



Ngôn ngữ R – Bài 1

Ths. Vũ Duy Khương

1

Giới thiệu về ngôn ngữ R

2

Cú pháp trong R

3

Các kiểu dữ liệu : Vector – Matrix – Chuỗi

4

Bài thực hành

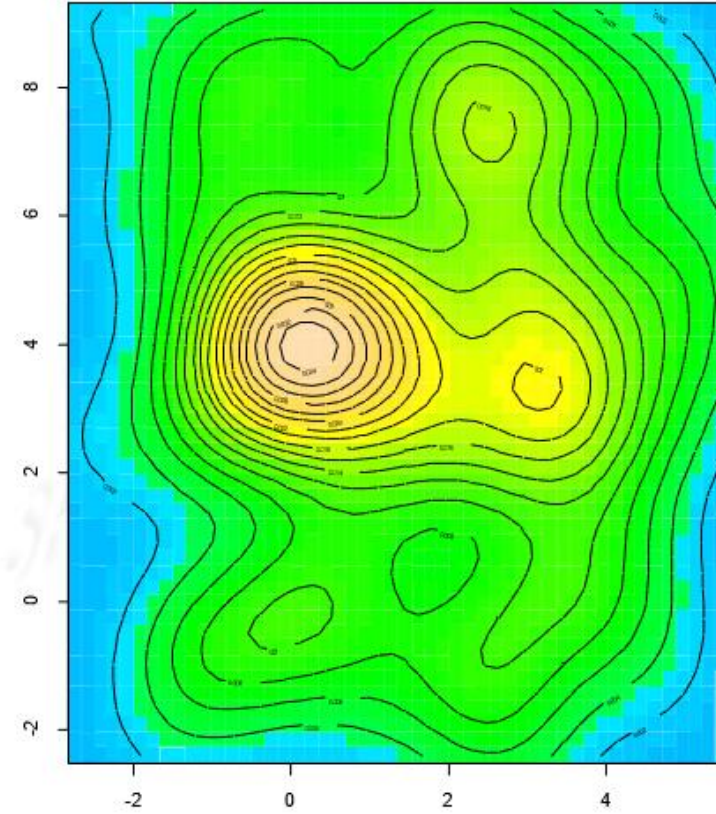
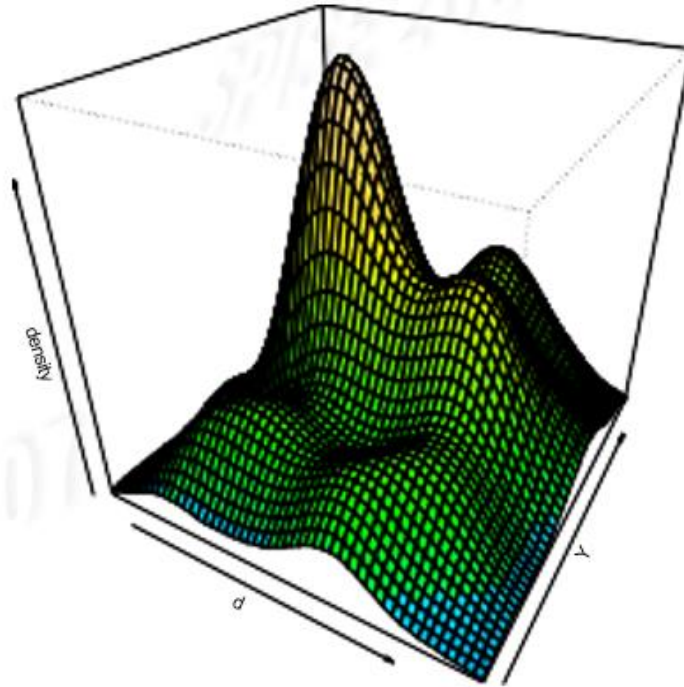
Ngôn ngữ R là gì

- R là
- Một ngôn ngữ lập trình
- Một phần mềm thống kê (a statistical package)
- Một thông dịch viên (an interpreter)
- Nguồn mở Open Source

Khởi động - Getting started

1. Tải và cài đặt R trong máy tính của bạn
Đăng nhập <http://cran.r-project.org/mirrors.html> để chọn một “mirror” (quốc gia) gần bạn nhất
2. Chọn hệ điều hành mà bạn đang sử dụng (Linux, Mac, hoặc Windows) (Click on your favorite operating system (Linux, Mac, or Windows))
3. Tải và cài đặt “base” (Download and install the “base”)
Cài đặt các packages phụ trợ
 - 3.1. Khởi động R trong máy tính của bạn
 - 3.2. Chọn đề mục thích hợp trong danh mục các “Package”

Cách cài đặt và sử dụng R



Cách cài đặt và sử dụng R

- Tính toán đơn giản, toán học giải trí (recreational mathematics), tính toán ma trận (matrix).
- Được phát triển thành các phần mềm chuyên môn cho một vấn đề tính toán cá biệt: các packages.
- R là một phần mềm sử dụng cho phân tích thống kê và vẽ biểu đồ.

Tài liệu tham khảo

1. Nguyễn Văn Tuấn, Phân tích số liệu và tạo biểu đồ bằng R, NXB KHKT, 2007, 340 trang
2. Michael O'Mahony, Sensory Evaluation of Food: Statistical Methods and Procedures, Marcel Dekker, New York, 1986, 487 p.
3. Douglas C. Montgomery, George C. Runger, Applied Statistics and Probability for engineers, John Wiley & Son, 2003, 706p
www.ykhoanet.com/r , kỹ thuật thống kê
www2.hcmut.edu.vn/~dzung/Rworkshop2006
www.r-project.org , download Phần mềm R.

Cài đặt và một số ứng dụng đơn giản

1. “Comprehensive R Archive Network” (CRAN): <http://cran.R-project.org>.
2. <http://cran.r-project.org/bin/windows/base/R-2.7.2-win32.exe>

Cài đặt và một số ứng dụng đơn giản

Setup file

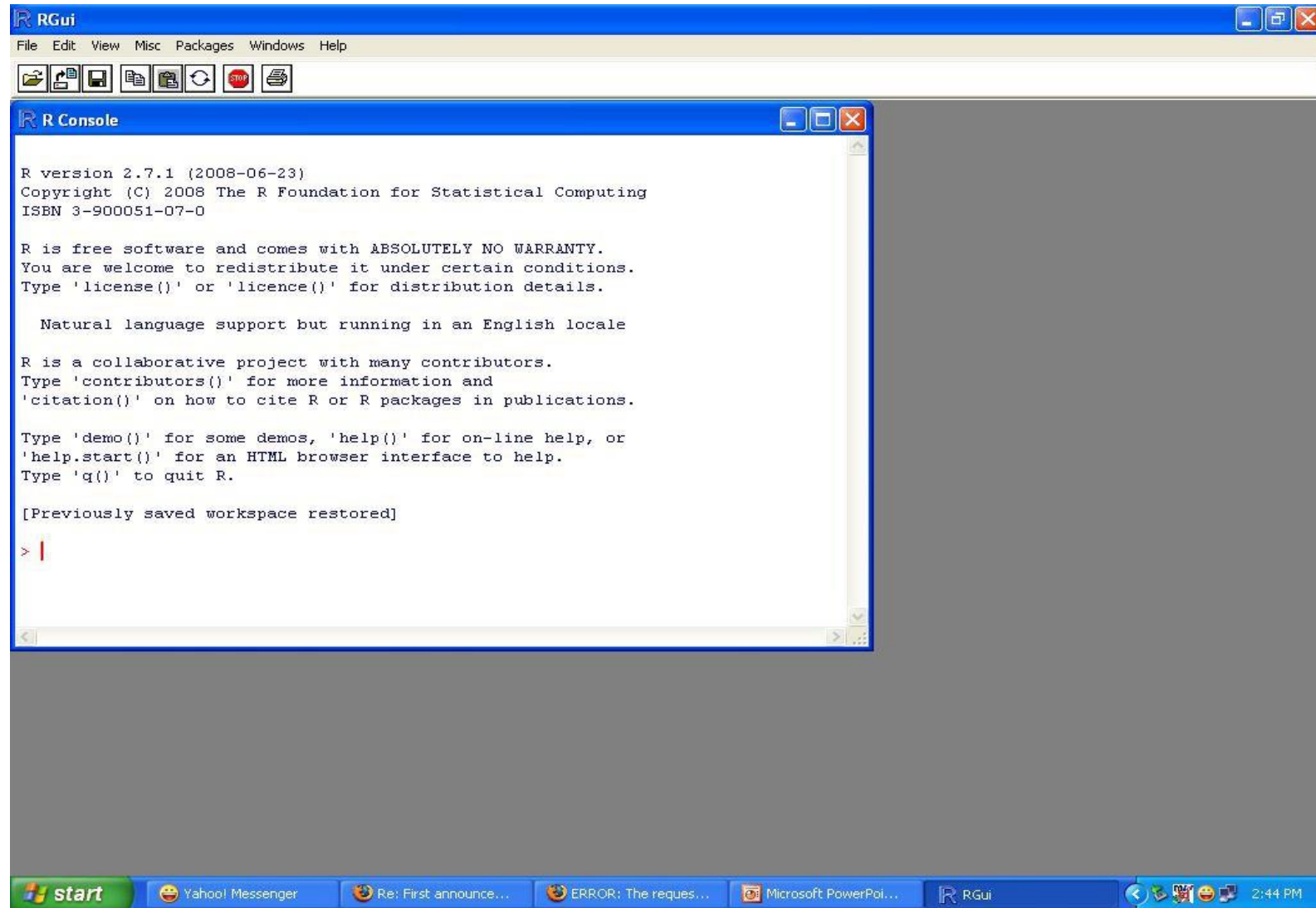


Icon on desktop



Window Screen

Cài đặt và một số ứng dụng đơn giản

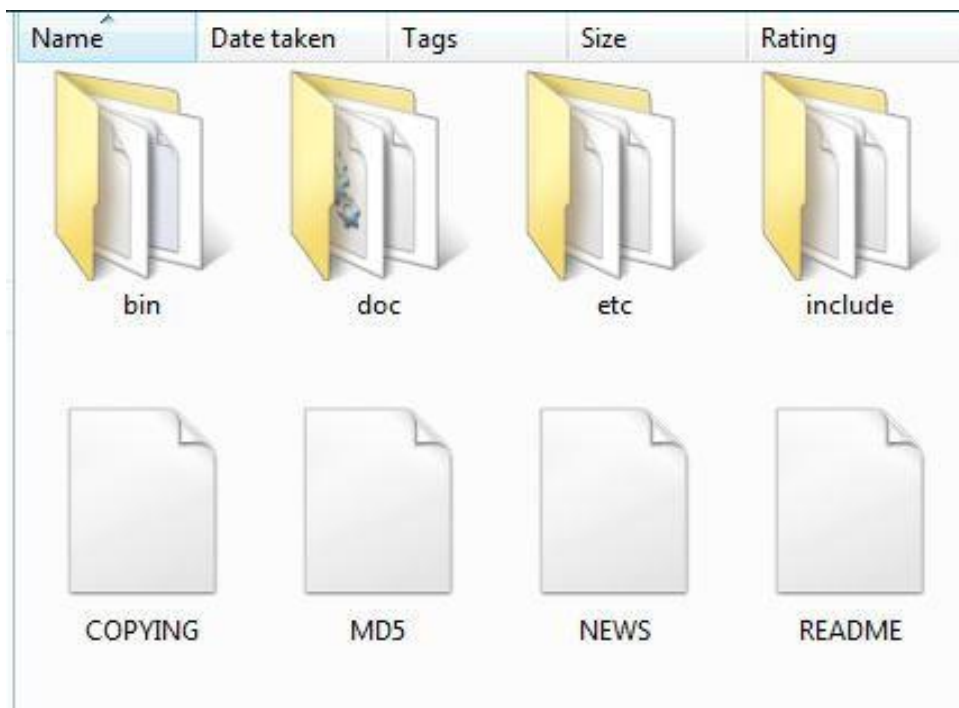


Cách cài đặt và sử dụng R

- Chương trình được cài đặt vào 1 thư mục.
- Sau đó chép toàn bộ thư mục về để sử dụng. Giả sử thư mục nằm trong ổ D của máy tính



Cách cài đặt và sử dụng R

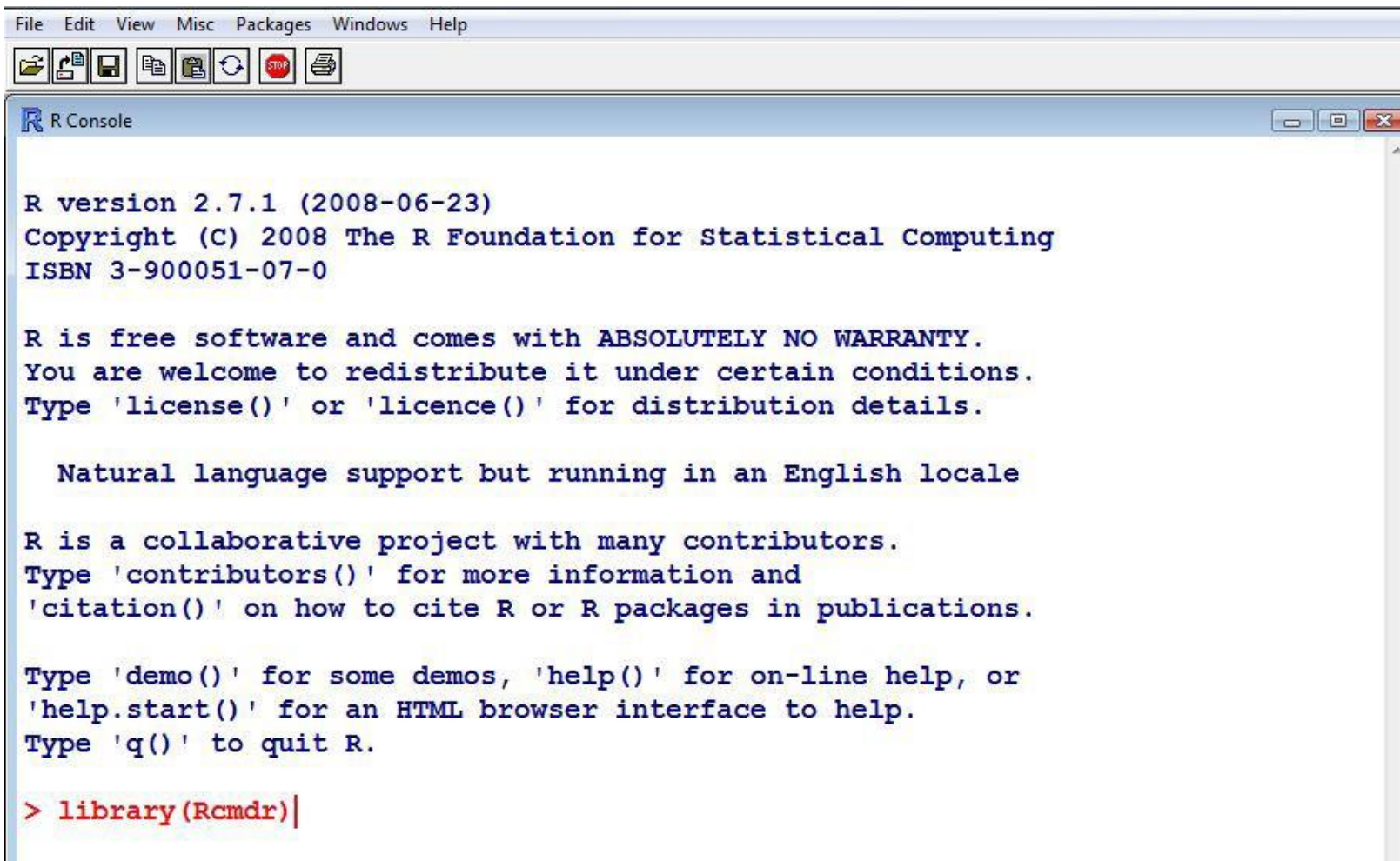


Cách cài đặt và sử dụng R

1. Click vào File name: **Rgui.exe**
2. Màn hình làm việc xuất hiện
3. Đánh câu lệnh
4. **library(Rcmdr)**



Cách cài đặt và sử dụng R



```
File Edit View Misc Packages Windows Help
R Console

R version 2.7.1 (2008-06-23)
Copyright (C) 2008 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.

> library(Rcmdr)|
```

Cách cài đặt và sử dụng R



Cài đặt và một số ứng dụng đơn giản

Prompt : >

Getting help: ?t.test or help(t.test)

Cú pháp trong R

object \leftarrow *function(arguments)*

Ví dụ: *reg* \leftarrow *lm*(*y* ~ *x*)

Cú pháp trong R

Phân biệt HOA và THƯỜNG

`a` \leftarrow 5

`A` \leftarrow 7

`B` \leftarrow `a`+`A`

Tên biến KHÔNG có khoảng trắng

`var a` \leftarrow 5

Nhưng có thể gán thêm bằng cách thêm dấu “.”

`var.a` \leftarrow 5

`var.b` \leftarrow 10

`var.c` \leftarrow `var.a` + `var.b`

Ứng dụng- ma trận

`a ← c (1,2,3,4,5,6,7,8,9)`

`A ← matrix (a,nrow=3)`

A

`a ← c (1,2,3,4,5,6,7,8,9)`

`A ← matrix (a,nrow=3, byrow=TRUE)`

A

→ Đọc thêm trong quyển “Phân tích số liệu và biểu đồ bằng R” Tác giả Nguyễn Văn Tuấn

Ứng dụng- ma trận

`a ← c (1,2,3,4,5,6,7,8,9)`

`A ← matrix (a,nrow=3)`

A

`a ← c (1,2,3,4,5,6,7,8,9)`

`A ← matrix (a,nrow=3, byrow=TRUE)`

A

➔ Đọc thêm trong quyển “Phân tích số liệu và biểu đồ bằng R” Tác giả Nguyễn Văn Tuấn

Đọc dữ liệu

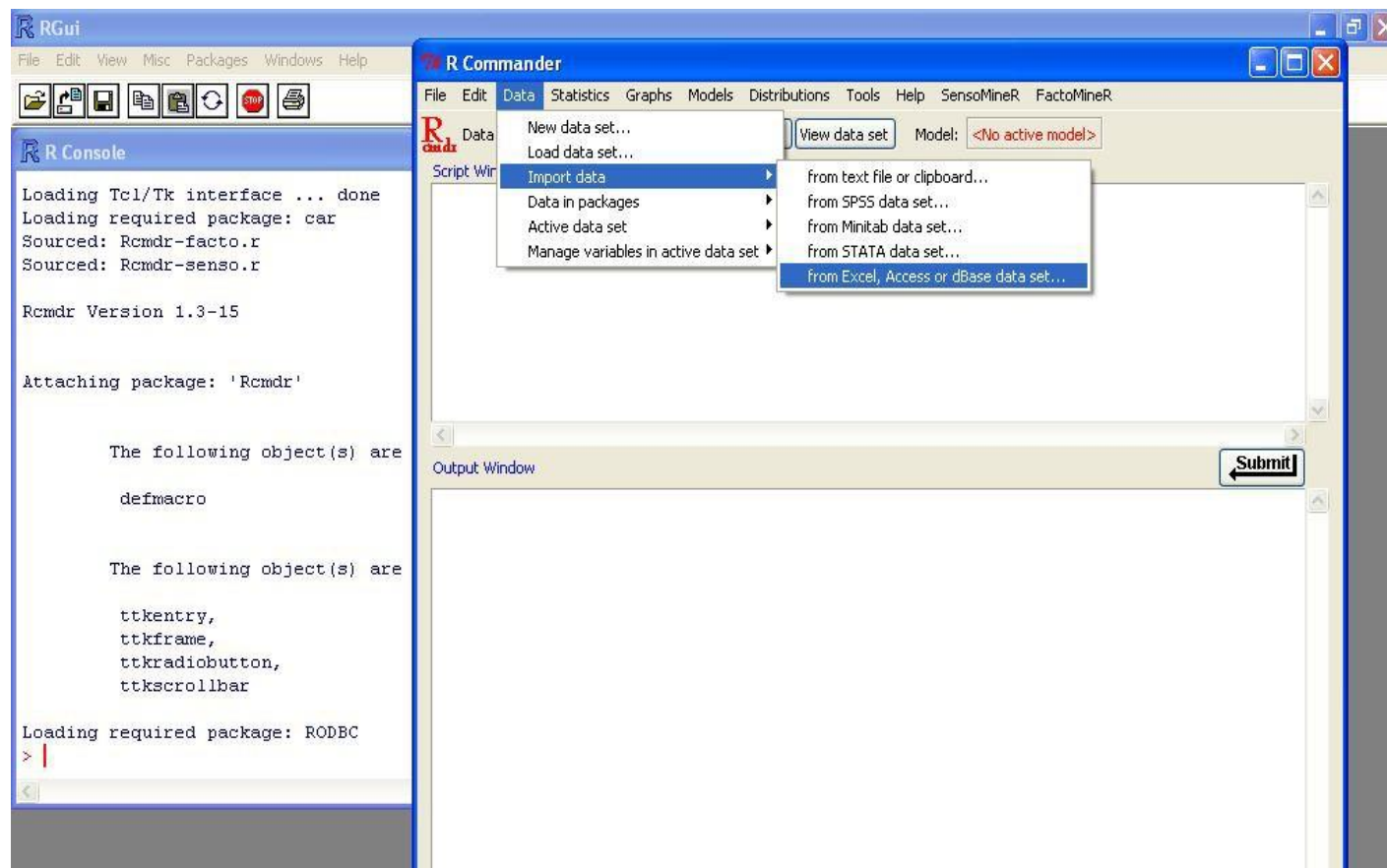
1. Cài đặt R
2. Một số phép tính đơn giản trên R.
3. Tạo chuỗi số với các lệnh: seq, rep, gl.
4. Vẽ lại phân bố chuẩn với số lượng người lần lượt là 10, 30, 150, và 1000.



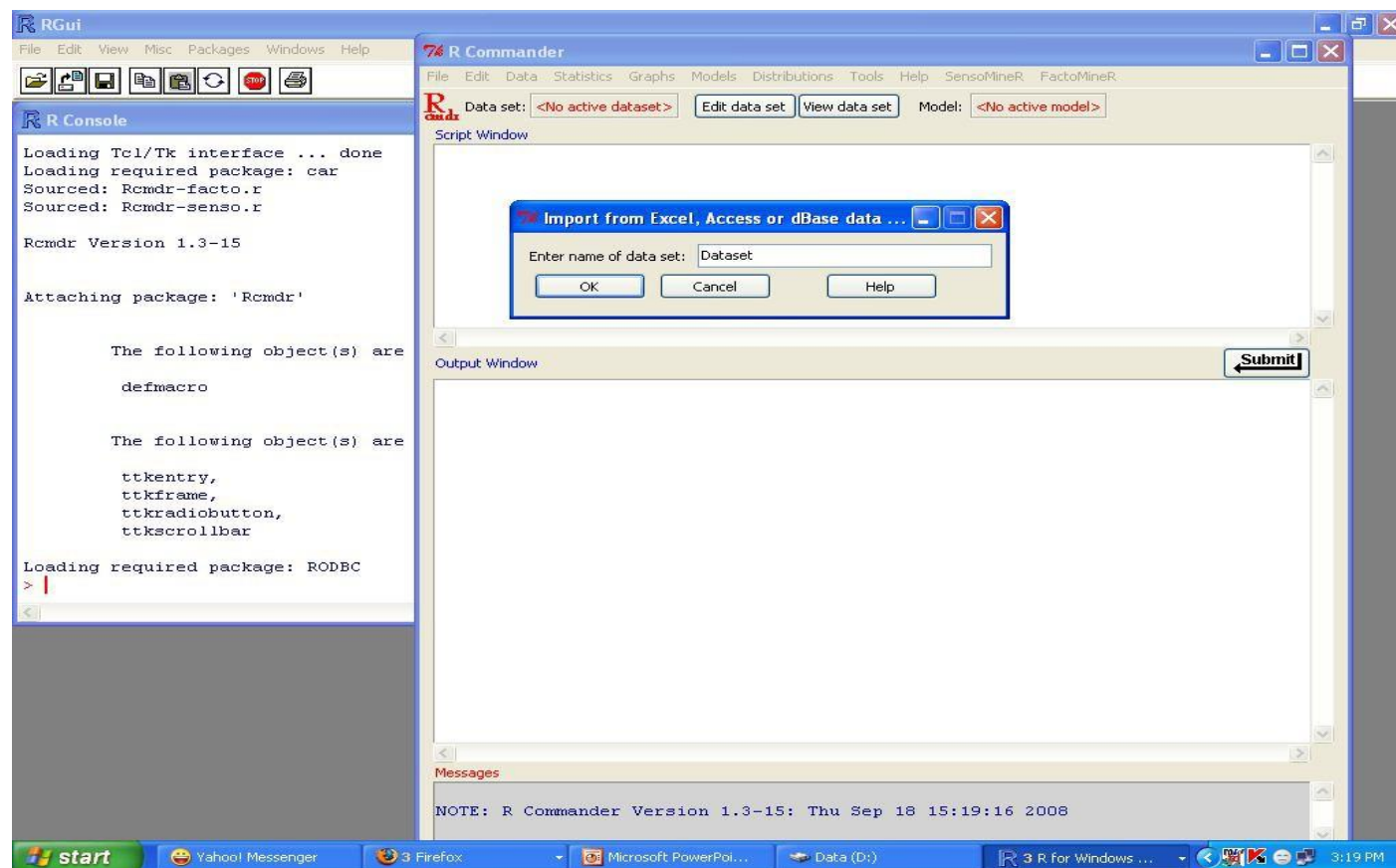
Đọc dữ liệu

```
age ← c(50,62,60,40,48,47,57,70,48,67)
bmi ← c(17,18,18,18,18,18,19,19,19,19)
thanh ← data.frame(age,bmi) attach(thanh)
thanh
```

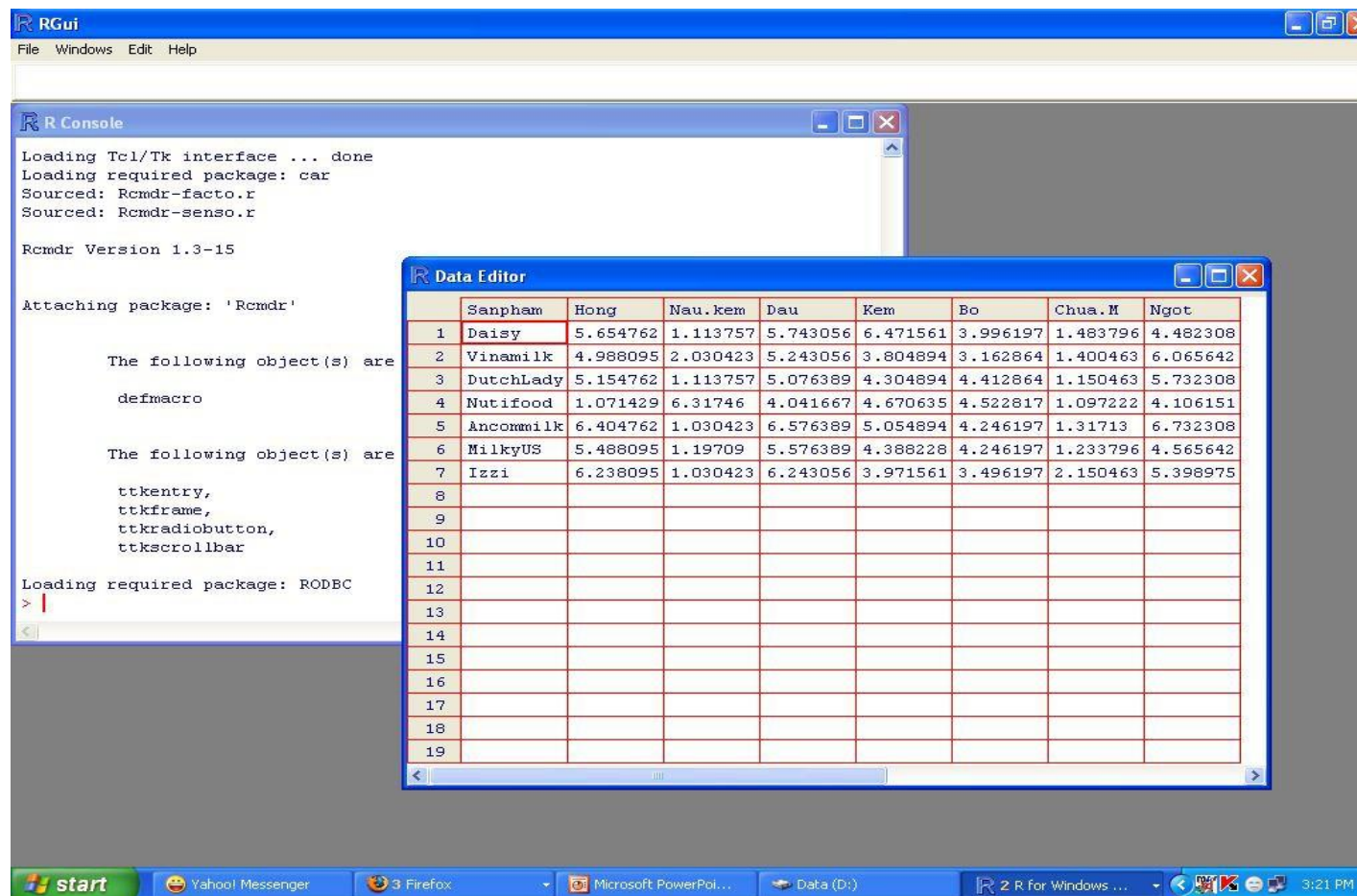

Đọc dữ liệu



Đọc dữ liệu



Đọc dữ liệu



The screenshot shows the RGui interface with the following components:

- R Console:** Displays the following text:


```

      Loading Tcl/Tk interface ... done
      Loading required package: car
      Sourced: Rcmdr-facto.r
      Sourced: Rcmdr-senso.r

      Rcmdr Version 1.3-15

      Attaching package: 'Rcmdr'

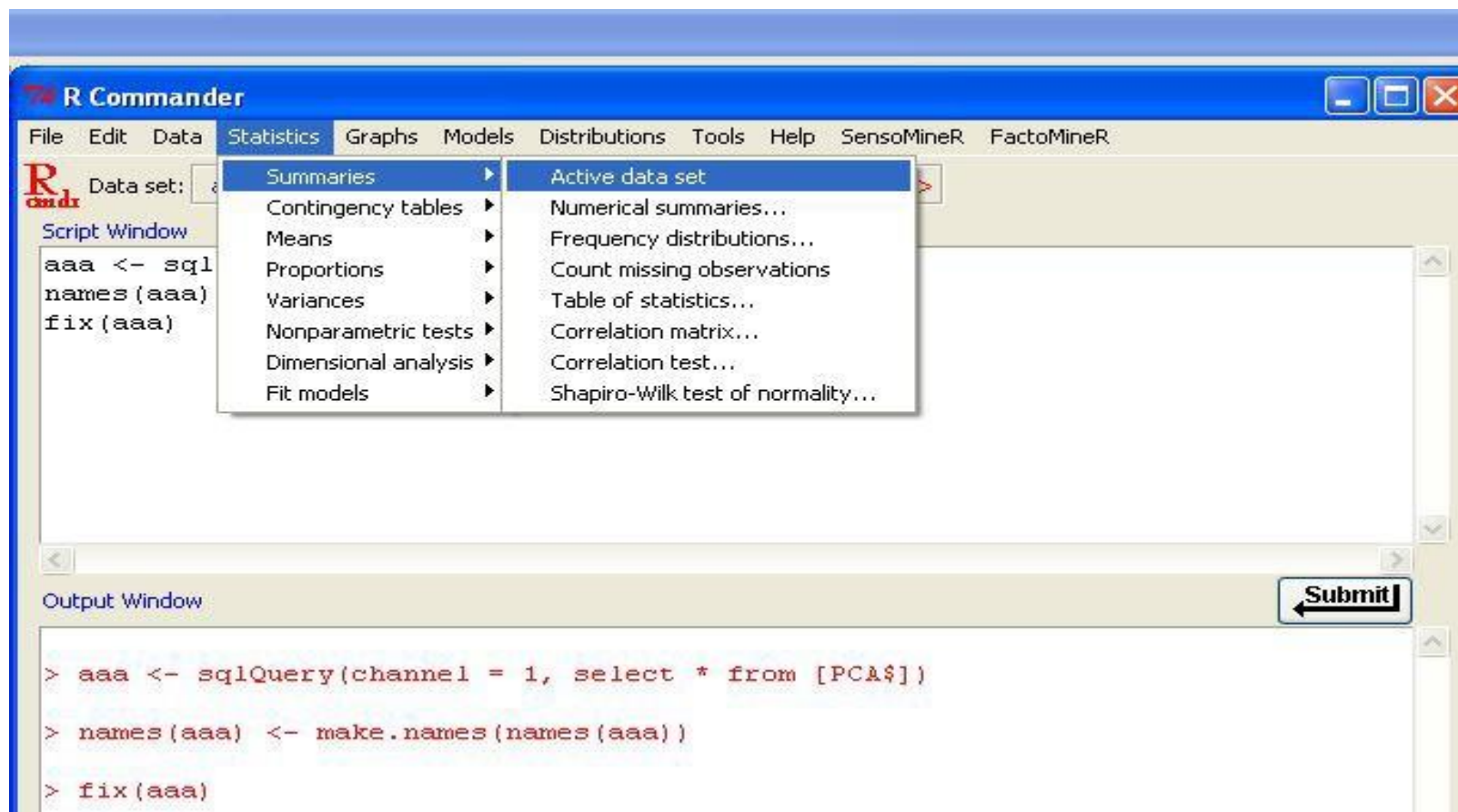
      The following object(s) are
      defined:
      defmacro

      The following object(s) are
      loaded into the workspace:
      ttkentry,
      ttkframe,
      ttkradiobutton,
      ttkscrollbar

      Loading required package: RODBC
      >
      
```
- Data Editor:** A window showing a data table with 19 rows and 9 columns. The columns are labeled: Sanpham, Hong, Nau.kem, Dau, Kem, Bo, Chua.M, and Ngot. The data is as follows:

	Sanpham	Hong	Nau.kem	Dau	Kem	Bo	Chua.M	Ngot
1	Daisy	5.654762	1.113757	5.743056	6.471561	3.996197	1.483796	4.482308
2	Vinamilk	4.988095	2.030423	5.243056	3.804894	3.162864	1.400463	6.065642
3	DutchLady	5.154762	1.113757	5.076389	4.304894	4.412864	1.150463	5.732308
4	Nutifood	1.071429	6.31746	4.041667	4.670635	4.522817	1.097222	4.106151
5	Ancomilk	6.404762	1.030423	6.576389	5.054894	4.246197	1.31713	6.732308
6	MilkyUS	5.488095	1.19709	5.576389	4.388228	4.246197	1.233796	4.565642
7	Izzi	6.238095	1.030423	6.243056	3.971561	3.496197	2.150463	5.398975
8								
9								
10								
11								
12								
13								
14								
15								
16								
17								
18								
19								

Đọc dữ liệu

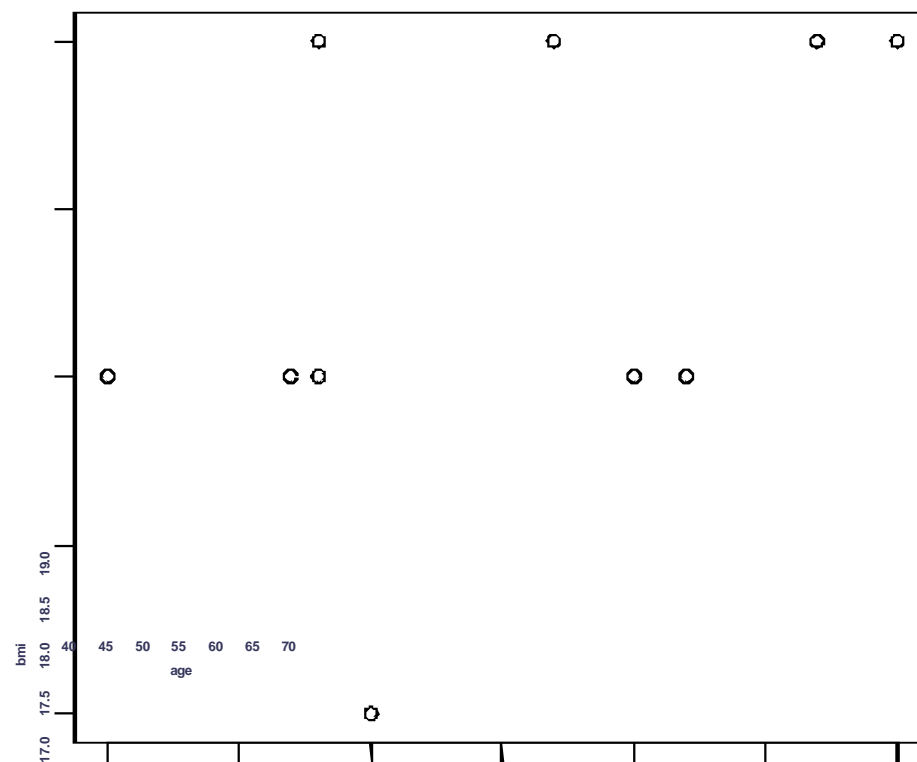


Đọc dữ liệu



1. Read file: igf, t.test, anova, rankingtest trong vidu.xls
2. Edit & attach <data file>.

Plot số liệu age-bmi



Mã hóa mẫu và trật tự mẫu

Mã hóa mẫu

```
sample(0:999,10,replace=FALSE)
```

```
[1] 667 926 888 511 475 889 404 184 713 770
```

Trật tự mẫu

```
Load packages (crossdes)
```

```
williams(4)
```

```
[,1] [,2] [,3] [,4]
```

```
[1,] 1 2 4 3
```

```
[2,] 2 3 1 4
```

```
[3,] 3 4 2 1
```

```
[4,] 4 1 3 2
```


R – Máy tính (calculator)

R có thể được dùng như một máy tính tay:

```
> 5 + (6 + 7) * pi^2
```

```
[1] 133.3049
```

```
> log(exp(1))
```

```
[1] 1
```

```
> log(1000, 10)
```

```
[1] 3
```

```
> sin(pi/3)^2 + cos(pi/3)^2
```

```
[1] 1
```

```
> Sin(pi/3)^2 + cos(pi/3)^2
```

Lỗi: không thể tìm được chức năng “Sin”
(Error: couldn't find function "Sin")

Các dạng số liệu cơ bản

- Lôgic (Logical)

```
>x <- T; y <-  
F
```

```
>x; y  
[1] TRUE  
[1] FALSE
```

- Số (Numerical)

```
>a <- 5; b <-  
sqrt(2)  
>a; b  
[1] 5  
[1] 1.414214
```

- Ký tự (Character)

```
>a <- "1"; b <- 1  
>a; b  
[1] "1"  
[1] 1  
>a <- "character"  
>b <- "a"; c <- a  
>a; b; c  
[1] "character"  
[1] "a"  
[1] "character"
```

Vector, Ma trận, Chuỗi

- Vector
 - Ordered collection of data of the same data type
 - Ví dụ:
 - + Họ của tất cả sinh viên trong lớp
 - + Các điểm số do một người thử đưa ra cho một tập hợp sản phẩm
 - Trong R, một chữ số là một vector có chiều dài bằng 1
- Ma trận
 - Bảng số liệu hình chữ nhật của các số liệu cùng loại
(Rectangular table of data of the same type)

Vector, Ma trận, Chuỗi

- Ví dụ:

+ Các điểm số do tất cả thành viên hội đồng đưa ra cho một tập hợp sản phẩm (sản phẩm = hàng, thành viên = cột)

• Chuỗi

+ Ma trận có nhiều chiều hơn (nghĩa là “đa chiều - Multiway”)

Vector

- Vector: tập các số liệu đồng dạng có thứ tự (Ordered collection of data of the same data type)

```
>x <- c(5.2, 1.7, 6.3)
```

```
>log(x)
```

```
[1] 1.6486586 0.5306283 1.8405496
```

```
>y <- 1:5
```

```
>z <- seq(1, 1.4, by = 0.1)
```

```
>y + z
```

```
[1] 2.0 3.1 4.2 5.3 6.4
```

```
>length(y)
```

```
[1] 5
```

```
>mean(y + z)
```

```
[1] 4.2
```

Ma trận

- Ma trận: Bảng số liệu dạng chữ nhật của số liệu cùng loại

```
>m <- matrix(1:12, 4, byrow = T); m      [,1] [,2] [,3]
```

```
[1,] 1      2      3
[2,] 4      5      6
[3,] 7      8      9
[4,] 10     11     12
```

```
>y <- -1:2
```

```
>m.new <- m + y
```

```
>t(m.new)
```

```
 [,1] [,2] [,3] [,4]
[1,]  0      4 8    12
[2,]  1      5 9    13
[3,]  2      6 10   14
```

```
>dim(t(m.new))
```

```
[1] 3 4
```

```
>dim(m)
```

```
[1] 4 3
```

Giá trị vắng mặt (Missing values)

- R được thiết kế để xử lý các số liệu thống kê (handle statistical data) và do đó cũng được xác định là phải đề cập đến các giá trị vắng mặt (deal with missing values)

- Các số không “sẵn có”

```
> x <- c(1, 2, 3, NA)
```

```
> x + 3
```

```
[1] 4 5 6 NA
```

- “không phải số” “Not a number”

```
> log(c(0, 1, 2))
```

```
[1] -Inf 0.00000000 0.6931472
```

```
> 0/0
```

```
[1] NaN
```


Tập con - Subsetting

- Việc trích lấy một tập con từ một vector hoặc ma trận thường là cần thiết
- R cung cấp một số các phương pháp hay để làm việc đó

```
>x <- c("a", "b", "c", "d", "e", "f", "g", "h")
>x[1]
>x[3:5]
>x[-(3:5)]
>x[c(T, F, T, F, T, F, T, F)]
>x[x <= "d"]
>m[,2]
>m[3,]
```

Những dạng đối tượng & số liệu khác

- Các hàm - Functions
- Các yếu tố - Factors
- Các danh sách - Lists
- Data frames

Chúng ta sẽ nói về những vấn đề này sau

Nhập/Xuất số liệu

- Nhập số liệu

- R có thể nhập số liệu từ các phần mềm ứng dụng khác
- Cách dễ nhất là nhập file.txt dạng “tab delimited”

```
>myData<-read.table("file",sep=",")
```

```
>myData <- read.table(file = "C:/myFile.txt", header =  
TRUE, quote = "", sep = "\t", comment.char="")
```

- Xuất số liệu

- R cũng có thể xuất số liệu dưới những định dạng khác nhau
- Tab delimited là dạng phổ biến nhất

```
>write.table(x, "filename")
```

Phân tích/Tóm tắt số liệu

- Trước hết, xem qua một chút ...
`> SimpleData[1:10,]`
- Trung bình, phương sai, độ lệch chuẩn, v.v...
`> mean(SimpleData[, 3])`
`> mean(log(SimpleData[, 3]))`
`> var(SimpleData[, 4])`
`> sd(SimpleData[, 3])`
`> cor(SimpleData[, 3:4])`
`> colMeans(SimpleData[3:14])`

Biểu diễn bằng biểu đồ

- Scatter plot

```
>plot(log(SimpleData[, "C1"]), log(SimpleData[, "W1"]), xlab =  
"channel 1", ylab = "channel 2")
```

- Histogram

```
>hist(log(SimpleData[, 7]))  
>hist(log(SimpleData[, 7]), nclass = 50, main = "Histogram of  
W3 (on log scale)")
```

- Boxplot

```
>boxplot(log(SimpleData[, 3:14]))  
>boxplot(log(SimpleData[, 3:14]), outline = F, boxwex = 0.5,  
col = 3, main = "Boxplot of SimpleData")
```

Trợ giúp ... và thoát

- Kiểm thông tin về một lệnh cụ thể
 - >help(rnorm)
 - >?rnorm
- Tìm các hàm liên quan đến một từ khoá
 - >help.search("boxplot")
- Khởi động những trang trợ giúp cài đặt R
 - >help.start()
- Thoát khỏi R
 - >q()

The Rcmdr package

- Tải “Rcmdr package”
>library(Rcmdr)





THANK YOU