

Introduction to data analysis in R

(STATS240.1: A short course in R)

Gunnar Stefánsson, Ásta Jenný Sigurðardóttir and Lorna Taylor

October 12, 2012

Starting up

Linux: Enter "R" at the shell prompt

Windows: Find R in the startup menu

Before you begin

">" is the R prompt

Comments are marked with #

To get detailed information about a function, type:

```
> help("function name")
```

Notice that most R commands are actually calls to **functions** which implies that they are followed by parentheses.

Data vectors in R

A typical session

```
> x<-42  
> x  
[1] 42  
>
```

The object `x` contains the single number 42. Typing the name of an object displays the content.

Some commands

```
x<-c(1,5,3,6)  
myseq<-1:5  
longseq<-10:100  
fractions<-(1:150)/100  
x<-1:4  
y<- -1:4
```

Deleting and listing objects

`ls()` - list objects

`rm()` - remove objects

Entering data into a file

First enter data using an editor

- Windows: Notepad or emacs
- Linux: Emacs or vi

Reading data into R

Data in a file:

abc	xyz
1	4
2	5
15	3

Read them into a data frame:

```
> mydata<-read.table("x.dat",header=T)
```

Type the name to display the data

```
> mydata
  abc xyz
1   1   4
2   2   5
3  15   3
```

Data summaries in R

mean Means of individual columns
var Variance of individual columns
sd Standard deviation of individual columns
median Median of individual columns

```
x<-mydat$x  
y<-mydat$y  
mean(x)  
var(x)  
sd(x)  
median(x)
```


Random numbers in R

Random numbers are very useful for checking out properties, sampling schemes etc.

Common functions:

```
rnorm()
```

```
runif()
```

Some examples:

Random normal (Gaussian), single number:

```
rnorm()
```

Uniform between 0 and 1, ten numbers:

```
runif(10)
```

Single observation from a binomial distribution with 5 trials each a probability 1/2 of success:

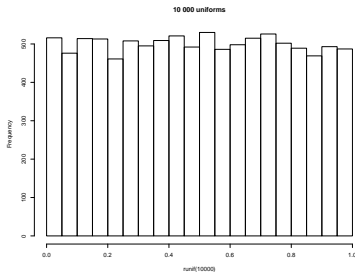
```
rbinom(1,5,.5)
```

Draw random sample of 5 with replacement from 10, with replacement:

```
sample(c("a","b","c","d","e","f","g","h"),10,replace=T)
```

Plot histogram of 100 numbers from U(0,1)

```
hist(runif(100))
```



Simple data plots in R

```
x<-1:10
z<-2+3*x
y<-z+rnorm(10)
plot(x,y,
      xlim=c(0,10),ylim=c(0,30),
      xlab="Cost",ylab="Profit")
lines(x,z)
```

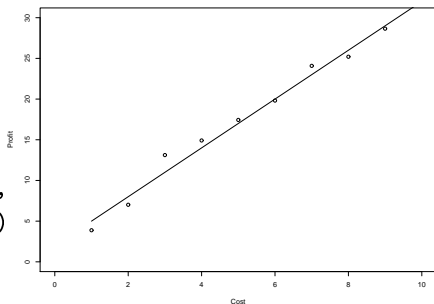
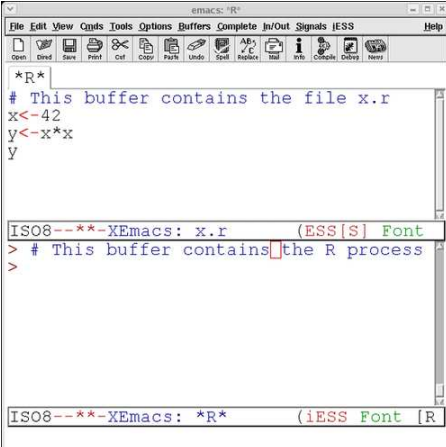


Figure: A typical scatterplot from R

R and emacs

R/S-mode for .r-files
Interactive R/S startup within
emacs!



The screenshot shows the Emacs editor window titled "emacs: 'R'". The menu bar includes File, Edit, View, Cmds, Tools, Options, Buffers, Complete, In/Out, Signals, iESS, and Help. The toolbar contains icons for Open, Dired, Save, Print, Cut, Copy, Paste, Undo, Spell, Replace, Mail, Info, Compile, Debug, and News. The main text area contains the following R code:

```
*R*  
# This buffer contains the file x.r  
x<-42  
y<-x*x  
y
```

Below the code, there are two status lines:

```
ISO8--**--XEmacs: x.r (ESS[S] Font  
> # This buffer contains the R process  
>
```

At the bottom of the window, another status line reads:

```
ISO8--**--XEmacs: *R* (iESS Font [R
```