

AI機器學習與深度學習應用 – 使用Python與R語言 2022(R)

大數據分析

- R/Python/Julia/SQL程式設計與應用
(R/Python/Julia/SQL Programming and Application)
- 資料視覺化 (Data Visualization)
- 機器學習 (Machine Learning)
- 統計品管 (Statistical Quality Control)
- 最佳化 (Optimization)



李明昌博士

alan9956@gmail.com

<http://rwepa.blogspot.com/>

個人簡介 <http://rwepa.blogspot.com/>

- 姓名：李明昌 (ALAN LEE)
- 現職：中華R軟體學會 常務理事
臺灣資料科學與商業應用協會 常務理事
- 學歷：中原大學 工業與系統工程所 博士
- 經歷：
 - 育達科技大學 資訊管理系(所) 專任助理教授
 - 佛光大學 兼任教師
 - 國立台北商業大學 兼任教師
 - 東吳大學 兼任教師
 - 崇友實業 行銷企劃專員
 - 國航船務代理股份有限公司 海運市場運籌管理員
- 大專院校、資策會、工業技術研究院、國家發展委員會、中央氣象局、公平交易委員會、各縣市政府與日本名古屋產業大學等公民營單位演講達300餘場，2600小時以上。
- 連絡資訊：alan9956@gmail.com



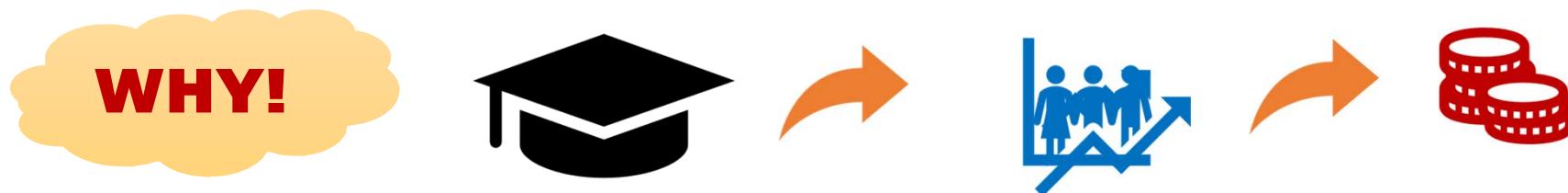
- iPAS 巨量資料分析師 證照推廣
- iPAS 營運智慧分析師 證照推廣

大綱

- 06.R與RStudio簡介,資料匯入與匯出
- 07.R資料物件,判斷式與函數,群組分析與繪圖graphics
- 08.資料操作dplyr,視覺化ggplot2,互動式表格視覺化
- 09.地理資料視覺化leaflet,機器學習與深度學習應用
- 10.基礎互動式shiny,進階互動式shiny Server 佈署

如何學習 Python & R?

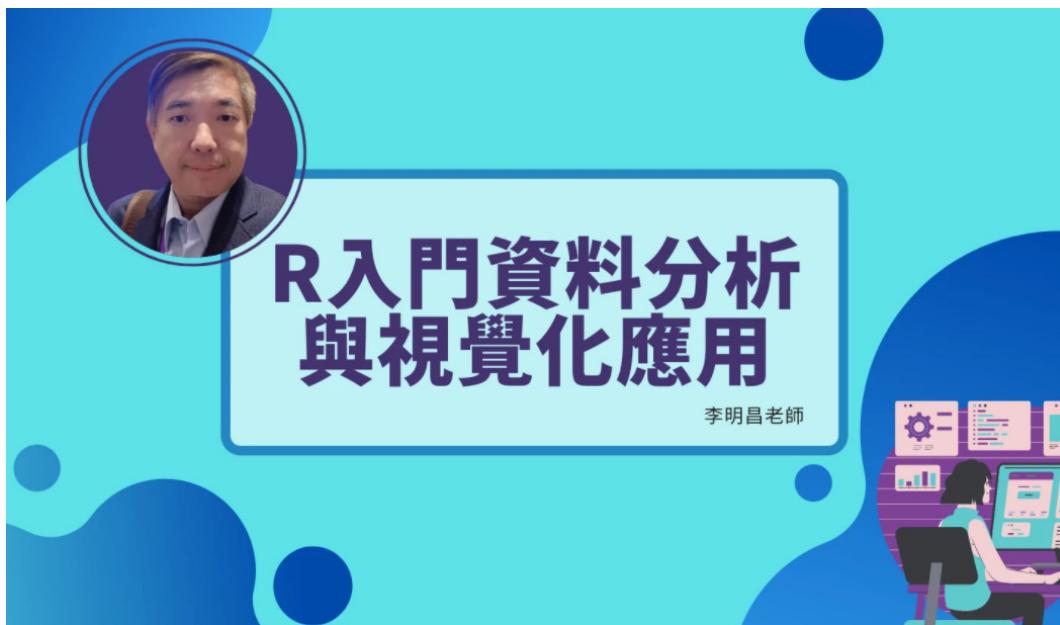
- 熟悉教材內容
- 將教材內容的資料集改為工作資料集
- 遇到問題時, 想辦法尋找答案
- 掌握 APC方法
- 掌握 摘要, 繪圖, 建模
- 參考網路應用文章 (進階) & 學術論文



R 入門資料分析與視覺化應用(7小時28分鐘)

- <https://mastertalks.tw/products/r?ref=MCLEE>

課程提供教學範例的原始程式檔案與資料集



- **主題**
 1. R, RStudio簡介與套件使用
 2. 認識資料物件
 3. 資料處理與分析
 4. 資料視覺化應用
- **特色**
 1. 資料分析的**關鍵八步**
 2. 提供必備**ggplot2**套件的應用知識與使用情境
 3. 提供日期時間**zoo, xts**套件的整合應用操作
 4. 提供**人力資源**資料與**銷售**資料，強化**實務資料**操作能力

R 商業預測應用(8小時53分鐘)

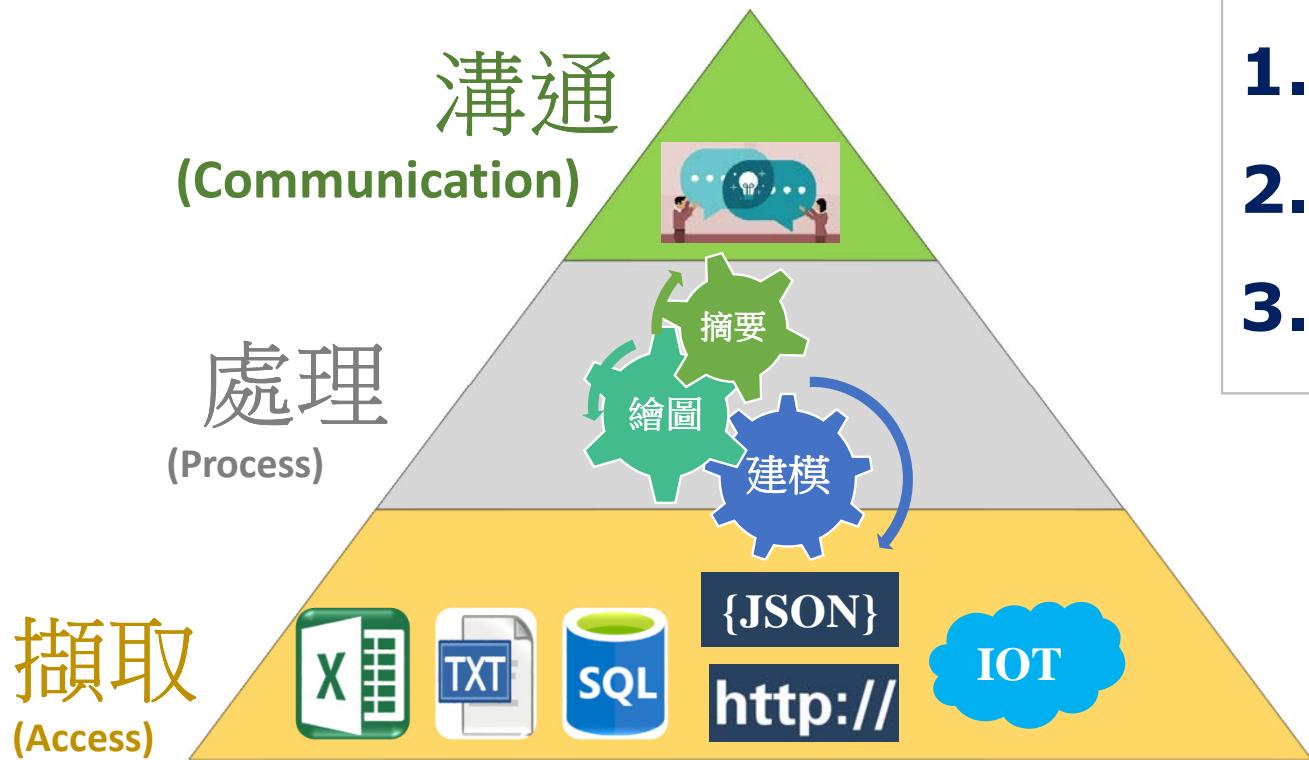
- <https://mastertalks.tw/products/r-2?ref=MCLEE>



課程提供教學範例的原始程式檔案與資料集

- **主題**
 1. R , RStudio工具操作
 2. 非監督式學習商業預測
 3. 監督式學習商業預測
 4. 財金資料預測應用
- **特色**
 1. 採用**最有效率**方式學習大數據R語言，並應用於**職場資料分析**與**商業預測應用**
 2. 提供**多元線性迴歸**的必備知識
 3. 提供**財金資料商業預測應用**的基礎與進階必學技能
 4. 提供學員人力資源資料與**台指期tick資料**預測演練

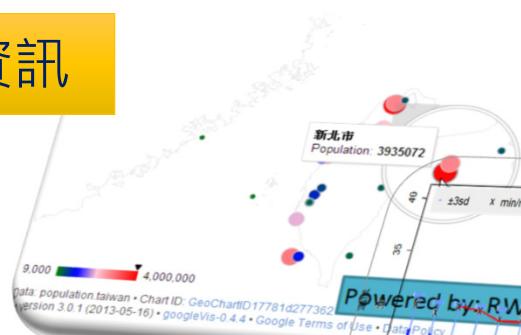
★★★資料分析架構→APC方法



資料分析/視覺化應用

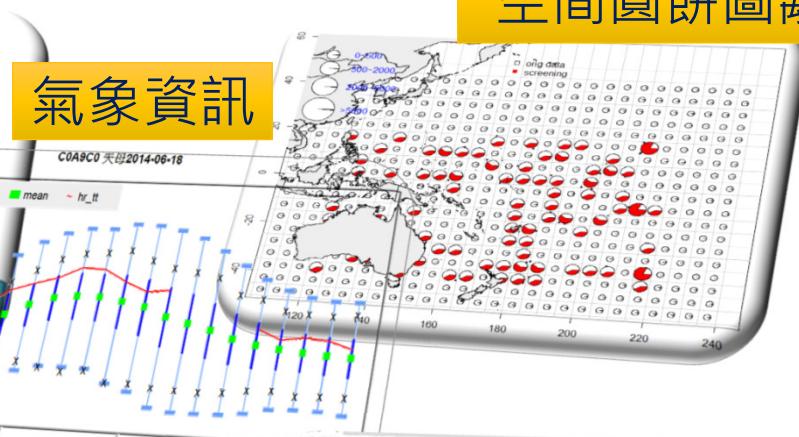
R + shiny → 互動式網頁

地理資訊



空間圓餅圖離群值分析

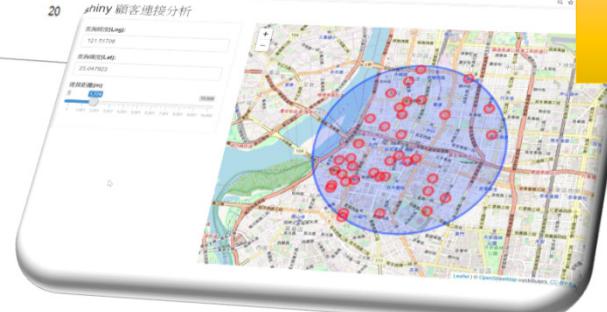
氣象資訊



保險預測



顧客連結資訊



中央氣象局 1,600萬筆資料

網頁呈現



客製化選單

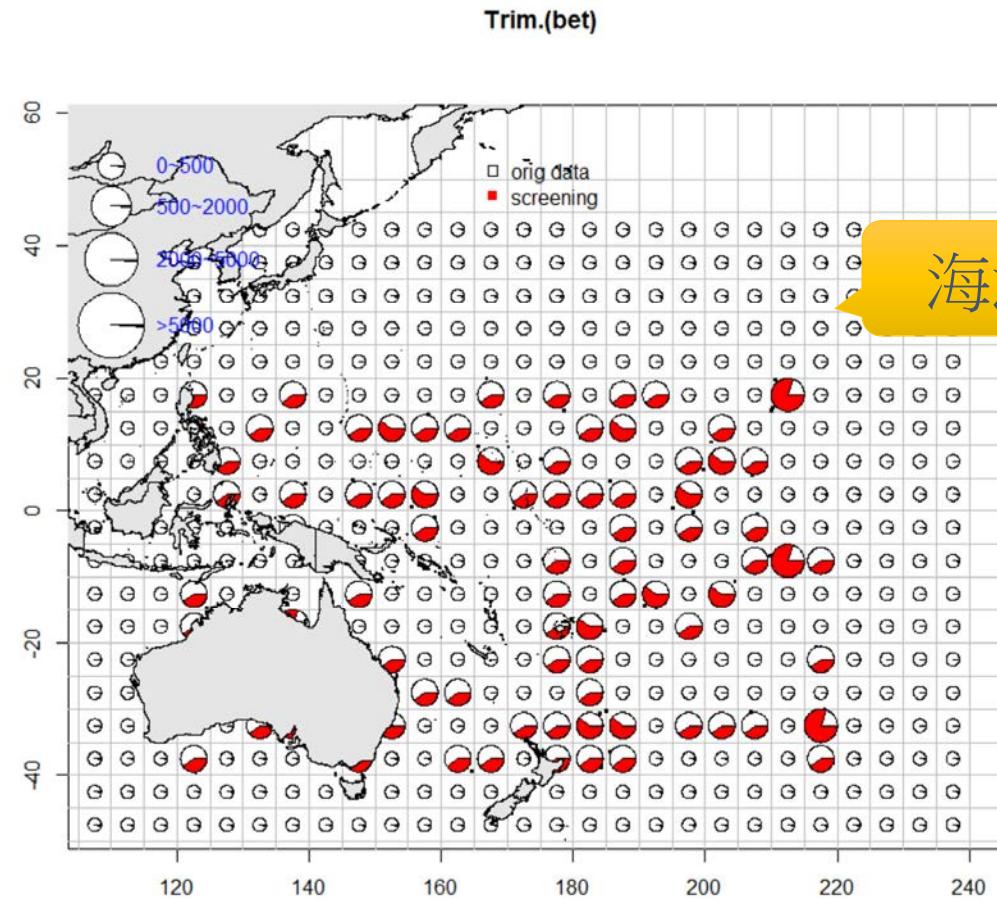
R統計運算

保險預測模型

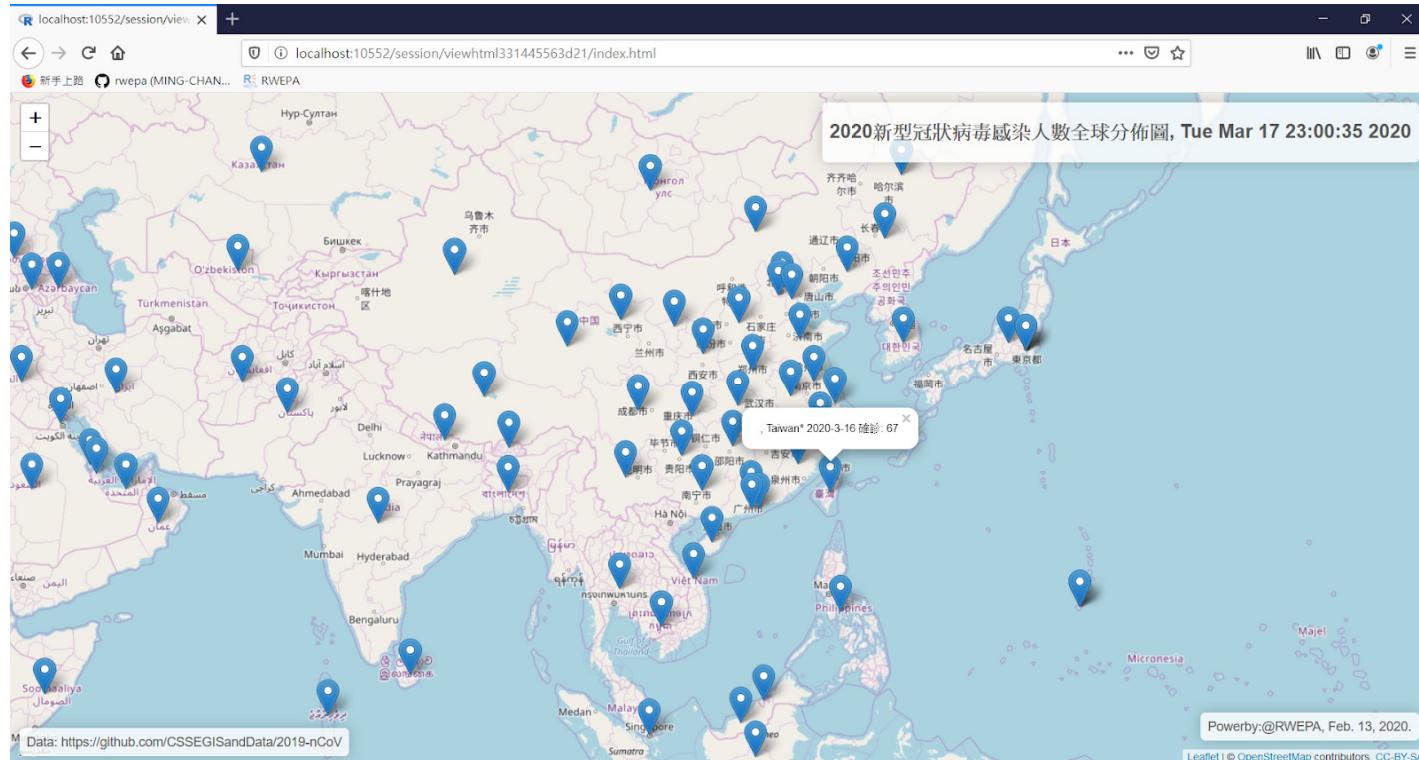
The screenshot shows the iinsurance interactive analysis platform version v.16.3.24. The interface includes a navigation bar with tabs like '檔案上傳', '資料處理', '統計圖表', '模型評估', and '預測模型'. A yellow callout bubble labeled '機率模型閾值調整' points to a slider labeled '機率模型閾值' with a value of 0.1. Another yellow callout bubble labeled '預測結果' points to a table showing prediction results for 12 entries. The table columns include gender, vehicle type, exposure risk, exposure risk count, discount rate, insured person age, private car 0, private car 1, private car 2, private car 0-1-2 combination, age 0-1-2 combination, predicted probability, and claim status. The claim status column shows values like '有' (Yes) and '無' (No). Red boxes highlight the '機率模型閾值' slider, the '檢視結果' button in the top right, and the '預測機率' and '理賠' columns in the table.

性別	女性	車輛種類	私家車	曝露風險		曝露風險對數		無索償折扣	被保險人年齡	私家車			私家車			私家車			私家車 車齡 0_1_2 組合	車齡 0_1_2 組合	預測機率	理賠
				0	1	私	家			車	一	車	齡	0	1	2	私	家				
M	0	A	1	0.9144422	-0.08944106	50	4	1	0	0	1	0	2	0.1069	有							
M	0	A	1	0.8158795	-0.20348856	20	4	0	0	1	1	2	2	0.1441	有							
3	M	0	A	1	0.8377823	-0.17699695	50	3	0	0	1	1	2	2	0.1866	有						
4	M	0	A	1	0.4325804	-0.83798702	50	6	0	1	0	1	1	2	0.0944	無						
5	M	0	A	1	0.7173169	-0.33223755	50	4	0	0	1	1	2	2	0.1218	有						
6	M	0	A	1	0.8377823	-0.17699695	50	4	0	0	1	1	2	2	0.1495	有						
7	M	0	A	1	0.8487337	-0.16400975	50	5	0	0	1	1	2	2	0.1422	有						
8	F	1	A	1	0.8268309	-0.19015503	10	3	0	0	1	1	2	2	0.1733	有						
9	M	0	A	1	0.7145791	-0.33606164	0	5	1	0	0	1	0	2	0.0694	無						
10	M	0	A	1	0.3340178	-1.09656101	0	3	0	0	1	1	2	2	0.0783	無						

空間圓餅圖離群值分析



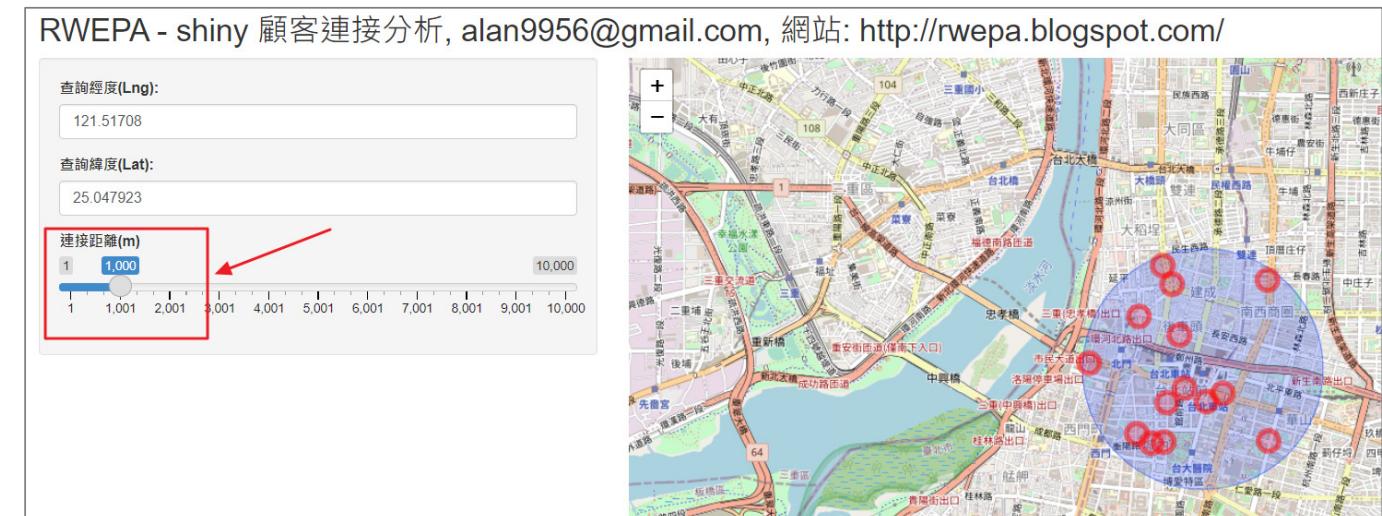
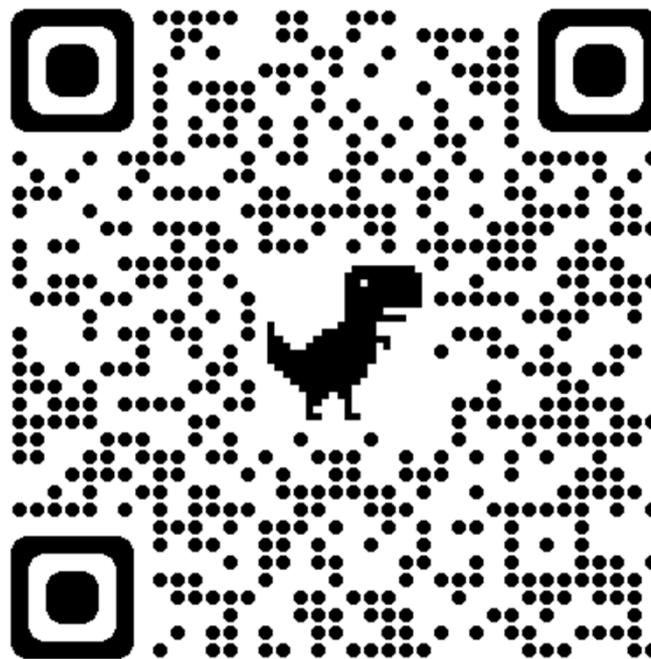
2020新型冠狀病毒視覺化



<http://rwepa.blogspot.com/2020/02/2019nCoV.html>

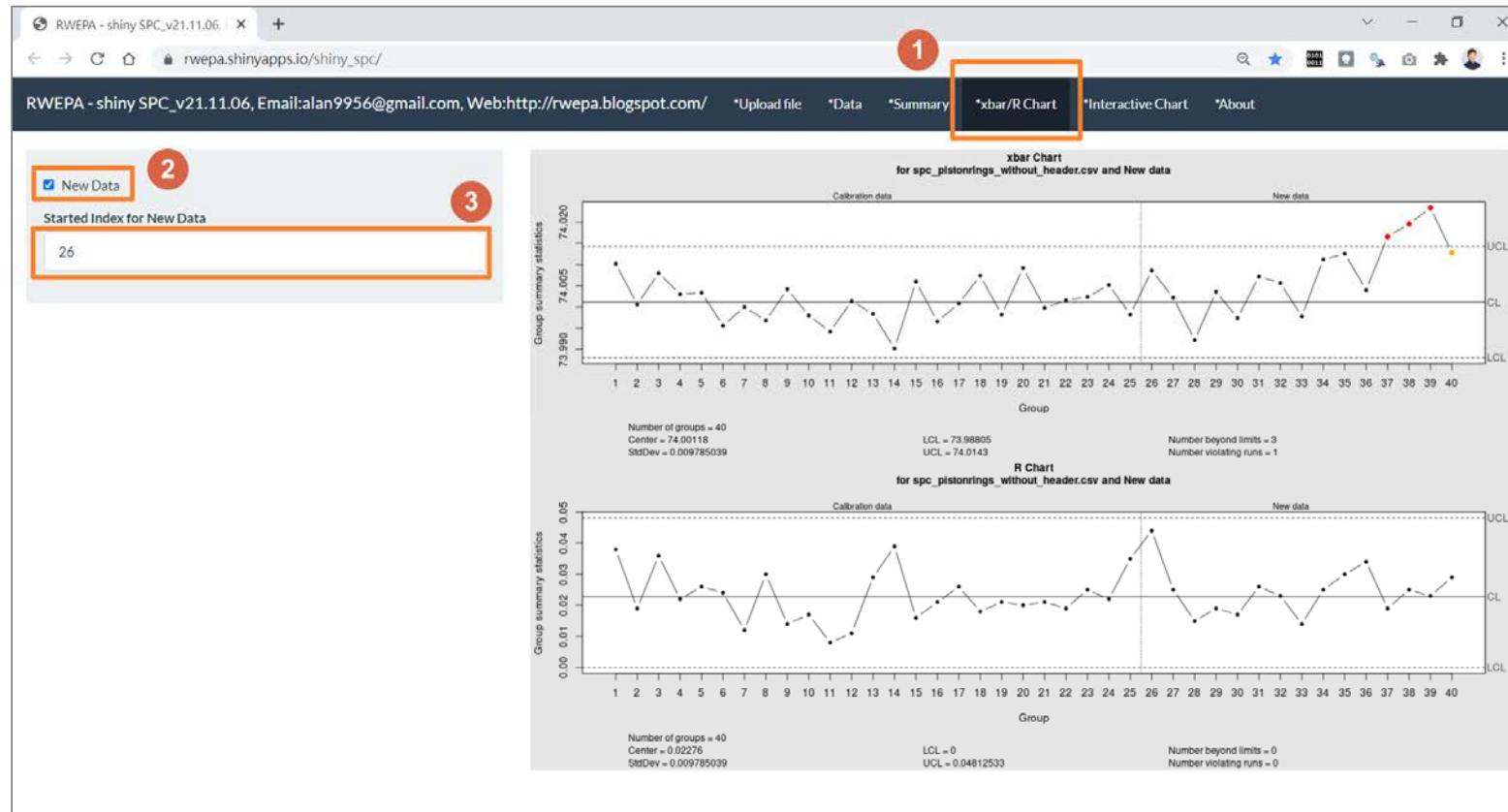
shiny 顧客連接分析

- <https://rwepa.shinyapps.io/shinyCustomerConnect/>



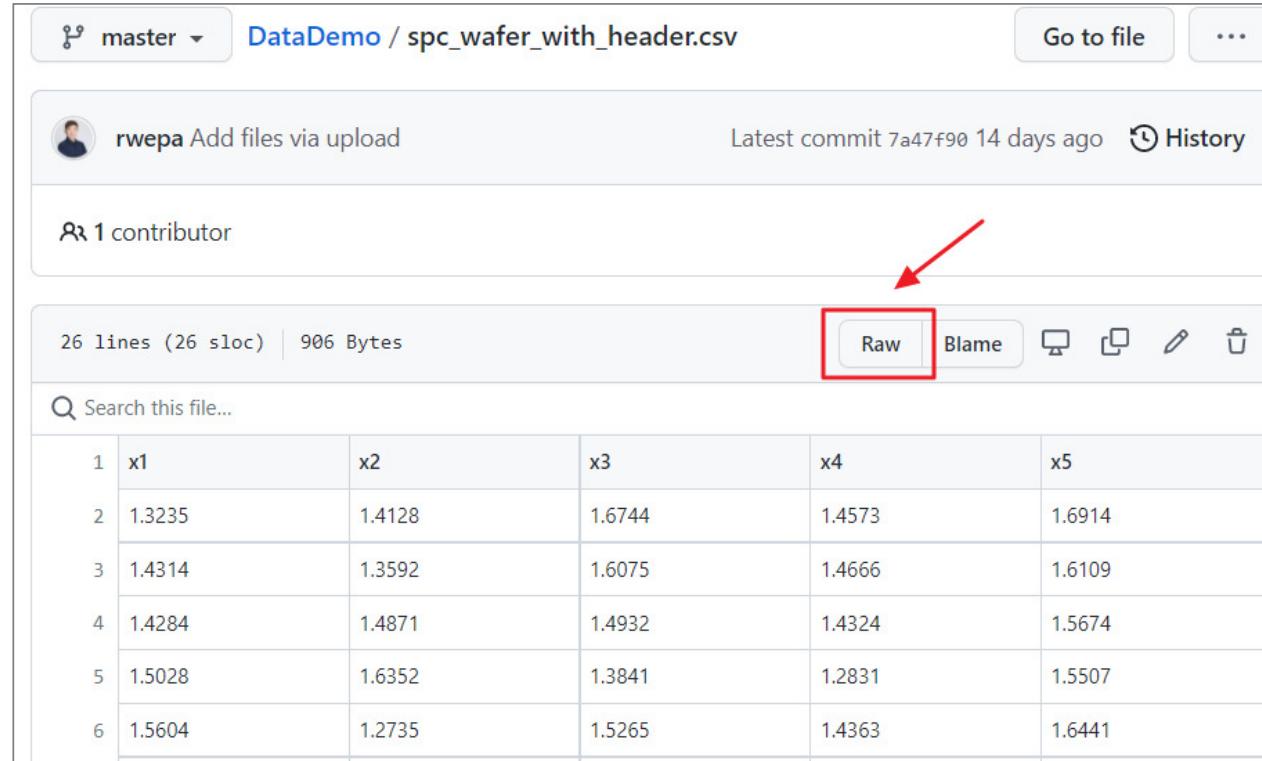
品質管制圖(quality control chart)應用

- <http://rwepa.blogspot.com/2021/10/r-shiny-quality-control-chart.html>



品質管制圖應用 (續)

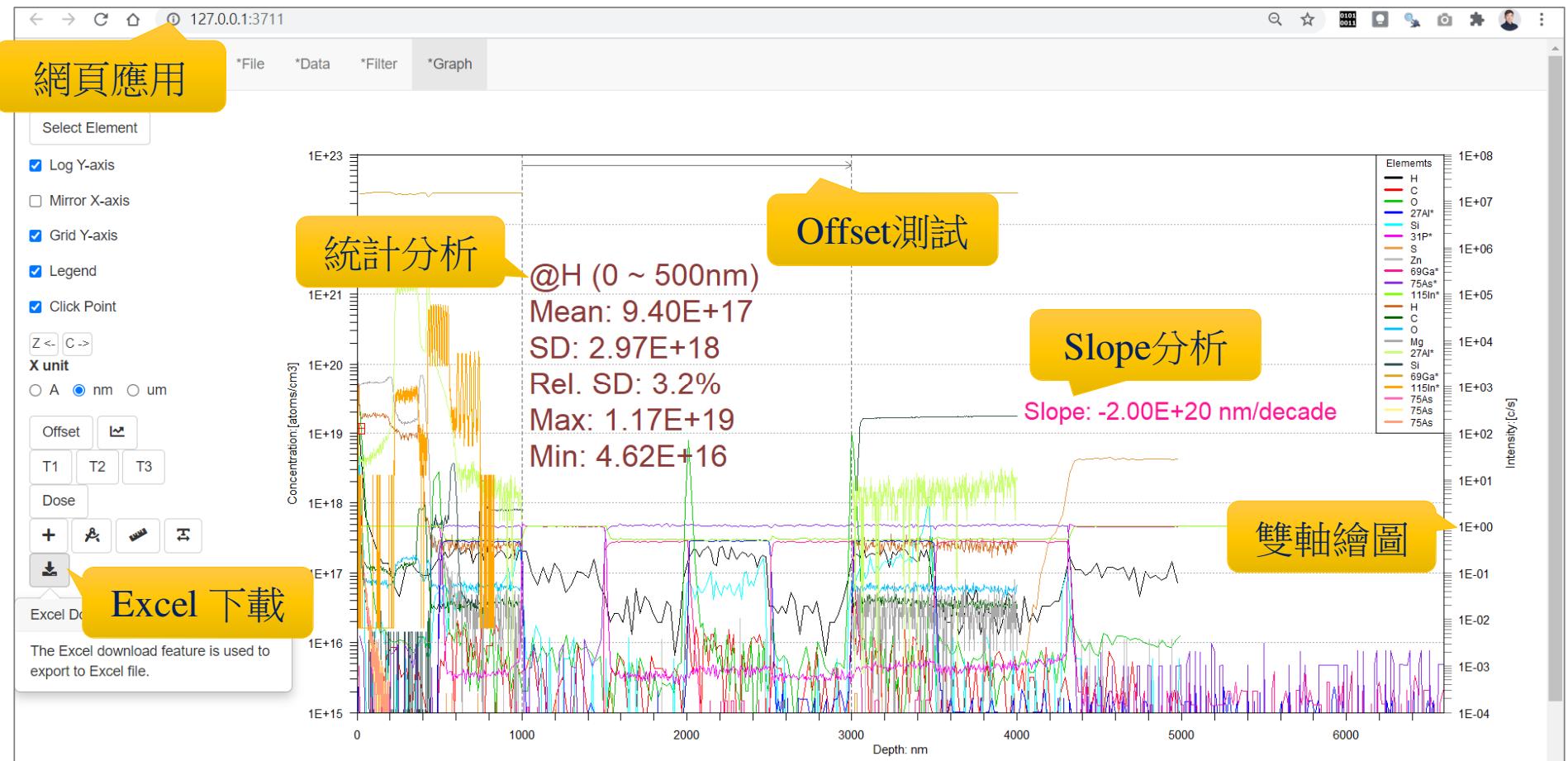
- https://github.com/rwepa/DataDemo/blob/master/spc_wafer_with_header.csv
- https://github.com/rwepa/DataDemo/blob/master/spc_pistonrings_without_header.csv



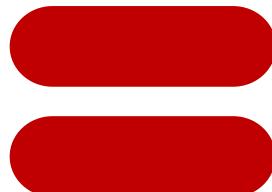
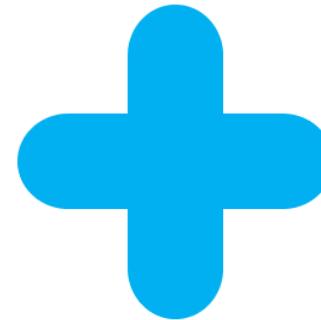
The screenshot shows a GitHub file page for 'spc_wafer_with_header.csv'. At the top, it displays the repository 'DataDemo' and the file path 'spc_wafer_with_header.csv'. Below this, it shows the author 'rwepa' and the commit message 'Add files via upload'. It indicates there is 1 contributor and a latest commit 14 days ago. A red arrow points to the 'Raw' button in the toolbar below the file stats. The table below contains 6 rows of data:

	x1	x2	x3	x4	x5
2	1.3235	1.4128	1.6744	1.4573	1.6914
3	1.4314	1.3592	1.6075	1.4666	1.6109
4	1.4284	1.4871	1.4932	1.4324	1.5674
5	1.5028	1.6352	1.3841	1.2831	1.5507
6	1.5604	1.2735	1.5265	1.4363	1.6441

離子資料分析與視覺化應用



學習目標



06.R與RStudio簡介,資料匯入與匯出



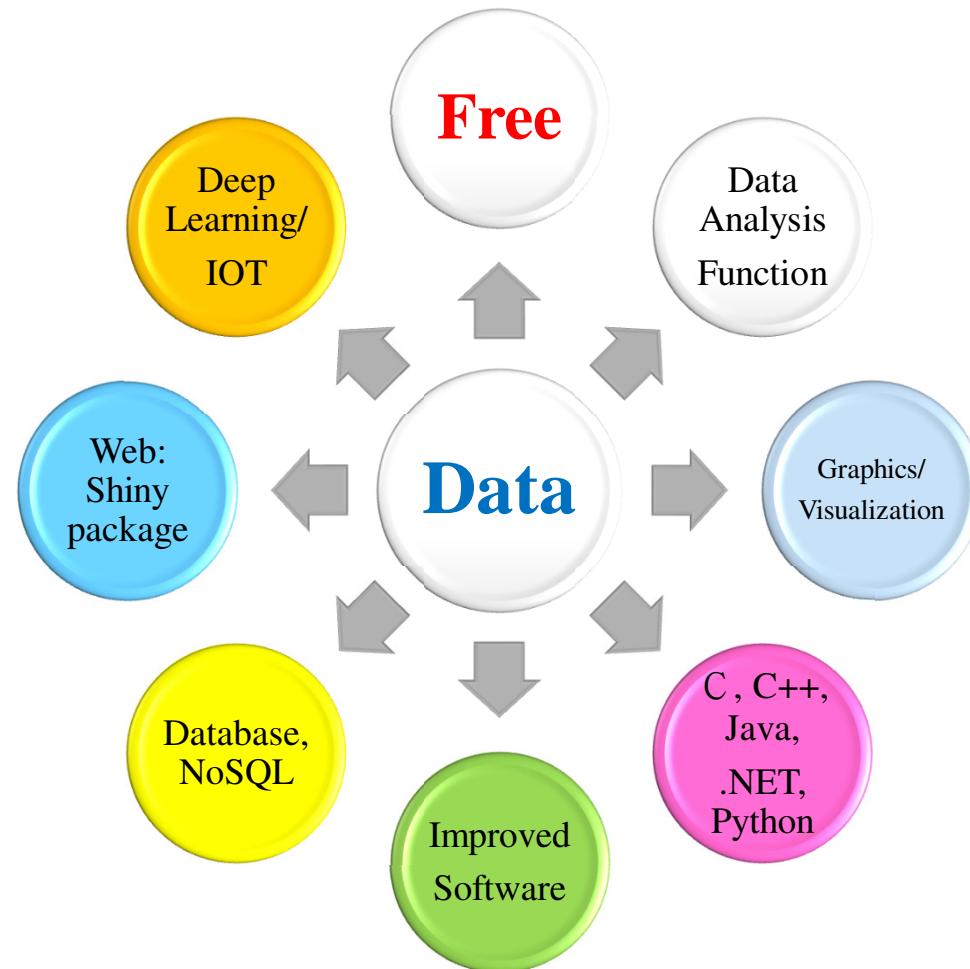
R 安裝與簡介

認識 R

- 1976 - 貝爾實驗室 John Chambers, Rick Becker, and Allan Wilks 研發S語言。
- 1993 - Ross Ihaka and Robert Gentleman, University of Auckland, New Zealand 研發R 語言。
 - R 是一種基於 S 語言所發展出具備統計分析、繪圖與資料視覺化的程式語言。
- 1997年 - R的核心開發團隊 (R development core team) 成立，專責R原始碼的修改與編寫。
 - 2000年2月 - R 1.0.0
 - 2013年3月 - R 2.15.3
 - 2021年11月 - R 4.1.2



R-八大功能



R官方網頁

[\[Home\]](#)[Download](#)[CRAN](#)**下載****繪圖**[R Project](#)[About R](#)[Logo](#)[Contributors](#)[What's New?](#)[Reporting](#)

The R Project for Statistical Computing

Getting Started

統計計算

R is a free software environment for statistical computing and graphics. It compiles and runs on a wide variety of UNIX platforms, Windows and MacOS. To [download R](#), please choose your preferred CRAN mirror.

If you have questions about R like how to download and install the software, or what the license terms are, please read our answers to [frequently asked questions](#) before you send an email.

..

R-下載

- 官網: <http://www.r-project.org/>
- 選取左側 Download \ CRAN
- 選取 Taiwan CRAN

Taiwan
<https://cran.csie.ntu.edu.tw/>

- 選取 Download R for Windows

- [Download R for Linux](#)
- [Download R for \(Mac\) OS X](#)
- [Download R for Windows](#)



R-下載 (續)

- 選取 base → 下載 [R-4.1.2-win.exe]



- R安裝路徑: 保留原路徑,不要修改
- https://github.com/rwepa/DataDemo/blob/master/windows_intall_R.pdf

Rtools 下載並安裝

- Rtools for Windows: 一定要保留原路徑 C:\rtools40



On Windows 64-bit: [rtools40v2-x86_64.exe](#)

- Save version information to registry
- Create start-menu icons to msys2 shells

2個都打勾

✓ 安裝 R

✓ 安裝 Rtools



R Manuals

The R Manuals

edited by the R Development Core Team.

The following manuals for R were created on Debian Linux and may differ from the manuals for Mac or Windows on platform-specific pages, but most parts version of the manuals for each platform are part of the respective R installations. The manuals change with R, hence we provide versions for the most recent version for the patched release version (R-patched) and finally a version for the forthcoming R version that is still in development (R-devel).

Here they can be downloaded as PDF files, EPUB files, or directly browsed as HTML:

Manual	R-release	R-patched
An Introduction to R is based on the former "Notes on R", gives an introduction to the language and how to use R for doing statistical analysis and graphics.	HTML PDF EPUB	HTML PDF EPUB
R Data Import/Export describes the import and export facilities available either in R itself or via packages which are available from CRAN.	HTML PDF EPUB	HTML PDF EPUB
R Installation and Administration	HTML PDF EPUB	HTML PDF EPUB
Writing R Extensions covers how to create your own packages, write R help files, and the foreign language (C, C++, Fortran, ...) interfaces.	HTML PDF EPUB	HTML PDF EPUB
A draft of The R language definition documents the language <i>per se</i> . That is, the objects that it works on, and the details of the expression evaluation process, which are useful to know when programming R functions.		
R Internals : a guide to the internal structures of R and coding standards the core team working on R itself.		
The R Reference Index : contains all help files of the R standard and recommended packages in printable form. (9MB, approx. 3500 pages)		

contributed documentation
(貢獻文件, 免費啦)

Translations of manuals into other languages than English are available from the [contributed documentation](#) section (only a few translations are available).

R Manuals (續)

Contributed Documentation

[English](#) --- [Other Languages](#)

Manuals, tutorials, etc. provided by users of R. The R core team does not take any responsibility for contents, but we appreciate the effort very much and encourage everybody to contribute to this list! To submit, follow the submission instructions on the [CRAN main page](#). All material below is available directly from CRAN, you may also want to look at the list of [other R documentation](#) available on the Internet.

Note: Please use the [directory listing](#) to sort by name, size or date (e.g., to see which documents have been updated lately).

English Documents

Documents with more than 100 pages:

- “**Visual Statistics. Use R!**” by Alexey Shipunov ([PDF](#), 2016-06-06, 301 pages) are accessible from [Alexey Shipunov's English R page](#).
- “**Using R for Data Analysis and Graphics - Introduction, Examples and Commentary**” by John Maindonald ([PDF](#), data sets and scripts are available at [JM's homepage](#)).
- “**Practical Regression and Anova using R**” by Julian Faraway ([PDF](#), data sets and scripts are available at the [book homepage](#)).

好書!



R 執行畫面

實作練習16

The screenshot shows the R GUI interface. The title bar says "R Gui (64-bit)". The menu bar includes "檔案", "命令歷程", "重設大小", and "視窗". Below the menu is a toolbar with icons for file operations. The main window is titled "R Console". It displays the R startup message, which includes:

```
R version 3.0.3 (2014-03-06) -->
Copyright (C) 2014 The R Foundation for Statistical Computing
Platform: x86_64-w64-mingw32/x64

R 是免費軟體，不提供任何擔保。  
在某些條件下您可以將其自由散佈  
用 'license()' 或 'licence()'

R 是個合作計劃，有許多人為之効  
用 'contributors()' 來看詳細的  
用 'citation()' 會告訴您如何在  
用 'demo()' 來看一些示範程式，  
用 'help.start()' 透過 HTML 準  
用 'q()' 離開 R。
```

Below the message, there is a command prompt:
> plot(runif(100), type="l")

A yellow callout bubble points to the text in the R message area with the text: "安裝R, 登入名稱: 1. 不要使用空格 2. 不要使用中文字型".

Below the console, three code snippets are shown in boxes:

- plot(runif(100), type="l", main= "R大數據分析")
- demo(graphics)
- demo(persp)

A yellow callout bubble points to the "demo(graphics)" snippet with the text: "大小寫須一致".

R 功能表

檔案



編輯



現行目錄 getwd()

輔助



儲存控制台-文字檔



新增R檔案練習

實作練習17

RGui (64-bit)

檔案 命令歷程 重設大小 視窗

C:\mydata\myfirst.R - R 編輯器
2021.08.10

```
plot(runif(10), type="b", main= "R大數據分析")
x <- rnorm(10)
x
```

● 步驟1: 選取程式碼

● 步驟2: 按 Run line or selection 或 Ctrl + R

● 步驟3: 按 File \ Save

R Graphics: Device 2 (ACTIVE)

R大數據分析

用 'license()' 或 'licence()' 來獲得散佈的詳細條件。

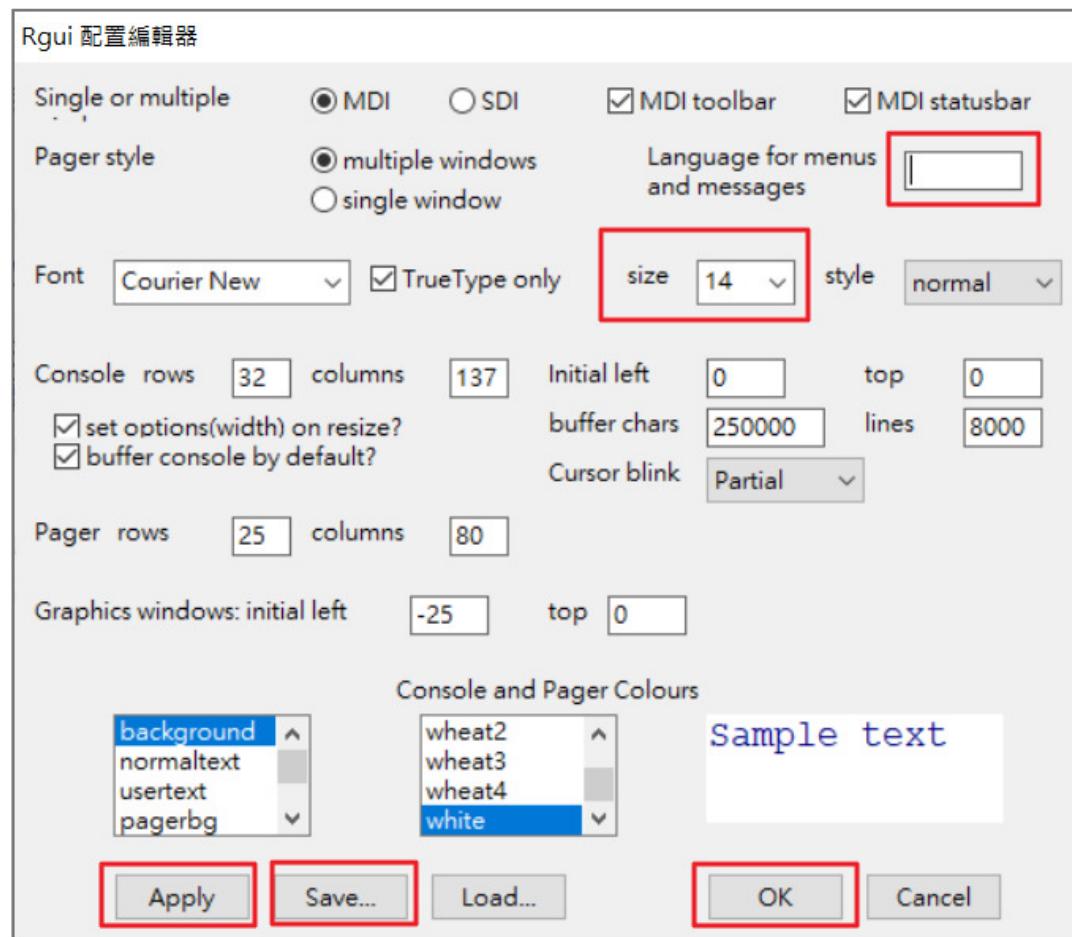
R 是個協作計劃，有許多人為之做出了貢獻。
用 'contributors()' 來看詳細的情況以及
用 'citation()' 會告訴您如何在出版品中正確地參照 R 或 R 套件。

用 'demo()' 來看一些示範程式，用 'help()' 來檢視線上輔助檔案，或
用 'help.start()' 透過 HTML 瀏覽器來看輔助檔案。
用 'q()' 離開 R。

```
> # 2021.08.10
> plot(runif(10), type="b", main= "R大數據分析")
> x <- rnorm(10)
> x
[1]  0.94387445 -1.51643566 -0.63187431 -1.40512487  0.10853098
[6] -0.37268350 -0.54767406 -0.39325153  0.01232859 -1.22981462
> |
```

R version 4.1.0 Patched (2021-07-16 r80639)

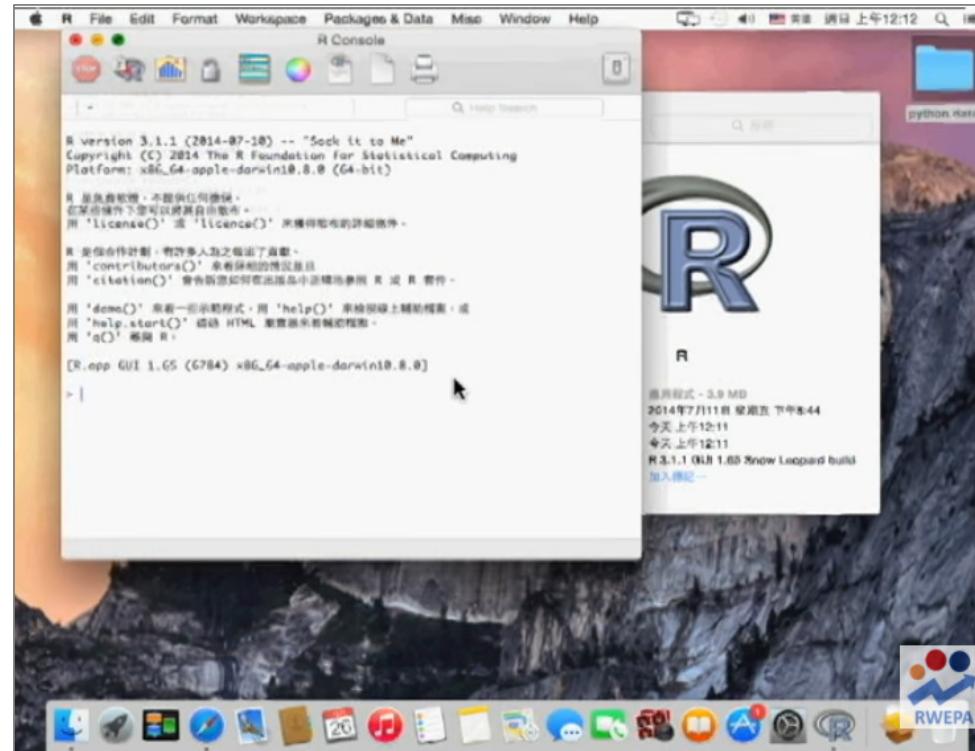
編輯 \ GUI 偏好設定



- Language: en 英文
- size: 字型大小

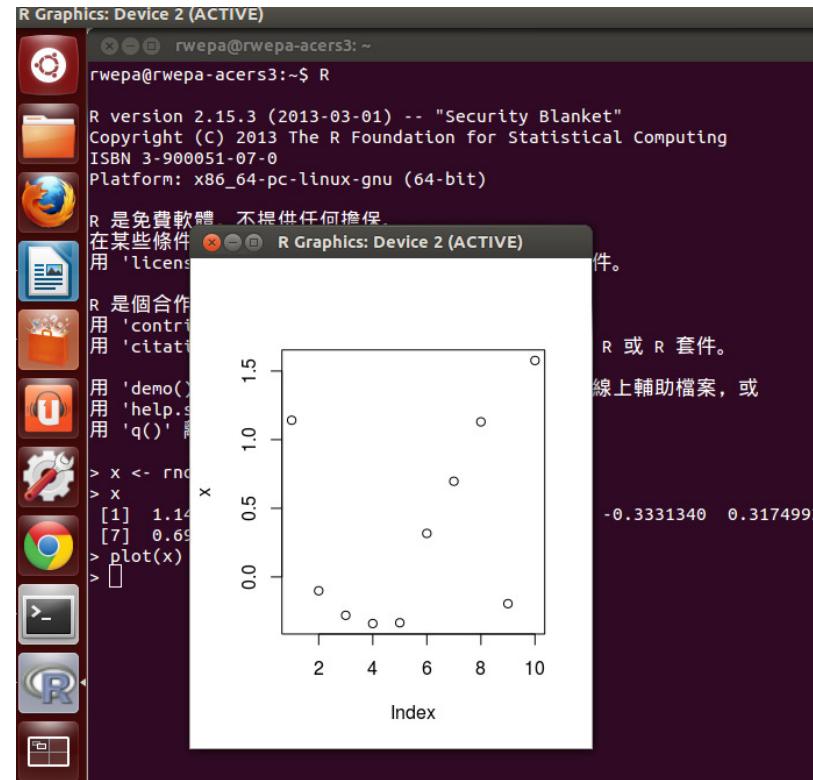
R for Mac

- <https://youtu.be/72MYRBNo5Bk>



R for Ubuntu

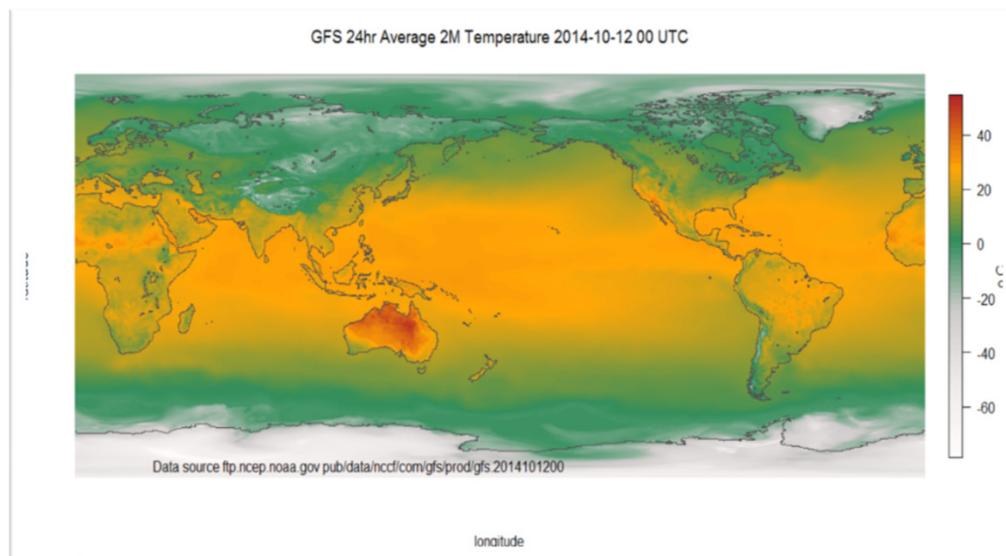
- <http://rwepa.blogspot.com/2013/05/ubuntu-r.html>



RStudio 安裝與簡介

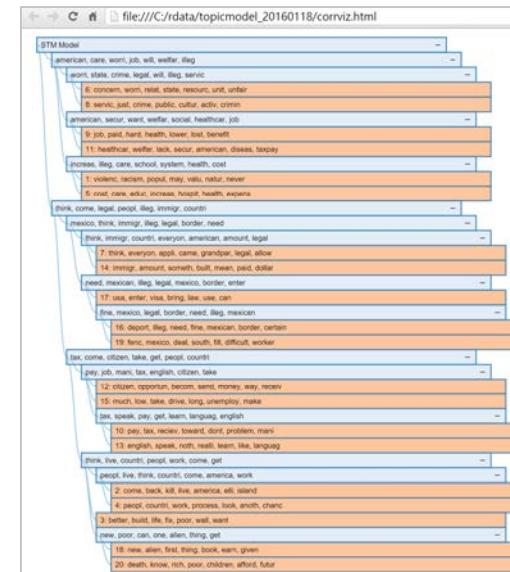
整合式開發環境 - RStudio

- <http://www.rstudio.com/>



視覺化應用

(全球2M氣溫圖)



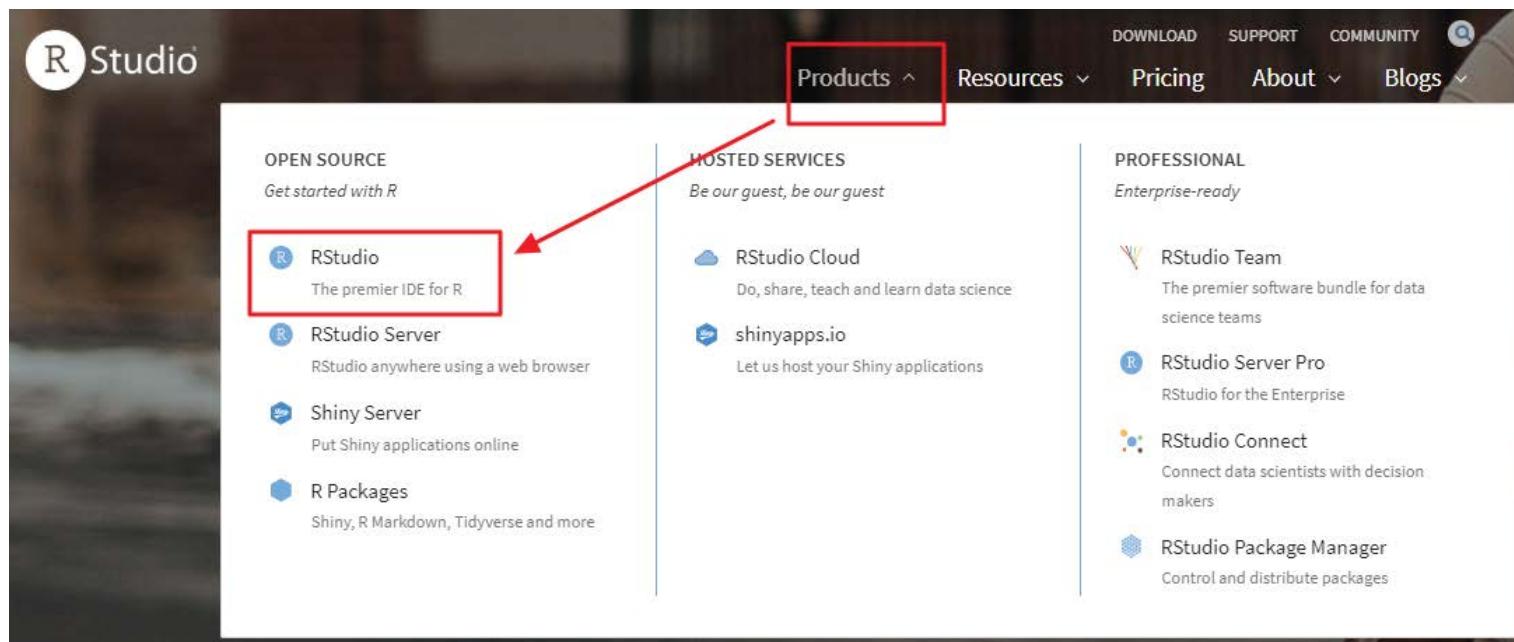
主題模型

RStudio – 特性

- 支援智慧輸入 (按Tab)
- 高亮度顯示程式碼
- 整合R程式, 控制台, 變數清單, 繪圖視窗
- 整合資料庫匯入 SQL, Spark
- 整合R套件: shiny, rmarkdown
- 安裝注意:
 - 先安裝R, 再安裝 RStudio
 - 安裝 RStudio時, 請先關閉R

RStudio 下載

- <http://www.rstudio.com/>



RStudio 下載 (續)

RStudio Desktop Open Source License Free	RStudio Desktop Commercial License \$995	RStudio Server Open Source License Free	RStudio Server Pro Commercial License \$4,975
---	---	--	--

單機版 /year
伺服器版本 /year
(Named Users)

免費版



Integrated Tools for R
Priority Support
Access via Web Browser
Enterprise Security
Project Sharing

Learn more  

Learn more  

Evaluation | Learn more

RStudio 下載 (續)

RStudio Desktop 2021.09.0+351 - [Release Notes](#)

- 1.** Install R. RStudio requires R 3.0.1+.
- 2.** Download RStudio Desktop. Recommended for your system:



Requires Windows 10 (64-bit)

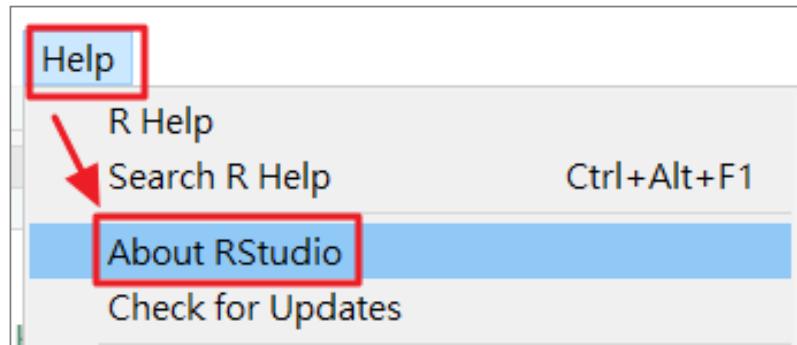
RStudio-2021.09.0-351.exe



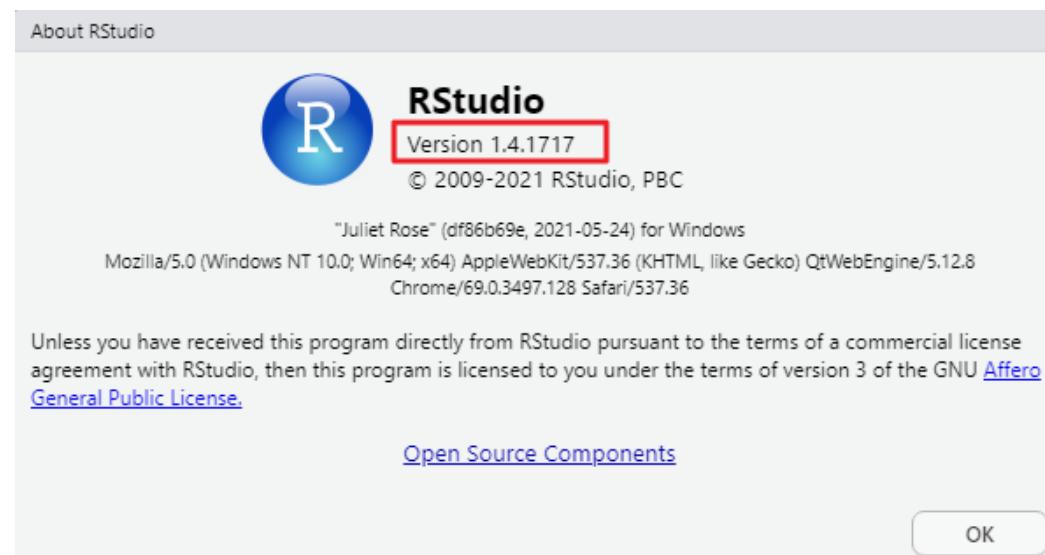
RStudio 安裝



RStudio 版本訊息

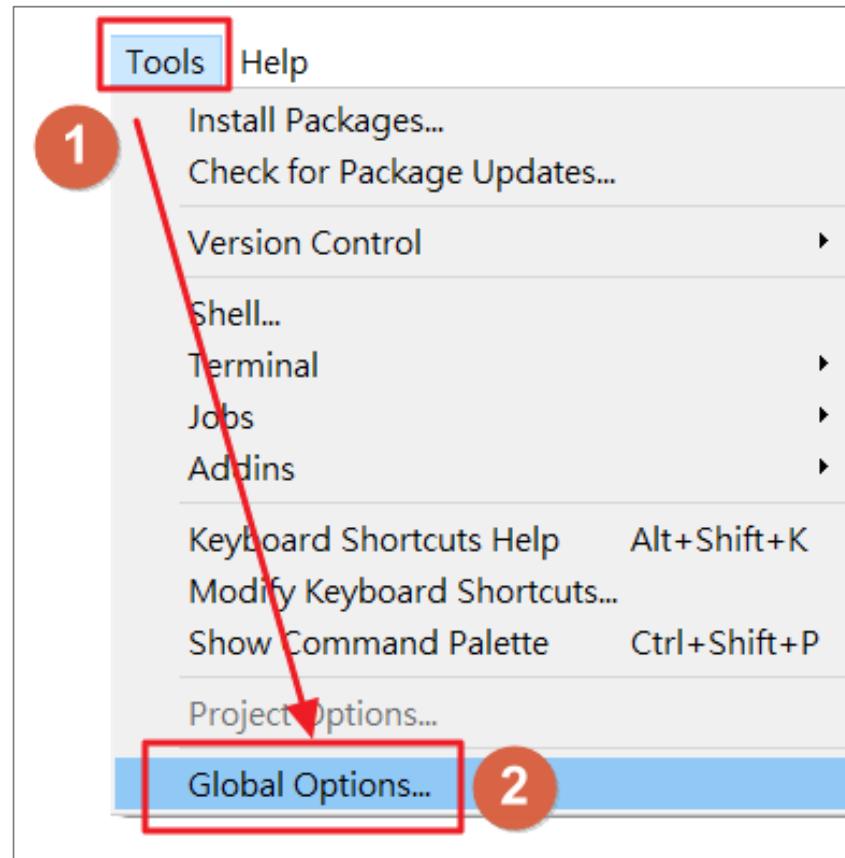


Help \ About RStudio

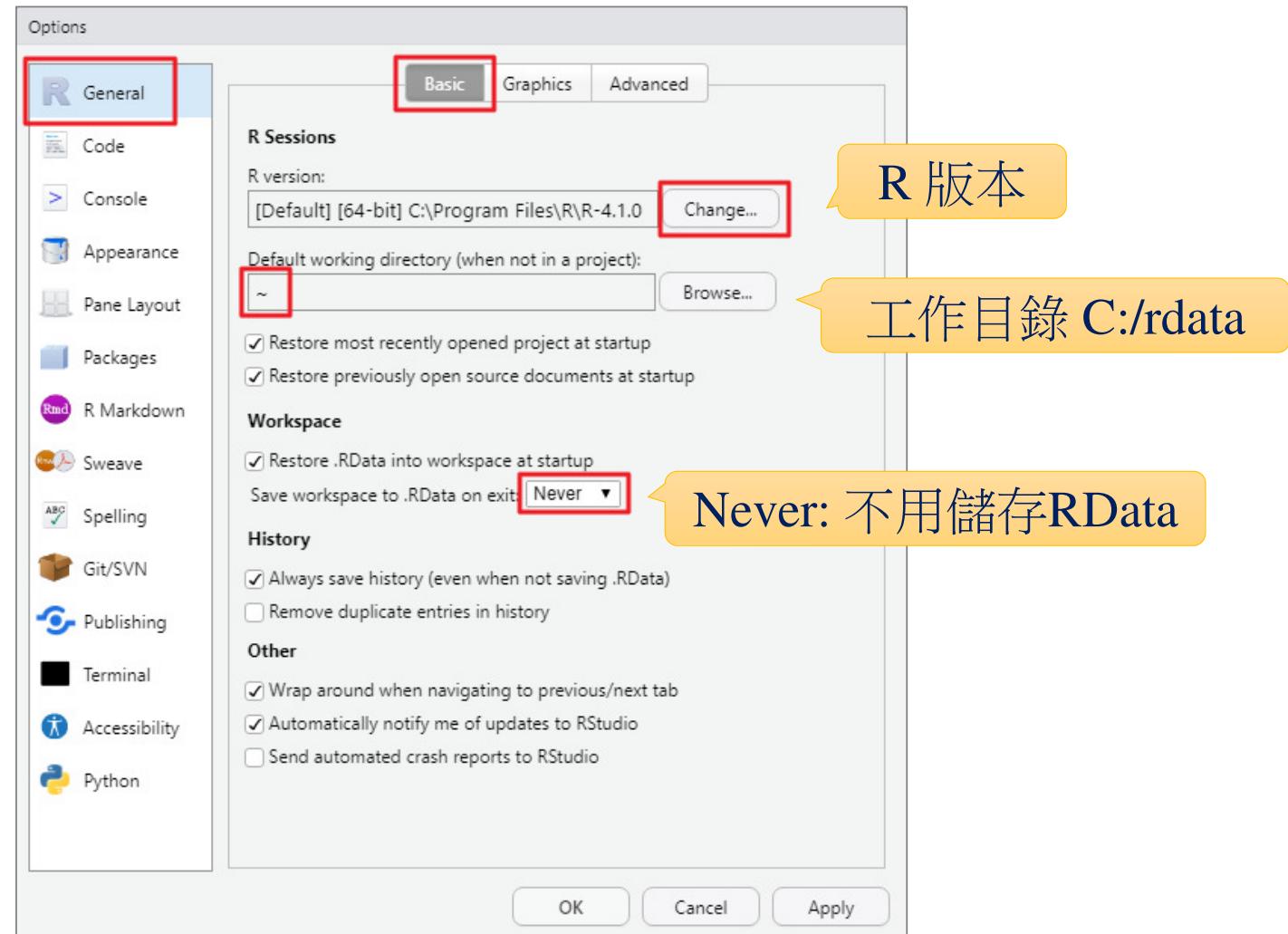


RStudio-選項設定

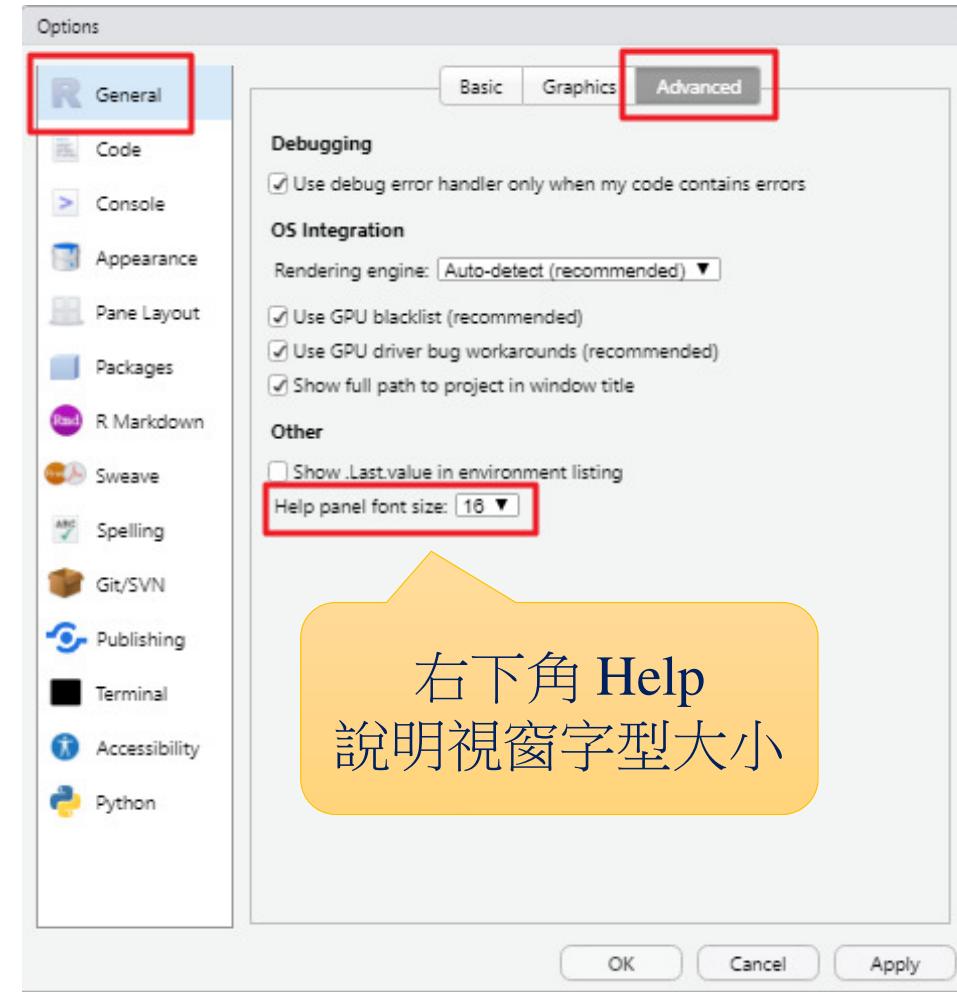
- Tools \ Global Options



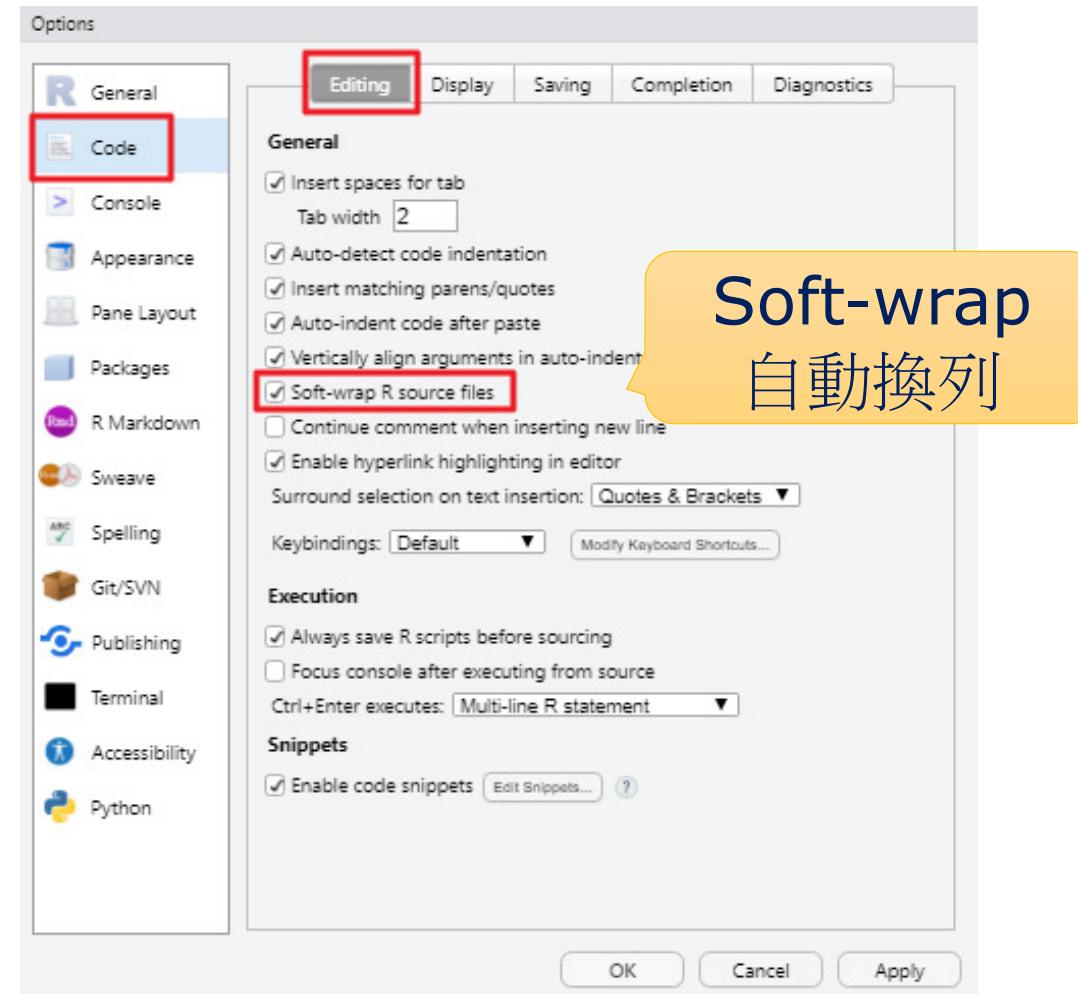
General \ Basic



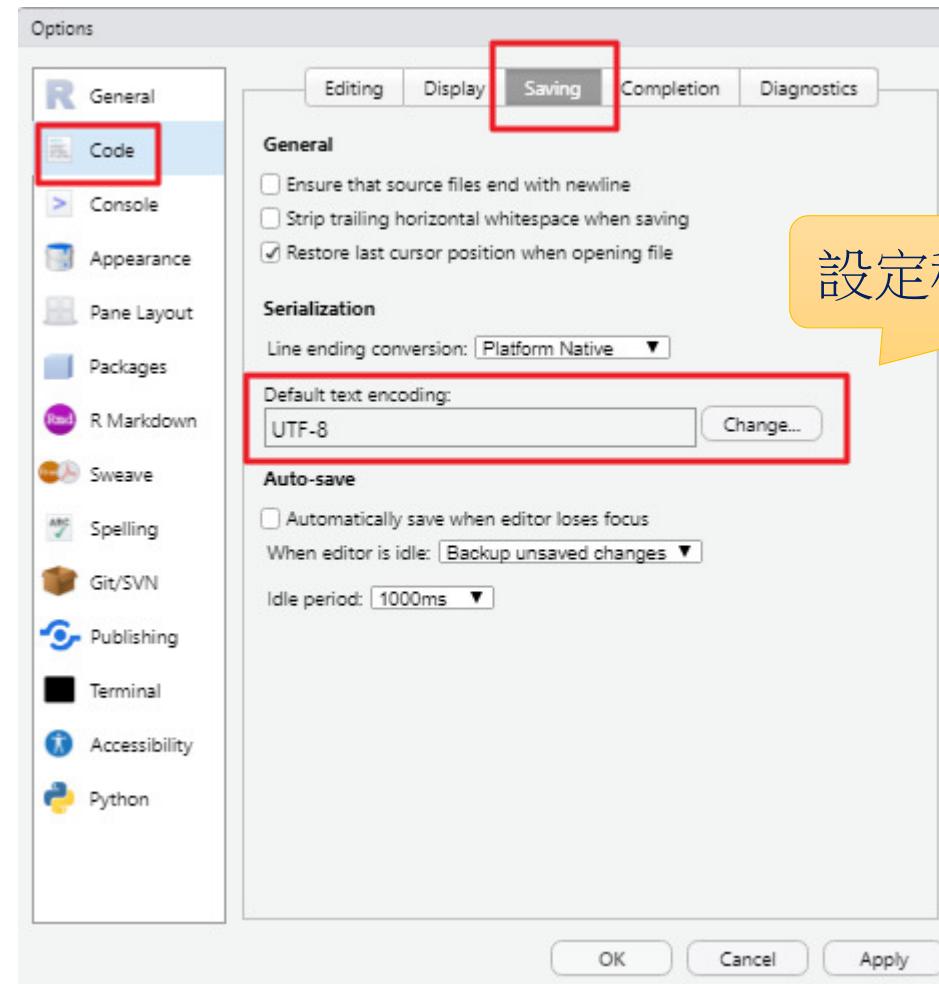
General \ Advanced



Code \ Editing



Code \ Saving

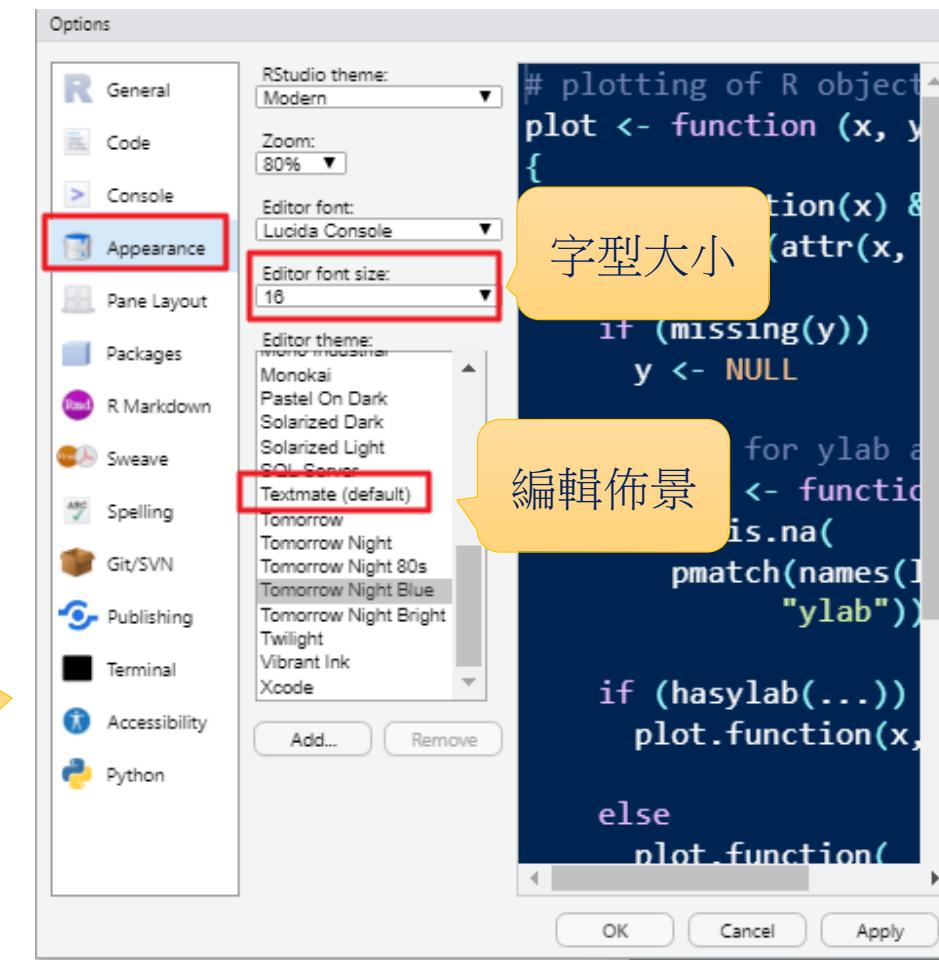


設定程式編碼 UTF-8

RStudio-選項設定(續)

- Appearance \ Editor theme
- 預設值:
TextMate

設定完成,須重
新啟動RStudio



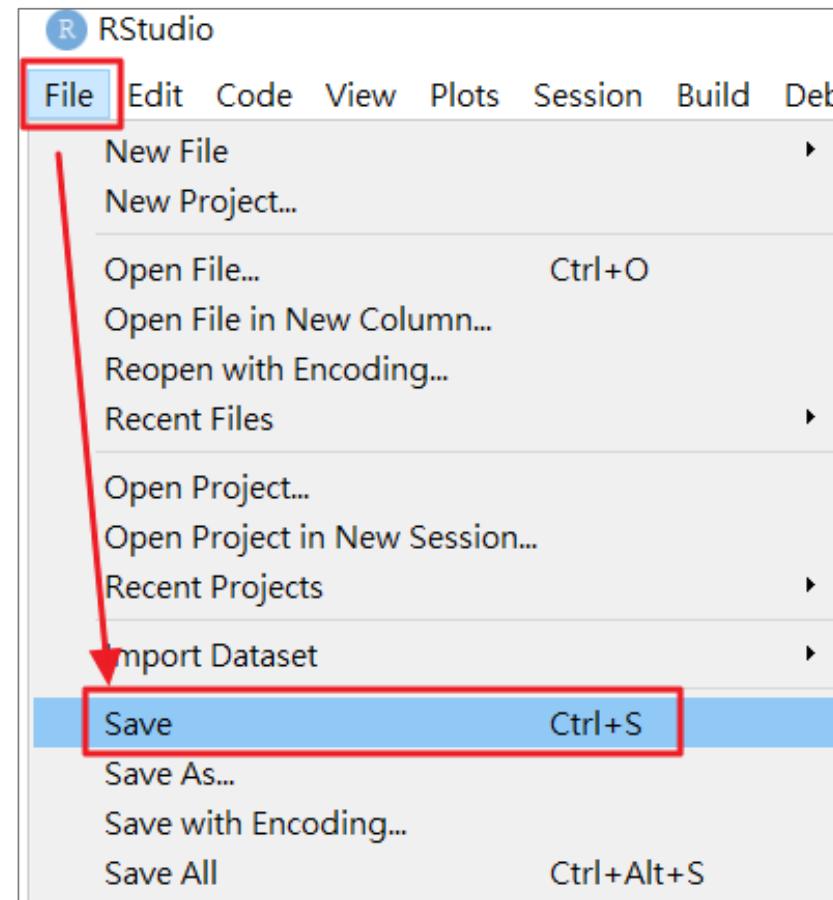
開啟檔案

- File \ Open File

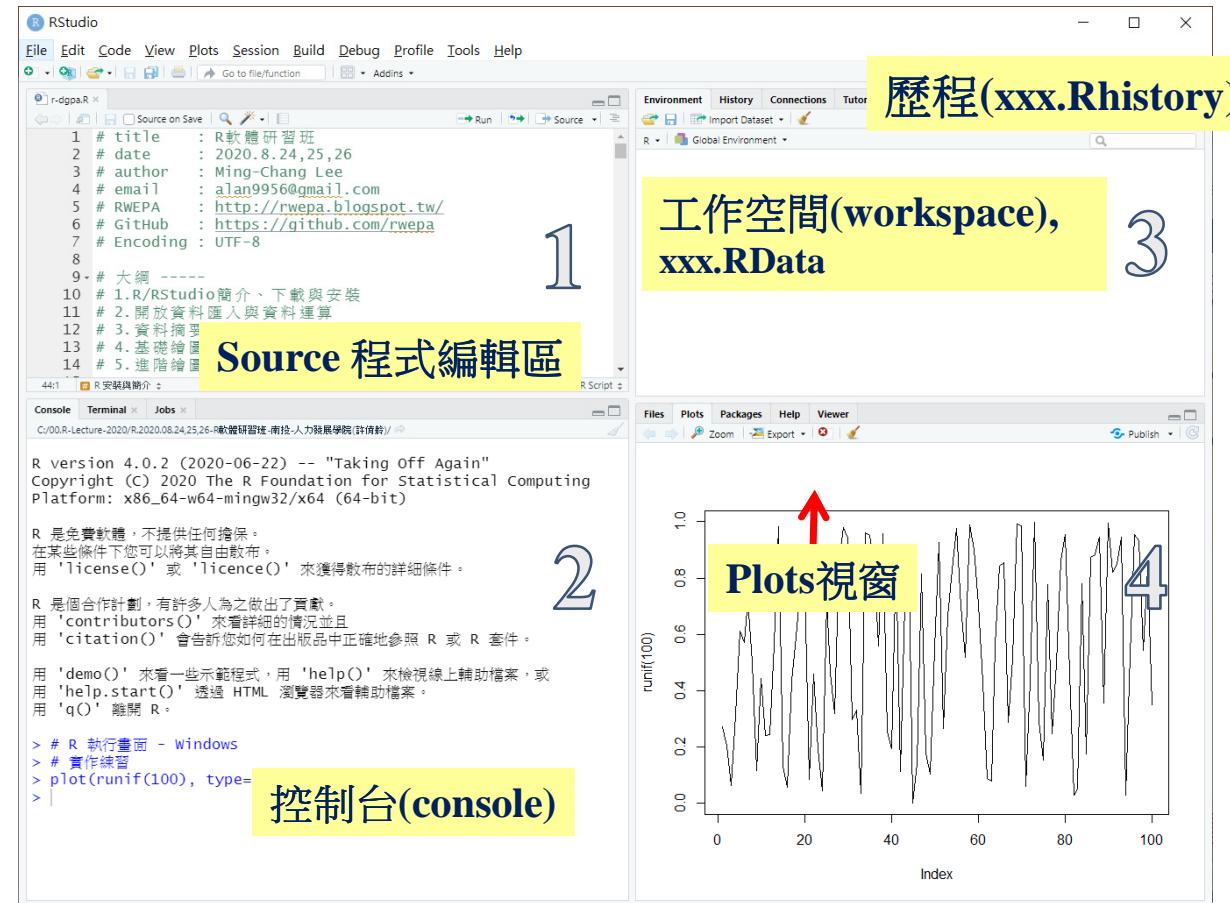


儲存檔案

- File \ Save
(CTRL + S)



R/RStudio環境的基礎觀念



Ctrl + Shift + F10: 重新啟動R

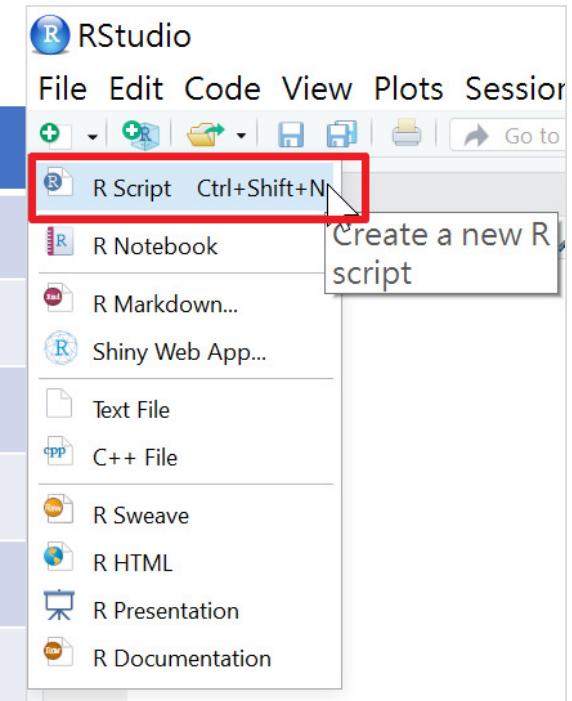
R + Editor

- R – 原生環境
- RStudio – IDE 整合介面
- Eclipse
 - StatET 4.x.0:
An Eclipse based IDE (integrated development environment) plug-in for R.
 - <https://projects.eclipse.org/projects/science.statet>
- 如何安裝 Visual Studio R 工具
 - R Tools for Visual Studio



RStudio 快速鍵

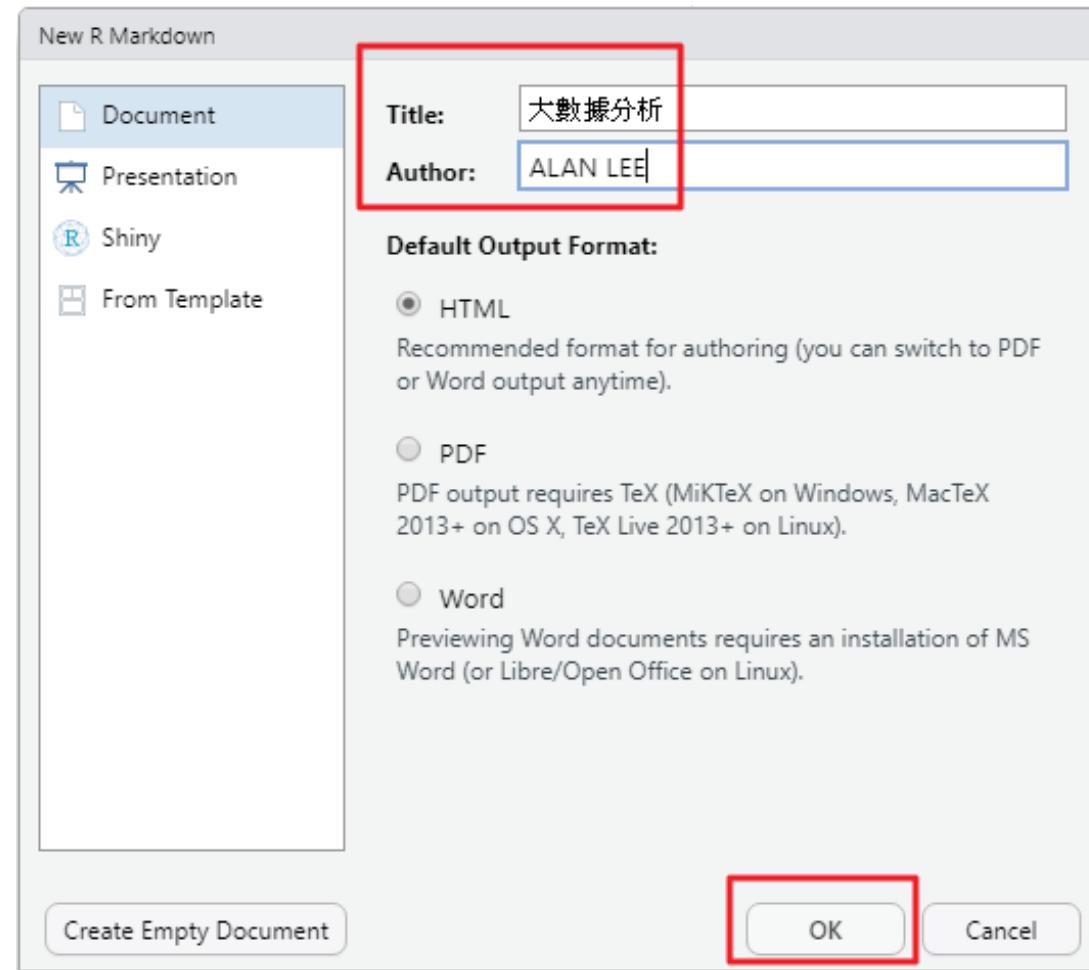
快速鍵	功能
Ctrl + Shift + N	建立新的R程式
Ctrl + S	儲存檔案
Ctrl + Shift + R	建立章節 (-----)
Alt + -	指派符號
Ctrl + Shift + C	註解
Ctrl + Enter	執行程式
Ctrl + Shift + F10	重新啟動R
Alt + Shift + K	快速鍵總表 (Esc 退出)



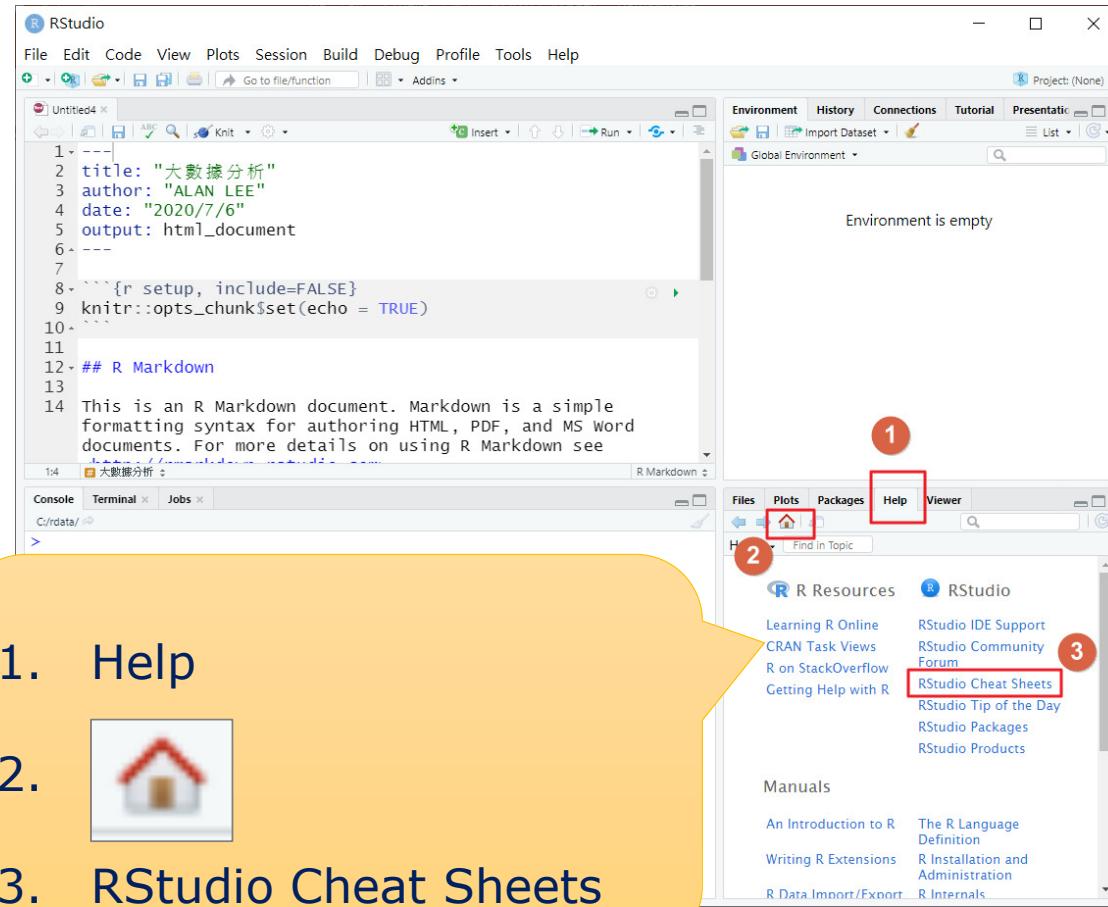
R Markdown

(R 標記語言)

RStudio - Markdown



RStudio - Markdown (續)



The screenshot shows the RStudio interface. On the left, a code editor displays an R Markdown document with the following content:

```
1 ---  
2 title: "大數據分析"  
3 author: "ALAN LEE"  
4 date: "2020/7/6"  
5 output: html_document  
6 ---  
7  
8 ````{r setup, include=FALSE}  
9 knitr::opts_chunk$set(echo = TRUE)  
10 ````  
11  
12 ## R Markdown  
13  
14 This is an R Markdown document. Markdown is a simple  
formatting syntax for authoring HTML, PDF, and MS Word  
documents. For more details on using R Markdown see  
15
```

The right side of the interface shows the Environment pane with the message "Environment is empty". A yellow callout box on the left contains three numbered steps:

1. Help
2. 
3. RStudio Cheat Sheets

Red boxes and numbers highlight specific elements in the RStudio interface:

1. A red circle with the number 1 points to the "Help" tab in the top navigation bar.
2. A red box with the number 2 points to the "Help" icon in the top navigation bar.
3. A red box with the number 3 points to the "RStudio Cheat Sheets" link in the "Help" sidebar.

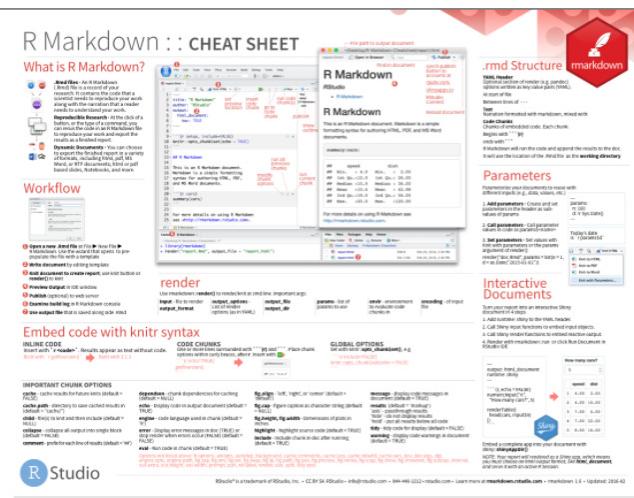
R Markdown Cheatsheet 線上說明

- <https://www.rstudio.com/resources/cheatsheets/>

R Markdown Cheatsheet

R Markdown is an authoring format that makes it easy to write reusable reports with R. You combine your R code with narration written in markdown (an easy-to-write plain text format) and then export the results as an html, pdf, or Word file. You can even use R Markdown to build interactive documents and slideshows. Updated February 16. ([Old Version](#).)

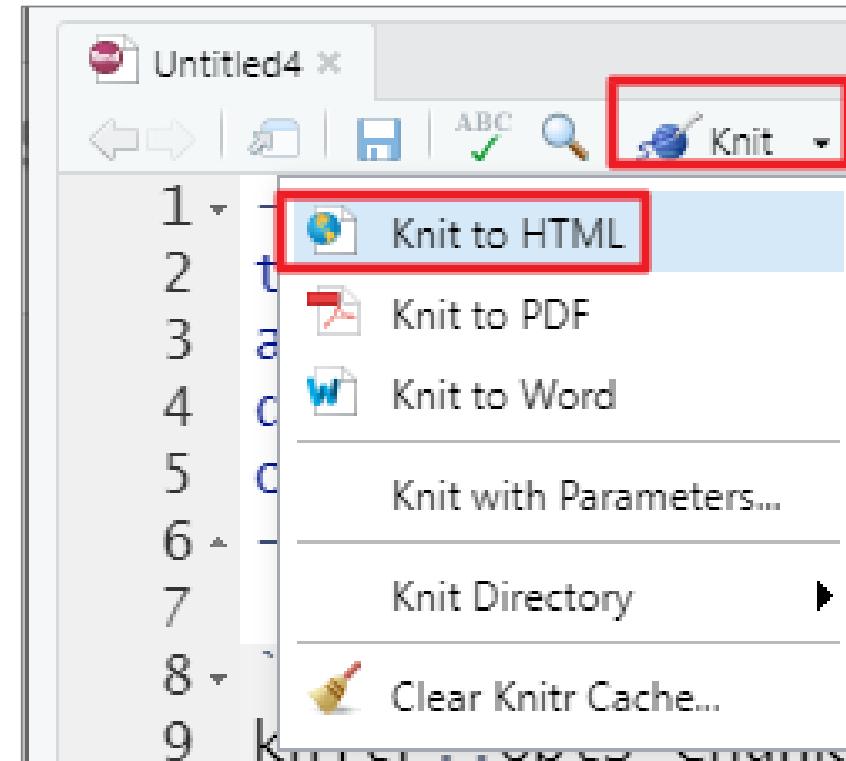
[DOWNLOAD](#)



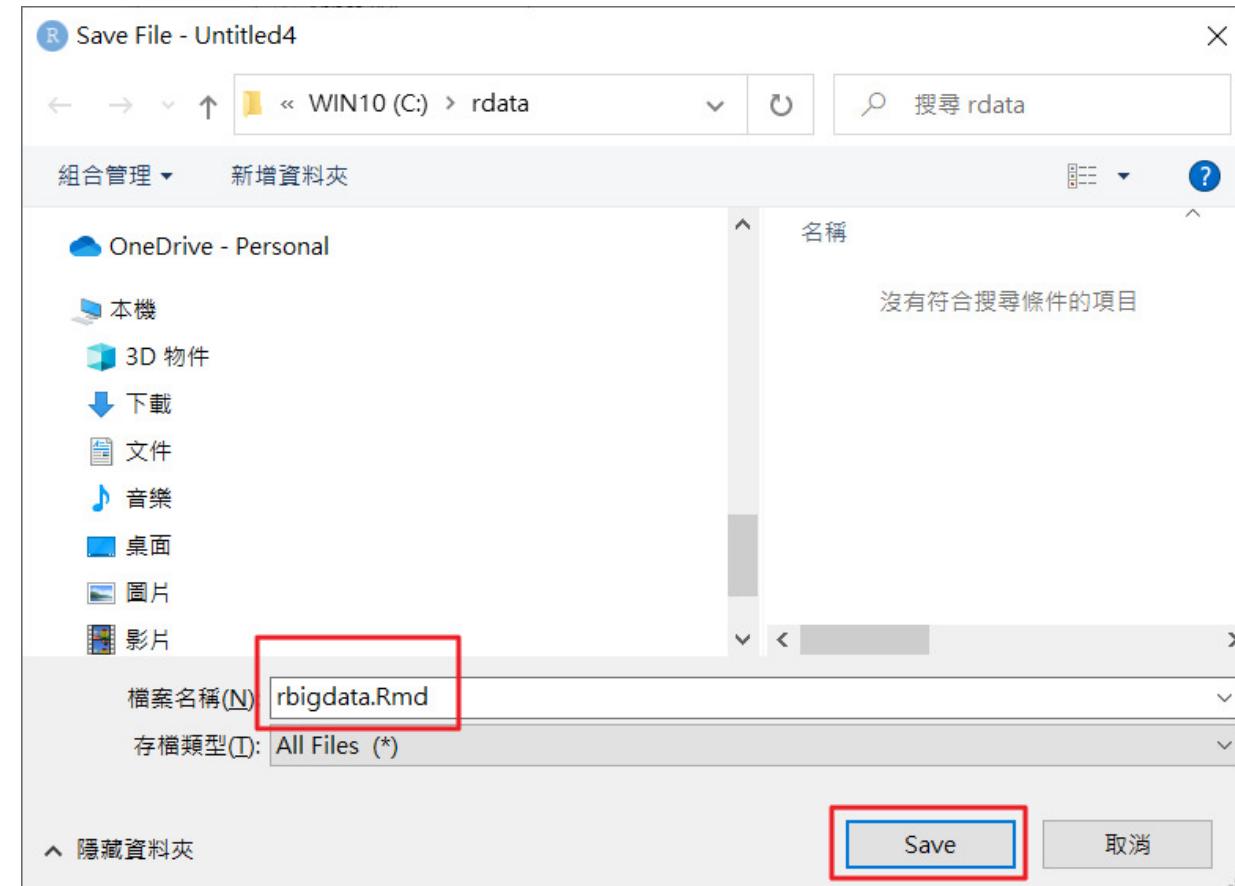
The screenshot shows the R Markdown Cheatsheet page. It includes a large central image of the cheatsheet itself, which is divided into several sections: 'Workflow' (describing the process from a new R Markdown file to various output formats like HTML, PDF, and Word), 'render' (options for rendering code blocks), 'Parameters' (for specifying parameters in R code), and 'Interactive Documents' (for creating shiny apps). Below the main image are three smaller sections: 'IMPORTANT CHUNK OPTIONS' (like echo, eval, and comment), 'CODE CHUNKS' (like tidy, eval, and comment), and 'GLOBAL OPTIONS' (like quiet, encoding, and error). At the bottom, there's a note about RStudio's support for R Markdown and a link to the full document.

RStudio - Markdown (續)

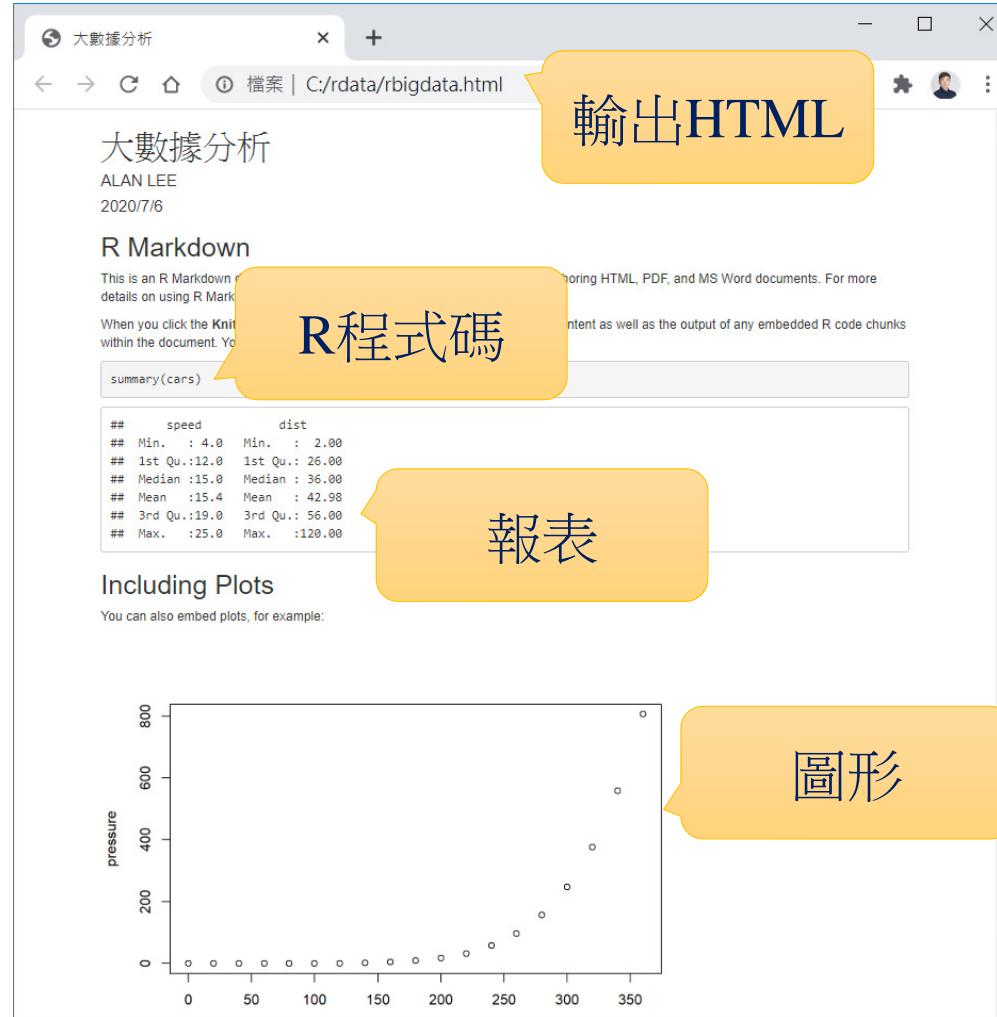
- Knit HTML
- Knit PDF
- Knit Word



RStudio - Markdown (續)



RStudio - Markdown : HTML



The screenshot shows an RStudio window displaying an R Markdown document. The title is "大數據分析" by ALAN LEE, dated 2020/7/6. The content includes an R code chunk for `summary(cars)` which outputs a table of car speeds and distances. Below this is a section titled "Including Plots" with a note about embedding plots. A scatter plot of pressure vs speed is shown at the bottom. Four yellow callout bubbles highlight different features: "輸出HTML" points to the title and author information; "R程式碼" points to the R code chunk; "報表" points to the table output; and "圖形" points to the scatter plot.

大數據分析
ALAN LEE
2020/7/6

R Markdown

This is an R Markdown document. You can render this document as a web page, PDF, or Microsoft Word document. For more details on using R Markdown, see the [R Markdown documentation](#).

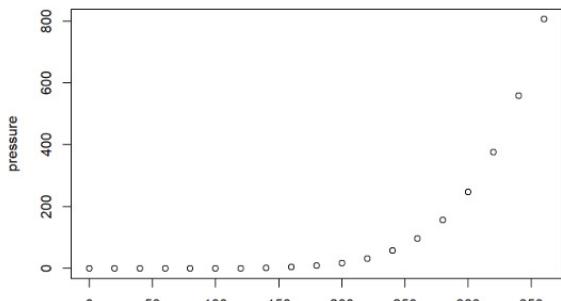
When you click the Knit button, a rendering process will convert your R Markdown document into a presentation slide, a PDF, or a Microsoft Word document, depending on what you have specified. The resulting document will contain both the content of the document as well as the output of any embedded R code chunks.

```
summary(cars)
```

	speed	dist
## Min.	4.0	2.00
## 1st Qu.	12.0	1st Qu.: 26.00
## Median	15.0	Median : 36.00
## Mean	15.4	Mean : 42.98
## 3rd Qu.	19.0	3rd Qu.: 56.00
## Max.	25.0	Max. :120.00

Including Plots

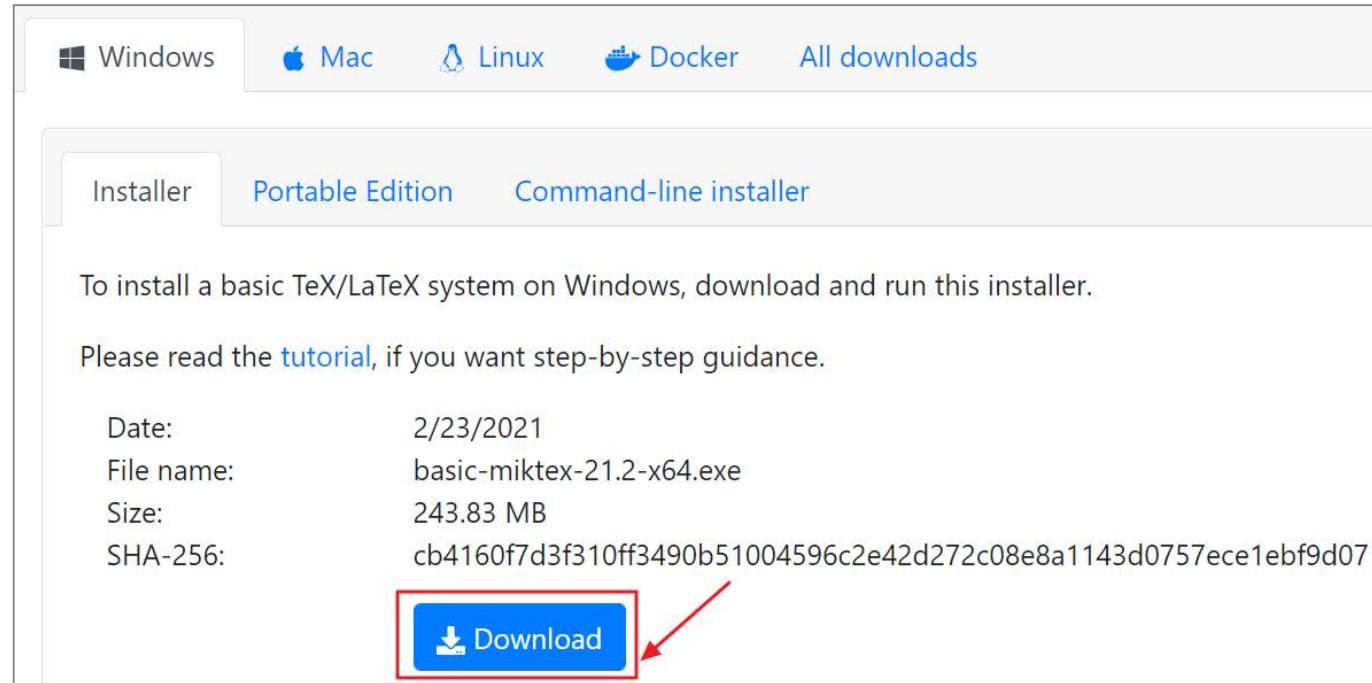
You can also embed plots, for example:



A scatter plot showing the relationship between pressure (y-axis, 0 to 800) and speed (x-axis, 0 to 350). The data points show a positive correlation, with most points clustered below 200 speed and 400 pressure, and a few outliers reaching up to 350 speed and 800 pressure.

RStudio - Markdown : PDF

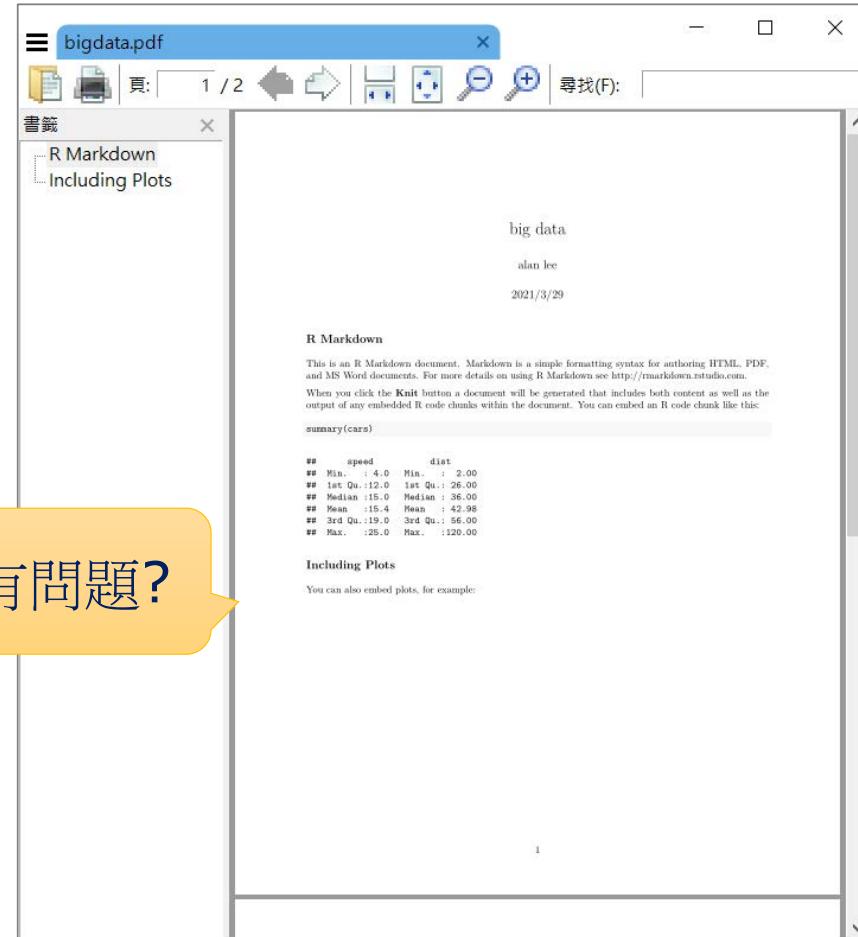
- 下載 Miktex: <https://miktex.org/download>
- basic-miktex-21.2-x64.exe (243.83MB)



Knit to PDF



中文可能有問題?



The screenshot shows a PDF viewer window titled 'bigdata.pdf'. The PDF contains the following text and code:

big data
alan lee
2021/3/29

R Markdown

This is an R Markdown document. Markdown is a simple formatting syntax for authoring HTML, PDF, and MS Word documents. For more details on using R Markdown see <http://rmarkdown.rstudio.com>.

When you click the Knit button a document will be generated that includes both content as well as the output of any embedded R code chunks within the document. You can embed an R code chunk like this:

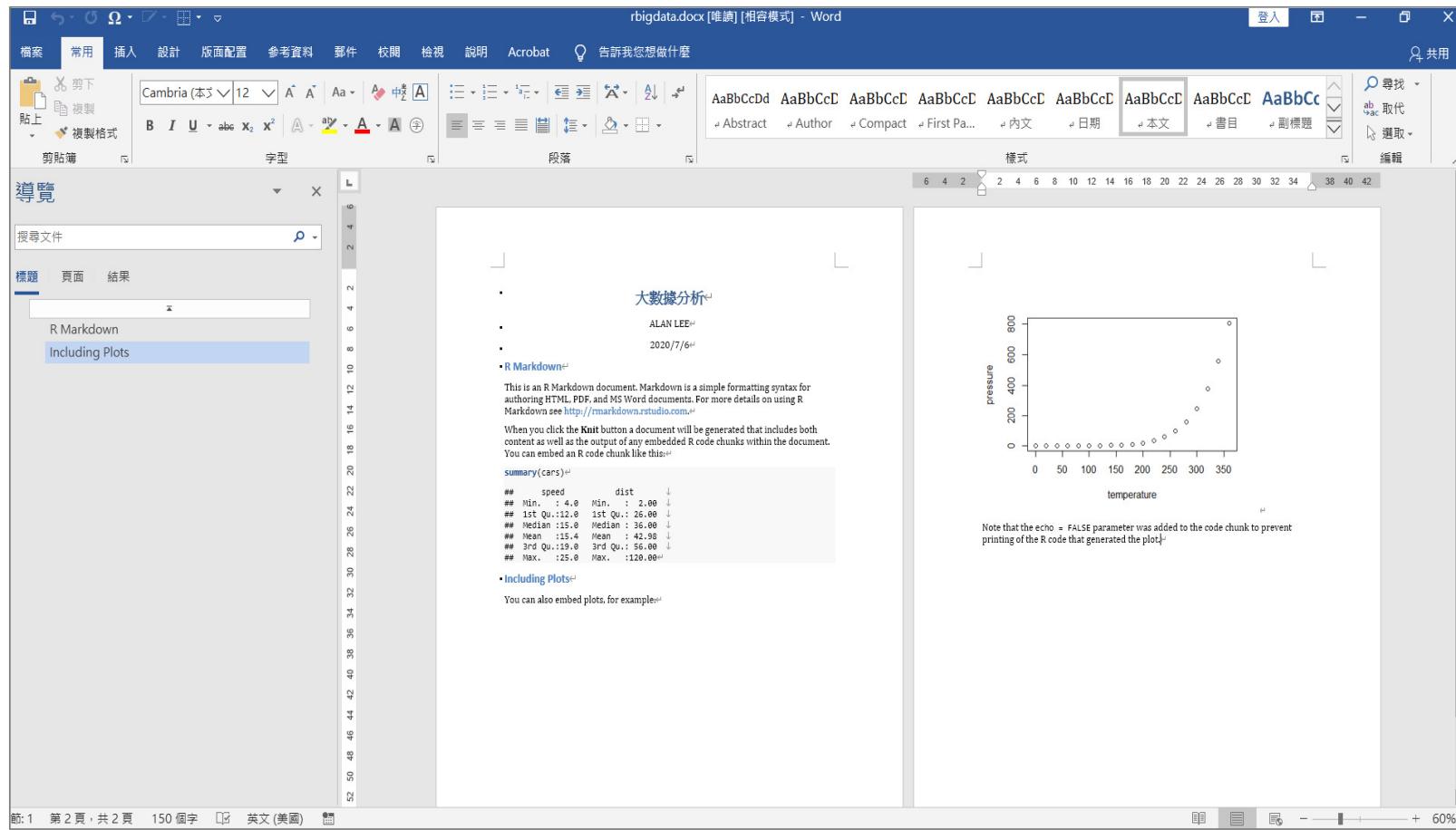
```
summary(cars)
```

speed dist
Min. : 4.0 Min. : 2.00
1st Qu.:12.0 1st Qu.: 26.00
Median :15.0 Median : 36.00
Mean :15.4 Mean : 42.98
3rd Qu.:19.0 3rd Qu.:56.00
Max. :28.0 Max. :120.00

Including Plots

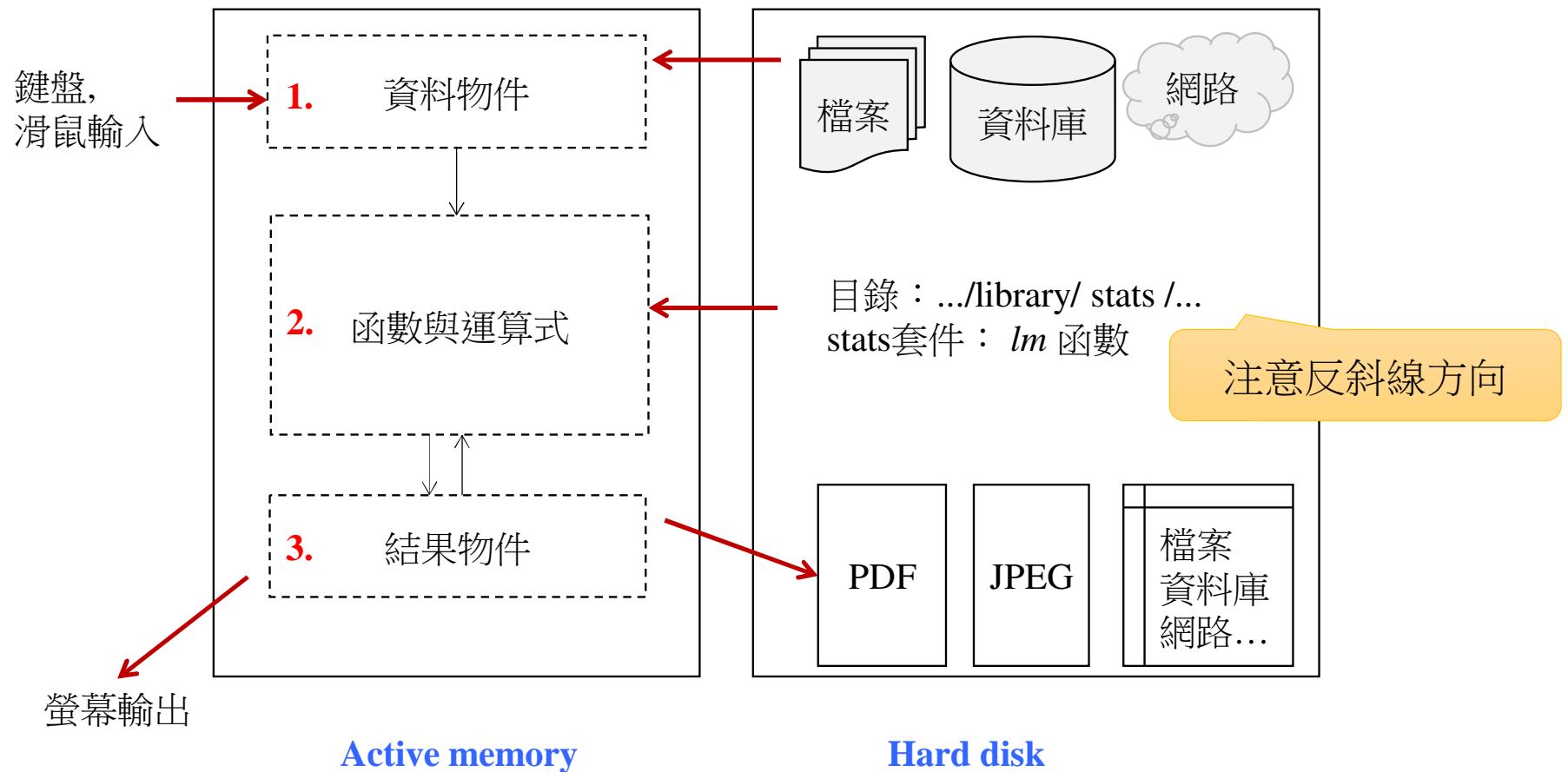
You can also embed plots, for example:

RStudio - Markdown 轉換為 Word



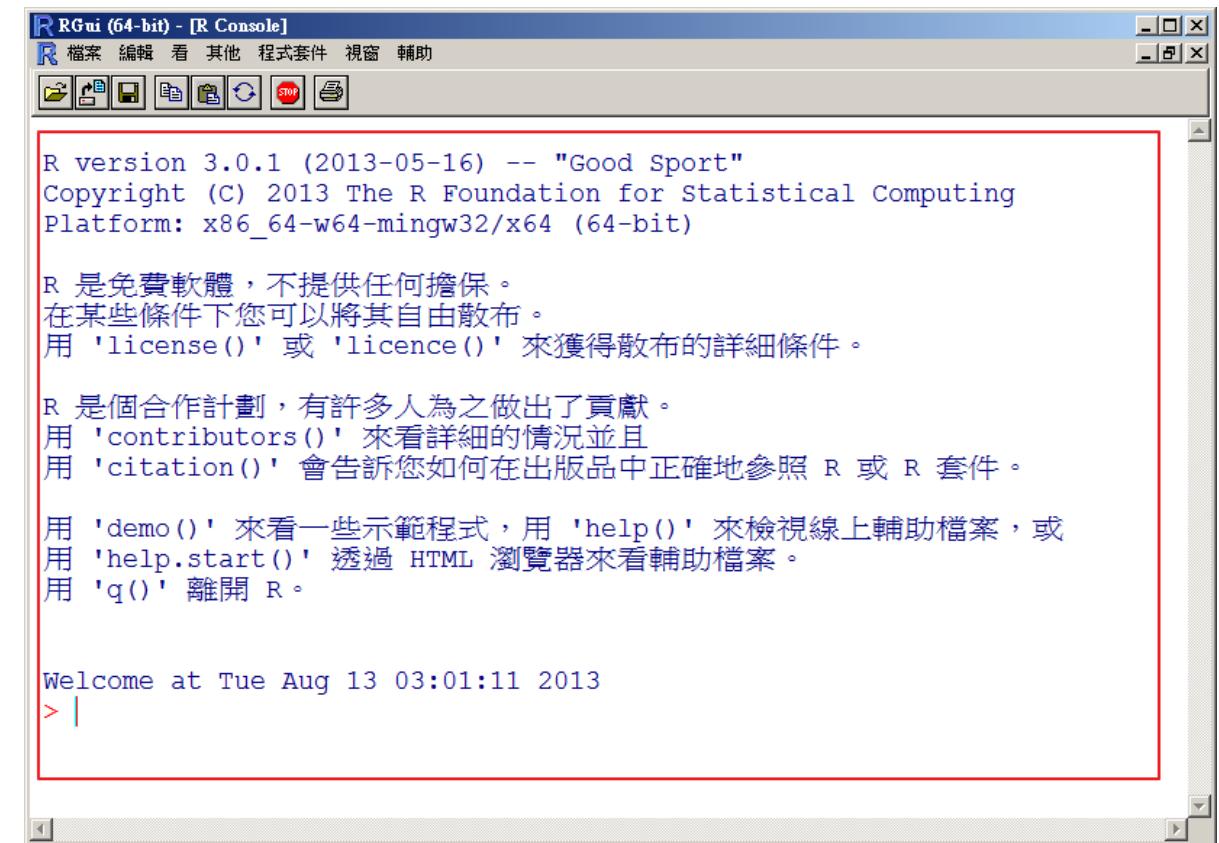
R使用環境

R運作方式



基本觀念

- 控制台(console)
- 歷程
 - xxx.Rhistory
- 套件(package)
- 工作空間(workspace)
 - xxx.RData
- 物件(object)



```
R Gui (64-bit) - [R Console]
R 檔案 編輯 看 其他 程式套件 視窗 輔助
R version 3.0.1 (2013-05-16) -- "Good Sport"
Copyright (C) 2013 The R Foundation for Statistical Computing
Platform: x86_64-w64-mingw32/x64 (64-bit)

R 是免費軟體，不提供任何擔保。
在某些條件下您可以將其自由散布。
用 'license()' 或 'licence()' 來獲得散布的詳細條件。

R 是個合作計劃，有許多人為之做出了貢獻。
用 'contributors()' 來看詳細的情況並且
用 'citation()' 會告訴您如何在出版品中正確地參照 R 或 R 套件。

用 'demo()' 來看一些示範程式，用 'help()' 來檢視線上輔助檔案，或
用 'help.start()' 透過 HTML 瀏覽器來看輔助檔案。
用 'q()' 離開 R。

Welcome at Tue Aug 13 03:01:11 2013
> |
```

控制台的特定符號

- 命令提示字元(大於) > (等待使用者輸入資料)
- 指令未完提示字元(加號) + (表示尚未輸入完成)
- 註解提示字元(井字號) # (不會編譯註解)
- 結果行列顯示編號

```
> iris$Sepal.Length  
[1] 5.1 4.9 4.7 4.6 5.0 5.4 4.6 4.6 5.0 4.4 4.9 5.4 4.8 4.8 4.3 5.8 5.7  
[17] 5.4 5.1 5.7 5.1 5.4 5.1 4.6 5.1 4.8 5.0 5.0 5.2 5.2 4.7 4.8 5.4  
[33] 5.2 5.5 4.9 5.0 5.5 4.9 4.4 5.1 5.0 4.5 4.4 5.0 5.1 4.8 5.1 4.6  
[49] 5.3 5.0 7.0 6.4 6.9 5.5 6.5 5.7 6.3 4.9 6.6 5.2 5.0 5.9 6.0 6.1  
[65] 5.6 6.7 5.6 5.8 6.2 5.6 5.9 6.1 6.3 6.1 6.4 6.6 6.8 6.7 6.0 5.7  
[81] 5.5 5.5 5.8 6.0 5.4 6.0 6.7 6.3 5.6 5.5 5.5 6.1 5.8 5.0 5.6 5.7  
[97] 5.7 6.2 5.1 5.7 6.3 5.8 7.1 6.3 6.5 7.6 4.9 7.3 6.7 7.2 6.5 6.4  
[113] 6.8 5.7 5.8 6.4 6.5 7.7 7.7 6.0 6.9 5.6 7.7 6.3 6.7 7.2 6.2 6.1  
[129] 6.4 7.2 7.4 7.9 6.4 6.3 6.1 7.7 6.3 6.4 6.0 6.9 6.7 6.9 5.8 6.8  
[145] 6.7 6.7 6.3 6.5 6.2 5.9
```

物件命名原則

- R的大小寫有差異: *a* 與 *A* 是不同的物件
- R 也保留一些物件與指令人稱, 如 c, C, T, F 等為保留字 (“reserved words”), 命名時避免重覆, 以免引起人類困擾.
- 物件名稱起始位置須以文字或 “.” (句點)
- 如果物件名稱以 “.” (句點) 為起始, 名稱第二個位置需為文字, 物件名稱其餘位置, 以文字 (A-Z 或 a-z), 數字 (0-9), / . - , 皆可.
- 中間不可有空格

Google' s R Style Guide

- <https://google.github.io/styleguide/Rguide.html>
- 函數使用 BigCamelCase

```
# Good
DoNothing <- function() {
  return(invisible(NULL))
}
```

- 不要使用 attach 函數
- 使用 **x <- 1**, 不要使用 x = 1
- = 用於函數之參數設定 plot(..., type = "b")
- 不要使用句點 Customer.Sales  改為 CustomerSales

套件 package

套件

- 使用套件兩部曲 - 先安裝, 再載入套件
 - `install.packages("套件名稱")` # 安裝套件(一生一次)
 - `library(套件名稱)` # 載入套件(每次使用)
- 範例: 新增與載入 e1071套件(machine learning)

```
> install.packages("e1071")
trying URL 'http://cran.cs.pu.edu.tw/bin/windows/contrib/3.0/e1071_1.6-1.zip'
Content type 'application/zip' length 514468 bytes (502 Kb)
opened URL
downloaded 502 Kb

package 'e1071' successfully unpacked and MD5 sums checked

The downloaded binary packages are in
  C:\Users\Administrator\AppData\Local\Temp\RtmpoHS0Ak\downloaded_packages
> library(e1071)
Loading required package: class
>
```

example(svm, package="e1071")

已載入的套件 search()

```
> # 已載入套件
> search()
[1] ".GlobalEnv"
[3] "tools:rstudio"
[5] "package:graphics" "package:graphics"
[7] "package:utils"
[9] "package:methods"
[11] "package:base"
>
```



```
"package:e1071" "package:e1071"
"package:stats"
"package:grDevices"
"package:datasets" "package:datasets"
"Autoloads"
```

R套件 - 41類別

(2021.10.31)

- <https://cran.csie.ntu.edu.tw/web/packages/index.html>

Contributed Packages

Available Packages

Currently, the CRAN package repository features 18364 available packages.

[Table of available packages, sorted by date of publication](#)

[Table of available packages, sorted by name](#)

Installation of Packages

Please type `help("INSTALL")` or `help("install.packages")` in R for information on how to install packages from this repository. The manual [R Installation and Administration](#) (also contained in the R base sources) explains the process in detail.

[CRAN Task Views](#) allow you to browse packages by topic and provide tools to automatically install all packages for special areas of interest. Currently, 41 views are available.

41類別 - 中文說明

2013年10月8日 星期二

Task Views - R套件區分成41個類別

更新日期: 2020.3.1

CRAN Task View : <https://cran.r-project.org/web/views/>

套件區分成 41 個類別，中文說明如下：

編號 主題 英文說明 中文說明

01, Bayesian, Bayesian Inference, 貝氏統計

02, ChemPhys, Chemometrics and Computational Physics, 計量化學學, 計量物理

03, ClinicalTrials, Clinical Trial Design, Monitoring, and Analysis, 臨床試驗設計, 監測和分析

04, Cluster, Cluster Analysis & Finite Mixture Models, 群集分析, 有限混合模型

05, Databases, Databases with R, R與資料庫連接

06, DifferentialEquations, Differential Equations, 微分方程

07, Distributions, Probability Distributions, 機率分配

08, Econometrics, Computational Econometrics, 計量經濟

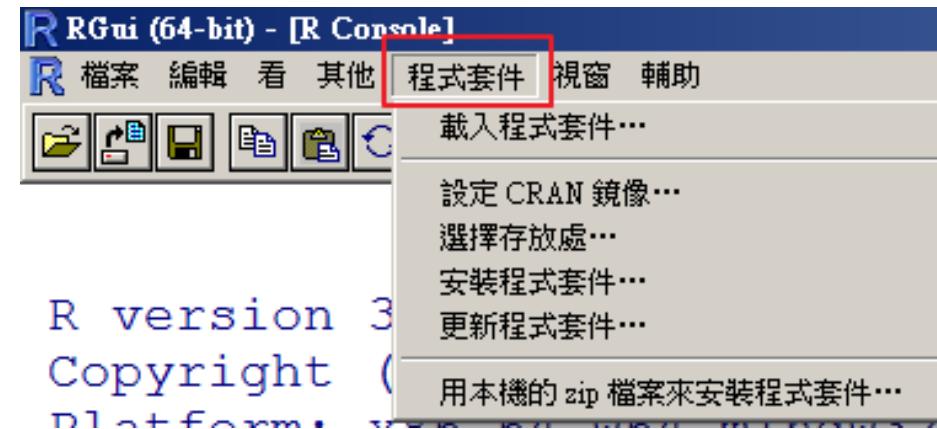
09, Environmetrics, Analysis of Ecological and Environmental Data, 生態, 環境資料分析

RWEPA → task

```
# 已安裝套件  
x <- installed.packages()  
x  
nrow(x)
```

```
library()
```

R 套件選單



- `update.packages("xxx") # 更新套件`
- `detach("package:xxx") # 卸離套件`
- `remove.packages("xxx") # 移除已安裝套件`
- 上述指令大部份可在 R / RStudio 執行

R對話資訊

- `sessionInfo()` → 理解R安裝訊息

```
> sessionInfo()
R version 4.1.0 Patched (2021-07-16 r80639)
Platform: x86_64-w64-mingw32/x64 (64-bit)
Running under: Windows 10 x64 (build 19043)

Matrix products: default

locale:
[1] LC_COLLATE=Chinese (Traditional)_Taiwan.950  LC_CTYPE=Chinese (Traditional)_Taiwan.950
[3] LC_MONETARY=Chinese (Traditional)_Taiwan.950 LC_NUMERIC=C
[5] LC_TIME=Chinese (Traditional)_Taiwan.950

attached base packages:
[1] stats      graphics   grDevices utils      datasets   methods    base

loaded via a namespace (and not attached):
[1] compiler_4.1.0 tools_4.1.0
```

套件安裝目錄

- .Library

```
> # 預設套件安裝目錄  
> .Library  
[1] "C:/PROGRA~1/R/R-41~1.0/library"
```

- .libPaths()

```
> # 套件安裝目錄  
> .libPaths()  
[1] "C:/Users/88697/Documents/R/win-library/4.1"  
[2] "C:/Program Files/R/R-4.1.0/library"
```

已安裝套件

```
> # 已安裝套件
> x <- installed.packages()
> class(x) # "matrix" "array"
[1] "matrix" "array"
> dim(x) # 626*16
[1] 625 16
> mypackage = x[, 1] # matrix[列, 行]
> mypackage[1:5]
      abind      ade4       AER      affy    affydata
"abind"    "ade4"    "AER"    "affy"  "affydata"
>
> library() # same as install.packages()
```

套件安裝目錄1

- C:\Program Files\R\R-4.1.0\library

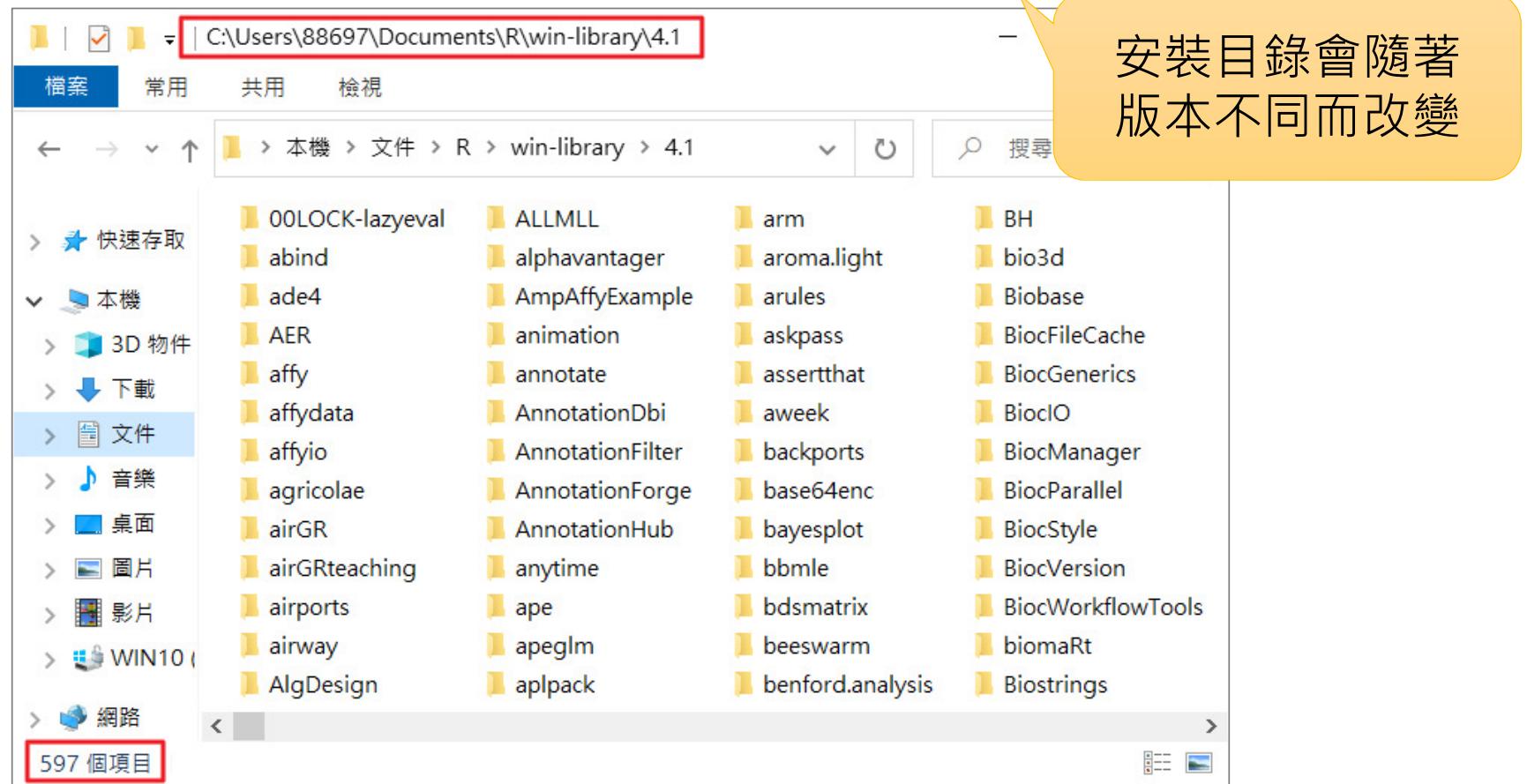
WIN10 (C) > Program Files > R > R-4.1.0 > library

base	grid	rpart
boot	KernSmooth	spatial
class	lattice	splines
cluster	MASS	stats
codetools	Matrix	stats4
compiler	methods	survival
datasets	mgcv	tcltk
foreign	nlme	tools
graphics	nnet	translations
grDevices	parallel	utils

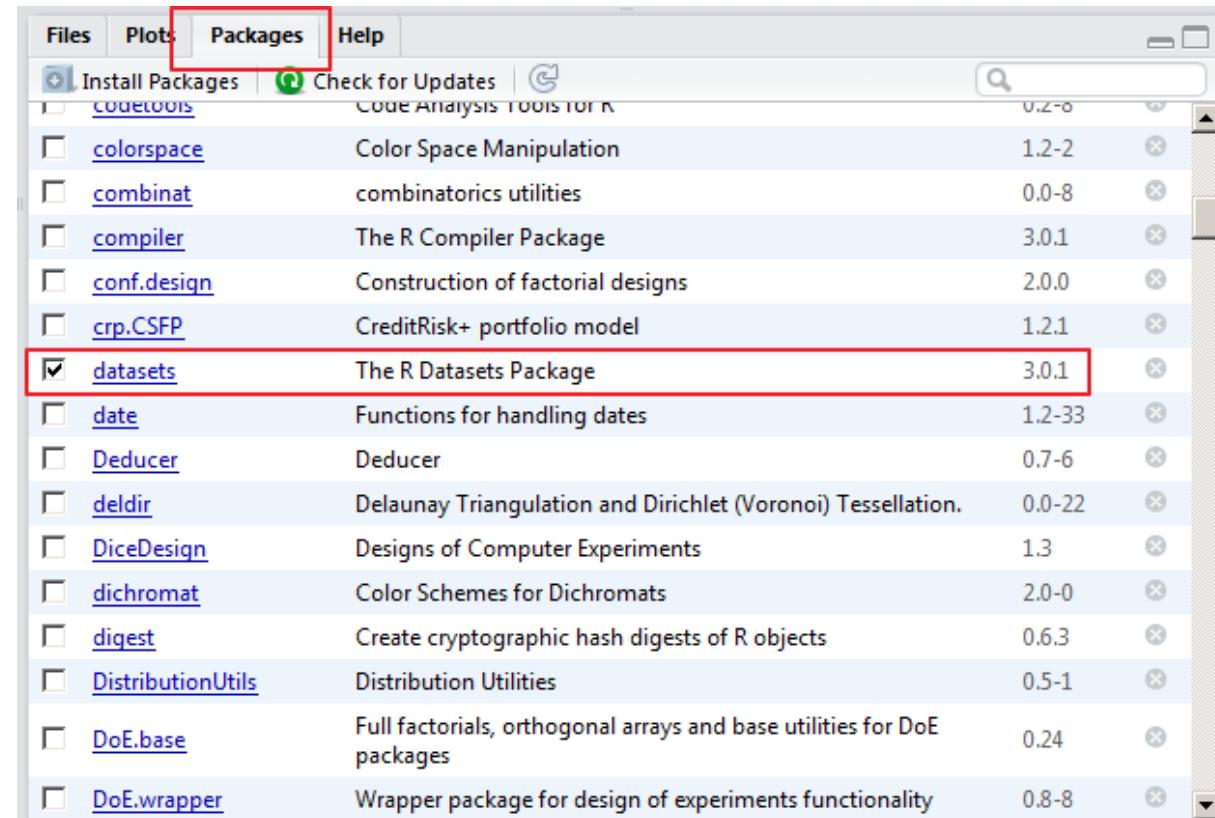
30個基礎套件，
不可刪除。

套件安裝目錄2

- C:\Users\88697\Documents\R\win-library\4.1



RStudio 套件管理



輔助說明 help

輔助說明

- 常用輔助說明方法
 - help.start()
 - ?plot
 - help(plot)
 - 選取 plot 按 F1
 - help.search()



```
> help.search("regression")
>
> ??regression
```

函數說明

1.函數

2.套件

3.R文件

4.簡單
說明

5.詳細說明

6.方法

7.參數

type="n"

plot {graphics}

① ②

Generic X-Y Plotting

R Documentation

Description

Generic function for plotting of R objects. For more details about the graphical parameter arguments, see [par](#).

For simple scatter plots, [plot.default](#) will be used. However, there are plot methods for many R objects, including [functions](#), [data.frames](#), [density](#) objects, etc. Use `methods(plot)` and the documentation for these.

Usage

`plot(x, y, ...)`

Arguments

x

the coordinates of points in the plot. Alternatively, a single plotting structure, function or *any R object with a plot method* can be provided.

數學計算

數學運算

- R 即是計算機
 - log, exp
- 算數操作 (arithmetic operator)
 - +, -, *, /, ^, %% , %/%, %*%
- 關係比較操作 (relation/comparison operator)
 - ==, !=, <, <=, >, >=
- 邏輯操作(logical operator)
 - !, &, |



- `x == "台北市"`
- `x == '台北市'`
- `y == 3.14`

特殊數值

- R 可以正確表示無窮大數值:
 - $+\infty$ (正無窮大): `Inf`
 - $-\infty$ (負無窮大): `-Inf`
- `NaN`: 不是一個數值(數學上無定義,例:0/0)
- `NA`: 表示遺漏值(missing values)或(Not Available)
- `is.finite(x)`判定是否為有限的
- `is.infinite(x)` 判定是否為無窮大
- `is.nan(x)` 判定是否為NaN
- `pi, letters, LETTERS, month.abb, month.name`

英文月份

資料型別

- 整數
- 數值
- 字串: 須使用 '台北市' 或 "台北市" 符號
- 邏輯值: 包括 TRUE, FALSE

資料匯入與匯出

資料建立與輸入輸出

- `read.table`
- `read.csv`
- `write.table`
- `write.csv`

資料整理與摘要

- 步驟 1. 設定工作目錄
- 步驟 2. 建立資料檔
- 步驟 3. 汇入資料 `read.table`
- 步驟 4. 資料處理
- 步驟 5. 汇出資料

步驟 1. 設定工作目錄

```
> # 預設工作目錄  
> getwd() # get working directory  
[1] "C:/Users/wepa9_000.VAIO/Documents"  
> workpath <- "C:/rdata"  
> workpath <- "C:\\rdata" # for windows  
> setwd(workpath)  
> # 已更改為 C:\\rdata 工作目錄  
> getwd()  
[1] "C:/rdata"
```

“C:/rdata”

步驟 2. 準備資料檔 – 範例

- 日空氣品質指標(AQI) – 下載CSV
- <https://data.gov.tw/dataset/40507>



日空氣品質指標(AQI)

環保署將每日空氣品質監測站小時測值，經計算之日AQI公布。

評分此資料集：
☆ ☆ ☆ ☆ ☆

平均 0.00 (0 人次投票)

瀏覽次數: 10230 下載次數: 3636 意見數: 3

列印

主要欄位說明
SiteId(測站編號)、SiteName(測站名稱)、MonitorDate(監測日期)、AQI(空氣品質指標)、SO2SubIndex(二氧化硫副指標)、COSubIndex(一氧化碳副指標)、O3SubIndex(臭氧副指標)、PM10SubIndex(懸浮微粒副指標)、NO2SubIndex(二氧化氮副指標)、O38SubIndex(臭氧8小時副指標)、PM25SubIndex(細懸浮微粒副指標)

資料資源下載網址

[JSON](#) 檢視資料 日空氣品質指標(AQI)-JSON

[CSV](#) (highlighted with a red border) 檢視資料 日空氣品質指標(AQI)-CSV

[XML](#) 檢視資料 日空氣品質指標(AQI)-XML

步驟 2. 準備資料檔(續)

	A	B	C	D	E	F	G	H	I	J	K
1	SiteId	SiteName	MonitorDate	AQI	SO2SubIndex	COSubIndex	O3SubIndex	PM10SubIndex	NO2SubIndex	O38SubIndex	PM25SubIndex
2	18	大園	2021/7/24	28	2	1		19	10	28	20
3	19	觀音	2021/7/24	35	2	1		25	8	35	22
4	20	平鎮	2021/7/24	24	0	1		12	17	24	17
5	21	龍潭	2021/7/24	25	0	1		10	10	25	10
6	22	湖口	2021/7/24	27	2	1				27	18
7	23	竹東	2021/7/24	23	0	1				23	8
8	24	新竹	2021/7/24	25	2	2		14		25	16
9	25	頭份	2021/7/24	23	18	1		17	23		15
10	26	苗栗	2021/7/24	19	2	1		14	10	19	12
11	27	三義	2021/7/24	26	2	1		12	8	26	11
12	56	前金	2021/7/24	37	2	1		16	7	37	24
13	57	前鎮	2021/7/24	23	2	1		19	23		23

遺漏值

步驟 3. 汇入資料 `read.table`

- `read.table` 函數將文字檔讀入R, 其回傳值是資料框(data.frame)
- 每一列表示一組觀測值(observation)
- 每直行表示一個變數(variable)
- `read.table` 函數預設以空白做為區隔變數

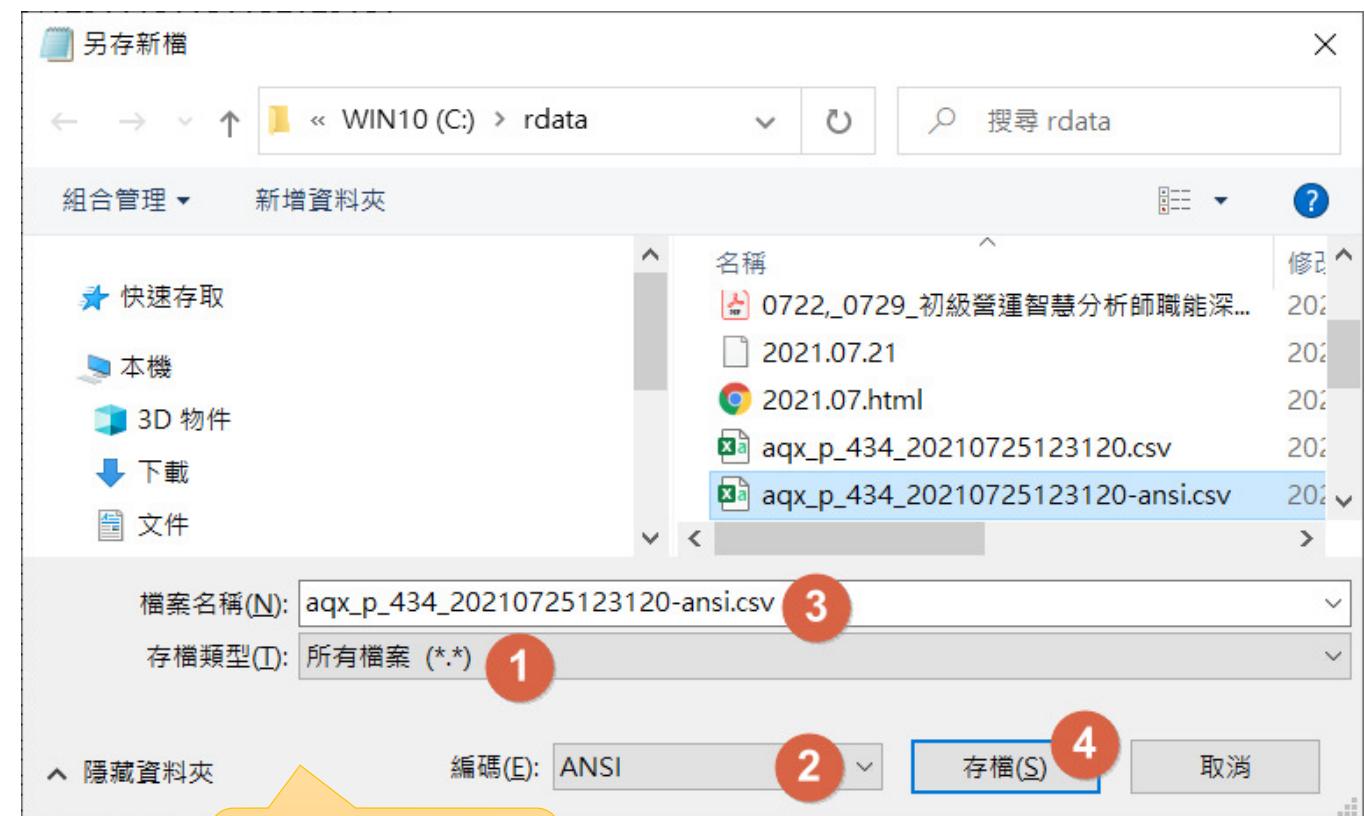
read.table("x.csv", header = TRUE, sep = ",")

```

> myfile <- "aqx_p_434_20210725123120.csv"
>
1 > aq <- read.table(myfile, header=TRUE, sep=",") # error!
2 Error in scan(file = file, what = what, sep = sep, quote = quote, dec = dec, :
3   line 2 did not have 11 elements
>
> aq <- read.table(myfile, header=TRUE, sep=",", fill=TRUE) # 亂碼!
> head(aq, n=3) # 檢視前6筆, 標題, 第2行有亂碼
  嘿燙iteId SiteName MonitorDate AQI SO2SubIndex cosubIndex O3SubIndex
1     18    慶批<9c><92> 2021-07-24  28          2       1      NA
2     19 閩\u0080<e9>\u2022,2021-07-24        35     2       1      NA      25
3     20    捲喲\u2022 2021-07-24  24          0       1      NA
  PM10SubIndex NO2SubIndex O38SubIndex PM25SubIndex
1     19         10        28        20
2     8          35        22        NA
3    12         17        24        17
>
> aq <- read.table(myfile, header=TRUE, sep=",", fill=TRUE, encoding="UTF-8")
> head(aq, n=3) # 第1個欄位名稱異常!
  X.U.FEFF.SiteId SiteName MonitorDate AQI SO2SubIndex cosubIndex O3SubIndex PM10SubIndex
1     18      大園 2021-07-24  28          2       1      NA      19
2     19      觀音 2021-07-24  35          2       1      NA      25
3     20      平鎮 2021-07-24  24          0       1      NA      12
  NO2SubIndex O38SubIndex PM25SubIndex
1     10        28        20
2     8          35        22
3    17         24        17

```

Windows 記事本 \ 編碼 ANSI



R demo

使用 ANSI 讀取資料格式正常

```
> # 將檔案另儲存為 ANSI 編碼格式
> myfileNew <- "aqx_p_434_20210725123120-ansi.csv"
> aq <- read.table(myfileNew, header=TRUE, sep=",",) # OK
> head(aq) # 第1個欄位名稱正常!
  siteId siteName MonitorDate AQI so2subIndex cosubIndex o3subIndex
1     18    大園 2021-07-24   28           2           1        NA
2     19    觀音 2021-07-24   35           2           1        NA
3     20    平鎮 2021-07-24   24           0           1        NA
4     21    龍潭 2021-07-24   25           0           1        NA
5     22    湖口 2021-07-24   27           2           1        NA
6     23    竹東 2021-07-24   23           0           1        NA
o38subIndex PM25subIndex
1             28            20
2             35            22
```

步驟 4. 資料處理 head, names

```
> head(aq, n=5)
  SiteId SiteName MonitorDate AQI SO2SubIndex COSubIndex O3SubIndex PM10SubIndex
1     40    朴子 2020-10-08   61         7         3       NA        51
2     84  富貴角 2020-10-08   54         0         2       NA        54
3     83    麥寮 2020-10-08   94         6         5       NA        68
4     80    關山 2020-10-08   42         1       NA       NA        17
5     78    馬公 2020-10-08   64         3         2       NA        35
  NO2SubIndex O38SubIndex PM25SubIndex
1           11          61          46
2            1          48          52
3           10          NA          94
4            6          42          26
5            2          64          50
>
> names(aq) # 欄位名稱
[1] "SiteId"      "SiteName"     "MonitorDate"   "AQI"         "SO2SubIndex"
[6] "COSubIndex"   "O3SubIndex"   "PM10SubIndex" "NO2SubIndex"  "O38SubIndex"
[11] "PM25SubIndex"
>
```

names 欄位名稱

步驟 4. 資料處理(as.Date)

```
> # 日期: 字串(chr)修正為日期(Date)
> str(aq)
'data.frame': 616 obs. of 11 variables:
$ SiteId      : int 40 84 83 80 78 77 75 72 71 70 ...
$ SiteName    : chr "朴子" "富貴角" "麥寮" "關山" ...
$ MonitorDate : chr "2020-10-08" "2020-10-08" "2020-10-08" "2020-10-08"
$ AQI         : int 61 54 94 42 64 90 101 77 72 40 ...
$ SO2SubIndex : int 7 0 6 1 3 9 3 3 13 7 ...
$ COSubIndex  : int 3 2 5 NA 2 3 1 5 7 11 ...
$ O3SubIndex  : logi NA NA NA NA NA NA ...
$ PM10SubIndex: int 51 54 68 17 35 42 31 30 47 31 ...
$ NO2SubIndex : int 11 1 10 6 2 8 4 12 40 29 ...
$ O38SubIndex : int 61 48 NA 42 64 90 101 77 NA NA ...
$ PM25SubIndex: int 46 52 94 26 50 53 45 55 72 40 ...
>
> aq$MonitorDate <- as.Date(aq$MonitorDate)
>
> str(aq)
'data.frame': 616 obs. of 11 variables:
$ SiteId      : int 40 84 83 80 78 77 75 72 71 70 ...
$ SiteName    : chr "朴子" "富貴角" "麥寮" "關山" ...
$ MonitorDate : Date[format: "2020-10-08" "2020-10-08" ...]
$ AQI         : int 61 54 94 42 64 90 101 77 72 40 ...
$ SO2SubIndex : int 7 0 6 1 3 9 3 3 13 7 ...
$ COSubIndex  : int 3 2 5 NA 2 3 1 5 7 11 ...
$ O3SubIndex  : logi NA NA NA NA NA NA ...
$ PM10SubIndex: int 51 54 68 17 35 42 31 30 47 31 ...
$ NO2SubIndex : int 11 1 10 6 2 8 4 12 40 29 ...
$ O38SubIndex : int 61 48 NA 42 64 90 101 77 NA NA ...
$ PM25SubIndex: int 46 52 94 26 50 53 45 55 72 40 ...
>
```

步驟 4. 資料處理(續)

```
> head(aq)
  SiteId SiteName MonitorDate AQI SO2SubIndex COsubIndex O3SubIndex PM10SubIndex NO2SubIndex O38SubIndex PM25SubIndex
1     40    朴子 2020-10-08   61         7        3       NA        51        11        61        46
2     84  富貴角 2020-10-08   54         0        2       NA        54         1        48        52
3     83    麥寮 2020-10-08   94         6        5       NA        68        10       NA        94
4     80    關山 2020-10-08   42         1       NA       NA        17         6        42        26
5     78    馬公 2020-10-08   64         3        2       NA        35         2        64        50
6     77    金門 2020-10-08   90         9        3       NA        42         8        90        53
>
> dim(aq) # 616列, 11行
[1] 616  11
>
```

步驟 4. 資料處理(續)

```
> AQI  
Error: object 'AQI' not found  
> aq$AQI  
[1] 44 51 26 49 60 47 36 40 37 43 44 40  
  
> attach(aq)          少用 attach  
> AQI  
[1] 44 51 26 49 60 47 36 40 37
```

步驟 4. 資料處理(續)

判斷是否相等使用 ==

```
# 篩選資料  
aq.Banqiao<- aq[aq$SiteName == "板橋", ]  
aq.Banqiao <- aq.Banqiao[order(aq.Banqiao$MonitorDate), ]  
  
aq.Xizhi <- aq[aq$SiteName == "汐止", ]  
aq.Xizhi <- aq.Xizhi[order(aq.Xizhi$MonitorDate), ]
```

步驟 4. 資料處理(續)

```
# 繪製直線圖
plot(aq.Banqiao$AQI, type = "b")
lines(aq.Xizhi$AQI, col = "red")
points(aq.Xizhi$AQI, col = "red")

ymin <- min(aq.Banqiao$AQI, aq.Xizhi$AQI) - 1
ymax <- max(aq.Banqiao$AQI, aq.Xizhi$AQI) + 3

plot(aq.Banqiao$AQI,
      type = "b",
      ylim = c(ymin, ymax),
      main = "2020年10月AQI-板橋vs.汐止",
      xaxt = "n",
      xlab = "日期")

axis(side = 1,
      at = 1:length(aq.Banqiao$AQI),
      labels = aq.Banqiao$MonitorDate,
      cex.axis = 0.8)

lines(aq.Xizhi$AQI, col="red")
points(aq.Xizhi$AQI, col="red")

legend("topleft", legend=c("板橋", "汐止"), col=c(1,2), lty=1)
```

- **ylim:** Y軸範圍
- **xaxt = “n”** 取消預設刻度
- **axis** 人工設定刻度
- **legend** 圖例

步驟 5. 汇出資料

rbind: 列合併

```
# 汇出資料  
aq.Banqiao.Xizhi <- rbind(aq.Banqiao, aq.Xizhi)
```

```
write.table(aq.Banqiao.Xizhi,  
            "aq.Banqiao.Xizhi.csv",  
            sep=",",  
            row.names=FALSE)
```

- **write.table :** 輸出檔案



實作練習18

RData 資料物件儲存/匯入

- `save(資料物件1, 資料物件2, file= "myData.RData")`
 - `load("myData.RData")`
-
- 儲存 `aq.Banqiao.Xizhi` 儲存為 `aq.Banqiao.Xizhi.RData`
 - 練習 載入 `aq.Banqiao.Xizhi.RData`

讀取 SAS 檔案

sas7bdat: SAS Database Reader (experimental)

Read SAS files in the `sas7bdat` data format.

Version: 0.5

Depends: R (≥ 2.10)

Published: 2014-06-04

Author: Matt Shotwell

Maintainer: Matt Shotwell <matt.shotwell at vanderbilt.edu>

License: [GPL-2](#) | [GPL-3](#) [expanded from: GPL (≥ 2)]

NeedsCompilation: no

Materials: [README](#)

CRAN checks: [sas7bdat results](#)

Downloads:

Reference manual: [sas7bdat.pdf](#)

Vignettes:

[sas7bdat](#)

R demo

SPSS/MySQL/SQL

- 汇入 SPSS {**foreign**}
- **foreign** 套件可读取以下档案格式:
 - Minitab
 - S
 - SAS
 - SPSS
 - Stata
 - Systat
 - Weka
 - dBase
- RMySQL 套件编译与建立
 - <http://rwepa.blogspot.com/2013/01/windows-rmysql.html>
- RODBC 与 SQL Server 资料汇入与写入
 - <http://rwepa.blogspot.com/2013/08/rodbc-sql-server.html>



07.R資料物件,判斷式與函數,群組分析與繪圖 graphics

資料物件

資料物件

向量 vector

北部	中部	南部
----	----	----

矩陣 matrix

1	3	5
2	4	6

陣列 array

1.1	4.4	7.7
2.2	5.5	8.8
3.3	6.6	9.9

資料框 data.frame

1	男	62
2	女	50
3	女	54
4	男	72

資料框：以串列方式儲存，但其長度相同。

串列 list

北部	中部	南部
1	3	5
2	4	6

1	男	62
2	女	50
3	女	54
4	男	72

向量

矩陣

資料框

串列：
每一個元素其資料型別與長度可以不相同。

矩陣預設採用
直行填入資料

資料物件重要觀念

- 向量是最基本的物件
 - 數值向量、字元向量
 - 所有資料的資料型別須相同
 - 因子是一種較特別的向量，儲存類別型變數
- 矩陣與陣列
 - 矩陣是二維陣列，陣列允許大於或等於2個維度
 - 所有資料的資料型別須相同
- 資料框、時間序列、串列均可同時存入數字與字串，但資料框與時間序列內的向量長度都相等
- 串列可包含陣列與資料框(二維，可多種資料型別)，函數回傳值以串列(list)物件為主

資料物件名稱中，英文皆可，建議英文，不可用數字開頭。

建立向量函數 c (concatenate)

- 將一群數字、**字串**、**邏輯值**結合成向量
- 將多個向量結合成向量
- 向量物件具有屬性長度 **length** 與型式**mode**
- 向量會將所有元素將強制(**coercion**)轉換成**單一相同型態**
- 因子 (**factor**) 是一種特別的向量，用於將資料依離散型變數做成分群 (**group**)

因子 factor - levels, labels

```
> # 因子 factor
> f1 <- factor(1:3)
> f2 <- factor(1:3, levels=1:5)
> f1
[1] 1 2 3
Levels: 1 2 3
> f2
[1] 1 2 3
Levels: 1 2 3 4 5
> f2[4] <- 5
> f2[5] <- 10
Warning message:
In `<-factor`(`*tmp*`, 5, value = 10) :
  invalid factor level, NA generated
> f2
[1] 1      2      3      5      <NA>
Levels: 1 2 3 4 5
>
```

factor 範例1

```
> eye.colors <- factor(c("brown", "blue", "blue", "green", "brown", "brown", "brown"))
>
> eye.colors
[1] brown blue  blue  green brown brown brown
Levels: blue brown green
>
> levels(eye.colors)
[1] "blue" "brown" "green"
>
> labels(eye.colors)
[1] "1" "2" "3" "4" "5" "6" "7"
>
```

factor 範例2

```
> gender <- factor(c("男", "女", "男", "男", "女"))
>
> gender
[1] 男 女 男 男 女
Levels: 女 男
>
> levels(gender)
[1] "女" "男"
>
> str(gender)
Factor w/ 2 levels "女", "男": 2 1 2 2 1
>
```

有序因子 (ordered factor)

- 有序因子表示有大小順序, 例: {大, 中, 小}

```
> clothSize <- ordered(c("L", "H", "L", "M", "H"),
+                           levels = c("L", "M", "H"))
>
> clothSize
[1] L H L M H
Levels: L < M < H
>
> levels(clothSize)
[1] "L" "M" "H"
>
> str(clothSize)
Ord.factor w/ 3 levels "L"<"M"<"H": 1 3 1 2 3
>
```

因子轉換

- `as.factor` 轉換為因子
- `as.numeric()` 轉換為數值
- `as.character()` 轉換為字串
- 因子內部儲存為 整數 {1, 2, 3, ...}, 整數表示顏色或使用 **colors()**

- 使用時機-例: 縣市別, 性別
- 使用 **cut** 函數: 數值→類別

向量 vector

- 類似 Excel 的一行或是一列
- 整數, 實數, 字元, 數值 + 字元?

```
> # 向量 vector -----
> # 整數
> v0 <- c(1:10)
> v0
[1] 1 2 3 4 5 6 7 8 9 10
> class(v0)
[1] "integer"
> typeof(v0)
[1] "integer"
>
```

R demo

矩陣 matrix

- 矩陣是將向量擴充至二個(或保持一個)維度。
- 建立矩陣
 - `matrix(data = NA, nrow = 1, ncol = 1,
byrow = FALSE, dimnames = NULL)`
- 轉換為矩陣
 - `as.matrix(x)`
- 判斷是否為矩陣
 - `is.matrix(x)`

預設採用直行
填入資料

矩陣的運算

- 判斷是否為矩陣 `is.matrix()`
- 轉換為矩陣 `as.matrix()`
- 矩陣運算 `+, -, %*%`
- 矩陣轉置 `t()`
- 取出對角線值 `diag()`

矩陣相乘

$a*X=b$
`?solve`

陣列 array

- 陣列是將向量擴充至二個(或以上)維度。
- 陣列表示多重維度且為相同資料型態。
- 產生陣列
 - `array(data = NA, dim = length(data), dimnames = NULL)`
- 轉換為陣列
 - `as.array(x)`
- 判斷是否為陣列
 - `is.array(x)`

資料框 data.frame

- 資料框是一種重要的資料物件型態。
- 一般R模型計算以資料框作為資料輸入。
- 資料框是二維資料物件，每一橫列表示一個觀測值，每一直行表示一個變數，變數資料型態可能不相同，但個數相同。

```
cars[2]      # data.frame  
cars["dist"] # data.frame  
cars[,2]     # vector
```

認識 iris, 150*5

```
> iris
```

	Sepal.Length	Sepal.Width	Petal.Length	Petal.Width	Species
1	5.1	3.5	1.4	0.2	setosa
2	4.9	3.0	1.4	0.2	setosa
3	4.7	3.2	1.3	0.2	setosa
4	4.6	3.1	1.5	0.2	setosa
5	5.0	3.6	1.4	0.2	setosa
6	5.4	3.9	1.7	0.4	setosa
7	4.6	3.4	1.4	0.3	setosa
8	5.0	3.4	1.5	0.2	setosa
9	4.4	2.9	1.4	0.2	setosa
10	4.9	3.1	1.5	0.1	setosa

資料結構 str

- head(iris) # 前6筆
- tail(iris) # 後6筆
- str(iris)

```
> str(iris)
'data.frame': 150 obs. of 5 variables:
 $ Sepal.Length: num 5.1 4.9 4.7 4.6 5 5.4 4.6 5 4.4 4.9 ...
 $ Sepal.Width : num 3.5 3 3.2 3.1 3.6 3.9 3.4 3.4 2.9 3.1 ...
 $ Petal.Length: num 1.4 1.4 1.3 1.5 1.4 1.7 1.4 1.5 1.4 1.5 ...
 $ Petal.Width : num 0.2 0.2 0.2 0.2 0.2 0.4 0.3 0.2 0.2 0.1 ...
 $ Species      : Factor w/ 3 levels "setosa","versicolor",...: 1 1 1 1 1 1 1 1 1 1 ...
```

資料摘要 summary

```
> summary(iris)
   Sepal.Length    Sepal.Width     Petal.Length    Petal.Width      Species
  Min. :4.300    Min. :2.000    Min. :1.000    Min. :0.100    setosa   :50
  1st Qu.:5.100  1st Qu.:2.800  1st Qu.:1.600  1st Qu.:0.300  versicolor:50
  Median :5.800  Median :3.000  Median :4.350  Median :1.300  virginica :50
  Mean   :5.843  Mean   :3.057  Mean   :3.758  Mean   :1.199
  3rd Qu.:6.400  3rd Qu.:3.300  3rd Qu.:5.100  3rd Qu.:1.800
  Max.   :7.900  Max.   :4.400  Max.   :6.900  Max.   :2.500
>
```



資料框的操作

實作練習19

- 取出1行 → 向量
- 取出1列 → 向量
- 取出1個數值 $x[\text{列}, \text{行}]$
 - 練習1. iris {datasets} 操作
找出 Sepal.Length 變數大於中位數的資料集
 - 練習2. Cars93 {MASS}操作, 資料篩選
`data(Cars93, package="MASS")`
 - 練習3. 計算 Luggage.room 的平均值?

日期與時間資料

資料物件 日期 as.Date

```
> # 字串轉換為日期 as.Date( )  
  
> mydates <- as.Date(c("2013-01-01", "2016-04-07"))  
  
> mydates  
[1] "2013-01-01" "2016-04-07"  
  
> class(mydates)  
[1] "Date"  
  
> # 日期相差  
  
> days <- mydates[2] - mydates[1]  
  
> days  
Time difference of 1192 days
```

日期/時間格式

Symbol	Meaning	Example
%d	day as a number (0-31)	01-31
%a	abbreviated weekday	Mon
%A	unabbreviated weekday	Monday
%m	month (00-12)	00-12
%b	abbreviated month	Jan
%B	unabbreviated month	January
%y	2-digit year	07
%Y	4-digit year	2007

%H 時

%M 分

%S 秒

```
> # 系統日期  
> today <- Sys.Date()  
> today  
[1] "2016-04-07"  
  
> format(today, format="%B %d %Y")  
[1] "四月 07 2016"
```

日期與字串轉換

```
> # 日期轉換 format 'mm/dd/yyyy'  
  
> strDates <- c("01/05/2015", "04/07/2016")  
  
> dates <- as.Date(strDates, "%m/%d/%Y")  
  
> dates  
[1] "2015-01-05" "2016-04-07"  
  
> # 日期轉換字串  
  
> strDates <- as.character(dates)  
  
> strDates  
[1] "2015-01-05" "2016-04-07"  
  
mydatetime <- strptime("2016-04-15 08-01-09", format="%Y-%m-%d %H-%M-%S")  
mydatetime  
class(mydatetime)
```

?strptime
?as.Date

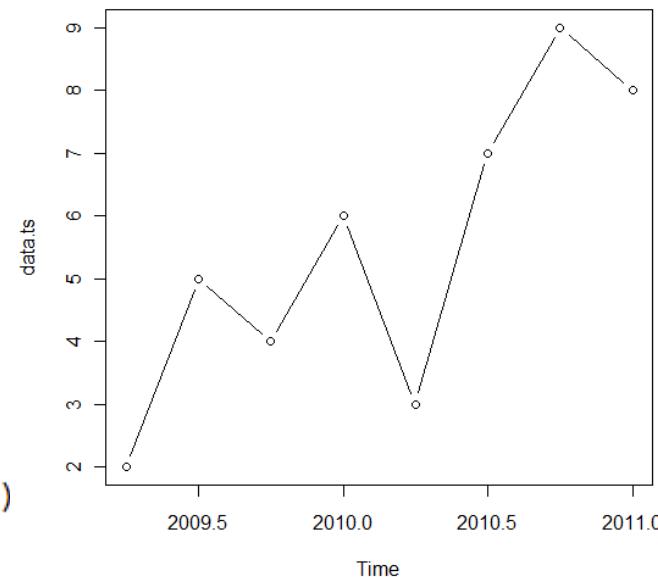
時間序列 ts

時間序列 ts → 日, 月, 季, 年

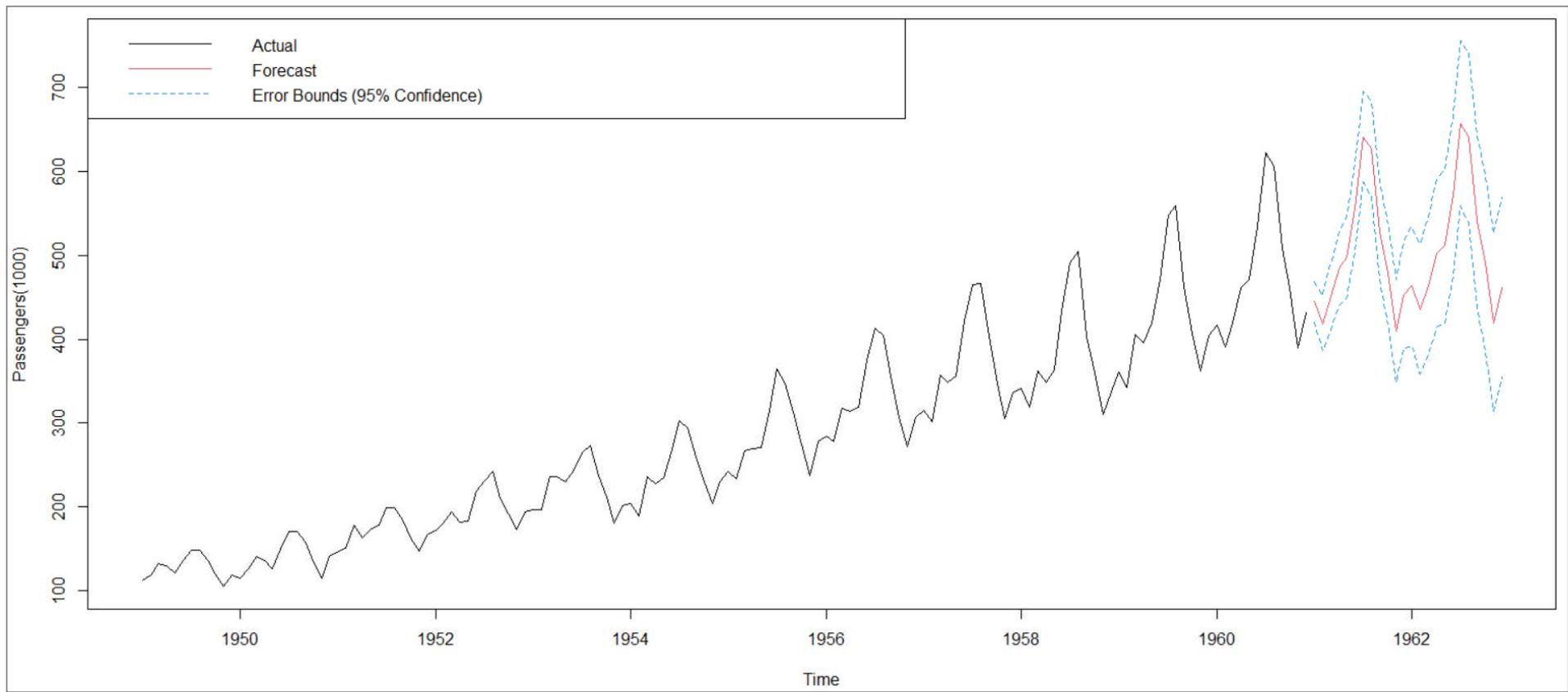
- 觀測變數隨時間變化並加以記錄，例如每日股票指數，每日平均溫度。
- 時間序列中的迴歸函數 (例：ar or arima) 採用 ts 物件。
- 時間序列物件包括多種特定繪圖方法。

時間序列 - 範例

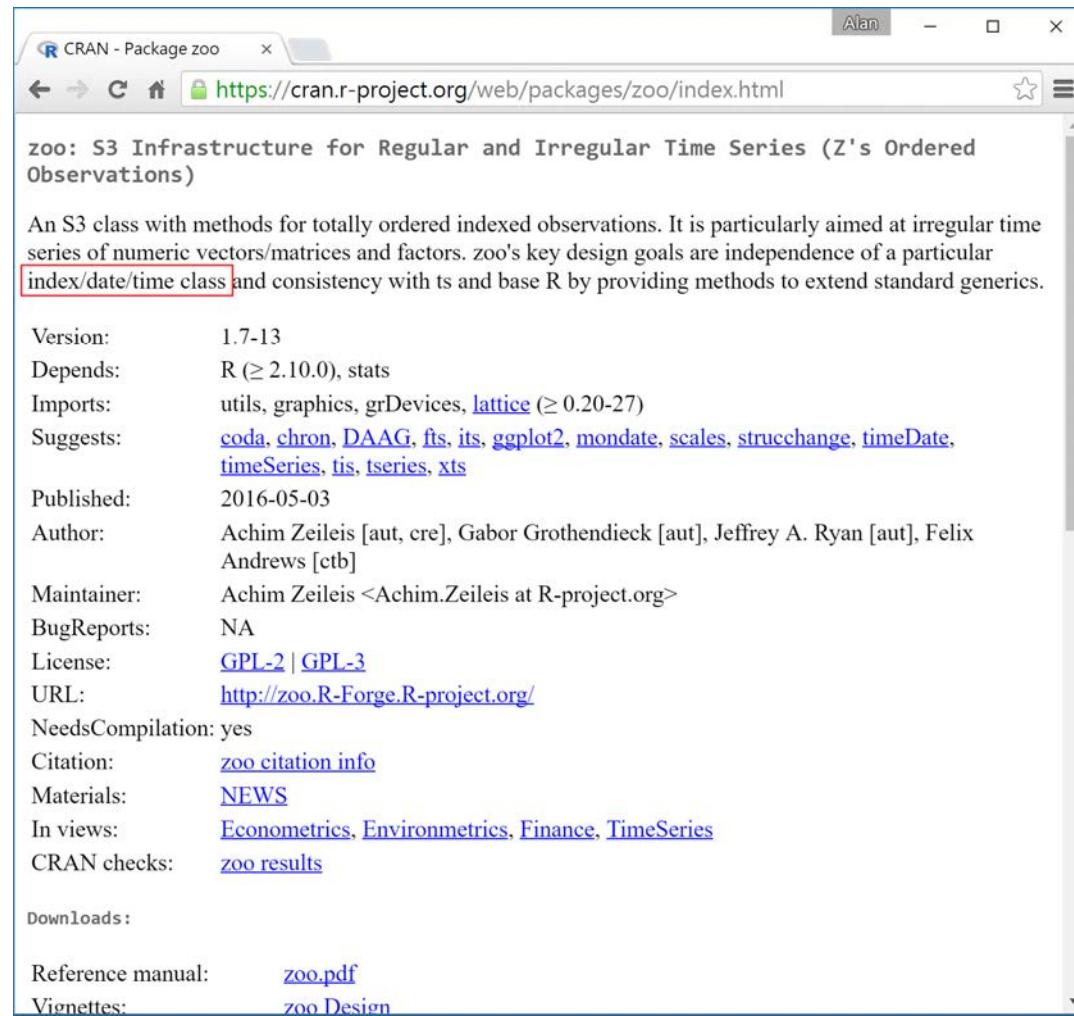
```
> # Time-series
> data.ts <- ts(c(2,5,4,6,3,7,9:8),start=c(2009,2),frequency=4)
> data.ts
      Qtr1 Qtr2 Qtr3 Qtr4
2009      2      5      4
2010      6      3      7      9
2011      8
> is.ts(data.ts)
[1] TRUE
> start(data.ts)
[1] 2009   2
> end(data.ts)
[1] 2011   1
> frequency(data.ts)
[1] 4
> deltat(data.ts) # 0.25 (=1/4)
[1] 0.25
> plot(data.ts, type="b")
> # end
```



ARIMA 視覺化



zoo 套件 – 日期/時間物件



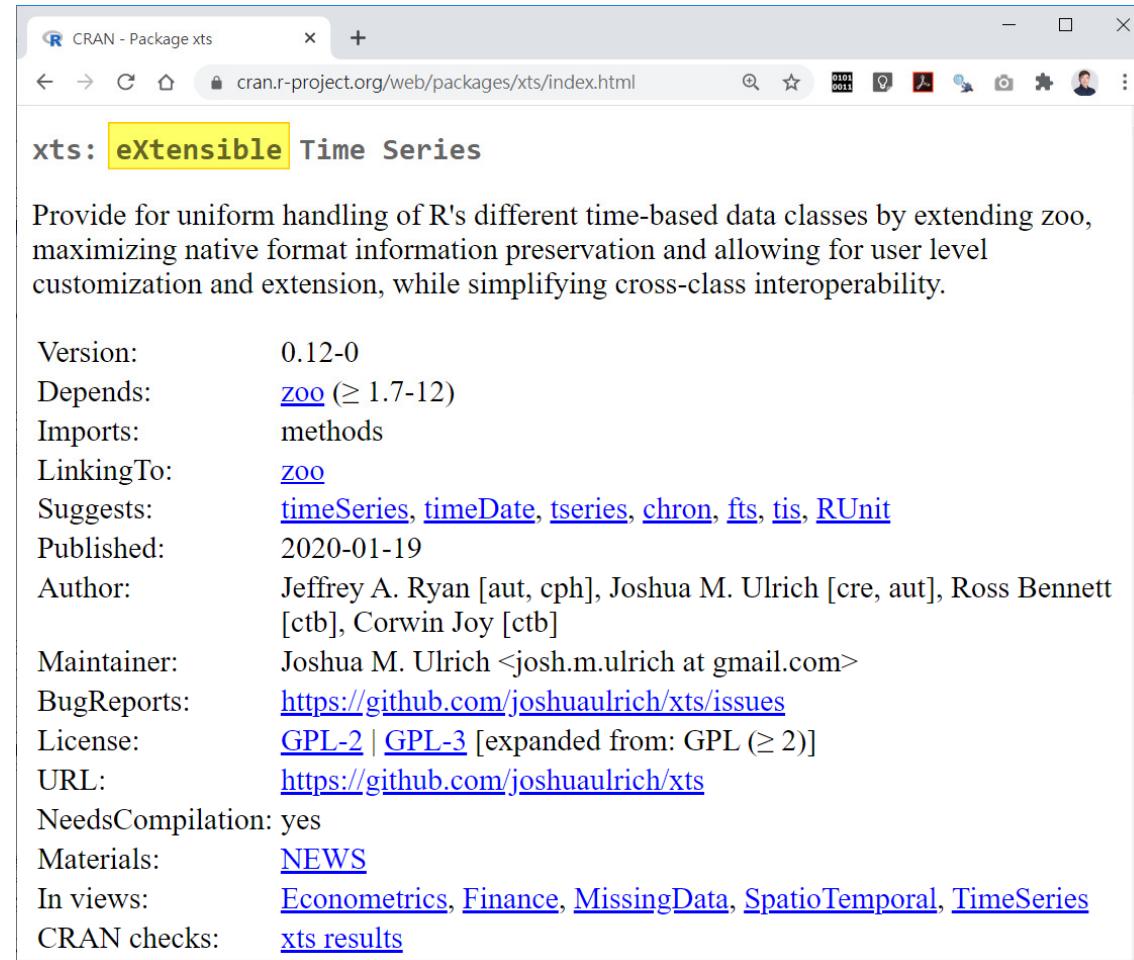
The screenshot shows a web browser window displaying the CRAN package page for 'zoo'. The URL in the address bar is <https://cran.r-project.org/web/packages/zoo/index.html>. The page content is as follows:

zoo: S3 Infrastructure for Regular and Irregular Time Series (Z's Ordered Observations)

An S3 class with methods for totally ordered indexed observations. It is particularly aimed at irregular time series of numeric vectors/matrices and factors. zoo's key design goals are independence of a particular index/date/time class and consistency with ts and base R by providing methods to extend standard generics.

Version:	1.7-13
Depends:	R (>= 2.10.0), stats
Imports:	utils, graphics, grDevices, lattice (>= 0.20-27)
Suggests:	coda , chron , DAAG , fts , its , ggplot2 , mondate , scales , strucchange , timeDate , timeSeries , tis , tseries , xts
Published:	2016-05-03
Author:	Achim Zeileis [aut, cre], Gabor Grothendieck [aut], Jeffrey A. Ryan [aut], Felix Andrews [ctb]
Maintainer:	Achim Zeileis <Achim.Zeileis at R-project.org>
BugReports:	NA
License:	GPL-2 GPL-3
URL:	http://zoo.R-Forge.R-project.org/
NeedsCompilation:	yes
Citation:	zoo citation info
Materials:	NEWS
In views:	Econometrics , Environmetrics , Finance , TimeSeries
CRAN checks:	zoo results
Downloads:	
Reference manual:	zoo.pdf
Vignettes:	zoo Design

xts 套件



The screenshot shows a web browser window displaying the CRAN package page for 'xts'. The title bar reads 'CRAN - Package xts'. The main content area has a yellow header 'xts: eXtensible Time Series'. Below it, a paragraph describes the package: 'Provide for uniform handling of R's different time-based data classes by extending zoo, maximizing native format information preservation and allowing for user level customization and extension, while simplifying cross-class interoperability.' A table lists various package details:

Version:	0.12-0
Depends:	zoo ($\geq 1.7-12$)
Imports:	methods
LinkingTo:	zoo
Suggests:	timeSeries , timeDate , tseries , chron , fts , tis , RUnit
Published:	2020-01-19
Author:	Jeffrey A. Ryan [aut, cph], Joshua M. Ulrich [cre, aut], Ross Bennett [ctb], Corwin Joy [ctb]
Maintainer:	Joshua M. Ulrich <josh.m.ulrich at gmail.com>
BugReports:	https://github.com/joshualrich/xts/issues
License:	GPL-2 GPL-3 [expanded from: GPL (≥ 2)]
URL:	https://github.com/joshualrich/xts
NeedsCompilation:	yes
Materials:	NEWS
In views:	Econometrics , Finance , MissingData , SpatioTemporal , TimeSeries
CRAN checks:	xts results

tidyquant 財金套件

2000年~2021年台灣加權股價指數收盤價



tidyquant – 台股繪圖

```
> # tidyquant 財金套件 -----
> library(tidyquant) # tq_get
>
> library(tidyverse) # 自動載入ggplot2
>
> TWII <- tq_get("^TWII", get = "stock.prices", from = "2000-01-01")
>
> # 台灣加權股價指數收盤價視覺化 -----
> TWII %>%
+   ggplot(aes(x = date, y = close)) +
+   geom_line(size = 1) +
+   labs(title = "2000年~2021年台灣加權股價指數收盤價",
+        x = "日期", y = "收盤價")
>
```

判斷式

if...else
ifelse
switch
for
while

```
?if # ERROR, 出現+
# unexpected token '?', expected 'LPAREN', LPAREN: Left PARENthesis 左括號
?"if" # OK
?`if` # OK
```

if 範例1

```
x <- 6
if (x > 0) {
  print("正數")
}
```

if (...) {
}

```
x <- -1
if(x > 0) {
  print("非負數")
} else {
  print("負數")
}
```

if 範例2

```
x <- 0
if (x < 0) {
  print("負數")
} else if (x > 0) {
  print("正數")
} else
  print("零")
```

ifelse 用法

```
> # ifelse 用法 -----
> x <- c(1, 3, NA, 5, NA)
> ifelse(is.na(x), 999, x)
[1]    1    3 999    5 999
>
```

switch 函數

```
> # switch 函數 -----
> centre <- function(x, type) {
+   switch(type, 1
+     mean = mean(x),
+     median = median(x),
+     trimmed = mean(x, trim = 0.1))
+ }
> (x <- rt(n=10, df=1)) # 隨機抽樣10個樣本,其自由度為1的t分配
[1]  0.3499267  1.9869349 -1.1893368 -0.2432037
[5] -5.9578435 -3.1969842 -0.2317308 -1.9788989
[9]  0.2586949 -27.1684901
> centre(x, "mean")
[1] -3.737093
> centre(x, "median")
[1] -0.7162703
> centre(x, "trimmed") # 結果與沒有 trimmed 有明顯差異
[1] -1.523672
```

for迴圈

```
> # for迴圈 -----
> (mydata <- c("R", "ggplot2", "shiny", "zoo", "leaflet"))
[1] "R"          "ggplot2"    "shiny"      "zoo"       "leaflet"
> for (i in mydata) {
+   print(nchar(i))
+
[1] 1
[1] 7
[1] 5
[1] 3
[1] 7
```

1 2 3

while迴圈

```
> # while迴圈 -----
> i <- 1
> while (i <= length(mydata)) {
+   print(paste0("字串 ", mydata[i], " 的長度為: ", nchar(mydata[i])))
+   i <- i + 1
+
[1] "字串 R 的長度為: 1"
[1] "字串 ggplot2 的長度為: 7"
[1] "字串 shiny 的長度為: 5"
[1] "字串 zoo 的長度為: 3"
[1] "字串 leaflet 的長度為: 7"
```

變數 *i* 須設定初值

函數

函數

- 函數：R的物件之一，用以評估一些輸入參數，並回傳結果作為輸出值。
- 函數架構

```
functionName <- function(參數1, 參數2, ...) {  
    函數主程式  
    return(輸出值)  
}
```

函數-範例

```
> oddcount <- function(x) {  
+   k <- 0 # assign 0 to k  
+   for (n in x) {  
+     if (n %% 2 == 1) k <- k+1 # %% is the modulo operator  
+   }  
+   return(k)  
+ }  
>  
> test <- c(1:100)  
> oddcount(test)  
[1] 50
```

計算奇數個數

撰寫程式時
“+”不用輸入



自訂函數

實作練習20

- 輸入iris 資料集
- 計算每列第2大數值的向量結果
- 使用 function 自訂函數

資料轉換

資料物件判斷與資料物件轉換

Test	Convert
<code>is.numeric()</code>	<code>as.numeric()</code>
<code>is.character()</code>	<code>as.character()</code>
<code>is.vector()</code>	<code>as.vector()</code>
<code>is.matrix()</code>	<code>as.matrix()</code>
<code>is.data.frame()</code>	<code>as.data.frame()</code>
<code>is.factor()</code>	<code>as.factor()</code>
<code>is.logical()</code>	<code>as.logical()</code>

判斷 is

轉換 as

數學轉換

- 數學轉換

- $\log(x)$
- $1/x$
- $x^{0.5}$
- $1/x^2$
- $1/x^{0.5}$

```
> # 數學轉換 -----
> x <- 100
> log(x)
[1] 4.60517
>
> log10(x)
[1] 2
>
> log2(x)
[1] 6.643856
>
> log(x, base = 5)
[1] 2.861353
>
> 1/x
[1] 0.01
>
> x^0.5
[1] 10
>
> 1/x^0.5
[1] 0.1
```

標準化 - scale

scale {base}

R Documentation

Scaling and Centering of Matrix-like Objects

Description

scale is generic function whose default method centers and/or scales the columns of a numeric matrix.

Usage

① ② ③
scale(x, center = TRUE, scale = TRUE)

Arguments

x

a numeric matrix(like object).

center

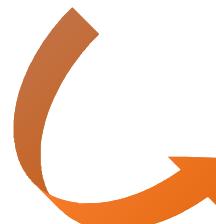
either a logical value or a numeric vector of length equal to the number of columns of x.

scale

either a logical value or a numeric vector of length equal to the number of columns of x.

標準化 - scale(續)

```
> # 2-3 標準化處理 ----  
> x <- matrix(1:30, ncol = 3)  
> x  
      [,1] [,2] [,3]  
[1,]    1   11   21  
[2,]    2   12   22  
[3,]    3   13   23  
[4,]    4   14   24  
[5,]    5   15   25  
[6,]    6   16   26  
[7,]    7   17   27  
[8,]    8   18   28  
[9,]    9   19   29  
[10,]   10   20   30
```



```
> scale(x) # mean=0, sd=1  
      [,1]      [,2]      [,3]  
[1,] -1.4863011 -1.4863011 -1.4863011  
[2,] -1.1560120 -1.1560120 -1.1560120  
[3,] -0.8257228 -0.8257228 -0.8257228  
[4,] -0.4954337 -0.4954337 -0.4954337  
[5,] -0.1651446 -0.1651446 -0.1651446  
[6,]  0.1651446  0.1651446  0.1651446  
[7,]  0.4954337  0.4954337  0.4954337  
[8,]  0.8257228  0.8257228  0.8257228  
[9,]  1.1560120  1.1560120  1.1560120  
[10,] 1.4863011  1.4863011  1.4863011  
attr("scaled:center")  
[1]  5.5 15.5 25.5  
attr("scaled:scale")  
[1] 3.02765 3.02765 3.02765
```

標準化 - 最小值/最大值法

```
# 標準化 - 最小值/最大值法 ----  
  
# R 練習-將 Cars93$Price 標準化至1~5之間  
x <- Cars93[1:10,]  
x$Price  
  
(max.old <- max(x$Price))  
(min.old <- min(x$Price))  
  
price.new <- ((x$Price - min.old)/(max.old - min.old))*(5 - 1) + 1  
price.new
```

使用R內建函數

- min
- which.min
- max
- which.max
- sort
- order
- which
- %in%

比較 sort, order

向量化運算思維

- R的運算採用向量化方式進行

```
> x <- c(1:10)
> (x.square <- x^2)
[1] 1 4 9 16 25 36 49 64 81 100
>
> x <- runif(23000000)
> system.time(x2 <- x^2)
使用者 系統 流逝
0.08 0.01 0.09

> system.time(x2 <- x^2)
user system elapsed
0.02 0.04 0.04
```

apply與lapply應用

apply 家族 -列,行的統計值

```
> # apply
> m <- matrix(c(1:8), ncol=2)
> m
      [,1] [,2]
[1,]    1    5
[2,]    2    6
[3,]    3    7
[4,]    4    8
> apply(m, 1, function(x) mean(x))
[1] 3 4 5 6
> as.matrix(apply(m, 1, function(x) mean(x)))
      [,1]
[1,]    3
[2,]    4
[3,]    5
[4,]    6
>
```

which, any, all 函數

which, any, all

```
> (x <- matrix(rnorm(10)*10, ncol=2))
      [,1]      [,2]
[1,]  5.691987 -15.789516
[2,]  9.560501 -21.025382
[3,] -12.430259  1.666184
[4,] -13.715679 -13.974480
[5,]  2.167013  -0.220889
> x>5
      [,1]      [,2]
[1,] TRUE FALSE
[2,] TRUE FALSE
[3,] FALSE FALSE
[4,] FALSE FALSE
[5,] FALSE FALSE
> # !(x>5)
>
> which(x>5) # return index
[1] 1 2
> which(x>5, arr.ind = TRUE) # return (row,column)
     row col
[1,]    1    1
[2,]    2    1
>
> any(x>5)
[1] TRUE
>
> all(x>5)
[1] FALSE
```

群組分析

iris鳶尾花資料集

```
> iris
   Sepal.Length Sepal.Width Petal.Length Petal.Width Species
1          5.1        3.5         1.4       0.2    setosa
2          4.9        3.0         1.4       0.2    setosa
3          4.7        3.2         1.3       0.2    setosa
4          4.6        3.1         1.5       0.2    setosa
5          5.0        3.6         1.4       0.2    setosa
6          5.4        3.9         1.7       0.4    setosa
7          4.6        3.4         1.4       0.3    setosa
8          5.0        3.4         1.5       0.2    setosa
9          4.4        2.9         1.4       0.2    setosa
10         4.9        3.1         1.5       0.1   setosa
```

iris + apply / lapply

- apply(iris[-5], 1, mean)

1

2

3

```
> apply(iris[-5], 1, mean)
```

```
[1] 2.550 2.375 2.350 2.350 2.550 2.850 2.425 2.525 2.225 2.400 2.700 2.500 2.325 2.125 2.800 3.000 2.750  
[18] 2.575 2.875 2.675 2.675 2.675 2.350 2.650 2.575 2.450 2.600 2.600 2.550 2.425 2.425 2.675 2.725 2.825  
[35] 2.425 2.400 2.625 2.500 2.225 2.550 2.525 2.100 2.275 2.675 2.800 2.375 2.675 2.350 2.675 2.475 4.075  
[52] 3.900 4.100 3.275 3.850 3.575 3.975 2.900 3.850 3.300 2.875 3.650 3.300 3.775 3.350 3.900 3.650 3.400  
[69] 3.600 3.275 3.925 3.550 3.800 3.700 3.725 3.850 3.950 4.100 3.725 3.200 3.200 3.150 3.400 3.850 3.600  
[86] 3.875 4.000 3.575 3.500 3.325 3.425 3.775 3.400 2.900 3.450 3.525 3.525 3.675 2.925 3.475 4.525 3.875  
[103] 4.525 4.150 4.375 4.825 3.400 4.575 4.200 4.850 4.200 4.075 4.350 3.800 4.025 4.300 4.200 5.100 4.875  
[120] 3.675 4.525 3.825 4.800 3.925 4.450 4.550 3.900 3.950 4.225 4.400 4.550 5.025 4.250 3.925 3.925 4.775  
[137] 4.425 4.200 3.900 4.375 4.450 4.350 3.875 4.550 4.550 4.300 3.925 4.175 4.325 3.950
```

```
>
```

lapply 回傳 list

```
> apply(iris[-5], 2, mean)
Sepal.Length  Sepal.Width  Petal.Length  Petal.Width
      5.843333     3.057333     3.758000     1.199333
>
> lapply(iris[-5], mean)
$Sepal.Length  ①  ②
[1] 5.843333

$Sepal.Width
[1] 3.057333

$Petal.Length
[1] 3.758

$Petal.Width
[1] 1.199333

>
```

資料處理

- 數值篩選
- 字串篩選
- 橫列篩選
- 直行篩選
- 橫列上下合併 `rbind` (row bind)
- 直行左右合併 `cbind` (column bind)

彙總計算

- `table` → 次數統計表
- `aggregate` → 類似 Excel 樞紐分析

table, aggregate

```
data(cars93, package = "MASS")
head(cars93)
str(cars93)
summary(cars93) # 有NA值





```

資料框排序

```
> df <- head(iris, n = 5)
>
> # 遞增排序
> df[order(df$Sepal.Length),]
  Sepal.Length Sepal.Width Petal.Length Petal.Width Species
4          4.6      3.1         1.5       0.2   setosa
3          4.7      3.2         1.3       0.2   setosa
2          4.9      3.0         1.4       0.2   setosa
5          5.0      3.6         1.4       0.2   setosa
1          5.1      3.5         1.4       0.2   setosa
>
> # 遞減排序
> df[order(df$Sepal.Length, decreasing = TRUE),]
  Sepal.Length Sepal.Width Petal.Length Petal.Width Species
1          5.1      3.5         1.4       0.2   setosa
5          5.0      3.6         1.4       0.2   setosa
2          4.9      3.0         1.4       0.2   setosa
3          4.7      3.2         1.3       0.2   setosa
4          4.6      3.1         1.5       0.2   setosa
```

群組 table (交叉表格)

```
> # 群組個數 table  
> table(Cars93$AirBags)
```

	Driver & Passenger	Driver only	None
	16	43	34

```
>
```

```
> # 群組個數 table-2個維度  
> table(Cars93$AirBags, Cars93$Origin)
```

	USA	non-USA
Driver & Passenger	9	7
Driver only	23	20
None	16	18

addmargins 新增邊界值

```
> # 群組邊界計算 addmargins-預設值為總和  
> addmargins(table(Cars93$AirBags, Cars93$Origin))
```

	USA	non-USA	Sum	
Driver & Passenger	9	7	16	
Driver only	23	20	43	
None	16	18	34	
Sum	48	45	93	

```
>  
> # 群組邊界計算 addmargins-mean  
> addmargins(table(Cars93$AirBags, Cars93$Origin), FUN = mean)
```

Margins computed over dimensions
in the following order:

1:

2:

1

2

	USA	non-USA	mean	
Driver & Passenger	9.0	7.0	8.0	
Driver only	23.0	20.0	21.5	
None	16.0	18.0	17.0	
mean	16.0	15.0	15.5	

```
>  
:
```

群組百分比 prop.table

```
> # 群組百分比計算 prop.table  
> prop.table(table(Cars93$AirBags, Cars93$Origin))
```

	USA	non-USA
Driver & Passenger	0.09677419	0.07526882
Driver only	0.24731183	0.21505376
None	0.17204301	0.19354839

table 多維度

```
> # 安全氣囊，進口別，傳動系統  
> table(Cars93$AirBags, Cars93$Origin, Cars93$DriveTrain)
```

		1	3	2	1
3					
Driver & Passenger	0	0			
Driver only	3	2			
None	2	3			

, , = Front

	USA	non-USA
Driver & Passenger	6	5
Driver only	15	13
None	13	15

, , = Rear

	USA	non-USA
Driver & Passenger	3	2
Driver only	5	5
None	1	0

aggregate (Excel樞紐分析)

```

> # 類別平均值計算
> aggregate(formula = Price ~ AirBags, data = Cars93, FUN = mean)
      AirBags    Price
1 Driver & Passenger 28.36875
2     Driver only 21.22326
3          None 13.17353
>
> aggregate(formula = Price ~ AirBags + Origin, data = Cars93, FUN = mean)
      AirBags   Origin    Price
1 Driver & Passenger     USA 24.57778
2     Driver only     USA 19.86957
3          None     USA 13.33125
4 Driver & Passenger non-USA 33.24286
5     Driver only non-USA 22.78000
6          None non-USA 13.03333
>

```

	A	B
1		
2		
3	列標籤	平均值 - Price
4	Driver & Passenger	28.36875
5	Driver only	21.22325581
6	None	13.17352941
7	總計	19.50967742

繪圖 {graphics}

繪圖

- R針對不同資料物件會有不同對應之繪圖方法。

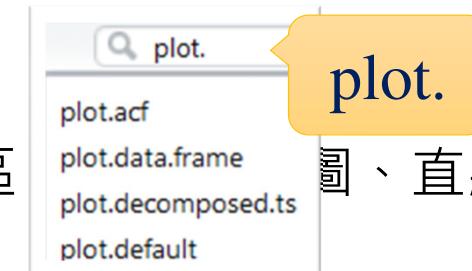
- 基礎繪圖包括二種不同類型：

- **高階繪圖**(high-level plotting) :

在圖形裝置(graphic device)上建立新的繪圖區
長條圖、直方圖、圓形圖、盒鬚圖與3D繪圖。

- **低階繪圖**(low-level plotting) :

在一個已經存在的繪圖上，加上其它圖形元素。例：額外的點、線、文字與圖例等。



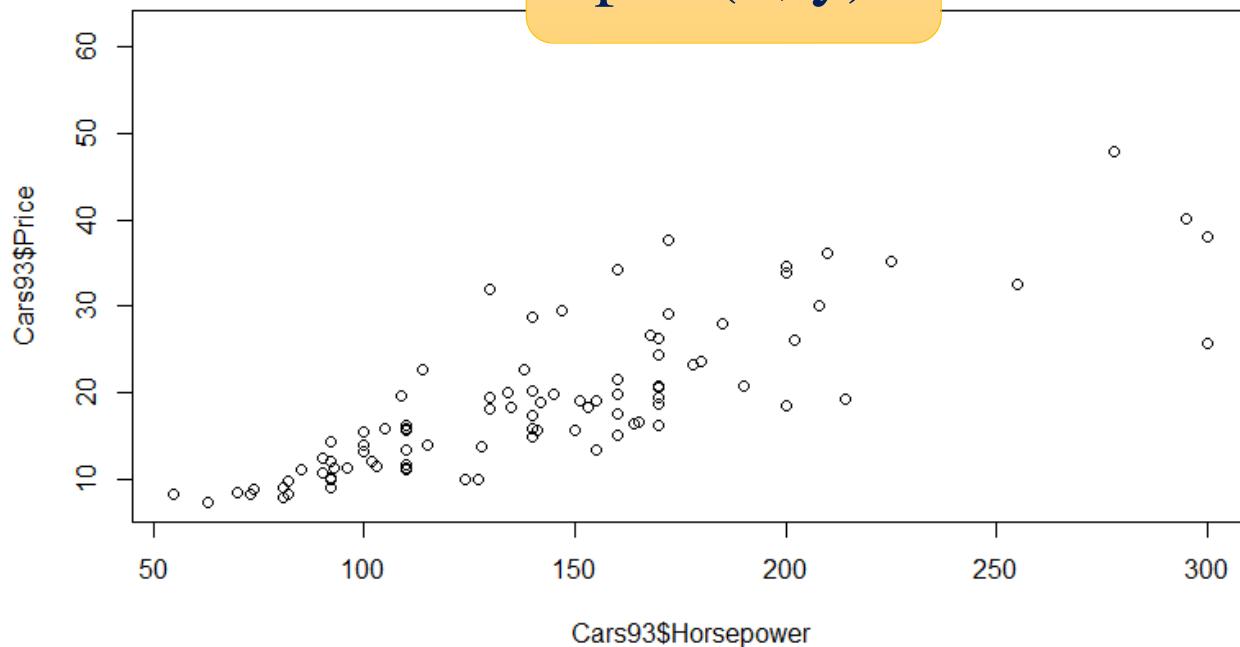
plot.

圖、直線圖、

散佈圖 plot

```
> data(Cars93, package="MASS")
> plot(Cars93$Horsepower, Cars93$Price)
```

plot(x, y)



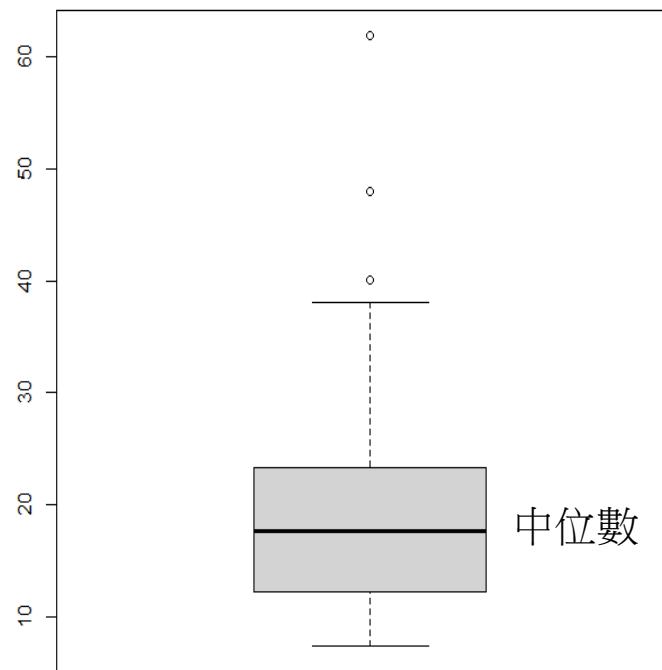
盒鬚圖 boxplot

```
boxplot(Cars93$Price)
```

```
Cars93_Price <- boxplot(Cars93$Price)
```

```
Cars93_Price
```

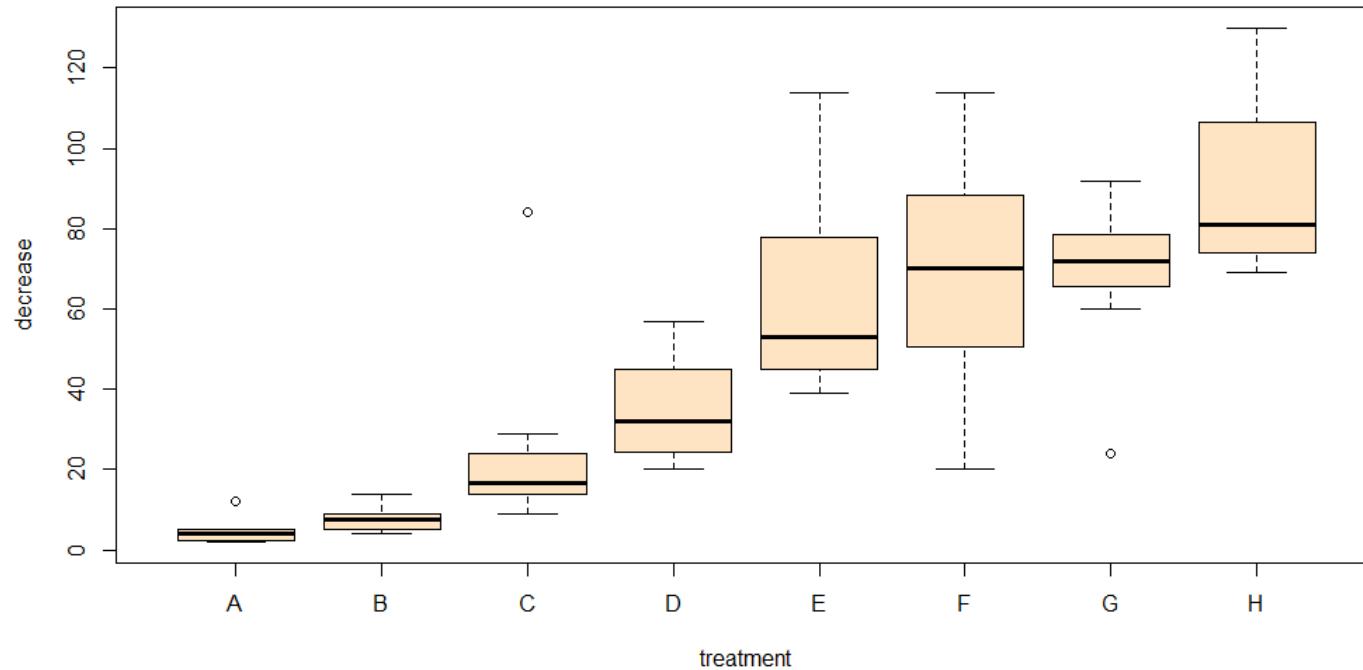
盒鬚圖詳解



```
> # 盒鬚圖 boxplot
> data(Cars93, package = "MASS")
> boxplot(Cars93$price)
> Cars93_Price <- boxplot(Cars93$Price)
> Cars93_Price
$stats
[1,] 7.4 下限
[2,] 12.2 Q1
[3,] 17.7 上限離群值:  $x > Q_3 + (Q_3 - Q_1) \times 1.5$ 
[4,] 23.3 Q3
[5,] 38.0 上下限離群值:  $x < Q_1 - (Q_3 - Q_1) \times 1.5$ 
$n
[1] 93
$conf
[1,]
[1,] 15.88139
[2,] 19.51861
$out
[1] 40.1 47.9 61.9
$group
[1] 1 1 1
$names
[1] "1"
```

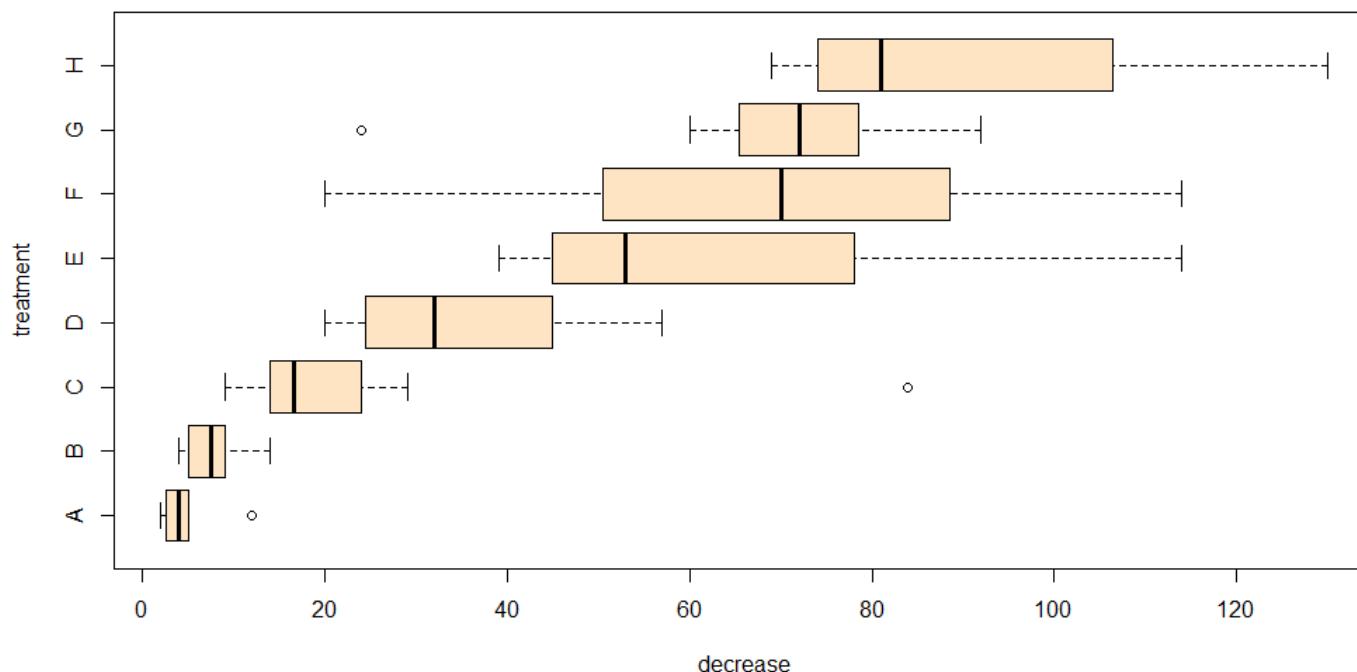
群組盒鬚圖

```
boxplot(decrease ~ treatment,  
        data = orchardsprays,  
        col = "bisque")
```



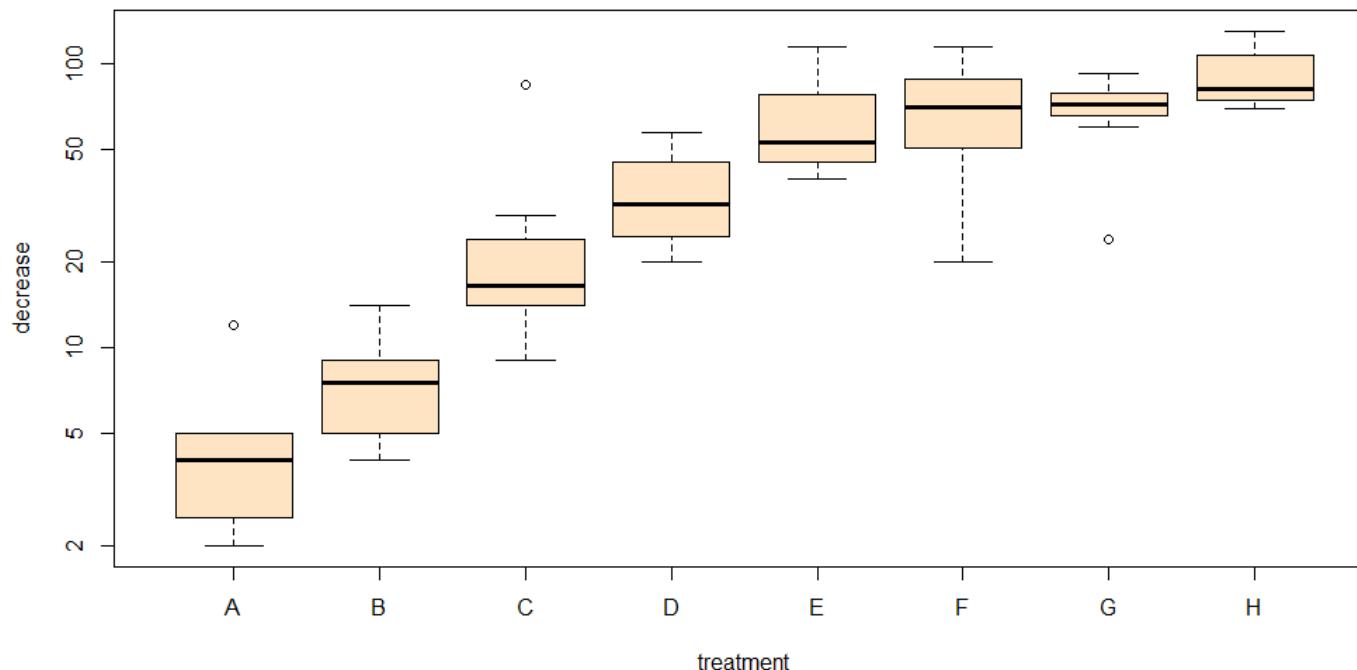
水平群組盒鬚圖

```
boxplot(decrease ~ treatment,  
        data = OrchardSprays,  
        col = "bisque",  
horizontal=TRUE)
```



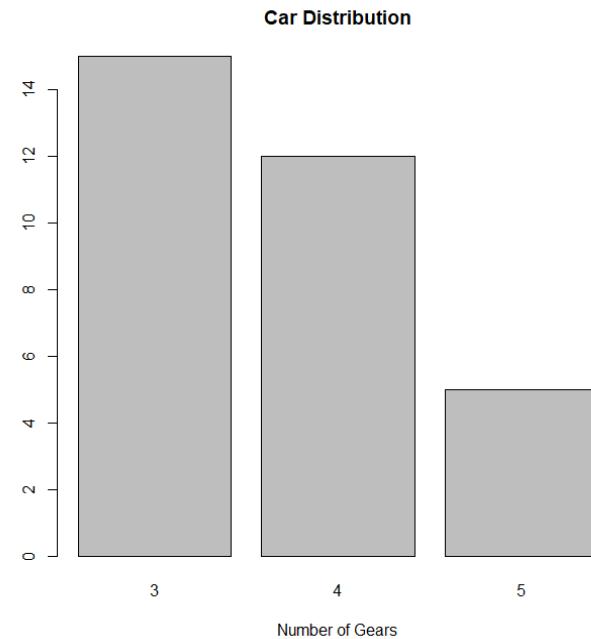
群組盒鬚圖-Y軸取log值

```
boxplot(decrease ~ treatment,  
        data = OrchardSprays,  
        col = "bisque",  
log = "y")
```

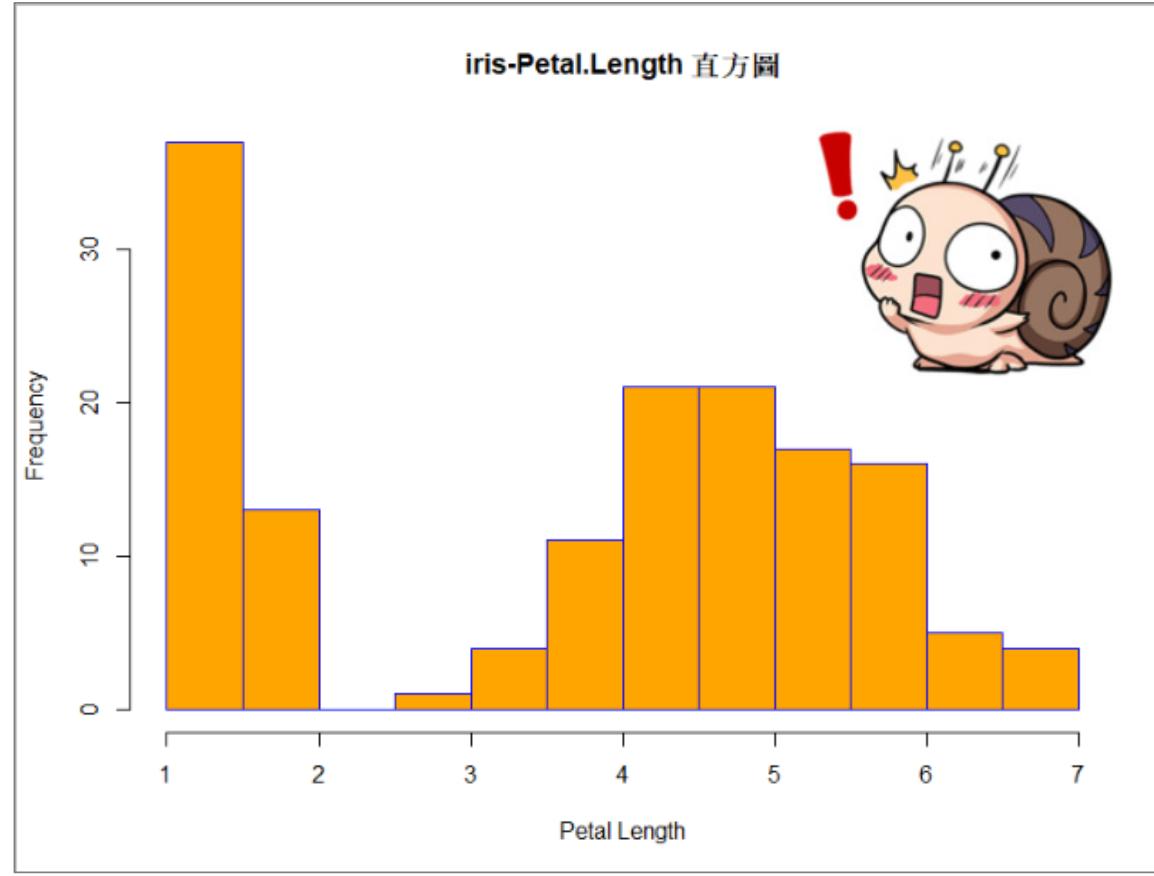


長條圖 barplot

```
counts <- table(mtcars$gear)
barplot(counts,
        main="Car Distribution",
        xlab="Number of Gears")
```



直方圖 hist

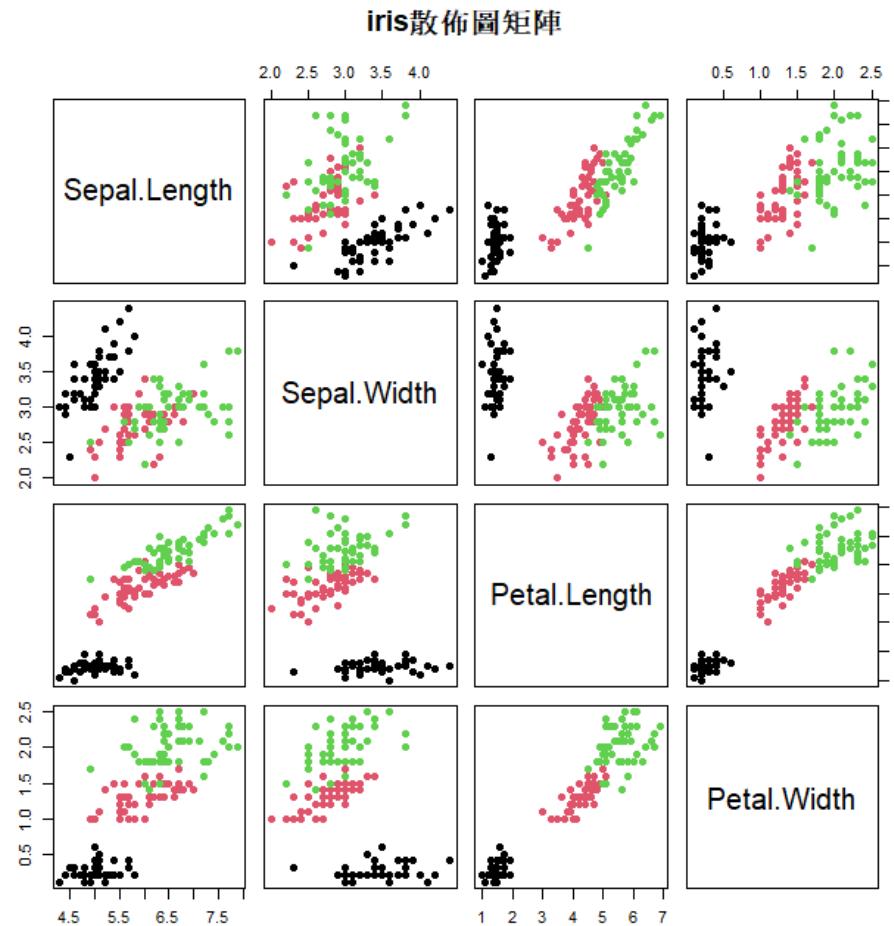


直方圖優化



散佈圖矩陣 pairs

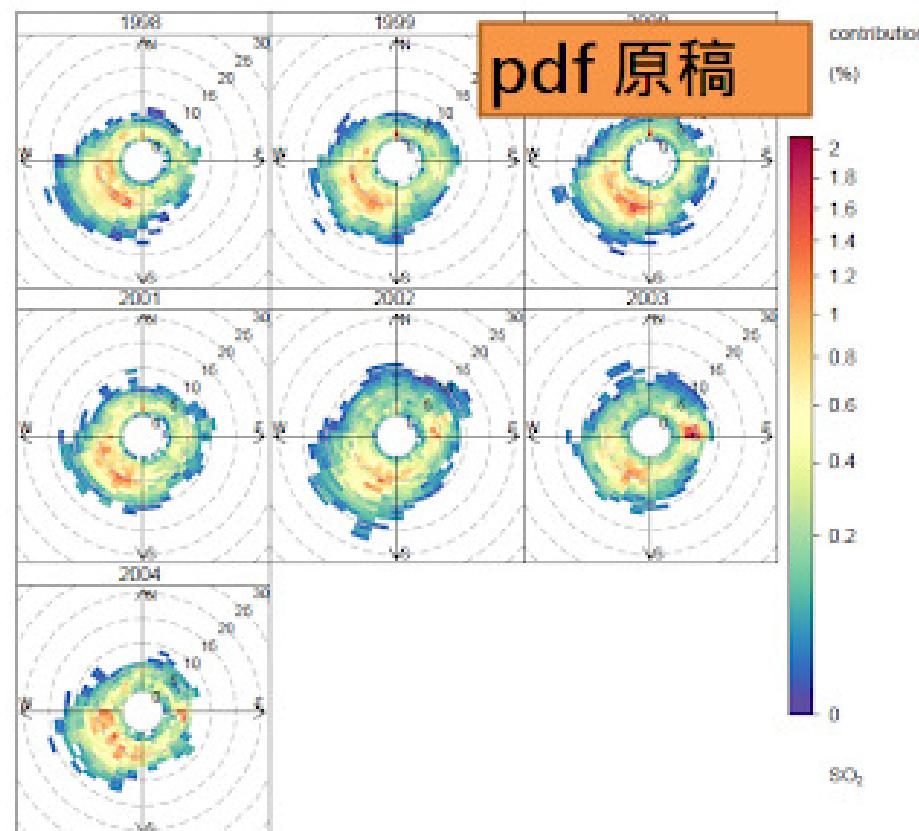
實作練習21



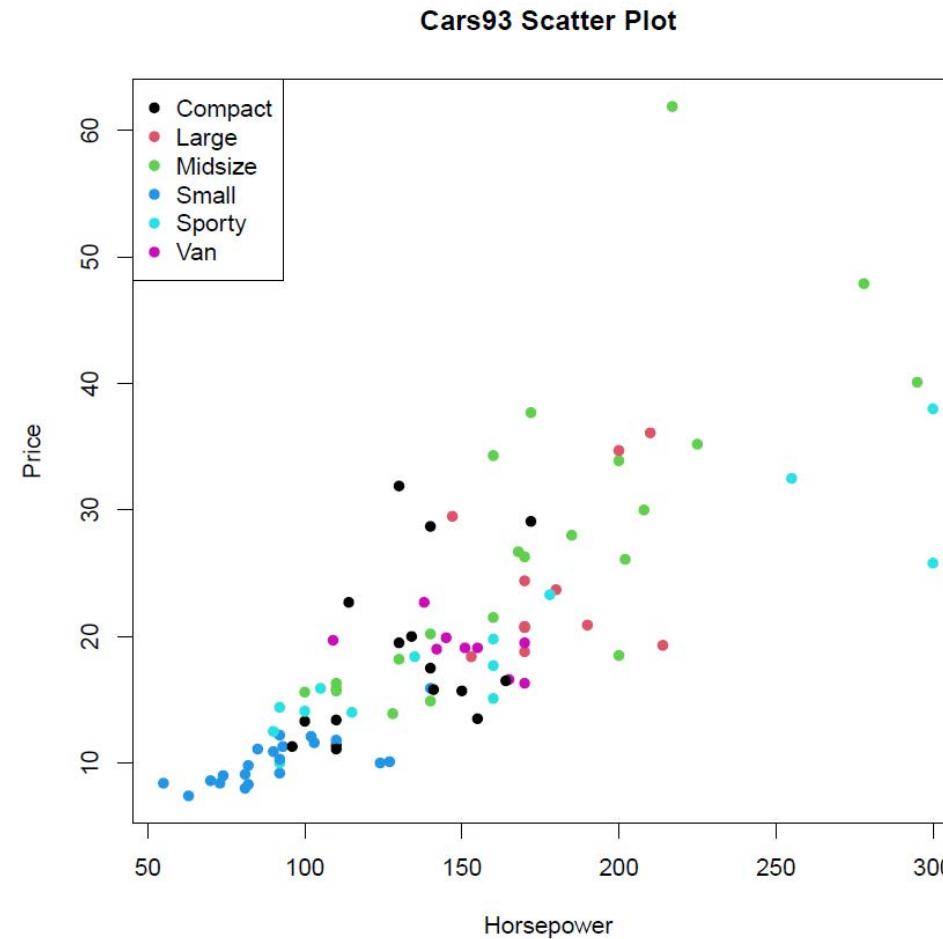
繪圖結果輸出

R的圖形輸出

- <http://rwepa.blogspot.com/2013/05/r.html>



繪圖結果輸出 - 範例



四大繪圖類型

- 四大繪圖類型
 - 比較 Comparison
 - 組成 Composition
 - 分配 Distribution
 - 關係 Relationship

參考資料: <http://www.tatvic.com/blog/7-visualizations-learn-r/>

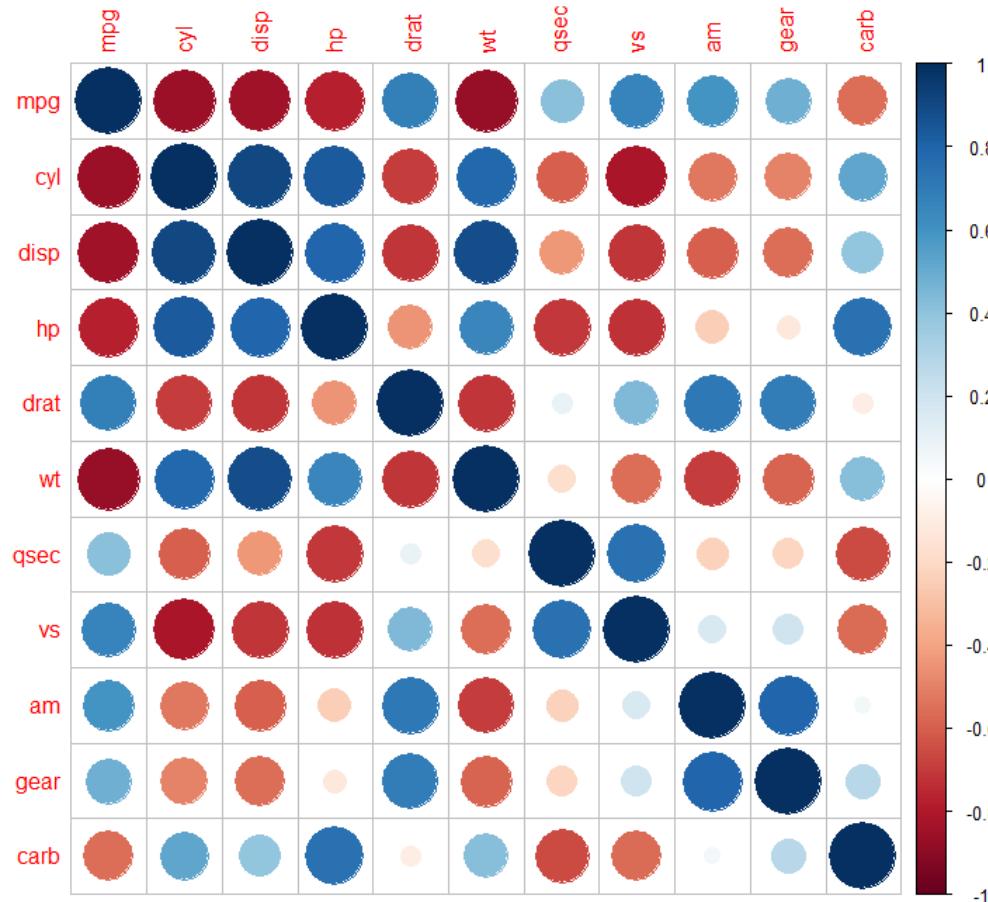
視覺化圖表 <https://datavizcatalogue.com/>



九大圖形

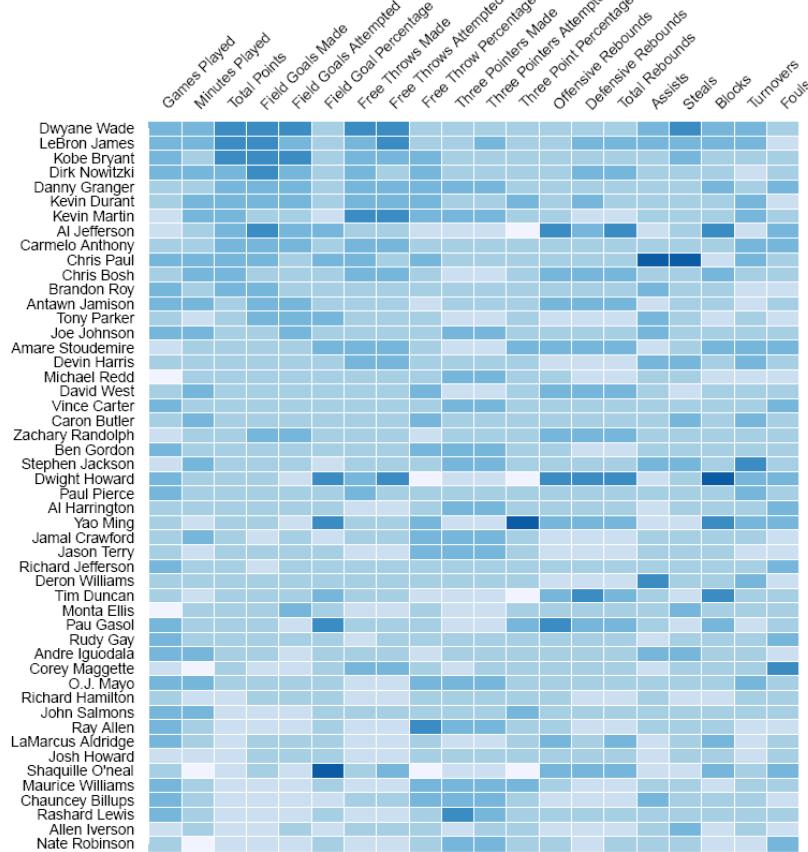
1. Scatter Plot 散佈圖 plot
2. Line Plot 線圖 plot
3. Histogram 直方圖 hist
4. Bar Chart 長條圖 barplot
5. Box Plot 盒鬚圖 boxplot
6. Pie Chart 圓形圖 pie
7. Heat Map 熱繪圖 image
8. Correlation 相關圖 corrplot {corrplot}
9. Interactive plot 互動式繪圖 {shiny}

相關圖 - corrplot 套件

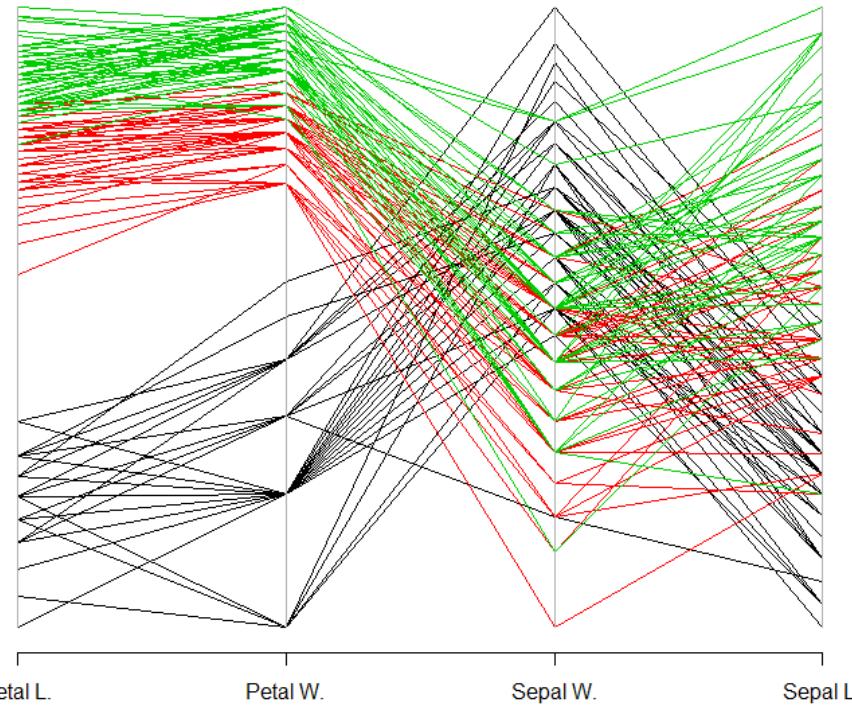


熱繪圖 image {graphics}

NBA per game performance of top 50 scorers



平行座標軸 parcoord {MASS}



個案討論

GFC案例

資料匯入

```
# https://github.com/rwepa/DataDemo/blob/master/gfc.csv
```

```
gfc <- read.table("gfc-ansi.csv", head=TRUE, sep=",")
```

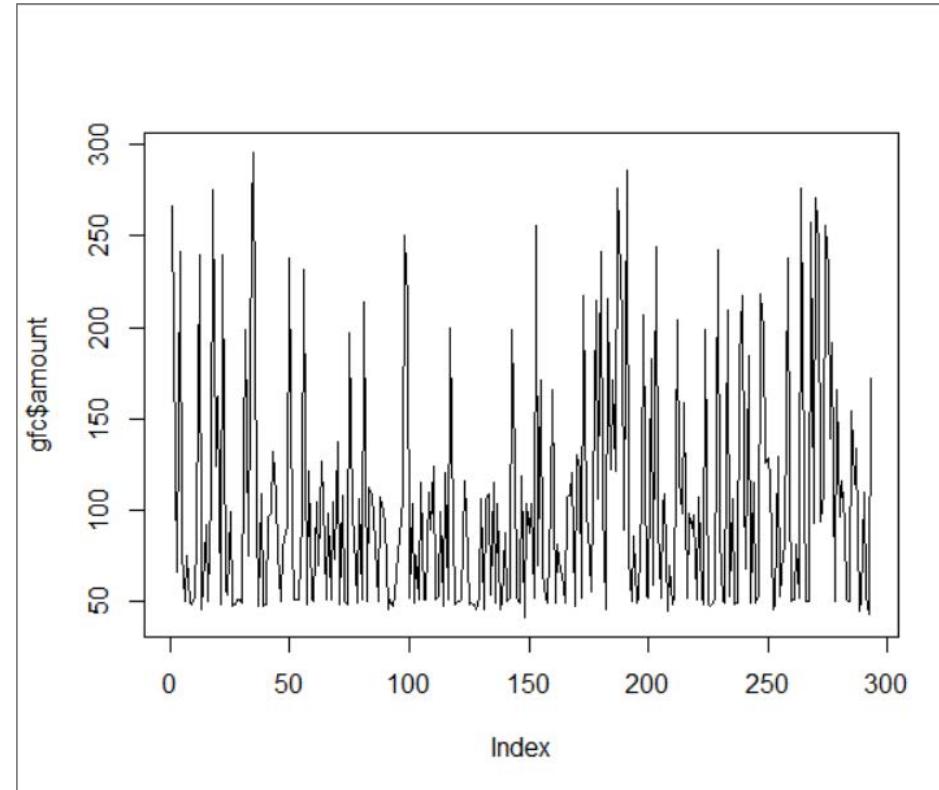
```
head(gfc)
```

```
str(gfc)
```

```
summary(gfc)
```

	orderdate	supplier	amount
1	2009/1/3	日亞	266
2	2009/1/4	廣達	123
3	2009/1/5	廣達	66
4	2009/1/6	日亞	241
5	2009/1/7	廣達	77
6	2009/1/8	科銳	50

繪圖



有什麼 Patterns?



08. 資料操作dplyr, 視覺化ggplot2, 互動式表格視覺化

資料操作 dplyr

- A Grammar of Data Manipulation 資料操作文法
- dplyr = data frame + plyr

dplyr 操作

- filter : 條件式篩選資料
- slice : 列的指標篩選資料
- arrange : 排序
- select : 選取行/更改欄位名稱
- rename : 選取所有行/更改欄位名稱
- distinct : 選取不重覆資料
- mutate : 新增欄位,保留原資料
- transmute : 新增欄位,不留原資料
- summarise : 群組計算

dplyr –tbl_df 類別

```
library(dplyr)

library(nycflights13) # 2013年NYC機場航班資料，33萬筆資料 ----

flights # 336776*19

class(flights) # "tbl_df" "tbl" "data.frame"
> # 如何轉換為tbl_df, 使用as.tbl -----
> mytbl <- as.tbl(iris) # deprecated in dplyr 1.0.0.
Warning message:
`as.tbl()` was deprecated in dplyr 1.0.0.
Please use `tibble::as_tibble()` instead.
This warning is displayed once every 8 hours.
Call `lifecycle::last_warnings()` to see where this warning was generated.
> mytbl <- tibble::as_tibble(iris)
>
> class(mytbl)
[1] "tbl_df"      "tbl"        "data.frame"
```

R demo

tidyverse 套件 – 長寬資料轉換

tidyr 長寬資料轉換套件

tidyr: Easily Tidy Data with 'spread()' and 'gather()' Functions

An evolution of 'reshape2'. It's designed specifically for data tidying (not general reshaping or aggregating) and works well with 'dplyr' data pipelines.

Version:	0.6.0
Depends:	R (\geq 3.1.0)
Imports:	tibble , dplyr (\geq 0.4), stringi , lazyeval , magrittr , Rcpp
LinkingTo:	Rcpp
Suggests:	knitr , testthat , covr , gapminder , rmarkdown
Published:	2016-08-12
Author:	Hadley Wickham [aut, cre], RStudio [cph]
Maintainer:	Hadley Wickham <hadley at rstudio.com>
BugReports:	https://github.com/hadley/tidyr/issues
License:	MIT + file LICENSE
URL:	https://github.com/hadley/tidyr
NeedsCompilation:	yes
Materials:	README NEWS
CRAN checks:	tidyr results

長寬資料轉換

- `tidyverse` 套件
 - `gather` : From wide to long
 - `spread` : From long to wide
- `reshape2` 套件
 - `melt` : From wide to long
 - `dcast` : From long to wide

See http://www.cookbook-r.com/Manipulating_data/Converting_data_between_wide_and_long_format/

長寬資料轉換 – 範例

```
> olddata_wide
  subject sex control cond1 cond2
  1       1   M     7.9  12.3  10.7
  2       2   F     6.3  10.6  11.1
  3       3   F     9.5  13.1  13.8
  4       4   M    11.5  13.4  12.9
> olddata_long
  subject sex condition measurement
  1       1   M      control      7.9
  2       1   M      cond1      12.3
  3       1   M      cond2      10.7
  4       2   F      control      6.3
  5       2   F      cond1      10.6
  6       2   F      cond2      11.1
  7       3   F      control      9.5
  8       3   F      cond1      13.1
  9       3   F      cond2      13.8
 10      4   M      control     11.5
 11      4   M      cond1      13.4
 12      4   M      cond2      12.9
>
```

tidyR – From wide to long

```
> library(tidyR)
> # gather: From wide to long
> data_long <- gather(olddata_wide, condition, measurement, control:cond2)
> data_long
```

	subject	sex	condition	measurement	
1	1	M	control	7.9	
2	2	F	control	6.3	
3	3	F	control	9.5	
4	4	M	control	11.5	
5	1	M	cond1	12.3	> olddata_wide
6	2	F	cond1	10.6	subject sex control cond1 cond2
7	3	F	cond1	13.1	1 1 M 7.9 12.3 10.7
8	4	M	cond1	13.4	2 2 F 6.3 10.6 11.1
9	1	M	cond2	10.7	3 3 F 9.5 13.1 13.8
10	2	F	cond2	11.1	4 4 M 11.5 13.4 12.9
11	3	F	cond2	13.8	
12	4	M	cond2	12.9	

1

2

3

Key, Value

寬欄位

tidyverse – From long to wide

```
> data_wide <- spread(olddata_long, condition, measurement)  
> data_wide
```

	subject	sex	cond1	cond2	control
1	1	M	12.3	10.7	7.9
2	2	F	10.6	11.1	6.3
3	3	F	13.1	13.8	9.5
4	4	M	13.4	12.9	11.5



Key, Value

	subject	sex	condition	measurement
1	1	M	control	7.9
2	1	M	cond1	12.3
3	1	M	cond2	10.7
4	2	F	control	6.3
5	2	F	cond1	10.6
6	2	F	cond2	11.1
7	3	F	control	9.5
8	3	F	cond1	13.1
9	3	F	cond2	13.8
10	4	M	control	11.5
11	4	M	cond1	13.4
12	4	M	cond2	12.9

字串處理套件 - stringr 套件

stringr 字串處理

stringr: Simple, Consistent Wrappers for Common String Operations

A consistent, simple and easy to use set of wrappers around the fantastic 'stringi' package. All function and argument names (and positions) are consistent, all functions deal with "NA"s and zero length vectors in the same way, and the output from one function is easy to feed into the input of another.

Version:	1.4.0
Depends:	R (\geq 3.1)
Imports:	glue (\geq 1.2.0), magrittr , stringi (\geq 1.1.7)
Suggests:	covr , htmltools , htmlwidgets , knitr , rmarkdown , testthat
Published:	2019-02-10
Author:	Hadley Wickham [aut, cre, cph], RStudio [cph, fnd]
Maintainer:	Hadley Wickham <hadley at rstudio.com>
BugReports:	https://github.com/tidyverse/stringr/issues

Reference manual: [stringr.pdf](#)

Vignettes:

[Regular expressions](#)

[Introduction to stringr](#)



判斷字串是否符合條件

- **str_detect()** detects the presence or absence of a pattern and returns a logical vector (similar to **grepl()**).
- **str_subset()** returns the elements of a character vector that match a regular expression (similar to **grep()** with value = TRUE)`.

str_detect, str_subset 範例1

```
> strings
[1] "apple"                      "219 733 8965"
[3] "329-293-8753"                "Work: 579-499-7527; Home: 543.355.3679"
> phone
[1] "([2-9][0-9]{2})[- .]([0-9]{3})[- .]([0-9]{4})"

> library(stringr)
> # stringr
> strings <- c(
+ "apple",
+ "219 733 8965",
+ "329-293-8753",
+ "Work: 579-499-7527; Home: 543.355.3679"
+ )
> phone <- "([2-9][0-9]{2})[- .]([0-9]{3})[- .]([0-9]{4})"
> # which strings contain phone numbers?
> str_detect(strings, phone)
[1] FALSE TRUE TRUE TRUE
> str_subset(strings, phone)
[1] "219 733 8965"                  "329-293-8753"
[3] "Work: 579-499-7527; Home: 543.355.3679"
>
```

str_detect, str_subset 範例2

```
> strings
[1] "apple"                               "219 733 8965"
[3] "329-293-8753"                         "work: 579-499-7527; Home: 543.355.3679"
> phone
[1] "([2-9][0-9]{2})[- .]([0-9]{3})[- .]([0-9]{4})"

> # str_detect, str_subset 範例2
> fruit <- c("apple", "banana", "pear", "pinapple", "ndc")
> str_detect(fruit, "a") # 包括字母a
[1] TRUE TRUE TRUE TRUE FALSE
> str_detect(fruit, "^a") # 開頭是字母a
[1] TRUE FALSE FALSE FALSE FALSE
> str_detect(fruit, "a$") # 結尾是字母a
[1] FALSE TRUE FALSE FALSE FALSE
> str_detect(fruit, "b") # 包括字母b
[1] FALSE TRUE FALSE FALSE FALSE
> str_detect(fruit, "pp") # 包括字母pp
[1] TRUE FALSE FALSE TRUE FALSE
> str_detect(fruit, "[aeiou]") # 包括字母a 或 e 或 i 或 o 或 u
[1] TRUE TRUE TRUE TRUE FALSE
```

尋找字串位址

- `str_locate()` locates the first position (第1個位置) of a pattern and returns a numeric matrix with columns start and end.
- `str_locate_all()` locates all matches (所有位置), returning a list of numeric matrices.
- Similar to `regexp()` and `gregexpr()`.

str_locate, str_locate_all 範例

```
> strings
[1] "apple"                               "219 733 8965"
[3] "329-293-8753"                         "work: 579-499-7527; Home: 543.355.3679"
> phone
[1] "([2-9][0-9]{2})[- .]([0-9]{3})[- .]([0-9]{4})"
> # Where in the string is the phone number located?
> (loc <- str_locate(strings, phone))
      start end
[1,]     NA  NA
[2,]     1 12
[3,]     1 12
[4,]     7 18
> str_locate_all(strings, phone)
[[1]]
      start end
[[2]]
      start end
[1,]     1 12
[[3]]
      start end
[1,]     1 12
[[4]]
      start end
[1,]     7 18
[2,]    27 38
```

選取(部份)字串

- **str_extract()** extracts text corresponding to the first match, returning a character vector.
- **str_extract_all()** extracts all matches and returns a list of character vectors.

R demo

選取(部份)群組字串-回傳矩陣

- **str_match()** extracts capture groups formed by () from the first match. It returns a character matrix with one column for the complete match and one column for each group.
- **str_match_all()** extracts capture groups from all matches and returns a list of character matrices. Similar to regmatches().

字串取代

- **str_replace()** replaces the first matched pattern and returns a character vector.
- **str_replace_all()** replaces all matches.
- Similar to sub() and gsub().

正規表示式 Regular expression

- ?regex
- http://stat545.com/block022_regular-expression.html



?regex

regex {base}

R Documentation

Regular Expressions as used in R

Description

This help page documents the regular expression patterns supported by [grep](#) and related functions [grepl](#), [regexpr](#), [gregexpr](#), [sub](#) and [gsub](#), as well as by [strsplit](#) and optionally by [agrep](#) and [agrepl](#).

Details

A ‘regular expression’ is a pattern that describes a set of strings. Two types of regular expressions are used in R, *extended* regular expressions (the default) and *Perl-like* regular expressions used by `perl = TRUE`. There is also `fixed = TRUE` which can be considered to use a *literal* regular expression.

STAT 545

- <https://stat545.com/character-vectors.html>

Chapter II Character vectors



II.1 Character vectors: where they fit in

We've spent a lot of time working with big, beautiful data frames. That are clean and wholesome, like the Gapminder data.

But real life will be much nastier. You will bring data into R from the outside world and discover there are problems. You might think: how hard can it be to deal with character data? And the answer is: it can be very hard!

- Stack Exchange outage
- Regexes to validate/match email addresses
- Fixing an Atom bug

Here we discuss common remedial tasks for cleaning and transforming character data, also known as "strings". A data frame or tibble will consist of one or more atomic vectors of a certain class. This lesson deals with things you can do with vectors of class `character`.

視覺化 ggplot2

圖形文法繪圖

參考資料 <https://ggplot2-book.org/index.html>

Welcome to ggplot2

- ggplot2是一個以圖層(layers)為基礎的繪圖套件，它實現了 Wilkinson (2005)的繪圖文法(Grammar of Graphics)概念。
- 一個圖形是由數個圖層所組成，其中一層包含了資料(data)
- Wilkinson認為圖形繪製須結合數據與繪製規範，規範並非是圖形視覺效果的名稱(例如：長條圖、散佈圖、直方圖等)。
- 規範應是一組共同決定圖形如何建立的規則 – a grammar of graphics

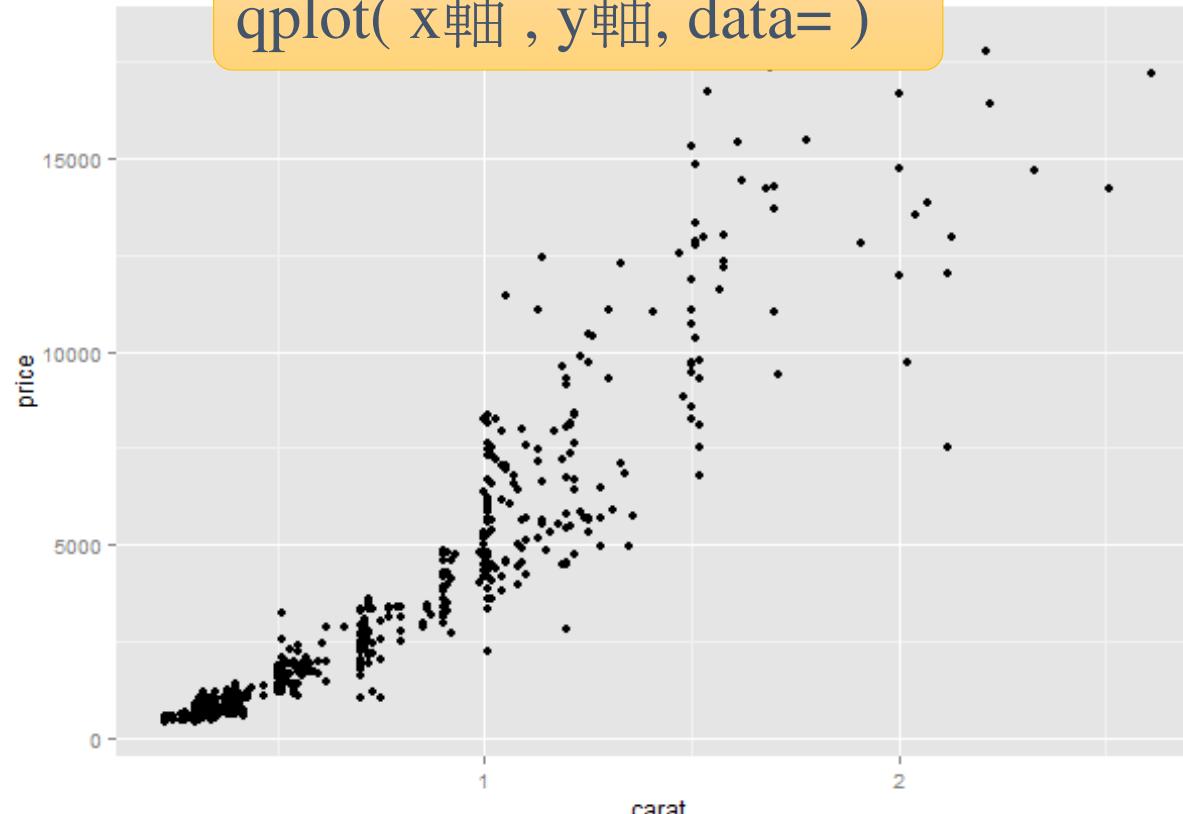
ggplot2核心概念

- 資料 + 畫布 + 圖層 + 規則
- 資料與映射(mapping)
 - 變項到圖形(面)屬性(雙軸、顏色、填滿、尺寸)之間的映射
- 幾何對象(geometric object, geom)
 - 定義展現數據的圖形(點、線、多邊形...)
- 統計變換(statistical transformation, stat)
 - 提煉或歸納原始數據
 - 迴歸直線、平滑曲線、預測區間、信賴區間

qplot: 散佈圖

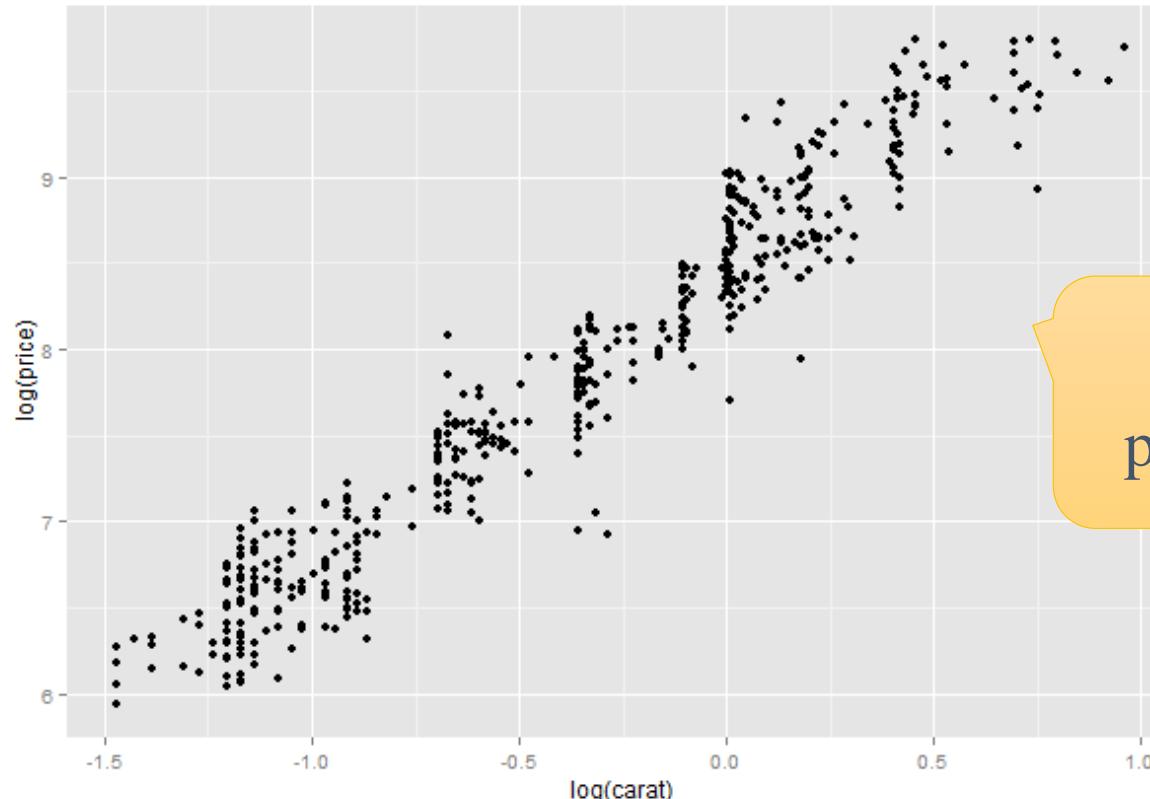
```
qplot(carat, price, data=dsmall)
```

qplot(x軸 , y軸, data=)



qplot: 散佈圖+變數轉換

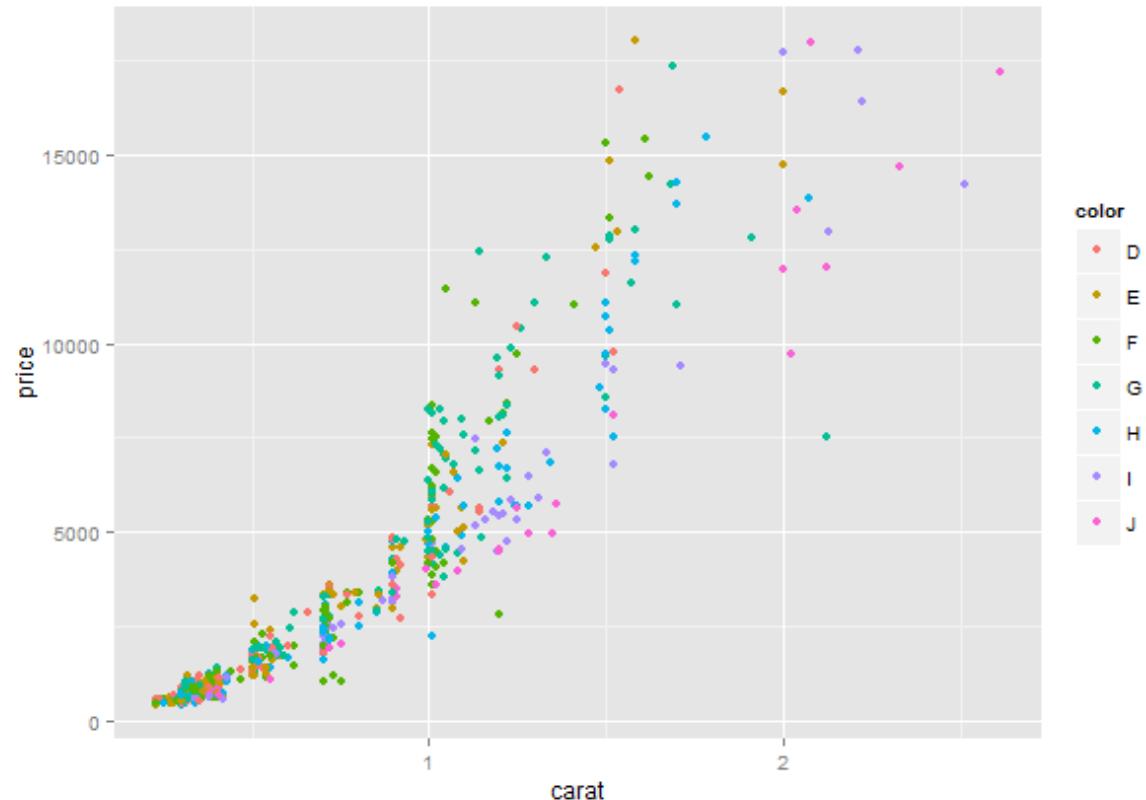
```
qplot(log(carat), log(price), data=dsma11)
```



發現
pattern?

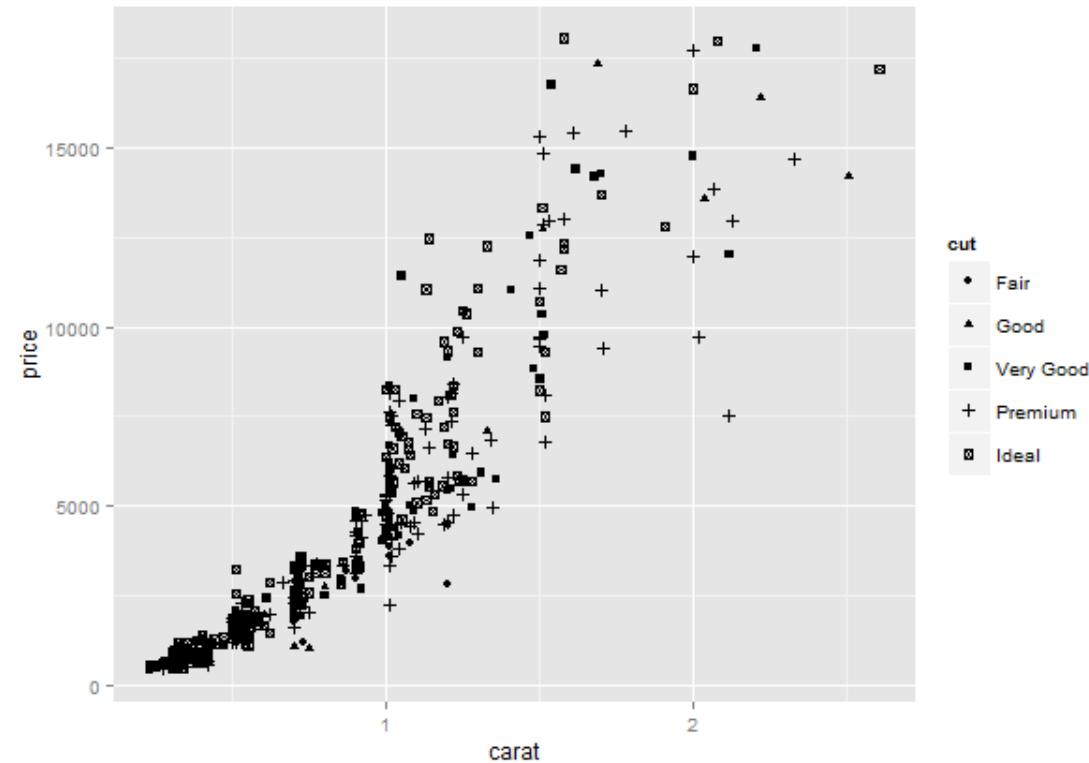
qplot: 散佈圖+群組顏色

```
qplot(carat, price, data=dsmall, colour=color)
```



qplot: 散佈圖+群組形狀

```
qplot(carat, price, data=dsmall, shape=cut)
```



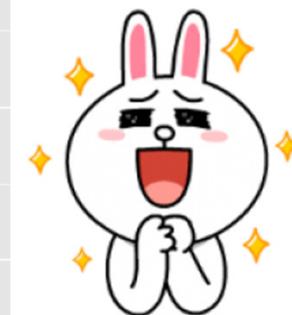
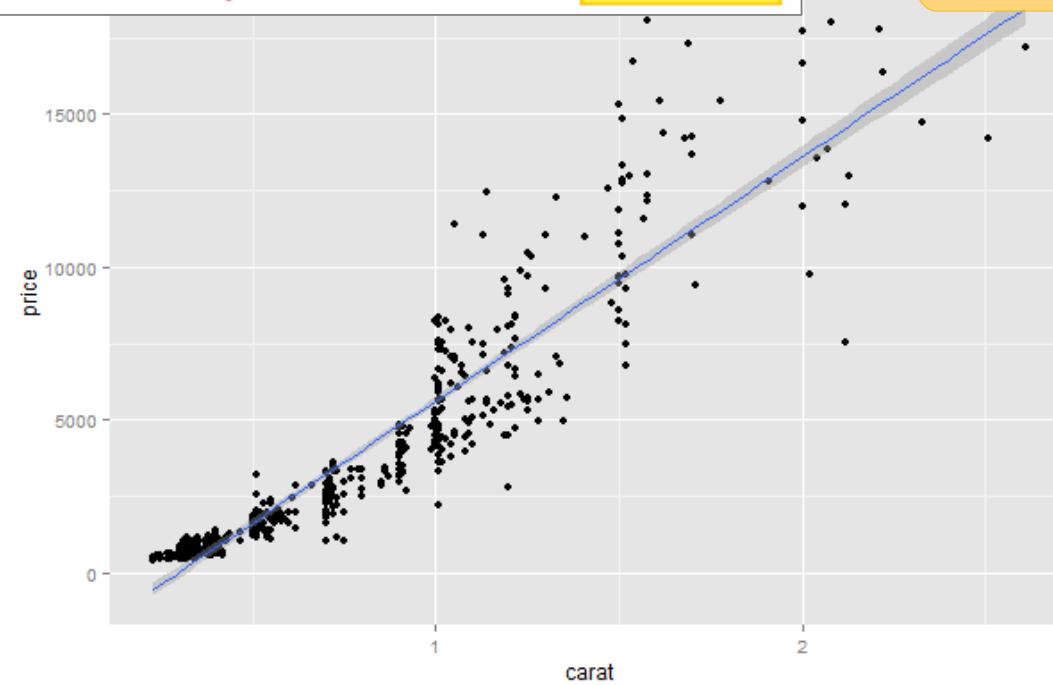
qplot: 散佈圖+迴歸模型

```
p <- qplot(carat, price, data=dsmall,  
           geom=c("point", "smooth"), method="lm")
```

Warning message:

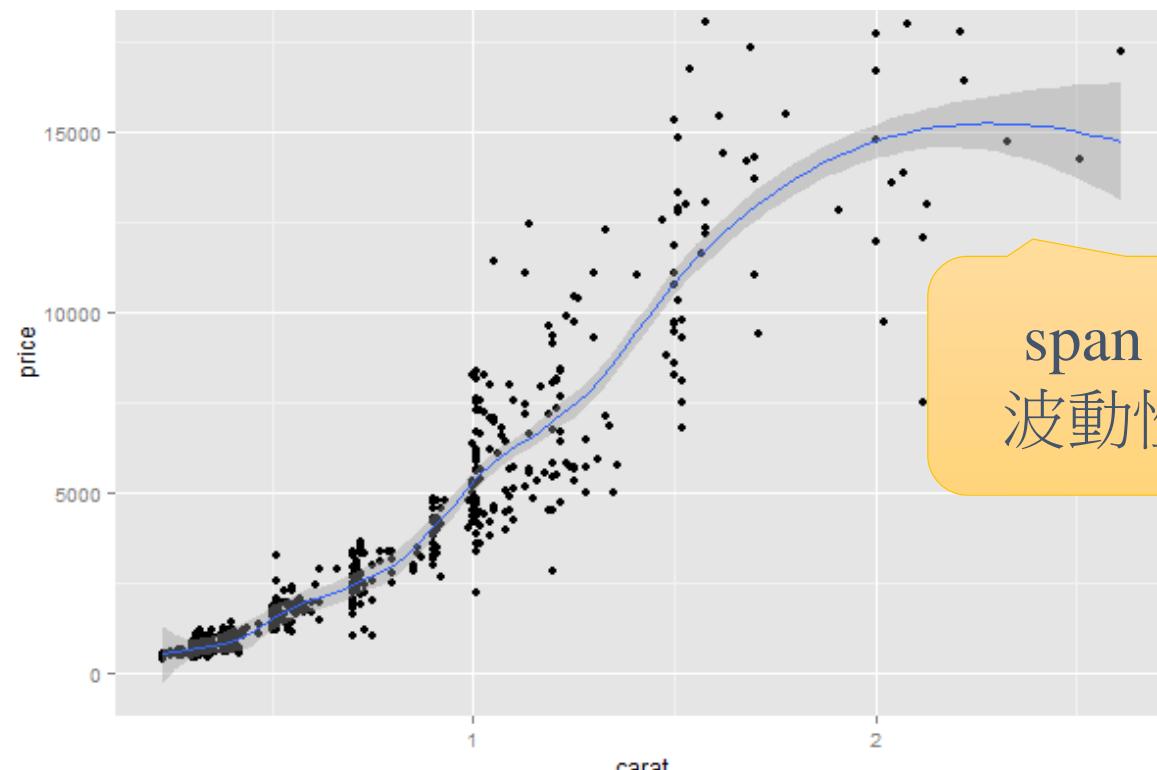
Ignoring unknown parameters: method

revised: stat_smooth()



qplot: 散佈圖+局域加權迴歸線

```
p <- qplot(carat, price, data=dsmall,  
           geom=c("point", "smooth"), span=0.4)  
p
```



span
可以修改

span 愈小，
波動性愈大

ggplot 函數

- 相較於qplot函數, ggplot函數更能展現ggplot2套件的完整功能
- 主要參數:
 - data :指定資料框
 - aes : 提供的美學映射

ggplot 函數 (續)

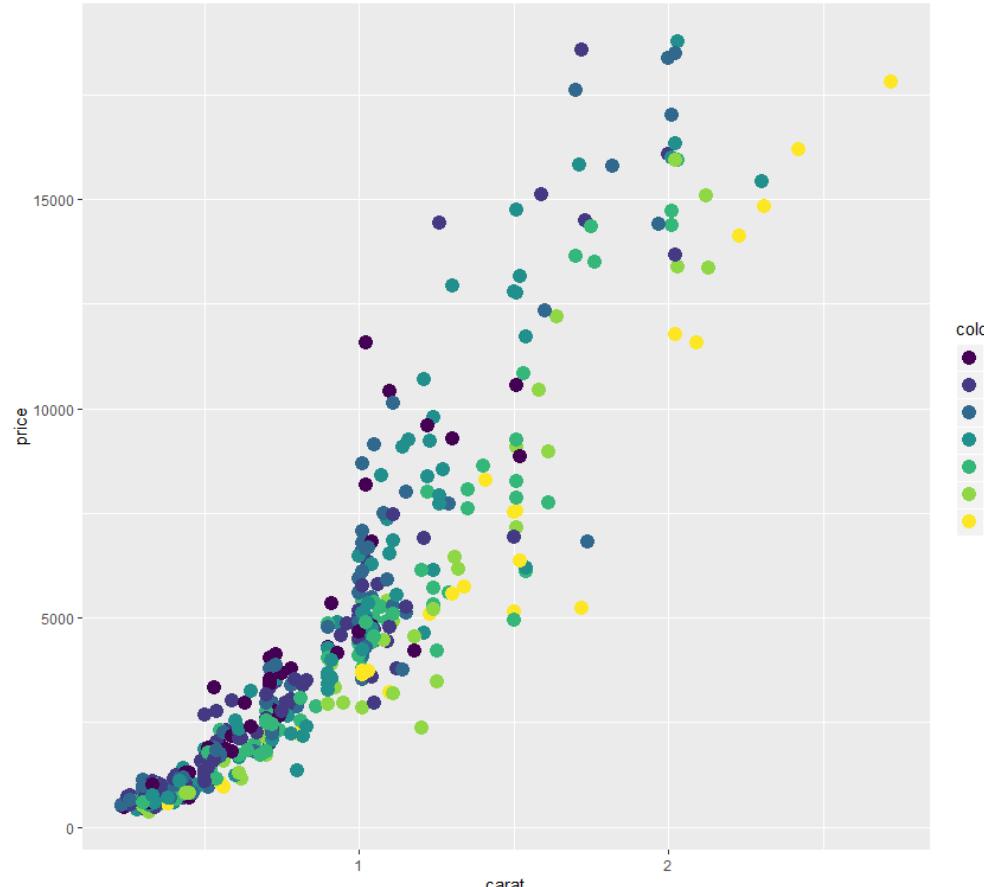
- ggplot 通用語法
 - `ggplot(data, mapping=aes(...)) + geom_*(...) + ... + stat_*(...) + ...`
- 圖層規範
 - `geom_*(mapping, data, ..., geom, position)`
 - `stat_*(mapping, data, ..., stat, position)`
- 額外元件
 - `scales`
 - `coordinates (coord)`
 - `facet`

基礎繪圖 vs. geom

Base function	ggplot2 layer
curve()	geom_curve()
hline()	geom_hline()
lines()	geom_line()
points()	geom_point()
polygon()	geom_polygon()
rect()	geom_rect()
rug()	geom_rug()
segments()	geom_segment()
text()	geom_text()
vline()	geom_vline()
abline(lm(y ~ x))	geom_smooth(method = "lm")
lines(density(x))	geom_density()
lines(loess(x, y))	geom_smooth()

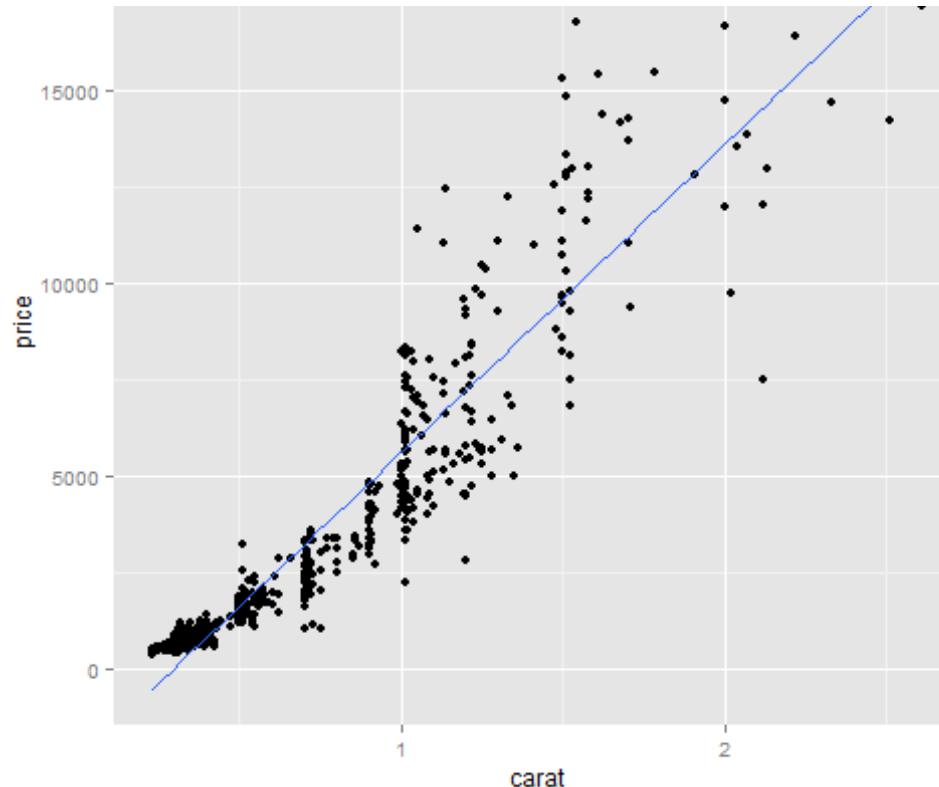
ggplot: 散佈圖

```
p <- ggplot(data=dsmall, mapping=aes(carat, price, color=color)) +  
  geom_point(size=4)  
p
```



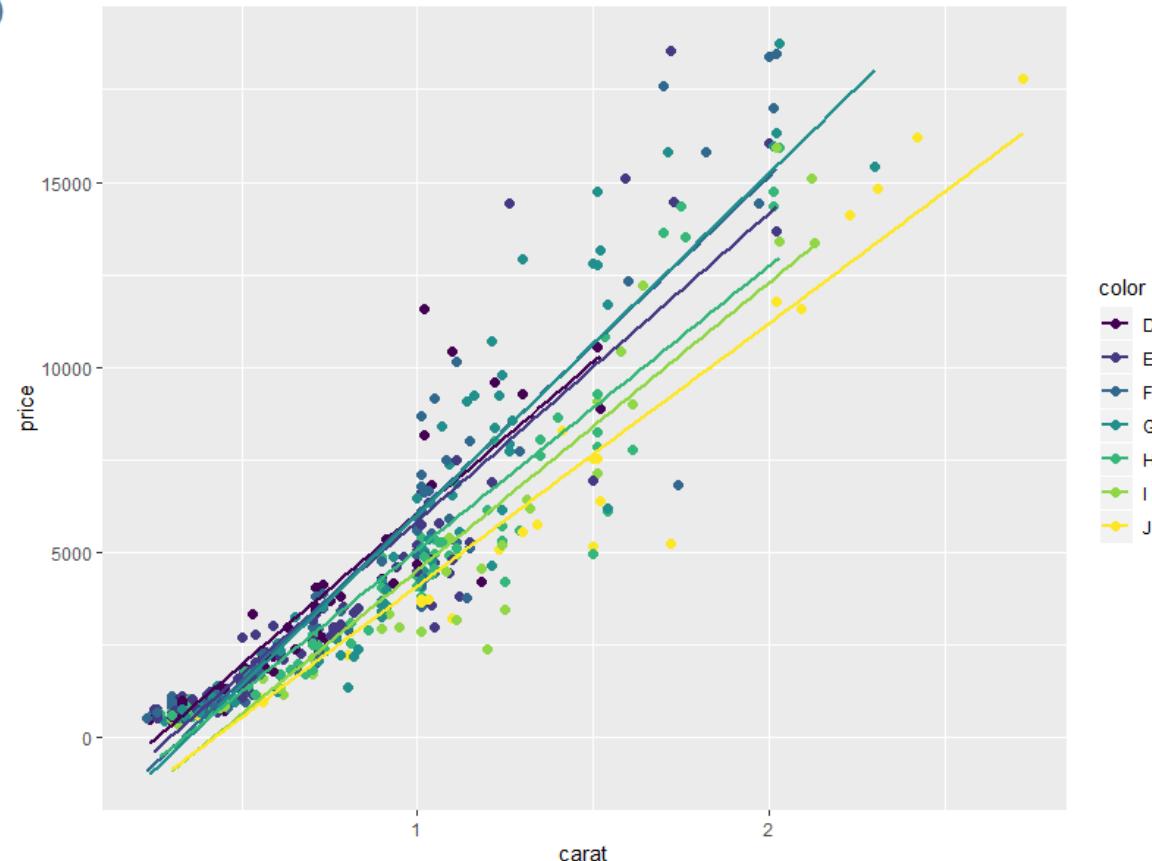
ggplot: 散點圖+線性迴歸

```
p <- ggplot(dsmaill, aes(carat, price)) + geom_point() +  
  geom_smooth(method="lm", se=FALSE)  
p
```



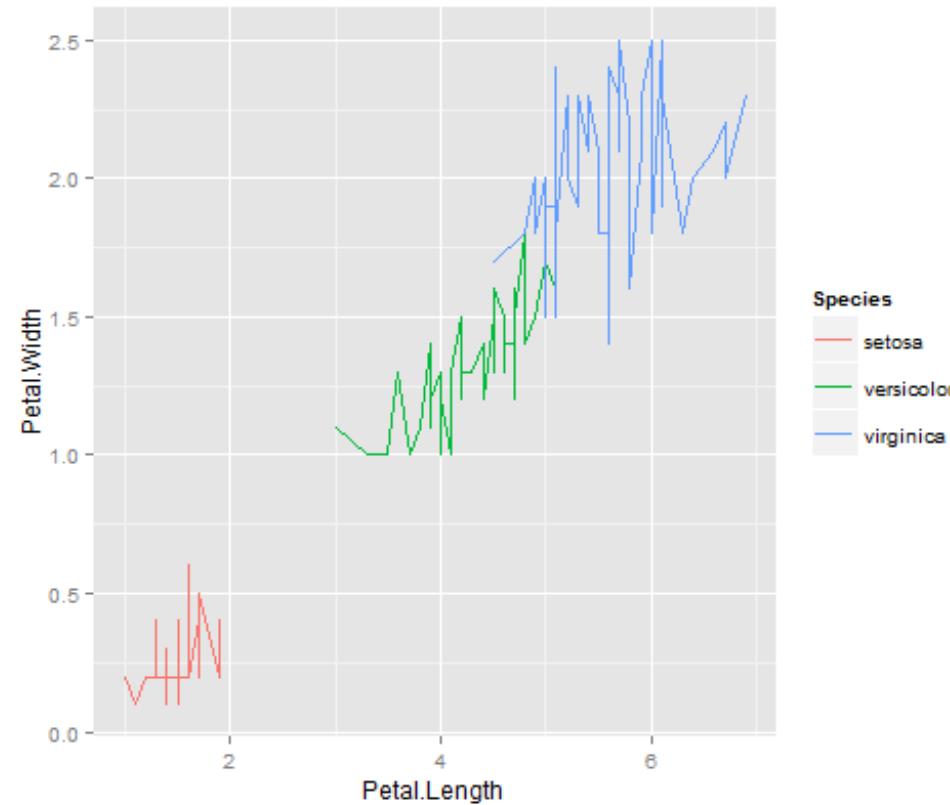
ggplot: 散點圖+群組線性迴歸

```
ggplot(dsmaill, aes(carat, price, group=color)) + geom_point(aes(color=color), size=2) + geom_smooth(aes(color=color), method="lm", se=FALSE)
```



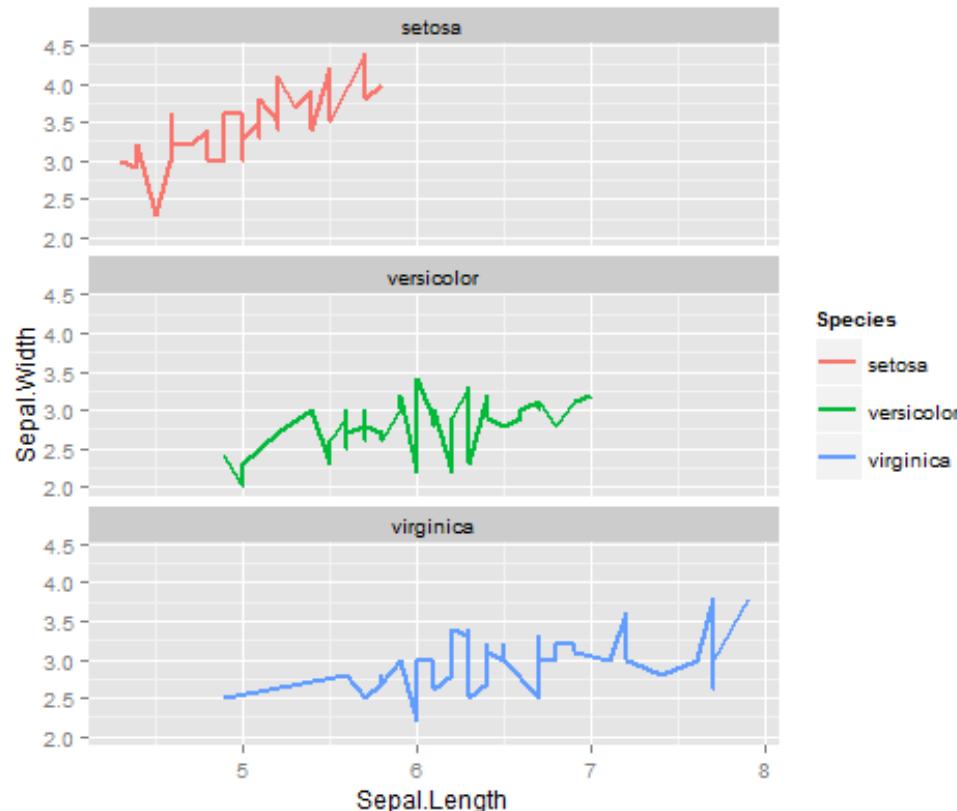
ggplot: 線圖

```
p <- ggplot(iris, aes(Petal.Length, Petal.Width, group=Species,  
color=Species)) + geom_line()  
p
```

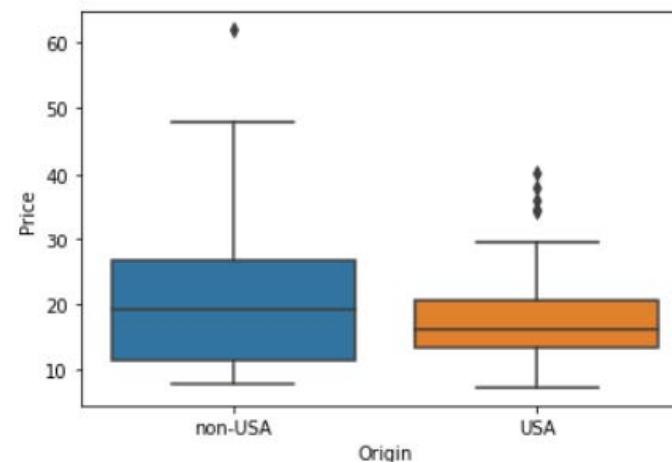


ggplot: 線圖+分面

```
p <- ggplot(iris, aes(Sepal.Length, Sepal.Width)) + geom_line(aes(color=Species), size=1) + facet_wrap(~Species, ncol=1)  
p
```

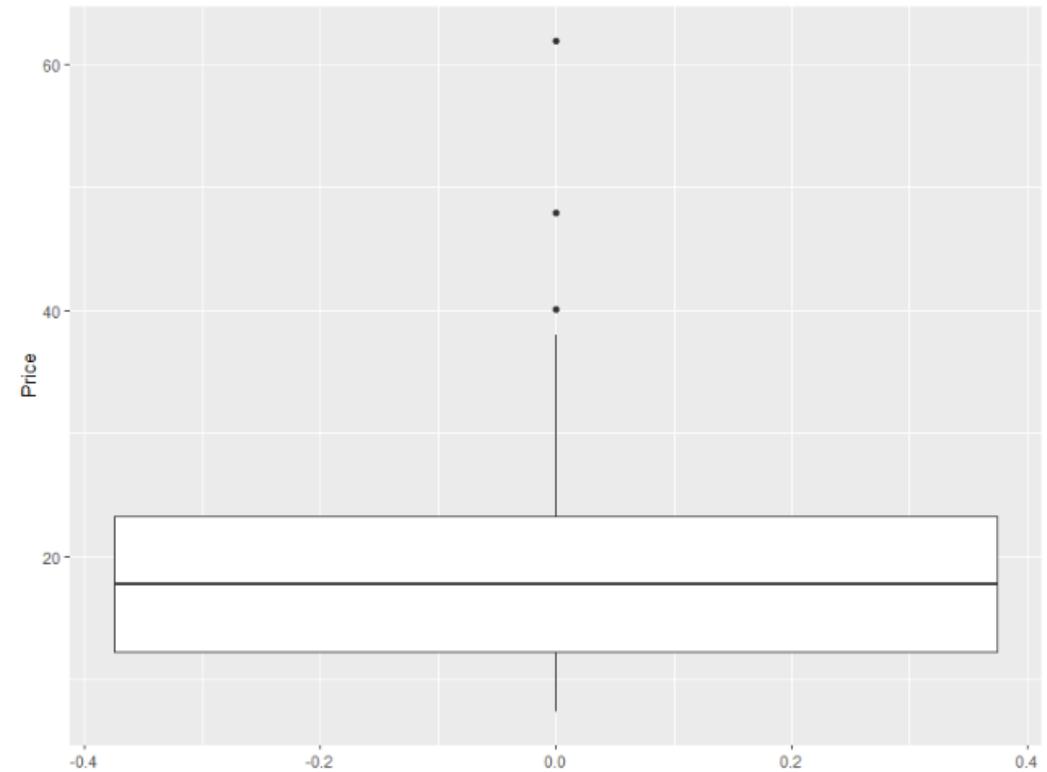


ggplot2 盒鬚圖



盒鬚圖 - ggplot2

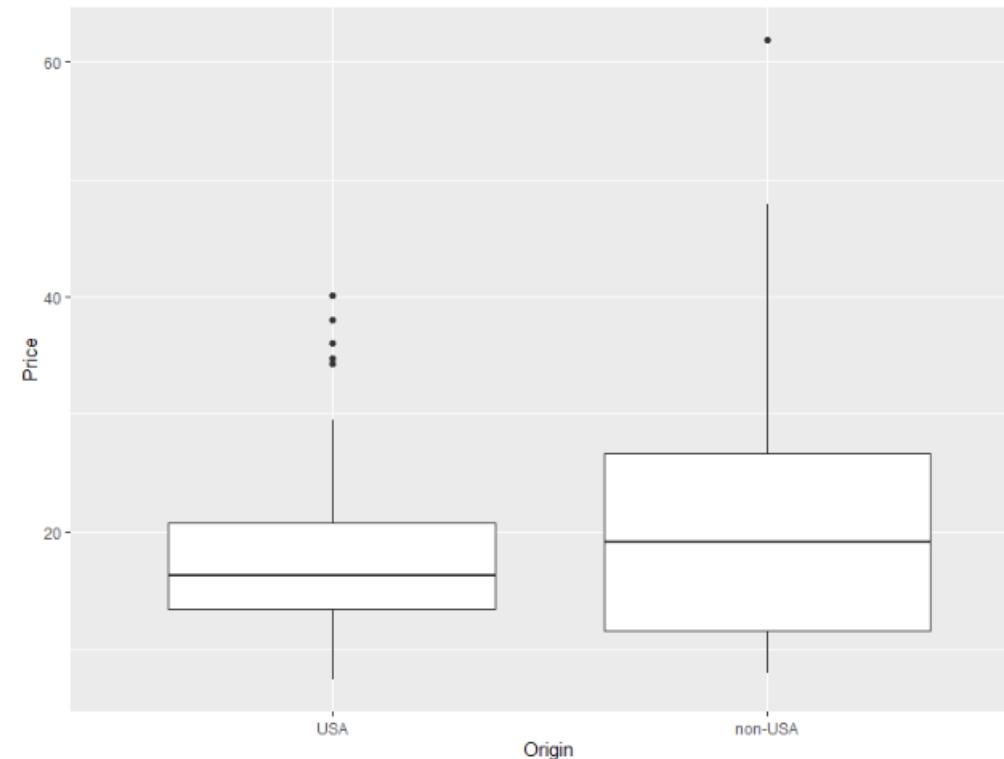
```
# 盒鬚圖 - ggplot2  
  
library(ggplot2)  
p <- ggplot(Cars93, aes(y = Price)) +  
  geom_boxplot()  
p
```



群組盒鬚圖 - ggplot2

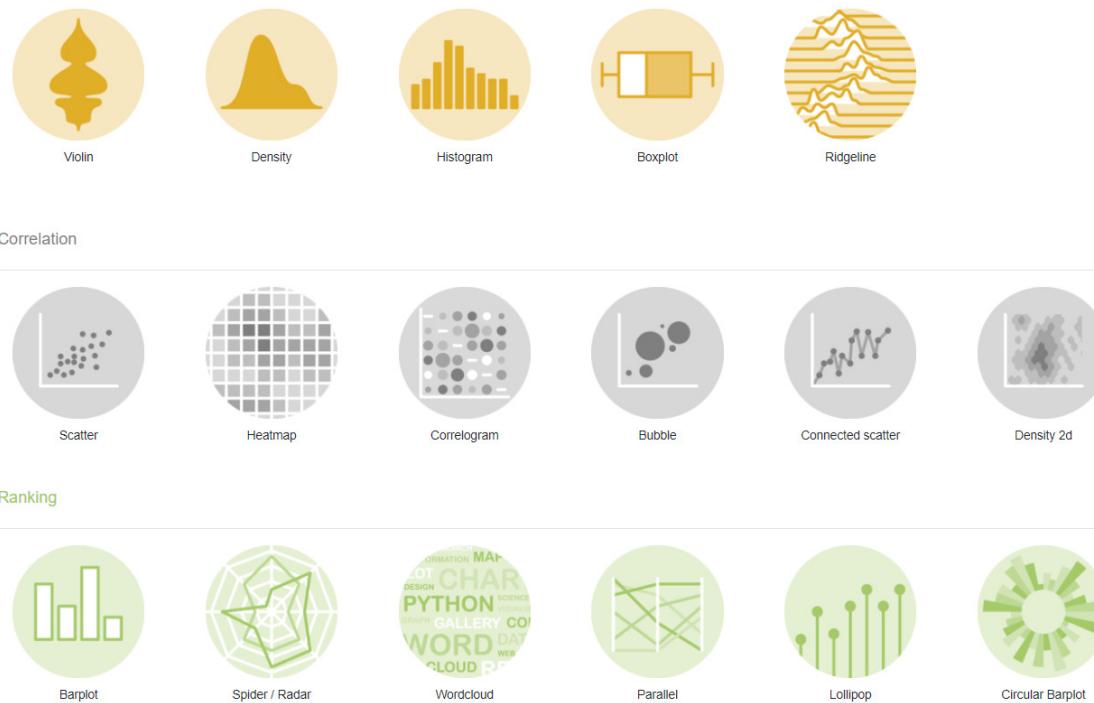
```
# 群組盒鬚圖 - ggplot2  
p <- ggplot(Cars93, aes(x = Origin, y = Price)) +  
  geom_boxplot()
```

```
p
```



ggplot2 gallery

- <https://www.r-graph-gallery.com/>



互動式表格視覺化

kableExtra

reactable

kableExtra 套件

CRAN - Package kableExtra

kableExtra: Construct Complex Table with 'kable' and Pipe Syntax

Build complex HTML or 'LaTeX' tables using 'kable()' from 'knitr' and the piping syntax from 'magrittr'. Function 'kable()' is a light weight table generator coming from 'knitr'. This package simplifies the way to manipulate the HTML or 'LaTeX' codes generated by 'kable()' and allows users to construct complex tables and customize styles using a readable syntax.

Version: 1.3.4
Depends: R (≥ 3.1.0)
Imports: knitr (≥ 1.16), magrittr, stringr (≥ 1.0), xml2 (≥ 1.1.1), rvest, rmarkdown (≥ 1.6.0), scales, viridis, stats, grDevices, htmltools, rstudioapi, glue, tools, webshot, digest, graphics, svglite, testthat, magick, formattable, sparkline
Suggests:
Published: 2021-02-20
Author: Hao Zhu [aut, cre], Thomas Traverso [ctb], Timothy Tsai [ctb], Will Beasley [ctb], Yihui Xie [ctb], GuangChuang Yu [ctb], Stéphane Laurent [ctb], Rob Shepherd [ctb], Yoni Sidi [ctb], Brian Salzer [ctb], George Gui [ctb], Yeliang Fan [ctb], Duncan Murdoch [ctb], Bill Evans [ctb]
Maintainer: Hao Zhu <haozhu233 at gmail.com>
BugReports: <https://github.com/haozhu233/kableExtra/issues>
License: MIT + file LICENSE
URL: <http://haozhu233.github.io/kableExtra/>, <https://github.com/haozhu233/kableExtra>
NeedsCompilation: no
Materials: README NEWS
CRAN checks: [kableExtra results](#)

Downloads:

Reference manual: [kableExtra.pdf](#)

Vignettes: [Create Awesome HTML Table with knitr::kable and kableExtra](#)
[Create Awesome PDF Table with knitr::kable and kableExtra](#)
[Best Practice for newline in LaTeX table](#)
[Legacy or Unrecommended Features](#)

建立 HTML , LaTeX 表格

套件說明

kableExtra 套件說明

Create Awesome HTML Table with knitr::kable and kableExtra

Hao Zhu

2021-02-19

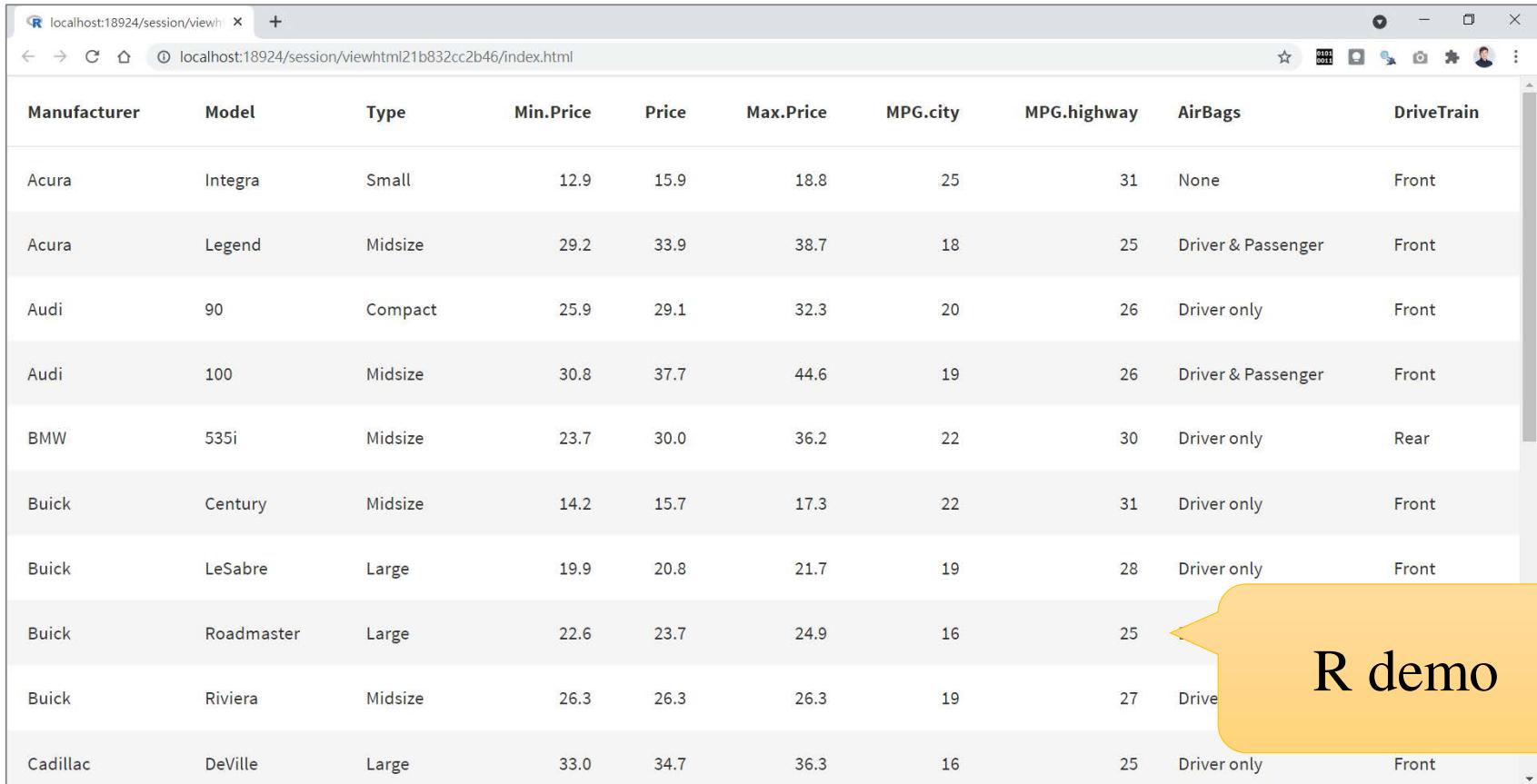
Please see the package [documentation site](#) for how to use this package in LaTeX.



Overview

The goal of `kableExtra` is to help you build common complex tables and manipulate table styles. It imports the pipe `%>%` symbol from `magrittr` and verbalize all the functions, so basically you can add “layers” to a kable output in a way that is similar with `ggplot2` and `plotly`.

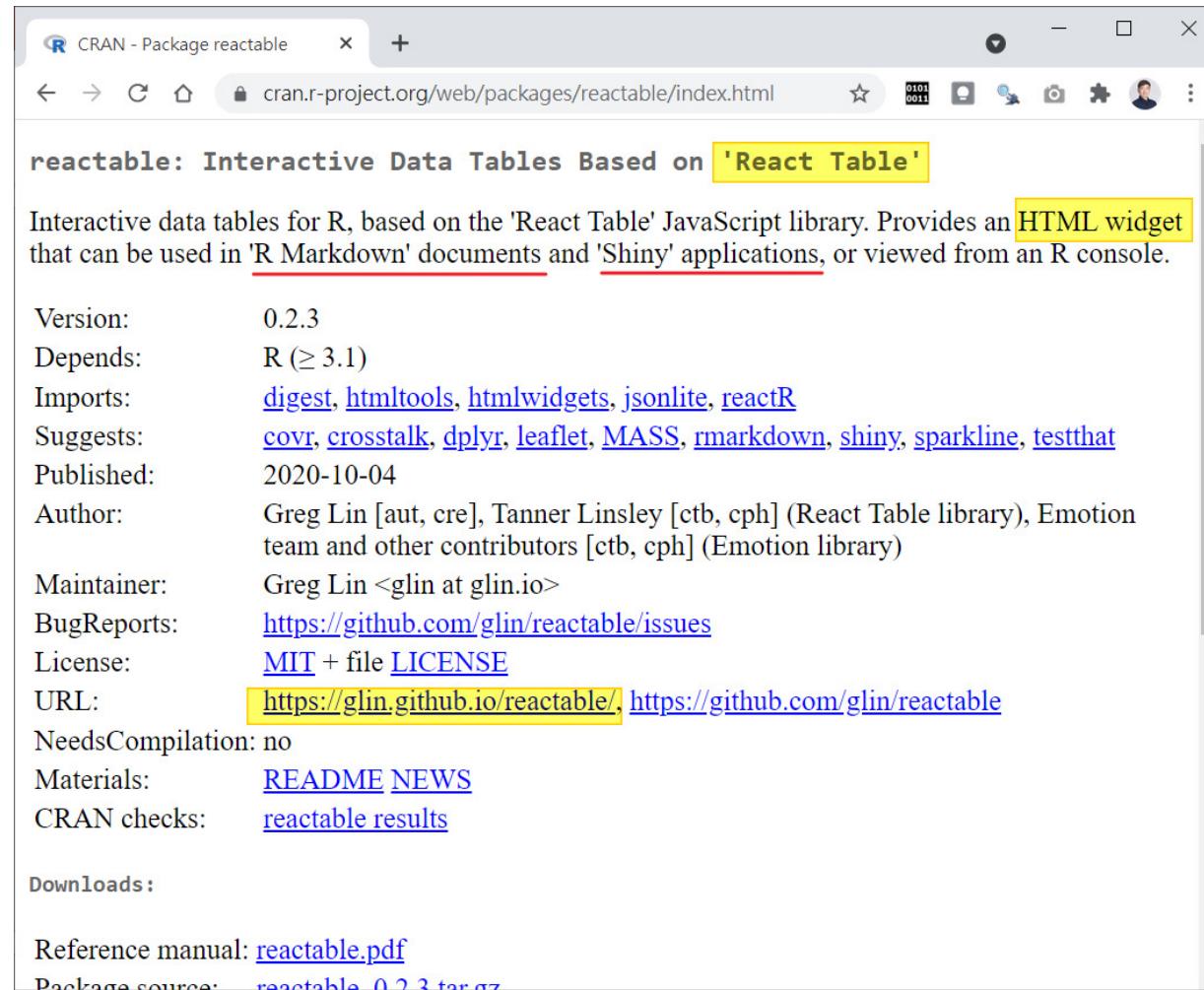
kableExtra 範例



Manufacturer	Model	Type	Min.Price	Price	Max.Price	MPG.city	MPG.highway	AirBags	DriveTrain
Acura	Integra	Small	12.9	15.9	18.8	25	31	None	Front
Acura	Legend	Midsize	29.2	33.9	38.7	18	25	Driver & Passenger	Front
Audi	90	Compact	25.9	29.1	32.3	20	26	Driver only	Front
Audi	100	Midsize	30.8	37.7	44.6	19	26	Driver & Passenger	Front
BMW	535i	Midsize	23.7	30.0	36.2	22	30	Driver only	Rear
Buick	Century	Midsize	14.2	15.7	17.3	22	31	Driver only	Front
Buick	LeSabre	Large	19.9	20.8	21.7	19	28	Driver only	Front
Buick	Roadmaster	Large	22.6	23.7	24.9	16	25	Driver only	Front
Buick	Riviera	Midsize	26.3	26.3	26.3	19	27	Driver only	Front
Cadillac	DeVille	Large	33.0	34.7	36.3	16	25	Driver only	Front

R demo

reactable 套件



reactable: Interactive Data Tables Based on 'React Table'

Interactive data tables for R, based on the 'React Table' JavaScript library. Provides an HTML widget that can be used in 'R Markdown' documents and 'Shiny' applications, or viewed from an R console.

Version: 0.2.3

Depends: R (\geq 3.1)

Imports: [digest](#), [htmltools](#), [htmlwidgets](#), [jsonlite](#), [reactR](#)

Suggests: [covr](#), [crosstalk](#), [dplyr](#), [leaflet](#), [MASS](#), [rmarkdown](#), [shiny](#), [sparkline](#), [testthat](#)

Published: 2020-10-04

Author: Greg Lin [aut, cre], Tanner Linsley [ctb, cph] (React Table library), Emotion team and other contributors [ctb, cph] (Emotion library)

Maintainer: Greg Lin <glin at glin.io>

BugReports: <https://github.com/glin/reactable/issues>

License: [MIT](#) + file [LICENSE](#)

URL: <https://glin.github.io/reactable/>, <https://github.com/glin/reactable>

NeedsCompilation: no

Materials: [README](#) [NEWS](#)

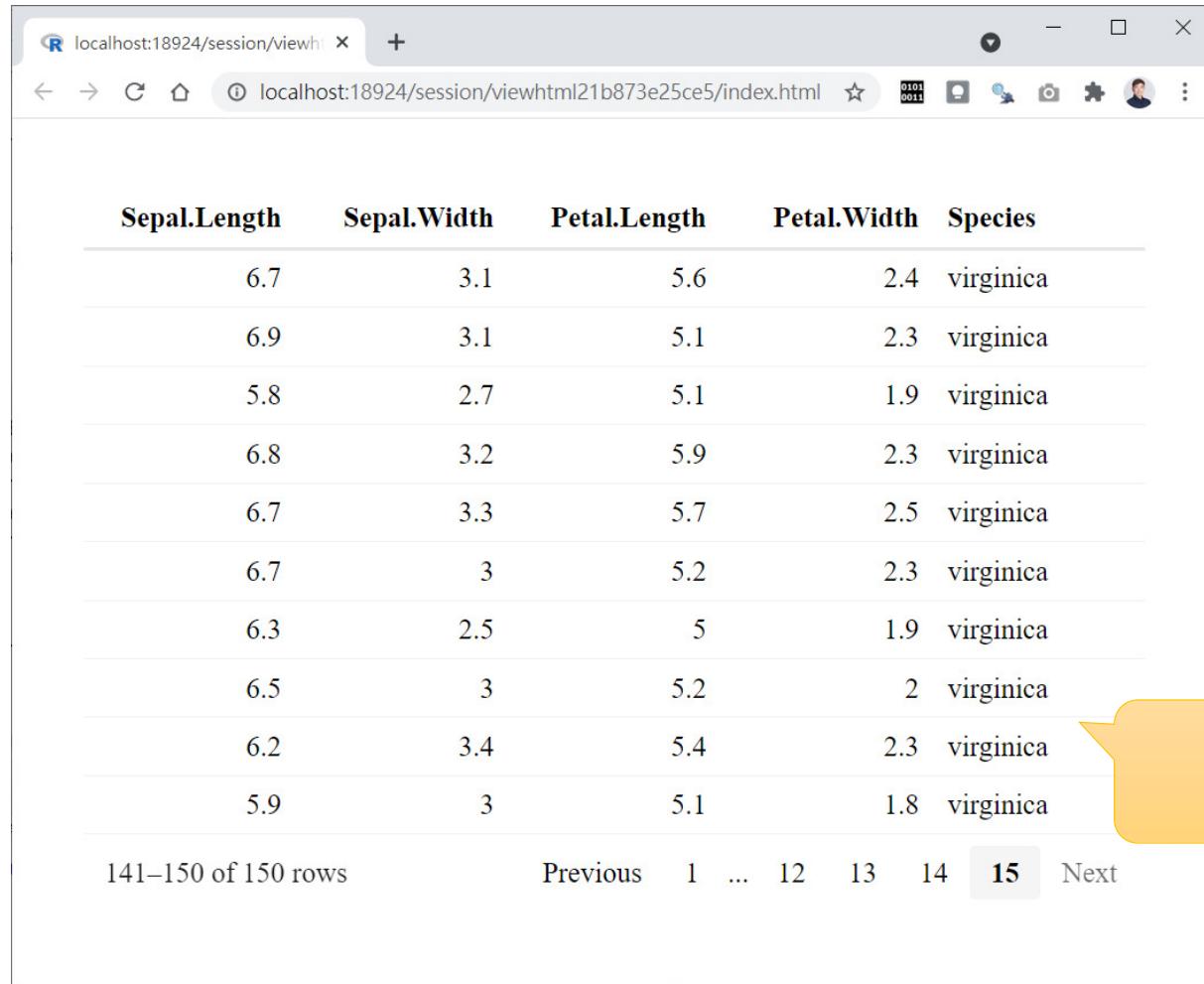
CRAN checks: [reactable results](#)

Downloads:

Reference manual: [reactable.pdf](#)

Package source: [reactable_0.2.3.tar.gz](#)

reactable 範例



Sepal.Length	Sepal.Width	Petal.Length	Petal.Width	Species
6.7	3.1	5.6	2.4	virginica
6.9	3.1	5.1	2.3	virginica
5.8	2.7	5.1	1.9	virginica
6.8	3.2	5.9	2.3	virginica
6.7	3.3	5.7	2.5	virginica
6.7	3	5.2	2.3	virginica
6.3	2.5	5	1.9	virginica
6.5	3	5.2	2	virginica
6.2	3.4	5.4	2.3	virginica
5.9	3	5.1	1.8	virginica

141–150 of 150 rows Previous 1 ... 12 13 14 15 Next

R demo

互動式繪圖 plotly

plotly 套件

plotly: Create Interactive Web Graphics via 'plotly.js'

Create interactive web graphics from 'ggplot2' graphs and/or a custom interface to the (MIT-licensed) JavaScript library 'plotly.js' inspired by the grammar of graphics.

Version:	4.9.2.1
Depends:	R (\geq 3.2.0), ggplot2 (\geq 3.0.0)
Imports:	tools, scales , httr , jsonlite (\geq 1.6), magrittr , digest , viridisLite , base64enc , htmltools (\geq 0.3.6), htmlwidgets (\geq 1.3), tidyverse , hexbin , RColorBrewer , dplyr , tibble , lazyeval (\geq 0.2.0), rlang , crosstalk , purrr , data.table , promises
Suggests:	MASS , maps , ggthemes , GGally , testthat , knitr , devtools , shiny (\geq 1.1.0), shinytest (\geq 1.3.0), curl , rmarkdown , vdiffr , Cairo , broom , webshot , listviewer , dendextend , sf , maptools , rgeos , png , IRdisplay , processx , plotlyGeoAssets , forcats
Published:	2020-04-04
Author:	Carson Sievert  [aut, cre], Chris Parmer [aut], Toby Hocking [aut], Scott Chamberlain [aut], Karthik Ram [aut], Marianne Corvellec  [aut], Pedro Despouy [aut], Plotly Technologies Inc. [cph]
Maintainer:	Carson Sievert <cpsievert1 at gmail.com>
BugReports:	https://github.com/ropensci/plotly/issues
License:	MIT + file LICENSE
URL:	https://plotly-r.com , https://github.com/ropensci/plotly#readme , https://plot.ly/r
NeedsCompilation:	no

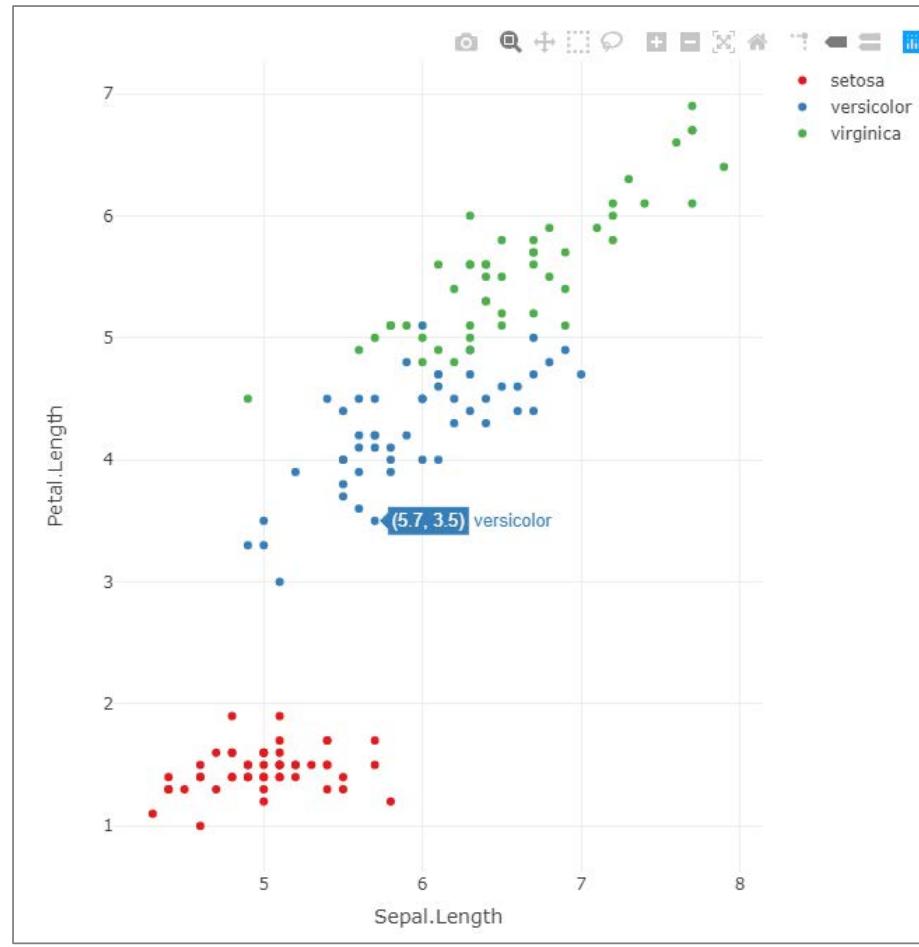
plot_ly {plotly}

```
library(plotly)

fig <- plot_ly(data = iris,
                 x = ~Sepal.Length,
                 y = ~Petal.Length,
                 color = ~Species,
                 colors = "Set1")

fig
```

plotly 應用



wordcloud2 套件 - demo

文字雲

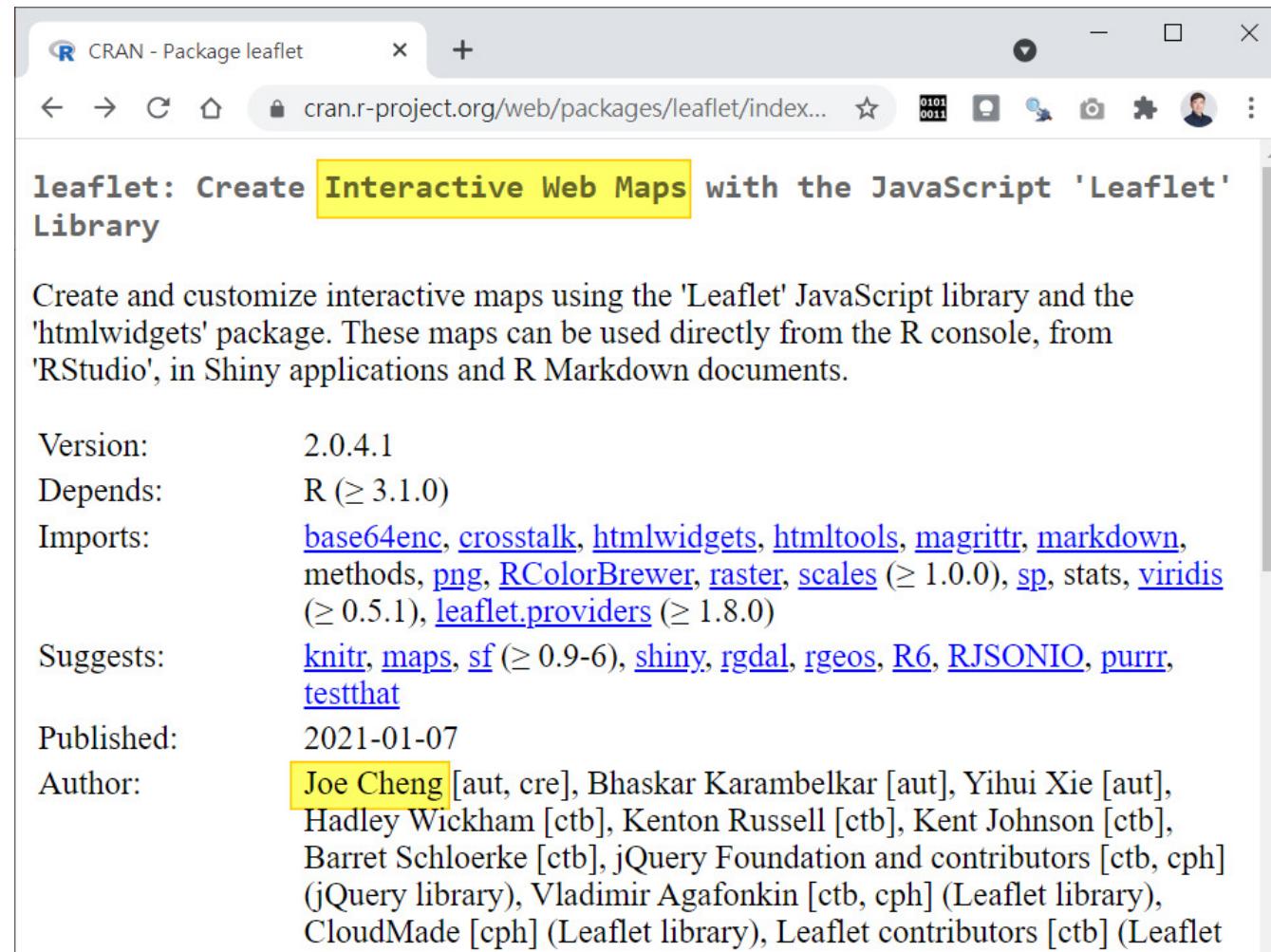




09. 地理資料視覺化leaflet, 機器學習與 深度學習應用

地理資料視覺化 leaflet

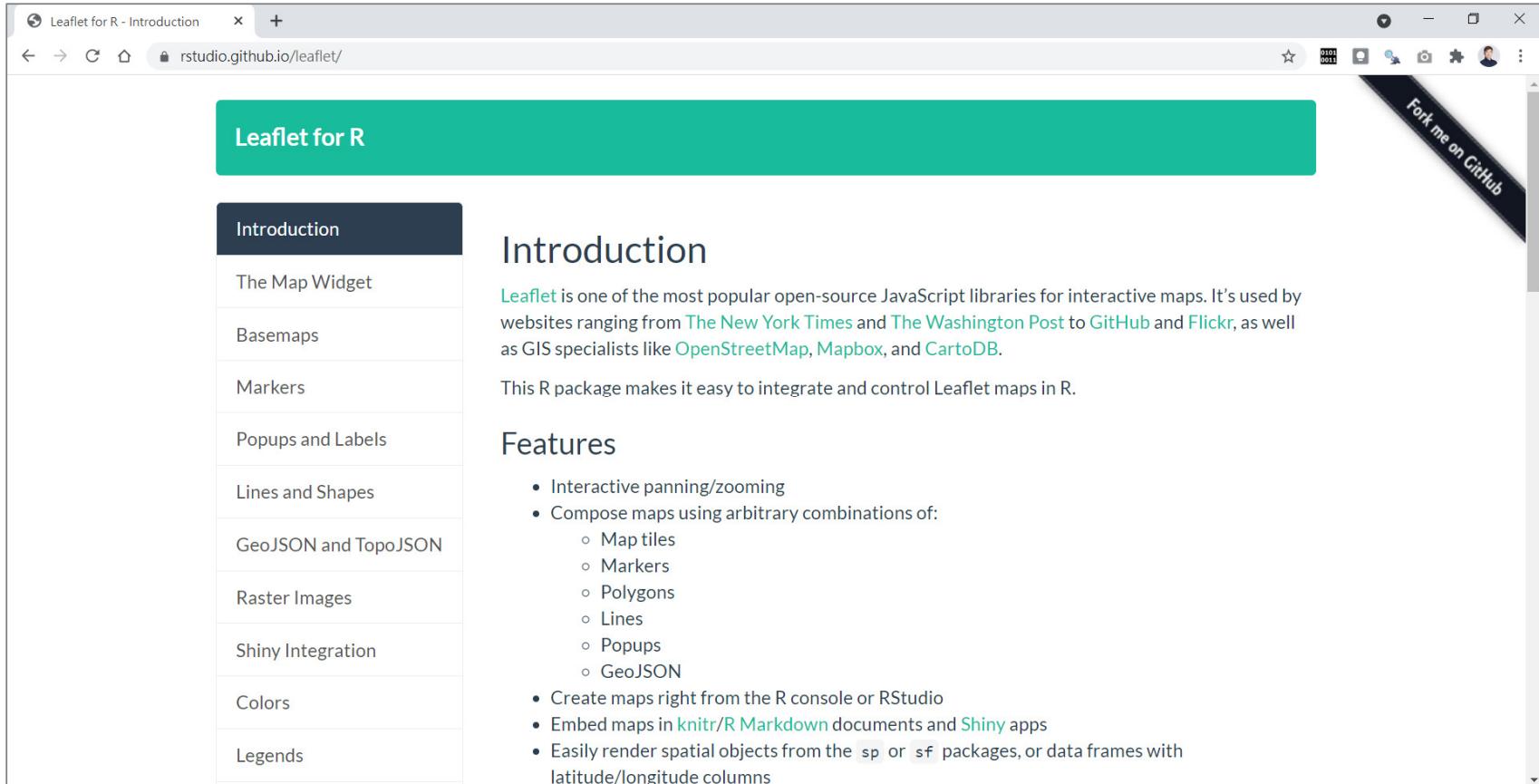
leaflet 套件



The screenshot shows a web browser window displaying the CRAN package page for 'leaflet'. The title bar says 'CRAN - Package leaflet'. The main content area has a yellow box highlighting the package name and its purpose: 'leaflet: Create Interactive Web Maps with the JavaScript 'Leaflet' Library'. Below this, a paragraph describes the package's function: 'Create and customize interactive maps using the 'Leaflet' JavaScript library and the 'htmlwidgets' package. These maps can be used directly from the R console, from 'RStudio', in Shiny applications and R Markdown documents.' A table lists package details:

Version:	2.0.4.1
Depends:	R (≥ 3.1.0)
Imports:	base64enc , crosstalk , htmlwidgets , htmltools , magrittr , markdown , methods, png , RColorBrewer , raster , scales (≥ 1.0.0), sp , stats, viridis (≥ 0.5.1), leaflet.providers (≥ 1.8.0)
Suggests:	knitr , maps , sf (≥ 0.9-6), shiny , rgdal , rgeos , R6 , RJSONIO , purrr , testthat
Published:	2021-01-07
Author:	Joe Cheng [aut, cre], Bhaskar Karambelkar [aut], Yihui Xie [aut], Hadley Wickham [ctb], Kenton Russell [ctb], Kent Johnson [ctb], Barret Schloerke [ctb], jQuery Foundation and contributors [ctb, cph] (jQuery library), Vladimir Agafonkin [ctb, cph] (Leaflet library), CloudMade [cph] (Leaflet library), Leaflet contributors [ctb] (Leaflet

<https://rstudio.github.io/leaflet/>



The screenshot shows a web browser window titled "Leaflet for R - Introduction". The URL in the address bar is "rstudio.github.io/leaflet/". The page has a green header bar with the text "Leaflet for R". On the left, there is a sidebar with a dark blue header "Introduction" and a list of topics: "The Map Widget", "Basemaps", "Markers", "Popups and Labels", "Lines and Shapes", "GeoJSON and TopoJSON", "Raster Images", "Shiny Integration", "Colors", and "Legends". The main content area has a large heading "Introduction". Below it, a paragraph explains that Leaflet is a popular open-source JavaScript library for interactive maps, used by major websites like The New York Times and The Washington Post, and by GIS specialists like OpenStreetMap, Mapbox, and CartoDB. It also states that this R package makes it easy to integrate and control Leaflet maps in R. A section titled "Features" lists several capabilities, including interactive panning/zooming, composing maps from various elements like map tiles, markers, polygons, lines, popups, and GeoJSON, creating maps from R console or RStudio, embedding maps in knitr/R Markdown documents, and rendering spatial objects from sp or sf packages.

Leaflet for R

Introduction

- The Map Widget
- Basemaps
- Markers
- Popups and Labels
- Lines and Shapes
- GeoJSON and TopoJSON
- Raster Images
- Shiny Integration
- Colors
- Legends

Introduction

Leaflet is one of the most popular open-source JavaScript libraries for interactive maps. It's used by websites ranging from [The New York Times](#) and [The Washington Post](#) to [GitHub](#) and [Flickr](#), as well as GIS specialists like [OpenStreetMap](#), [Mapbox](#), and [CartoDB](#).

This R package makes it easy to integrate and control Leaflet maps in R.

Features

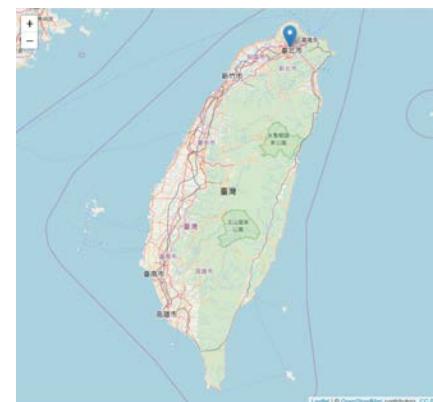
- Interactive panning/zooming
- Compose maps using arbitrary combinations of:
 - Map tiles
 - Markers
 - Polygons
 - Lines
 - Popups
 - GeoJSON
- Create maps right from the R console or RStudio
- Embed maps in [knitr](#)/[R Markdown](#) documents and [Shiny](#) apps
- Easily render spatial objects from the `sp` or `sf` packages, or data frames with latitude/longitude columns

leaflet 套件練習

library(套件)

```
# leaflet  
library(leaflet)  
m1 <- leaflet() %>%  
  addTiles() %>%  
  addMarkers(lng = 121.542384, lat = 25.058035, popup = "台北") %>%  
  setView(lng = 120.974, lat = 23.6, zoom = 8)  
m1
```

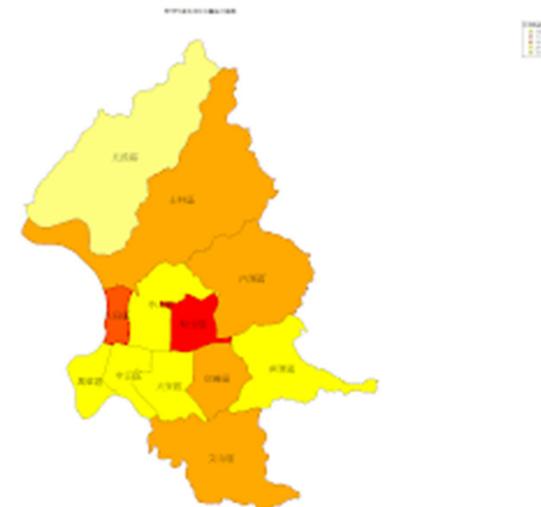
資料物件 <- 函數(參數1= " " , 參數2= ...)



主題式地圖(Thematic map) – 政府開放資料為例

主題式地圖

主題式地圖(Thematic map) - 政府開放資料為例



```
# 主題式地圖  
# Thematic map  
# 開放式資料  
# open data  
# 地圖資料與社會經濟資料合併  
# rgdal 套件  
# tmap 套件
```

R demo

機器學習

機器學習 Machine learning

- 監督式學習 (Supervised learning)
 - Telling the algorithm what to predict
- 非監督式學習 (Unsupervised learning)
 - No label or target value given for the data
- 半監督學習 (Semi-supervised learning)
 - 具有少量標記資料
- 強化學習 (Reinforcement learning)
 - 為了達成目標，隨著環境的變動，而逐步調整其行為，並評估每一個行動之後所到的回饋是正向的或負向的。
- 深度學習 (Deep learning)



監督式學習 vs. 非監督式學習

- 監督式學習 Supervised learning - 執行 $X \rightarrow$ 預測 $\rightarrow Y$
 - 迴歸分析 Regression analysis
 - 廣義線性模型 General linear model (GLM)
 - 天真貝氏法 Naïve-Bayes
 - K近鄰法 k-nearest neighbors (KNN)
 - 決策樹 Decision tree
 - 支持向量機 Support vector machine (SVM)
 - 類神經網路 Neural network (NN)
 - 集成學習 Ensemble learning: 使用多種學習算法來獲得比單獨使用演算法更好預測結果
- 非監督式學習 Unsupervised learning
 - 集群法 Clustering
 - 關聯規則 Association rule
 - 主成分分析 Principal Component Analysis

CRISP-DM標準流程

資料探勘生命週期 - CRISP-DM

- 跨產業資料探勘標準作業流程 (Cross Industry Standard Process for Data Mining)
- 資料探勘方法論
- CRISP-DM是於1990年起，由SPSS以及NCR兩大廠商在合作戴姆克萊斯勒-賓士(Daimler Benz)的資料倉儲以及資料探勘過程中發展出來的。

CRISP-DM 資料探勘流程(續)

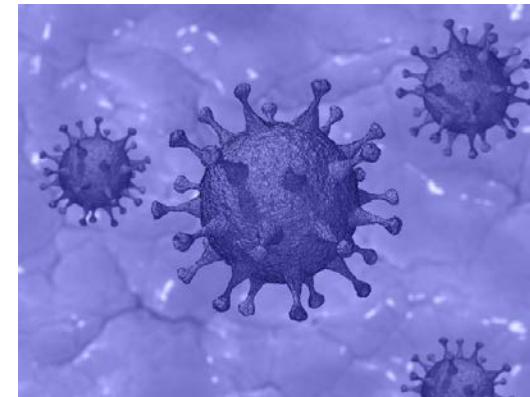
- 步驟 1：商業理解
- 步驟 2：資料理解
- 步驟 3：資料準備
- 步驟 4：模式建立
- 步驟 5：評估與測試
- 步驟 6：佈署應用

} 佔整專案時間的
~80%

- 訓練資料70%
- 測試資料30%

商業理解

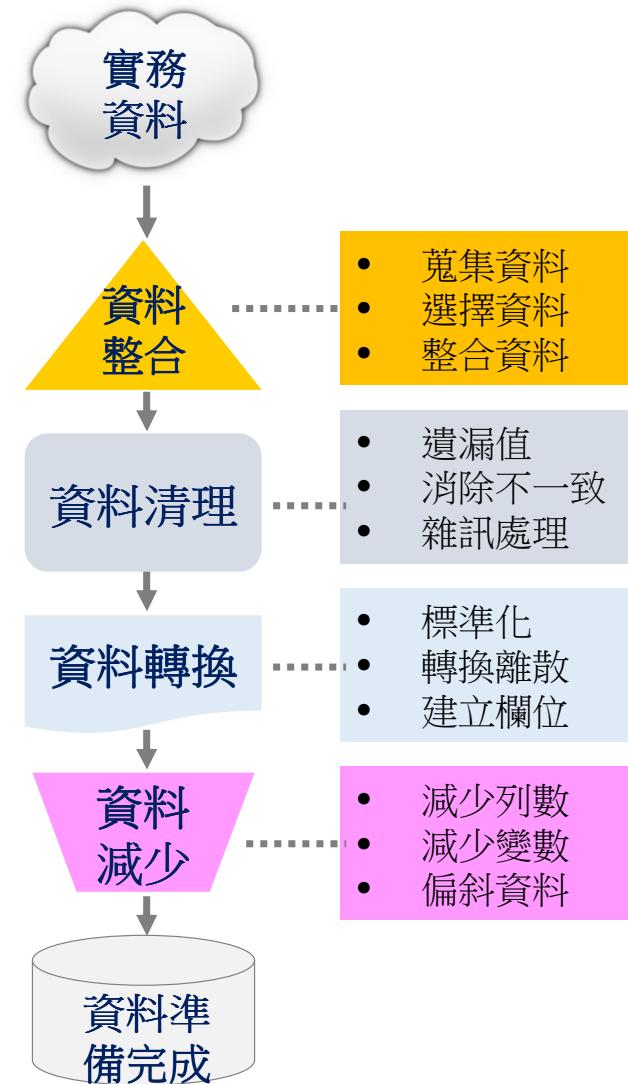
- 終極目標是要解決具體的產業問題，諸如提高購買率、找出詐欺交易、銷售預測與異常偵測等，因此以專業知識 (domain knowledge)進行商業理解是重要的第一步，處理重點：
 - 擬定商業目標
 - 進行當前處境評估
 - 決定資料探勘目標/成本
 - 產生專案計劃
 - 解決顧客問題



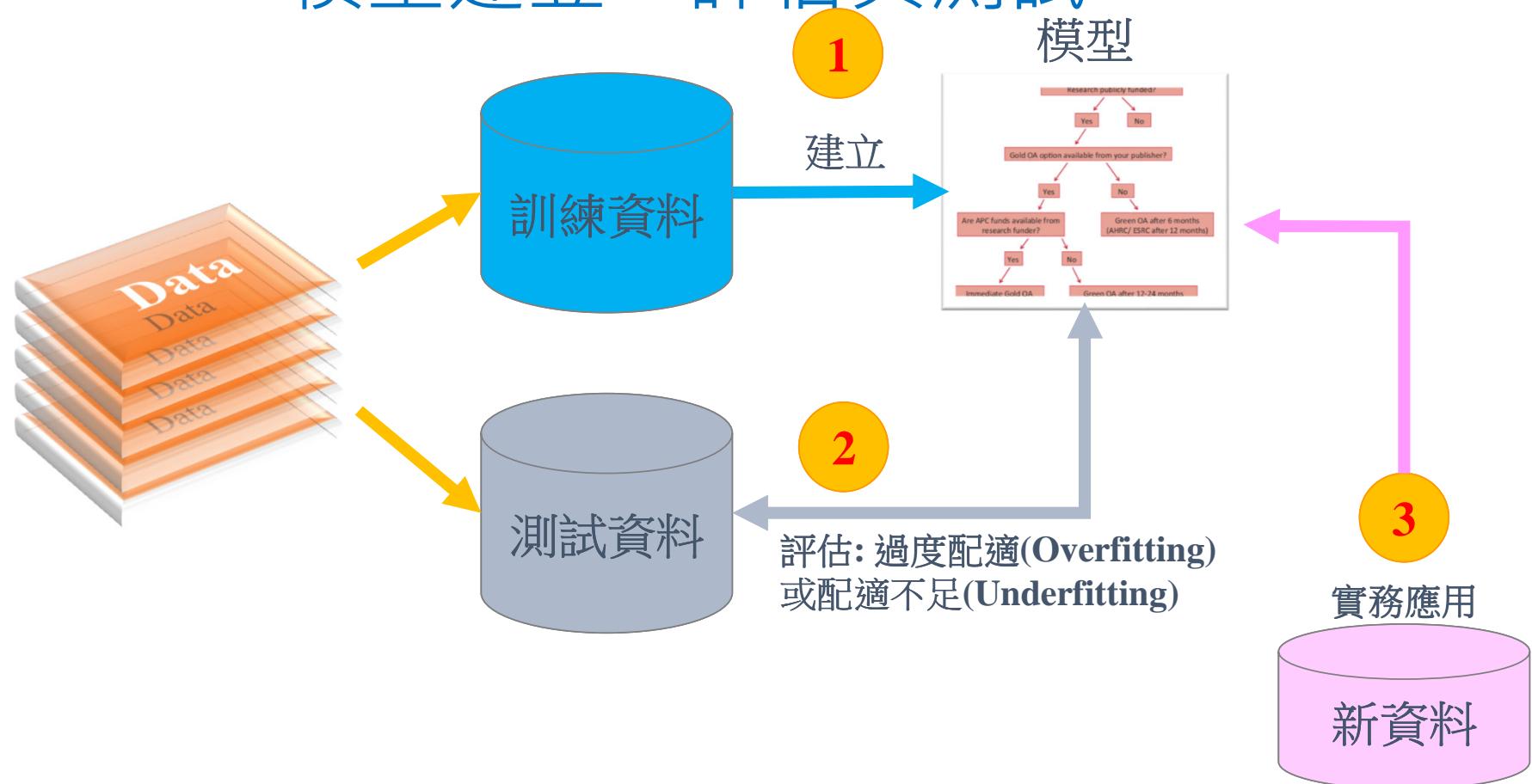
資料理解

- 包括描述資料、探索資料、核驗資料品質
- 敘述統計分析
 - 六力分析(summary函數)
- 繪圖
 - 依群組特性
 - 依時間特性
 - 新增評估欄位
- 趨勢
- 離群值 (outlier)
- 散佈圖、散佈圖矩陣
- 盒鬚圖

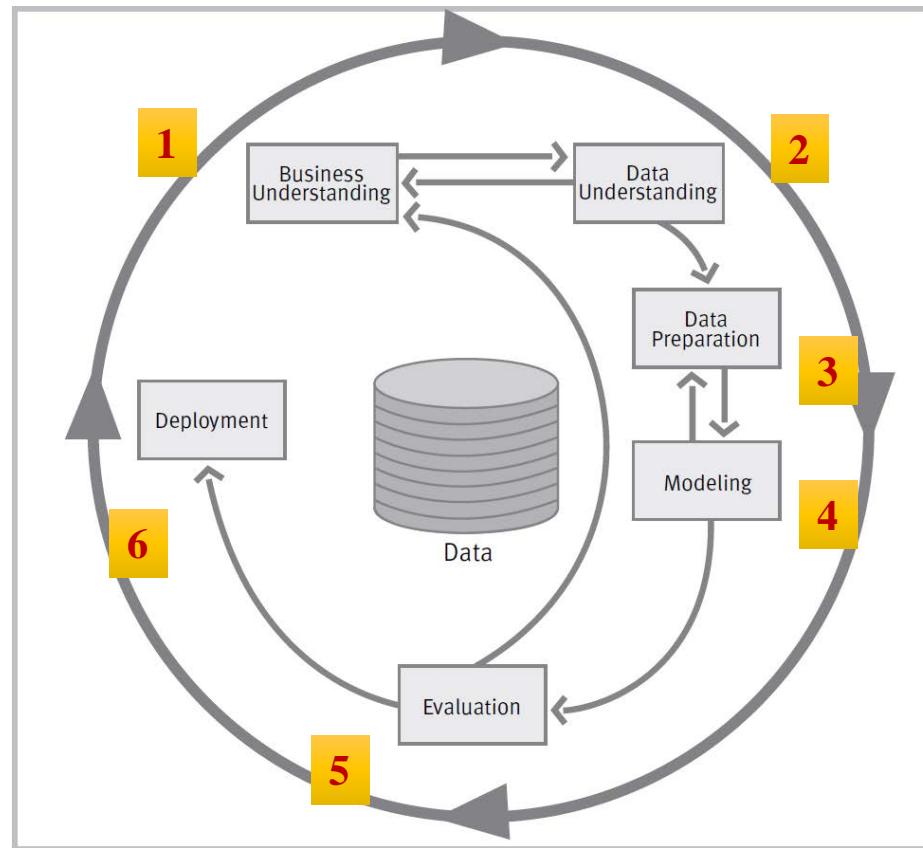
資料準備



模型建立、評估與測試



CRISP-DM 資料探勘流程(續)



參考 https://en.wikipedia.org/wiki/Cross_Industry_Standard_Process_for_Data_Mining

數值模型績效指標

- 不可直接使用誤差的算術平均!

$$\text{Total error} = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)$$



- 均方誤差 (Mean Squared Error, MSE)

$$MSE = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2$$

- 均方根誤差 (Root Mean Squared Error, RMSE)

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2}$$

- 平均絕對誤差 (Mean Absolute Error, MAE)

$$MAE = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i|$$

類別模型績效指標

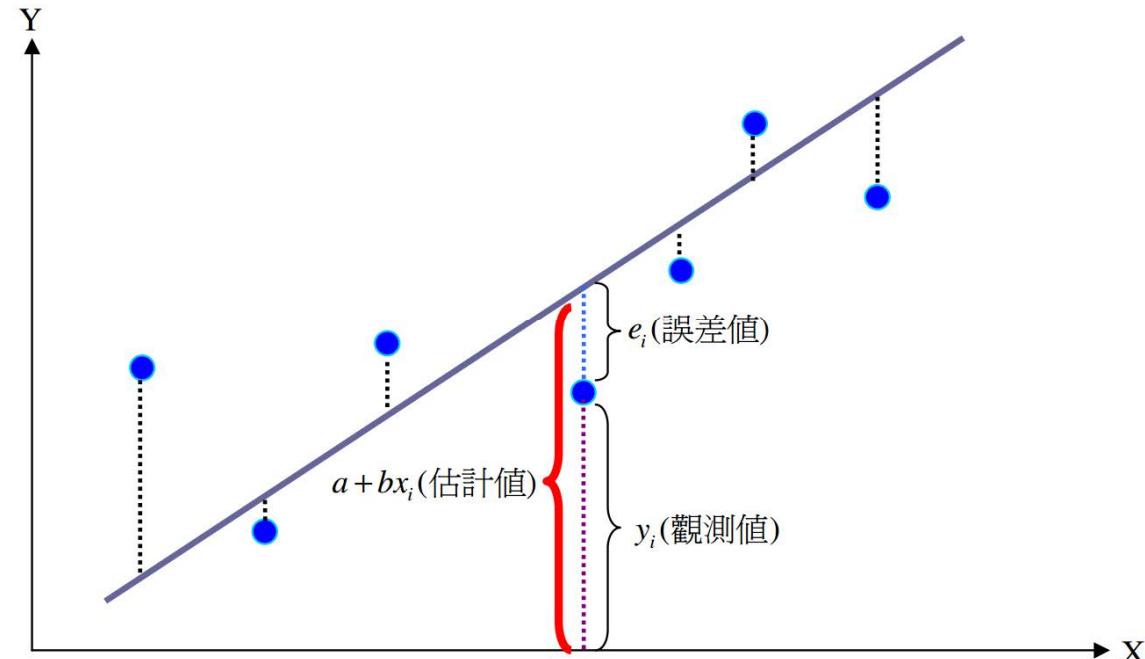
- <http://rwepa.blogspot.com/2013/01/rocr-roc-curve.html>

```
#          | 真實P類別 真實N類別
# ****|*****
# 預測P類別 | TP真陽數 FP假陽數
# 預測N類別 | FN假陰數 TN真陰數
# ****|*****
#          | P         N

# 1.TPR(True positive rate) 真陽性率，愈大愈好 -----
# =TP/ (TP+FN)
# =TP/P
# =Sensitivity 積敏度
# =Recall 召回率
# =Probability of detection
# =Power
# 實際為陽性的樣本中，判斷為陽性的比例。
# 例如真正有生病的人中，被醫院判斷為有生病者的比例。
```

混淆矩陣
(Confusion Matrix)

迴歸分析



參考:

- https://github.com/rwepa/DataDemo/blob/master/regression_01.pdf
- 迴歸分析繪圖 R code

迴歸分析 Regression analysis

依變數

- Model: $\hat{Y}_i = \hat{\beta}_0 + \hat{\beta}_1 X_{1i} + \dots + \hat{\beta}_k X_{ki} \quad i = 1 \dots n$
 - Y: dependent, outcome, or response variable
 - X: independent, predictor, or explanatory variables

自變數

- Regression analysis can be used to
 - identify the **explanatory variables**(解釋變數) that are related to **a response variable** (反應變數)
 - to describe the form of the **relationships** involved
 - to provide an equation for **predicting** the response variable from the explanatory variables.

迴歸分析範例

- 休閒運動
 - Y : 跑步消耗卡路里
 - X : 時間, 心跳比率, 速度, 年齡, 性別, BMI
- 財務風險管理
 - Y : 財務危機
 - X : 獲利, 經營績效, 穩健程度, 經營管理
- 企業預測
 - Y : 明年電梯市場總量預測
 - X : 出貨數, 建照發照數, 市場占有率, 台灣生活指標, 世界經濟指標

迴歸模型

lm

迴歸種類	反應變數	解釋變數	備註
Simple linear 簡單線性迴歸	1	1	$y \sim x$
Polynomial 多項式迴歸	1	1	n次多項式 $y \sim x + x^2$
Multiple linear 多元線性迴歸	1	≥ 2	$y \sim x_1 + x_2 + x_3$
Multivariate 多變數迴歸	≥ 2		
glm	Logistic 邏輯斯迴歸	1個類別型變數	
	Poisson 卜瓦松迴歸	1個計數型變數	
nls	Nonlinear 非線性迴歸	1	$y \sim 1/\exp(x_1+\log x_2)$
arima	Time series 時間序列		

最小平方法 Ordinary Least Square (OLS)

- 統計假設：

- 常態性 • Normality —For fixed values of the independent variables, the dependent variable (Y) is normally distributed.
- 獨立性 • Independence —The Y_i values are independent of each other.
- 線性關係 • Linearity —The dependent variable is linearly related to the independent variables.
- 同質性 • Homoscedasticity —The variance of the dependent variable doesn't vary with the levels of the independent variables.

線性模型 linear model: *lm*

```
> # Simple linear regression  
> ?lm  
> # my.lm <- lm(formula, data="xxx")  
> # formula: y ~ x1 + x2 + ... +xn  
> # end
```

R公式(formula)之符號

Symbol	Usage
\sim	Separates response variables on the left from the explanatory variables on the right. For example, a prediction of y from x , z , and w would be coded $y \sim x + z + w$.
$+$	Separates predictor variables.
$:$	Denotes an interaction between predictor variables. A prediction of y from x , z , and the interaction between x and z would be coded $y \sim x + z + x:z$.
$*$	A shortcut for denoting all possible interactions. The code $y \sim x * z * w$ expands to $y \sim x + z + w + x:z + x:w + z:w + x:z:w$.
\wedge	Denotes interactions up to a specified degree. The code $y \sim (x + z + w)^2$ expands to $y \sim x + z + w + x:z + x:w + z:w$.
\cdot	A place holder for all other variables in the data frame except the dependent variable. For example, if a data frame contained the variables x , y , z , and w , then the code $y \sim .$ would expand to $y \sim x + z + w$.
$-$	A minus sign removes a variable from the equation. For example, $y \sim (x + z + w)^2 - x:w$ expands to $y \sim x + z + w + x:z + z:w$.
-1	Suppresses the intercept. For example, the formula $y \sim x -1$ fits a regression of y on x , and forces the line through the origin at $x=0$.
$I()$	Elements within the parentheses are interpreted arithmetically. For example, $y \sim x + (z + w)^2$ would expand to $y \sim x + z + w + z:w$. In contrast, the code $y \sim x + I((z + w)^2)$ would expand to $y \sim x + h$, where h is a new variable created by squaring the sum of z and w .
function	Mathematical functions can be used in formulas. For example, $\log(y) \sim x + z + w$ would predict $\log(y)$ from x , z , and w .

迴歸模型 – women 資料集

```
> # women: Average Heights and Weights for American Women  
> # y: weight  
> # x: height  
> fit.lm <- lm(weight ~ height, data=women)  
> summary(fit.lm)
```

Call:

```
lm(formula = weight ~ height, data = women)
```

Residuals:

Min	1Q	Median	3Q	Max
-1.7333	-1.1333	-0.3833	0.7417	3.1167

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	-87.51667	5.93694	-14.74	1.71e-09 ***
height	3.45000	0.09114	37.85	1.09e-14 ***

Signif. codes: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 1.525 on 13 degrees of freedom

Multiple R-squared: 0.991, Adjusted R-squared: 0.9903

F-statistic: 1433 on 1 and 13 DF, p-value: 1.091e-14

```
> # weight = -87.52+3.45*height
```

```
> # verify residuals
> names(fit.lm)
[1] "coefficients"   "residuals"      "effects"       "rank"
[5] "fitted.values"  "assign"        "qr"           "df.residual"
[9] "xlevels"         "call"          "terms"        "model"
> women$weight    # actual
[1] 115 117 120 123 126 129 132 135 139 142 146 150 154 159 164
> fitted(fit.lm) # predicted
     1      2      3      4      5      6      7      8
112.5833 116.0333 119.4833 122.9333 126.3833 129.8333 133.2833 136.7333
     9     10     11     12     13     14     15
140.1833 143.6333 147.0833 150.5333 153.9833 157.4333 160.8833
> residuals(fit.lm) # residual=actual-predicted
     1      2      3      4      5      6
2.41666667 0.96666667 0.51666667 0.06666667 -0.38333333 -0.83333333
     7      8      9     10     11     12
-1.28333333 -1.73333333 -1.18333333 -1.63333333 -1.08333333 -0.53333333
     13     14     15
0.01666667 1.56666667 3.11666667
> women$weight - fitted(fit.lm)
     1      2      3      4      5      6
2.41666667 0.96666667 0.51666667 0.06666667 -0.38333333 -0.83333333
     7      8      9     10     11     12
-1.28333333 -1.73333333 -1.18333333 -1.63333333 -1.08333333 -0.53333333
     13     14     15
0.01666667 1.56666667 3.11666667
>
```

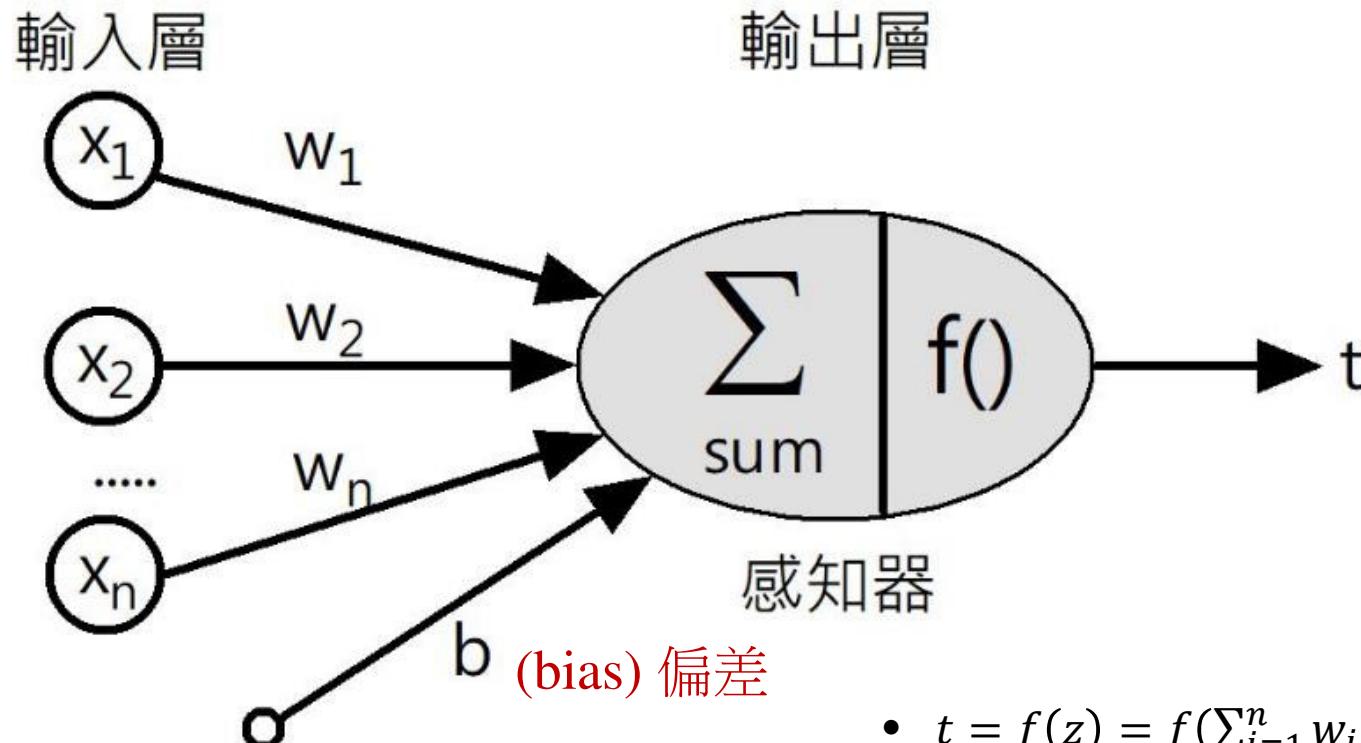
深度學習應用

深度學習發展史



- 1943年：美國數學家 Walter Pitts 和心理學家 Warren McCulloch 提出人工神經元。
- 1957年：美國心理學家 Frank Rosenblatt 提出了感知器(Perceptron)。
- 1980年：多層類神經網路失敗，淺層機器學習方法(SVM等)興起。
- 2006年：Geoffrey Hinton 成功訓練多層神經網路(限制玻爾茲曼機, RBM)，命名為深度學習。
- 2012年：ImageNet 比賽讓深度學習重回學界視野，開啟 NVIDIA GPU 為重要運算硬體。

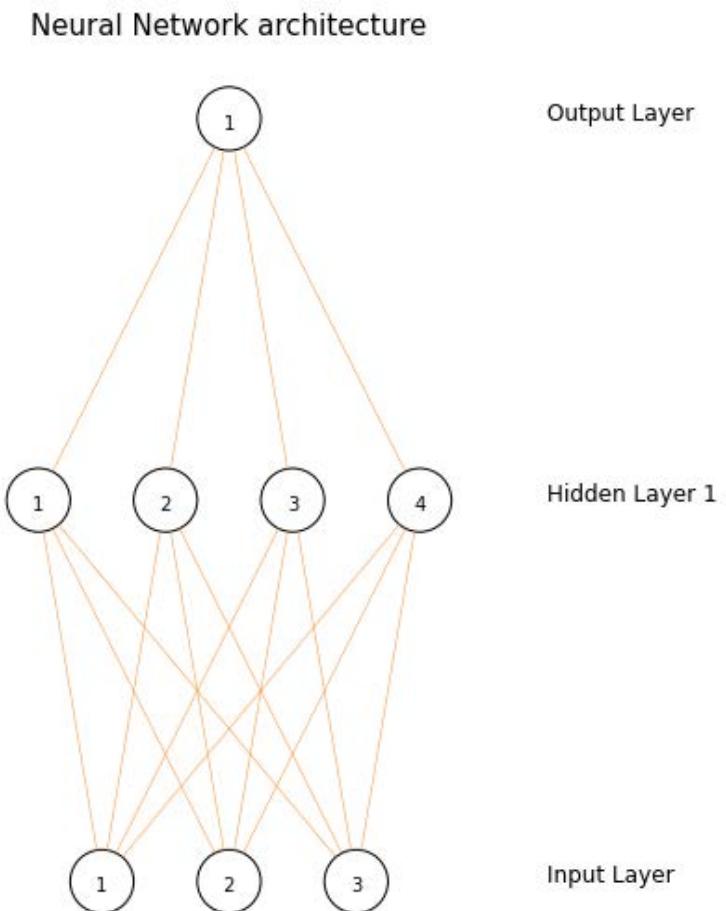
感知器(Perceptron) – 只有二層



- $t = f(z) = f(\sum_{i=1}^n w_i x_i + b)$
- 函數 f 稱為啟動函數或激活函數 Activation function

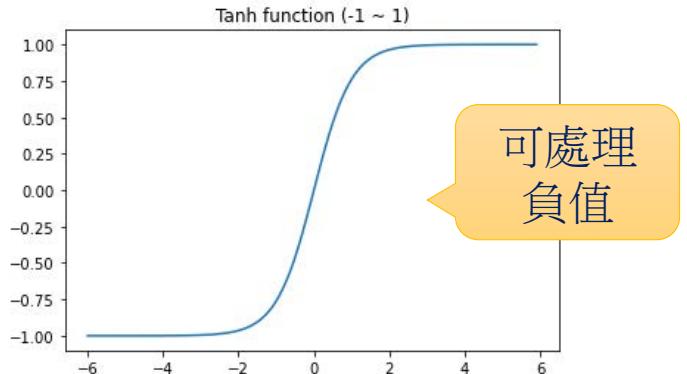
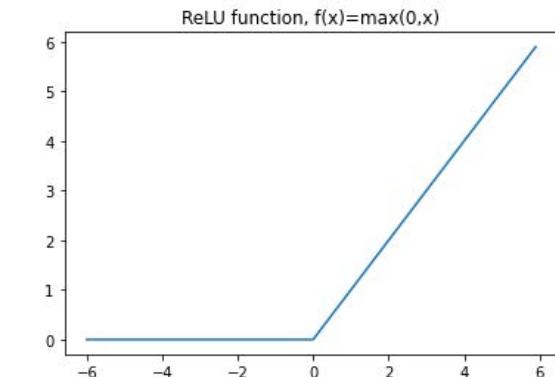
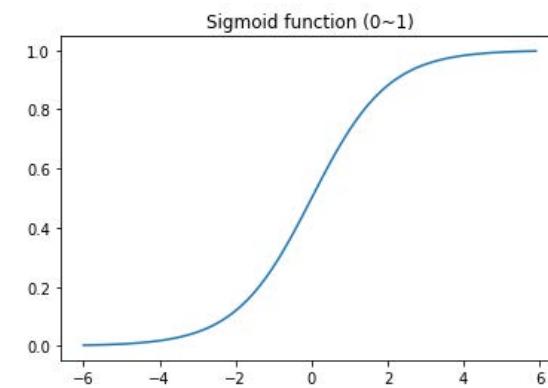
多層感知器 (Multilayer Perceptron, MLP)

- 前饋式神經網路
(Feedforward Neural Network)
- 擴充感知器,加入隱藏層.
- 如果MLP有2層隱藏層,則稱為深度
神經網路 (Deep Neural
Network).



啟動函數 (Activation function)

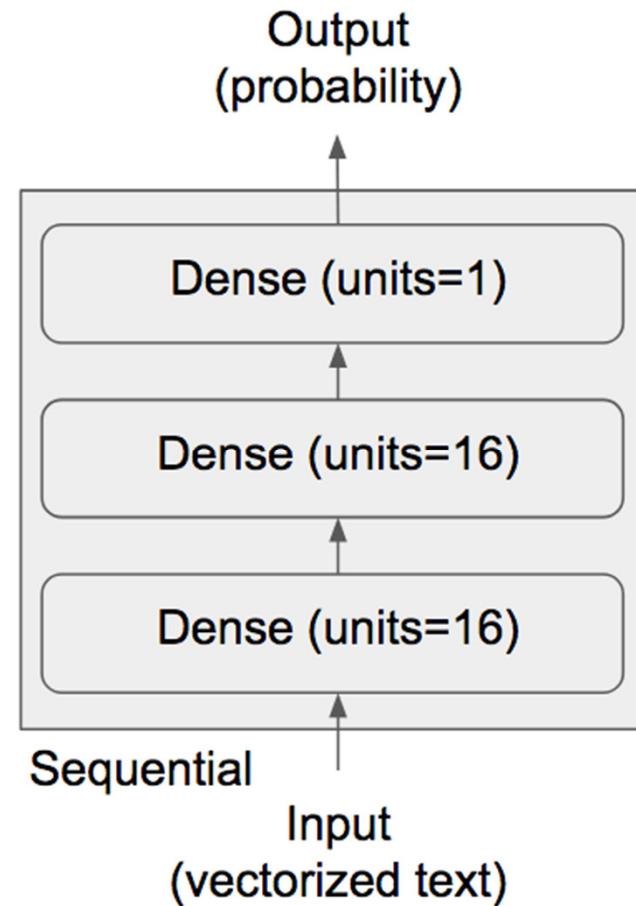
- 隱藏層
 - ReLU (Rectified Linear Unit) , $f(x) = \max(0, x)$
- 輸出層
 - 二元分類
 - Sigmoid , $f(x) = \frac{1}{1+e^{-x}}$
 - Tanh , $f(x) = \frac{\sinh(x)}{\cosh(x)}$
 - 多元分類
 - Softmax, $f(x) = \frac{e^x}{\sum e^x}$



IMDB 網路電影資料庫

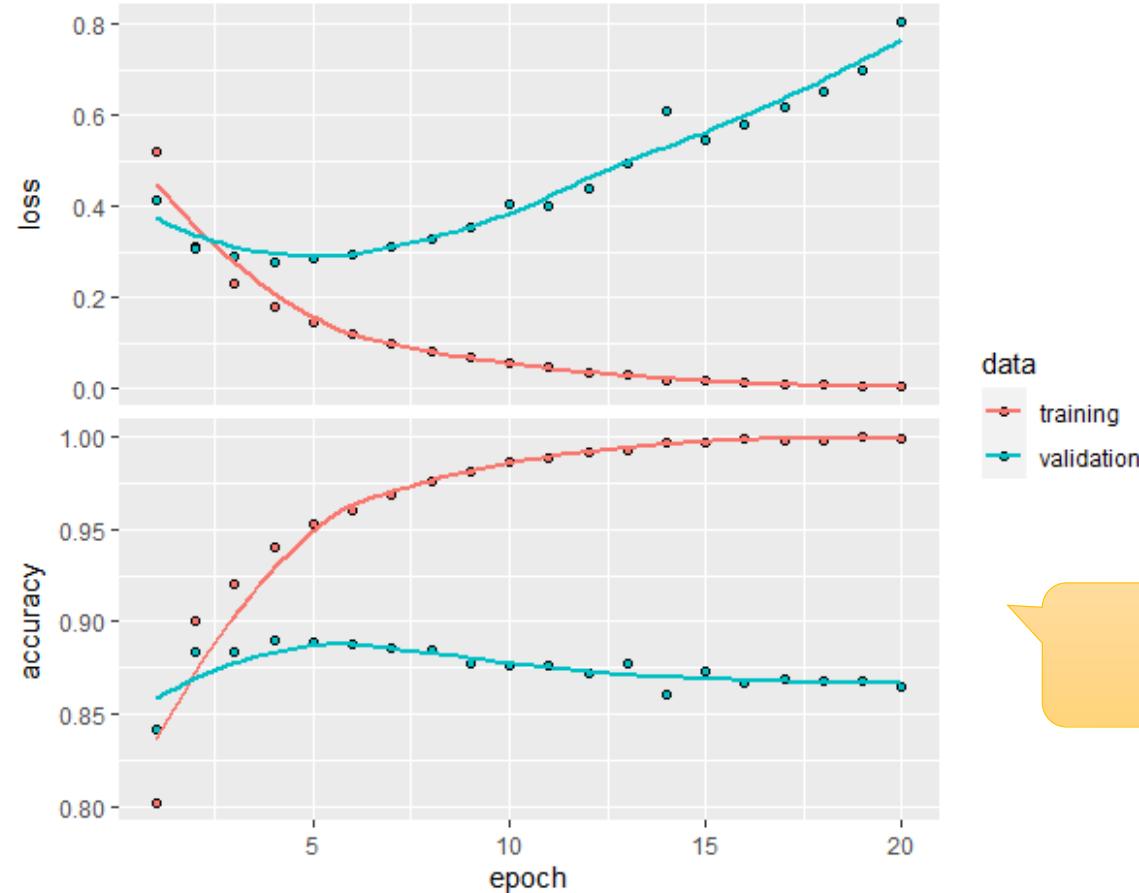
- IMDB 包括 50,000 筆 “影評文字” 資料集。
- 訓練資料與測試資料各 25,000 筆，每一筆 “影評文字” 都被標記成 “正面評價” 或 “負面評價” 。
- 目標：建立一個模型用於預測 “影評文字” 是 “正面評價” 或 “負面評價” 。
- 參考：Francois Chollet and J.J. Allaire , Deep Learning with R, Manning Publications, 2018.
<https://www.manning.com/books/deep-learning-with-r>

3 layer network



- 全連接層僅使用兩層 16 個神經元的網路。
- 啟動函數設定為 `relu`。
- 最後使用一個神經元輸出（表示正評或負評）。
- 最後一層輸出啟動函數為 `sigmoid`。

training and validation



R demo

大型資料處理 & 大量資料處理

2300萬筆資料處理

大型資料準備

```
# 大型資料處理 (data.table)
# 建立2300萬筆模擬資料
working <- "C:/rdata"
setwd(working)
getwd()
datasize <- 23000000
mydata <- matrix(c(NA), nrow=datasize, ncol=5)
set.seed(168)
mydata[,1] <- sample(c(1:17770), datasize, replace = TRUE)
mydata[,2] <- sample(c(1:480189), datasize, replace = TRUE)
mydata[,3] <- sample(c(1:5), datasize, replace = TRUE)
mydata[,4] <- sample(c(1999:2014), datasize, replace = TRUE)
mydata[,5] <- sample(c(1:12), datasize, replace = TRUE)
colnames(mydata) <- c("movie", "customer", "rating", "year", "month")
write.table(mydata, file="bigdata.txt", sep=" ", row.names=FALSE, col.names=TRUE)
# 2300萬*5, 491MB

# Ctrl + Shift + F10: 重新啟動R
> system.time(bigdata1 <- read.csv("C:/rdata/bigdata.txt", header=TRUE))
    user  system elapsed
  86.78     1.36   88.14
```

讀取資料 1分鐘多

fread {data.table} 1.x秒

```
> # Ctrl + Shift + F10: 重新啟動R
> library(data.table)
data.table 1.14.0 using 4 threads (see ?getDTthreads). Latest news: r-datarable.com
> system.time(movies <- fread("C:/rdata/bigdata.txt"))
      user   system elapsed
    3.40     0.42   1.22
> dim(movies) # 23000000*5
[1] 23000000      5
> summary(movies)
      movie           customer        rating       year      month
Min. : 1 Min. : 1 Min. :1 Min. :1999 Min. : 1.0
1st Qu.: 4445 1st Qu.:120045 1st Qu.:2 1st Qu.:2002 1st Qu.: 4.0
Median : 8887 Median :240080 Median :3 Median :2007 Median : 7.0
Mean   : 8887 Mean   :240080 Mean   :3 Mean   :2006 Mean   : 6.5
3rd Qu.:13330 3rd Qu.:360077 3rd Qu.:4 3rd Qu.:2010 3rd Qu.: 9.0
Max.  :17770  Max.  :480189 Max.  :5 Max.  :2014 Max.  :12.0
> class(movies)
[1] "data.table" "data.frame"
```

大量資料處理

```
> # 大量資料處理
> working_path <- "C:/rdata"
> setwd(working_path)
> getwd()
[1] "C:/rdata"
> sample1 <- iris[sample(1:nrow(iris),10),]
> sample2 <- iris[sample(1:nrow(iris),10),]
> sample3 <- iris[sample(1:nrow(iris),10),]
> write.table(sample1, file="sample1.csv", sep=",", row.names=FALSE)
> write.table(sample2, file="sample2.csv", sep=",", row.names=FALSE)
> write.table(sample3, file="sample3.csv", sep=",", row.names=FALSE)
> files <- dir(getwd(), pattern="sample.*.csv", recursive=TRUE, full.names=TRUE)
> files
[1] "C:/rdata/sample1.csv" "C:/rdata/sample2.csv" "C:/rdata/sample3.csv"
> tables <- lapply(files, read.table, header=TRUE, sep=",") # list
> sample.all <- do.call(rbind, tables) # data.frame
> sample.all
   Sepal.Length Sepal.Width Petal.Length Petal.Width    Species
1          5.1        3.4         1.5       0.2     setosa
2          6.1        3.0         4.9       1.8  virginica
3          6.0        3.4         4.5       1.6 versicolor
4          6.7        3.3         5.7       2.5  virginica
5          5.2        3.4         1.4       0.2     setosa
6          5.7        2.8         4.1       1.3 versicolor
```

10. 基礎互動式shiny, 進階互動式shiny Server佈署



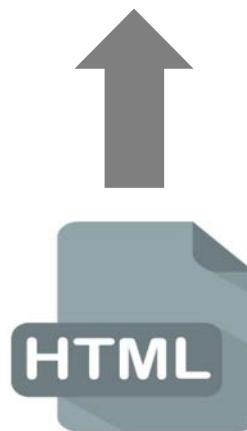
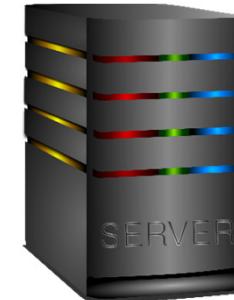
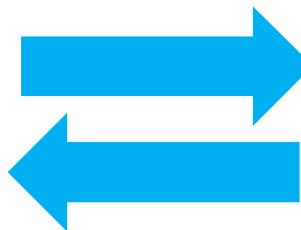
基礎互動式shiny

- <https://shiny.rstudio.com/>
- <https://www.rstudio.com/resources/webinars/>

shiny 簡介

- 隨著資訊科技技術的進步，如何提供WEB化的應用服務。
- shiny 套件提供更方便，互動式與動態更新等應用
 - 輸入 - 文字方塊、下拉式選單、按鈕。
 - 處理 - 執行 R 運算、建立模型。
 - 輸出 - 網頁中呈現文字摘要、表格與圖表等。

shiny (server) 架構



ui.R



server.R

CRAN shiny 套件

shiny: Web Application Framework for R

互動式網頁應用

Makes it incredibly easy to build interactive web applications with R. Automatic "reactive" binding between inputs and outputs and extensive prebuilt widgets make it possible to build beautiful, responsive, and powerful applications with minimal effort.

Version: 1.6.0

Depends: R (\geq 3.0.2), methods

Imports: utils, grDevices, [httpuv](#) (\geq 1.5.2), [mime](#) (\geq 0.3), [jsonlite](#) (\geq 0.9.16), [xtable](#), [digest](#) (\geq 0.6.25), [htmltools](#) (\geq 0.5.0.9001), [R6](#) (\geq 2.0), [sourcetools](#), [later](#) (\geq 1.0.0), [promises](#) (\geq 1.1.0), tools, [crayon](#), [rlang](#) (\geq 0.4.9), [fastmap](#) (\geq 1.0.0), [withr](#), [commonmark](#) (\geq 1.7), [glue](#) (\geq 1.3.2), [bslib](#) (\geq 0.2.2.9002), [cachem](#), [ellipsis](#), [lifecycle](#) (\geq 0.2.0)

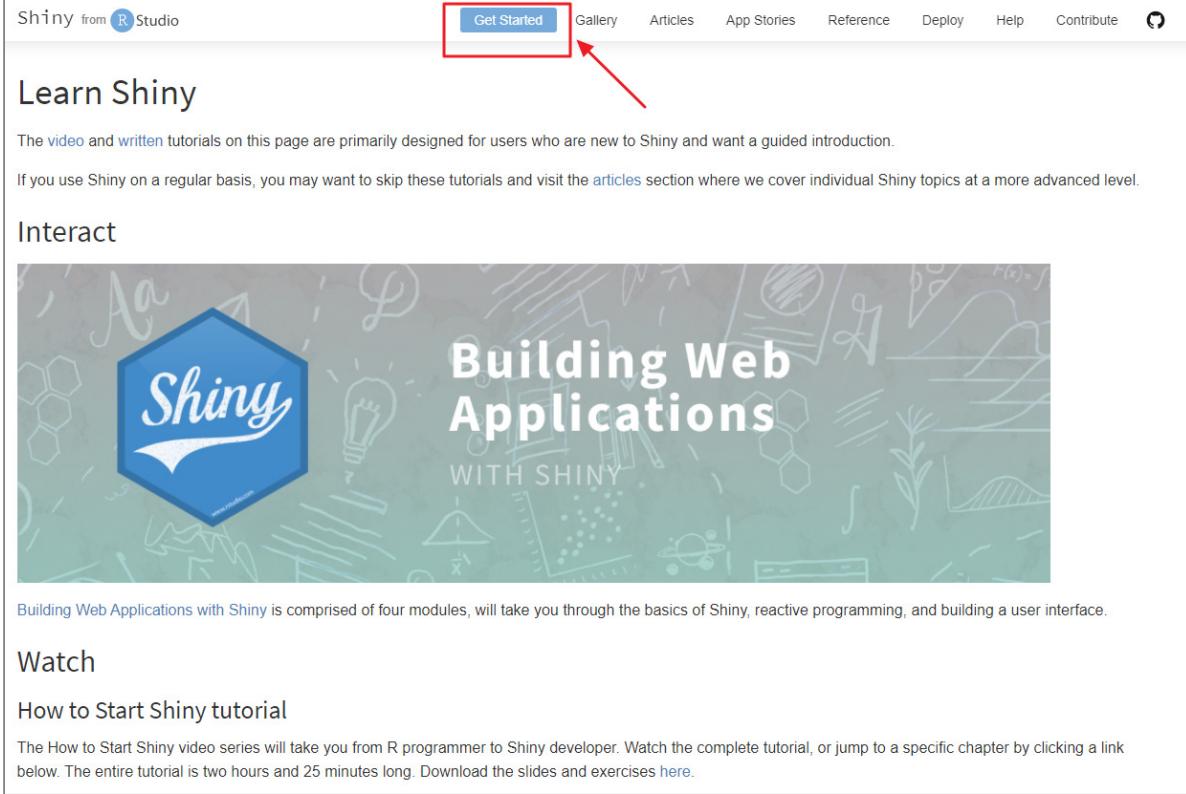
Suggests: datasets, [Cairo](#) (\geq 1.5-5), [testthat](#) (\geq 2.1.1), [knitr](#) (\geq 1.6), [markdown](#), [rmarkdown](#), [ggplot2](#), [reactlog](#) (\geq 1.0.0), [magrittr](#), [shinytest](#) (\geq 1.4.0.9003), [yaml](#), [future](#), [dygraphs](#), [ragg](#), [showtext](#), [sass](#)

Published: 2021-01-25

Author: Winston Chang [aut, cre], Joe Cheng [aut], JJ Allaire [aut], Carson Sievert [aut], Barret Schloerke [aut], Yihui Xie [aut], Jeff Allen [aut],

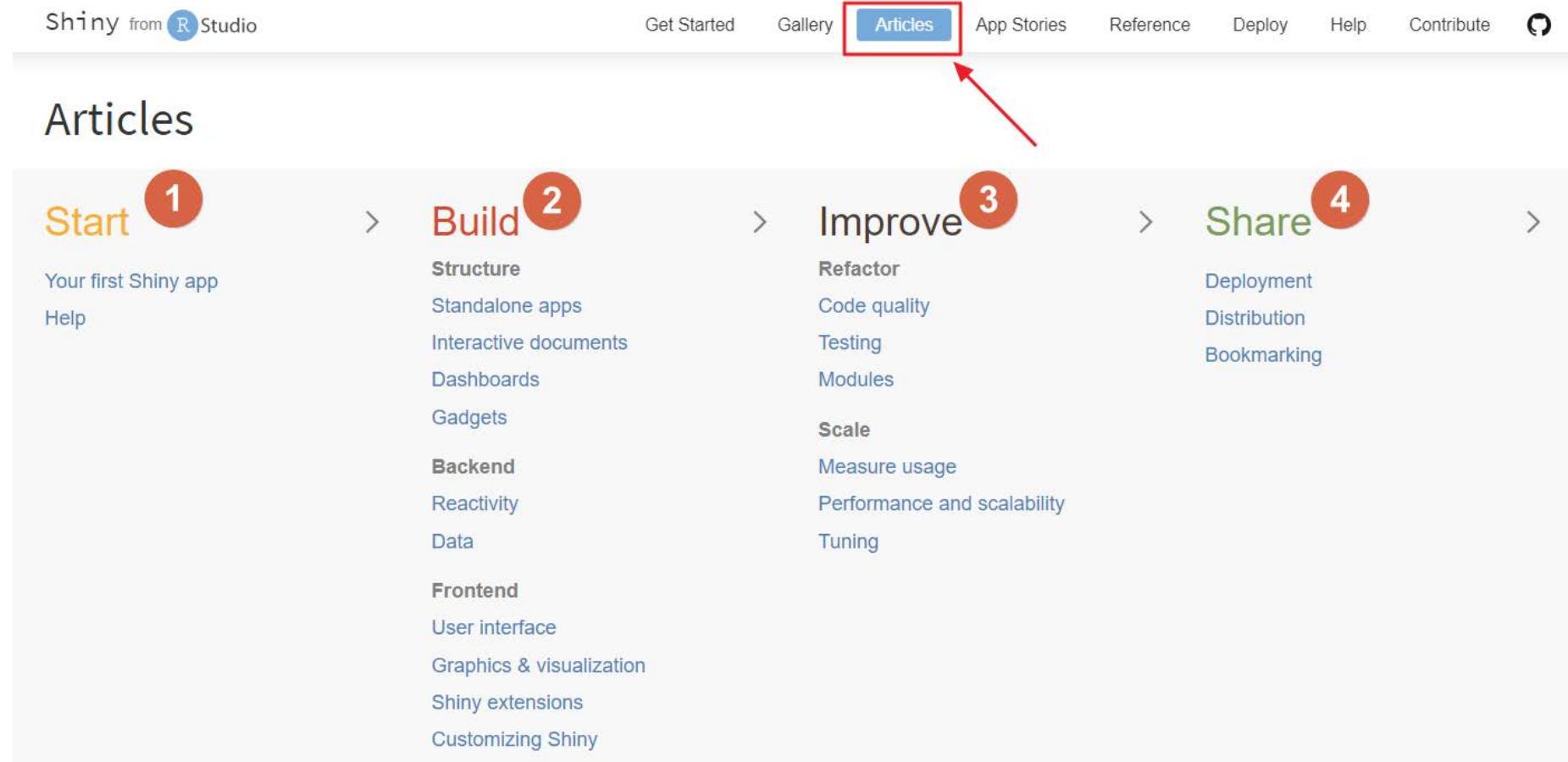
Shiny 參考資料-Get Started

- <http://shiny.rstudio.com/tutorial/>



The screenshot shows the 'Get Started' section of the Shiny website. At the top, there's a navigation bar with links: 'Get Started' (which is highlighted with a red box and has a red arrow pointing to it), 'Gallery', 'Articles', 'App Stories', 'Reference', 'Deploy', 'Help', and 'Contribute'. Below the navigation, the page title is 'Learn Shiny'. It contains two paragraphs of text: 'The video and written tutorials on this page are primarily designed for users who are new to Shiny and want a guided introduction.' and 'If you use Shiny on a regular basis, you may want to skip these tutorials and visit the [articles](#) section where we cover individual Shiny topics at a more advanced level.' Below this, there's a section titled 'Interact' featuring a large image of the Shiny logo (a blue hexagon with the word 'Shiny' in white) overlaid on a background of various scientific and mathematical icons like a lightbulb, a graph, and chemical structures. To the right of the logo, the text 'Building Web Applications WITH SHINY' is visible. At the bottom of the page, there's a section titled 'Watch' with a link to 'How to Start Shiny tutorial'. A note below it says: 'The How to Start Shiny video series will take you from R programmer to Shiny developer. Watch the complete tutorial, or jump to a specific chapter by clicking a link below. The entire tutorial is two hours and 25 minutes long. Download the slides and exercises [here](#)'.

Shiny 參考資料-Articles



Shiny from RStudio

Get Started Gallery **Articles** App Stories Reference Deploy Help Contribute 

Articles

Start ① > **Build** ② > **Improve** ③ > **Share** ④ >

Start Your first Shiny app Help	Build Structure Standalone apps Interactive documents Dashboards Gadgets Backend Reactivity Data Frontend User interface Graphics & visualization Shiny extensions Customizing Shiny	Improve Refactor Code quality Testing Modules Scale Measure usage Performance and scalability Tuning	Share Deployment Distribution Bookmarking
--	--	--	---

shiny 內建範例資料

- 範例(examples)目錄

R\R-4.x.x\library\shiny\examples

```
> # 顯示內建11個範例
> dir(paste0(.libPaths(), "/shiny/examples"))
[1] "01_hello"          "02_text"           "03_reactivity"   "04_mpg"
[5] "05_sliders"        "06_tabssets"       "07_widgets"      "08_html1"
[9] "09_upload"         "10_download"      "11_timer"
>
```

內建11個shiny範例

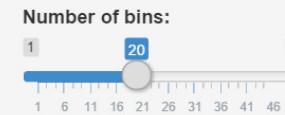
shiny 範例 - 01_hello

```
> # shiny example - 01_hello  
> library(shiny)  
> runExample("01_hello")
```

- 左側: user-interface

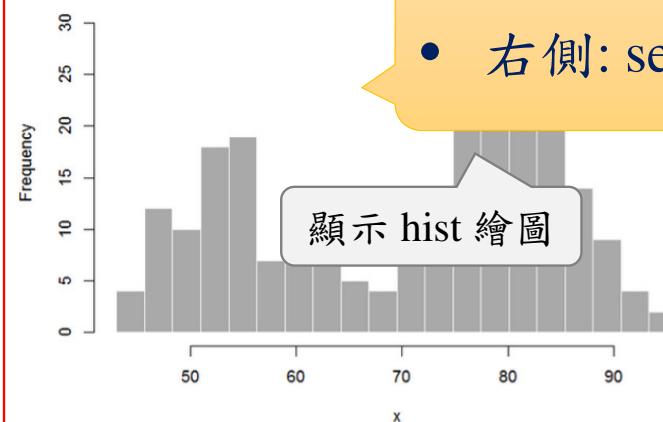
- runExample("範例名稱")
- 區分ui.R, server.R
- 亦可合併為 app.R

Hello Shiny!



slider: 可調
整樣本大小

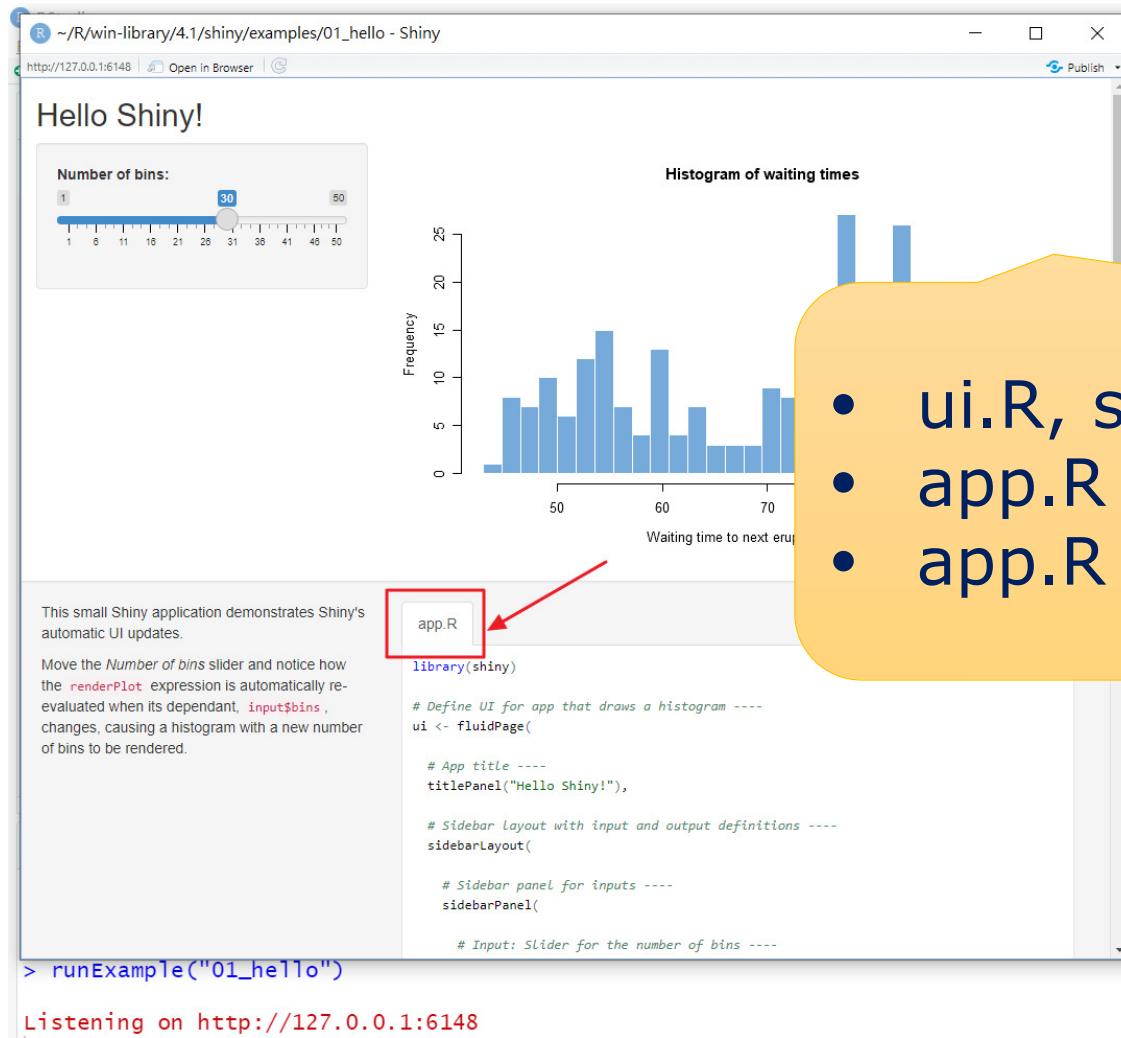
Histogram of x



- 右側: server-interface

顯示 hist 繪圖

方法1-app.R



~ /R/win-library/4.1/shiny/examples/01_hello - Shiny
http://127.0.0.1:6148 | Open in Browser | Publish

Hello Shiny!

Number of bins:

Histogram of waiting times

Frequency

Waiting time to next eruption

This small Shiny application demonstrates Shiny's automatic UI updates.

Move the Number of bins slider and notice how the `renderPlot` expression is automatically re-evaluated when its dependant, `input$bins`, changes, causing a histogram with a new number of bins to be rendered.

```
library(shiny)

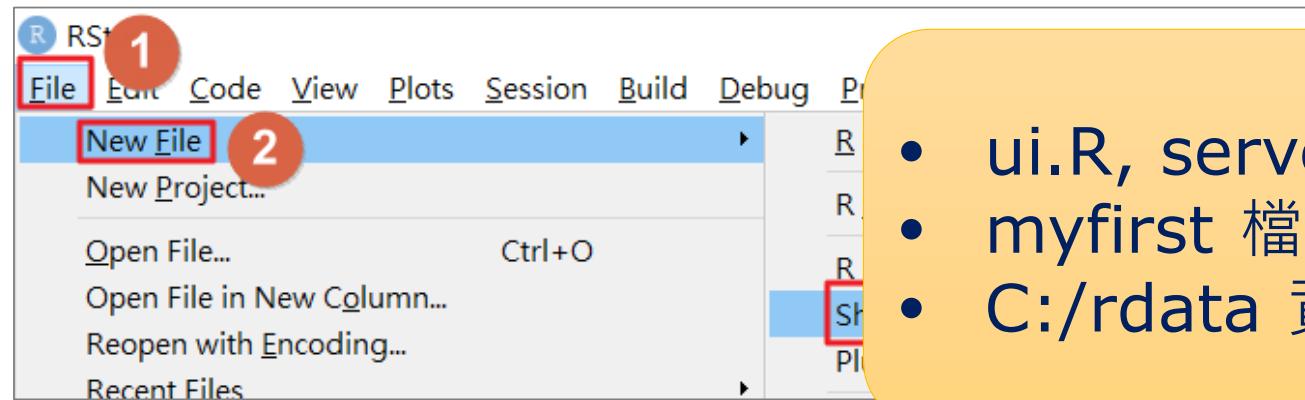
# Define UI for app that draws a histogram ----
ui <- fluidPage(
  # App title ----
  titlePanel("Hello Shiny!"),

  # Sidebar Layout with input and output definitions ----
  sidebarLayout(
    # Sidebar panel for inputs ----
    sidebarPanel(
      # Input: Slider for the number of bins ----
      sliderInput("bins", "Number of bins:", min = 1, max = 50, value = 30)
    ),
    # Main panel for displaying outputs ----
    mainPanel(
      # Output: Histogram ----
      renderPlot(histogram(~eruptions, bins = input$bins))
    )
  )
)
```

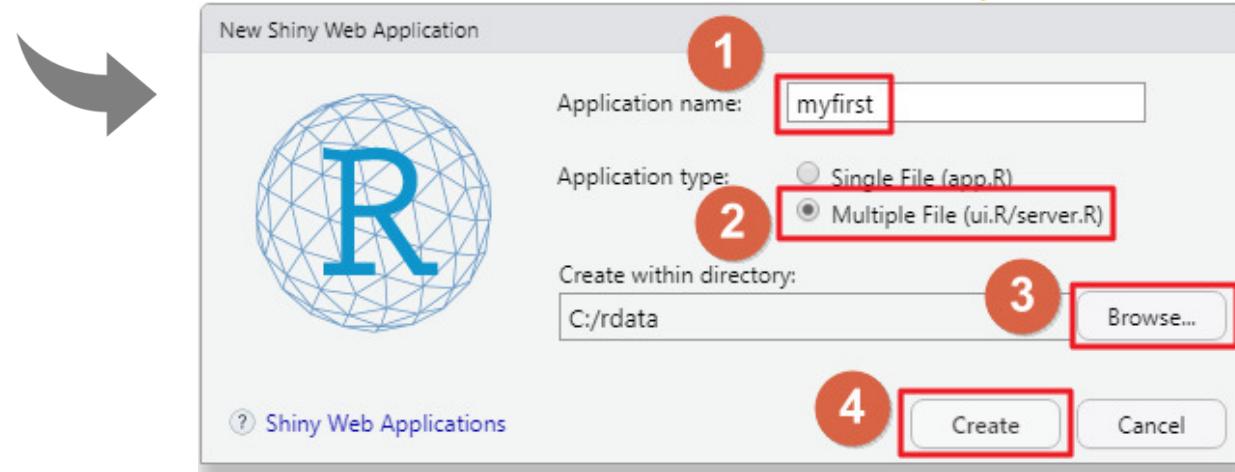
> runExample("01_hello")
Listening on http://127.0.0.1:6148

- ui.R, server.R 檔名不可更換
- app.R 檔名可以更換
- app.R 適用於小程序式方便測試

方法2-Multiple File(ui.R/server.R)



- ui.R, server.R 檔名不可更換
- myfirst 檔名可以更換
- C:/rdata 資料夾名稱可更換



ui.R

```
10 library(shiny)          fluidPage( )  
11  
12 # Define UI for application that draws a histogram  
13 shinyUI(fluidPage(  
14  
15   # Application title  
16   titlePanel("old Faithful Geyser Data"),  
17  
18   # sidebar with a slider input for number of bins  
19   sidebarLayout(  
20     sidebarPanel(  
21       sliderInput("bins",  
22         "Number of bins:",  
23         min = 1,  
24         max = 50,  
25         value = 30)  
26     ),  
27  
28   # Show a plot of the generated distribution  
29   mainPanel(  
30     plotOutput("distPlot")  
31   )  
32 )  
33 ))
```

1

titlePanel: 標題

2

sidebarInput: 滑桿(輸入功能)

輸入參數 bins

value 預設值

輸出採用 distPlot 參數

server.R

```
10 library(shiny)
11
12 # Define server logic required to draw a histogram
13 shinyServer(function(input, output) {
14
15   output$distPlot <- renderPlot({
16
17     # generate bins based on input$bins from ui.R
18     x      <- faithful[, 2]
19     bins <- seq(min(x), max(x), length.out = input$bins + 1)
20
21     # draw the histogram with the specified number of bins
22     hist(x, breaks = bins, col = 'darkgray', border = 'white')
23
24   })
25
26 })
```

- distPlot 繪圖參數

renderPlot

- 輸入參數: **x** 內建資料集 faithful
- 輸入參數: 使用者輸入 **bins**

shiny 範例 - 02_text

Shiny Text

Choose a dataset:

rock

Number of observations to view:

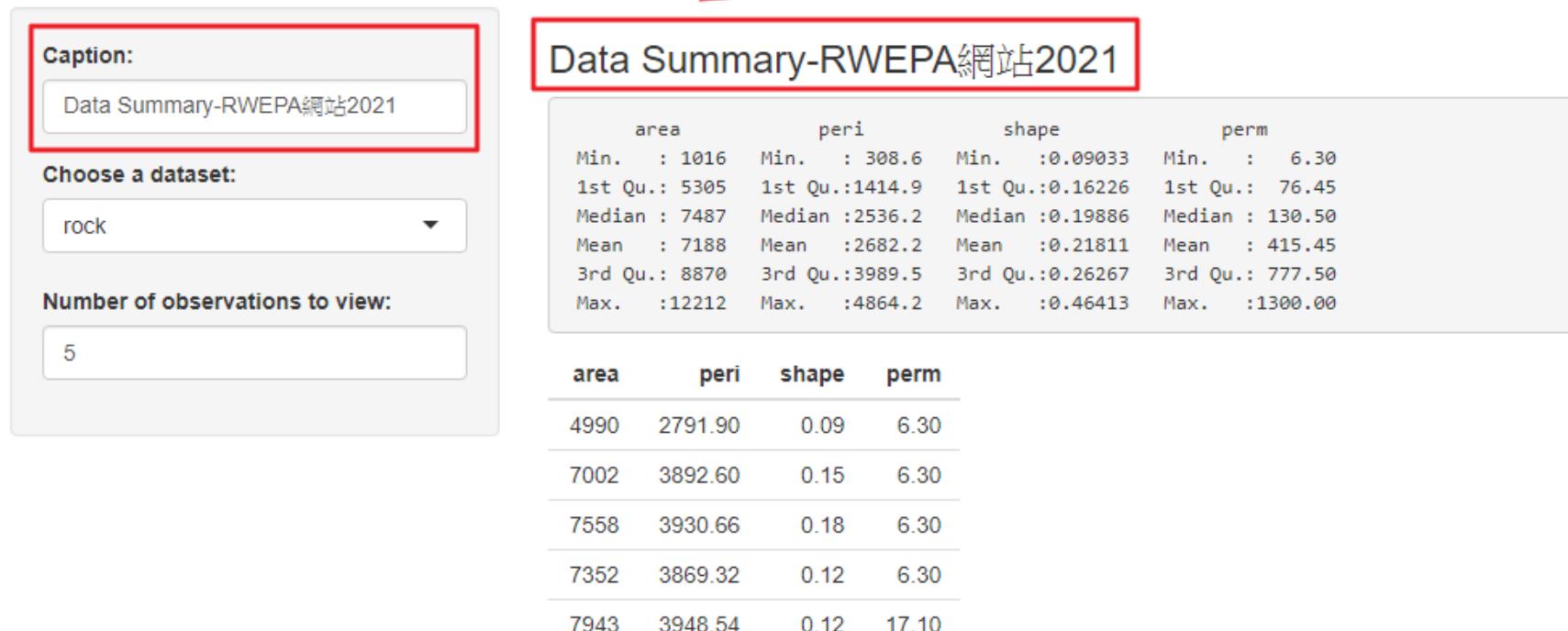
10

area	peri	shape	perm
Min. : 1016	Min. : 308.6	Min. : 0.09033	Min. : 6.30
1st Qu.: 5305	1st Qu.: 1414.9	1st Qu.: 0.16226	1st Qu.: 76.45
Median : 7487	Median : 2536.2	Median : 0.19886	Median : 130.50
Mean : 7188	Mean : 2682.2	Mean : 0.21811	Mean : 415.45
3rd Qu.: 8870	3rd Qu.: 3989.5	3rd Qu.: 0.26267	3rd Qu.: 777.50
Max. :12212	Max. : 4864.2	Max. : 0.46413	Max. : 1300.00

area	peri	shape	perm
4990	2791.90	0.09	6.30
7002	3892.60	0.15	6.30
7558	3930.66	0.18	6.30
7352	3869.32	0.12	6.30
7943	3948.54	0.12	17.10
7979	4010.15	0.17	17.10
9333	4345.75	0.19	17.10
8209	4344.75	0.16	17.10
8393	3682.04	0.20	119.00
6425	3098.65	0.16	119.00

shiny 範例 - 03_reactivity

Reactivity



The diagram illustrates the concept of reactivity in a shiny application. A red arrow points from the 'Caption' input field in the left panel to the 'Data Summary-RWEPA網站2021' output in the right panel, indicating that the output is dynamically generated based on the user's input.

Caption:
Data Summary-RWEPA網站2021

Choose a dataset:
rock

Number of observations to view:
5

Data Summary-RWEPA網站2021

area	peri	shape	perm
Min. : 1016	Min. : 308.6	Min. : 0.09033	Min. : 6.30
1st Qu.: 5305	1st Qu.: 1414.9	1st Qu.: 0.16226	1st Qu.: 76.45
Median : 7487	Median : 2536.2	Median : 0.19886	Median : 130.50
Mean : 7188	Mean : 2682.2	Mean : 0.21811	Mean : 415.45
3rd Qu.: 8870	3rd Qu.: 3989.5	3rd Qu.: 0.26267	3rd Qu.: 777.50
Max. : 12212	Max. : 4864.2	Max. : 0.46413	Max. : 1300.00

area	peri	shape	perm
4990	2791.90	0.09	6.30
7002	3892.60	0.15	6.30
7558	3930.66	0.18	6.30
7352	3869.32	0.12	6.30
7943	3948.54	0.12	17.10

建立第一個shiny網頁程式

myFirstShiny

- 以「02_text」為基礎，複製到「C:\rdata\myFirstShiny」資料夾，完成以下功能：
 1. 修改標題為 myFirstShiny
 2. 整合「01_hello」: sidebarInput
 3. 整合「01_hello」: renderPlot
(繪製第1個變數之直方圖)
 4. 將直方圖標題改為選取之資料集名稱



1

myFirstShiny-完成畫面

myFirstShiny

Choose a dataset:

rock

Number of observations to view:

7

Number of bins:

1



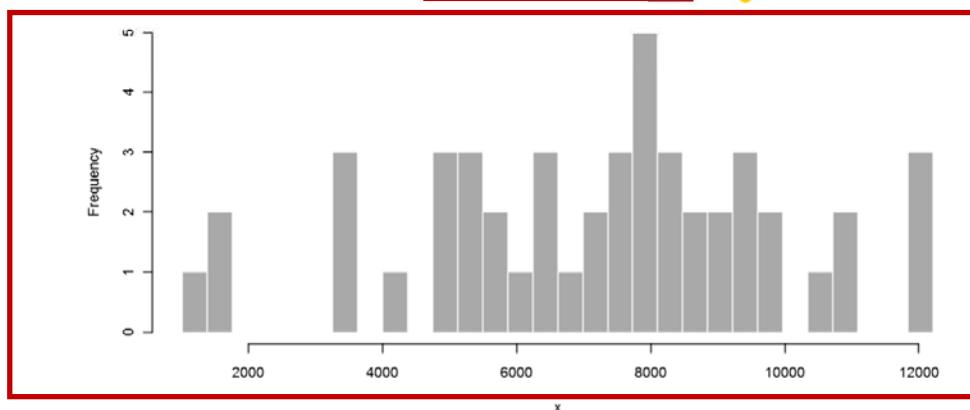
area	peri	shape	perm
Min. : 1016	Min. : 308.6	Min. : 0.09033	Min. : 6.30
1st Qu.: 5305	1st Qu.: 1414.9	1st Qu.: 0.16226	1st Qu.: 76.45
Median : 7487	Median : 2536.2	Median : 0.19886	Median : 130.50
Mean : 7188	Mean : 2682.2	Mean : 0.21811	Mean : 415.45
3rd Qu.: 8870	3rd Qu.: 3989.5	3rd Qu.: 0.26267	3rd Qu.: 777.50
Max. : 12212	Max. : 4864.2	Max. : 0.46413	Max. : 1300.00

實作練習22

2

area	peri	shape	perm
4990	2791.90	0.09	6.30
7002	3892.60	0.15	6.30
7558	3930.66	0.18	6.30
7352	3869.32	0.12	6.30
7943	3948.54	0.12	17.10
7979	4010.15	0.17	17.10
9333	4345.75	0.19	17.10

3

上課練習1
^ ^

4

rock

輸入控制項

輸入控制項

函數	功能
actionButton	每次按鈕+1
checkboxGroupInput	核取方塊-群組
checkboxInput	核取方塊 (用於YES/NO)
dateInput	選取單一日期
dateRangeInput	選取日期範圍 (開始-結束)
fileInput	上傳檔案
helpText	輔助說明
numericInput	數值輸入
radioButtons	選取按鈕 (單選題)
selectInput	下拉式選單
sliderInput	左右滑桿
submitButton	送出按鈕
textInput	文字方塊

selectInput 下拉式選單

selectInput {shiny}

R Documentation

Create a select list input control

Description

Create a select list that can be used to choose a single or multiple items from a list of values.

Usage

① ② ③

```
selectInput(inputId, label, choices, selected = NULL, multiple = FALSE,  
           selectize = TRUE, width = NULL, size = NULL)  
  
selectizeInput(inputId, ..., options = NULL, width = NULL)
```

Arguments

inputId

The input slot that will be used to access the value.

label

Display label for the control, or `NULL` for no label.

choices

List of values to select from. If elements of the list are named then that name rather than the value is displayed to the user.

輸入資料須為list

selectInput 下拉式選單(續)

```
1 ## selectInput example
2 library(shiny)
3 if (interactive()) {
4
5   ui <- fluidPage(
6     selectInput("variable", "variable:",
7                 c("cylinders" = "cy1",
8                   "Transmission" = "am",
9                   "Gears" = "gear")),
10    tableOutput("data")
11  )
12
13  server <- function(input, output) {
14    output$data <- renderTable({
15      mtcars[, c("mpg", input$variable), drop = FALSE]
16    }, rownames = TRUE)
17  }
18
19 shinyApp(ui, server)
20 }
```



萬一資料
筆數很多!@#

檔案上傳

shiny 範例 – 09_upload

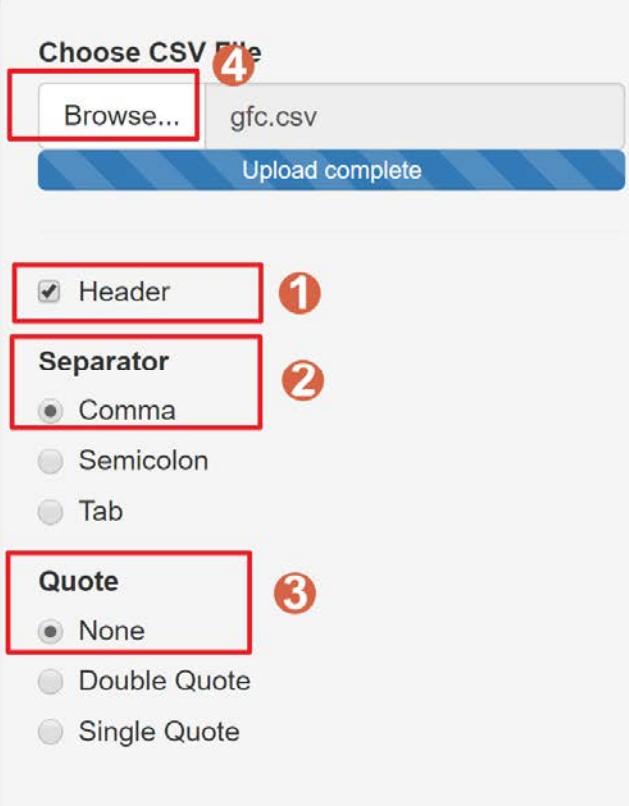
Uploading Files

Choose CSV File **④**
 gfc.csv
Upload complete

Header **①**

Comma **②**
 Semicolon
 Tab

③
 None
 Double Quote
 Single Quote



orderdate	supplier	amount
2009/1/3	日亞	266
2009/1/4	廣達	123
2009/1/5	廣達	66
2009/1/6		
2009/1/7		
2009/1/8		
2009/1/9	廣達	75
2009/1/10	科銳	50
2009/1/10	科銳	48
2009/1/12	科銳	52
2009/1/13	廣達	88
2009/1/14	日亞	240
2009/1/15	科銳	46

• R-程式解說

09_upload - ui.R

```
1 library(shiny)
2
3 fluidPage(
4   titlePanel("Uploading Files"),
5   sidebarLayout(
6     sidebarPanel(
7       fileInput('file1', 'choose CSV File',
8                 accept=c('text/csv',
9                           'text/comma-separated-values,text/plain',
10                          '.csv')),
11      tags$hr(),
12      checkboxInput('header', 'Header', TRUE),
13      radioButtons('sep', 'Separator',
14                    c(Comma=',',
15                      Semicolon=';',
16                      Tab='\t'),
17                    ','),
18      radioButtons('quote', 'Quote',
19                    c(None='',
20                      'Double Quote'="\"",
21                      'single Quote'="'")),
22                    ""))
23    ),
24    mainPanel(
25      tableOutput('contents')
26    )
27  )
28 )
```

fileInput()

09_upload - server.R

```
library(shiny)

function(input, output) {
  output$contents <- renderTable({  
  
  # input$file1 will be NULL initially. After the user selects  
  # and uploads a file, it will be a data frame with 'name',  
  # 'size', 'type', and 'datapath' columns. The 'datapath'  
  # column will contain the local filenames where the data can  
  # be found.  
  
  inFile <- input$file1  
  if (is.null(inFile))  
    return(NULL)  
  
  read.csv(inFile$datapath, header=input$header, sep=input$sep,  
           quote=input$quote)
})  
}
```

技巧:
return(NULL)

輸出控制項-文字/表報



render函數對照表

使用時機	ui.R (依字母排序)	server.R
table	dataTableOutput	renderDataTable
raw HTML	htmlOutput	renderUI
image	imageOutput	renderImage
plot	plotOutput	renderPlot
table	tableOutput	renderTable
text	textOutput	renderText
raw HTML	uiOutput	renderUI
text	verbatimTextOutput	renderPrint

文字/報表

- 參考 `runExample("02_text")` 比對以下表格

ui.R (mainPanel)	server.R
<pre># 文字 verbatimTextOutput("summary"),</pre>	<pre>output\$summary <- renderPrint({ ... })</pre>
<pre># 報表 tableOutput("view")</pre>	<pre>output\$view <- renderTable({ ... })</pre>

輸出控制項 - ggplot2

server.R 採用 renderPlot

ggplot2 使用 renderPlot

```
# 統計圖
output$basicplot <- renderPlot({
  load("data/mydata.RData")
  loadData()

  if (input$select_independent==1) {
    ggplot(mydata, aes(曝露風險, colour=性別, group=性別)) +
      geom_histogram(alpha=0.7, binwidth = 0.01) +
      facet_wrap(~性別, ncol=1)

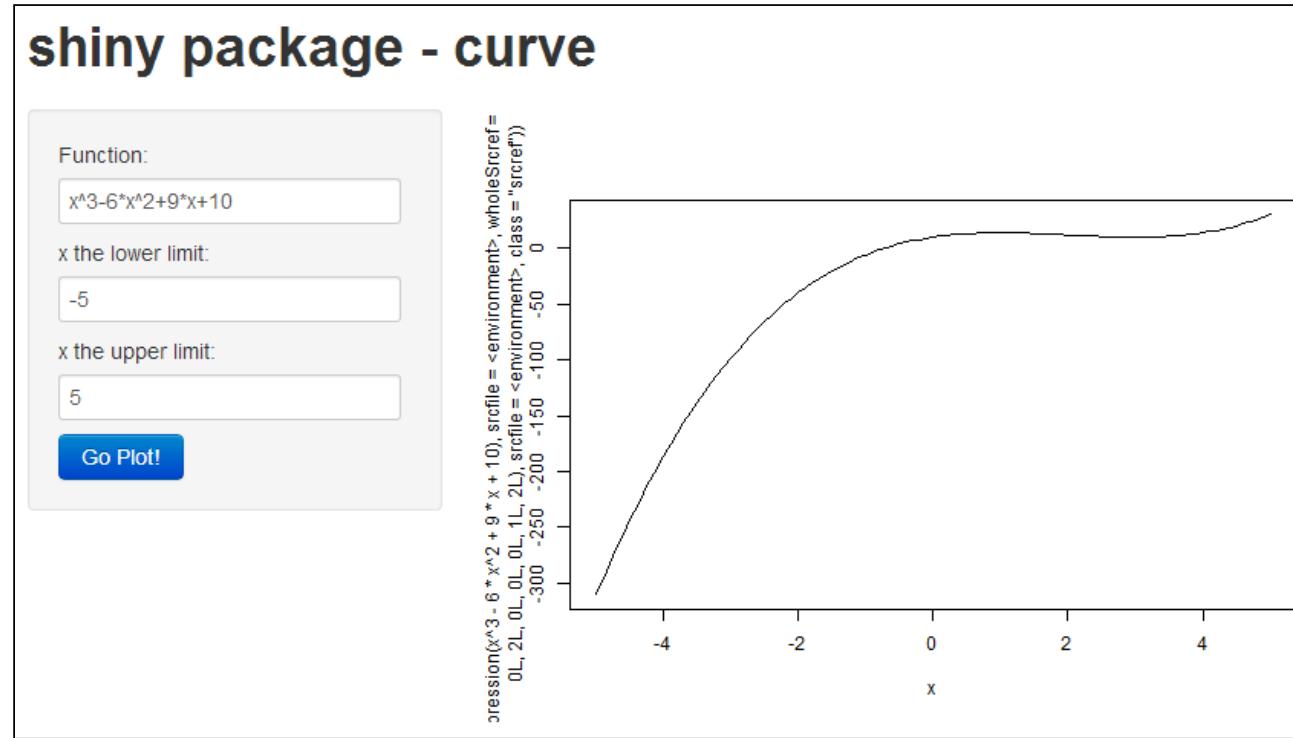
  } else if (input$select_independent==2) {
    ggplot(mydata, aes(曝露風險, colour=車輛種類, group=車輛種類)) +
      geom_histogram(alpha=0.7, binwidth = 0.01) +
      facet_wrap(~車輛種類, ncol=1)

  } else if (input$select_independent==3) {
    ggplot(mydata, aes(曝露風險, colour=被保險人年齡, group=被保險人年齡)) +
      geom_histogram(alpha=0.7, binwidth = 0.01) +
      facet_wrap(~被保險人年齡, ncol=1)

  } else {
    ggplot(mydata, aes(曝露風險, colour=車齡, group=車齡)) +
      geom_histogram(alpha=0.7, binwidth = 0.01) +
      facet_wrap(~車齡, ncol=1)
  }
})|
```

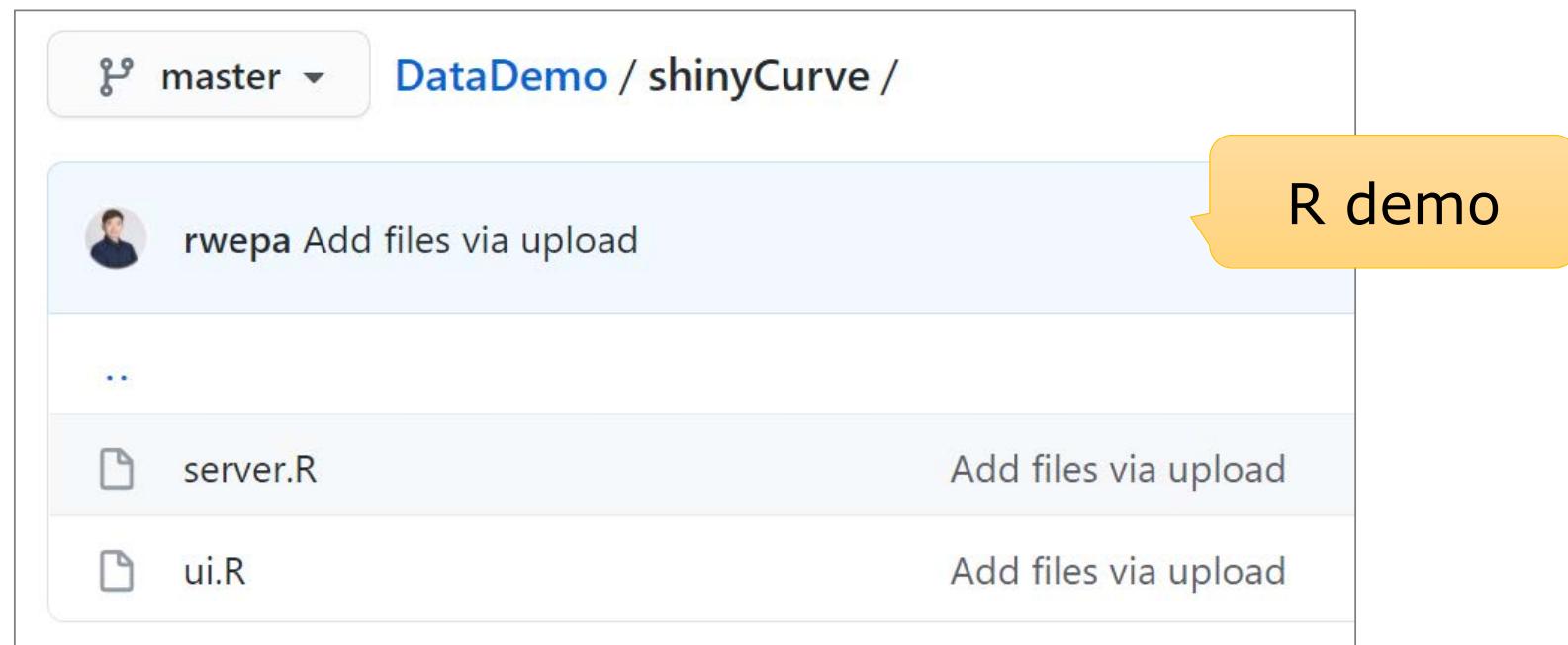
shiny 範例 – shinyCurve - 展示

- <https://rwepa.shinyapps.io/shinyCurve/>



shiny 範例 – shinyCurve - 程式碼

- <https://github.com/rwepa/DataDemo/tree/master/shinyCurve>
- 將 server.R 與 ui.R 複製到工作目錄的單一資料夾中, 例如
「C:\rdata\shinyCurve」



進階互動式shiny Server佈署

版面配置

<http://shiny.rstudio.com/articles/layout-guide.html>

側邊佈局 Sidebar Layout

Hello Shiny!

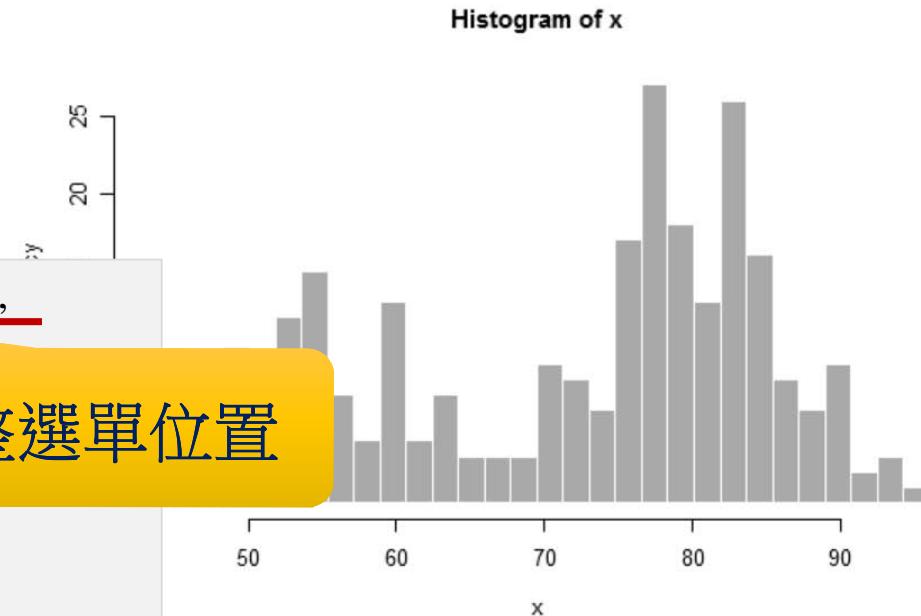
Number of bins:

1 6 11 16 21 26 31 36 41 46 50

```
sidebarLayout(position = "right",
```

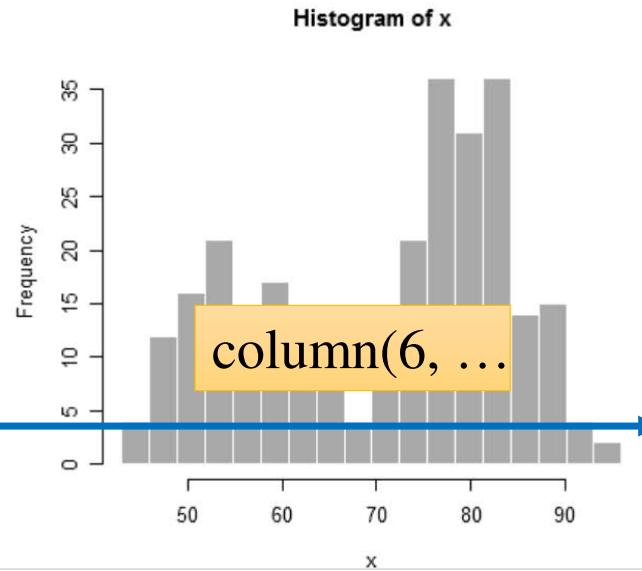
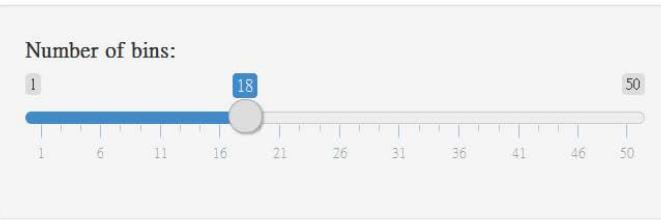
```
  sidebarPanel(  
    # Inputs  
  mainPanel(  
    # Outputs  
)
```

調整選單位置



網格佈局 Grid Layout

Hello Shiny - Grid Layout!



column(6, ...)

Grid Layout – ui.R

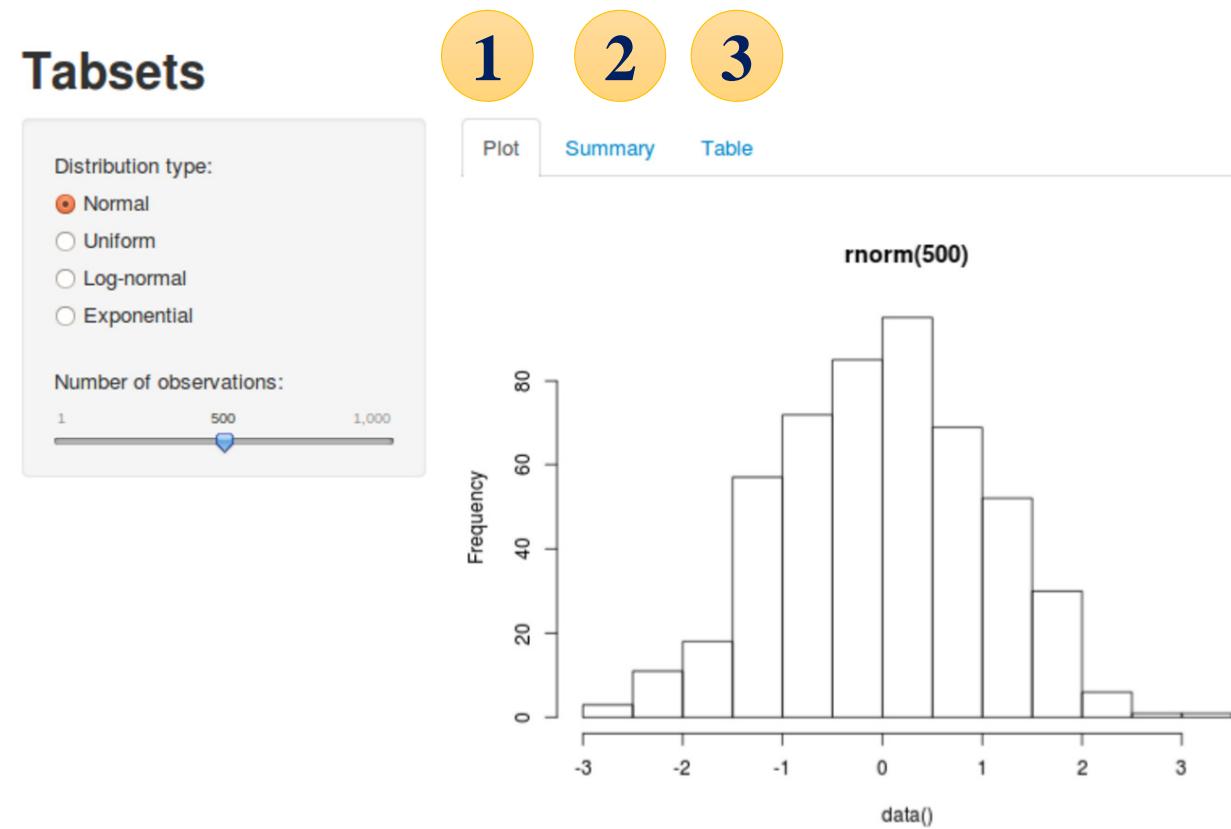
```
library(shiny)

# Define UI for application that draws a histogram
shinyUI(fluidPage(

  titlePanel("Hello shiny - Grid Layout!"),
  fluidRow(
    column(6,
      wellPanel(
        sliderInput("bins", "Number of bins:",
                   min = 1, max = 50, value = 30)
      )
    ),
    column(6,
      plotOutput("distPlot")
    )
  )
))
```

- $6+6=12$
- try: $4+8=12$

分頁佈局 Tabssets Layout



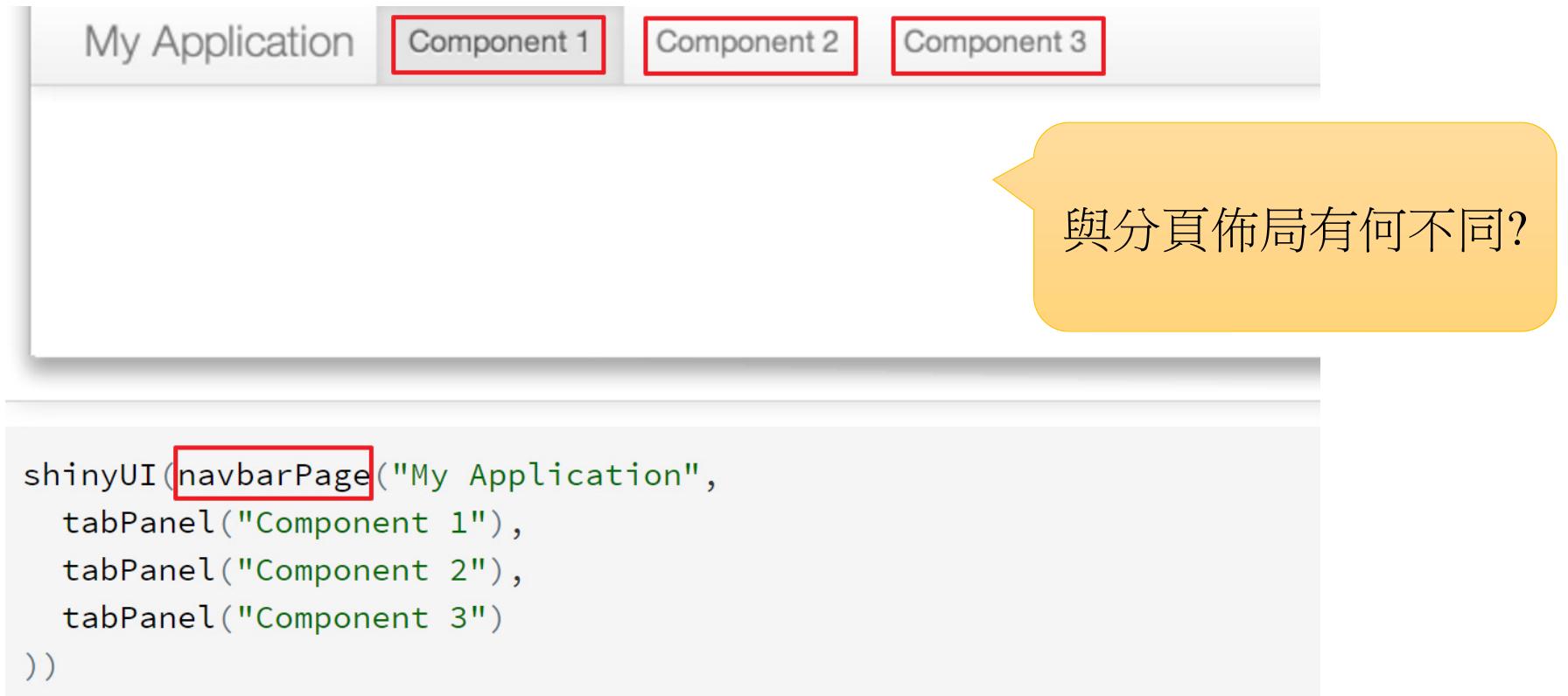
Tabssets Layout – ui.R

```
shinyUI(fluidPage(  
  titlePanel("Tabssets"),  
  sidebarLayout(  
    sidebarPanel(  
      # Inputs excluded for brevity  
    ),  
    mainPanel(  
      tabsetPanel(  
        tabPanel("Plot", plotOutput("plot")),  
        tabPanel("Summary", verbatimTextOutput("summary")),  
        tabPanel("Table", tableOutput("table"))  
      )  
    )  
  ))
```

tabsetPanel()

3個分頁 tabPanel

瀏覽選單佈局 Navbar Pages

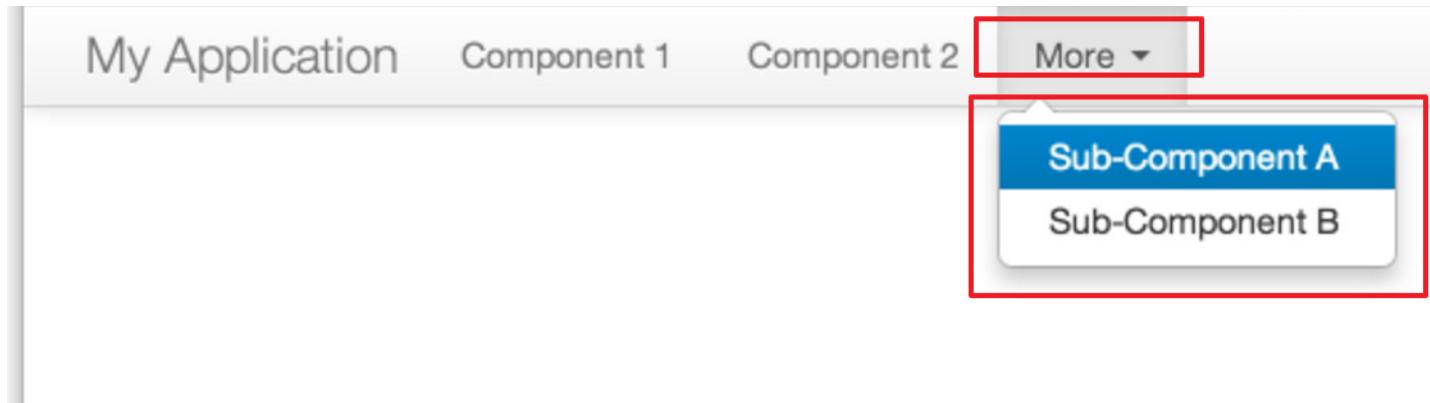


My Application Component 1 Component 2 Component 3

與分頁佈局有何不同?

```
shinyUI(navbarPage("My Application",
  tabPanel("Component 1"),
  tabPanel("Component 2"),
  tabPanel("Component 3")
))
```

瀏覽下拉式選單佈局



The screenshot shows a Shiny application interface. At the top, there is a navigation bar with tabs labeled "My Application", "Component 1", "Component 2", and "More". A red box highlights the "More" tab, which has a dropdown menu. The dropdown menu contains two items: "Sub-Component A" and "Sub-Component B", also highlighted with a red box.

```
shinyUI(navbarPage("My Application",
  tabPanel("Component 1"),
  tabPanel("Component 2"),
  navbarMenu("More",
    tabPanel("Sub-Component A"),
    tabPanel("Sub-Component B"))
))
```

navbarMenu()

HTML UI

客製化選單, 使用時機?

HTML UI- ui.R

```
# 3.QC圖 hr_tt.RData -----
tabPanel("●QC圖",
  fluidRow(
    column(2,
      uioutput("qc_station"),
      uioutput("qc_date")
    ),
    column(10,
      htmloutput("qc_message"),
      plotoutput("qc_chart"),
      DT::dataTableOutput("qc_table_hr_tt"),
      DT::dataTableOutput("qc_table")
    )
  )
),
```

**ui.R 宣告
uiOutput()**

HTML UI- server.R

```
# 選取觀測站 -----
output$qc_station <- renderUI({  
  
  if (is.null(input$file_upload) == FALSE) {  
    load("data/mystationlist.RData")  
  } else {  
    load("data/TT_target_stationlist.RData")  
  }  
  
  id_name <- levels(factor(paste(mystation$stn_id, mystation$station_name, sep=" ")))  
  
  selectInput("select_qc_station",  
             label="選取測站",  
             choices=id_name)  
})
```

**server.R 實作
renderUI()**

反應型函數 Reactive function

reactive

```
runExample("03_reactivity")
```

```
datasetInput <- reactive({  
  switch(input$dataset,  
    "rock" = rock,  
    "pressure" = pressure,  
    "cars" = cars)  
})
```

isolate

- 使用者不希望達到 reactive
- 使用者按下按鈕後, 才會更新繪圖結果
- isolation: 避免相依性

isolate - demo

ui.R

```
1 shinyUI(pageWithSidebar(  
2   headerPanel("click the button"),  
3   sidebarPanel(  
4     sliderInput("obs", "Number of observations:",  
5                  min = 0, max = 1000, value = 500),  
6     actionButton("goButton", "Go!")  
7   ),  
8   mainPanel(  
9     plotOutput("distPlot")  
10  )  
11 ))
```

server.R

```
1 shinyServer(function(input, output) {  
2   output$distPlot <- renderPlot({  
3     # Take a dependency on input$goButton  
4     input$goButton  
5  
6     # Use isolate() to avoid dependency on input$obs  
7     dist <- isolate(rnorm(input$obs))  
8     hist(dist)  
9   })  
10 })
```

shape檔案的輸入與處理

CRAN Task View: Analysis of Spatial Data

- <https://cran.r-project.org/web/views/Spatial.html>

CRAN Task View: **Analysis of Spatial Data**

Maintainer: Roger Bivand

Contact: Roger.Bivand at nhh.no

Version: 2021-06-23

URL: <https://CRAN.R-project.org/view=Spatial>

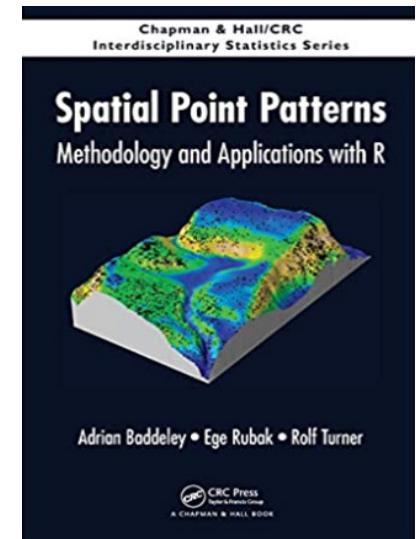
讀取, 視覺化, 分析

Base R includes many functions that can be used for reading, visualising, and analysing spatial data. The focus in this view is on "geographical" spatial data, where observations can be identified with geographical locations, and where additional information about these locations may be retrieved if the location is recorded with care.

Base R functions are complemented by contributed packages, some of which are on CRAN, and others are still in development. One location is [Github](#). Some key packages including [sf](#) and [stars](#) are grouped under [r-spatial](#), others including [raster](#) and [terra](#) under [rspatial](#). Maintenance of the [sp](#) is continuing here: [sp](#).

CRAN Task View: Analysis of Spatial Data (續)

1. Classes for spatial data
2. Handling spatial data
3. Reading and writing spatial data
4. Reading and writing spatial data - other packages
5. Visualization 視覺化
6. Point pattern analysis 點樣式分析
7. Geostatistics 地理統計
8. Disease mapping and areal data analysis 疾病繪圖和區域資料分析
9. Spatial regression 空間迴歸
10. Ecological analysis 生態分析



空間地理視覺化 – maps 套件

地圖 Maps

- 地圖是了解地理資料的重要方法之一
- 地圖-
 - y 軸 - 緯度 latitude
 - x 軸 - 經度 longitude
 - 加上時間維度?



繪圖 - maps 套件

- library(maps) 取得地圖資料

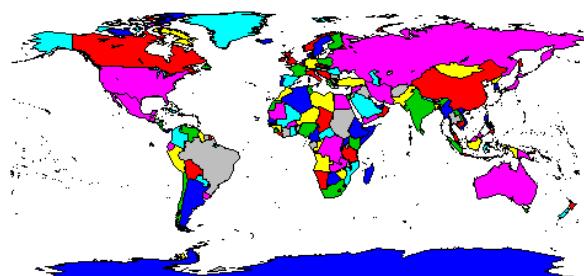
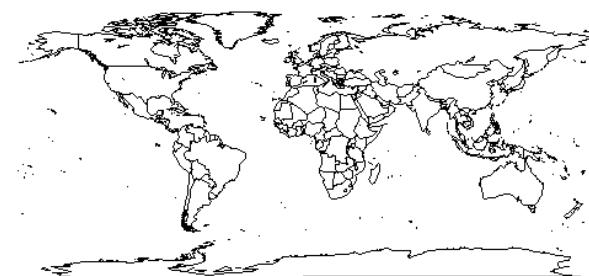
maps 套件
說明

```
> #####  
> # package: maps  
> #####  
> library(maps)  
> help(package="maps")  
> help(map, package="maps")
```

繪製地圖

```
> # draw low resolution map of the world  
> map()  
> map('world', fill=TRUE, col=2:8)
```

map() 繪製地圖



R demo

繪製地圖(續)

```
> # country map  
> map("world", "China")  
> map.cities(country = "China", capitals = 2)
```

map.cities() 加入城市



台灣地圖

```
> # taiwan map  
> map("world", xlim=c(117,123), ylim=c(21.5,26), mar=c(3,2,1,1))  
> # map.cities(country = "Taiwan")  
> map.axes()  
> # str(x.taiwan)  
>  
> # taiwan cities  
> data(world.cities)  
> x <- world.cities  
> x.taiwan <- x[x$country.etc=="Taiwan", ]  
> x.taiwan.city <- x.taiwan[x.taiwan$name %in%  
c("Taipei","Taichung","Kaohsiung"),]  
> map.cities(x.taiwan.city, capital=1) # taipei  
> map.cities(x.taiwan.city, label=TRUE)
```

設定範圍

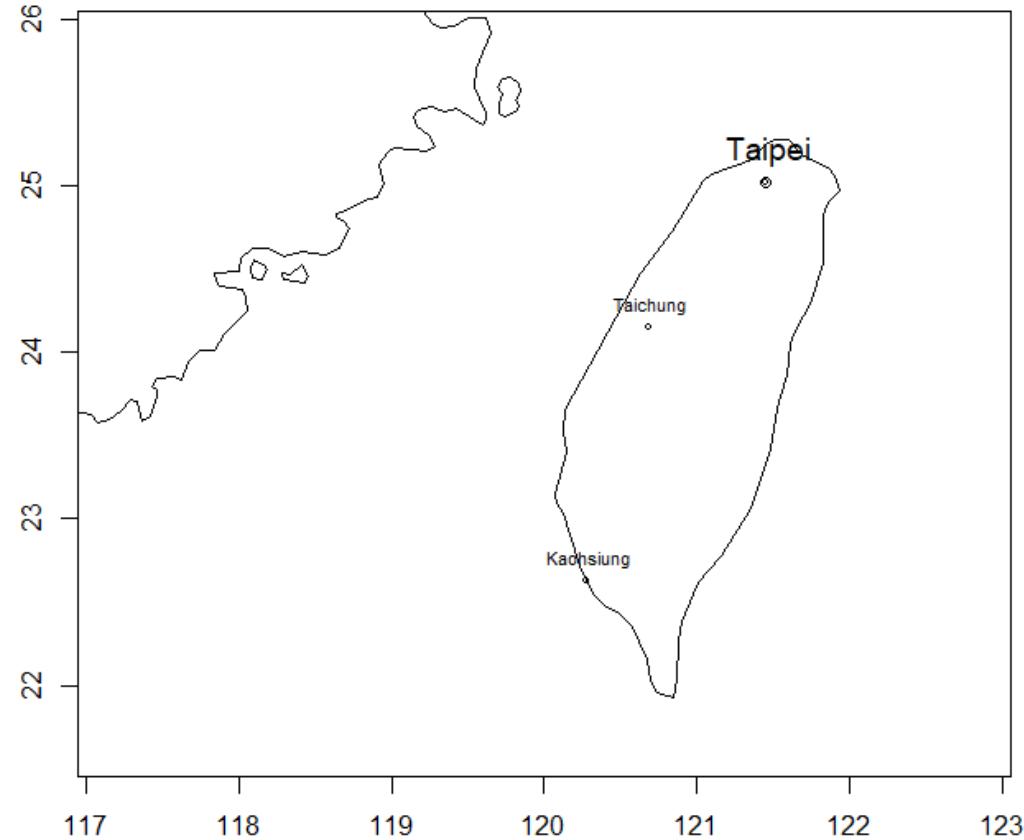
mar 設定邊界

== "Taiwan"

取出 Taiwan 城市

%in%
子集合

台灣地圖(續)



shapefile

- ESRI Shapefile (shp) ，或簡稱shapefile，是美國環境系統研究所公司（ESRI）開發的一種空間資料開放格式。該檔案格式已經成為了地理資訊軟體界的一個開放標準。
- 三個必備檔案(**shp, shx, dbf**):
 - .shp — 圖形格式，用於儲存元素的幾何實體。
 - .shx — 圖形索引格式。幾何體位置索引，記錄每一個幾何體在shp檔案之中的位置，能夠加快向前或向後搜尋一個幾何體的效率。
 - .dbf — 屬性資料格式，以dBase IV的資料表格式儲存每個幾何形狀的屬性資料。

參考資料: <https://zh.wikipedia.org/wiki/Shapefile>

空間地理視覺化—sp 套件

sp 套件

sp: Classes and Methods for Spatial Data

Classes and methods for spatial data; the classes document where the spatial location information resides, for 2D or 3D data. Utility functions are provided, e.g. for plotting data as maps, spatial selection, as well as methods for retrieving coordinates, for subsetting, print, summary, etc.

Version: 1.4-5

Depends: R (\geq 3.0.0), methods

Imports: utils, stats, graphics, grDevices, [lattice](#), grid

Suggests: [RColorBrewer](#), [rgdal](#) (\geq 1.2-3), [rgeos](#) (\geq 0.3-13), [gstat](#), [maptools](#), [deldir](#)

Published: 2021-01-10

Author: Edzer Pebesma [aut, cre], Roger Bivand [aut], Barry Rowlingson [ctb], Virgilio Gomez-Rubio [ctb], Robert Hijmans [ctb], Michael Sumner [ctb], Don MacQueen [ctb], Jim Lemon [ctb], Finn Lindgren [ctb], Josh O'Brien [ctb], Joseph O'Rourke [ctb]

Maintainer: Edzer Pebesma <edzer.pebesma@uni-muenster.de>

sp class

	data type	class	attributes	contains
點	points	SpatialPoints	No	Spatial
	points	SpatialPointsDataFrame	data.frame	SpatialPoints
	multipoints	SpatialMultiPoints	No	Spatial
	multipoints	SpatialMultiPointsDataFrame	data.frame	SpatialMultiPoints
	pixels	SpatialPixels	No	SpatialPoints
	pixels	SpatialPixelsDataFrame	data.frame	SpatialPixels SpatialPointsDataFrame
網格	full grid	SpatialGrid	No	SpatialPixels
	full grid	SpatialGridDataFrame	data.frame	SpatialGrid
線	line	Line	No	
	lines	Lines	No	Line list
	lines	SpatialLines	No	Spatial, Lines list
	lines	SpatialLinesDataFrame	data.frame	SpatialLines
面	polygons	Polygon	No	Line
	polygons	Polygons	No	Polygon list
	polygons	SpatialPolygons	No	Spatial, Polygons list
	polygons	SpatialPolygonsDataFrame	data.frame	SpatialPolygons

空間地理視覺化— maptools 套件

- 鄉鎮市區界線(TWD97經緯度) --- 較詳細
→ <https://data.gov.tw/dataset/7441>
- 直轄市、縣市界線(TWD97經緯度)
→ <https://data.gov.tw/dataset/7442>

台灣 shapefiles

- 鄉鎮市區界線(TWD97經緯度)

<https://data.gov.tw/dataset/7441>

瀏覽次數: 164527 下載次數: 51463 意見數: 38

主要欄位說明

*粗體欄位為資料標準欄位

TOWNID、TOWNCODE、COUNTYNAME、TOWNNAME、TOWNENG、COUNTYID、COUNTYCODE

資料資源下載網址

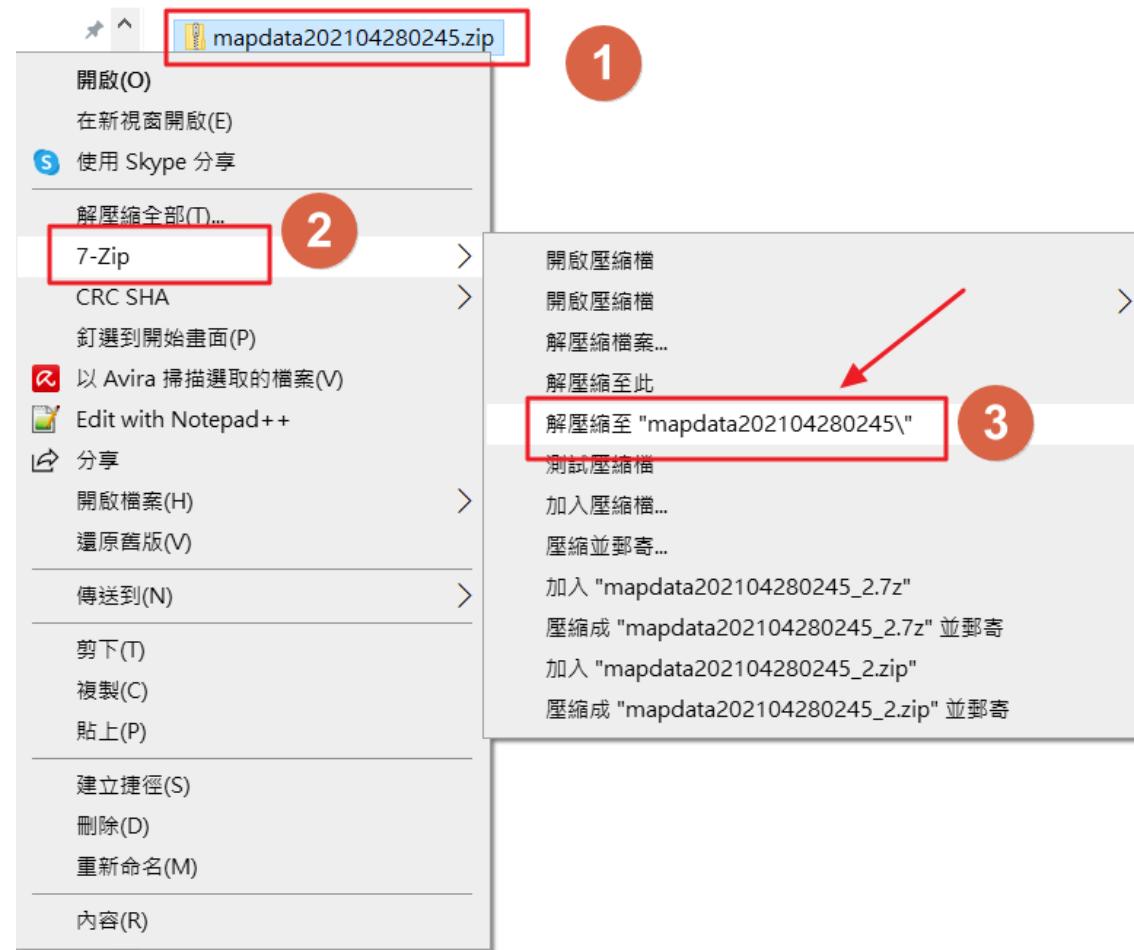


[檢視資料 鄉\(鎮、市、區\)界線\(TWD97經緯度\)SHP格式](#)



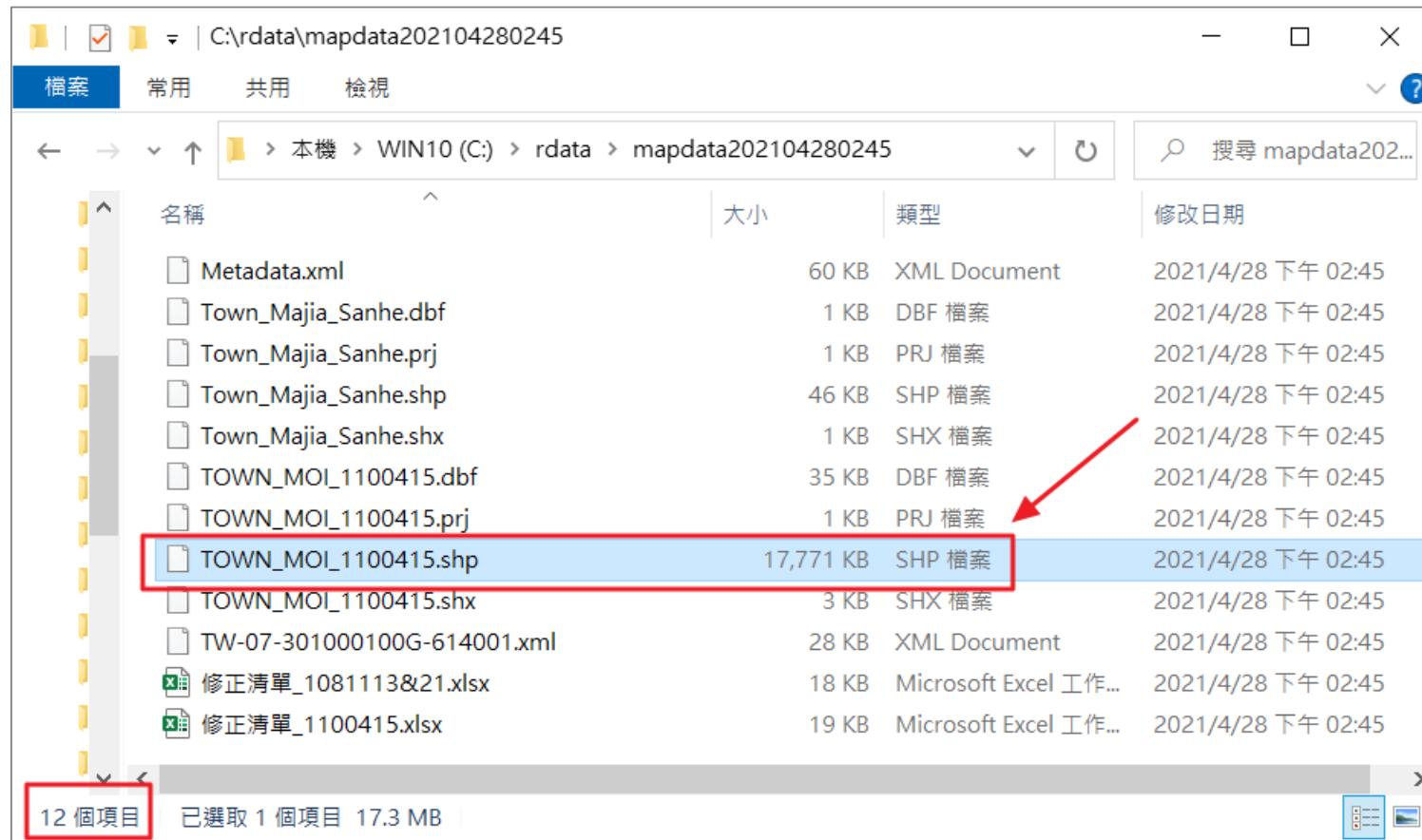
mapdata202104280245.zip

zip解壓縮至 C:/rdata/mapdata202104280245 資料夾



台灣 shapefiles (續)

- TWD97經緯度: 採用「南投埔里之虎子山」為大地基準點.



台灣 shapefiles (續)

- TOWN_MOI_1071226.prj (投影檔)

```
TOWN_MOI_1100415.prj x
1 GEOGCS ["GCS_TWD_1997", DATUM ["D_TWD_1997", SPHEROID ["GRS_1980", 637
8137.0, 298.257222101]], PRIMEM ["Greenwich", 0.0], UNIT ["Degree", 0.0
174532925199433]]
```

下載世界地圖

- <http://www.diva-gis.org/gdata>



The screenshot shows the DIVA-GIS website's "Download data by country" page. The top navigation bar includes links for "Download program", "Documentation", and "Free Spatial Data". Below the navigation is a "Home" link. The main content area features a heading "Download data by country" and a sub-instruction "Select and download free geographic (GIS) data for any country in the world". There are two dropdown menus: one for "Country" set to "Taiwan" and another for "Subject" set to "Administrative areas". A large "OK" button is at the bottom.

readOGR {rgdal}

```
library(maptools)
twn <- readShapePoly("c:/rdata/mapdata202104280245/TOWN_MOI_1100415.shp")
# warning message:
# readShapePoly is deprecated; use rgdal::readOGR or sf::st_read
```

```
library(rgdal)
twn <- readOGR(dsn="c:/rdata/mapdata202104280245", "TOWN_MOI_1100415")
```

1. 資料夾名稱
2. 檔案名稱,不用加 .shp

SpatialPolygonsDataFrame 物件

```
summary(twn)
class(twn) # SpatialPolygonsDataFrame
names(twn) # 7個欄位
names(attributes(twn)) # 7個屬性
```

亂碼?

> head(twn@data) # 有亂碼

	TOWNID	TOWNCODE	COUNTYNAME
0	V02	10014020	<e8>□<e6>□蝮<a3>
1	T21	10013210	撘□□蝮<a3>
2	P13	10009130	<e9>□<e6><9e>□腦
3	V11	10014110	<e8>□<e6>□蝮<a3>
4	V16	10014160	<e8>□<e6>□蝮<a3>
5	N07	10007120	喇啲<8c>□腦

	TOWNNAME
	<e6><88>□<8a>□□
	霆喳□<e9><84><89>
	噪瓦祚<e9><84><89>
	蝮□雀<e9><84><89>
	<e8>□攔潮<84><89>
	<e7>□鉢剝□

	TOWNENG	COUNTYID	COUNTYCODE
	Chenggong Township	V	10014
	Jiadong Township	T	10013
	Mailiao Township	P	10009
	Ludao Township	V	10014
	Lanyu Township	V	10014
	Tianzhong Township	N	10007

@data

```
> # 轉換縣市為正常中文顯示  
> twn@data$COUNTYNAME <- iconv(twn@data$COUNTYNAME, from = "UTF-8", to="UTF-8")  
>  
> # 轉換區鄉縣鎮為正常中文顯示  
> twn@data$TOWNNAME <- iconv(twn@data$TOWNNAME, from = "UTF-8", to="UTF-8")  
> head(twn@data$COUNTYNAME)  
[1] "臺東縣" "屏東縣" "雲林縣" "臺東縣" "臺東縣" "  
> head(twn@data$TOWNNAME)  
[1] "成功鎮" "佳冬鄉" "麥寮鄉" "綠島鄉" "蘭嶼鄉" "  
> str(twn@data)  
'data.frame': 368 obs. of 7 variables:  
 $ TOWNID    : chr  "v02" "T21" "P13" "V11" ...  
 $ TOWNCODE   : chr  "10014020" "10013210" "10009130" "10014110" ...  
 $ COUNTYNAME: chr  "臺東縣" "屏東縣" "雲林縣" "臺東縣"  
 $ TOWNNAME   : chr  "成功鎮" "佳冬鄉" "麥寮鄉" "綠島鄉" ...  
 $ TOWNENG    : chr  "Chenggong Township" "Jia Dong Township" "Ma Liao Township" "Lan Yu Township" ...  
 $ COUNTYID   : chr  "V" "T" "P" "V" ...  
 $ COUNTYCODE: chr  "10014" "10013" "10009" "10014" ...
```

iconv(x, from= , to=)

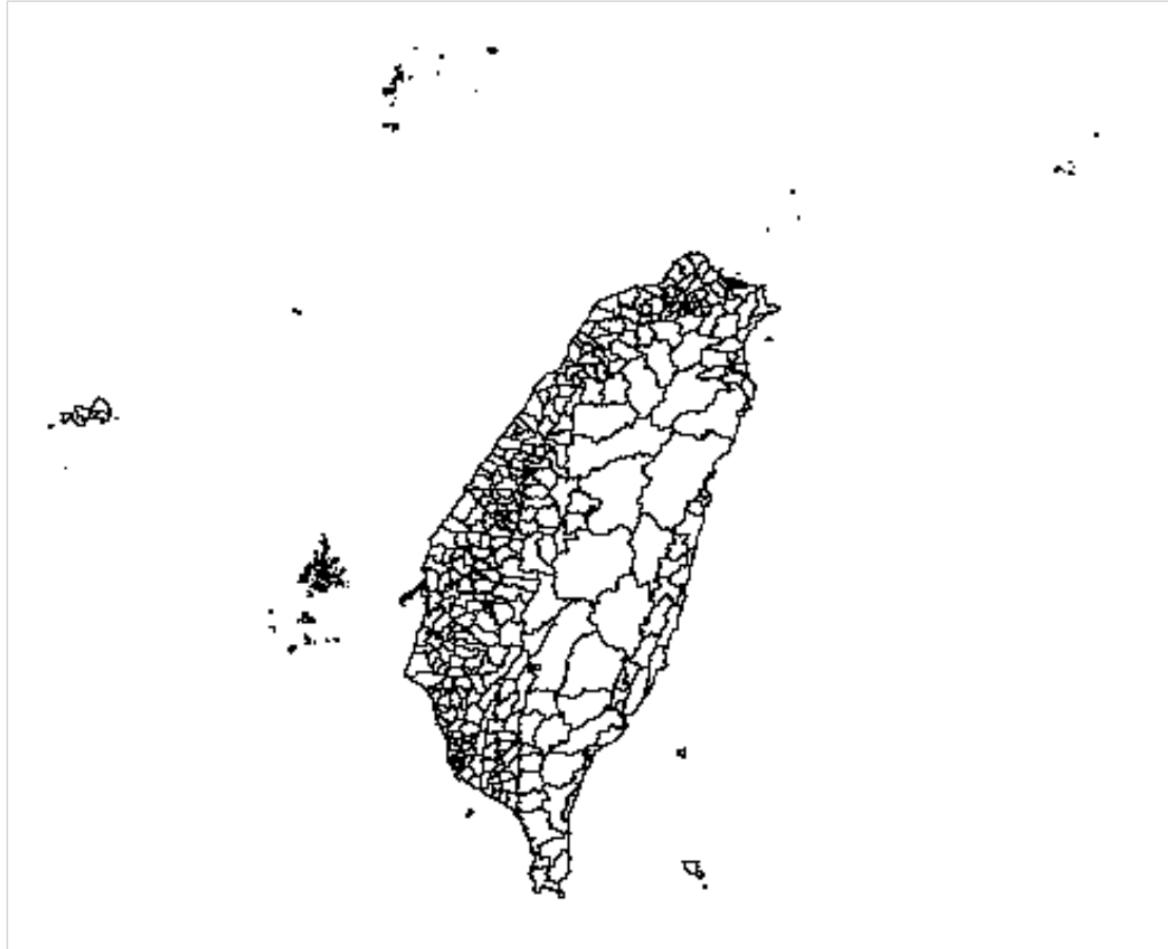
• 地理資料:twn@data

twn@polygons[[1]]

```
> str(twn@polygons[[1]])  
Formal class 'Polygons' [package "sp"] with 5 slots  
..@ Polygons :List of 1  
.. ..$ :Formal class 'Polygon' [package "sp"] with 5 slots  
.. .. . . .@ labpt : num [1:2] 121.4 23.1  
.. .. . . .@ area : num 0.0129  
.. .. . . .@ hole : logi FALSE  
.. .. . . .@ ringDir: int 1  
.. .. . . .@ coords : num [1:3528, 1:2] 121 121 121 121 121 ...  
..@ plotOrder: int 1  
..@ labpt : num [1:2] 121.4 23.1  
..@ ID : chr "0"  
..@ area : num 0.0129  
> twn@polygons[[1]]  
An object of class "Polygons"  
Slot "Polygons":  
[[1]]  
An object of class "Polygon"  
Slot "labpt":  
[1] 121.35410 23.12639  
  
Slot "area":  
[1] 0.0129003  
  
Slot "hole":  
[1] FALSE  
  
Slot "ringDir":  
[1] 1  
  
Slot "coords":  
[,1] [,2]  
[1,] 121.4098 23.21369  
[2,] 121.4098 23.21366  
[3,] 121.4099 23.21362  
[4,] 121.4099 23.21361  
[5,] 121.4100 23.21351  
[6,] 121.4098 23.21346  
[7,] 121.4097 23.21354  
[8,] 121.4096 23.21364  
[9,] 121.4097 23.21375  
[10,] 121.4096 23.21386
```

Polygons 類別

台灣鄉鎮市區界線圖

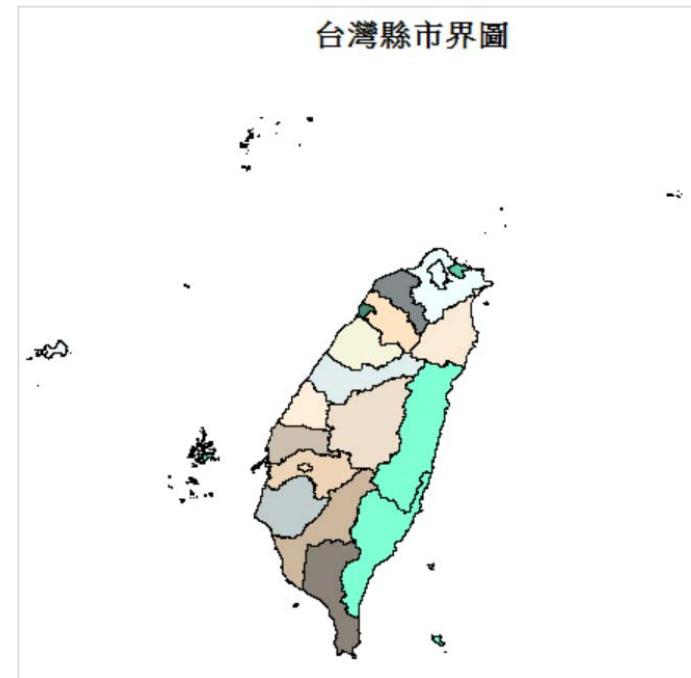




實作練習23

繪製台灣縣市界圖

- <https://data.gov.tw/dataset/7442>
- mapdata202008310842.zip



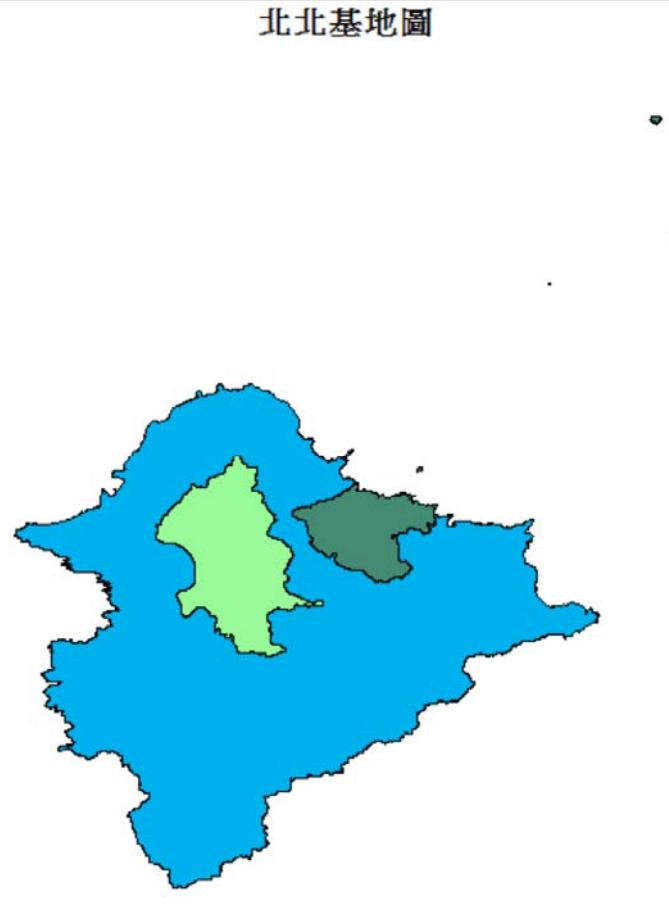
加入地理中心標籤

實作練習24



繪製北北基地圖

實作練習25



地理資料 - shiny進階網頁程式

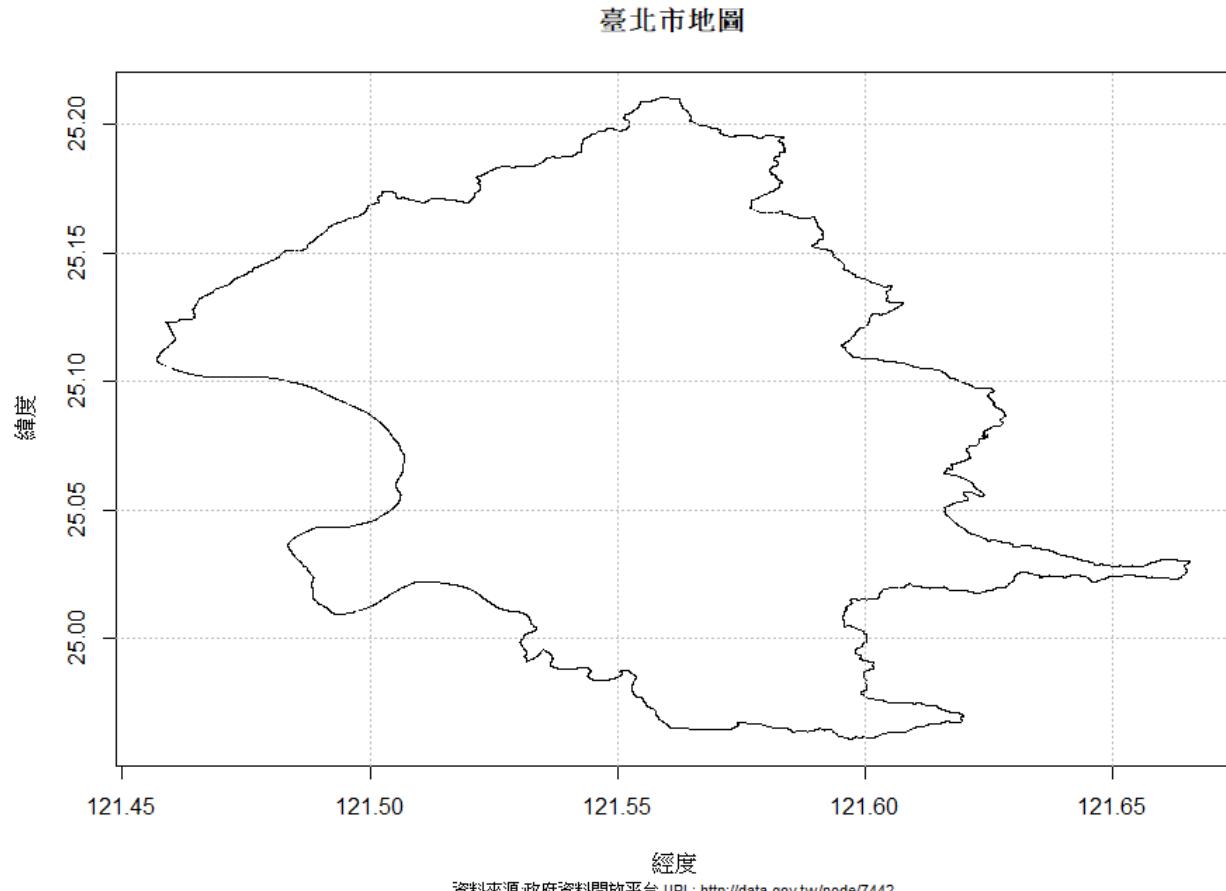
rgdal 套件

readOGR：中文亂碼

```
> library(rgdal)
Loading required package: sp
rgdal: version: 1.2-8, (SVN revision 663)
  Geospatial Data Abstraction Library extensions to
  Loaded GDAL runtime: GDAL 2.0.1, released 2015/09/15
  Path to GDAL shared files: C:/R/R-3.4.1/library/rgdal/gdal
  Loaded PROJ.4 runtime: Rel. 4.9.2, 08 September 2015, [PJ_VERSION: 492]
  Path to PROJ.4 shared files: C:/R/R-3.4.1/library/rgdal/proj
  Linking to sp version: 1.2-5
> twn <- readOGR(dsn="C:/rdata/mapdata201706280308/COUNTY_MOI_1060601", "COUNTY_MOI_1060525")
OGR data source with driver: ESRI Shapefile
Source: "C:/rdata/mapdata201706280308/COUNTY_MOI_1060601", layer: "COUNTY_MOI_1060525"
with 22 features
It has 4 fields
> twn@data
   COUNTYID COUNTYCODE COUNTYNAME      COUNTYENG
0        Z     09007 連江縣 Lienchiang County
1        W     09020 金門縣 Kinmen County
2        G    10002 宜蘭縣 Yilan County
3        N    10007 彰化縣 Changhua County
4        M    10008 南投縣 Nantou County
5        P    10009 雲林縣 Yunlin County
6        T    10013 屏東縣 Pingtung County
7        V    10014 臺東縣 Taitung County
8        U    10015 花蓮縣 Hualien County
9        X    10016 澎湖縣 Penghu County
10       C    10017 基隆市 Keelung City
11       O    10018 新竹市 Hsinchu City
12       A    63000 臺北市 Taipei City
13       F    65000 新北市 New Taipei City
14       B    66000 臺中市 Taichung City
15       D    67000 臺南市 Tainan City
16       H    68000 桃園市 Taoyuan City
17       K    10005 苗栗縣 Miaoli County
18       J    10004 新竹縣 Hsinchu County
19       I    10020 嘉義市 Chiayi City
20       Q    10010 嘉義縣 Chiayi County
21       E    64000 高雄市 Kaohsiung City
> class(twn)
[1] "SpatialPolygonsDataFrame"
attr(,"package")
[1] "sp"
>
```

配合 <https://data.gov.tw/dataset/7442>

rgdal 繪製台北市地圖

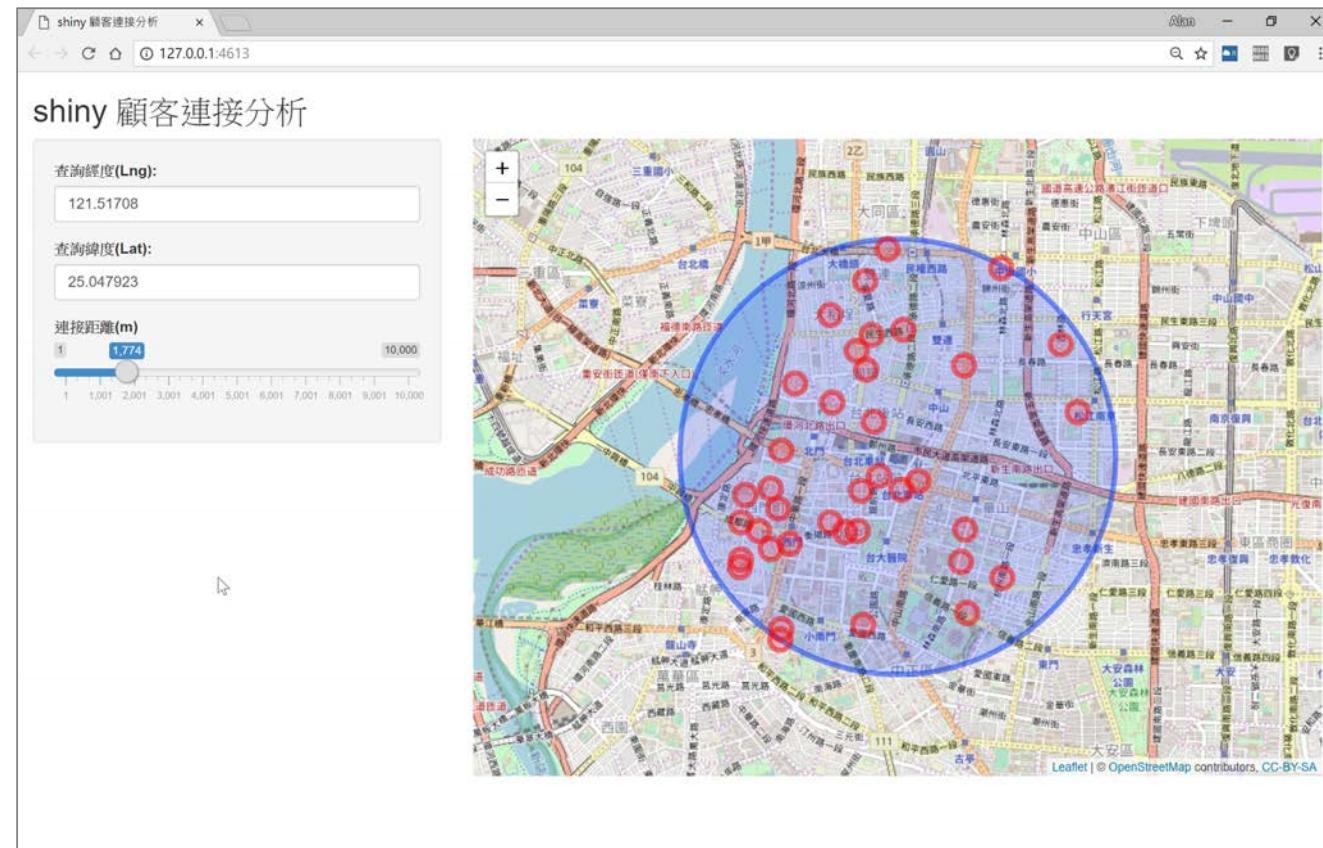


R demo

shiny 顧客連接分析

shiny 顧客連接分析

- <https://github.com/rwepa/DataDemo/tree/master/shinyCustomerConnect>



ui.R

```
library(shiny)
library(leaflet)

fluidPage(
  # Application title
  titlePanel("shiny 顧客連接分析"),
  # sidebar with a slider input for the number of distance
  sidebarLayout(
    sidebarPanel(
      ① numericInput("focusLng", label="查詢經度(Lng): ", value=121.517080),
      numericInput("focusLat", label="查詢緯度(Lat): ", value=25.047923),
      sliderInput("dists",
                 "連接距離(m)",
                 min = 1,
                 max = 10000,
                 value = 1000)
    ),
    ② # Show a plot of the generated distribution
    mainPanel(
      leafletOutput("distPlot", width="100%", height=600)
    )
  )
)
```

```

library(shiny)
library(geosphere) # distm 計算大圓距離
library(leaflet) # map
# x軸: 經度 longitude
# y軸: 緯度 latitude

# setwd("C:/rdata/shinyCustomerConnect")
customer <- read.csv("data/711_門市清單_台北市_lonlat.csv", header=TRUE, stringsAsFactors=FALSE) # 275*8

```

1.shiny, geosphere, leaflet 套件

server.R

```

function(input, output) {

  output$distPlot <- renderLeaflet({

    # 松山火車站 25.049398, 121.578248
    # 台北火車站 25.047923, 121.517080
    # distm(c(121.578248, 25.049398), c(121.517080, 25.047923), fun = distHaversine) # 公尺

    # input <- list(dists=1000, focusLng=121.517080, focusLat=25.047923)
    setradius <- input$dists # 半徑
    setlng <- input$focusLng # 經度
    setlat <- input$focusLat # 緯度

    distselect <- c()

    for (i in 1:nrow(customer)) {
      distselect[i] <- distm(c(setlng, setlat), c(customer$lon[i], customer$lat[i]))
    }
    customer$distm <- distselect
    customerSelect <- customer[customer$distm <= setradius, ]
  })
}

```

2.計算大圓距離

3.篩選資料

```

# plot(customerSelect$lon, customerSelect$lat, pch=16)
# points(121.517080, 25.047923, pch=4, col="red")
# text(customerSelect$lon + 0.0005, customerSelect$lat - 1:nrow(customerSelect), cex=1)
# symbols(121.517080, 25.047923, 100)

popname <- paste0(setradius, "m")
m1 <- leaflet(customer) %>%
  addTiles() %>%
  setView(lng = setlng, lat = setlat, zoom=14) %>%
  addCircles(lng = ~setlng, lat = ~setlat, radius = ~setradius, popup = paste0(setradius, "m")) %>%
  addCircleMarkers(data=customerSelect, lng= ~lon, lat= ~lat, popup=~店名, col="red")

m1
})

```

4.繪製地圖

如何擷取繪圖資料

<http://shiny.rstudio.com/articles/plot-interaction.html>

Interactive plots

LAST UPDATED: 30 MAY 2017

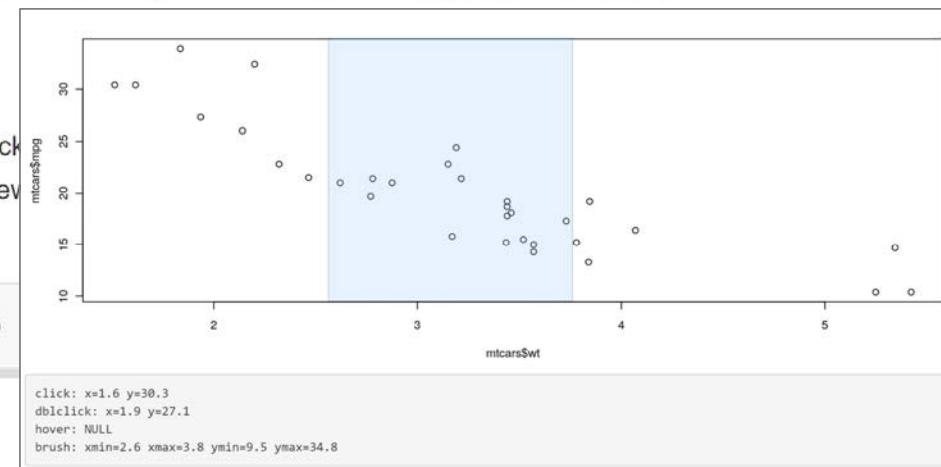
As of version 0.12.0, Shiny has built-in support for interacting with static plots generated by R's base graphics functions, and those generated by ggplot2.

This makes it easy to add features like selecting points and regions, as well as zooming in and out of images.

Basics

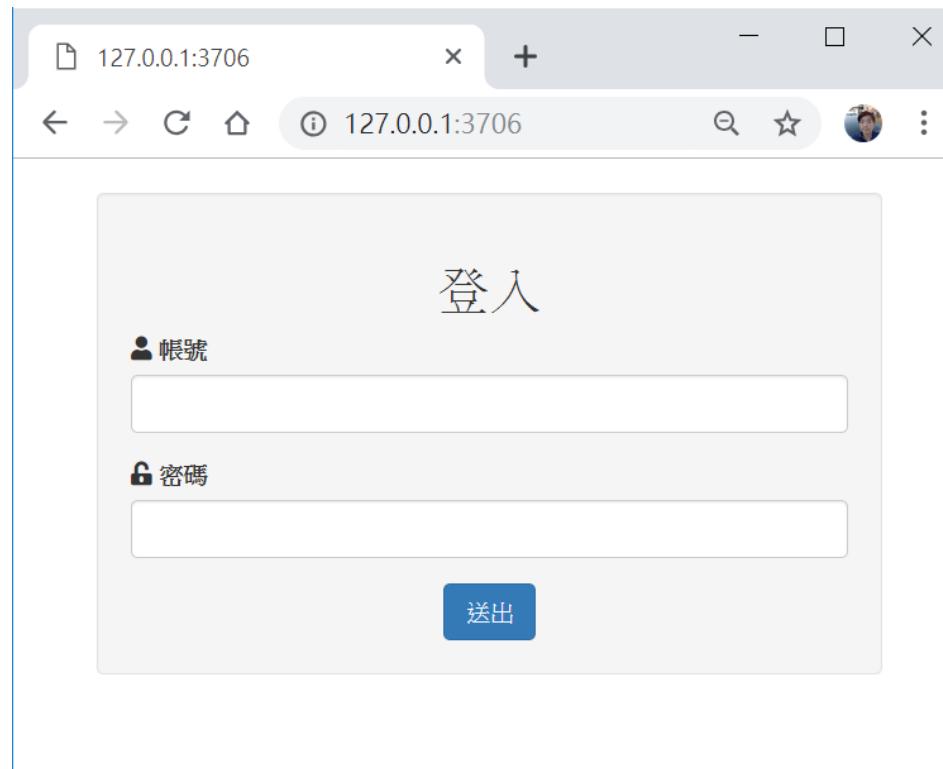
To get the position of the mouse when a plot is clicked or double-clicked, use `plotOutput()`. For example, this will define a new output object named `plot1`:

```
plotOutput("plot1",
  click = "plot_click",
  dblclick = "plot_dblclick",
  hover = "plot_hover",
  brush = "plot_brush"
),
```



shiny 認證 (Authentication)

shinyauthr 套件



Reference: <https://github.com/PaulC91/shinyauthr>

shiny 二階段選取



R ~ - Shiny
http://127.0.0.1:7314 | Open in Browser | C

選取資料集

mtcars

選取第1階段

欄位名稱

mpg

依照第1階段,自動顯示
第2階段的值

cyl
disp
hp
drat
wt
qsec
vs

R demo

shiny App 佈署

- 方法1: shinyapps.io免費版/付費版
- 方法2: shiny server 免費版(安裝於住家/公司) + Linux
 - <https://www.rstudio.com/products/shiny/shiny-server/>
- 方法3: shiny Server Pro 付費版

shiny server 免費版



Download RStudio Shiny Server v1.5.9.923



Ubuntu 14.04 or later

We currently only provide a pre-built binary for the 64-bit architecture. Running on other architectures will require [building from source](#).

Before installing Shiny Server, you'll need to install R and the Shiny package. To install the latest version of R you should first add the CRAN repository to your system as described here:

- [Ubuntu Packages for R](#)

You can then install R using the following command:

```
$ sudo apt-get install r-base
```

使用 www.shinyapps.io

- <https://www.shinyapps.io/>

The screenshot shows the shinyapps.io pricing page with five plans:

- FREE**: \$0/month. Includes 5 Applications, 25 Action Items, Community Support, and Shiny Broadcasting.
- STARTER**: \$9/month (or \$100/year). Includes 25 Applications, 100 Action Items, Premium Support, and Shiny Dev Center.
- BASIC**: \$39/month (or \$440/year). Includes Unlimited Applications, 500 Action Items, Performance Boost, Premium Support, and Shiny Dev Center.
- STANDARD**: \$99/month (or \$1,100/year). Includes Unlimited Applications, 2,000 Action Items, Authentication, Premium Support, and Shiny Dev Center.
- PROFESSIONAL**: \$299/month (or \$3,300/year). Includes Unlimited Applications, 10,000 Action Items, Authentication, Action Rating, Performance Boost, Custom Domains, and Premium Support.

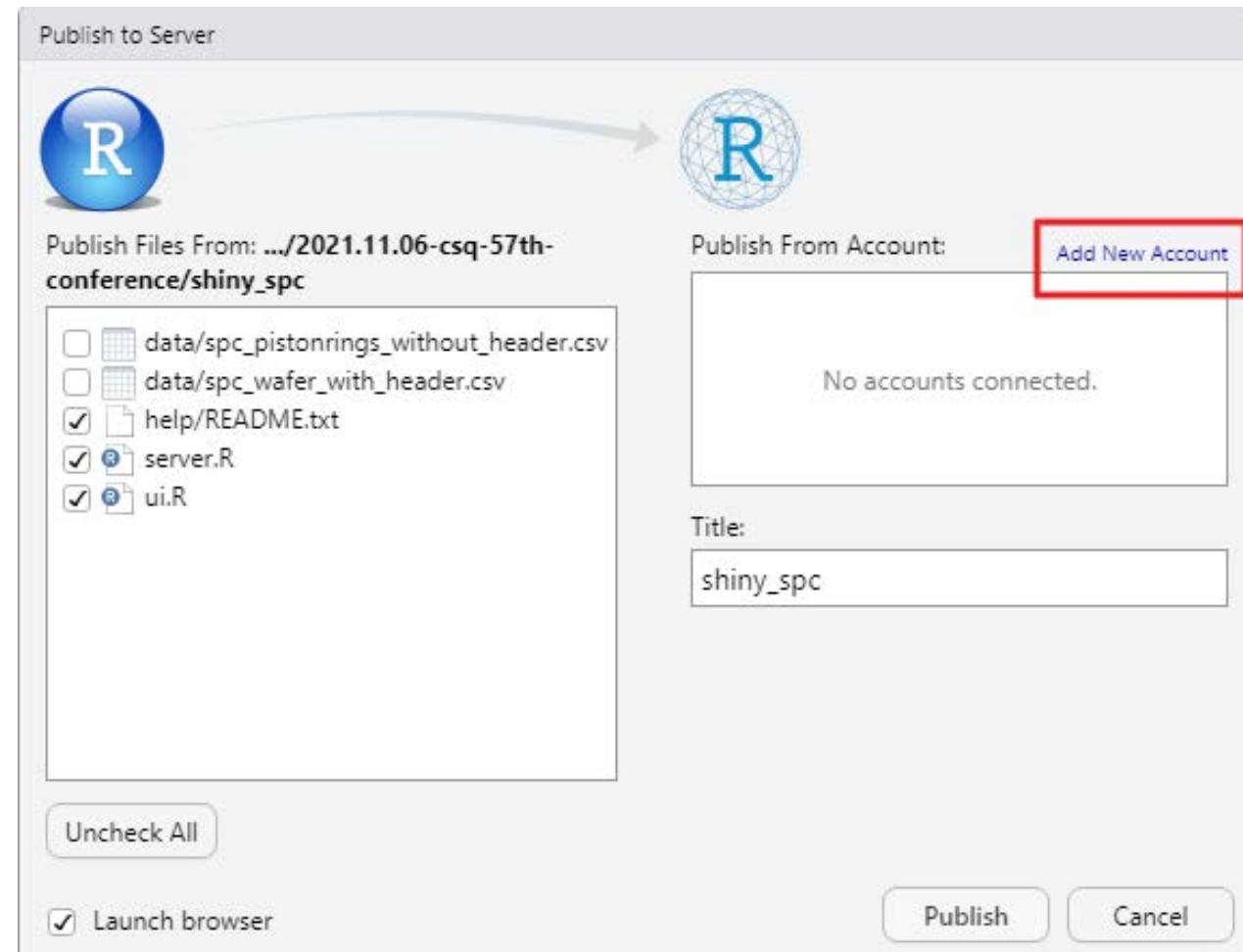
Each plan has a "Sign Up Now" button. Below the plans, there are sections for Documentation, Community Support, and Premium Support. A yellow speech bubble on the left says "免費".

線上說明

- rsconnect 套件
- devtools 套件

<https://docs.rstudio.com/shinyapps.io/getting-started.html>

發佈至 rwepa.shinyapp.io\shiny_spc



Connect Account

Connect Account

Connect Publishing Account

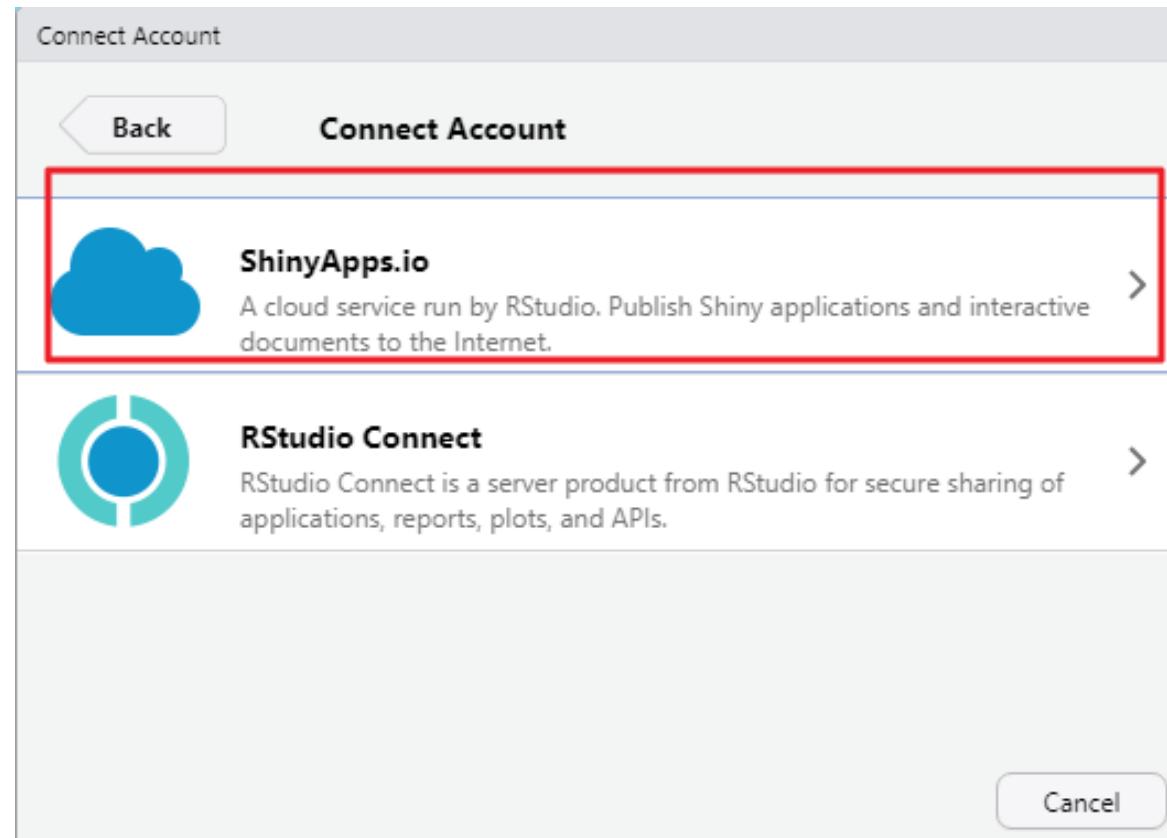


To publish content, you first need to connect RStudio to an account on the service you want to publish to.

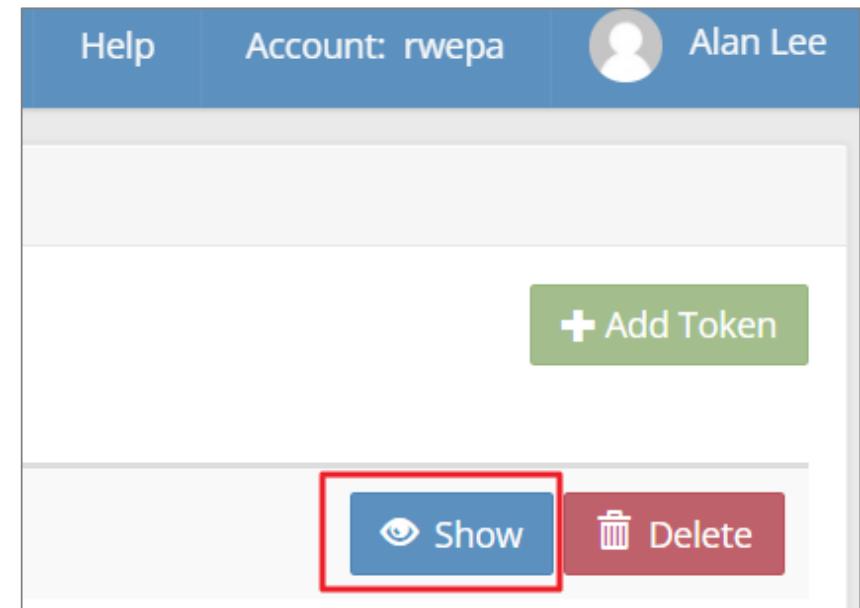
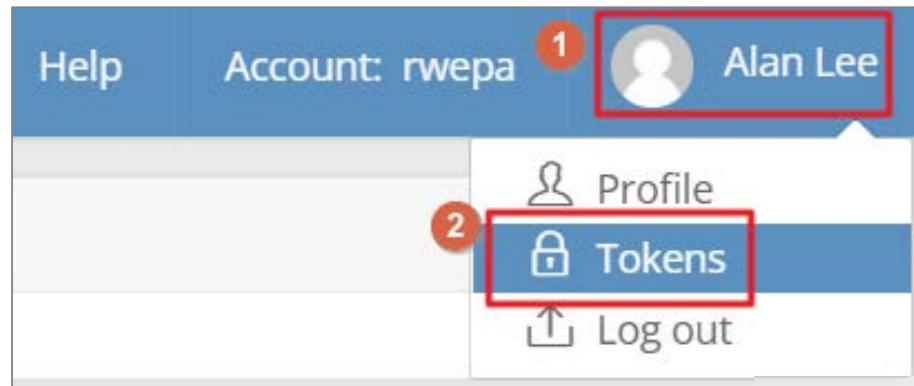
Once you've authorized this computer to publish content to an account, you can publish any time without re-entering your credentials.

Next **Cancel**

ShinyApps.io



複製 Tokens



按 CTRL + C

貼上 Tokens

Connect Account

Back Connect ShinyApps.io Account



Go to [your account on ShinyApps](#) and log in.
Click your name, then choose **Tokens** from your account menu.
Click **Show** on the token you want to use, then **Show Secret** and **Copy to Clipboard**. Paste the result here:



按 CTRL + V

Need a ShinyApps.io account? [Get started here.](#)

Connect Account Cancel

按 Publish

Publish to Server

已經新增帳號

Publish Files From: .../2021.11.06-csq-57th-conference/shiny_spc

data/spc_pistonrings_without_header.csv
 data/spc_wafer_with_header.csv
 help/README.txt
 server.R
 ui.R

Uncheck All

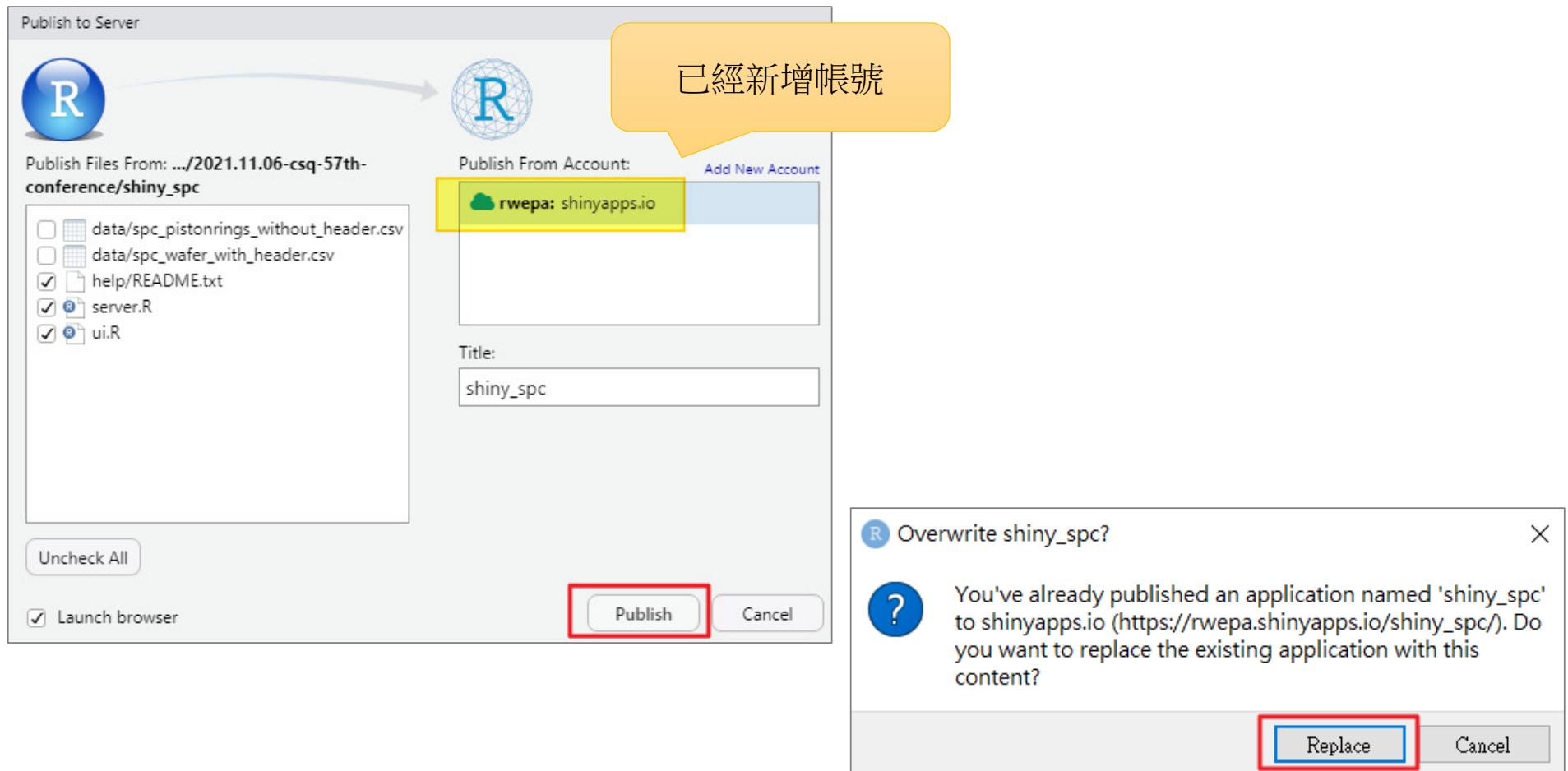
Launch browser

Publish From Account: **rwepa: shinyapps.io**

Title: shiny_spc

Overwrite shiny_spc?

You've already published an application named 'shiny_spc' to shinyapps.io (https://rwepa.shinyapps.io/shiny_spc/). Do you want to replace the existing application with this content?



Python vs. R

功能	Python {sklearn}	R
1. 繪圖	matplotlib	plot {graphics}
2. 迴歸	linear_model.LinearRegression	lm {stats}
3. 廣義線性模型(邏輯斯迴歸)	linear_model.LogisticRegression	glm {stats}
4. 決策樹	tree	rpart {rpart} ctree {party}
5. 隨機森林法	ensemble.RandomForestClassifier	randomForest {randomForest}
6. 關聯規則分析	mlxtend.frequent_patterns.apriori	apriori {arules}
7. 特徵選擇	feature_selection	train {caret}
8. 主成份分析 (非監督學習)	decomposition.PCA	prcomp {stats}
9. 集群 (非監督學習)	cluster	kmeans {stats}
10. 支持向量機 (SVM)	svm.SVC	svm {e1071}
11. 天真貝氏 (Naive Bayes)	naive_bayes.MultinomialNB	naiveBayes {e1071}
12. k近鄰法 (kNN)	neighbors.KNeighborsClassifier	knn {class}

彩蛋題



RFM Analysis

Recency, Frequency, and Monetary

- RFM是一種用於分析客戶價值的方法
- 常用於銷售資料庫,零售和服務分析之中,包括3項重要的指標:
 - 1.R 最近消費(Recency) - 天數
 - 2.F 消費頻率(Frequency) - 次數
 - 3.M 消費金額(Monetary) - 元數

Online Retail.xlsx

	A	B	C	D	E	F	G	H
1	InvoiceNo	StockCode	Description	Quantity	InvoiceDate	UnitPrice	CustomerID	Country
2	536365	85123A	WHITE HANGING HEART T-LIGHT HOLDER	6	2010/12/1 08:26	2.55	17850	United Kingdom
3	536365	71053	WHITE METAL LANTERN	6	2010/12/1 08:26	3.39	17850	United Kingdom
4	536365	84406B	CREAM CUPID HEARTS COAT HANGER	8	2010/12/1 08:26	2.75	17850	United Kingdom
5	536365	84029G	KNITTED UNION FLAG HOT WATER BOTTLE	6	2010/12/1 08:26	3.39	17850	United Kingdom
6	536365	84029E	RED WOOLLY HOTTIE WHITE HEART.	6	2010/12/1 08:26	3.39	17850	United Kingdom
7	536365	22752	SET 7 BABUSHKA NESTING BOXES	2	2010/12/1 08:26	7.65	17850	United Kingdom
8	536365	21730	GLASS STAR FROSTED T-LIGHT HOLDER	6	2010/12/1 08:26	4.25	17850	United Kingdom
9	536366	22633	HAND WARMER UNION JACK	6	2010/12/1 08:28	1.85	17850	United Kingdom
10	536366	22632	HAND WARMER RED POLKA DOT	6	2010/12/1 08:28	1.85	17850	United Kingdom
11	536367	84879	ASSORTED COLOUR BIRD ORNAMENT	32	2010/12/1 08:34	1.69	13047	United Kingdom

R 套件

模組	功能	
dplyr	A grammar of data manipulation	
data.table	Extension of data.frame	
ggplot2	Create Elegant Data Visualizations Using the Grammar of Graphics	
shiny	Web application framework	
caret	Classification and Regression Training	
mlr3	Provides R6 objects for efficient, object-oriented programming on the building blocks of machine learning	
Tensorflow	R Interface to 'TensorFlow'	
Bioconductor	Tools for the analysis genomic data	
SingleCellExperiment	Orchestrating Single-Cell Analysis with Bioconductor	

參考資料

- RWEPA
 - <http://rwepa.blogspot.com/>
- Python 程式設計-李明昌 <免費電子書>
 - <http://rwepa.blogspot.com/2020/02/pythonprogramming.html>
- R入門資料分析與視覺化應用教學(付費)
 - <https://mastertalks.tw/products/r?ref=MCLEE>
- R商業預測與應用(付費)
 - <https://mastertalks.tw/products/r-2?ref=MCLEE>

謝謝您的聆聽

Q & A



李明昌

alan9956@gmail.com

<http://rwepa.blogspot.tw/>