# 第 06 天

## 把變數集結起來（2）

第六天我們會介紹**二維**這個大類的資料結構。不同的資料結構是 R 語言為了讓使用者更便利地收藏變數，我們應該要用正向的心態看待它們，而非因為類型繁複而感到煩悶。這些資料結構的使用就像工具書（如字典、百科全書與技術手冊）的使用，先有一個概念，然後碰到需要使用的時候知道該怎麼查找就沒問題了。

* 一維
  + 向量（vector）
  + 因素向量（factor）
* **二維**
  + **矩陣（matrix）**
  + **資料框（data.frame）**
* 多維
  + 陣列（array）
  + 清單（list）

### 矩陣

矩陣是能夠儲存列（Row）與欄（Column）的資料結構，如果讀者跟我一樣對於分辨行、列與欄這些中文字有障礙的話，我非常推薦用英文來記憶：**Row** 指的就是水平方向資料，**Column** 指的就是垂直方向資料。

我們通用的習慣是**先 Row 後 Column**，這句話是什麼意思？假如我們現在和您說要建立一個 2x3 的矩陣，意思就是兩個水平方向搭配三個垂直方向的矩陣，外觀長得像這樣：

> my\_mat <- matrix(1:6, nrow = 2)  
> my\_mat  
 [,1] [,2] [,3]  
[1,] 1 3 5  
[2,] 2 4 6  
> class(my\_mat)  
[1] "matrix"

我們使用 matrix() 函數並指定參數 nrow = 2 將一維的數字向量（1 到 6）轉換成一個 2x3 的矩陣；往後如果聽到一個矩陣是 mxn 或者 (m, n) 的外觀，我們的心中就會知道，這個矩陣具有 m 個水平方向資料，n 個垂直方向資料。matrix() 函數還有一個 byrow 的參數可以指定要用什麼順序擺放原先在向量中的元素：

> my\_mat <- matrix(1:6, nrow = 2, byrow = TRUE)  
> my\_mat  
 [,1] [,2] [,3]  
[1,] 1 2 3  
[2,] 4 5 6

眼尖的您很快就發現雖然這個矩陣的外觀仍然是 2x3，但六個數字擺放的方向變為水平的。byrow 參數的預設值為 FALSE，意思是如果我們沒有特別指定，就是以垂直的方向來擺放矩陣。

矩陣同樣能夠以 [] 搭配索引值選出裡面的變數，只不過跟向量不同的是，現在有兩個維度的索引值必須指定，所以要用 [m, n] 兩個索引值來搭配選擇：

> my\_mat <- matrix(1:6, nrow = 2)  
> my\_mat  
 [,1] [,2] [,3]  
[1,] 1 3 5  
[2,] 2 4 6  
> my\_mat[2, 3] # 選出位於（2, 3）這個位置的 6  
[1] 6  
> my\_mat[2, ] # 選出所有第二列（2nd row）的元素  
[1] 2 4 6  
> my\_mat[, 3] # 選出所有第三欄（2nd column）的元素  
[1] 5 6

我們也可以透過**判斷運算子**來對矩陣進行篩選，選出介於 1 與 6 之間的數字：

> my\_mat <- matrix(1:6, nrow = 2)  
> filter <- my\_mat < 6 & my\_mat > 1  
> my\_mat[filter]  
[1] 2 3 4 5

矩陣與向量有一點很相似的特性，那就是包含**一種**變數類型，假如我們將數值與邏輯值一同放入矩陣，則邏輯值亦會被自動轉換成數值：

> my\_mat <- matrix(c(1, 2, TRUE, FALSE, 3, 4), nrow = 2)  
> my\_mat  
 [,1] [,2] [,3]  
[1,] 1 1 3  
[2,] 2 0 4  
> class(my\_mat[, 2]) # 原本第二欄（2nd column）的輸入是兩個邏輯值  
[1] "numeric"

### 資料框

資料框絕對是我們最需要關注的一種資料結構！它能夠容許不同的欄位有不同的變數類型，讀取外部資料之後，也通常預設以資料框的格式儲存。我們可以運用 data.frame() 函數手動創造資料框，讓我們建立一個很簡單的資料框叫做 great\_nba\_teams，這個資料框有隊名、勝場數、敗場數、是否獲得總冠軍與球季：

team\_name <- c("Chicago Bulls", "Golden State Warriors")  
wins <- c(72, 73)  
losses <- c(10, 9)  
is\_champion <- c(TRUE, FALSE)  
season <- c("1995-96", "2015-16")  
  
great\_nba\_teams <- data.frame(team\_name, wins, losses, is\_champion, season)  
View(great\_nba\_teams)

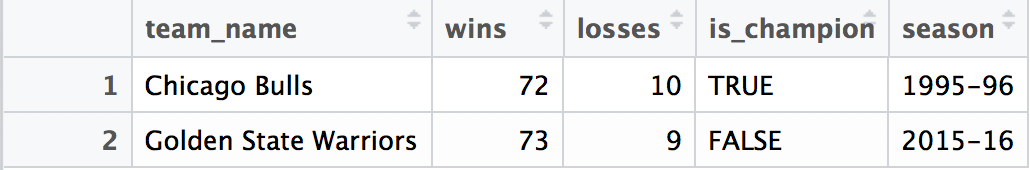


圖 6-1 使用 View() 函數瀏覽資料框

除了將資料框直接輸出在命令列（Console），我們可以使用 View() 函數瀏覽資料框的外觀與內容。我們習慣使用觀測值（Observations，Obs）來稱呼資料框中水平方向的資料，使用變數（Variables）來稱呼資料框中垂直方向的資料，我們用一個簡單的表格和矩陣比對一下：

|  |  |  |
| --- | --- | --- |
| 資料方向 | 矩陣 | 資料框 |
| 水平 | Rows | Observations |
| 垂直 | Columns | Variables |

資料框同樣能夠以 [m, n] 兩個索引值來搭配選擇出變數：

> team\_name <- c("Chicago Bulls", "Golden State Warriors")  
> wins <- c(72, 73)  
> losses <- c(10, 9)  
> is\_champion <- c(TRUE, FALSE)  
> season <- c("1995-96", "2015-16")  
>   
> great\_nba\_teams <- data.frame(team\_name, wins, losses, is\_champion, season)  
> great\_nba\_teams[1, 1] # 選出第一個變數的第一個觀測值 "Chicago Bulls"  
[1] Chicago Bulls  
Levels: Chicago Bulls Golden State Warriors  
> great\_nba\_teams[1, ] # 選出第一個觀測值  
 team\_name wins losses is\_champion season  
1 Chicago Bulls 72 10 TRUE 1995-96  
> great\_nba\_teams[, 1] # 選出第一個變數  
[1] Chicago Bulls Golden State Warriors  
Levels: Chicago Bulls Golden State Warriors

眼尖的您不知道有沒有發現，資料框預設會將文字的內容以因素向量儲存，如果您希望修正為文字可以有兩種作法，一種是建立的時候在 data.frame() 函數指定 stringsAsFactors = FALSE：

> team\_name <- c("Chicago Bulls", "Golden State Warriors")  
> wins <- c(72, 73)  
> losses <- c(10, 9)  
> is\_champion <- c(TRUE, FALSE)  
> season <- c("1995-96", "2015-16")  
>   
> great\_nba\_teams <- data.frame(team\_name, wins, losses, is\_champion, season, stringsAsFactors = FALSE)  
> great\_nba\_teams[, 1] # 選出第一個變數  
[1] "Chicago Bulls" "Golden State Warriors"  
> great\_nba\_teams[, 5] # 選出第五個變數  
[1] "1995-96" "2015-16"

另一種是事後使用 as.character() 函數進行變數類型的轉換：

> team\_name <- c("Chicago Bulls", "Golden State Warriors")  
> wins <- c(72, 73)  
> losses <- c(10, 9)  
> is\_champion <- c(TRUE, FALSE)  
> season <- c("1995-96", "2015-16")  
>   
> great\_nba\_teams <- data.frame(team\_name, wins, losses, is\_champion, season)  
> great\_nba\_teams[, 1] <- as.character(great\_nba\_teams[, 1])  
> great\_nba\_teams[, 5] <- as.character(great\_nba\_teams[, 5])  
> great\_nba\_teams[, 1] # 選出第一個變數  
[1] "Chicago Bulls" "Golden State Warriors"  
> great\_nba\_teams[, 5] # 選出第五個變數  
[1] "1995-96" "2015-16"

特別值得一提的是，資料框支援使用**變數名稱**來選擇，我們可以用 $變數名稱 或者 [, "變數名稱"] 這兩種寫法：

> team\_name <- c("Chicago Bulls", "Golden State Warriors")  
> wins <- c(72, 73)  
> losses <- c(10, 9)  
> is\_champion <- c(TRUE, FALSE)  
> season <- c("1995-96", "2015-16")  
>   
> great\_nba\_teams <- data.frame(team\_name, wins, losses, is\_champion, season, stringsAsFactors = FALSE)  
> great\_nba\_teams$team\_name  
[1] "Chicago Bulls" "Golden State Warriors"  
> great\_nba\_teams[, "season"]  
[1] "1995-96" "2015-16"

我們也可以透過**判斷運算子**來對資料框進行篩選，選出最終有獲得總冠軍的隊伍：

> team\_name <- c("Chicago Bulls", "Golden State Warriors")  
> wins <- c(72, 73)  
> losses <- c(10, 9)  
> is\_champion <- c(TRUE, FALSE)  
> season <- c("1995-96", "2015-16")  
>   
> great\_nba\_teams <- data.frame(team\_name, wins, losses, is\_champion, season, stringsAsFactors = FALSE)  
> filter <- great\_nba\_teams$is\_champion == TRUE  
> great\_nba\_teams[filter, ] # 注意這個逗號  
 team\_name wins losses is\_champion season  
1 Chicago Bulls 72 10 TRUE 1995-96

請讀者在這裡一定要注意我們的篩選是針對觀測值（水平方向），所以要將判斷運算子判斷後的結果（filter）放在逗號前面，逗號後面留空（我們沒有要選擇變數。）

資料框不像矩陣僅能包含**一種**變數類型，您可以發現在我們的 great\_nba\_teams 資料框中，有文字（team\_name、season）、整數（wins、losses）與邏輯值（is\_champion），我們可以用 str() 函數觀察：

> team\_name <- c("Chicago Bulls", "Golden State Warriors")  
> wins <- c(72, 73)  
> losses <- c(10, 9)  
> is\_champion <- c(TRUE, FALSE)  
> season <- c("1995-96", "2015-16")  
>   
> great\_nba\_teams <- data.frame(team\_name, wins, losses, is\_champion, season, stringsAsFactors = FALSE)  
> str(great\_nba\_teams)  
'data.frame': 2 obs. of 5 variables:  
 $ team\_name : chr "Chicago Bulls" "Golden State Warriors"  
 $ wins : num 72 73  
 $ losses : num 10 9  
 $ is\_champion: logi TRUE FALSE  
 $ season : chr "1995-96" "2015-16"

### 小結

好啦！第六天的內容就是這麼多，我們認識了 R 語言的兩種資料結構：矩陣與資料框。

### 練習題

###### 1. 我們有一個矩陣叫做 my\_mat，它是一個 3x3 的矩陣，裡面有 1 到 9 這些數字，請您利用 [m, n] 把 8 選出來

my\_mat <- matrix(1:9, nrow = 3)  
my\_mat[\_\_\_, \_\_\_]

###### 2. 同樣的一個矩陣，請您利用判斷運算子來對矩陣進行篩選，選出奇數（1、3、5、7、9）

my\_mat <- matrix(1:9, nrow = 3)  
filter <- my\_mat %% 2 == \_\_\_  
my\_mat[filter]

###### 3. 我們繼續使用 great\_nba\_teams 這個資料框，請您分別利用 $變數名稱 與 [, "變數名稱"] 將 is\_champion 變數挑出來

team\_name <- c("Chicago Bulls", "Golden State Warriors")  
wins <- c(72, 73)  
losses <- c(10, 9)  
is\_champion <- c(TRUE, FALSE)  
season <- c("1995-96", "2015-16")  
great\_nba\_teams <- data.frame(team\_name, wins, losses, is\_champion, season, stringsAsFactors = FALSE)  
  
# 利用 `$變數名稱`  
great\_nba\_teams$\_\_\_  
  
# 利用`[, "變數名稱"]`  
greate\_nba\_teams[, "\_\_\_"]

### 延伸資訊

* [R 語言導論](https://www.datacamp.com/community/open-courses/r-%E8%AA%9E%E8%A8%80%E5%B0%8E%E8%AB%96#gs.29lL1x4)
* [Introduction to R](https://www.datacamp.com/courses/free-introduction-to-r)