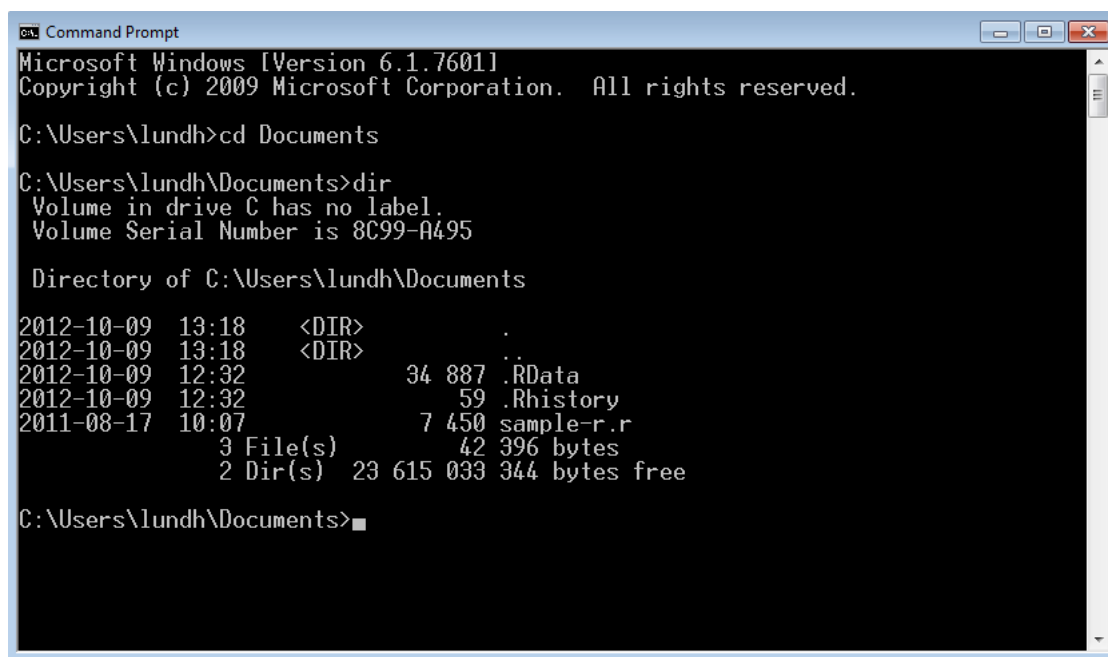
We are going to use `R.exe` but before we can do that we need to secure that R is started from the correct directory (i.e., the directory where the script file is located). The following command prompt commands are useful for the navigation:¹⁶

`dir` Lists the content of a directory.

`cd..` Change to the parent directory. Note that there is no space before the two dots.

¹⁶See footnote 12 for a reference additional commands.

Figure 8: Changing directory and list directory content at the command prompt.



```
Microsoft Windows [Version 6.1.7601]
Copyright (c) 2009 Microsoft Corporation. All rights reserved.

C:\Users\lundh>cd Documents

C:\Users\lundh\Documents>dir
Volume in drive C has no label.
Volume Serial Number is 8C99-A495

Directory of C:\Users\lundh\Documents

2012-10-09 13:18 <DIR>          .
2012-10-09 13:18 <DIR>          ..
2012-10-09 12:32             34 887 .RData
2012-10-09 12:32             59 .Rhistory
2011-08-17 10:07             7 450 sample-r.r
                3 File(s)      42 396 bytes
                2 Dir(s)  23 615 033 344 bytes free

C:\Users\lundh\Documents>
```

`cd dirname` Change directory to child directory *dirname*.

Using these commands we can list the content of directories and navigate up and down in the directory tree.

In Figure 3 we see that the command prompt starts in the home directory of the user. Suppose scripts are in the child directory **Documents** directly below the home directory of the user. Assume we have restarted the command prompt after we provided the search path to the binaries. We then change directory and list the contents of the directory by invoking the commands; see also Figure 8 on page 11.

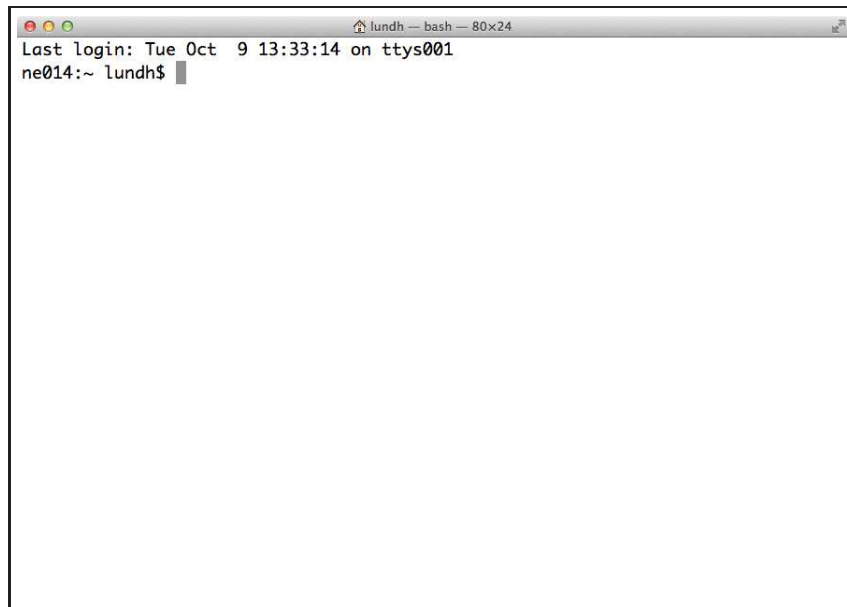
```
cd Documents
dir
```

In this directory we find the script file **sample-r.r**. We execute in batch mode using the vanilla option:

```
R --vanilla <sample-r.r >sample-r.out
```

All code comments, messages and regular output is now directed into the file **sample-r.out** which is created by R in the working directory.

Figure 9: The terminal in Mac OS X.



3.2 Mac OS X and the terminal

The Mac OS X correspondence to the command prompt is called the *terminal* and is found under **Applications - Utilities**; i.e., the **Terminal.app**. It can be started either via Finder or via Launchpad. If started it will generate a window like in Figure 9

In the terminal we can navigate up and down the directory tree and list content in a similar way to command prompt in Windows:¹⁷

`ls` Lists the content of a directory.

`ls -la` Lists the content of a directory but more information about files.

`cd ..` Change to the parent directory. Note that there is a space before the two dots.

`cd dirname` Change directory to child directory **emphdirname**.

`cd /` Change directory to the root (top) directory.

`cd` Change directory to the users home directory.

¹⁷The terminal runs a so called BASH (Bourn Again Shell) shell and is more versatile than the command prompt. For more information about BASH commands and BASH scripting, see <http://tldp.org/HOWTO/Bash-Prog-Intro-HOWTO.html>.

Once we navigated to the directory containing the script file we can use the same procedures as under Windows:¹⁸

```
R --vanilla <sample-r.r >sample-r.out
```

Note that we need to bother about the search path on MAC OS X systems, because installation procedures are different and the R binaries are already in the search path.

¹⁸Users with Mac OS X 10.6 (Snow Leopard) or later may want to consider the app **Go2Shell** available on App Store. It allows the user to open the terminal in a directory selected via Finder.