



# 資料物件、物件導向 與程式撰寫技巧

主辦單位：台灣資料科學與商業應用協會

主講顧問：Andrew Tang (湯明軒)

任職：中強光電雲端服務應用處 資料研發工程師



# 大綱

- 資料與物件(data and object)
  - 常見資料類型(type, mode)
  - 常見物件類型(class)
- 物件導向：R泛型函數介紹
  - 看懂R泛型函數
  - 實作R泛型函數
- 程式撰寫技巧
  - 各種好用函數介紹
  - 網頁資料擷取技巧
  - Code style

# 前置作業 for rJava

- 安裝JRE

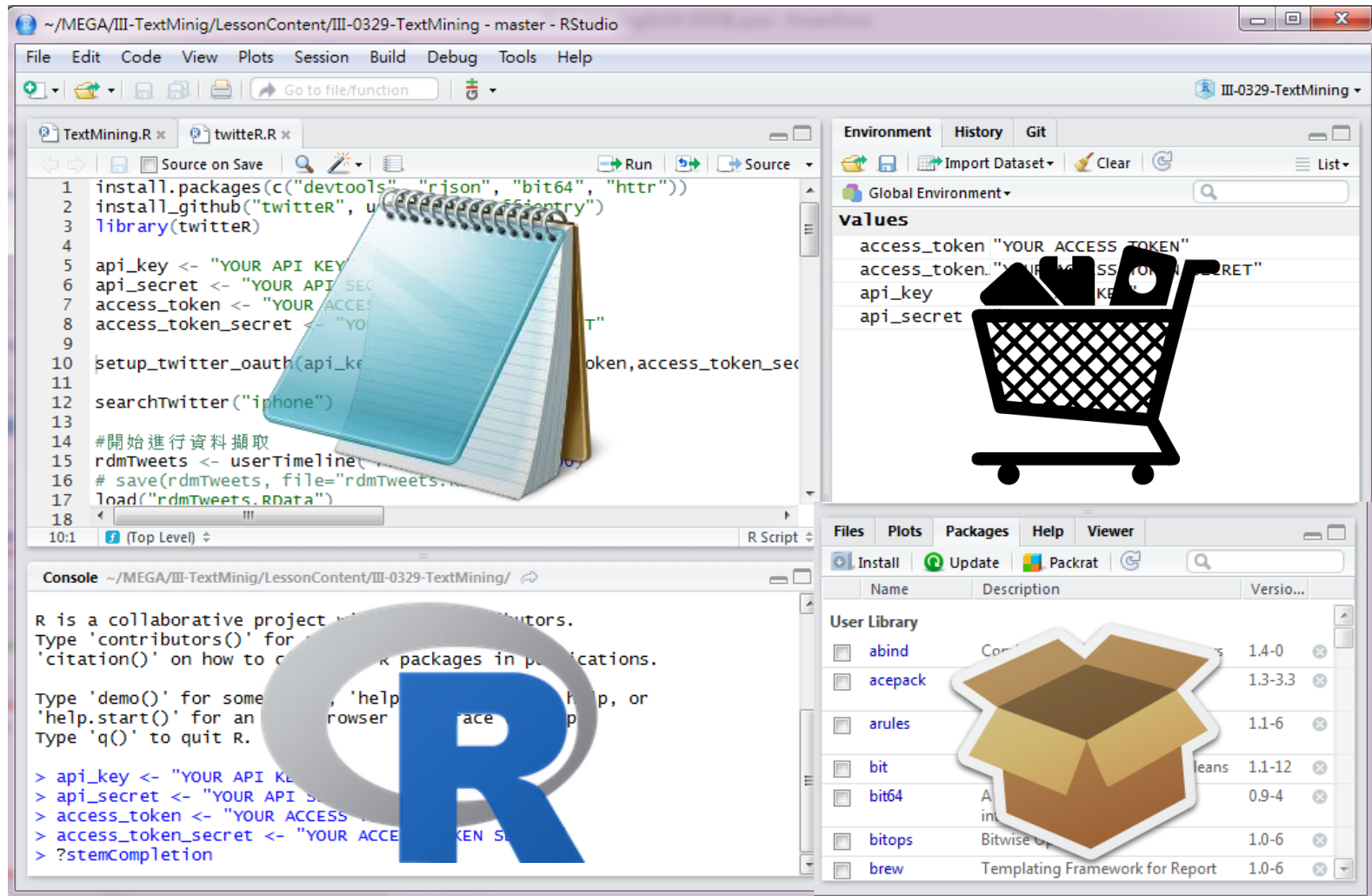
1. 以Google搜尋Java並選擇

- [下載免費Java 軟體](#)



The screenshot shows the official Java download page. At the top, there's a red banner with the Java logo and the word "Java" in white. Below the banner, there's a section titled "所有 Java 下載" (All Java Downloads) with a sub-header "如果您想為其他電腦或作業系統下載 Java，請按一下以下連結" (If you want to download Java for other computers or operating systems, click the link below). The link "所有 Java 下載" is circled in red. To the right of this section, there's a large red button that says "免費 Java 下載" (Free Java Download). Below the button, there's a link "» 什麼是 Java ? » 我有 Java 嗎 ? » 需要說明嗎 ?".

# RStudio介面概述





# RStudio 快捷鍵

- 給予變數值符號「<-」
  - Alt + 「-」
- 刪除整行
  - Ctrl + D
- 區塊註解
  - Ctrl + Shift + C
- 清空Console code
  - Ctrl + L

# 前置作業 load data

- <https://github.com/sulaxd/DSIA0605>
- 載入資料 hospital.xlsx
  - library(XLConnect) #若無安裝請先安裝
  - xlsx <- loadWorkbook("hospital.xlsx")
  - hospitalData <- readWorksheet(xlsx,1)
  - head(hospitalData)
- 資料說明：
  - KingNet國家網路醫院的醫護陣容資料，經過網頁資料擷取、處理並萃取出與醫院相關資料整理而成，透過Google API將地址轉為經緯度並以方圓五公里為門檻判斷附近是否有停車場。



Everything in R is an object;  
Every object in R has a class.

## 資料與物件

# 資料類型 (1/2)

- 對R語言來說，最精確描述類型的函數
- `typeof()`
  - 含整數值的向量(`integer`)
    - `c(1, 2, 3, 4, 5)`
  - 含實數值的向量(`double`)
    - `c(1.0, 2.2, 3.06, 1.111, 5.1)`
  - 含複數值的向量(`complex`)
    - `c(1i, 1i+5, 3i+1)`
  - 含字元值的向量(`character`)
    - `c("apple", "123", "0.12", "中文")`
  - 含邏輯值的向量(`logical`)
    - `c(TRUE, FALSE, F, T)`



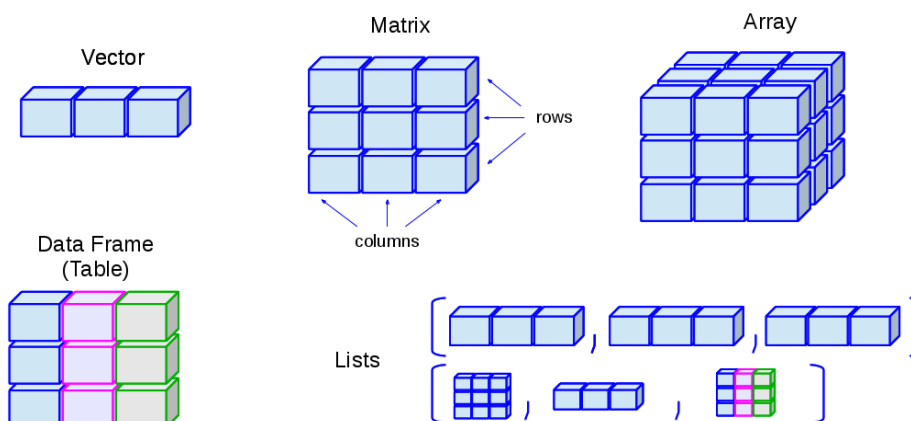
# 資料類型 (2/2)

- 較不精確的描述，但通常為R User所考慮，故此課程中我們稱mode函數所回傳之值為資料類型
- mode()
  - 數值向量(integer)
    - c(1, 2, 3.06, 1.111, 5)
  - 複數向量(complex)
    - c(1i, 1i+5, 3i+1)
  - 字元向量(character)
    - c("apple", "123", "0.12", "中文")
  - 邏輯向量(logical)
    - c(TRUE, FALSE, F, T)

# 物件類型 (1/2)

- `class()`
- 影響順序：Character > Complex > Numeric > Logical

|     | (資料類型) 同質            | (資料類型) 異質         |
|-----|----------------------|-------------------|
| 1d. | 原子向量 (Atomic Vector) | 串列 (List)         |
| 2d. | 矩陣 (Matrix)          | 資料框架 (Data frame) |
| nd. | 陣列 (Array)           |                   |



# 物件類型 (2/2)

- 向量：
  - 相同資料類型的一維度陣列，`class()`回傳值為其資料類型
- 因子：
  - 為方便處理類別資料的向量
- 矩陣：
  - 相同資料類型的二維度陣列，`mode()`回傳值為其資料類型
- 陣列：
  - 相同資料類型的多維度陣列，`mode()`回傳值為其資料類型
- 串列：
  - 特殊的向量，其向量元素為物件，每個元素可為不同資料類型
- 資料框架：
  - 通常每一列(`row`)表示個體，欄(`column`)表示變數，每個欄位可為不同資料類型
- 函數：
  - User自定義的運算流程



泛型函數 - 根據傳入物件的類型決定調用哪個具體的方法

## 物件導向：R泛型函數介紹



# R物件導向概論(1/2)

- R語言提供了3種物件導向設計(**Object-oriented Programming**)的底層物件類型，一種是S3類型，一種是S4類型，還有一種是RC類型。
- S3是R的第一個物件導向系統，物件簡單、具有動態性、結構化特徵不明顯；S4物件結構化、功能強大；RC物件是2.12版本後使用的新類型，用於解決S3, S4很難實現的物件。
- 透過pryr套件，可以檢測物件為S3，S4還是其他類型



# R物件導向概論(2/2)

- S3類型的物件導向設計是基於泛型函數(**Generic Function**)的概念實現
- 泛型函數 - 函數裡面包含了多個**方法(Method)**
- 泛型函數的工作是針對不同的輸入類型(**Class**)，決定以什麼樣的方法執行 - **實際上就是一個分派機制**
- 換言之，同一泛型函數會針對物件的類別做出特化的行為

# 操作重點

- 查詢泛型函數對應的物件類型及物件類型對應的泛型函數
  - `methods(generic.function, class)`
- 實際測試泛型函數對於不同物件類型使用的不同方法
  - `plot`: `default`, `table`, `factor`, `hist`
- 設計一個自己的泛型函數
  - `h <- list(a="Print me", b="Don't print")`
  - `class(h) <- "myclass"`
  - `print.myclass<-function(x){`  
`cat("A is:",x$a,"\n")}`



80%的問題，Google都有解答

# 程式撰寫技巧



# 程式撰寫技巧(1/4)

- 列出當前目錄下的檔案
  - `list.files(".')`
- 瀏覽資料夾並選擇欲載入檔案
  - `file.choose()`
- 只保留向量中不等於1的值，以邏輯值為索引值(index)
  - `c <- c(1, 1, 2, 2, 3, 4, 1)`
  - `c <- c[c!=1]`
- 互動式取得Plot中的某一點作標
  - `plot(0)`
  - `locator()`

# 程式撰寫技巧(2/4)

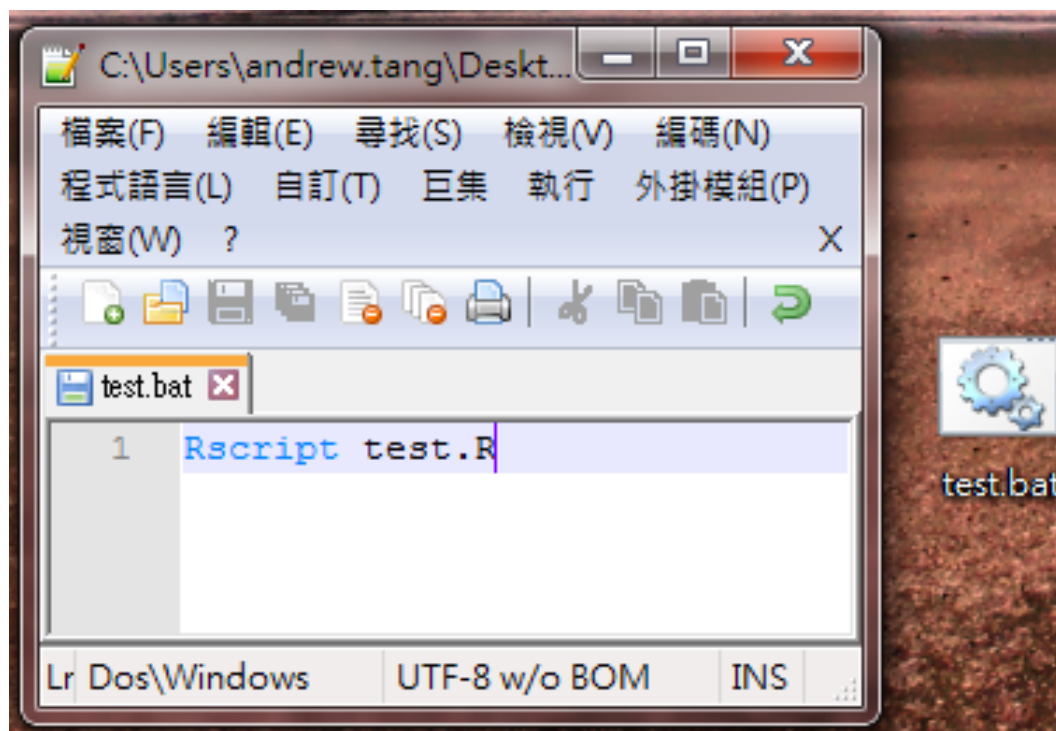
- 瀏覽某個指定的網頁
  - `browseURL("URL")`
- 下載網絡文件到本地
  - `download.file("URL", "存檔名稱", mode="wb")`
- 執行某個檔案
  - `shell.exec("C:/hospital.xlsx")`
- 對話框(only for windows)
  - `winDialog("yesno", "愛上R了嗎?")`
- 讓User輸入資料
  - `print("請問您幾歲")`
  - `age <- readLines(n=1)`

# 程式撰寫技巧 (3/4)

- 設定輸出的小數點位數
  - `options(digits = 10)`
- 避免以科學符號呈現過長數值 (`8.8888888889e+84`)
  - `options(scipen = 84)`
- 從剪貼簿載入資料及輸出資料至剪貼簿
  - `readLines("clipboard")`
  - `writeClipboard(str = '欲輸出字串')`
- 禁止字串被自動轉為因子，通常在載入資料框架時使用
  - `stringsAsFactors = FALSE`
- 暫時停止程式
  - `Sys.sleep(秒)`

# 程式撰寫技巧(4/4)

- 以R製作可執行程式 .bat檔
  - 將R.exe所在路徑加入環境變數



# - 網頁資料擷取技巧 -

## Google play APP評論擷取

- 善用F12開發人員工具

參考程式碼

GooglePlayCommentExtraction.R





# style guide(1)

- 檔案命名方式
- # Good
  - fit-models.R
  - utility-functions.R
- 若有順序關係時
  - 0-download.R
  - 1-parse.R
  - 2-explore.R
- # Bad
  - foo.r
  - stuff.r

# style guide(2)

- 物件命名方式
  - # Good
    - day\_one
    - dayOne
  - # Bad
    - first\_day\_of\_the\_month
    - dayone
    - djm1
- 切忌
  - T <- FALSE
  - c <- 10
  - mean <- function(x) sum(x)



# style guide(3)

- 程式碼句法
  - # Good
    - average `<-` mean(feet `/` 12 `+` inches, na.rm `=` TRUE)
  - # Bad
    - average<-mean(feet/12+inches,na.rm=TRUE)
- |             |               |
|-------------|---------------|
| – # Good    | – # Bad       |
| – x <- 1:10 | – x <- 1 : 10 |
| – base::get | – base :: get |



# style guide(4)

- 大括弧排版

# Good

```
if (y < 0 && debug) {  
    message("Y is negative")  
}
```

```
if (y == 0) {  
    log(x)  
} else {  
    y ^ x  
}
```



# style guide(5)

# Load data -----

# Plot data =====

# 經驗談

- 善用R套件(目前官方6719個)
- 避免使用Windows處理中文資料
- 編碼轉換：Notepad++
- 80%的問題，網路都找的到
  - Stack Overflow、统计之都：COS论坛
- 免費API
  - 如：地圖資訊(Google Map)
    - <http://maps.googleapis.com/maps/api/geocode/json?address=地址>
- 開放資料
  - 如：天氣資料(10分鐘雨量觀測資料)
    - <http://opendata.cwb.gov.tw/datadownload?dataid=0-A0002-001>



# Reference

- <http://bioankeyang.blogspot.tw/search/label/R>
- <http://www.biosino.org/R/R-doc/>
- <http://cran.r-project.org/doc/contrib/Liu-FAQ.pdf>
- <http://venus.ifca.unican.es/Rintro/dataStruct.html>
- <http://blog.fens.me/r-class-s3/>
- <https://sites.google.com/site/rnotewush/wu-jian-dao-xiang>
- <http://www.everdark.info/2013/02/r-r.html>
- <http://google-styleguide.googlecode.com/svn/trunk/Rguide.xml>
- <http://cos.name/cn/>
- <http://stackoverflow.com/>



# Q & A

[sulaxd@gmail.com](mailto:sulaxd@gmail.com)  
<http://www.r-software.org/>

