# 數據類型

## 建立數組

> X<-5:9

> X

[1] 5 6 7 8 9

## 建立向量:C函式

> x <- c(0.5,0.6)

> x

[1] 0.5 0.6

> x <- c(T,F)

> x

[1] TRUE FALSE

## 建立向量:VECTOR 函式

> x <- vector("numeric",length = 10)

> x

[1] 0 0 0 0 0 0 0 0 0 0

## 查詢向量類型

> x <-5:10

> class(x)

[1] "integer"

## 顯式轉型

### as.numeric(x)

> as.numeric(x)

[1] 5 6 7 8 9 10

### as.logical(x)

> as.logical(x)

[1] TRUE TRUE TRUE TRUE TRUE TRUE

### as.character(x)

> as.character(x)

[1] "5" "6" "7" "8" "9" "10"

## 建立矩陣向量

> m <- matrix(nrow =2, ncol=3)

> m

[,1] [,2] [,3]

[1,] NA NA NA

[2,] NA NA NA

## 建立賦值的矩陣向量(注意：它的特性是按列構造的)

> m <- matrix(1:6,nrow =2, ncol=3)

> m

[,1] [,2] [,3]

[1,] 1 3 5

[2,] 2 4 6

## 建立賦值的矩陣向量2：對向量賦與維度屬性

> m <-1:10

> dim(m)<-c(2,5)

> m

[,1] [,2] [,3] [,4] [,5]

[1,] 1 3 5 7 9

[2,] 2 4 6 8 10

## 建立賦值的矩陣向量 3: 使用綁定方法

### cbind()

> x <- 1:3

> y <- 10:12

> cbind(x,y)

x y

[1,] 1 10

[2,] 2 11

[3,] 3 12

### rbind()

> rbind(x,y)

[,1] [,2] [,3]

x 1 2 3

y 10 11 12

## 取得矩陣的維度屬性

> dim(m)

[1] 2 3

## 查詢向量中的屬性

> attributes(m)

$dim

[1] 2 3

## 建立列表LIST

> x<-list(1,"a",T,1+1)

> x

[[1]]

[1] 1

[[2]]

[1] "a"

[[3]]

[1] TRUE

[[4]]

[1] 2

## 列表和向量最大的不同在於列表可以存放不同類型的數據類型，而向量則必須存放相同的數據類型

## 建立因素factor

> x<-factor(c("yes","yes","no","yes","no"))

> x

[1] yes yes no yes no

Levels: no yes

> table(x)

x

no yes

2 3

> unclass(x) <--去掉分類信息

[1] 2 2 1 2 1

attr(,"levels")

[1] "no" "yes"

## 改變因素的水平順序

> x<-factor(c("yes","yes","no","yes","no"),levels=c("yes","no"))

> x

[1] yes yes no yes no

Levels: yes no

## 檢測缺失值missing values

### is.na

> x<-c(1,2,NA,10,3)

> is.na(x)

[1] FALSE FALSE TRUE FALSE FALSE

> x<-c(1,2,NaN,NA,4)

> is.na(x)

[1] FALSE FALSE TRUE TRUE FALSE

### is.nan

> x<-c(1,2,NA,10,3)

> is.nan(x)

[1] FALSE FALSE FALSE FALSE FALSE

> x<-c(1,2,NaN,NA,4)

> is.nan(x)

[1] FALSE FALSE TRUE FALSE FALSE

## 數據框data frames

> x<-data.frame(foo=1:4,bar=c(T,T,F,F))

> x

foo bar

1 1 TRUE

2 2 TRUE

3 3 FALSE

4 4 FALSE

> nrow(x)

[1] 4

> ncol(x)

[1] 2

## 命名names 向量

|  |
| --- |
| > x<-1:3  > names(x)  NULL  > names(x)<-c("foo","bar","norf")  > x  foo bar norf  1 2 3  > names(x)  [1] "foo" "bar" "norf" 命名 NAMES 列表 |
| x<-list(a=1,b=2,c=3)  x  $a  [1] 1  $b  [1] 2  $c  [1] 3 維度名字dimname > m<-matrix(1:4,nrow=2,ncol=2)  > dimnames(m)<-list(c("a","b"),c("c","d"))  > m  c d  a 1 3  b 2 4 |
| |  | | --- | |  | |

# 提取子集 subsetting

## 從向量取子集

> x <- c("a","b","c","c","d","a")

> x[1]

[1] "a"

> x[2]

[1] "b"

> x[1:4]

[1] "a" "b" "c" "c"

> x[x>"a"]

[1] "b" "c" "c" "d"

> u<-x>"a"

> u

[1] FALSE TRUE TRUE TRUE TRUE FALSE

> x[u]

[1] "b" "c" "c" "d"

## 從矩陣取子集subsetting a matrix

> x<-matrix(1:6,2,3)

> x[1,2]

[1] 3

> x[2,1]

[1] 2

> x[1,]

[1] 1 3 5

> x[,2]

[1] 3 4

> x[1, ,drop=F]

[,1] [,2] [,3]

[1,] 1 3 5

## 從列表隻子集 subsetting lists

> x<-list(foo=1:4,bar=0.6)

> x[1]

$foo

[1] 1 2 3 4

> x[[1]]

[1] 1 2 3 4

> x$bar

[1] 0.6

> x[["bar"]]

[1] 0.6

> x["bar"]

$bar

[1] 0.6

> x[c(1,2)]

$foo

[1] 1 2 3 4

$bar

[1] 0.6

> x[c(1)]

$foo

[1] 1 2 3 4

### 使用變量取得對應的索引列表

> name <-"foo"

> x[[name]]

[1] 1 2 3 4

> x$name

NULL

### 嵌套提取

> x<-list(a=list(10,12,14),b=c(3.14,2.81))

> x[[c(1,3)]]

[1] 14

> x[[1]][[3]]

[1] 14

> x[[c(2,1)]]

[1] 3.14

### 部份匹配

> x<-list(aardvark=1:5)

> x$a

[1] 1 2 3 4 5

### 移除缺失值 removing na values

> x<-c(1,2,NA,4,NA,5)

> bad<-is.na(x)

> x[!bad]

[1] 1 2 4 5

### 移除缺失值2 取得同樣有值的位置

> y<-c("a","b",NA,"d",NA,"f")

> good<-complete.cases(x,y)

> good

[1] TRUE TRUE FALSE TRUE FALSE TRUE

> x[good]

[1] 1 2 4 5

> y[good]

[1] "a" "b" "d" "f"

## 向量運算 vectorized operations

> x<-1:4; y<-6:9

> x+y

[1] 7 9 11 13

> x>2

[1] FALSE FALSE TRUE TRUE

> x>=2

[1] FALSE TRUE TRUE TRUE

> y==8

[1] FALSE FALSE TRUE FALSE

> x\*y

[1] 6 14 24 36

> x/y

[1] 0.1666667 0.2857143 0.3750000 0.4444444

## 矩陣運算 vectorized Matrix Operations

> x<-matrix(1:4,2,2);y<-matrix(rep(10,4),2,2)

> x\*y

[,1] [,2]

[1,] 10 30

[2,] 20 40

> x/y

[,1] [,2]

[1,] 0.1 0.3

[2,] 0.2 0.4

> x %\*%y //真正的矩陣相乘

[,1] [,2]

[1,] 40 40

[2,] 60 60

## 讀取檔案函數

### read.table,read.csv ,for reading tabular data (以下是其參數說明)

#### 文件名(file,the name of a file,or a connection)

#### 標頭(header,logical indicating if the file has a header line)-如果HEADER是宣告變量，就不是我們需要的數據

#### 分隔符(sep,astring indicating how the columns are separated)

#### 字符向量(colclasses,a character vector indicating the class of each column in the dataset)

#### 數據集的行數(nrows, the number of rows in the dataset)

#### 注釋標記(comment.char,a character string indicating the comment character)-一般是#號

#### 忽略(skip,the number of lines to skip from the beginning)-跳過幾行

#### 是否把字符變量變成因子(stringsasfactors,should character variabies be coded as factors)

Data <- read.table(“foo.txt”)

### readlines,for reading lines of a text file

### source,for reading in r code files(inverse of dump)

### dget,for reading in r code files(inverse of dput)

### load,for reading in saved workspaces

### unserialize,for reading single r objects in binary form

## 寫入檔案函數

### write.table

### writelines

### dump

### dput

### save

### serialize