The R Project for Statistical Computing

R Language Data Objects
物件、資料輸入輸出

vvn Weian Chao (趙韋安)

https://ce.nctu.edu.tw/member/teachers/23

Department of Civil Engineering, National Yang Ming Chiao Tung University, Taiwan

R: An Introduction
Accessing data.
Data formats

Data types

Data structures

Data types

Numeric-(integer, single, double)
Character
Logical
Complex
…

Data structures

Vector
Matrix/Array
Data frame
List
Others... (Date/POSIXct/POSIXlt/ts)

Vector
-1+numbers in a 1D array
-All same data type
-R basic data object

Matrix
-Two dimensions
-Same length
-Same data class
-Columns not named (index number)

Array
-Identical to a matrix
But 3+dimensions

Data frame
-Can have vectors of multiple data types
-All same length
-Closest R analogue to spreadsheet
-Special functions

類似 電子試算表

List
-Most flexible
-Ordered collection of elements
-Any class, length, or structure
-Can include lists

Naming Rules

- A and a are different
- All alphanumeric symbols are allowed

(A-Z, a-z, 0-9) . _

- Name must start with . or a letter

Be meaningful

Naming Rules

Wrong

```
3x
3_x
3-x
3.X
.3variable
```

Correct

```
x_3
x3
x.3
taipei.x3
.variable
```

Function name

GOOD: CalculateAvg
BAD: calculate_avg
BAD: calculateAvg

Variable name

GOOD: avg.clicks
OK: avgClicks
BAD: avg_Clicks

Naming Rules

Use "<-" to assign values to a variable

Naming Rules

Use “#” to be Prompt symbol

R: An Introduction
Accessing data.

Objects

objects(), ls(), rm(), object.size()
print(object.size(), units = “b”)

TRY it in R

R: An Introduction
Accessing data.

R_data_objects_a.R

R: An Introduction
Accessing data.
Coercing

Table 2.4. Functions for testing (is) the attributes of different categories of object (arrays, lists, etc.) and for coercing (as) the attributes of an object into a specified form. Neither operation changes the attributes of the object.

Type	Testing	Coercing
Array	is.array	as.array
Character	is.character	as.character
Complex	is.complex	as.complex
Dataframe	is.data.frame	as.data.frame
Double	is.double	as.double
Factor	is.factor	as.factor
List	is.list	as.list
Logical	is.logical	as.logical
Matrix	is.matrix	as.matrix
Numeric	is.numeric	as.numeric
Raw	is.raw	as.raw
Time series (ts)	is.ts	as.ts
Vector	is.vector	as.vector

Coercing Types

as.integer
as.numeric
as.data.frame

TRY it in R

R: An Introduction
Accessing data.

R_data_objects_b.R

R: An Introduction
Accessing data.

Factors

An “attribute” of a vector
that specifies the possible values
and
their order

TRY it in R

R: An Introduction
Accessing data.

R_data_objects_c.R

R: An Introduction
Accessing data.

Special Functions for Vector

which

Return location index when the logical is True

Vector Arithmetic

+ - * / ^
%% 餘數
%/% 整數商
log(x) logb(x,b) 以b為底數
pi exp(x)
sin cos tan
abs(x) sqrt(x)
length(x)
prod(x) 所有數列乘積
factorial(x) 階乘
sqrt()

sort

sort()

rank()-
顯示排序後元素的名次、最小值為第一名

order()-
排序由小至大的數值，顯示落在原向量第幾元素的位置

Character vector

Single/Double quotes

paste()

Logical vector

TRUE FALSE

T, F

< <= > >=

==

!=

&: and

| : or (pipe)

length

Number of elements

TRY it in R

R: An Introduction
Accessing data.

R_data_objects_d.R

R: An Introduction
Accessing data.

Special Functions for Matrix

Matrix

dim
diag
A*B
A %*% t(B)

Matrix

Some matrix functions	
t	Transpose
diag	Diagonal
%*%	Inner (dot) product of two vectors $x^t y$, matrix multiplication
%o%	Outer product of two vectors xy^t
crossprod, tcrossprod	Cross products $x^t y$ and xy^t of matrices
det	Determinant
solve	Inverse
eigen	Eigenvalues and eigenvectors
svd	Singular value decomposition
qr	QR decomposition
chol	Choleski decomposition

TRY it in R

R: An Introduction
Accessing data.

R_data_objects_e.R

R: An Introduction
Accessing data.
Rounding of Numbers

Rounding of Numbers

ceiling-不可小於x中元素的最小整數

floor-不可大於x中元素的最大整數

trunc-去掉小數位

round-可控制保留小數位數

getOption(‘digits’)

options(digits = 3)

TRY it in R

R: An Introduction
Accessing data.

R_data_objects_f.R

課堂練習1: 學號-姓名-objects.R

有一班級80位學生之數學成績如下
(假設成績已按座號排序)

```
1 set.seed(1)
2 math.score <- sample(0:100, 80, replace = TRUE)
3 math.score
```

(a)計算座號1-30號同學之成積平均數、標準差
(b)共有多少人及格? 及格同學的座號為何?
(c)全班最高分與最低分為何? 對應的同學座號為何?
(d)計算班上分數排行前十名(由高分至低分)之成積平均數標準差
(e)屏幕顯示80位同學分數資料的第一個四分位數(Hint: summary())