# 第 07 天

## 把變數集結起來（3）

終於來到資料結構的第三天，今天我們要切入多維度的資料結構：

* 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）**

### 陣列

陣列（array）是矩陣的強化版，除了原有的水平方向資料（Row）與垂直方向資料（Column）我們可以再多指定一個維度。簡單來說，就是在一個陣列的物件之中，可以允許我們儲存多個矩陣。舉例來說我們就可以用陣列將 1 到 20 這 20 個數字儲存在 5 個 2x2 的矩陣之中：

> my\_arr <- array(1:20, dim = c(2, 2, 5))  
> my\_arr  
, , 1  
  
 [,1] [,2]  
[1,] 1 3  
[2,] 2 4  
  
, , 2  
  
 [,1] [,2]  
[1,] 5 7  
[2,] 6 8  
  
, , 3  
  
 [,1] [,2]  
[1,] 9 11  
[2,] 10 12  
  
, , 4  
  
 [,1] [,2]  
[1,] 13 15  
[2,] 14 16  
  
, , 5  
  
 [,1] [,2]  
[1,] 17 19  
[2,] 18 20

因為我們現在有了三個維度，所以在使用索引值選擇時，就需要增加第三個維度的索引值：

> my\_arr <- array(1:20, dim = c(2, 2, 5))  
> my\_arr[, , 1] # 第一個矩陣  
 [,1] [,2]  
[1,] 1 3  
[2,] 2 4  
> my\_arr[, , 2] # 第二個矩陣  
 [,1] [,2]  
[1,] 5 7  
[2,] 6 8  
> my\_arr[, , 3] # 第三個矩陣  
 [,1] [,2]  
[1,] 9 11  
[2,] 10 12  
> my\_arr[, , 4] # 第四個矩陣  
 [,1] [,2]  
[1,] 13 15  
[2,] 14 16  
> my\_arr[, , 5] # 第五個矩陣  
 [,1] [,2]  
[1,] 17 19  
[2,] 18 20

既然可以選擇出裡頭的五個矩陣，自然也可以深入選擇出其中數字、其中一個 row 或其中一個 column，像是選出 7 可以使用 [1, 2, 2] 、選出第二個矩陣的第一個 row 可以使用 [1, , 2]、選出第二個矩陣的第二個 column 可以使用 [, 2, 2]。

> my\_arr <- array(1:20, dim = c(2, 2, 5))  
> my\_arr[1, 2, 2] # 選出 7  
[1] 7  
> my\_arr[1, , 2] # 選出第二個矩陣的第一個 row  
[1] 5 7  
> my\_arr[, 2, 2] # 選出第二個矩陣的第二個 column  
[1] 7 8

### 清單

清單（list）是一個終極的**巨大容器**，它可以收納所有的 R 語言物件，包含單變數、一維資料結構、二維資料結構與多維資料結構。讓我們先建立幾個物件（文字、向量、矩陣以及資料框）出來：

title <- "Great NBA Teams"  
teams <- c("Chicago Bulls", "Golden State Warriors")  
wins <- c(72, 73)  
losses <- c(10, 9)  
winning\_percentage <- wins / (wins + losses)  
season <- c("1995-96", "2015-16")  
wins\_losses <- matrix(c(wins, losses), nrow = 2)  
df <- data.frame(Teams = teams, Winning\_Percentage = winning\_percentage, Season = season)

利用 list() 函數將其中幾個物件收納到一個 great\_nba\_teams 之中。

> title <- "Great NBA Teams"  
> teams <- c("Chicago Bulls", "Golden State Warriors")  
> wins <- c(72, 73)  
> losses <- c(10, 9)  
> winning\_percentage <- wins / (wins + losses)  
> season <- c("1995-96", "2015-16")  
> wins\_losses <- matrix(c(wins, losses), nrow = 2)  
> df <- data.frame(Teams = teams, Winning\_Percentage = winning\_percentage, Season = season)  
> great\_nba\_teams <- list(title, teams, wins\_losses, df)  
> great\_nba\_teams  
[[1]]  
[1] "Great NBA Teams"  
  
[[2]]  
[1] "Chicago Bulls" "Golden State Warriors"  
  
[[3]]  
 [,1] [,2]  
[1,] 72 10  
[2,] 73 9  
  
[[4]]  
 Teams Winning\_Percentage Season  
1 Chicago Bulls 0.8780488 1995-96  
2 Golden State Warriors 0.8902439 2015-16

這時候我們就可以清楚認知到清單強大的收納能力，從單變數到資料框都可以儲存在裡頭，這時候如果要選擇裡面的物件，跟先前 [] 一層中括號略有不同，改使用 [[]] 兩層中括號。

> title <- "Great NBA Teams"  
> teams <- c("Chicago Bulls", "Golden State Warriors")  
> wins <- c(72, 73)  
> losses <- c(10, 9)  
> winning\_percentage <- wins / (wins + losses)  
> season <- c("1995-96", "2015-16")  
> wins\_losses <- matrix(c(wins, losses), nrow = 2)  
> df <- data.frame(Teams = teams, Winning\_Percentage = winning\_percentage, Season = season)  
> great\_nba\_teams <- list(title, teams, wins\_losses, df)  
> great\_nba\_teams[[1]] # 選出清單中的第一個物件  
[1] "Great NBA Teams"  
> great\_nba\_teams[[2]] # 選出清單中的第二個物件  
[1] "Chicago Bulls" "Golden State Warriors"  
> great\_nba\_teams[[3]] # 選出清單中的第三個物件  
 [,1] [,2]  
[1,] 72 10  
[2,] 73 9  
> great\_nba\_teams[[4]]  
 Teams Winning\_Percentage Season  
1 Chicago Bulls 0.8780488 1995-96  
2 Golden State Warriors 0.8902439 2015-16

既然可以選擇出清單中裝的四個物件，自然也可以深入選擇出每個物件中的元素：

> title <- "Great NBA Teams"  
> teams <- c("Chicago Bulls", "Golden State Warriors")  
> wins <- c(72, 73)  
> losses <- c(10, 9)  
> winning\_percentage <- wins / (wins + losses)  
> season <- c("1995-96", "2015-16")  
> wins\_losses <- matrix(c(wins, losses), nrow = 2)  
> df <- data.frame(Teams = teams, Winning\_Percentage = winning\_percentage, Season = season)  
> great\_nba\_teams <- list(title, teams, wins\_losses, df)  
> great\_nba\_teams[[3]][1, ] # 選出芝加哥公牛隊 1995-96 的戰績  
[1] 72 10  
> great\_nba\_teams[[4]]$winning\_percentage # 選出創紀錄的球季勝率  
[1] 0.8780488 0.8902439

建立清單的同時，如果指定了物件的命名，就可以在選擇時使用 $物件名稱。

> title <- "Great NBA Teams"  
> teams <- c("Chicago Bulls", "Golden State Warriors")  
> wins <- c(72, 73)  
> losses <- c(10, 9)  
> winning\_percentage <- wins / (wins + losses)  
> season <- c("1995-96", "2015-16")  
> wins\_losses <- matrix(c(wins, losses), nrow = 2)  
> df <- data.frame(Teams = teams, Winning\_Percentage = winning\_percentage, Season = season)  
> great\_nba\_teams <- list(title = title, teams = teams, wins\_losses = wins\_losses, df = df)  
> great\_nba\_teams$teams # 選出 teams 這個向量  
[1] "Chicago Bulls" "Golden State Warriors"  
> great\_nba\_teams$df # 選出 df 這個資料框  
 Teams Winning\_Percentage Season  
1 Chicago Bulls 0.8780488 1995-96  
2 Golden State Warriors 0.8902439 2015-16

R 語言有很多非常便利的函數，這些函數的輸出多半是一個清單。假如我們有兩個向量 x 與 y，兩個向量的關係是 y = 2x + 5，我們可以利用 lm() 這個函數建立一個 lm\_fit 清單：

x <- 1:10  
y <- 2 \* x + 5  
lm\_fit <- lm(formula = y ~ x)

而從這個清單中，我們可以取出裡面的 coefficients 這個物件（是一個向量），就可以知道 x 跟 y 的關係，跟我們建立時的設定完全相同：

> x <- 1:10  
> y <- 2 \* x + 5  
> lm\_fit <- lm(formula = y ~ x)  
> lm\_fit$coefficients  
(Intercept) x   
 5 2   
> lm\_fit$coefficients[1] # 截距為 5  
(Intercept)   
 5   
> lm\_fit$coefficients[2] # x 係數為 2  
x   
2

### 小結

好啦！第七天的內容就是這麼多，我們又認識了 R 語言的兩種資料結構：陣列與清單，總算把所有的資料結構都討論完畢了。

### 練習題

###### 1. 我們把 1 到 1000 儲存在 10 個 10x10 的矩陣，並且收納在一個陣列 my\_arr 之中，請你練習用索引值將 315 這個數字選出來

my\_arr <- array(1:1000, dim = c(10, 10, 10))  
my\_arr[\_\_\_, \_\_\_, \_\_\_]

###### 2. 我們已經建立好幾個物件，請您幫我們收納進一個清單中叫做 worst\_nba\_teams，並利用 [[索引值]] 從清單中選出 wins 這個向量

title <- "Worst NBA Teams"  
teams <- c("Charlotte Bobcats", "Philadelphia 76ers")  
wins <- c(7, 9)  
losses <- c(59, 73)  
worst\_nba\_teams <- list(\_\_\_, \_\_\_, \_\_\_, \_\_\_)  
worst\_nba\_teams[[\_\_\_]]

###### 3. 同樣的幾個物件，請您建立清單 worst\_nba\_teams 的時候為每個物件命名，並且利用 $物件名稱 從清單中選出 teams 這個向量

title <- "Worst NBA Teams"  
teams <- c("Charlotte Bobcats", "Philadelphia 76ers")  
wins <- c(7, 9)  
losses <- c(59, 73)  
worst\_nba\_teams <- list(\_\_\_ = title, \_\_\_ = teams, \_\_\_ = wins, \_\_\_ = losses)  
worst\_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)