

程式的基本函數撰寫

- 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

<https://www.youtube.com/@alan9956>

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

alan9956@gmail.com

113/07/17 (三) 程式的基本函數撰寫

- 1.RWEPA簡介
- 2.資料分析暨視覺化應用
- 3.R與RStudio簡介(輔助說明,套件)
- 4.開放資料分析
- 5.函數範例
- 6.補充篇:Rcmdr

下載: https://github.com/rwepa/r_data_scientist/

1.RWEPA簡介

RWEPA簡介: <http://rwepa.blogspot.com/>

- 姓名 : 李明昌 (ALAN LEE)
- 現職 : 中華R軟體學會 理事
臺灣資料科學與商業應用協會 理事
- 學歷 : 中原大學 工業與系統工程所 博士
- 經歷 :
 - 淡江大學 兼任教師
 - 育達科技大學 兼任教師
 - 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
 - 育達科技大學 資訊管理系(所) 專任助理教授
 - 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、衛生福利部、縣市政府與日本名古屋產業大學等公民營單位演講達344場，3151小時。
- 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

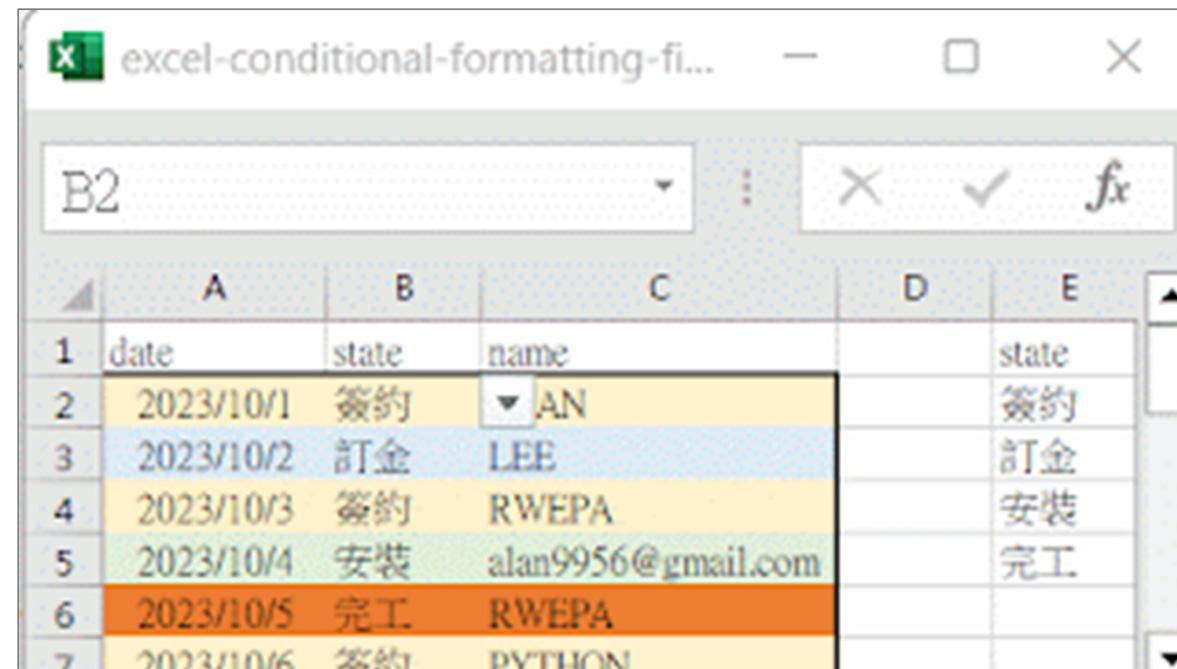
2. 資料分析暨視覺化應用

★★★資料分析架構→APC方法



Excel 下拉式選單與條件式格式設定教學

- ✿ YouTube (包括中文字幕) : <https://youtu.be/OVA4dvkrsBM>
- LINK: <https://rwepa.blogspot.com/2023/10/excel-drop-down-list-and-conditional-formatting.html>

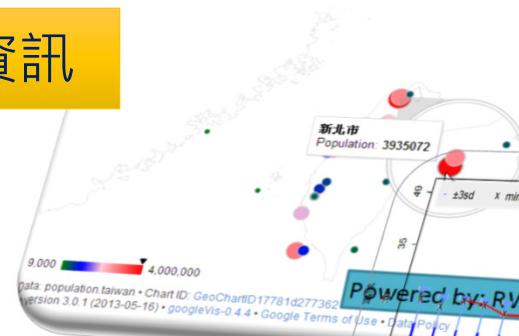


	A	B	C	D	E
1	date	state	name		state
2	2023/10/1	簽約	AN		簽約
3	2023/10/2	訂金	LEE		訂金
4	2023/10/3	簽約	RWEPA		安裝
5	2023/10/4	安裝	alan9956@gmail.com		完工
6	2023/10/5	完工	RWEPA		
7	2023/10/6	簽約	PYTHON		

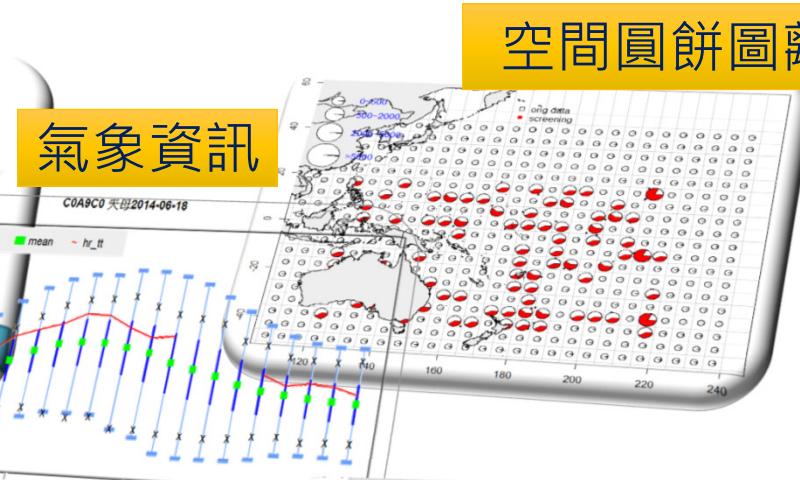
Excel 限制?

R+Shiny, Python+Streamlit 互動式平台

地理資訊



氣象資訊

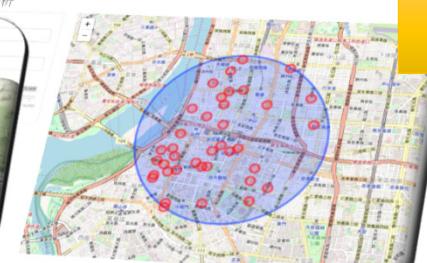


空間圓餅圖離群值分析

保險預測



顧客連結資訊



登山路線視覺化

中央氣象局 1,600萬筆資料(14,328檔案)

網頁呈現



客製化選單

R統計運算

保險預測模型

機率模型閥值調整

The screenshot shows the iInsurance interactive analysis platform version v.16.3.24. A yellow callout box labeled "機率模型閥值調整" points to a slider control labeled "機率模型閥值" with a value of 0.1. Another yellow callout box labeled "預測結果" points to a red-bordered section titled "檢視結果" which displays a table of 10 entries. The table includes columns for編號 (Entry ID), 性別 (Gender), 女性 (Female), 車輛種類 (Vehicle Type), 私家車 (Private Car), 曝露風險 (Exposure Risk), 曝露風險對數 (Exposure Risk Log), 無索償折扣 (No Claim Discount), 被保險人年齡 (Insured Person Age), 私家車一車齡 (Private Car Age 1), 私家車二車齡 (Private Car Age 2), 私家車三車齡 (Private Car Age 3), 私家車車齡組合 (Private Car Age Combination), 車齡 (Car Age), 預測機率 (Prediction Probability), and 理賠 (Claim). The last two columns are highlighted with red boxes. The table shows various entries with predicted probabilities ranging from 0.0694 to 0.1866, and claim outcomes marked as "有" (Yes) or "無" (No).

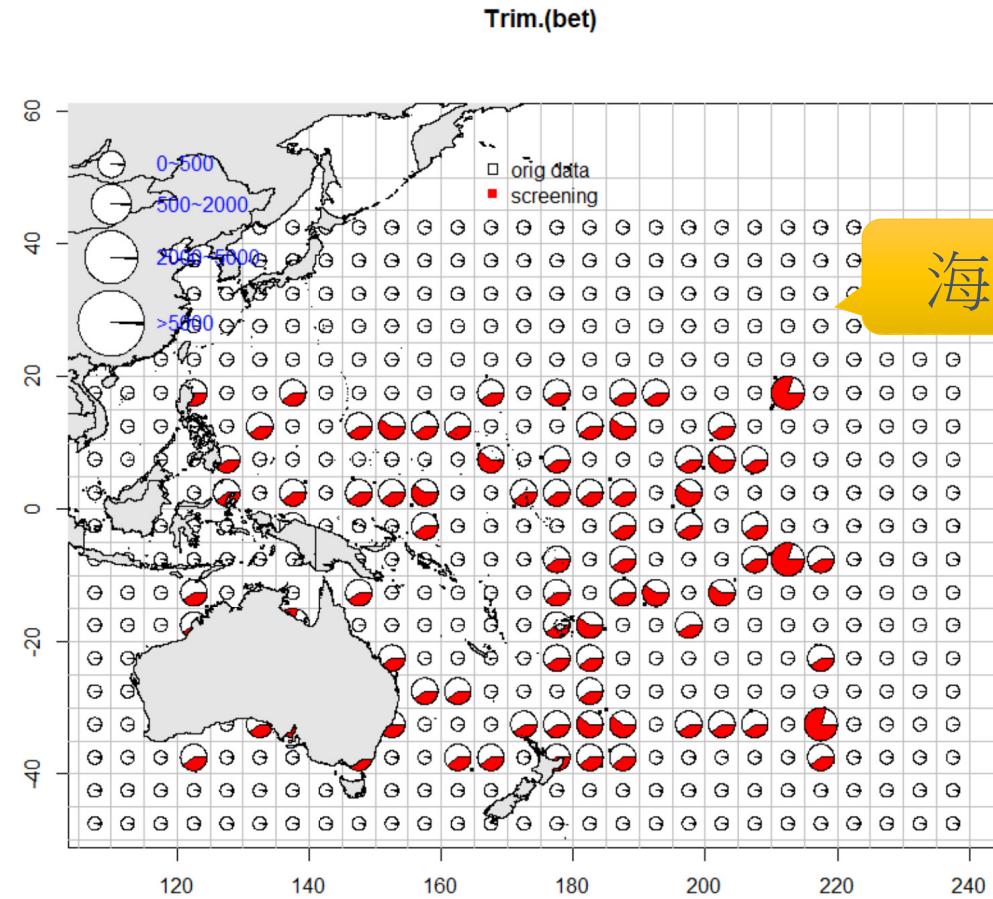
編號	性別	女性	車輛種類	私家車	曝露風險	曝露風險對數	無索償折扣	被保險人年齡	私家車一車齡 0	私家車二車齡 1	私家車三車齡 2	私家車車齡組合 0_1_2	車齡	預測機率	理賠	
1	M	0	A	1	0.9144422	-0.08944106	50	4	1	0	0	1	0	2	0.1069	有
2	M	0	A	1	0.8158795	-0.20348856	20	4	0	0	1	1	2	2	0.1441	有
3	M	0	A	1	0.8377823	-0.17699695	50	3	0	0	1	1	2	2	0.1866	有
4	M	0	A	1	0.4325804	-0.83798702	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

Showing 1 to 10 of 12 entries

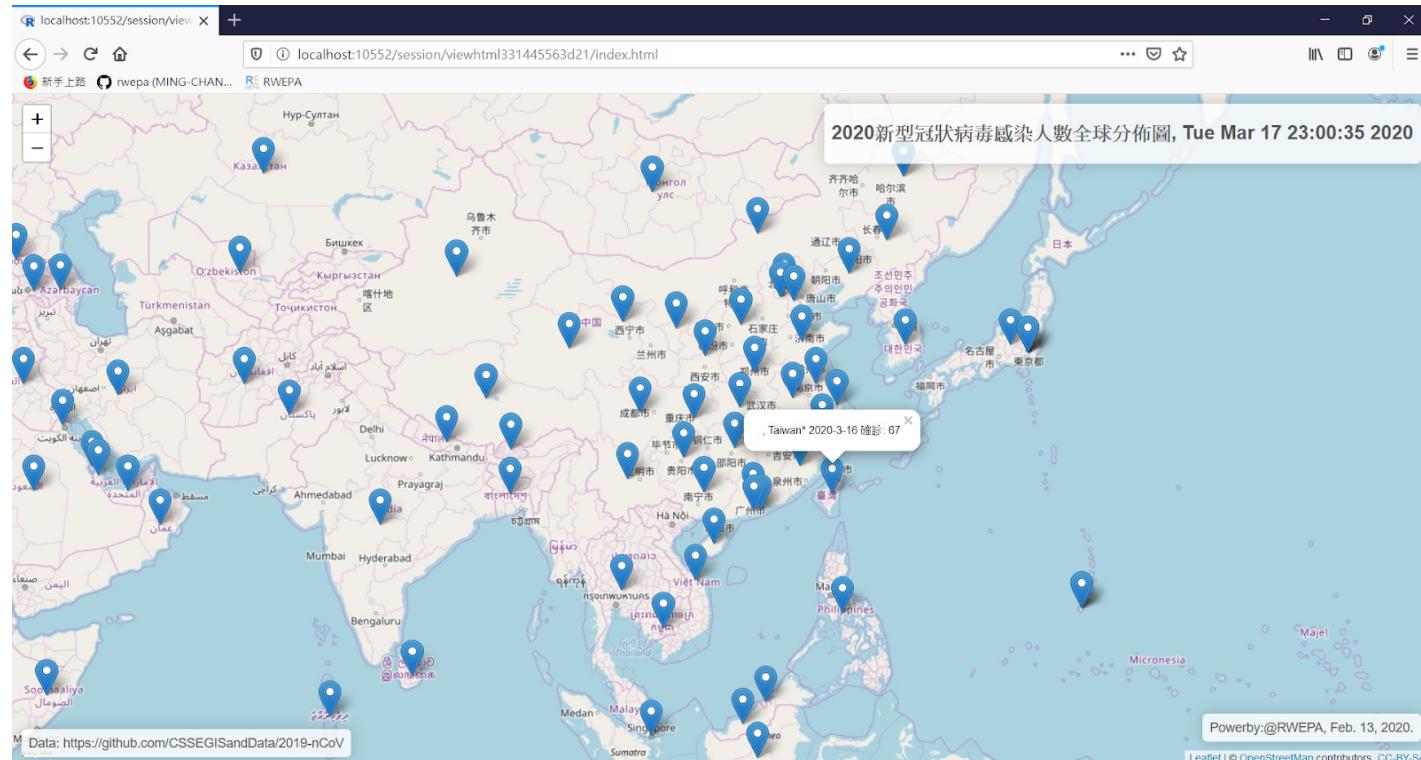
Previous 1 2 Next

127.0.0.1:6177/#tab-9487-2

空間圓餅圖離群值分析



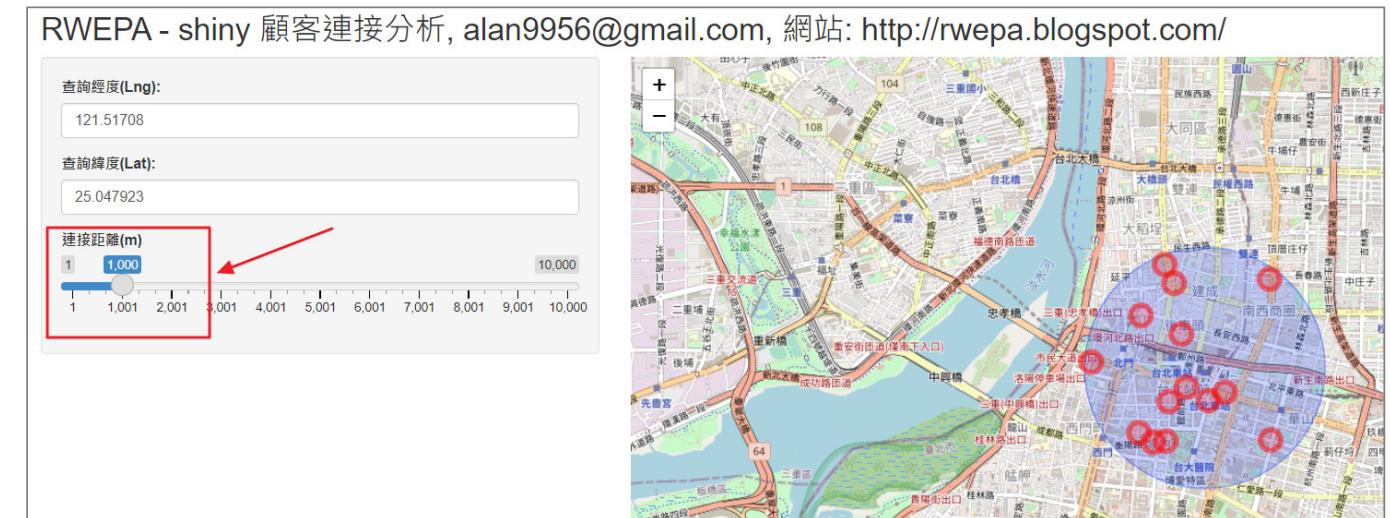
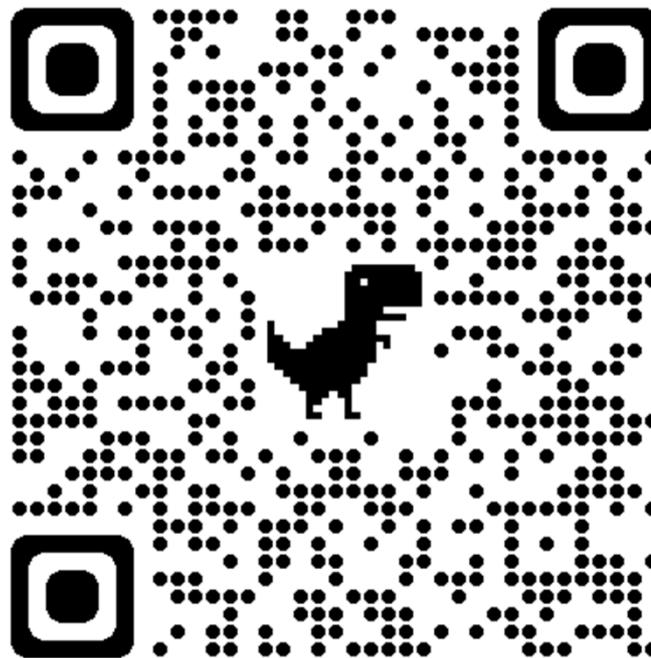
2020新型冠狀病毒視覺化



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

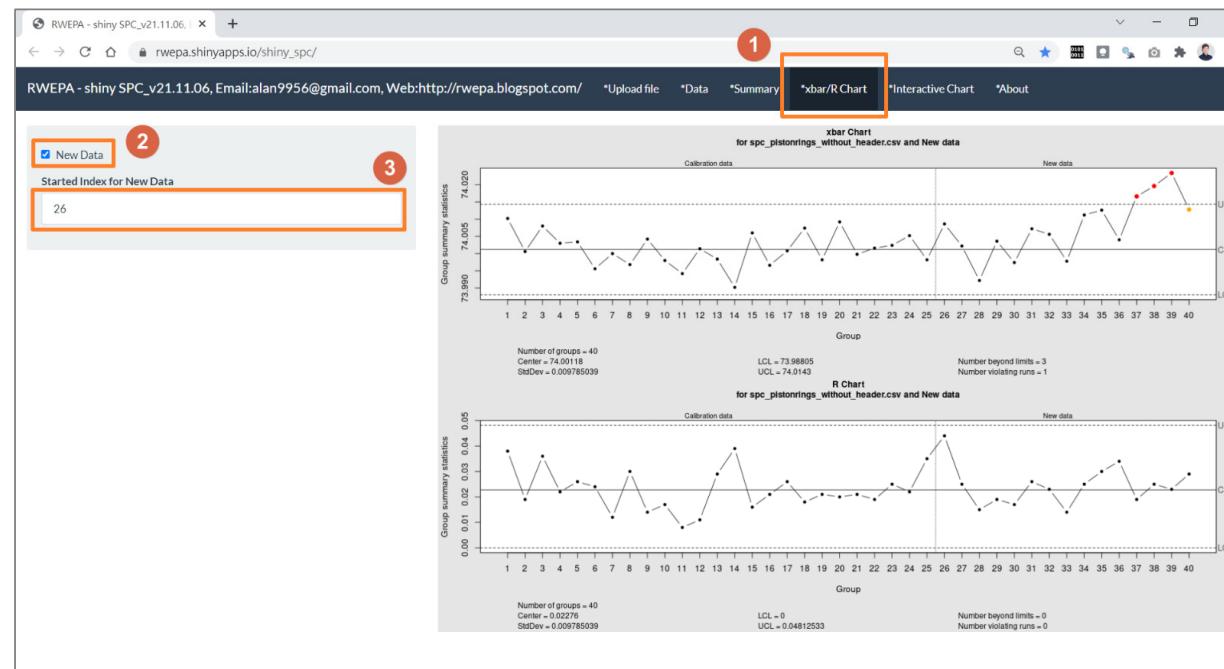
shiny 顧客連接分析

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



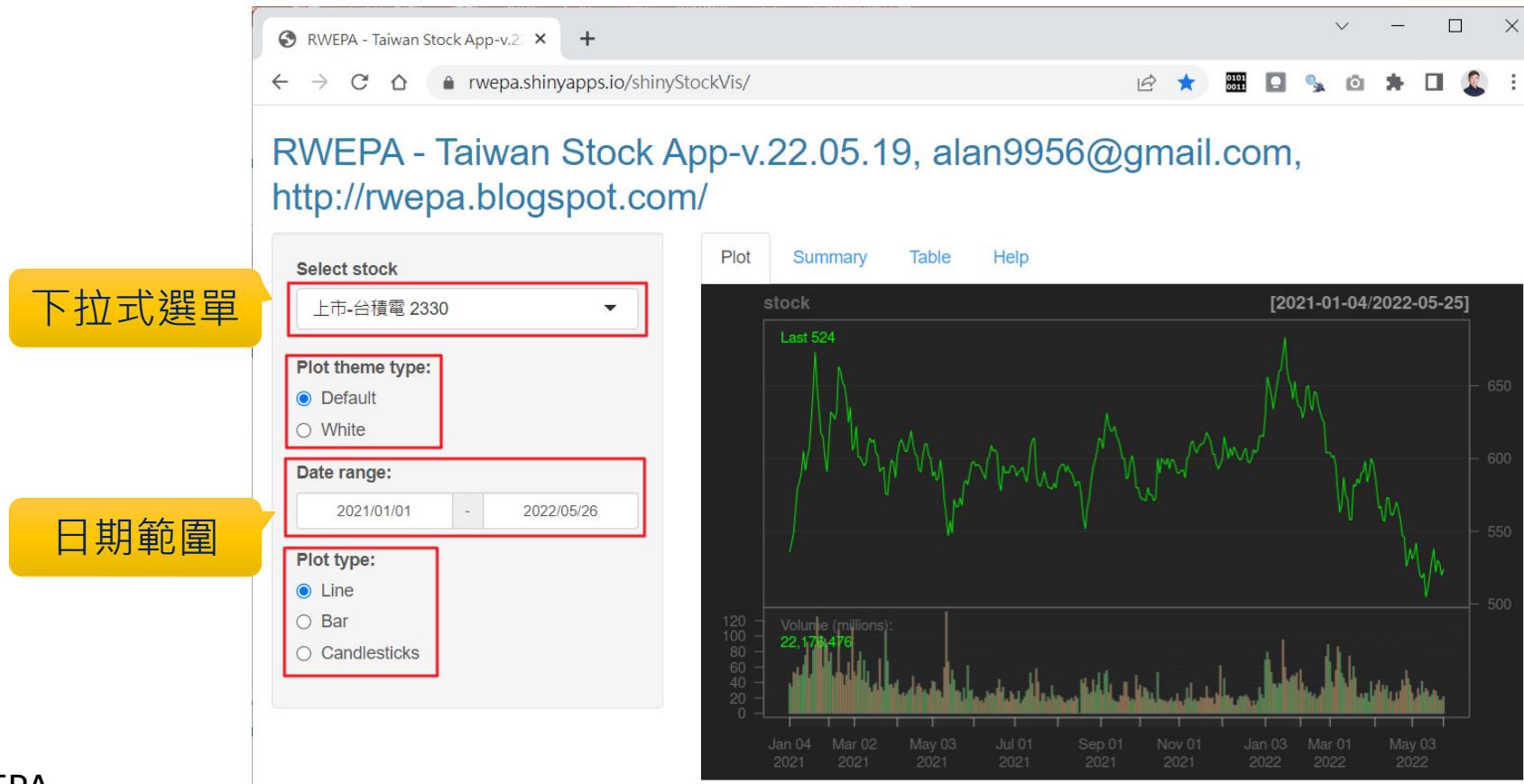
品質管制圖(quality control chart)應用

- 說明: <http://rwepa.blogspot.com/2021/10/r-shiny-quality-control-chart.html>
- 資料1: https://github.com/rwepa/shiny_spc/blob/main/data/spc_wafer_with_header.csv
- 資料2: https://github.com/rwepa/shiny_spc/blob/main/data/spc_pistonrings_without_header.csv
- 線上示範: https://rwepa.shinyapps.io/shiny_spc/



Taiwan Stock App

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

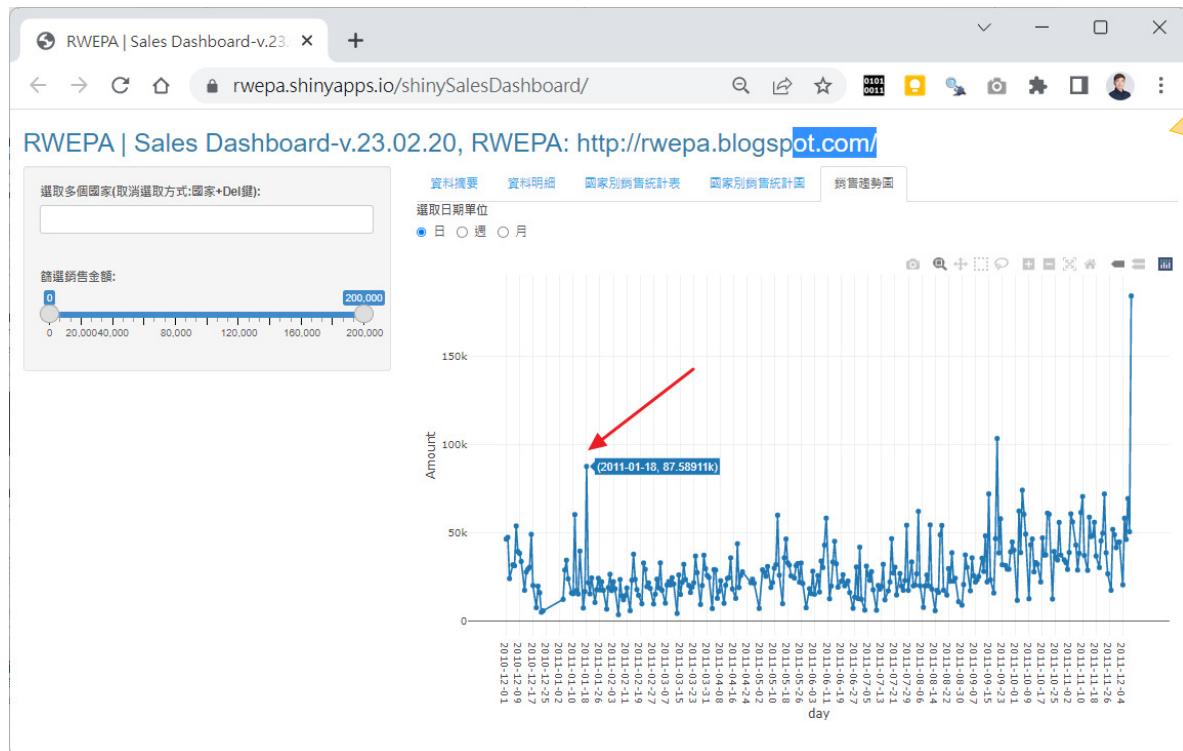


離子資料分析與視覺化應用



RWEPA | shiny企業實務應用 第4集-shiny銷售儀表板

- Shiny: <https://rwepa.shinyapps.io/shinySalesDashboard/>
- YouTube: <https://youtu.be/4GgZlf8heQk>



謝謝 ^_ ^

訂閱 + 讚 + 開啟小鈴鐺

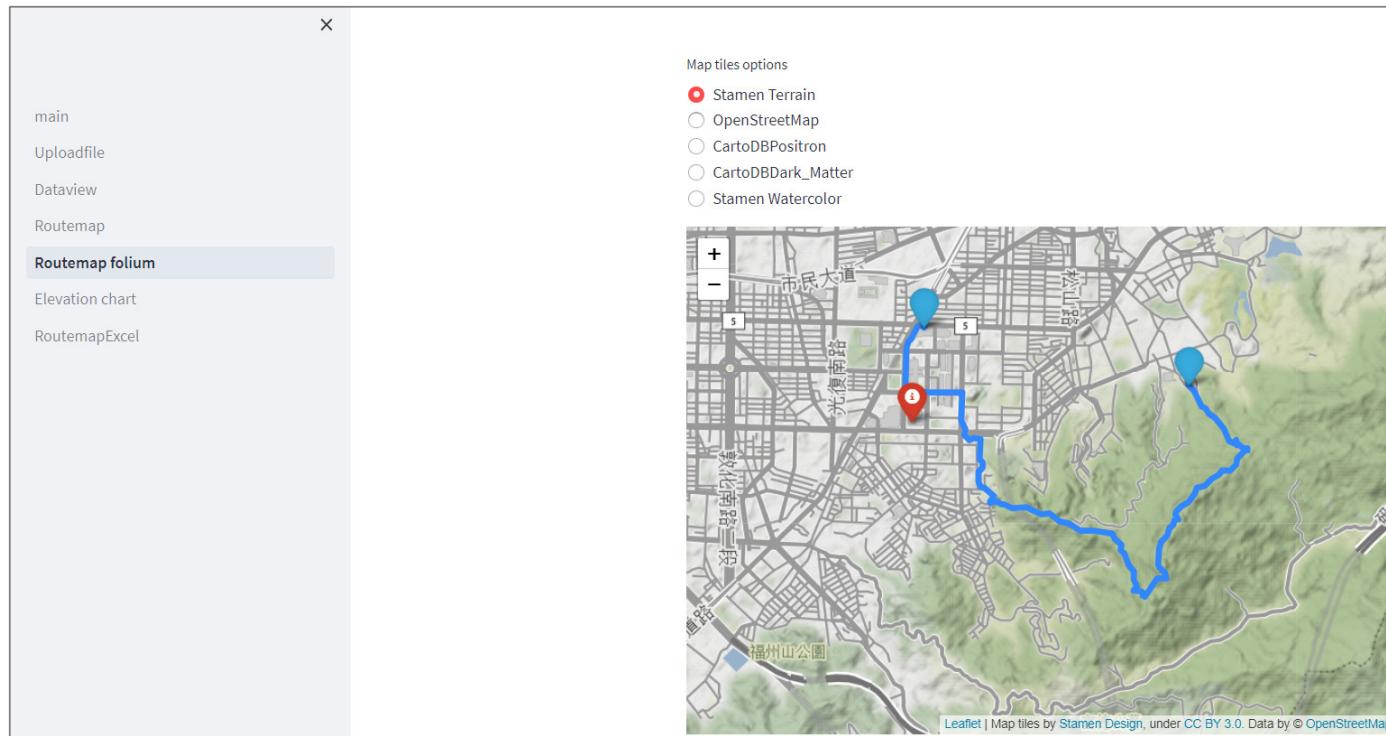
shiny企業實務應用 第6集-小明算命師(下) - 第1季完結篇

- Ubuntu Shiny Server: <https://shiny.rwepa.net/shiny-hr-teller/>
- YouTube: <https://youtu.be/rrD6KV3eV-w>



登山路線視覺化分析平台 (Python + Streamlit)

- YouTube : https://youtu.be/-_zghs2qrlg
- 系統展示 <https://rwepa-climb.streamlit.app/>



Power BI - 客戶輪廓分析

PowerBI_customer_profile_analysis - Power BI Desktop

客戶輪廓Overview

客戶數 666 平均年齡 34.8

• FEMALE • MALE

年齡組距

職業

職業	百分比
其他未分類	21%
製造/營造業	17%
資訊科技業	13%
金融相關業	11%
餐飲服務/休閒業	11%
批發零售業	10%
軍/警/公/教	9%
醫療相關業	5%
運輸倉儲業	2%
專業人員	1%
服務業	0%
宗教、慈善團體	0%
政府官員/公務員	0%
教育業	0%

縣市別

縣市	百分比
新北市	29%
台北市	23%
桃園市	11%
台中市	8%
高雄市	8%
臺南市	5%
彰化縣	4%
基隆市	3%
新竹市	2%
新竹縣	2%
苗栗縣	1%
南投縣	1%
花蓮縣	1%
雲林縣	1%

區域別

區域	百分比
新店區	4%
內湖區	4%
中和區	3%
松山區	3%
大安區	3%
三重區	3%
中山區	3%
板橋區	3%
新莊區	3%
桃園區	3%
東區	3%
信義區	3%
中壢區	2%
中正區	2%

資料

視覺效果

DAX量值管理表
分行資料表
日期對照表
刷卡交易資料表
刷卡國別對照表
刷卡類別對照表
客戶資料表
理財產品代碼表
理財產品交易資料表
業務員資料表
Dynamic_Calendar

值
於此處新增資料欄位

鑽研
跨報表
保留所有篩選
在此處新增鑽研欄位

Power BI – RFM分析

- 🌸 YouTube : <https://youtu.be/Lkr9HmzLTtg>
- LINK: <https://rwepa.blogspot.com/2023/07/rwepa-rfm-analysis-using-power-bi.html>

Customer Segmentation Using RFM Analysis, 2023



最近消費 (recency) :
顧客上次消費時間愈近，用戶價值愈大。

消費頻率 (frequency) :
顧客在一段時間中，總購買次數，購買頻率愈高，用戶價值愈大。

消費金額 (monetary) :
顧客總消費金額，消費金額愈高，用戶價值愈大。

Author : Ming-Chang Lee
YouTube : <https://www.youtube.com/@alan9956>
RWEPA : <http://rwepa.blogspot.tw/>
GitHub : <https://github.com/rwepa>
Email : alan9956@gmail.com

RFM分析 X RFM標準化分析 RECENTY FREQUENCY Monetary +

Tableau - Superstore

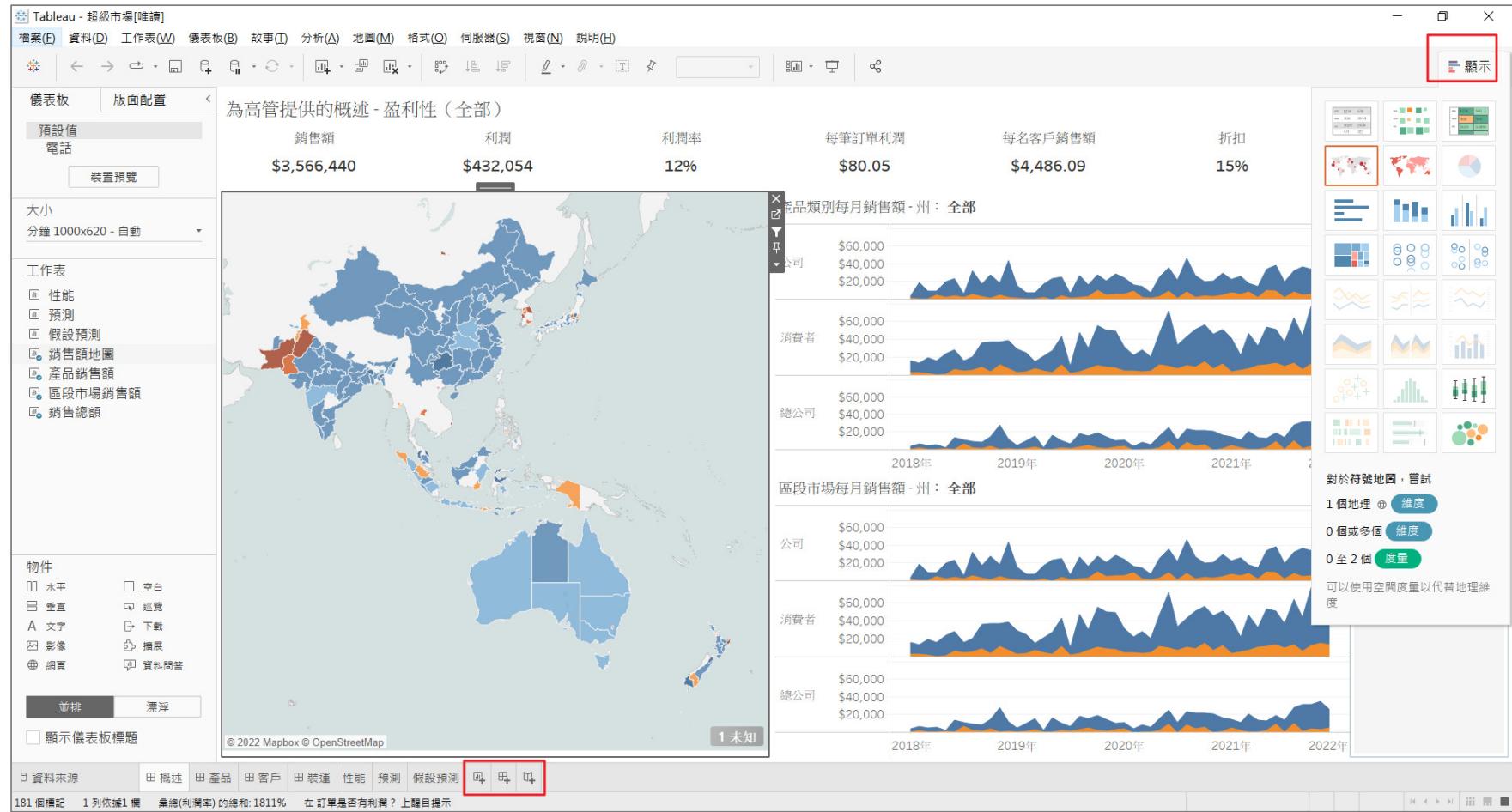
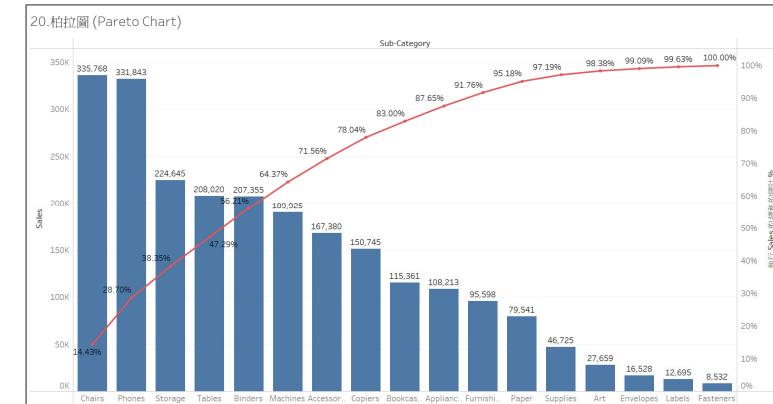
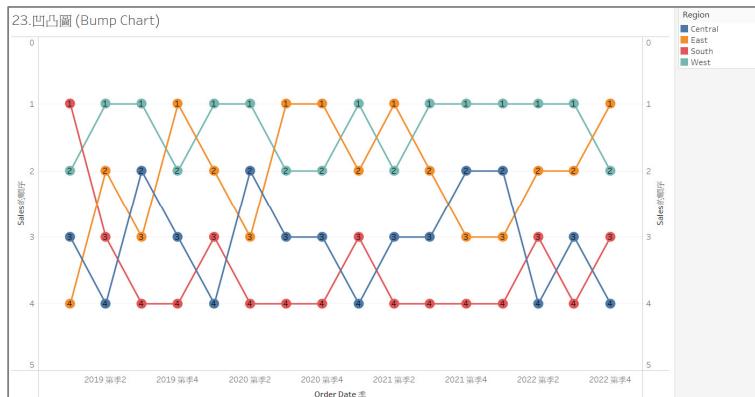
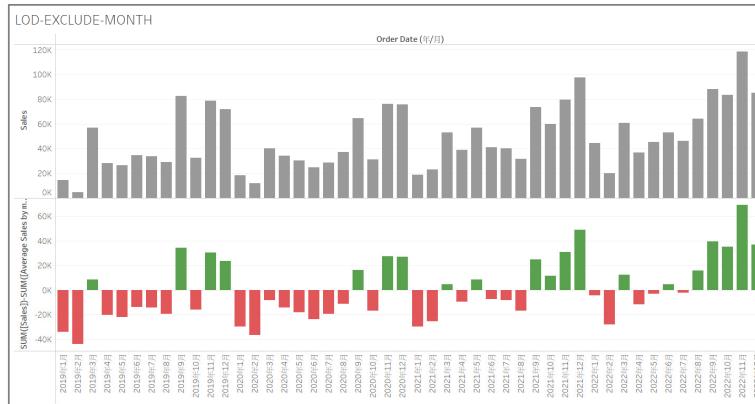


Tableau - 智慧製造應用

- <https://github.com/rwepa/Talks>
- <https://public.tableau.com/app/profile/ming.chang.lee/vizzes>



A Market Basket Analysis table showing the frequency of items purchased together. The columns represent different categories like Accessories, Appliances, Art, Binders, etc., and the rows show the count of each item.

Sub-Categ..	Access..	Applian..	Art	Binders	Bookcas..	Chairs	Copiers	Envole..	Fasten..	Furnishin..	Machin..	Paper	Phones	Storage	Supplies	Tables		
Accessories	718	60	89	161	29	64	7	32	27	115	47	22	153	118	103	25	44	
Appliances	60	459	63	135	13	49	9	21	23	82	21	9	110	78	56	24	32	
Art	89	63	756	159	159	34	87	7	29	30	105	47	12	152	124	100	29	41
Binders	161	135	159	1,339	56	126	19	54	62	200	82	30	276	199	201	43	74	
Bookcas..	29	13	34	56	228	21	2	11	7	30	21	6	48	42	37	10	9	
Chairs	64	49	87	126	21	591	10	29	28	107	36	19	133	91	85	19	36	
Copiers	6	9	7	19	2	10	70	4	5	12	6	2	20	11	14	6	4	
Envole..	32	21	29	54	11	29	4	251	11	32	11	6	59	41	38	5	12	
Fasteners	27	23	30	62	7	28	5	11	226	48	15	8	59	39	39	13	14	
Furnishin..	115	82	105	200	30	107	12	32	48	919	57	29	181	154	140	31	46	
Labels	47	21	47	82	21	36	6	11	15	57	348	8	80	59	61	18	17	
Machin..	22	9	12	30	6	19	2	6	8	29	8	114	28	22	21	1	6	
Paper	153	110	152	276	48	133	20	59	59	181	80	28	1,205	179	178	45	54	
Phones	116	78	124	199	42	91	11	41	39	154	58	22	179	826	117	22	54	
Storage	103	56	100	201	37	95	14	38	39	140	61	21	178	117	797	33	50	
Supplies	25	24	29	43	10	19	6	5	13	31	18	1	45	22	33	189	13	
Tables	44	32	41	74	9	36	4	12	14	46	17	6	54	54	50	13	314	

Tableau 教學

- Tableau資料分析與視覺化工具實作教師工作坊(初階)
 - https://github.com/rwepa/Talks/blob/main/tableau_tutorial_basic.pdf
- Tableau資料分析與視覺化工具實作教師工作坊(進階)
 - https://github.com/rwepa/Talks/blob/main/tableau_tutorial_advanced.pdf
- Tableau與R語言實務應用
 - https://github.com/rwepa/Talks/blob/main/tableau_r.pdf
- Tableau與MySQL資料庫實務應用
 - https://github.com/rwepa/Talks/blob/main/tableau_mysql.pdf



Python 程式設計-李明昌 免費電子書

- <http://rwepa.blogspot.com/2020/02/pythonprogramminglee.html>

主題: Python 程式設計-李明昌 - ipynb

檔名: Python_Programming_Lee_ipynb.zip

包括 Python 程式設計-李明昌電子書的原始 ipynb 檔案, 圖檔, 部分資料集

下載: https://github.com/rwepa/DataDemo/blob/master/Python_Programming_Lee_ipynb.zip



Python_Programming_Lee_ipynb.zip > python.book.lee >	
名稱	類型
.ipynb_checkpoints	檔案資料夾
data	檔案資料夾
img	檔案資料夾
Python程式設計-李明昌.ipynb	IPYNB 檔案

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 +中文字幕



- **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>



- **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

課程提供教學範例的原始程式檔案與資料集 +中文字幕

如何學習？

- 熟悉教材內容
- 將教材內容的資料集改為工作資料集(企業, 學術)
- 遇到問題時, 想辦法**尋找答案**
- 掌握 APC方法
- 掌握 ①摘要 ②繪圖 ③建模
- 參考網路應用文章 (進階) & 學術論文

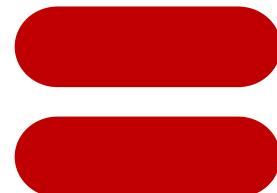
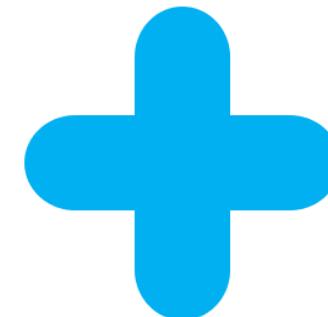
```
> 尋找答案 <- list(方法1 = c("同事", "同學", "朋友等"),
+                         方法2 = "Google",
+                         方法3 = "alan9956@gmail.com")
> print(尋找答案)
$方法1
[1] "同事"    "同學"    "朋友等"

$方法2
[1] "Google"

$方法3
[1] "alan9956@gmail.com"
```



學習目標



3.R與RStudio簡介(輔助說明,套件)

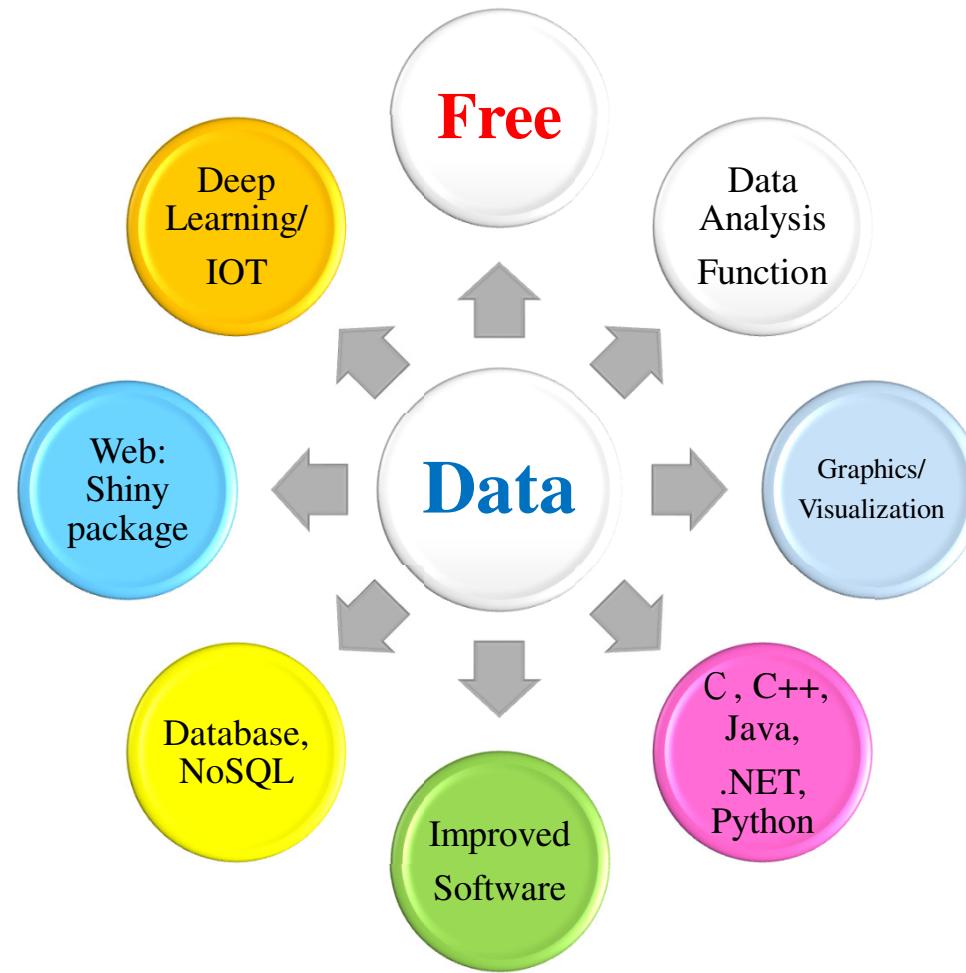
R 簡介與安裝

認識 R

- 1976 - 貝爾實驗室 John Chambers, Rick Becker, and Allan Wilks 將 S 語言研發為 R 。
- 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
 - 2013年4月 - R 3.0.0
 - 2024年6月 - R 4.4.1



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: <https://cran.csie.ntu.edu.tw/>
- 選取 Download R for Windows

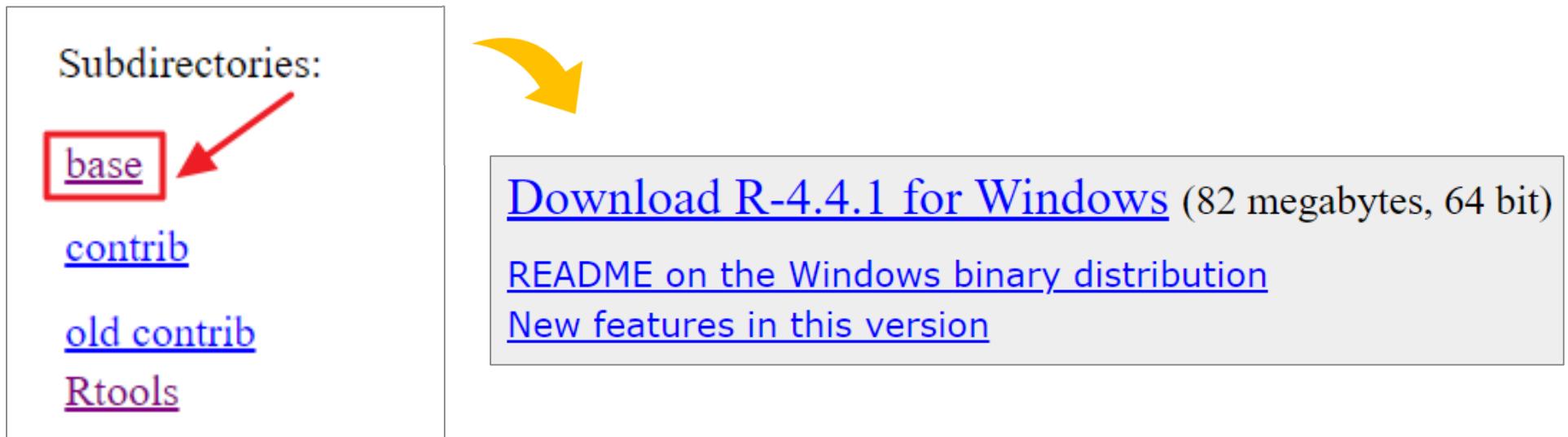


- [Download R for Linux](#) ([Debian](#), [Fedora/Redhat](#), [Ubuntu](#))
- [Download R for macOS](#)
- [Download R for Windows](#)



R-下載 (續)

- 選取 base → 下載 [R-4.4.1-win.exe]



- R安裝路徑: 保留原路徑,不要修改
- 安裝參考說明, 2006
https://github.com/rwepa/DataDemo/blob/master/windows_intall_R.pdf

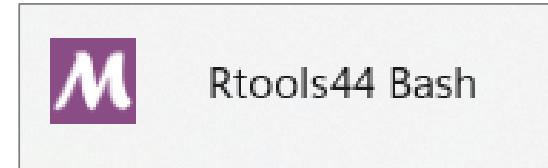
Rtools 下載與安裝

- Rtools for Windows: 保留預設安裝路徑 C:\rtoolsXX
- <https://cran.csie.ntu.edu.tw/bin/windows/Rtools/>

Subdirectories:	<u>RTools 4.4</u> 	for R versions from 4.4.0 (R-release and R-devel)
	<u>RTools 4.3</u>	for R versions 4.3.x (R-oldrelease)
	<u>RTools 4.2</u>	for R versions 4.2.x
	<u>RTools 4.0</u>	for R from version 4.0.0 to 4.1.3
	<u>old versions of RTools</u>	for R versions prior to 4.0.0

[base](#)
[contrib](#)
[old contrib](#)
[Rtools](#) 

- 安裝完成：程式集 \ Rtoolsxx Bash



Rtools44 Bash

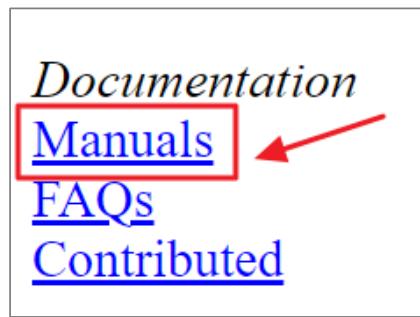
✓ 安裝 R

✓ 安裝 Rtools



R Manuals (使用手冊)

- <https://cran.csie.ntu.edu.tw/manuals.html>



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 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 used by 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)		

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

contributed documentation
(貢獻文件, 免費啦)

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參考文獻

```
> citation()
```

To cite R in publications use:

R Core Team (2024). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. <<https://www.R-project.org/>>.

LaTeX 的使用者的 BibTeX 條目是

正確引用

R, ~~https://www.R-project.org/~~, 2004

錯誤引用

實作
練習

R 執行畫面

The screenshot shows the RGui (64-bit) interface. The title bar says "RGui (64-bit)". The menu bar includes "檔案", "命令歷程", "重設大小", and "視窗". The toolbar has icons for file operations. The "R Console" window displays R version 3.0.3 (2014-03-06) and copyright information. The "R Graphics: Device 2 (ACTIVE)" window shows a plot of 100 random uniform numbers (runif(100)). A yellow callout bubble points to the plot area.

```
plot(runif(100), type="l", main= "R大數據分析")
```

```
demo(graphics)
```

```
demo(persp)
```

大小寫
須一致

參考文獻:
citation()

R 功能表

檔案



編輯



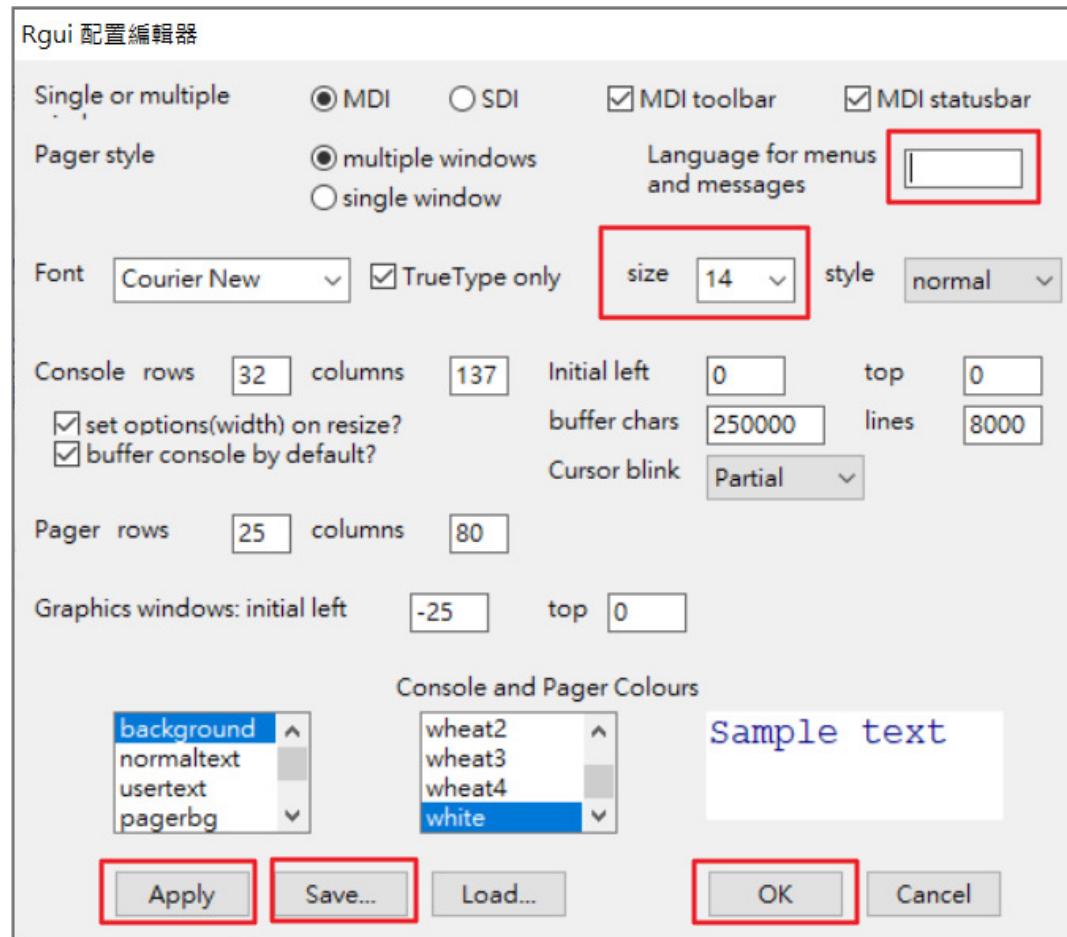
現行目錄 getwd()

儲存控制台-文字檔

輔助



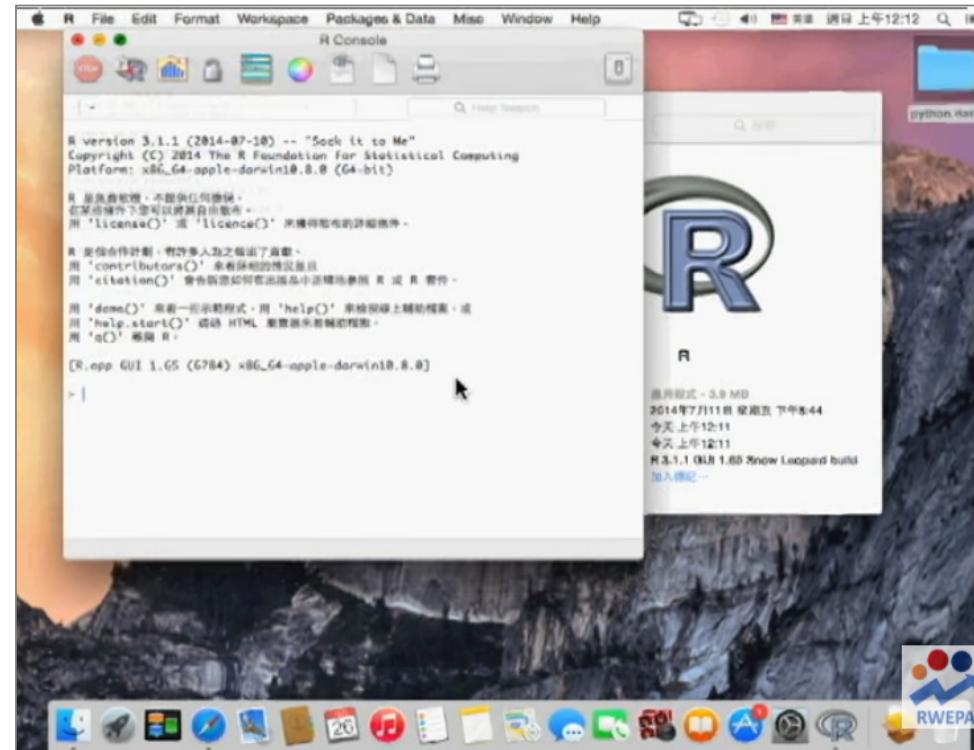
編輯 \ GUI 偏好設定



- Language: en 英文
- size: 字型大小

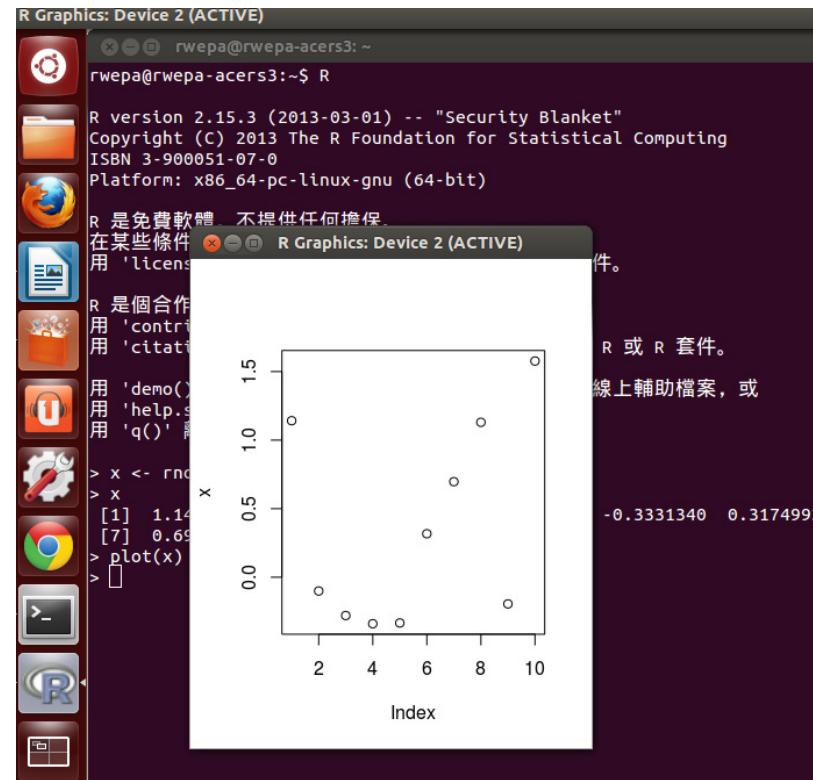
R for Mac

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



R for Ubuntu

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



實作
練習

新增R檔案練習

The screenshot shows the RGui interface. On the left is the R Editor window containing R code to generate a scatter plot matrix for the Iris dataset. In the center is the R Console window showing the command history and the generated scatter plot matrix titled "RWEPA-iris資料集散佈圖矩陣". On the right is the R Graphics window displaying the scatter plot matrix. A yellow callout bubble highlights the steps for running the code.

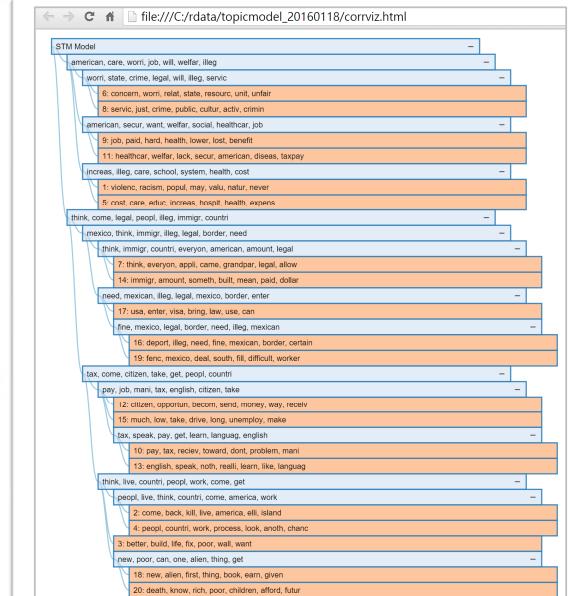
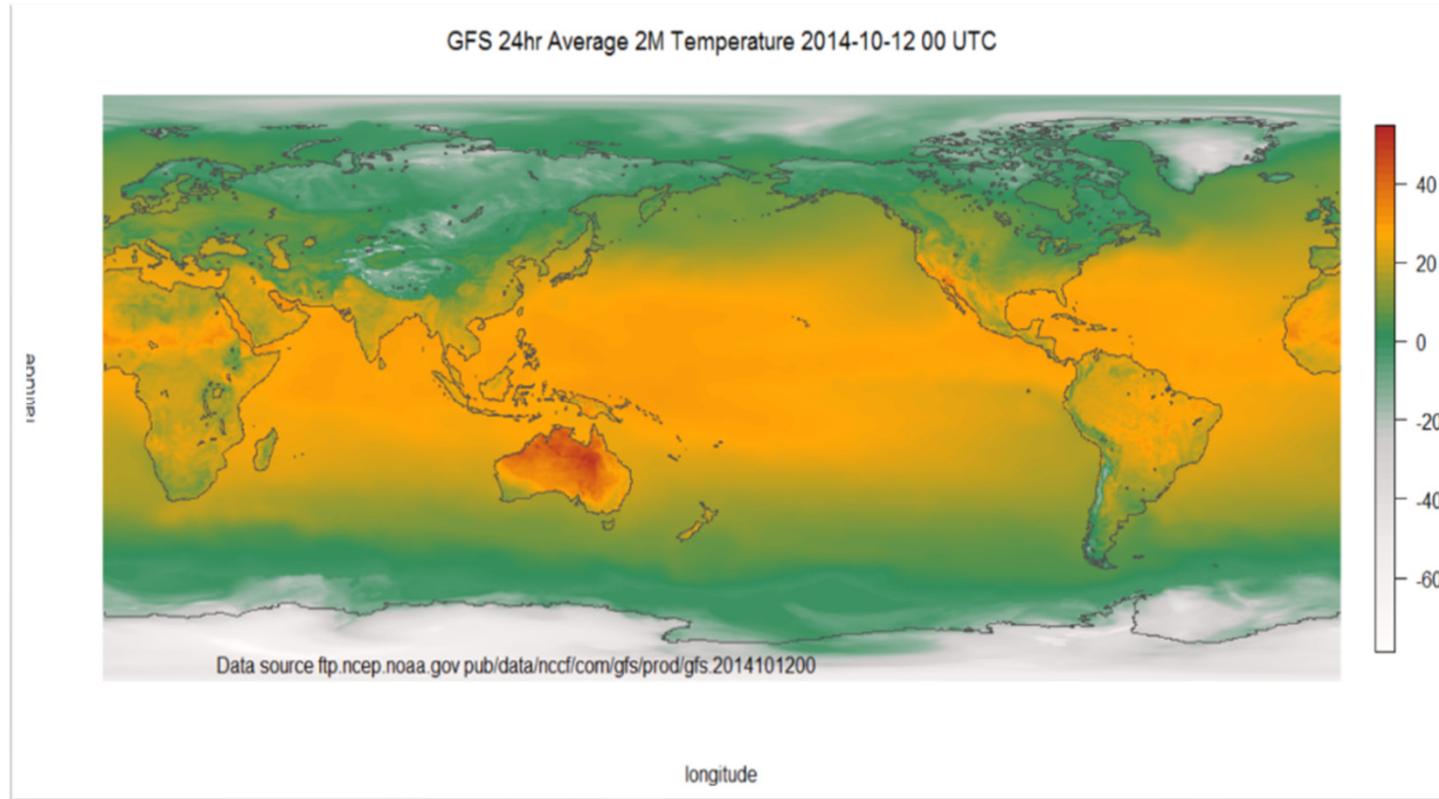
- 步驟1: 選取程式碼
- 步驟2: 按 Run line or selection 或 $\text{Ctrl} + \text{R}$
- 步驟3: 按 File \ Save \ MyFirstProject.R

```
plot(runif(10), type="b", main= "R大數據分析")
x <- rnorm(10)
x
pairs(iris[-5],
      pch=16,
      col=iris$Species,
      main="RWEPA-iris資料集散佈圖矩陣")
# end
```

```
[1] > pairs(iris[-5],
+         pch=16,
+         col=iris$Species,
+         main="RWEPA-iris資料集散佈圖矩陣")
+ 
```

RStudio 簡介與安裝

整合式開發環境 - RStudio



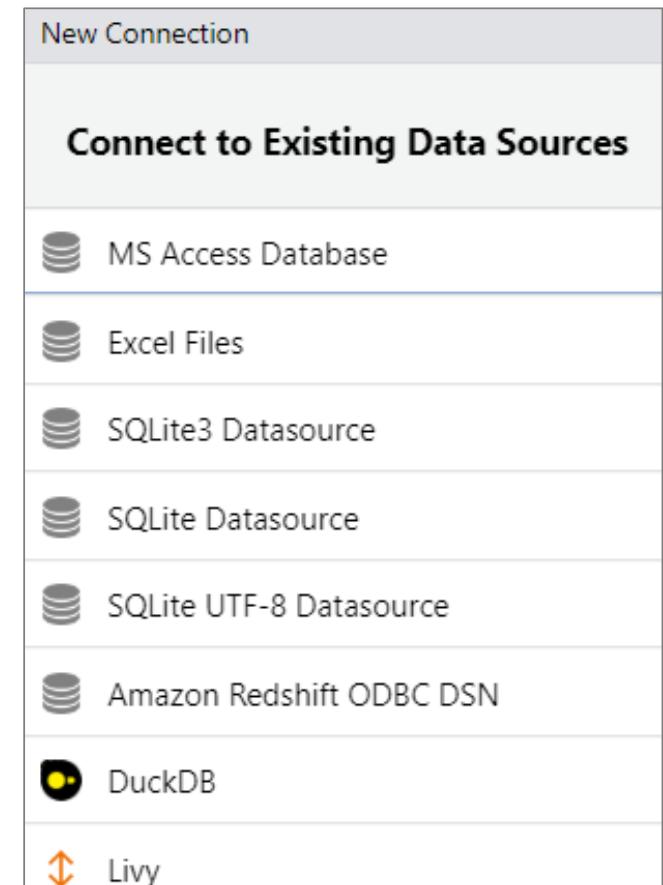
主題模型

視覺化應用

(全球2M氣溫圖)

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
- 整合連接資料庫: SQL, Spark
- 整合R套件: shiny, rmarkdown, Quarto
- 支援 RStudio外掛程式 (Addins)
- 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/>~~



<https://posit.co/>

The screenshot shows the posit.co homepage. At the top, there is a navigation bar with links: PRODUCTS (highlighted with a red box), SOLUTIONS, LEARN & SUPPORT, EXPLORE MORE, PRICING, and a search icon. To the right of the search icon is a blue button labeled "DOWNLOAD RSTUDIO" with a red box around it. A red arrow points from the crossed-out URL in the list above to this button. Below the navigation bar, the main content features a large, bold heading "Deployment made easy". Underneath the heading, there is a paragraph of text: "Deploy all of your work, including Shiny, Streamlit, and Dash applications. Models. Quarto documents. Jupyter Notebooks. Reports. Dashboards. Even APIs. With customizable access controls and authentication options that make IT happy." To the right of the text, there is a photograph of a person working at a computer, with the computer screen displaying a dashboard of data visualizations related to penguins.

RStudio 下載 (續)

PRODUCTS
Explore our open source, cloud, and enterprise products

Open Source > Enterprise > Cloud >

RStudio IDE	Posit Team	Posit Cloud
RStudio Server	Posit Workbench	Connect Cloud
Shiny	Posit Connect	Public Package Manager
Shiny Server	Posit Package Manager	shinyapps.io
R Packages	Posit Academy	Amazon SageMaker
Quarto		MS Azure

單機版

伺服器

伺服器版本



RStudio-2024.04.2-764.exe下載

- <https://posit.co/download/rstudio-desktop/>

2: Install RStudio

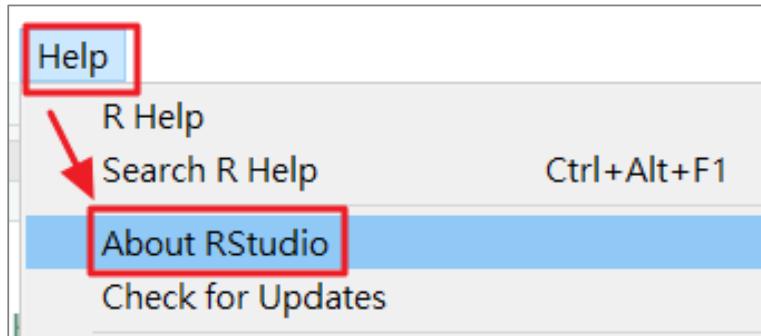
[DOWNLOAD RSTUDIO DESKTOP FOR WINDOWS](#)

Size: 262.79 MB | [SHA-256: 09E1E38A](#) | Version: 2024.04.2+764 |
Released: 2024-06-10

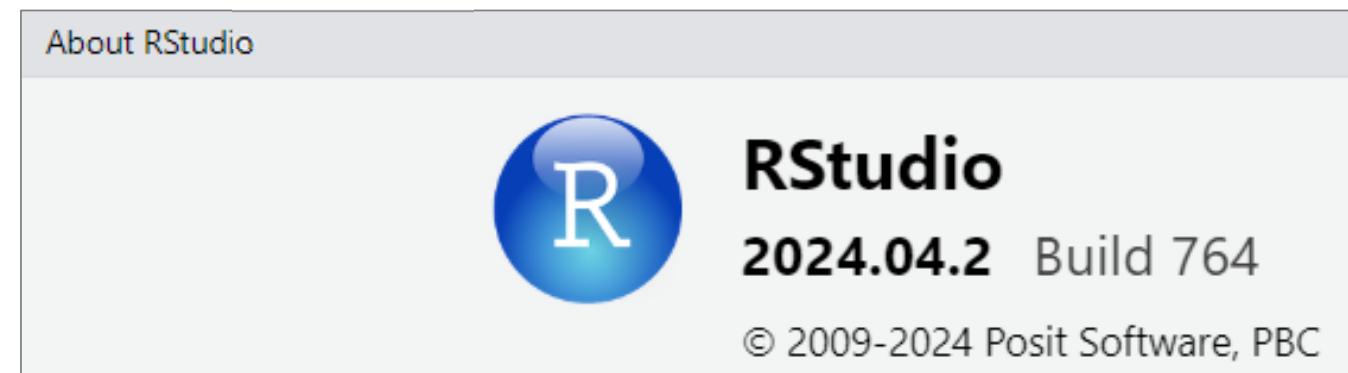
RStudio 安裝



RStudio 版本訊息

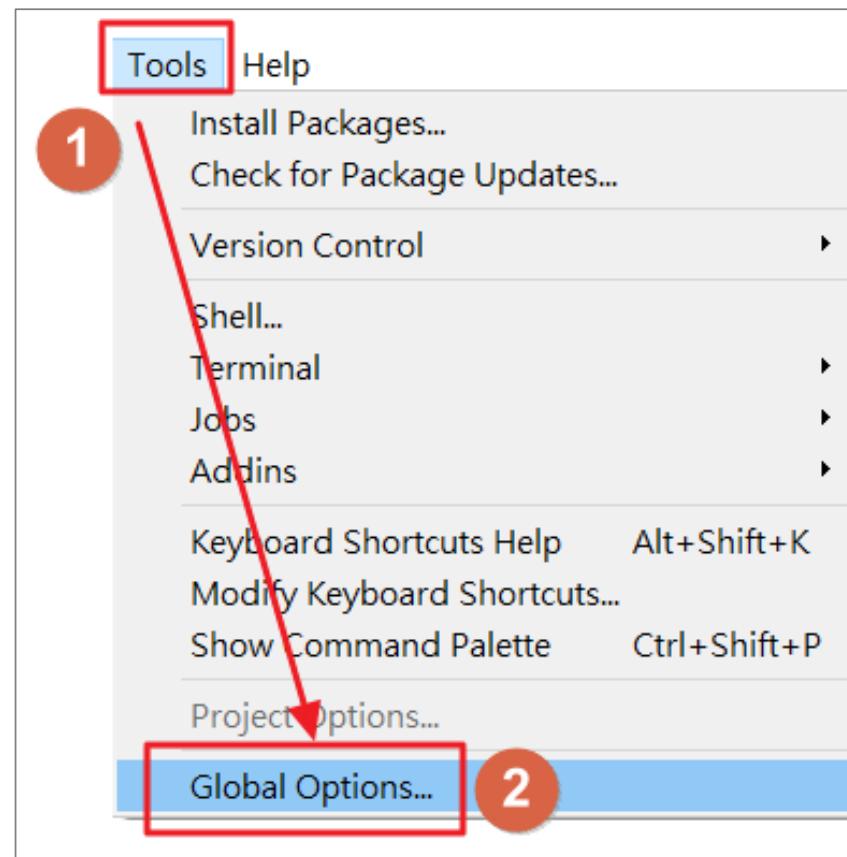


Help \ About RStudio

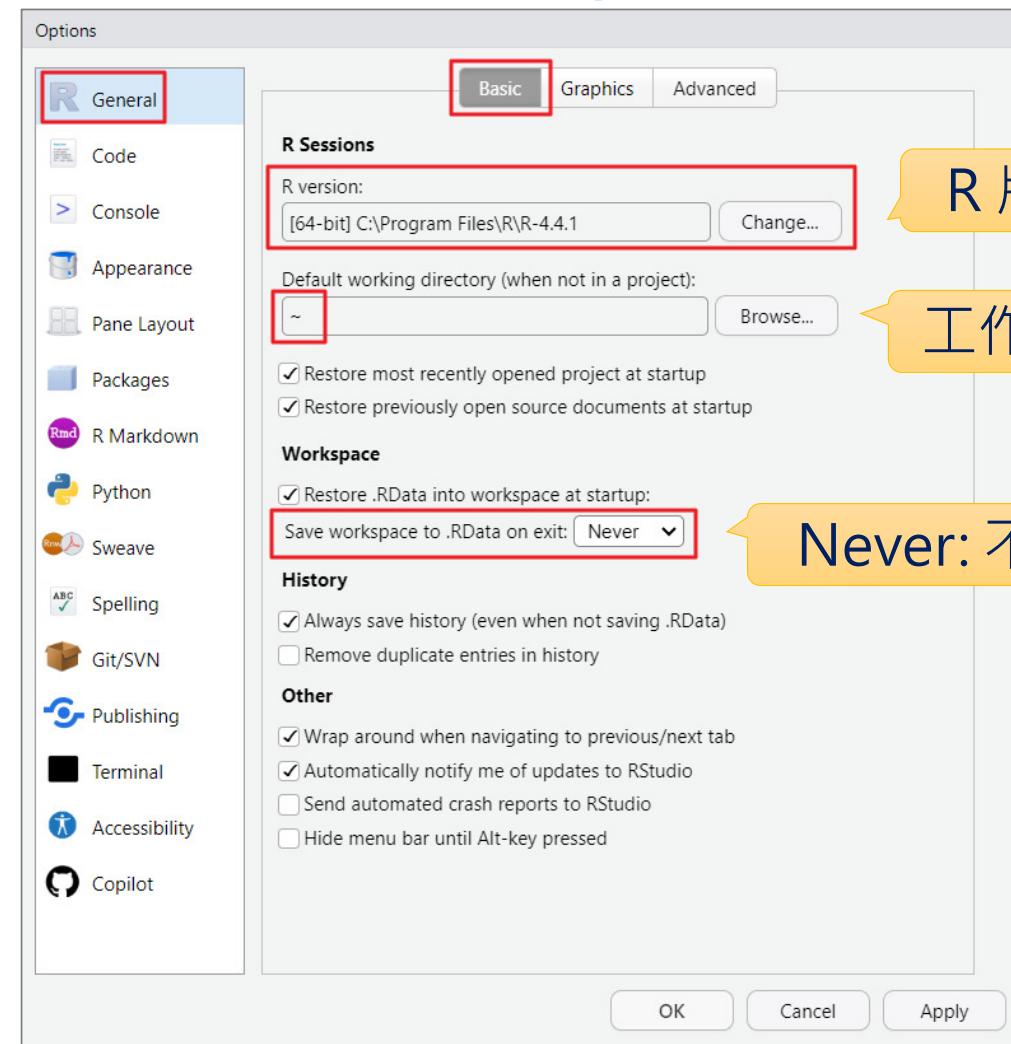


RStudio-選項設定

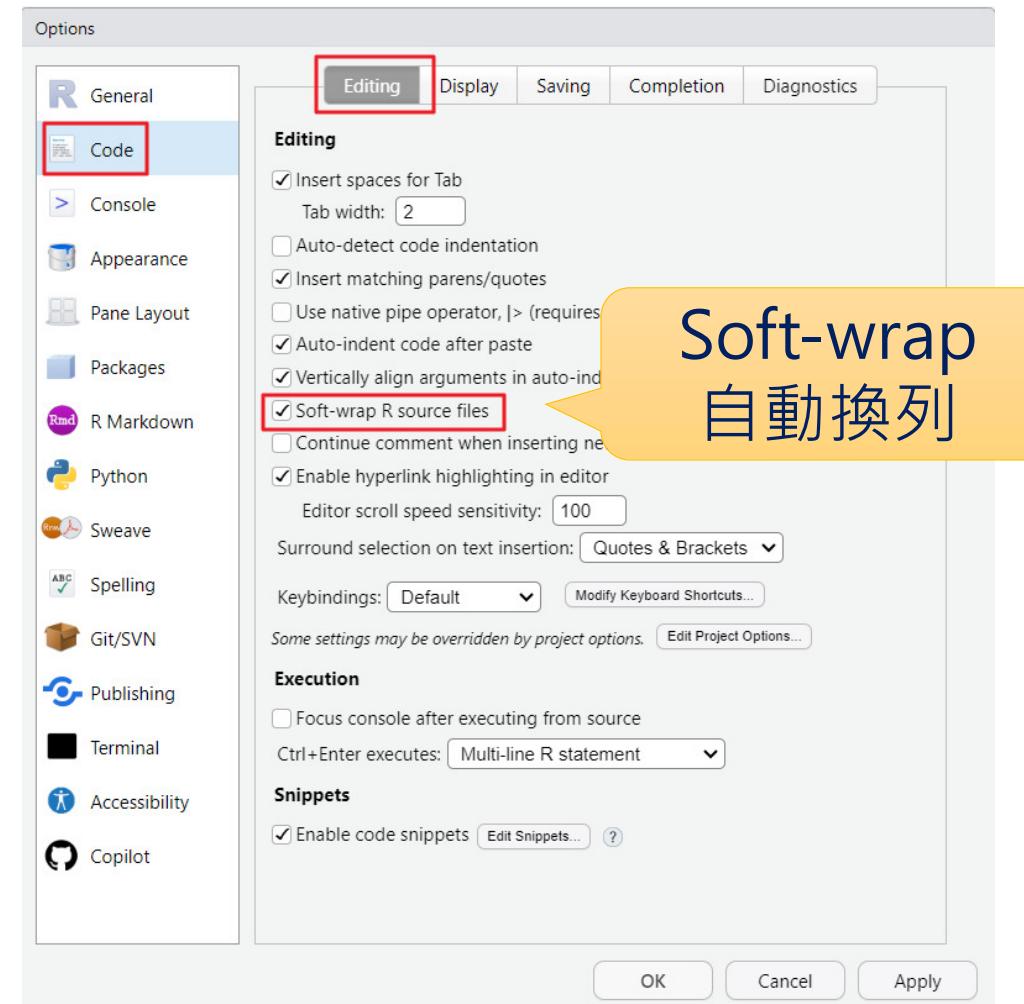
- Tools \ Global Options



General \ Basic

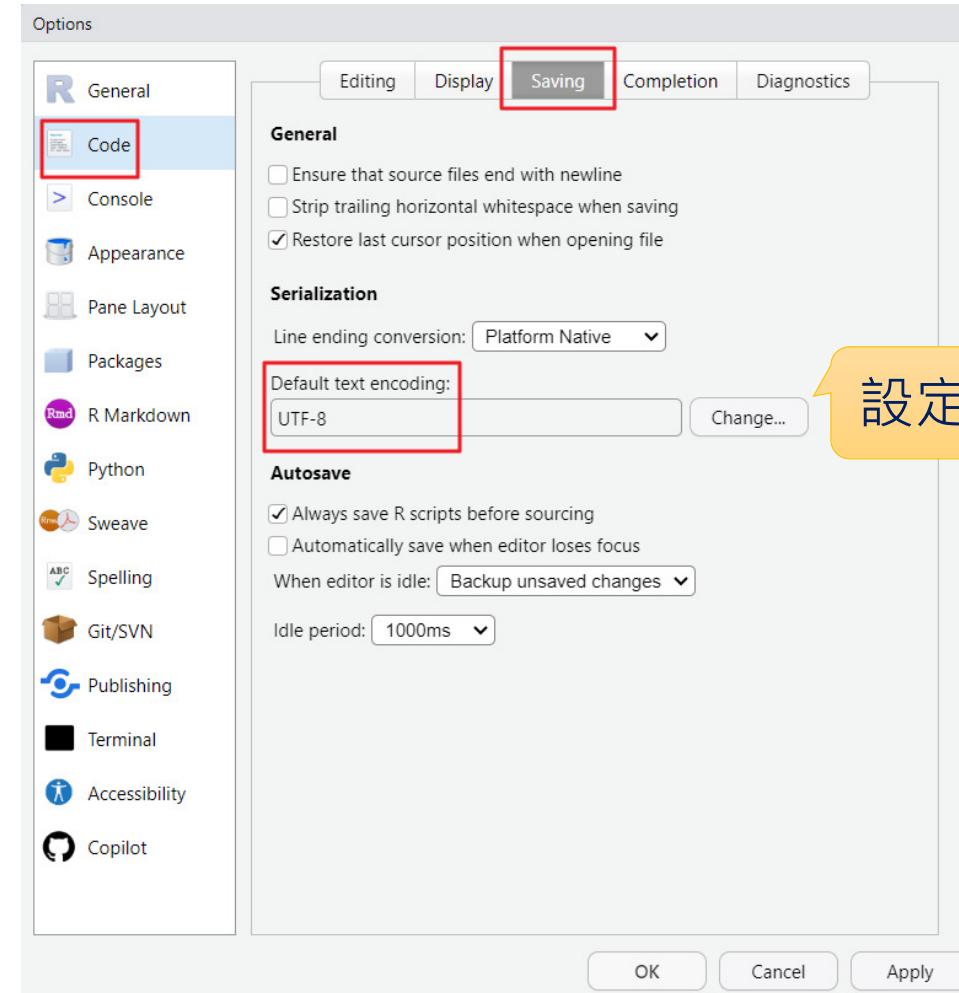


Code \ Editing



Soft-wrap
自動換列

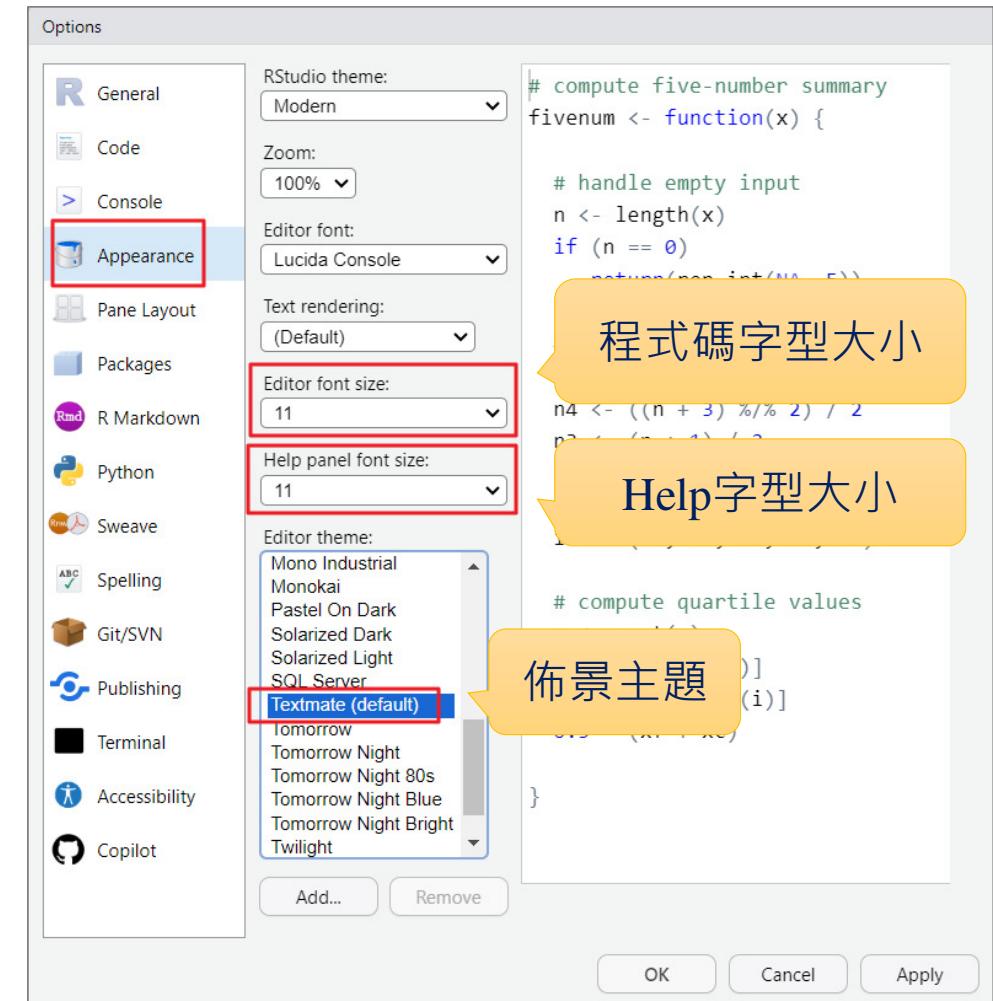
Code \ Saving



RStudio-選項設定(續)

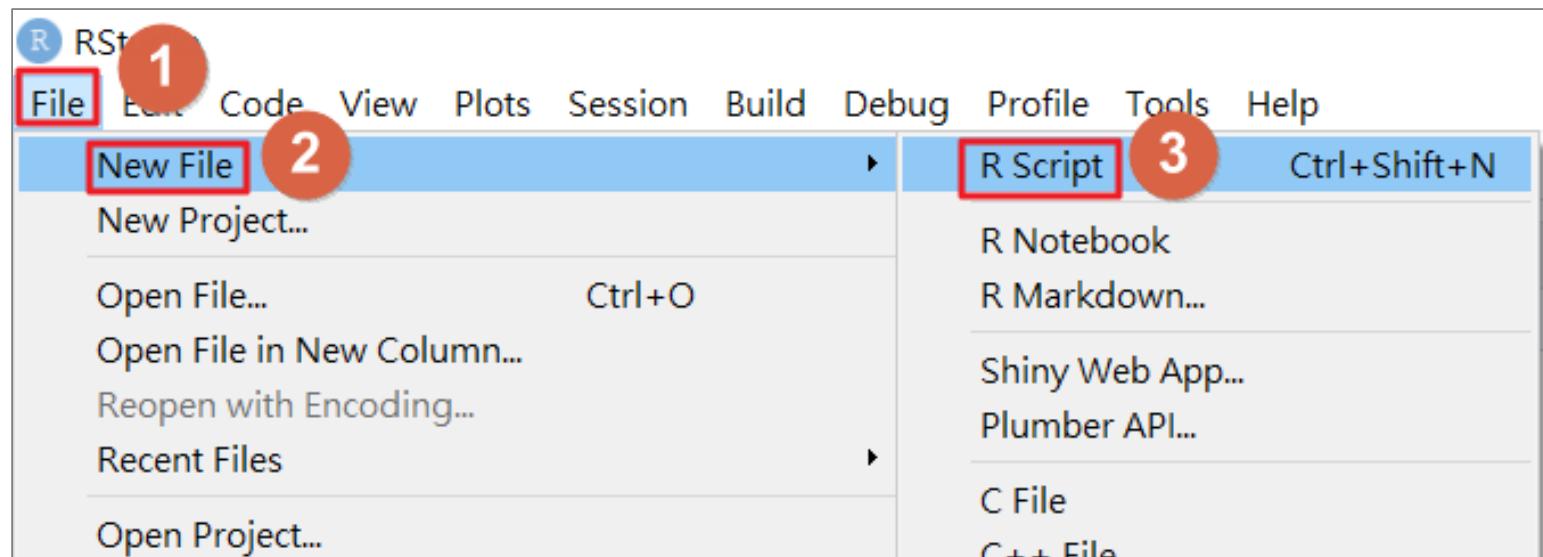
- Appearance \ Editor font size
- Appearance \ Help panel font size
- Editor theme \ TextMate

設定完成，
可能須重新啟動RStudio



新增檔案

- File \ New File \ R Script
(CTRL + SHIFT + N)





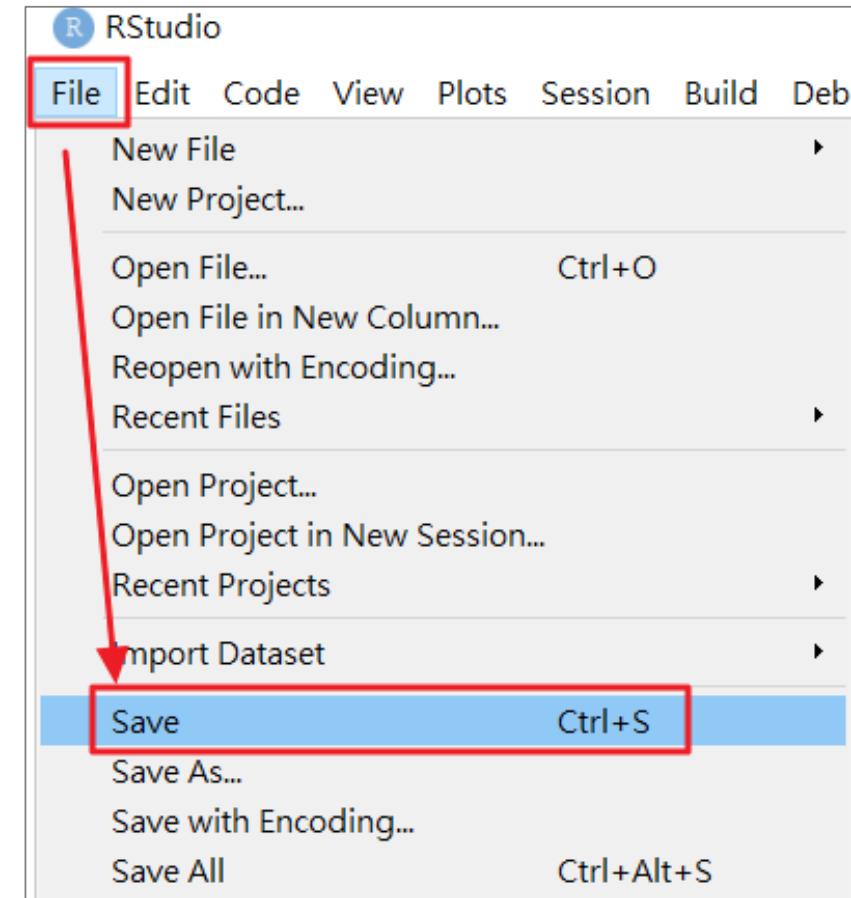
儲存檔案

- 輸入「新增R檔案練習」

實作
練習

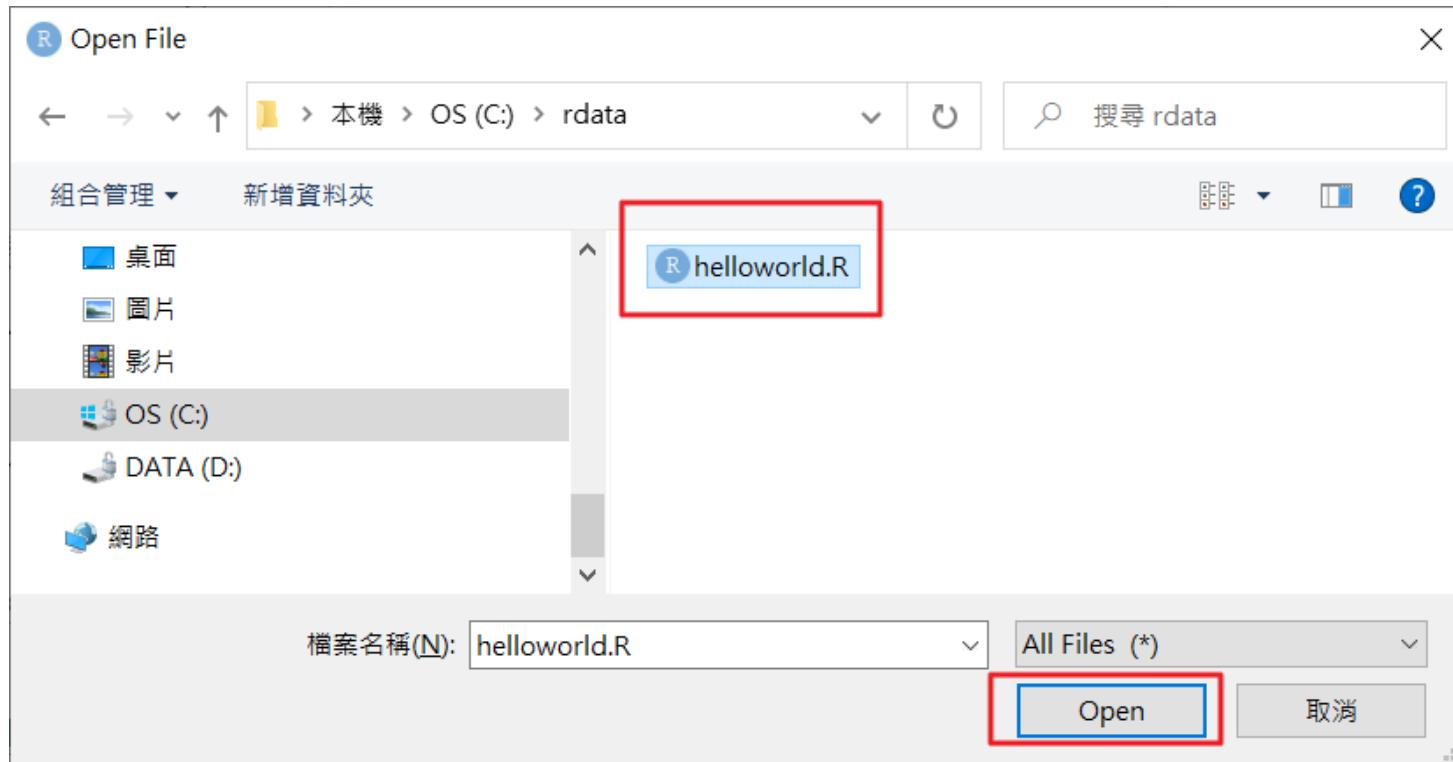
```
1 plot(runif(10), type="b", main= "R大數據分析")
2 x <- rnorm(10)
3 x
4 pairs(iris[-5],
5       pch=16,
6       col=iris$Species,
7       main="RWEPA-iris資料集散佈圖矩陣")
```

- File \ Save (CTRL + S) →
C:\rdata\helloworld.R

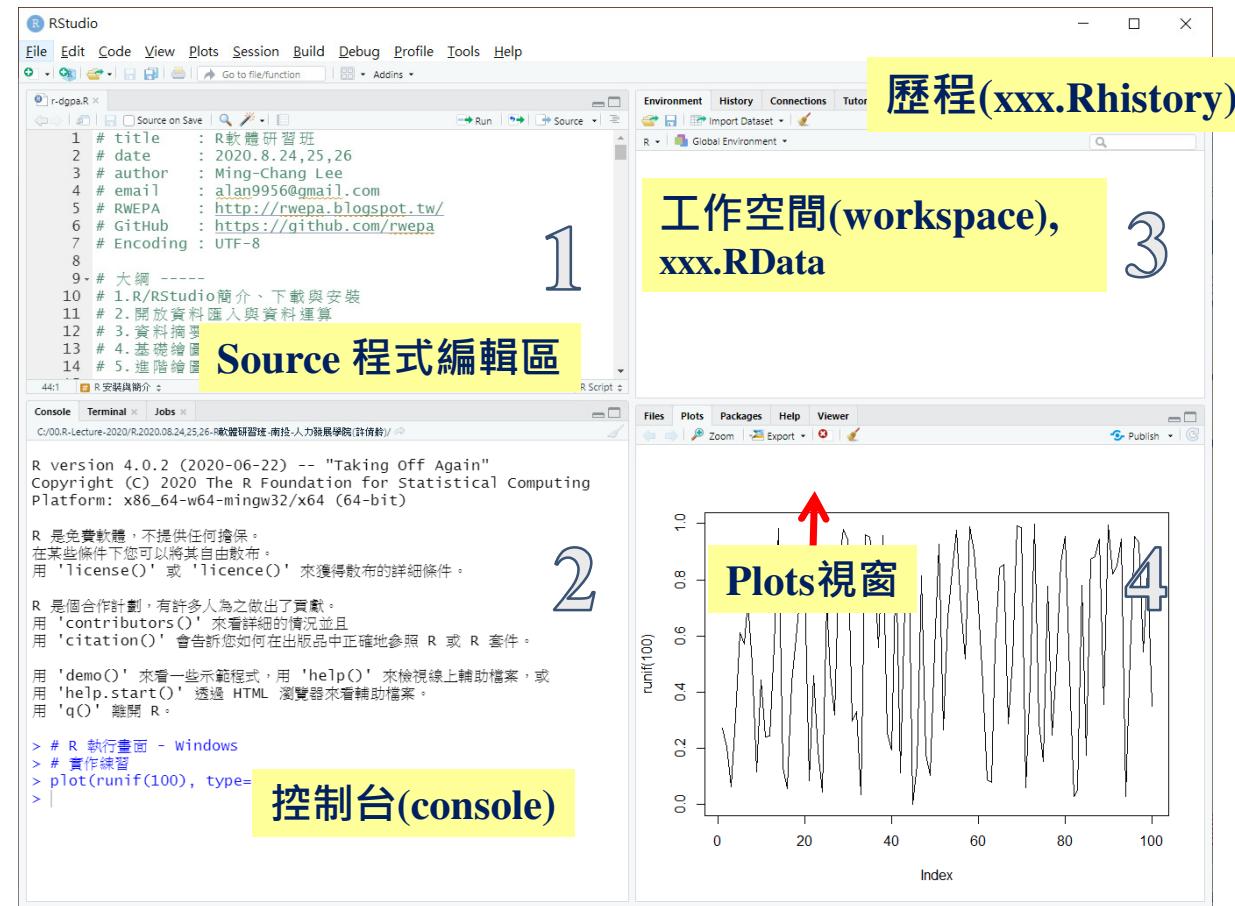


開啟檔案

- File \ Open File \ helloworld.R



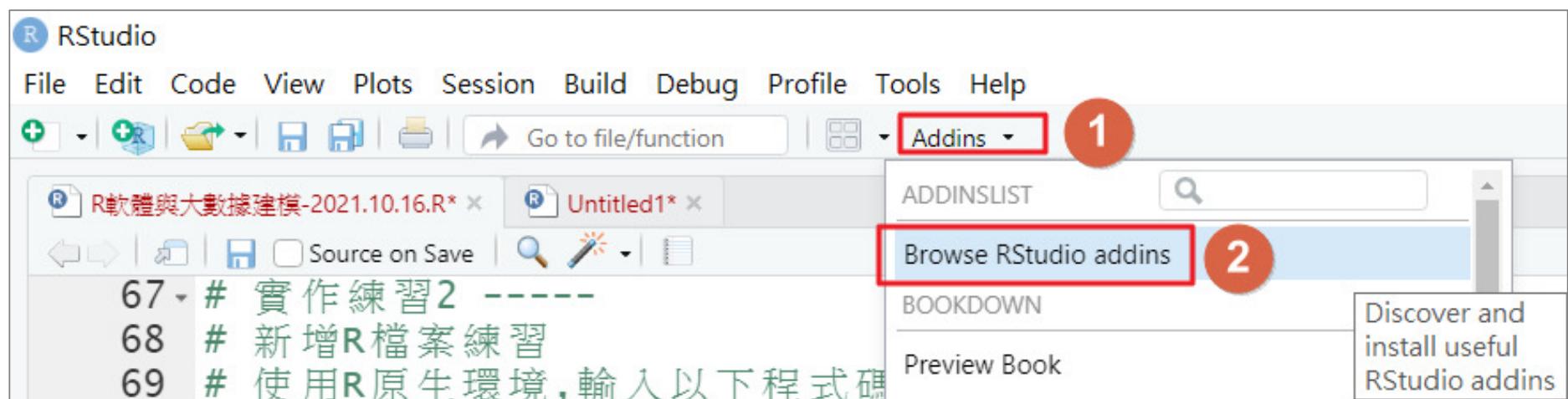
R/RStudio環境的基礎觀念



Ctrl + Shift + F10: 重新啟動R

RStudio Addins (外掛功能)

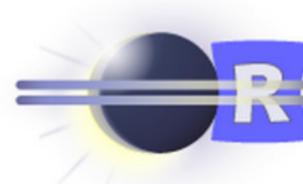
- `install.packages("addinslist")`
- Addins \ Browse RStudio addins



參考: <https://docs.posit.co/ide/user/ide/guide/productivity/add-ins.html>

R + Editor

- R – 原生環境
- RStudio – IDE 整合介面
- Eclipse
 - StatET 4.10.0
An Eclipse based IDE (integrated development environment) plug-in for R.
 - <https://projects.eclipse.org/projects/science.statet>
- Jupyter-notebook
- R 在 Visual Studio Code 中
 - <https://vscode.dev.org.tw/docs/languages/r>



RStudio 快速鍵

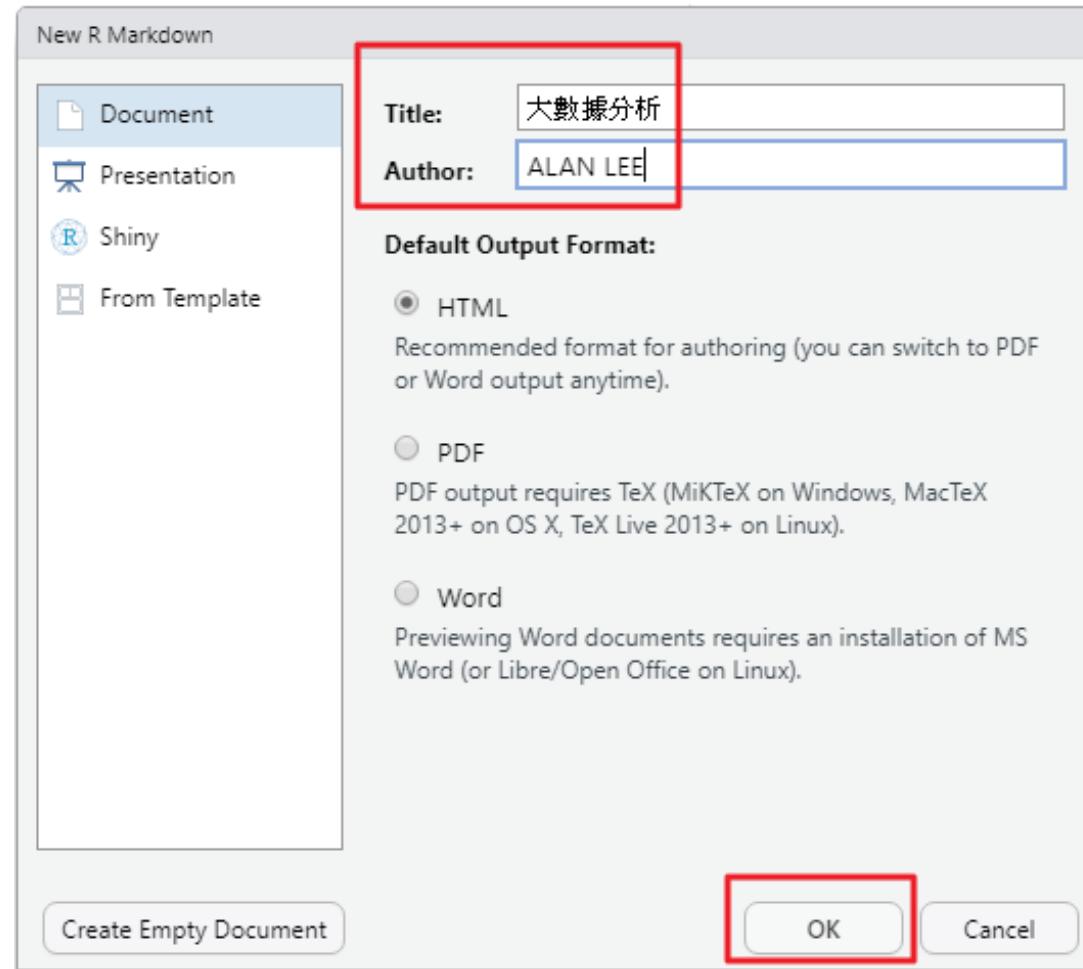
快速鍵	功能
Ctrl + Shift + N	建立新的R程式
Ctrl + S	儲存檔案
Ctrl + Shift + R	建立章節 (-----)
Alt + -	指派符號
Ctrl + Shift + C	註解
Ctrl + Enter	執行程式
Ctrl + Shift + F10	重新啟動R
Alt + Shift + K	快速鍵總表 (Esc 退出)

- 章節功能可以快速切換程式碼

R Markdown

(R 標記語言)

RStudio - Markdown



RStudio - Markdown (續)

The screenshot shows the RStudio interface with a Markdown file open in the left pane. The code includes R Markdown setup and a brief introduction. The right pane shows an empty environment. A yellow callout bubble on the left lists three resources:

1. Help
- 2.
3. RStudio Cheat Sheets

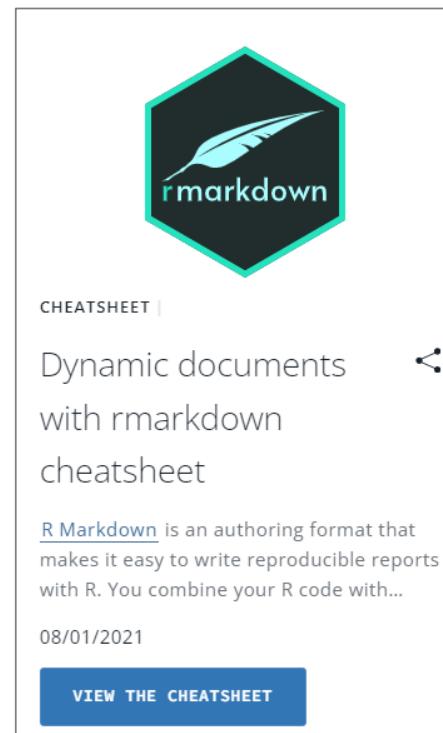
The RStudio Help menu is highlighted with red boxes and numbered circles:

1. Help tab
2. Home icon in the sub-menu
3. RStudio Cheat Sheets link in the sub-menu

```
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  
1:4 大数据分析
```

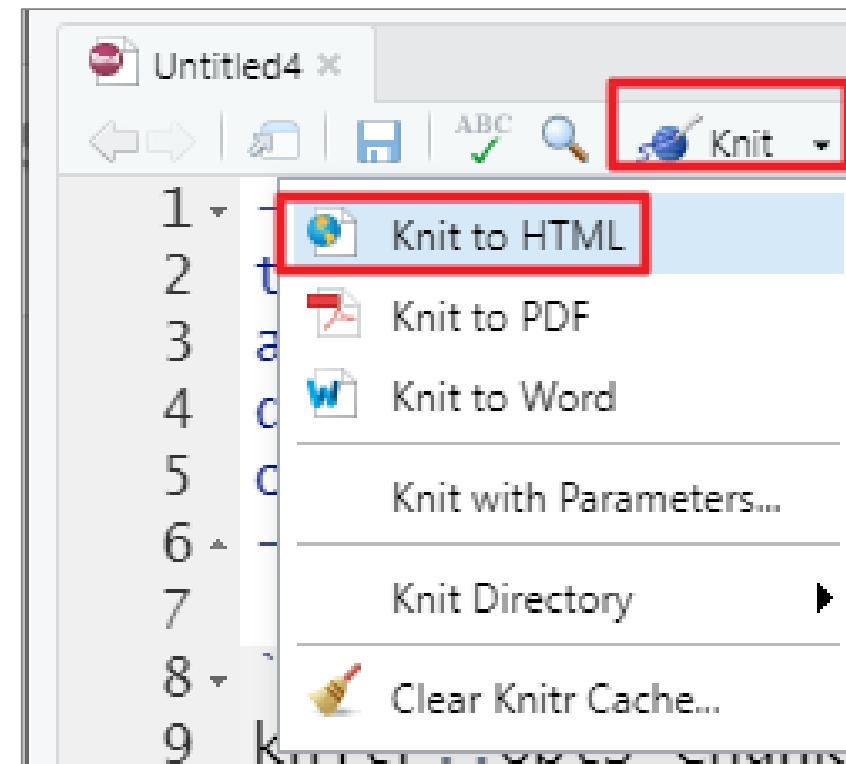
R Markdown Cheatsheet 線上說明

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

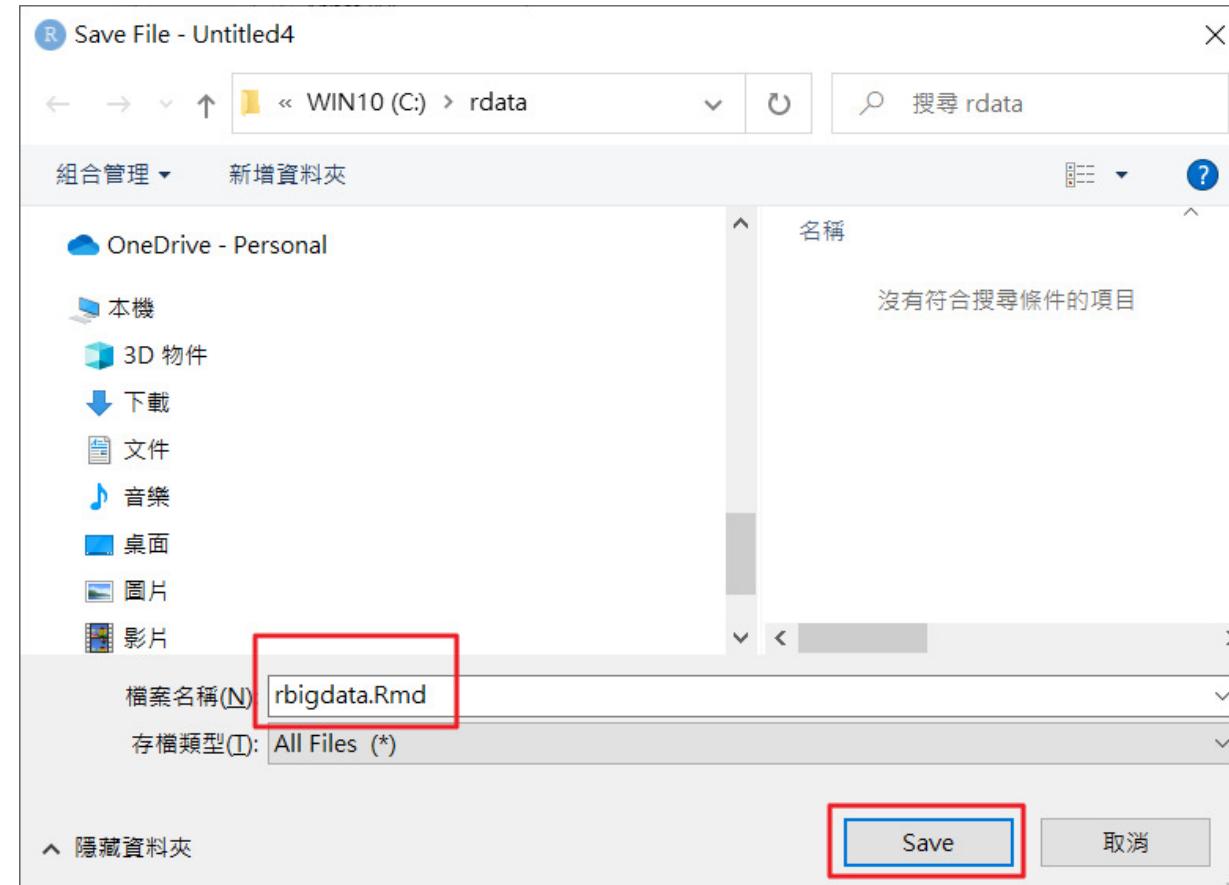


RStudio - Markdown (續)

- Knit HTML
- Knit PDF
- Knit Word



RStudio - Markdown (續)



RStudio - Markdown : HTML

The screenshot shows a browser window displaying an R Markdown document as HTML. The page title is "大數據分析" by ALAN LEE, dated 2020/7/6. The main content is titled "R Markdown" and includes a code chunk with the command `summary(cars)`. A yellow callout bubble labeled "R程式碼" points to this code chunk. Below it, there's a section titled "Including Plots" with a note about embedding plots. A scatter plot of "pressure" vs "speed" is shown, with data points forming a positive correlation. A yellow callout bubble labeled "圖形" points to this plot. The browser window has a tab bar at the top with "大數據分析" and "檔案 | C:/rdata/rbigdata.html". The RStudio interface is visible at the bottom.

大數據分析
ALAN LEE
2020/7/6

R Markdown

This is an R Markdown document. You can edit this document using RStudio's WYSIWYG editor, or you can edit it as plain text using a standard text editor like Notepad or Word. You can also embed R code chunks into your document. When you click the Knit button, RStudio will render the document into a variety of formats, including HTML, PDF, and MS Word documents. For more details on using R Markdown, see the [R Markdown documentation](#).

When you click the Knit button, RStudio will render the document into a variety of formats, including HTML, PDF, and MS Word documents. For more details on using R Markdown, see the [R Markdown documentation](#).

```
summary(cars)
```

	speed	dist
## Min.	4.0	2.00
## 1st Qu.	12.0	18.00
## Median	15.0	36.00
## Mean	15.4	42.98
## 3rd Qu.	19.0	56.00
## Max.	25.0	120.00

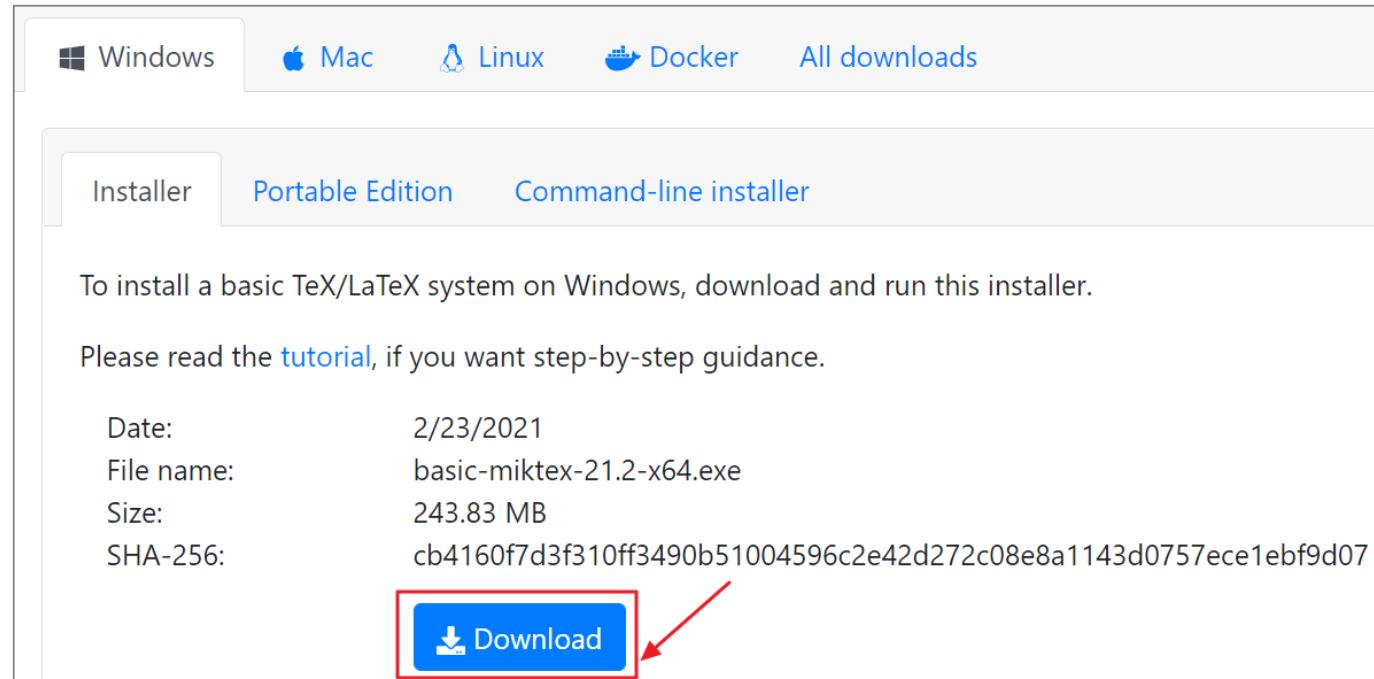
Including Plots

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 strong positive correlation, with most points clustered below 200 speed and below 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



The screenshot shows a PDF viewer window titled "bigdata.pdf". The document contains the following text:

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)
```

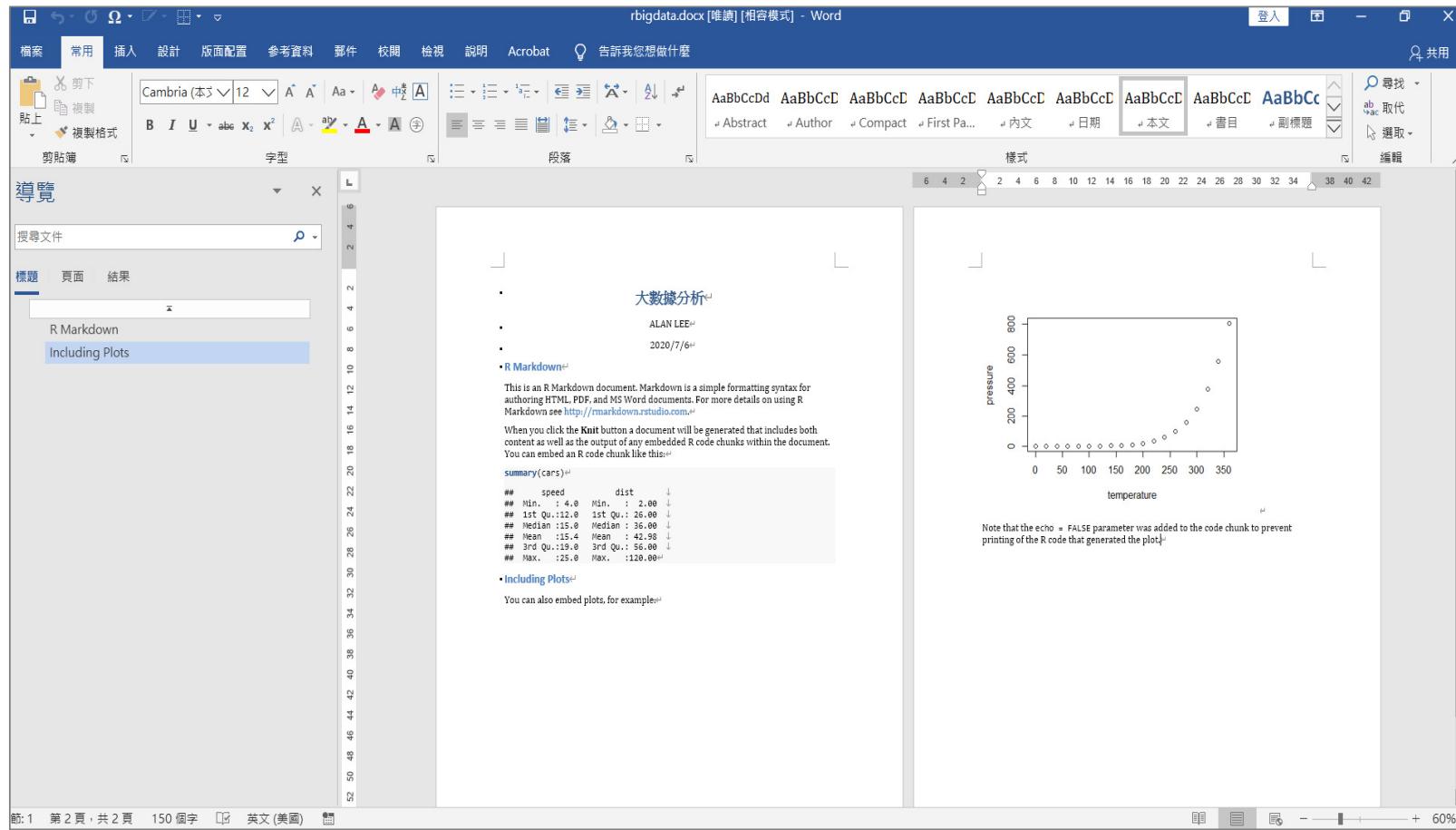
speed dist
Min. : 4.0 Min. : 2.00
1st Qu.:12.0 1st Qu.: 26.00
Median :15.0 Median : 36.00
Mean :15.4 Mean : 42.98
3rd Qu.:19.0 3rd Qu.:56.00
Max. :28.0 Max. :120.00

Including Plots

You can also embed plots, for example:

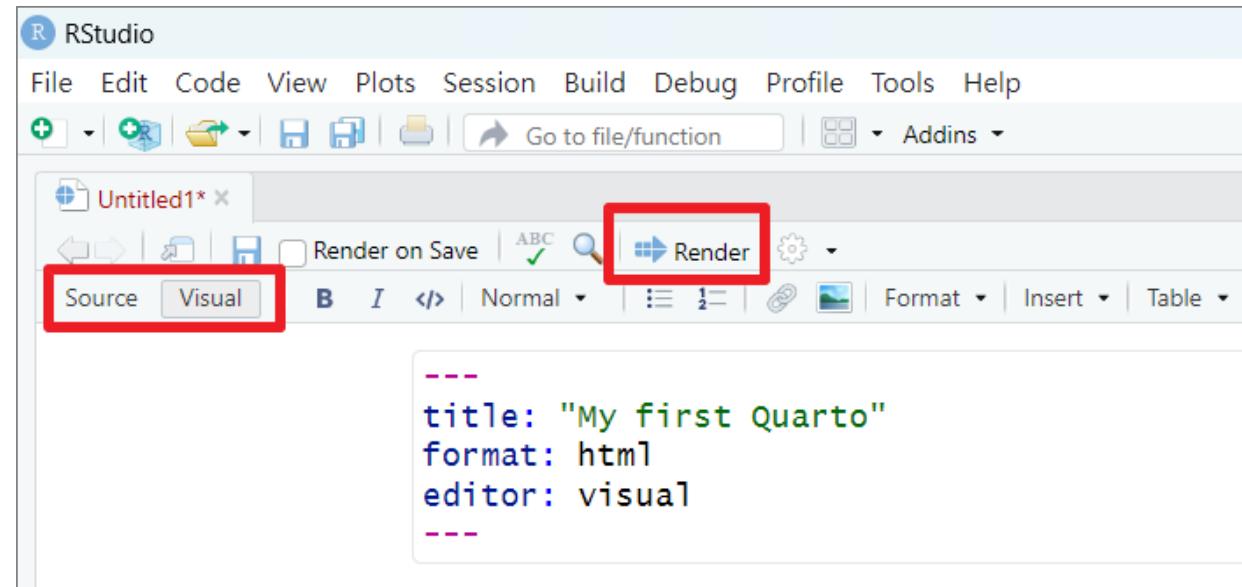
PDF轉換,中文可能有問題?

RStudio - Markdown 轉換為 Word



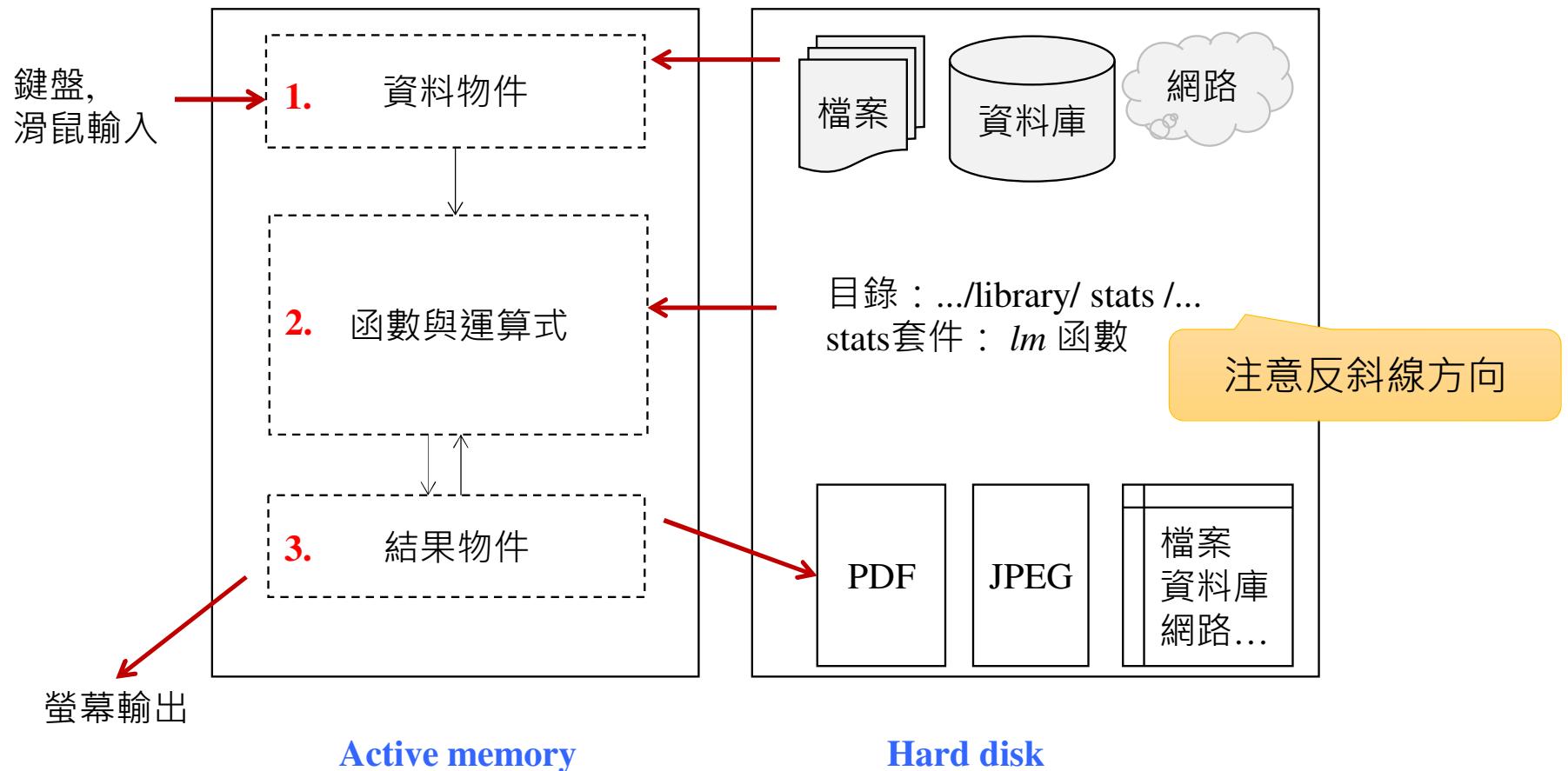
Quarto, Since 2022

- An open-source scientific and technical publishing system
- File \ New File \ Quarto Document...



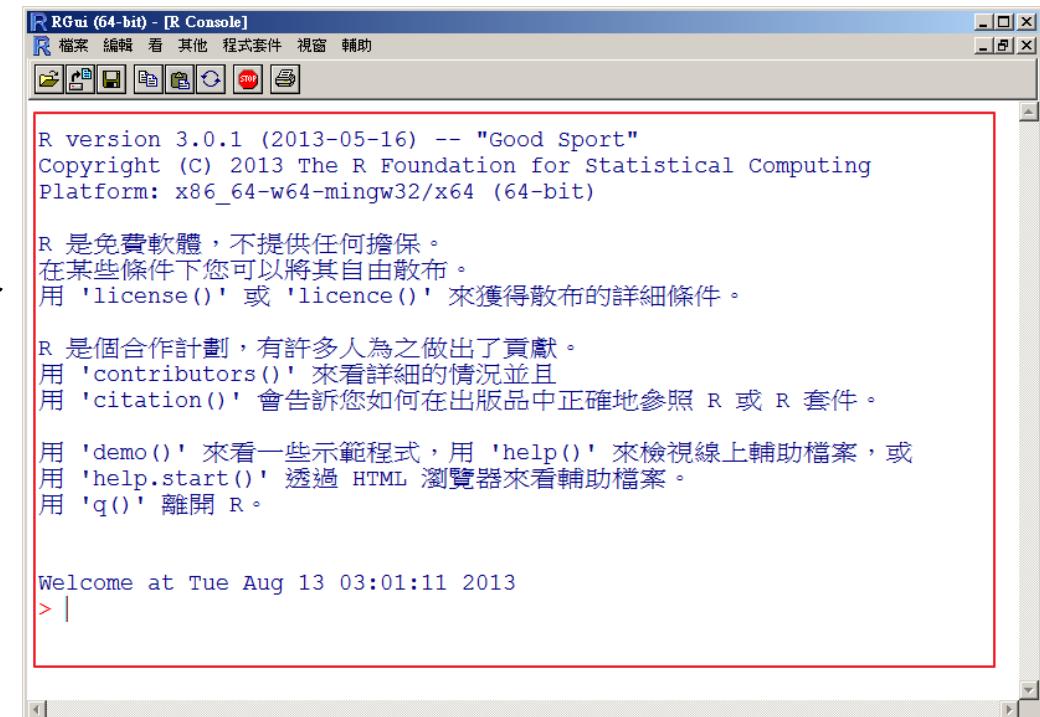
參考: <https://quarto.org/>, 2022.7.28.

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) – 變數集合
 - xxx.RData
- 物件(object)-物件導向程式設計



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。
- 物件名稱起始字元須以文字或 “.” (句點)。
- 物件名稱起始字元不可為數字。
- 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

輔助說明

輔助說明

- `help.start()` # 開啟輔助說明的首頁
- `?plot` # plot 函數說明
- `help(plot)` # plot 函數說明
- 選取 plot 按 F1 # plot 函數說明
- `help.search("regression")` # 搜尋關鍵字 regression
- `??regression` # 搜尋關鍵字 regression



Google

在以下這些程式套件裡找到了關於 'plot' 主題的說明:

[The Default Scatterplot Function](#)



(in package [graphics](#) in library C:/Program Files/R/R-4.4.1/library)

[Generic X-Y Plotting](#)

(in package [base](#) in library C:/PROGRA~1/R/R-44~1.1/library)

輔助說明 (續)

1.函數

plot.default {graphics}

2.套件

1

2

The Default Scatterplot Function

3

R Documentation

3.R文件

type="n"

Description

Draw a scatter plot with decorations such as axes and titles in the active graphics window.

Usage

```
## Default S3 method:  
plot(x, y = NULL, type = "p", xlim = NULL, ylim = NULL,  
      log = "", main = NULL, sub = NULL, xlab = NULL, ylab = NULL,  
      ann = par("ann"), axes = TRUE, frame.plot = axes,  
      panel.first = NULL, panel.last = NULL, asp = NA,  
      xgap.axis = NA, ygap.axis = NA,  
      ...)
```

6

方法

Arguments

x, y

the x and y arguments provide the x and y coordinates for the plot. Any reasonable way of defining the coordinates is acceptable. See the function [xy.coords](#) for details. If supplied separately, they must be of the same length.

5

詳細說明

4.簡單說明

7.參數

3.R文件

套件 Package (模組 Module)

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 - 44類別

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

Contributed Packages

(2024.7.13)

Available Packages

Currently, the CRAN package repository features 21075 available packages.

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

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

[CRAN Task Views](#) aim to provide some guidance which packages on CRAN are relevant for tasks related to a certain topic. They provide tools to automatically install all packages from each view. Currently, 44 views are available.

44類別 - 中文說明

- <http://rwepa.blogspot.com/2013/10/packages-list-32.html>

2013年10月8日 星期二

Task Views - R套件

RWEPA → task

更新日期: 2023.9.4 - 44個套件類別

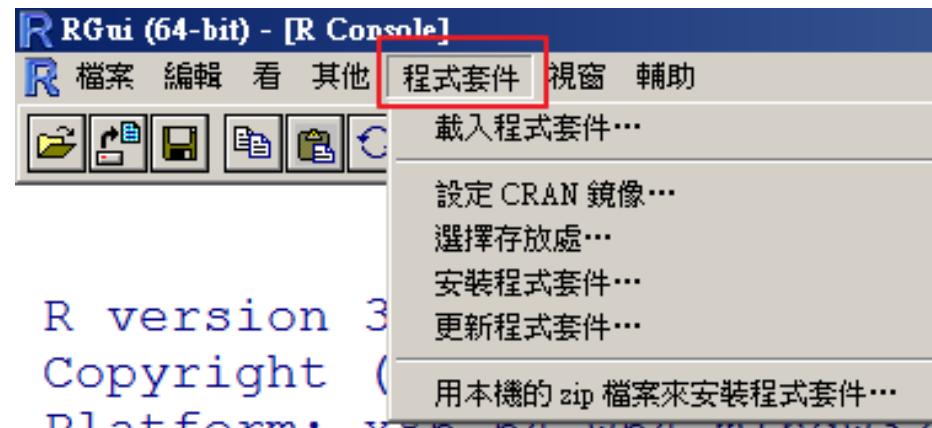
CRAN Task View:
<https://cran.csie.ntu.edu.tw/web/views/>

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

選取 CRAN 網站左側 [Packages]，套件區分成以下類別，中文說明如下:

1;ActuarialScience;Actuarial Science;精算學
2;Agriculture;Agricultural Science;農業學
3;Bayesian;Bayesian Inference;貝氏統計

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: R版本, 作業系統, 載入套件

```
> # R對話資訊 -----
> sessionInfo()
R version 4.4.1 (2024-06-14 ucrt)
Platform: x86_64-w64-mingw32/x64
Running under: Windows 11 x64 (build 22631)

Matrix products: default

locale:
[1] LC_COLLATE=Chinese (Traditional)_Taiwan.utf8
[2] LC_CTYPE=Chinese (Traditional)_Taiwan.utf8
[3] LC_MONETARY=Chinese (Traditional)_Taiwan.utf8
[4] LC_NUMERIC=C
[5] LC_TIME=Chinese (Traditional)_Taiwan.utf8

time zone: Asia/Taipei
tzcode source: internal

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

loaded via a namespace (and not attached):
[1] compiler_4.4.1    tools_4.4.1       rstudioapi_0.16.0
```

套件安裝目錄

- .Library

```
> # 預設套件安裝目錄  
> .Library  
[1] "C:/PROGRA~1/R/R-44~1.1/library"  
>
```

- .libPaths()

- 可能全部安裝在 R \ library

```
> # 套件安裝目錄  
> # 可能全部安裝在 R\library  
> .libPaths()  
[1] "C:/Users/asus/AppData/Local/R/win-library/4.4"  
[2] "C:/Program Files/R/R-4.4.1/library"
```

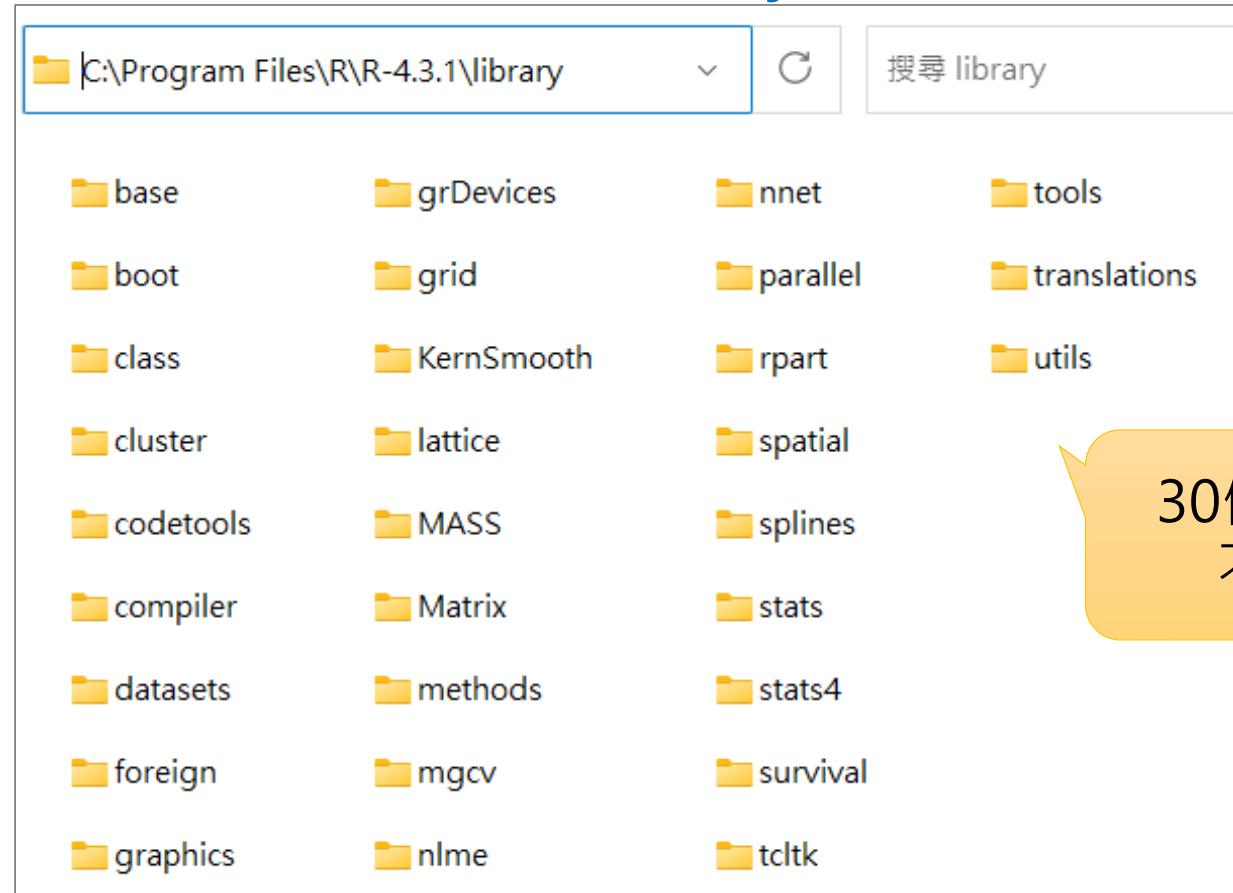
已安裝套件

```
> # 已安裝套件
> myinstalled <- installed.packages()
> class(myinstalled) # "matrix" "array"
[1] "matrix" "array"
> dim(myinstalled)    # 626*16
[1] 719  16
> mypackage <- myinstalled[, 1] # matrix[列, 行]
> mypackage[1:10]
      abind      addinslist       ade4          AER         affy
      "abind"    "addinslist"   "ade4"        "AER"       "affy"
      affydata     affyio       agricolae
      "affydata"   "affyio"    "agricolae"
                                         "agricolae"
```

```
library() # same as installed.packages()
```

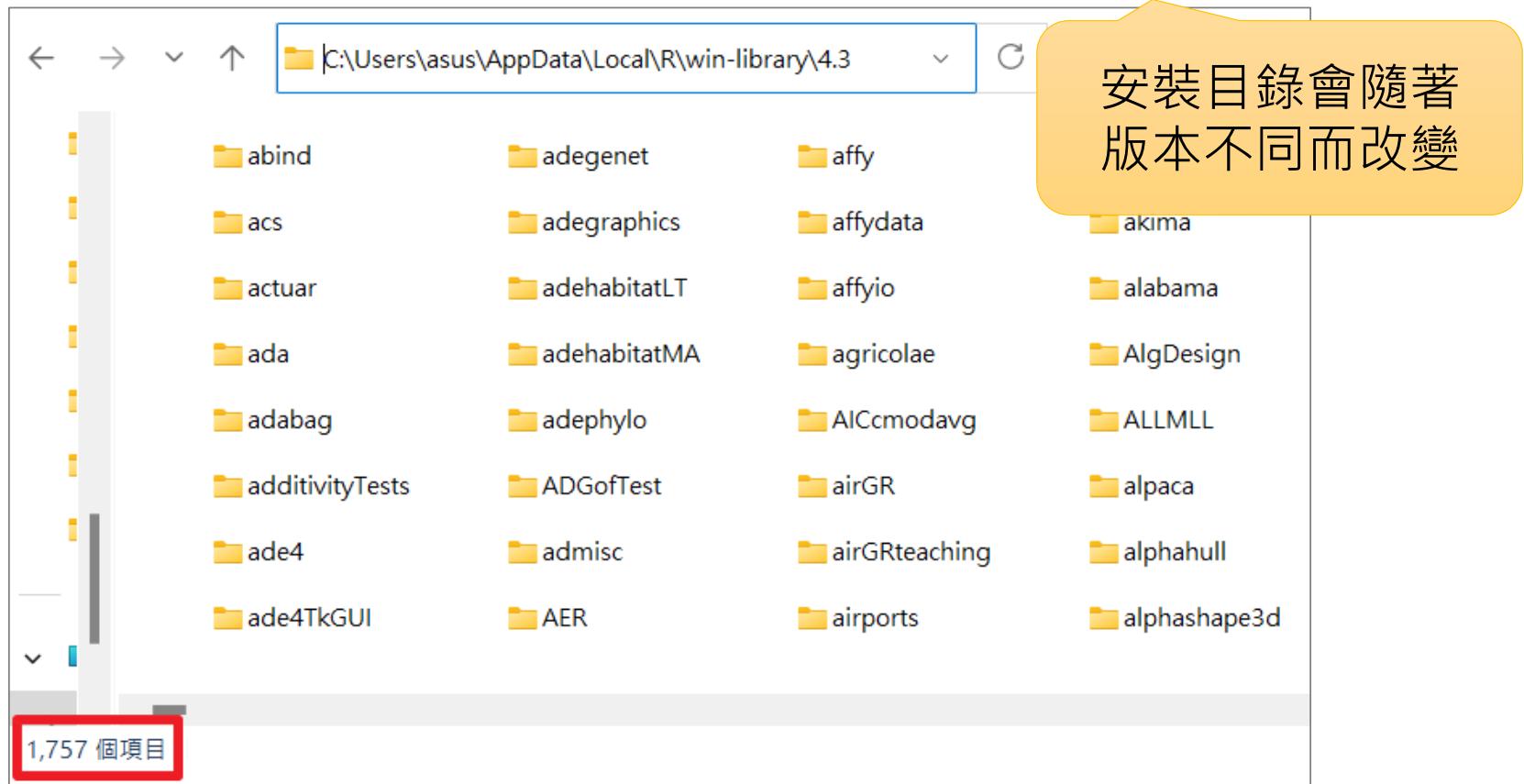
套件安裝目錄1

- C:/Program Files/R/R-4.3.1/library



套件安裝目錄2

- C:/Users/asus/AppData/Local/R/win-library/4.3



RStudio 套件管理

The screenshot shows the RStudio interface with the 'Packages' tab selected. A yellow callout bubble points to the 'base' package in the 'System Library' section, indicating it is the currently loaded package.

Name	Description	Version	Actions
zeallot	Maintained by 'YuLab-SMU' Multiple, Unpacking, and Destructuring Assignment	0.1.0	🌐 X
zip	Cross-Platform 'zip' Compression	2.3.0	🌐 X
zlibbioc	An R packaged zlib-1.2.5	1.46.0	🌐 X
zoo	S3 Infrastructure for Regular and Irregular Time Series (Z's Ordered Observations)	1.8-12	🌐 X

System Library			
<input checked="" type="checkbox"/>	base	The R Base Package	4.3.1
<input type="checkbox"/>	boot	Bootstrap Functions (Originally by Angelo Canty for S)	1.3-28.1
<input type="checkbox"/>	class	Functions for Classification	7.3-22

打勾表示已經
載入套件

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

R demo

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

4. 開放資料分析

open data

- RWEPA → open data
- RWEPA資料下載
 - <https://github.com/rwepa/DataDemo>
- 政府資料開放平台
 - <https://data.gov.tw/>
- UCI Machine Learning Repository
 - <https://archive.ics.uci.edu/datasets>
- Google Dataset Search
 - <https://datasetsearch.research.google.com/>
- Kaggle Dataset
 - <https://www.kaggle.com/datasets>
- World Bank Open Data
 - <https://data.worldbank.org/>

Service : alan9956@gmail.com

大數據分析,資料視覺化,R,PYTHON, Tableau,PowerBI程式設計,統計品管,最佳化,企業服務,業師協同教學.

RWEPA 搜尋此網誌 (例: task)

資料建立與輸入輸出

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

資料分析

- 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 1  
> getwd()  
[1] "c:/Users/asus/Documents"  
>  
> # 設定工作目錄  
> workpath <- "c:/rdata"  
> setwd(workpath) 2  
>  
> # 已更改變為 "c:/rdata" 工作目錄  
> getwd() 3  
[1] "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 4.38 (8 人次投票)

瀏覽次數: 20139 下載次數: 6867 意見數: 5

主要欄位說明
*粗體欄位為資料標準欄位

siteid(測站編號)、sitename(測站名稱)、monitordate(監測日期)、aqi(空氣品質指標)、so2subindex
cosubindex(一氧化碳副指標)、o3subindex(臭氧副指標)、pm10subindex(懸浮微粒副指標)、no2su
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

CSV 檢視資料 日空氣品質指標(AQI)-CSV

JSON 檢視資料 日空氣品質指標(AQI)-JSON

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	85	大城	2023/10/16	150	0	3		46	8	150	65
3	84	富貴角	2023/10/16	174	0	2		48	3	174	62
4	83	麥寮	2023/10/16	74	8	3		51	17		74
5	80	關山	2023/10/16	61	2	3		27	8	61	41
6	78	馬公	2023/10/16	140	2	3		34	5	140	66
7	77	金門	2023/10/16	156	8	5				156	84
8	75	馬祖	2023/10/16	161	8	6				161	85
9	72	埔里	2023/10/16	62	2	5				45	62
10	71	復興	2023/10/16	104	8	10		58	57		104
11	70	永和	2023/10/16	69	5	5		38	23		69

遺漏值

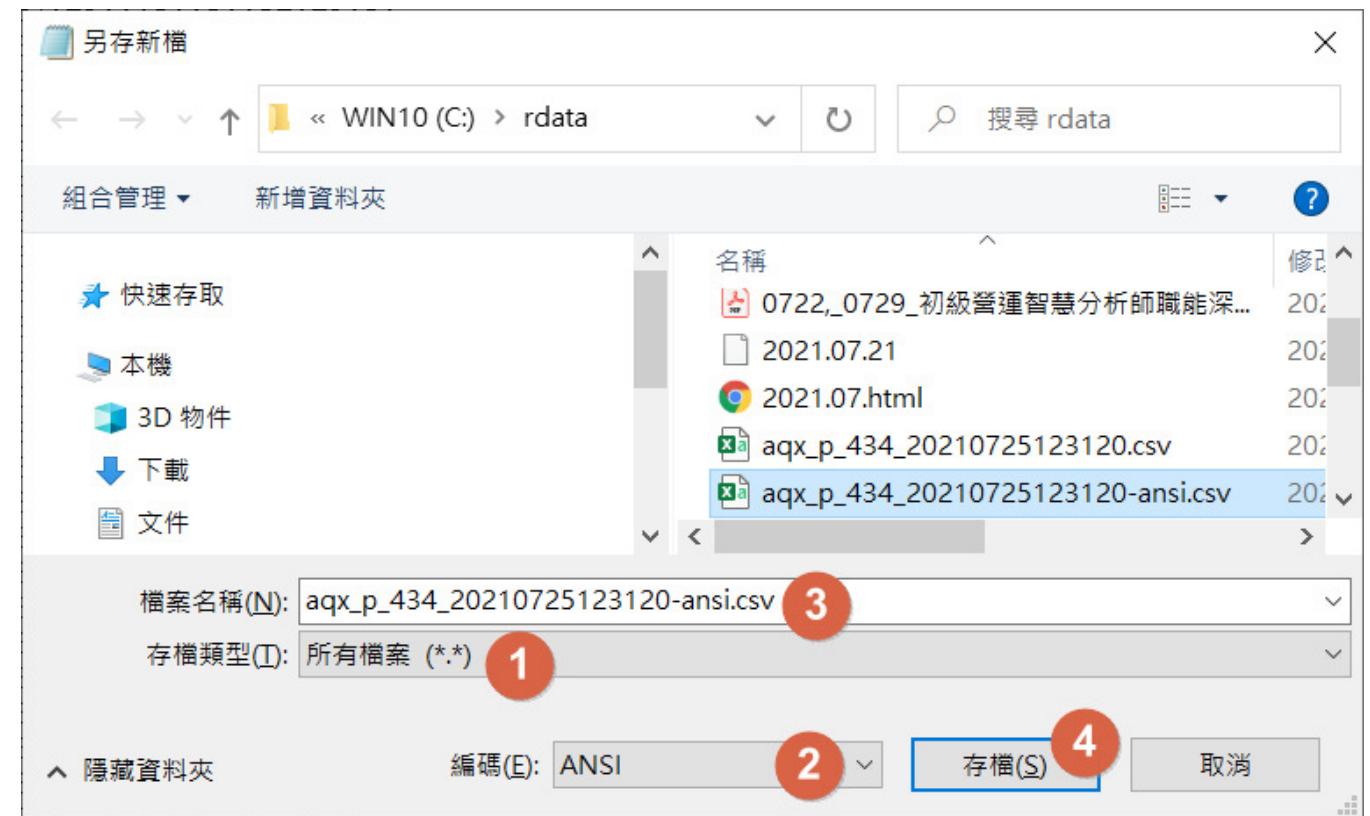
步驟 3. 汇入資料 `read.table`

- `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

R-v3.x早期版本匯入 UTF-8 文字檔會有錯誤

```
> myfile <- "aqx_p_434_20210725123120.csv"  
>  
1> aq <- read.table(myfile, header=TRUE, sep=",") # error!  
2Error 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  
4 嘿燙iteId      SiteName MonitorDate AQI SO2SubIndex cosubIndex O3SubIndex  
5 18          懶批<9c><92> 2021-07-24 28 2 1 NA  
6 19 閩\u0080<e9>\u20 ,2021-07-24 35 2 1 NA 25  
7 20          搞喲\ 2021-07-24 24 0 1 NA  
8 PM10SubIndex NO2SubIndex O38SubIndex PM25SubIndex  
9 19 10 28 20  
10 8 35 22 NA  
11 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-v3.x早期版本 - 正常匯入ANSI 文字檔

```
> # 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
```

資料檢視 head

- NA表示遺漏值

```
> # 資料檢視
> head(aq)
```

	siteid	sitename	monitordate	aqi	so2subindex	cosubindex	o3subindex	pm10subindex
1	85	大城	2023-10-16	150	0	3	NA	46
2	84	富貴角	2023-10-16	174	0	2	NA	48
3	83	麥寮	2023-10-16	74	8	3	NA	51
4	80	關山	2023-10-16	61	2	3	NA	27
5	78	馬公	2023-10-16	140	2	3	NA	34
6	77	金門	2023-10-16	156	8	5	NA	48
				no2subindex	o38subindex	pm25subindex		
1				8	150	65		
2				3	174	62		
3				17	NA	74		
4				8	61	41		
5				5	140	66		
6				20	156	84		

R-v4.3.1版本正常
匯入UTF-8文字檔

欄位名稱 names

```
> # 欄位名稱  
> names(aq)  
[1] "siteid"          "sitename"        "monitordate"    "aqi"           "so2subindex"  
[6] "cosubindex"      "o3subindex"      "pm10subindex"   "no2subindex"   "o38subindex"  
[11] "pm25subindex"
```

步驟 4. 資料處理 - 資料結構 str

```
> # 資料結構
> str(aq)
'data.frame': 1000 obs. of 11 variables:
 $ siteid      : int  85 84 83 80 78 77 75 72 71 70 ...
 $ sitename     : chr "大城" "富貴角" "麥寮" "關山" ...
 $ monitordate : chr "2023-10-16" "2023-10-16" "2023-10-16" "2023-10-16" ...
 $ aqi          : int 150 174 74 61 140 156 161 62 104 69 ...
 $ so2subindex  : int 0 0 8 2 2 8 8 2 8 5 ...
 $ cosubindex   : int 3 2 3 3 3 5 6 5 10 5 ...
 $ o3subindex   : logi NA NA NA NA NA NA ...
 $ pm10subindex: int 46 48 51 27 34 48 45 29 58 38 ...
 $ no2subindex  : int 8 3 17 8 5 20 12 15 57 23 ...
 $ o38subindex  : int 150 174 NA 61 140 156 161 45 NA NA ...
 $ pm25subindex: int 65 62 74 41 66 84 85 62 104 69 ...
```

資料摘要 summary

```
> # 資料摘要
> summary(aq)

  siteid      sitename      monitordate        aqi      so2subindex
Min.   : 1.00  Length:1000    Length:1000    Min.   :-1.00  Min.   : 0.000
1st Qu.:21.00  class  :character  Class  :character  1st Qu.: 41.00  1st Qu.: 2.000
Median  :40.00  Mode   :character  Mode   :character  Median  : 48.00  Median  : 2.000
Mean    :40.65
3rd Qu.:60.00
Max.   :85.00

  cosubindex    o3subindex    pm10subindex    no2subindex    o38subindex
Min.   : 0.000  Mode:logical  Min.   : 0.00  Min.   : 0.00  Min.   : 12.00
1st Qu.: 2.000  NA's:1000    1st Qu.: 16.50  1st Qu.:13.00  1st Qu.: 40.00
Median  : 3.000
Mean    : 3.248
3rd Qu.: 5.000
Max.   :16.000
NA's   :8

  pm25subindex
Min.   : 0.0
1st Qu.: 23.0
Median  : 35.0
Mean    : 38.8
3rd Qu.: 53.0
Max.   :128.0
NA's   :37

  NA's   :5

```

NA 遺漏值

轉換為日期 as.Date

```
> # 日期: 字串(chr)修正為日期(Date)
> aq$monitordate <- as.Date(aq$monitordate)
> str(aq)
```

'data.frame': 1000 obs. of 11 variables:

	siteid	sitename	monitordate	aqi	so2subindex	cosubindex	o3subindex	pm10subindex	no2subindex	o38subindex	pm25subindex
\$ siteid	: int 85 84 83 80 78 77 75 72 71 70 ...										
\$ sitename	: chr "大城" "富貴角" "麥寮" "關山" ...										
\$ monitordate	: Date, format: "2023-10-16" "2023-10-16" ...										
\$ aqi	: int 150 174 74 61 140 156 161 62 104 69 ...										
\$ so2subindex	: int 0 0 8 2 2 8 8 2 8 5 ...										
\$ cosubindex	: int 3 2 3 3 3 5 6 5 10 5 ...										
\$ o3subindex	: logi NA NA NA NA NA NA ...										
\$ pm10subindex	: int 46 48 51 27 34 48 45 29 58 38 ...										
\$ no2subindex	: int 8 3 17 8 5 20 12 15 57 23 ...										
\$ o38subindex	: int 150 174 NA 61 140 156 161 45 NA NA ...										
\$ pm25subindex	: int 65 62 74 41 66 84 85 62 104 69 ...										

monitordate 原為 chr 資料類型

資料維度 dim

```
> # 資料處理
> head(aq, n=3)
  siteid sitename monitordate aqi so2subindex cosubindex o3subindex pm10subindex
1     85    大城 2023-10-16 150          0            3        NA         46
2     84   富貴角 2023-10-16 174          0            2        NA         48
3     83    麥寮 2023-10-16  74          8            3        NA         51
  no2subindex o38subindex pm25subindex
1           8          150            65
2           3          174            62
3          17             NA            74
> dim(aq) # 1000列11行
[1] 1000   11
```

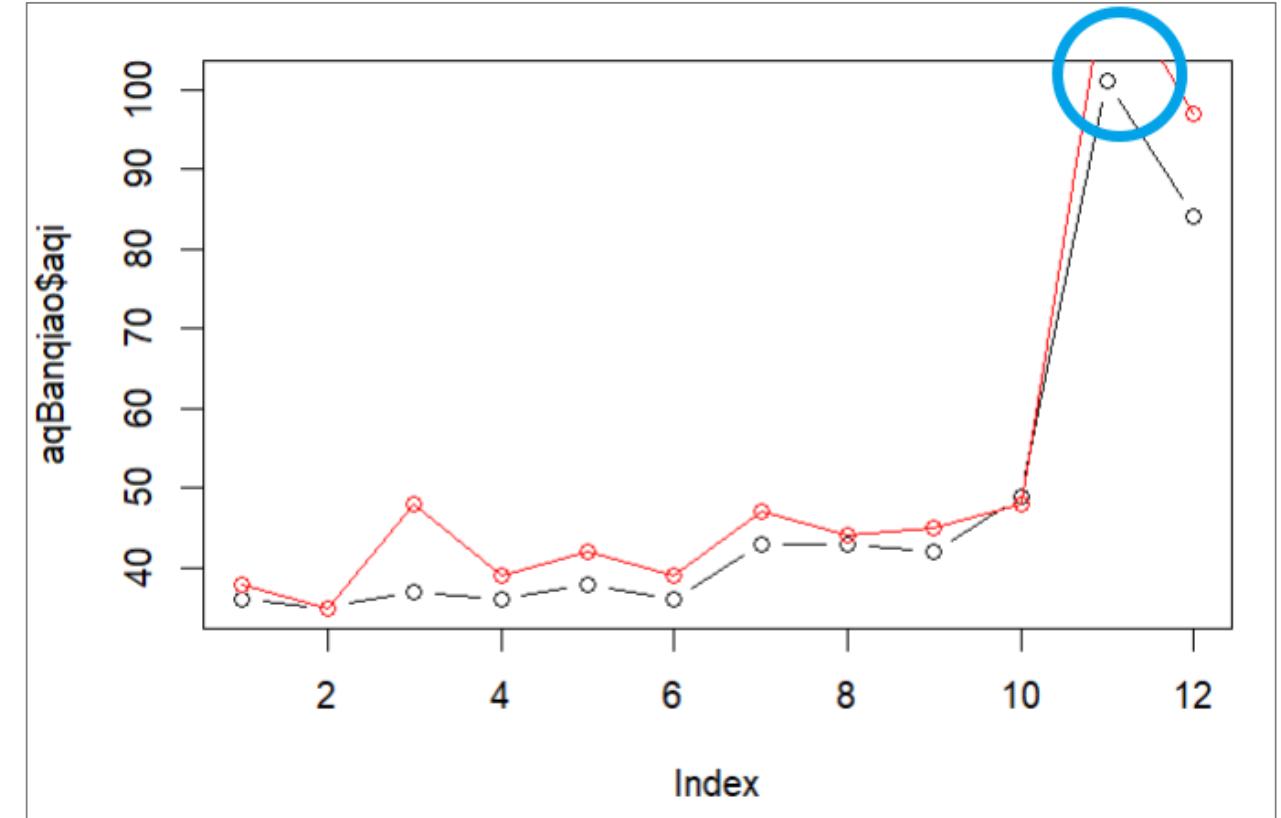
資料篩選與排序

`==` 表示判斷是否相等

```
> # 篩選 板橋 資料  
> aqBanqiao<- aq[aq$sitename == "板橋",]  
>  
> # 依照 monitordate 欄位由小至大遞增排序  
> aqBanqiao <- aqBanqiao[order(aqBanqiao$monitordate),]  
>  
> # 篩選 汐止 資料  
> aqXizhi <- aq[aq$sitename == "汐止",]  
>  
> # 依照 monitordate 欄位由小至大遞增排序  
> aqXizhi <- aqXizhi[order(aqXizhi$monitordate),]
```

板橋暨汐止AQI趨勢圖

```
> # 繪製板橋暨汐止AQI趨勢圖  
> plot(aqBanqiao$aqi, type="b")  
> lines(aqxizhi$aqi, col="red")  
> points(aqxizhi$aqi, col="red")
```



優化趨勢圖

```
> # 客製化Y軸最小值,最大值  
> ymin <- min(aqBanqiao$aqi, aqxizhi$aqi) - 1  
> ymax <- max(aqBanqiao$aqi, aqxizhi$aqi) + 3  
>  
> # 優化趨勢圖  
> plot(aqBanqiao$aqi,  
+       type = "b",  
+       ylim = c(ymin, ymax),  
+       main = paste0(aq$monitordate[1], "AQI 板橋vs.汐止"))  
> lines(aqxizhi$aqi, col="red")  
> points(aqxizhi$aqi, col="red")  
> legend("topleft", legend=c("板橋", "汐止"), col=c(1,2), lty=1)
```

- plot 繪圖
- ylim Y軸範圍
- lines 線
- points 點
- legend 圖例

步驟 5. 汇出資料

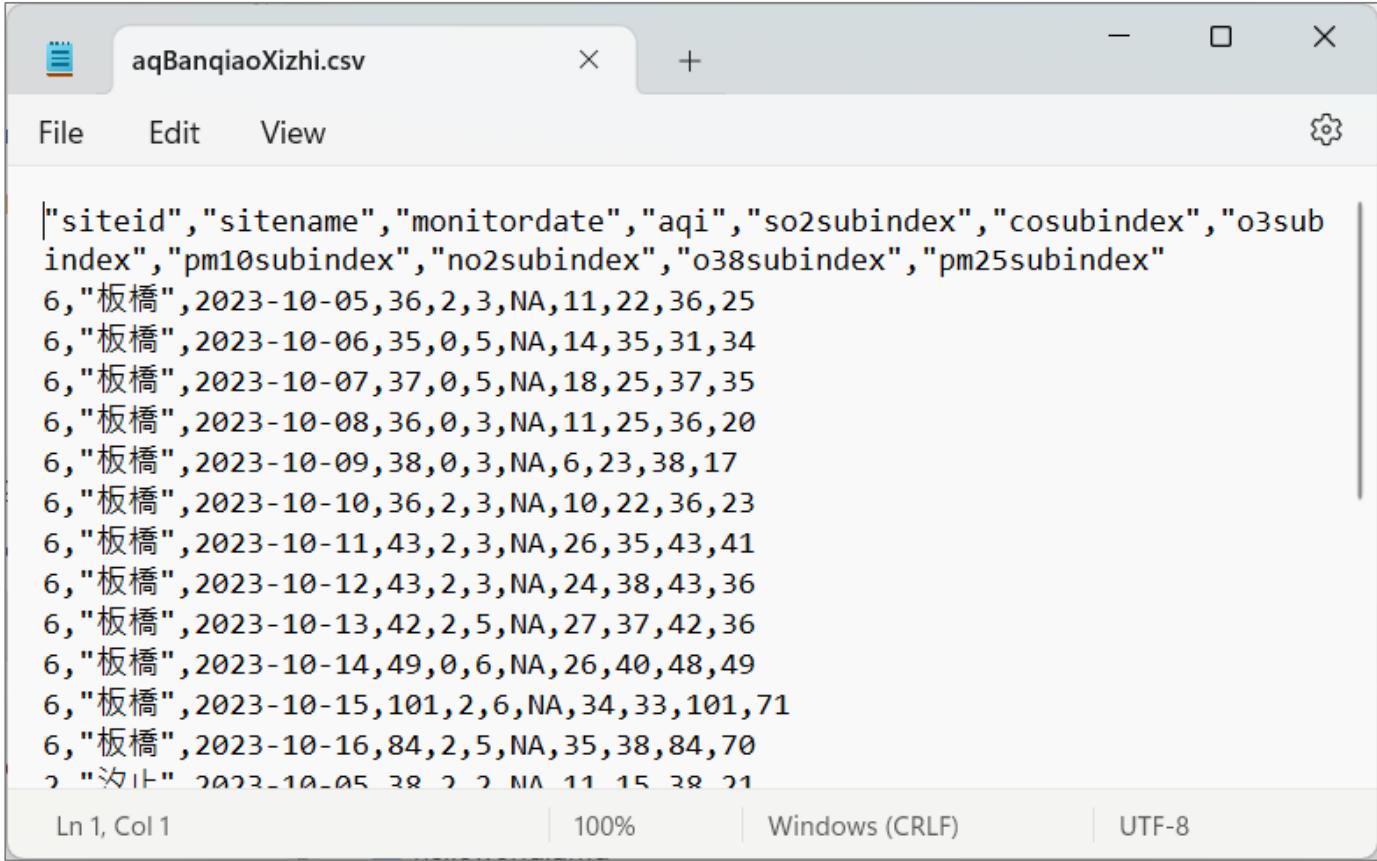
rbind: 列合併

```
aqBanqiaoXizhi <- rbind(aqBanqiao, aqXizhi)

# 汇出結果記事本OK, Excel亂碼
write.table(x = aqBanqiaoXizhi,
             file = "aqBanqiaoXizhi.csv",
             sep = ",",
             row.names = FALSE)
```

write.table : 輸出檔案

記事本開啟CSV顯示正常



The screenshot shows a Windows Notepad window with the title "aqBanqiaoXizhi.csv". The content of the file is a CSV dataset with the following structure:

	siteid	sitename	monitor date	aqi	so2subindex	cosubindex	o3subindex	pm10subindex	no2subindex	o38subindex	pm25subindex
6	"板橋"		2023-10-05	36	2	3	NA	11	22	36	25
6	"板橋"		2023-10-06	35	0	5	NA	14	35	31	34
6	"板橋"		2023-10-07	37	0	5	NA	18	25	37	35
6	"板橋"		2023-10-08	36	0	3	NA	11	25	36	20
6	"板橋"		2023-10-09	38	0	3	NA	6	23	38	17
6	"板橋"		2023-10-10	36	2	3	NA	10	22	36	23
6	"板橋"		2023-10-11	43	2	3	NA	26	35	43	41
6	"板橋"		2023-10-12	43	2	3	NA	24	38	43	36
6	"板橋"		2023-10-13	42	2	5	NA	27	37	42	36
6	"板橋"		2023-10-14	49	0	6	NA	26	40	48	49
6	"板橋"		2023-10-15	101	2	6	NA	34	33	101	71
6	"板橋"		2023-10-16	84	2	5	NA	35	38	84	70
2	"汐止"		2023-10-05	38	2	2	NA	11	15	38	21

Excel開啟CSV會有亂碼!



	A	B	C	D	E	F	G	H	I	J	K
1	siteid	sitename	monitordate	aqi	so2subindex	cosubindex	o3subindex	pm10subindex	no2subindex	o38subindex	pm25subindex
2	6	?蹕?	2023/10/5	36	2	3	NA		11	22	36
3	6	?蹕?	2023/10/6	35	0	5	NA		14	35	31
4	6	?蹕?	2023/10/7	37	0	5	NA		18	25	37
5	6	?蹕?	2023/10/8	36	0	3	NA		11	25	36
6	6	?蹕?	2023/10/9	38	0	3	NA		6	23	38
7	6	?蹕?	2023/10/10	36	2	3	NA		10	22	36
8	6	?蹕?	2023/10/11	43	2	3	NA		26	35	43
9	6	?蹕?	2023/10/12	43	2	3	NA		24	38	43
10	6	?蹕?	2023/10/13	42	2	5	NA		27	37	42
11	6	?蹕?	2023/10/14	49	0	6	NA		26	40	48
12	6	?蹕?	2023/10/15	101	2	6	NA		34	33	101
13	6	?蹕?	2023/10/16	84	2	5	NA		35	38	84
14	2	癩 迫	2023/10/5	38	2	2	NA		11	15	38
15	2	癩 迫	2023/10/6	35	2	2	NA		12	35	31

Excel匯入CSV檔案正常顯示方法

- Excel \ 資料 \ 從文字/CSV



The screenshot shows the 'From Text/CSV' import dialog box for the file 'aqBanqiaoXizhi.csv'. The '檔案原點' (File Origin) is set to '65001: Unicode (UTF-8)'. The '分隔符號' (Delimiter) is set to '逗號' (Comma). The '資料類型偵測' (Data Type Detection) is set to '依據前 200 個列' (Based on first 200 columns). The preview area shows the first four rows of the CSV file:

siteid	sitename	monitordate	aqi	so2subindex	cosubindex	o3subindex	pm10subindex	no2subindex	o38subindex
6	板橋	2023/10/5	36	2	3	NA		11	22
6	板橋	2023/10/6	35	0	5	NA		14	35
6	板橋	2023/10/7	37	0	5	NA		18	25
6	板橋	2023/10/8	36	0	3	NA		11	25

RData 資料物件

實作
練習

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`

匯入sas7bdat

讀取 sas7bdat 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](#)

sas7bdat – 範例

```
> # 讀取 SAS 檔案 ----  
> library(sas7bdat)  
>  
> # h_nhi_ipdte103.sas7bdat 103年模擬全民健保處方及治療明細檔_西醫住院檔  
> # 下載 https://github.com/rwepa/DataDemo/blob/master/h\_nhi\_ipdte103.sas7bdat  
>  
> system.time(dd2014 <- read.sas7bdat("h_nhi_ipdte103.sas7bdat"))
```

使用者 系統 流逝
46.44 0.05 46.73

```
> head(dd2014)
```

	ID	PRSN_ID	HOSP_ID	FEE_YM	APPL_TYPE	APPL_DATE	CASE_TYPE	SEQ_NO
1)])})#*+; [*<	%%%**#==_	~~ 000070353	201405		1 00006337		1 6368
2)])})+-^\$]\$[/	~%~@&[>#*^#_	000162274	201403		1 00034650		1 1081
3)])})+/_~^(&_	*]/*>=)&}:+}~	000181716	201405		2 00007779		5 687
4)])})~\$<_&#>]])]:+\$/_\$};:	000185617	201412		1 00056426		1 4466
5)])})<@%/-@/\`	+%+[-"-*%]<<	000069997	201412		1 00056426		1 304
6)])})!/?>--##	(%(("#])];=\`@	000178535	201412		1 00015473		1 2012

R demo
使用 haven 套件

匯入SPSS

SPSS

- 汇入 SPSS {foreign}
- foreign 套件可读取以下档案格式:
 - Minitab
 - S
 - SAS
 - **SPSS**
 - Stata
 - Systat
 - Weka
 - dBase

R demo

R 連結資料庫

R 連結資料庫

- RMySQL 套件編譯與建立連結 MySQL
 - <http://rwepa.blogspot.com/2013/01/windows-rmysql.html>
- RODBC 與 SQL Server 資料匯入與寫入
 - <http://rwepa.blogspot.com/2013/08/rodbc-sql-server.html>
- Oracle
 - ROracle: OCI Based Oracle Database Interface for R
- PostgreSQL
 - RPostgreSQL: R Interface to the 'PostgreSQL' Database System
- SQLite
 - RSQLite: 'SQLite' Interface for R

5.函數範例

函數

- 函數是R的物件之一，用以輸入參數，並回傳計算結果作為輸出值。
- 函數包括二大類型：
 - 內建函數：例: plot, pairs, install.packages, library, sessionInfo
 - 自訂函數：使用 function()為架構
- 內建函數的輔助說明使用問號，例: ?plot
- 自訂函數 function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



實作
練習

自訂函數

- 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

6.補充篇:Rcmdr

- 6.1 Rcmdr簡介
- 6.2 Rcmdr安裝
- 6.3 Rcmdr環境
- 6.4 Rcmdr範例-marketing.csv(1.資料匯入, 2.處理與分析, 3.報表匯出)

6.1 Rcmdr簡介

Rcmdr

- John Fox 教授於2003年在加拿大麥克馬斯特大學（McMaster University）社會科學學院發表Rcmdr套件，Rcmdr全名是 R Commander。
- Rcmdr套件基於免費 tcltk 框架並使用 R 語言撰寫的視窗操作統計分析套件。tcltk 表示 Tcl (Tool Command Language) 和 Tk (Graphical User Interface Toolkit) 統稱。
- Rcmdr套件可以透過視窗滑鼠與鍵盤選取完成統計分析功能。
- Rcmdr支援資料摘要分析、T檢定、變異數分析、迴歸、廣義線性模型、主成分分析與集群分析等統計模型。
- Rcmdr支援增益集的使用，例：存活分析等。

Rcmdr 增益集：<https://github.com/rwepa/teaching-Rcmdr/blob/main/README.md#rcmdrplugin>

Rcmdr 網頁

- 在 Google 首頁輸入「r cran rcmdr」一般第一個結果為連結至官方網頁。
- <https://cran.r-project.org/web/packages/Rcmdr/index.html>

Rcmdr: R Commander

A platform-independent basic-statistics GUI (graphical user interface) for R, based on the `tcltk` package.

Version: 2.9-2

- Author: John Fox
- URL, Citation, Vignettes

6.2 Rcmdr安裝

Rcmdr安裝

- 步驟1 安裝 R
 - R 網站 \ <https://www.r-project.org/> \ Download \ CRAN
 - <https://cloud.r-project.org/bin/windows/base/>
- 步驟2 安裝Rtools for Windows (macOS, Linux 不用安裝Rtools)
 - Rtools 軟體可用於編譯R或套件時使用，一般建議安裝。
 - 預設安裝路徑不可更改。
 - <https://cloud.r-project.org/bin/windows/Rtools/>
- 步驟3 安裝 Rcmdr 套件
- 步驟4 載入程式套件

步驟1 安裝 R

- R 網站 \ <https://www.r-project.org/> \ Download \ CRAN
- 選取第1個 CRAN: <https://cloud.r-project.org/>
- 選取 Download R for Windows: <https://cloud.r-project.org/bin/windows/base/>
- 選取 base \ 下載並安裝R

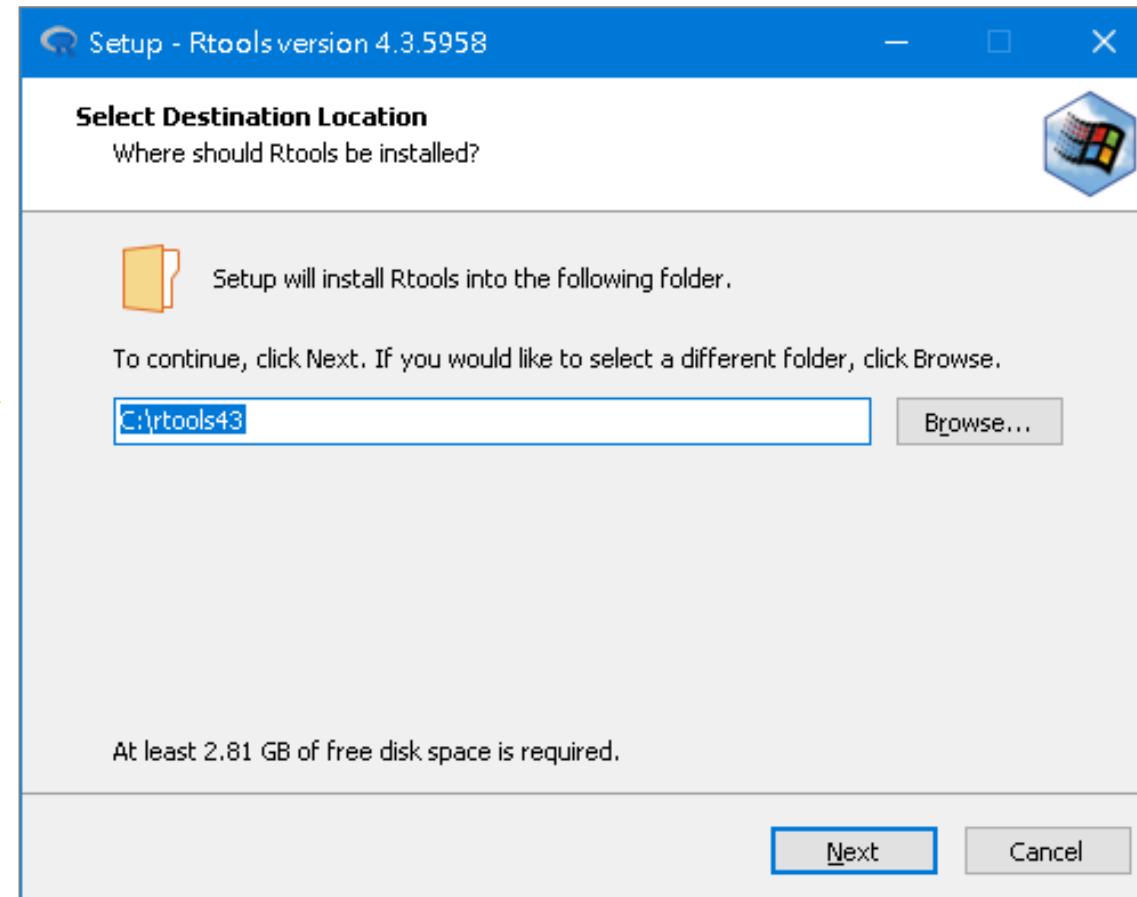


- [Download R for Linux \(Debian, Fedora/Redhat, Ubuntu\)](#)
- [Download R for macOS](#)
- [Download R for Windows](#)

Subdirectories:

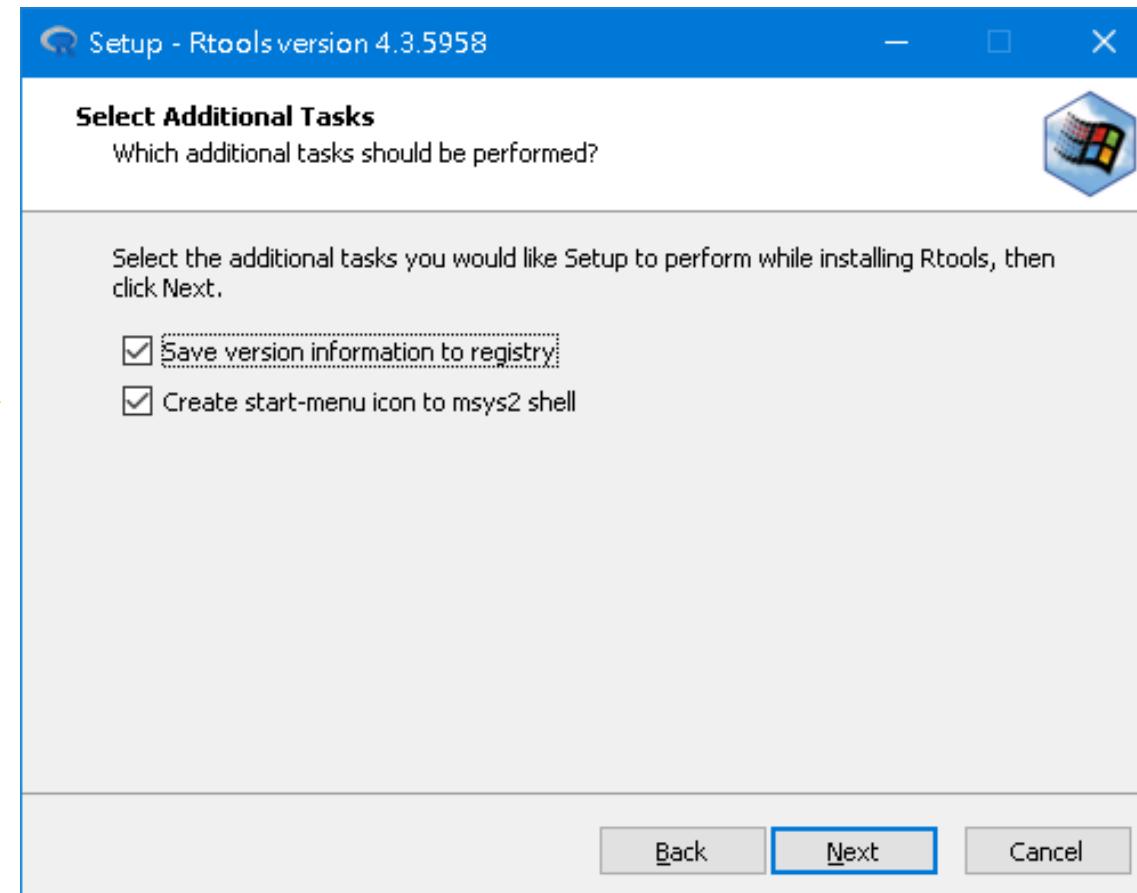
- [base](#)
- [contrib](#)
- [old contrib](#)
- [Rtools](#)

步驟2 Rtools 安裝



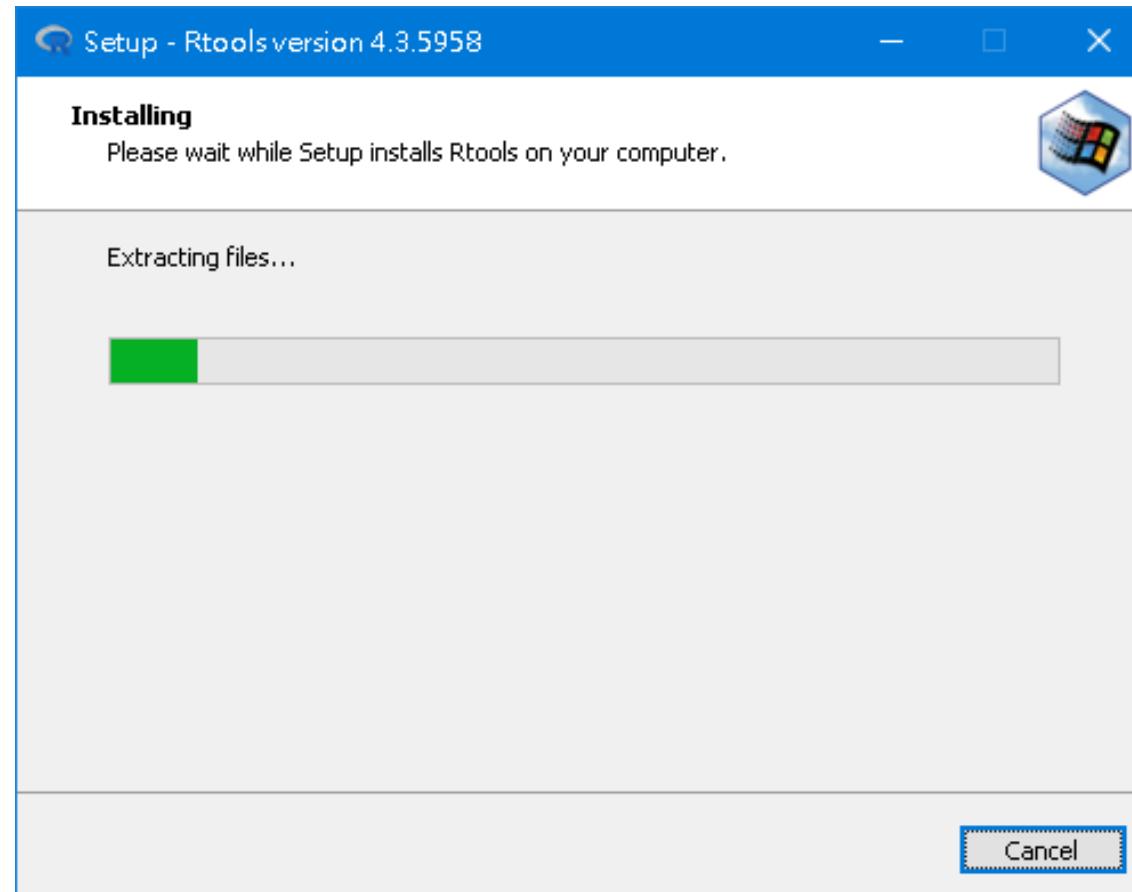
路徑不可修改

Rtools 安裝 (續)

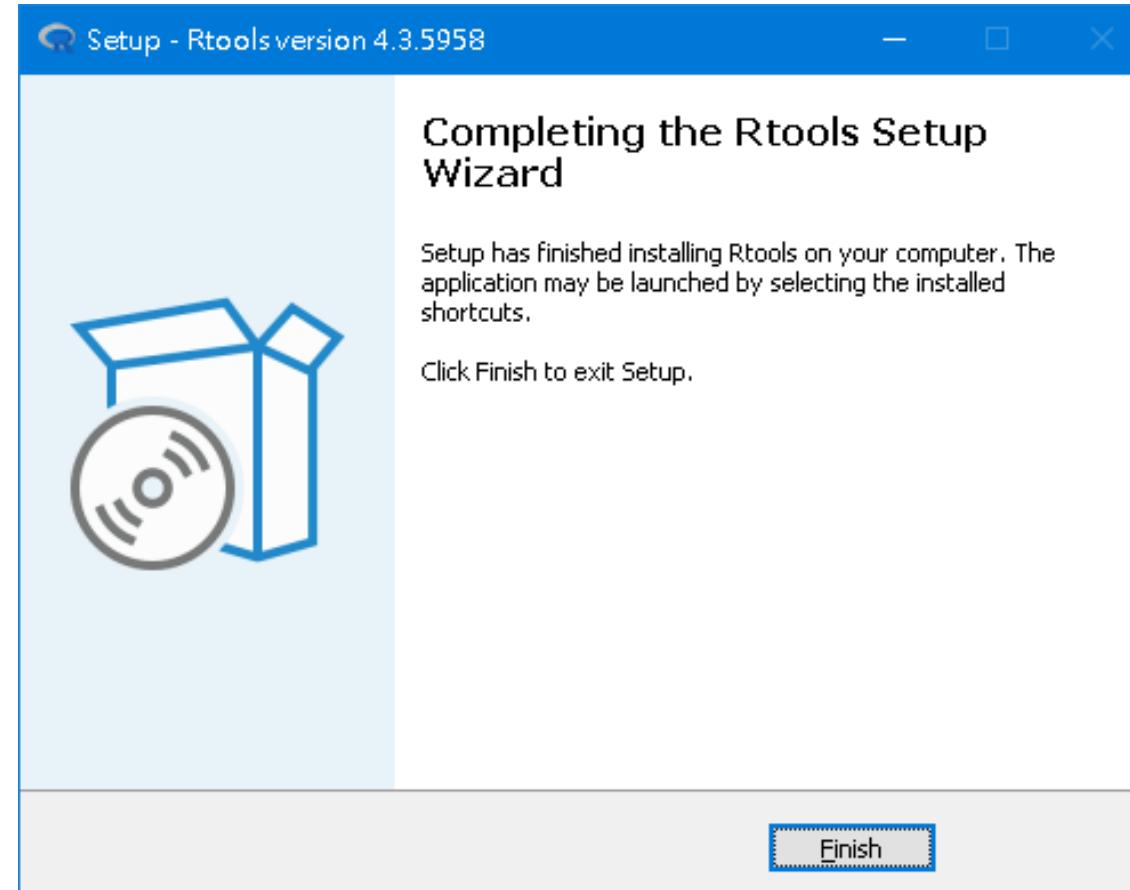


選項皆選取

Rtools 安裝 (續)

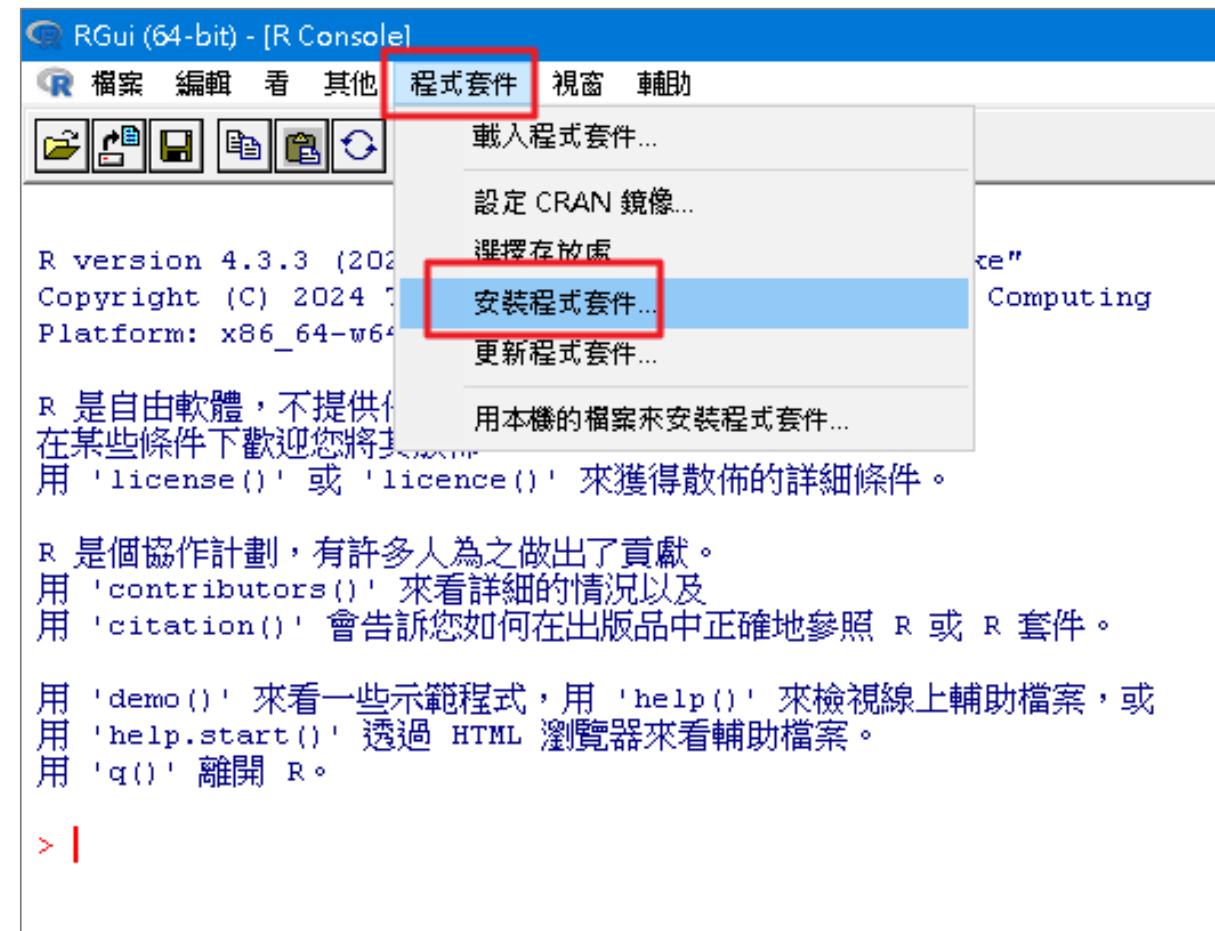


Rtools 安裝 (續)



步驟3 安裝 Rcmdr 套件

- 開啟R
- 程式套件 \ 安裝程式套件



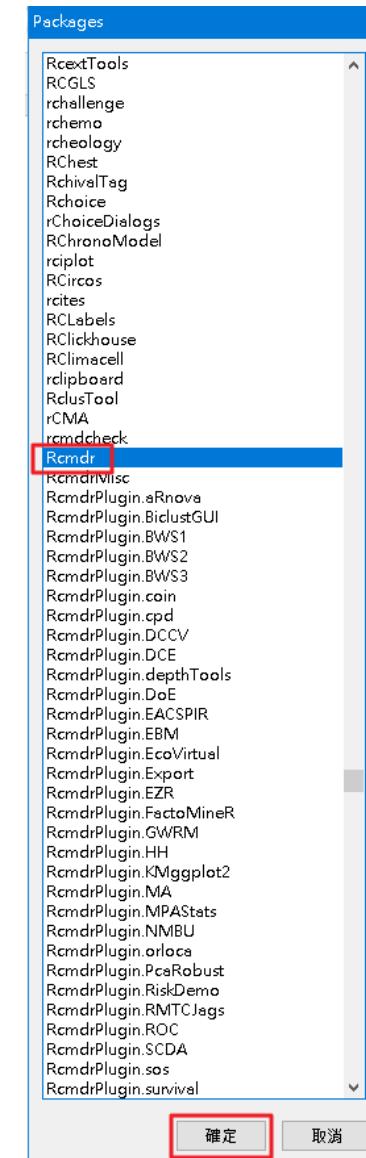
Secure CRAN mirrors

- 選取第1個 CRAN 或其他CRAN \ 確定



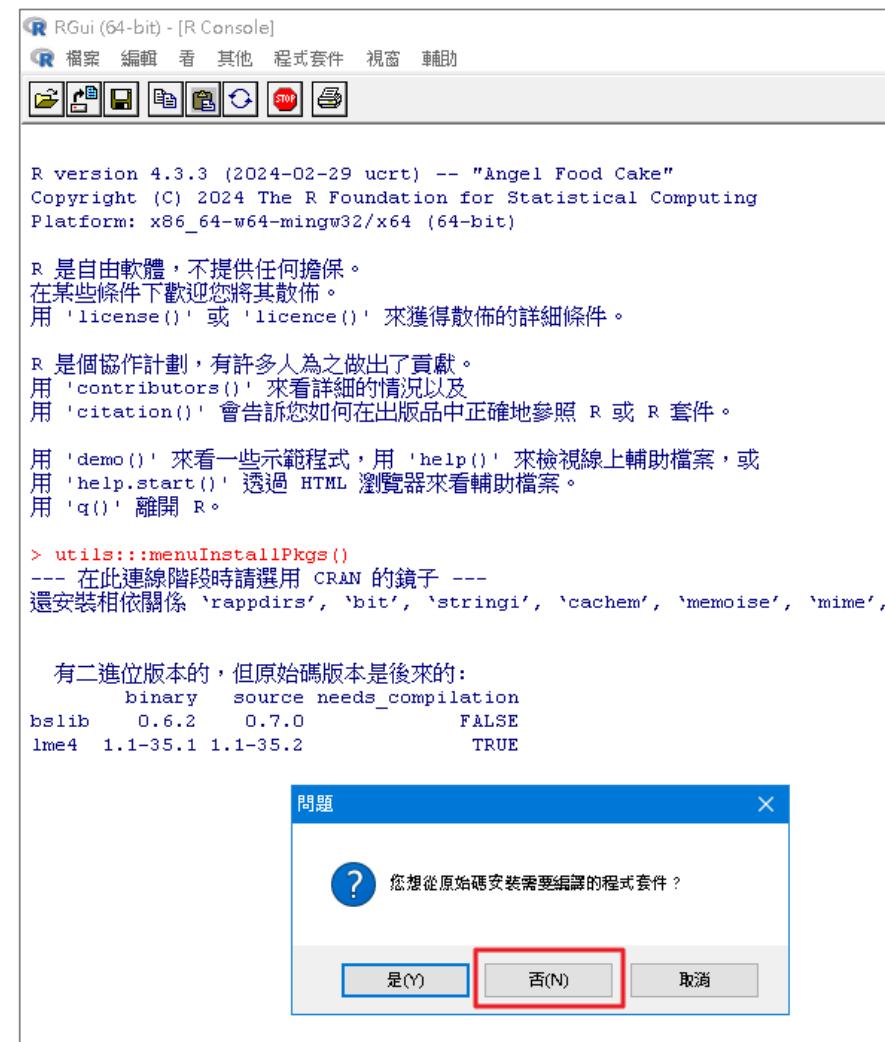
Packages

- 選取 Rcmdr 套件 \ 確定

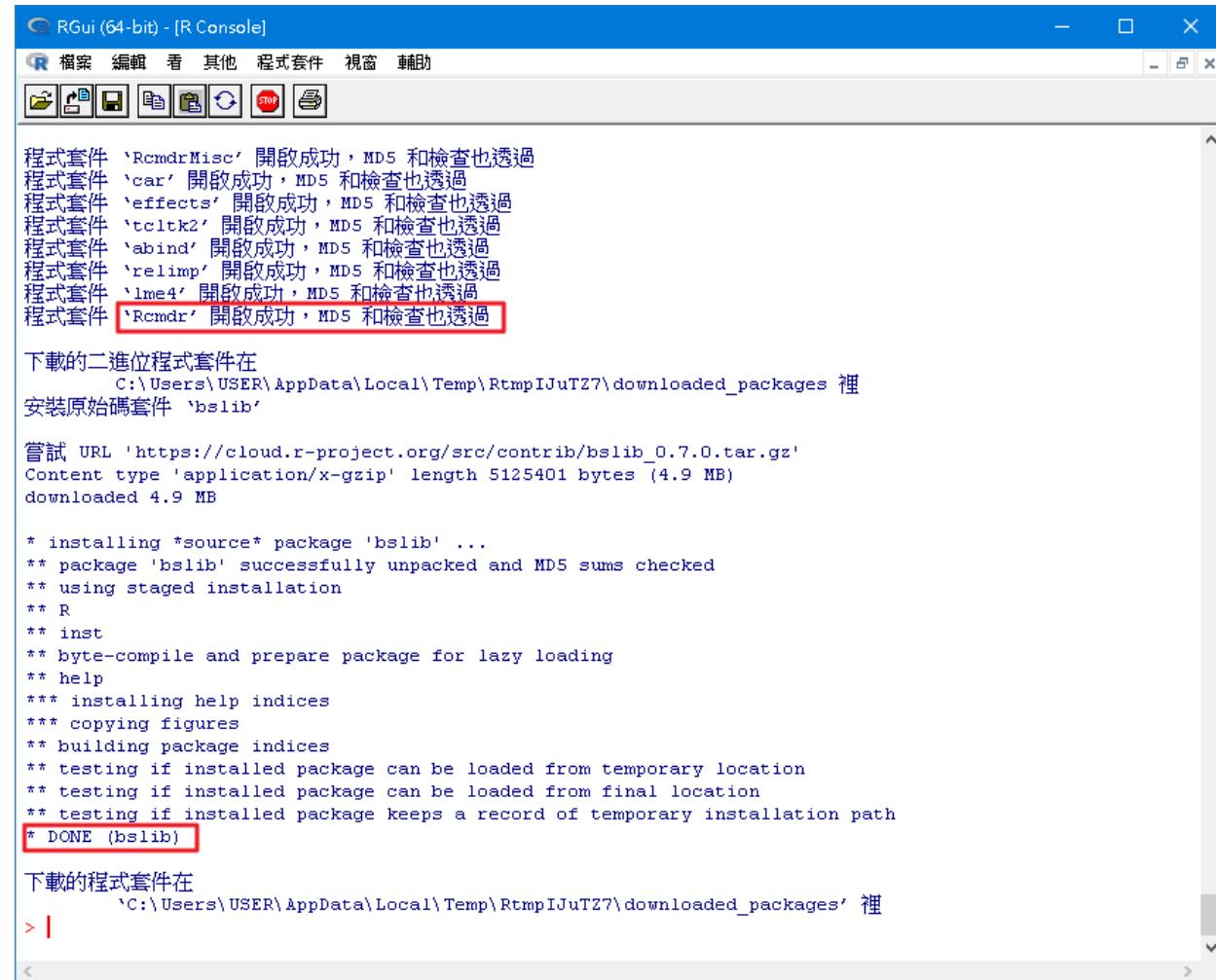


問題

- 您想從原始碼...? \ 否(N)



安裝完成



RGui (64-bit) - [R Console]

檔案 編輯 看 其他 程式套件 視窗 輔助

程式套件 'RcmdrMisc' 開啟成功，MD5 和檢查也透過
程式套件 'car' 開啟成功，MD5 和檢查也透過
程式套件 'effects' 開啟成功，MD5 和檢查也透過
程式套件 'tcltk2' 開啟成功，MD5 和檢查也透過
程式套件 'abind' 開啟成功，MD5 和檢查也透過
程式套件 'relimp' 開啟成功，MD5 和檢查也透過
程式套件 'lme4' 開啟成功，MD5 和檢查也透過
程式套件 'Rcmdr' 開啟成功，MD5 和檢查也透過

下載的二進位程式套件在
C:\Users\USER\AppData\Local\Temp\RtmpIJuT27\downloaded_packages 裡
安裝原始碼套件 'bslib'

嘗試 URL 'https://cloud.r-project.org/src/contrib/bslib_0.7.0.tar.gz'
Content type 'application/x-gzip' length 5125401 bytes (4.9 MB)
downloaded 4.9 MB

* installing *source* package 'bslib' ...
** package 'bslib' successfully unpacked and MD5 sums checked
** using staged installation
** R
** inst
** byte-compile and prepare package for lazy loading
** help
*** installing help indices
*** copying figures
** building package indices
** testing if installed package can be loaded from temporary location
** testing if installed package can be loaded from final location
** testing if installed package keeps a record of temporary installation path
* DONE (bslib)

下載的程式套件在
'C:\Users\USER\AppData\Local\Temp\RtmpIJuT27\downloaded_packages' 裡

> |

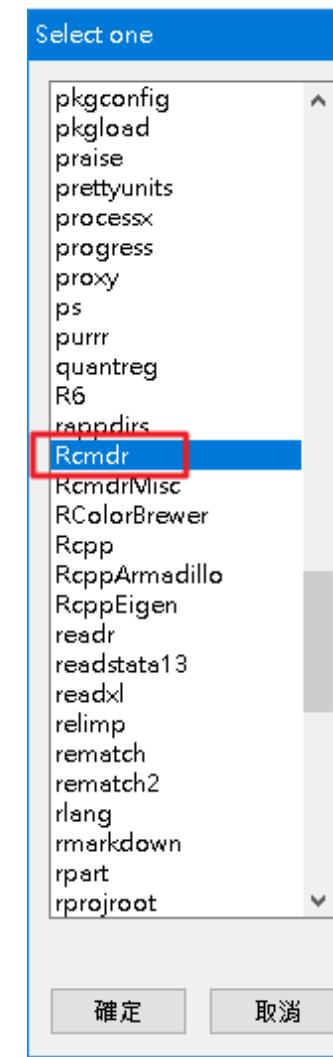
步驟4 載入程式套件

- 程式套件 \ 載入程式套件



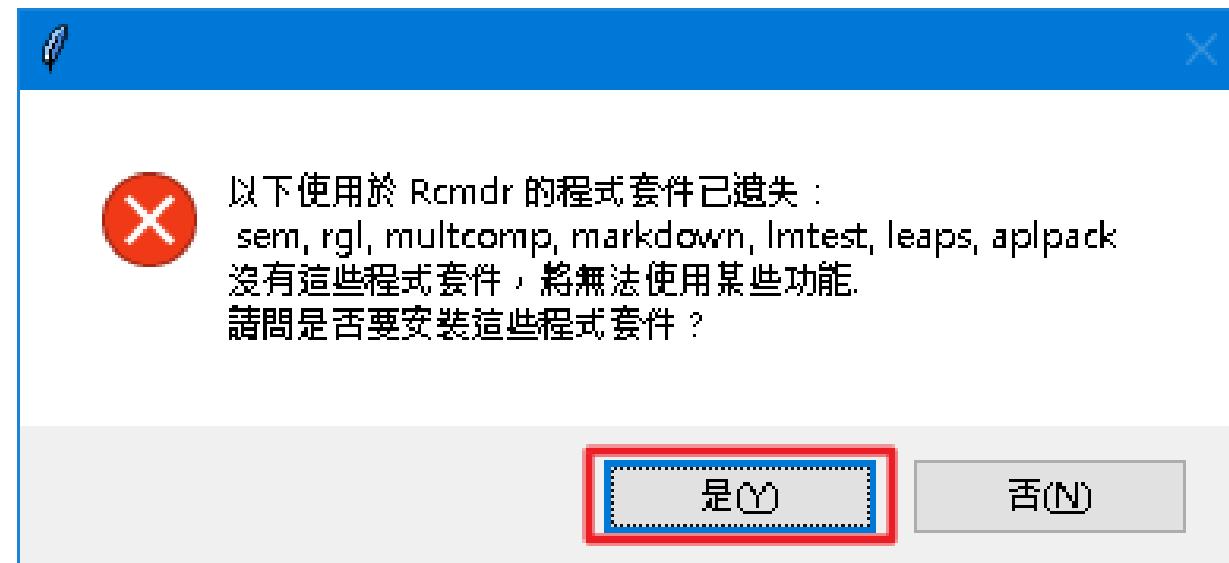
Select one

- 選取 Rcmdr \ 確定



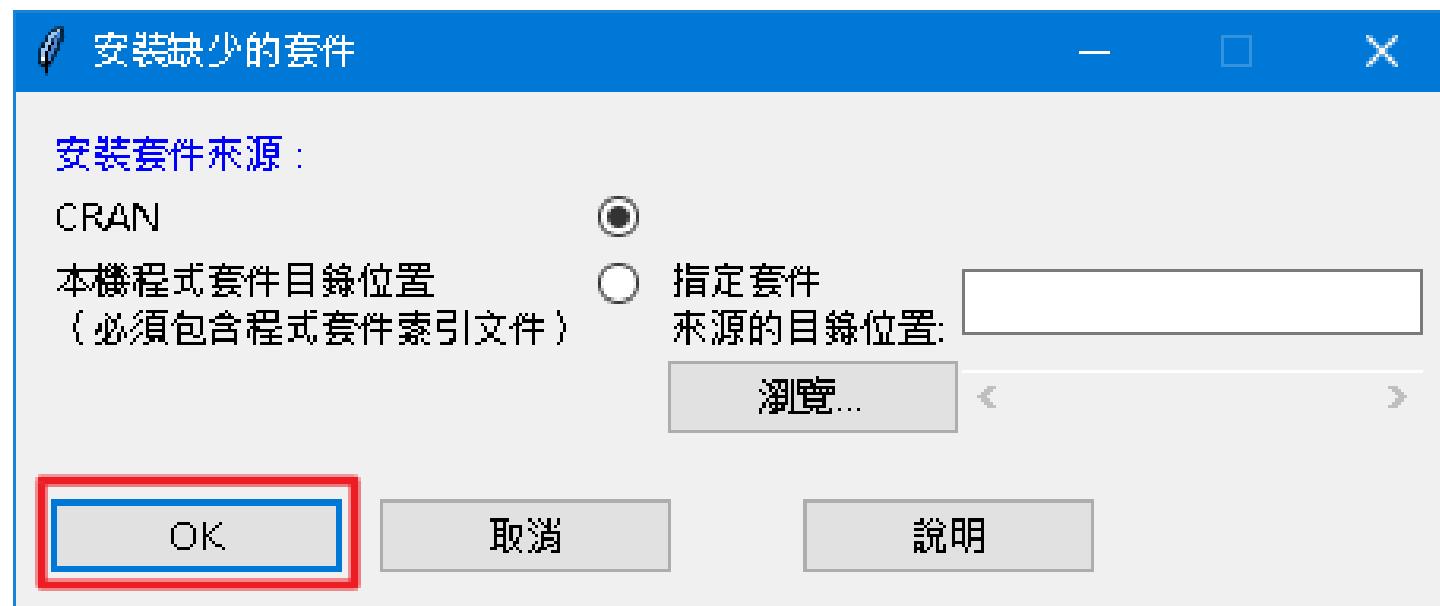
問題

- 請問是否要安裝這些程式套件? \ 是(Y)

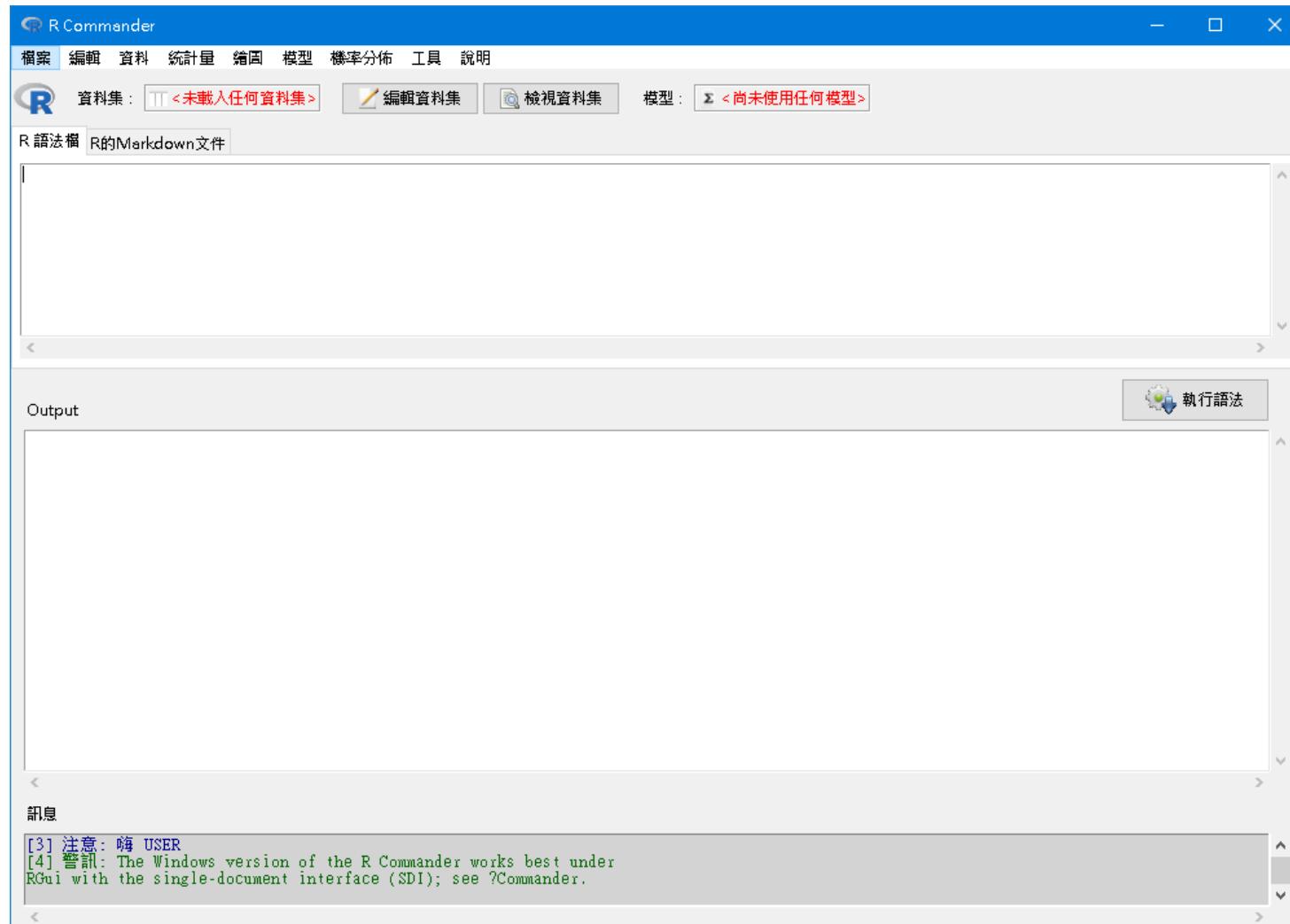


安裝缺少的套件

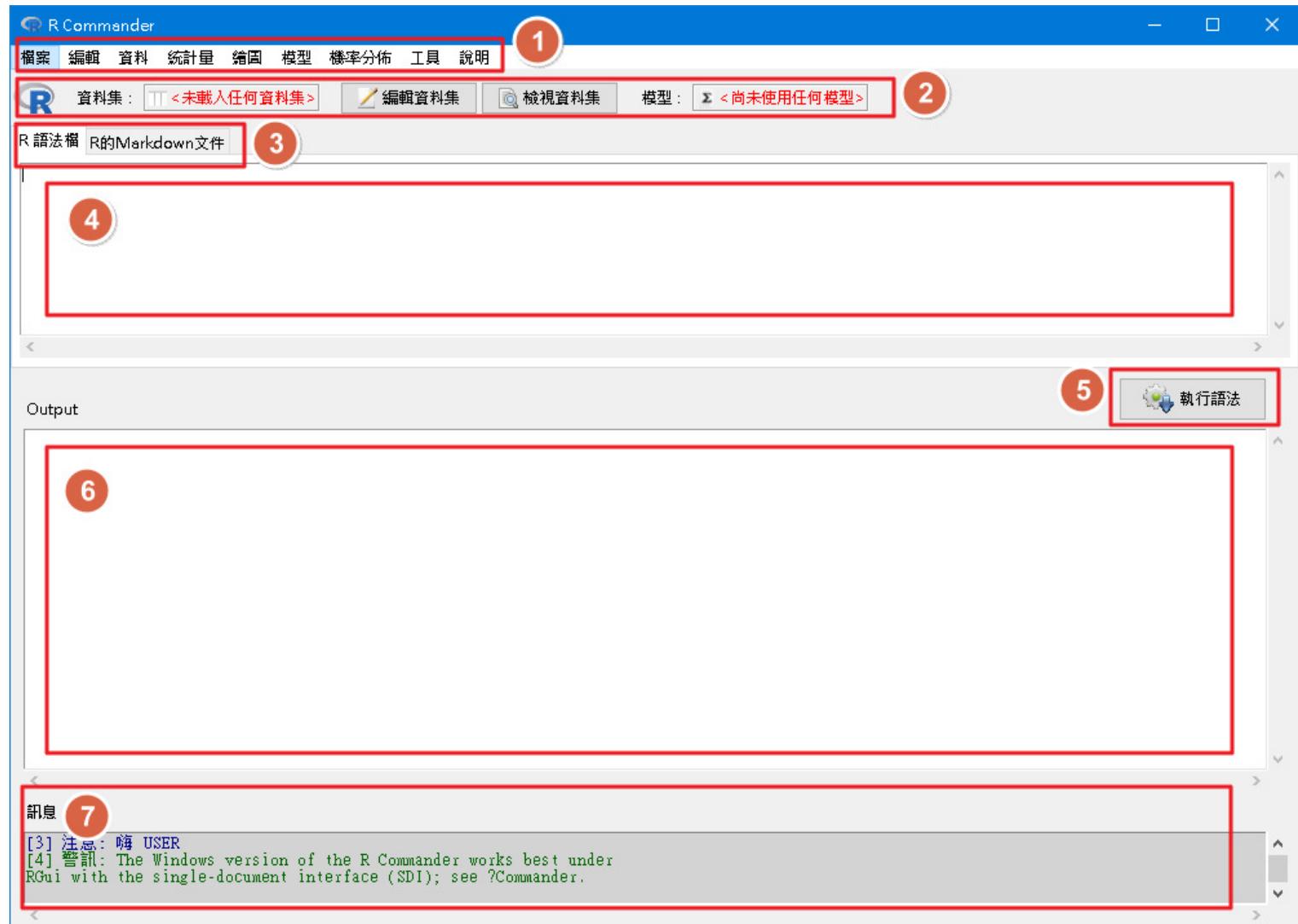
- 按 OK



Rcmdr 啟動視窗



6.3 Rcmdr環境

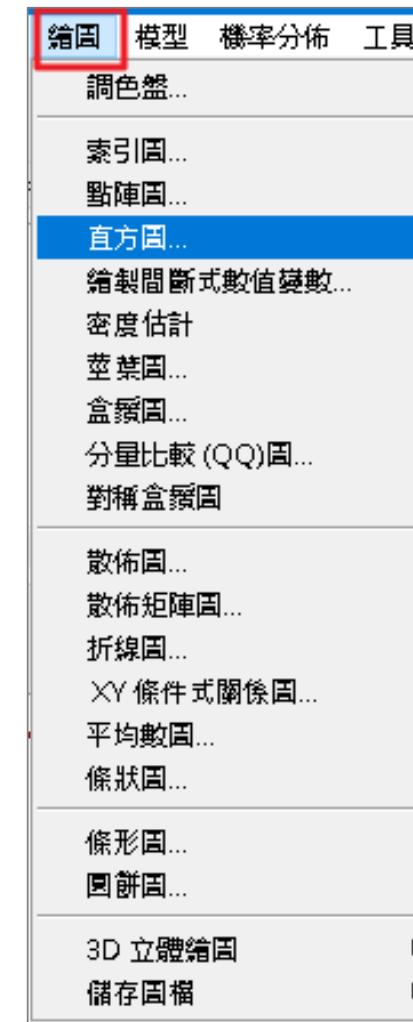
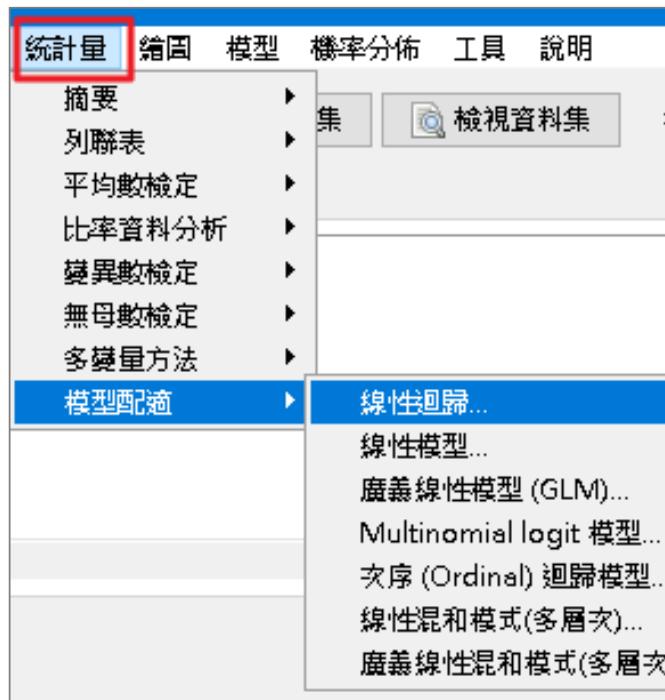


1. 功能表列
2. 工具列
3. R/Markdown語法切換
4. 輸入區
5. 執行語法按鈕
6. 輸出區
7. 訊息區

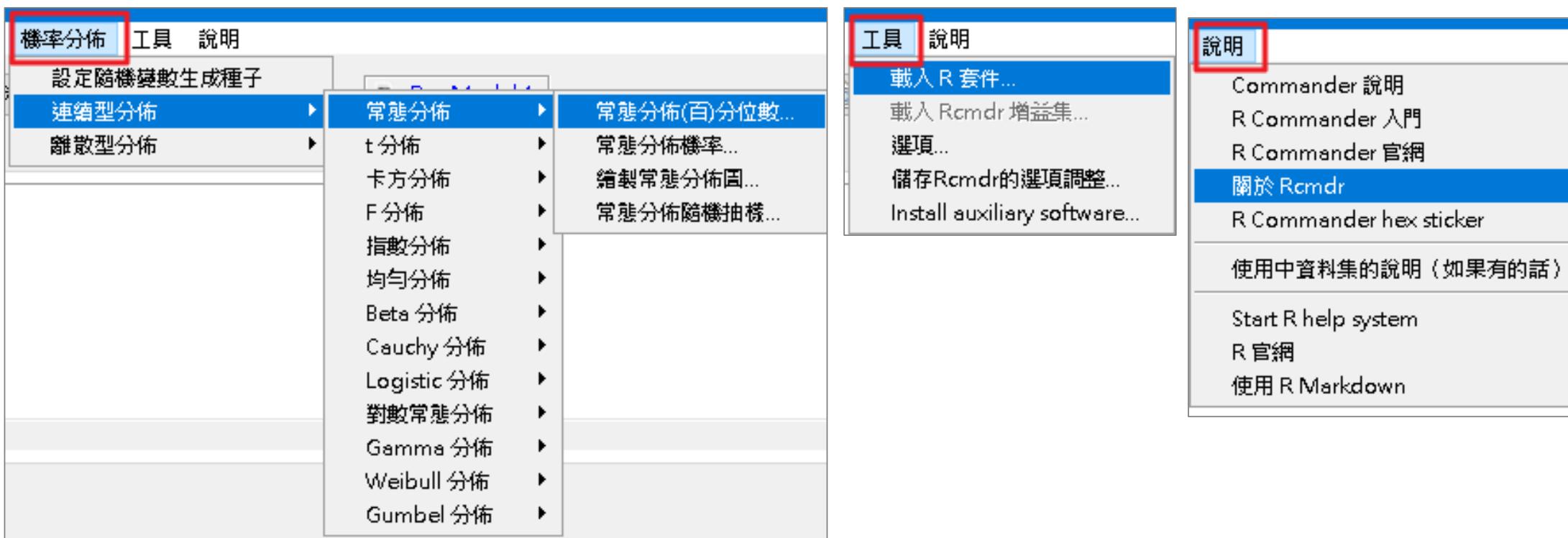
功能表列(檔案,編輯,資料)



功能表列(統計量,繪圖,模型)

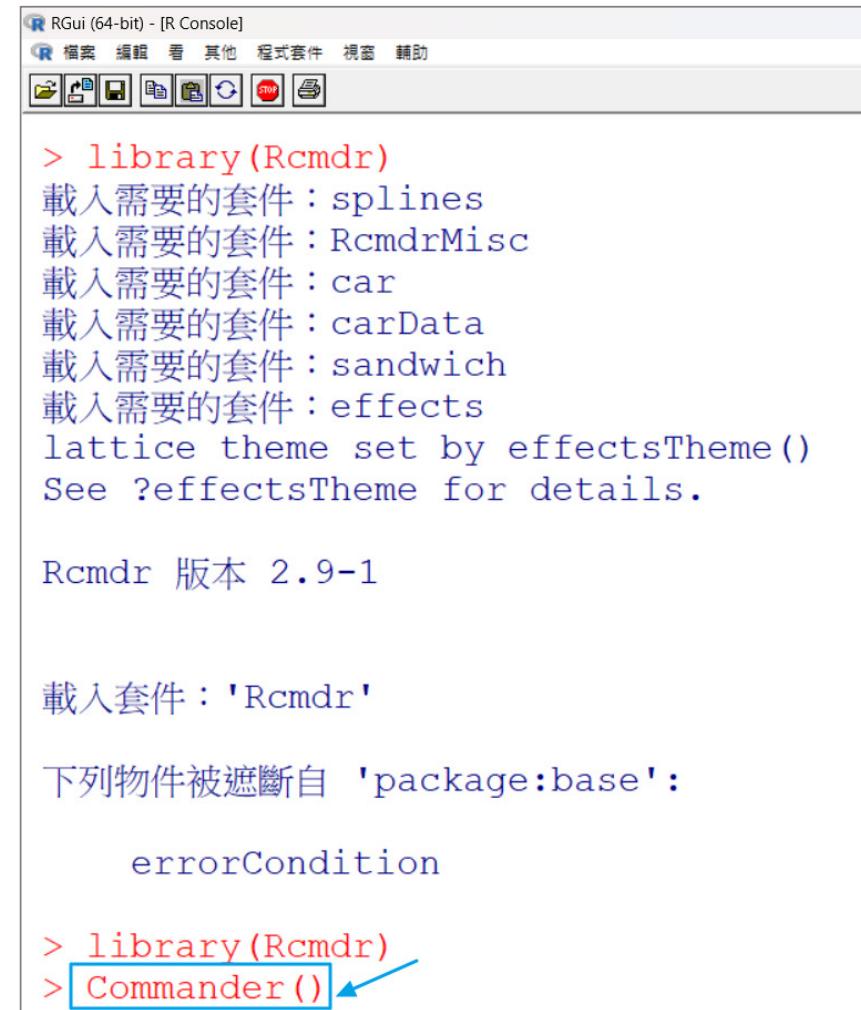


功能表列(機率分配,工具,說明)



再次登入Rcmdr

- Commander()



RGui (64-bit) - [R Console]

檔案 編輯 看 其他 程式套件 視窗 幫助

```
> library(Rcmdr)
載入需要的套件: splines
載入需要的套件: RcmdrMisc
載入需要的套件: car
載入需要的套件: carData
載入需要的套件: sandwich
載入需要的套件: effects
lattice theme set by effectsTheme()
See ?effectsTheme for details.

Rcmdr 版本 2.9-1

載入套件: 'Rcmdr'

下列物件被遮斷自 'package:base':
    errorCondition

> library(Rcmdr)
> Commander() ←
```

6.4 Rcmdr範例-marketing.csv

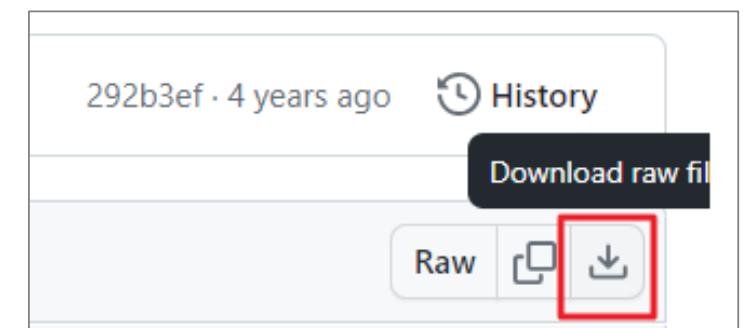
1. 資料匯入
2. 處理與分析
3. 報表匯出

1. 資料匯入

marketing.csv

- 下載 marketing.csv
- <https://github.com/rwepa/DataDemo/blob/master/marketing.csv>
- 按右側 [Download raw file] 並儲存檔案
- 資料有遺漏值 (missing values)

	A	B	C	D
1	youtube	facebook	newspaper	sales
2	276.12	45.36	83.04	26.52
3	53.4		54.12	12.48
4	20.64	55.08	83.16	11.16
5	181.8	49.56	70.2	22.2
6	216.96	12.96	70.08	15.48



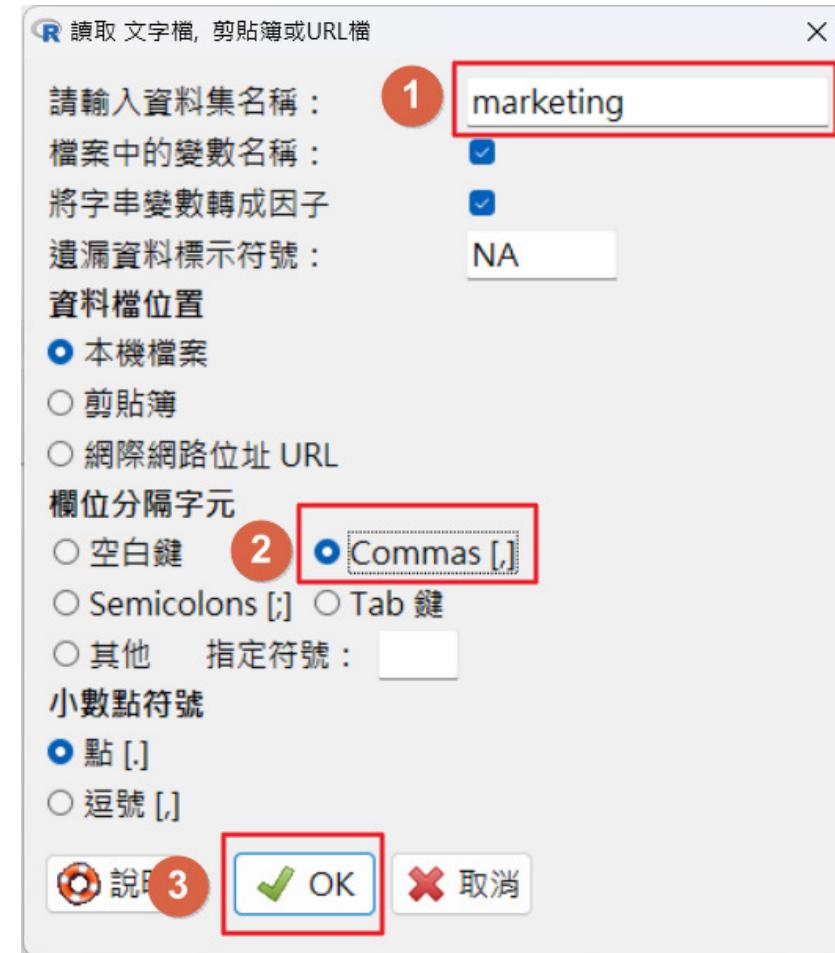
匯入文字檔

- 資料 \ 汇入資料 \ 汇入文字檔...



匯入文字檔(續)

- 資料集名稱: marketing
- 欄位分隔字元: Commas[,]
- 按 [OK]



資料匯入完成

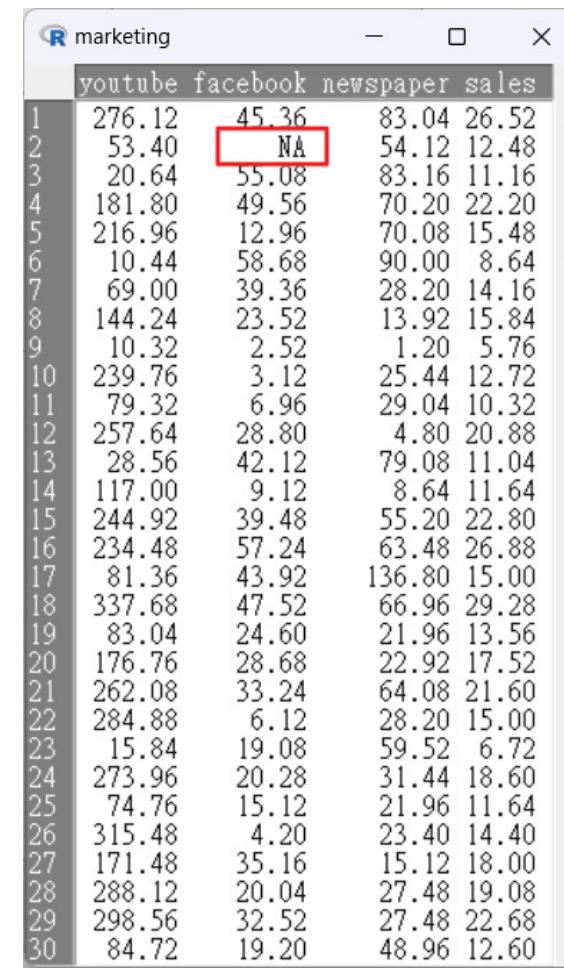
The screenshot shows the R Commander application window. At the top, the menu bar includes 檔案, 編輯, 資料, 統計量, 繪圖, 模型, 機率分佈, 工具, and 說明. Below the menu is a toolbar with icons for 資料集 (selected), 編輯資料集, and 檢視資料集. The status bar indicates '模型 : <尚未使用任何模型>'. The main area displays R code for loading a CSV file:

```
marketing <- read.table("C:/Users/asus/Downloads/marketing.csv", header=TRUE, stringsAsFactors=TRUE, sep=",", na.strings="NA", dec=".," strip.white=TRUE)
```

The code is highlighted with a red box. Below the code editor is an 'Output' panel containing the same R command. A '執行語法' (Execute) button is located next to the output panel. At the bottom, there is a '訊息' (Message) panel with two entries:

- [4] 警訊: The Windows version of the R Commander works best under RGui with the single-document interface (SDI); see ?Commander.
- [5] 注意: 此 marketing 資料集有 200 個列與 4 個欄.

檢視資料集



The screenshot shows an RStudio viewer pane titled "marketing". The data frame has 30 rows and 5 columns. The columns are labeled "youtube", "facebook", "newspaper", "sales", and an unnamed column containing row indices from 1 to 30. The cell at row 2, column 3 ("newspaper") contains the value "NA", which is highlighted with a red rectangular border.

	youtube	facebook	newspaper	sales
1	276.12	45.36	83.04	26.52
2	53.40	NA	54.12	12.48
3	20.64	55.08	83.16	11.16
4	181.80	49.56	70.20	22.20
5	216.96	12.96	70.08	15.48
6	10.44	58.68	90.00	8.64
7	69.00	39.36	28.20	14.16
8	144.24	23.52	13.92	15.84
9	10.32	2.52	1.20	5.76
10	239.76	3.12	25.44	12.72
11	79.32	6.96	29.04	10.32
12	257.64	28.80	4.80	20.88
13	28.56	42.12	79.08	11.04
14	117.00	9.12	8.64	11.64
15	244.92	39.48	55.20	22.80
16	234.48	57.24	63.48	26.88
17	81.36	43.92	136.80	15.00
18	337.68	47.52	66.96	29.28
19	83.04	24.60	21.96	13.56
20	176.76	28.68	22.92	17.52
21	262.08	33.24	64.08	21.60
22	284.88	6.12	28.20	15.00
23	15.84	19.08	59.52	6.72
24	273.96	20.28	31.44	18.60
25	74.76	15.12	21.96	11.64
26	315.48	4.20	23.40	14.40
27	171.48	35.16	15.12	18.00
28	288.12	20.04	27.48	19.08
29	298.56	32.52	27.48	22.68
30	84.72	19.20	48.96	12.60

2.處理與分析

摘要

- 統計量 \ 摘要 \ 使用中的資料集



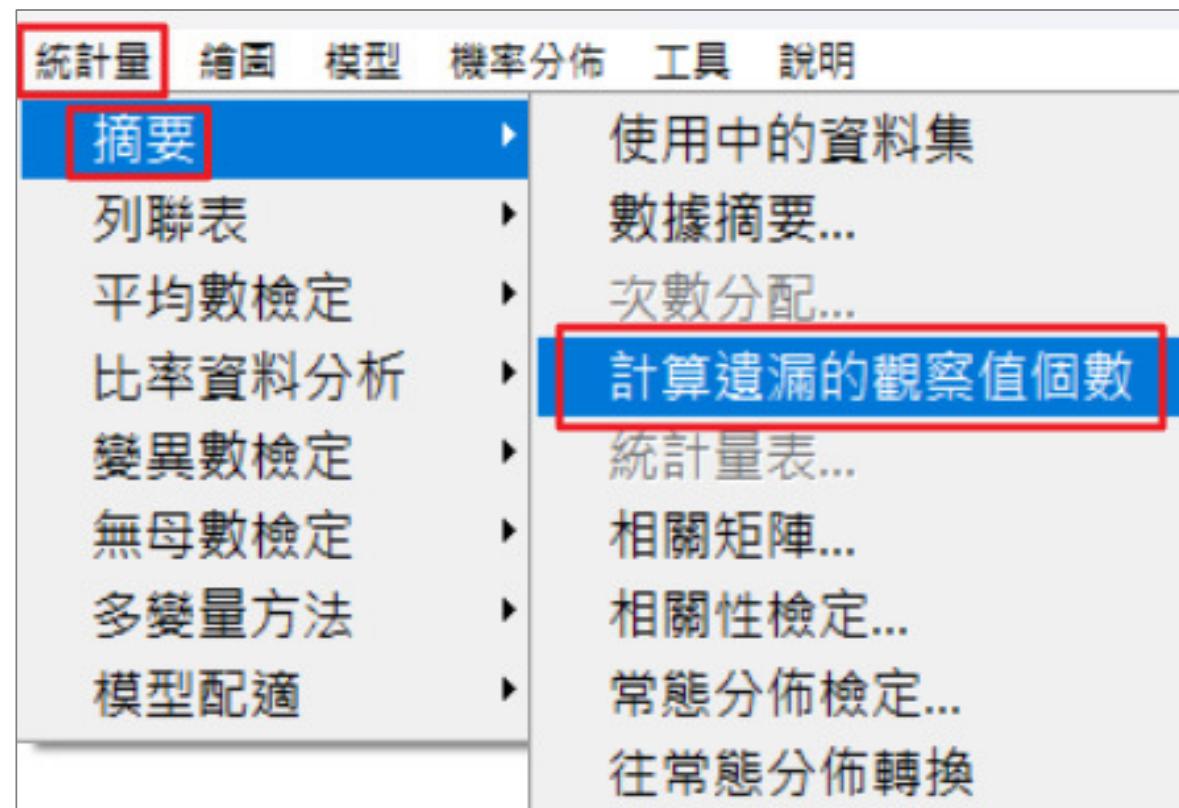
summary 函數

```
> summary(marketing)
```

youtube	facebook	newspaper	sales
Min. : 0.84	Min. : 0.00	Min. : 0.36	Min. : 1.92
1st Qu.: 89.25	1st Qu.:11.94	1st Qu.: 15.30	1st Qu.:12.45
Median :179.70	Median :27.00	Median : 30.90	Median :15.48
Mean :176.45	Mean :27.82	Mean : 36.66	Mean :16.83
3rd Qu.:262.59	3rd Qu.:43.68	3rd Qu.: 54.12	3rd Qu.:20.88
Max. :355.68	Max. :59.52	Max. :136.80	Max. :32.40
	NA's :1		

遺漏值

- 統計量 \ 摘要 \ 計算遺漏的觀察值個數



is.na(x) 函數

- is.na(x) 函數 – 判斷是否為 NA

```
> sapply(marketing, function(x)(sum(is.na(x)))) # NA counts
  youtube facebook newspaper      sales
        0         1         0         0
```

相關性矩陣

- 統計量 \ 摘要 \ 相關矩陣



相關係數

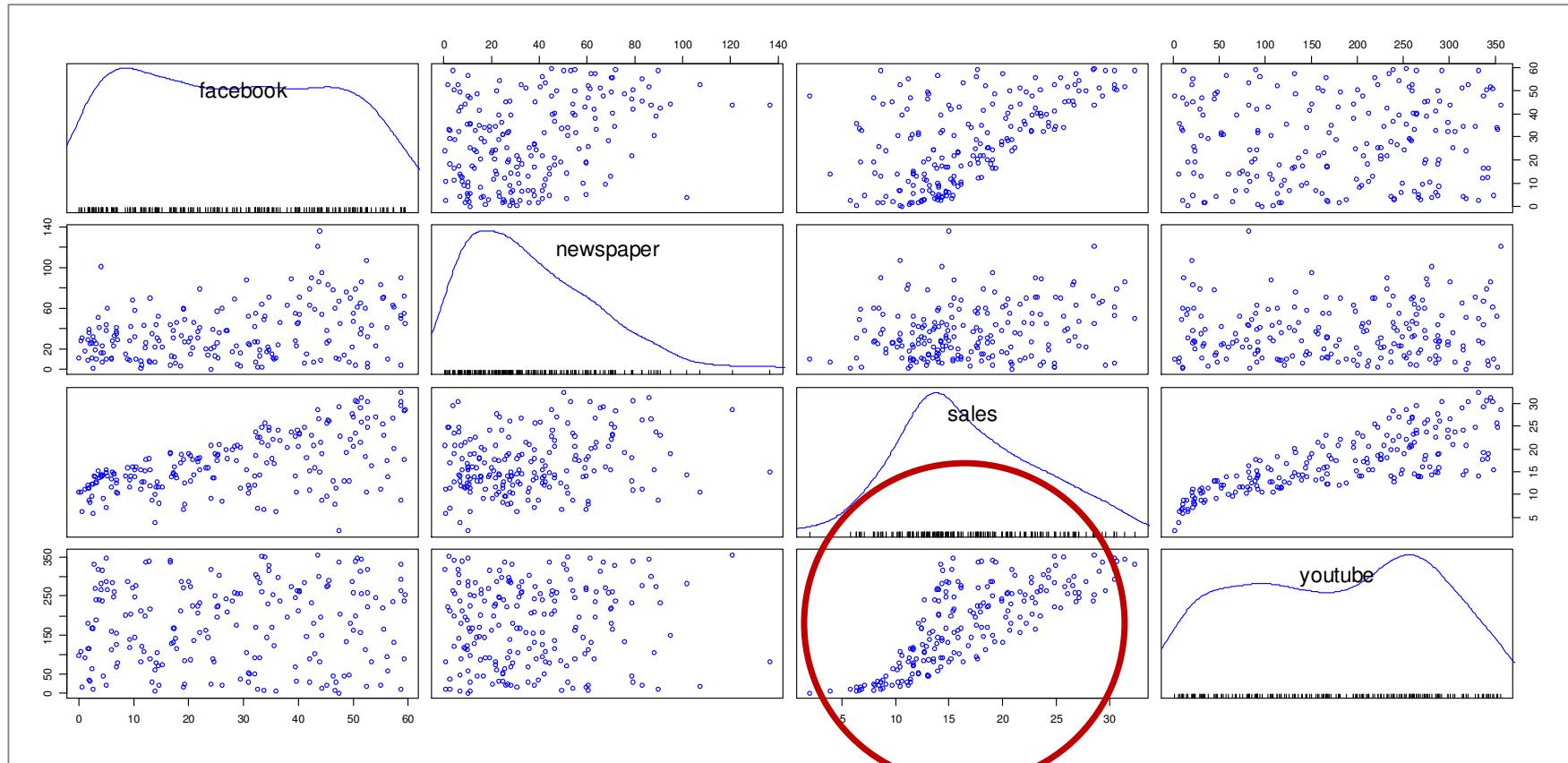
- $-1 \leq \text{相關係數} \leq 1$
- 相關係數為0，表示沒有線性相關，可能有非線性相關。

```
> cor(marketing[,c("facebook", "newspaper", "sales", "youtube")], use="complete")
      facebook newspaper sales youtube
facebook 1.0000000 0.35189404 0.5824356 0.06172810
newspaper 0.3518940 1.00000000 0.2311829 0.06096518
sales     0.5824356 0.23118286 1.0000000 0.78180655
youtube   0.0617281 0.06096518 0.7818065 1.00000000
```

3. 報表匯出

資料視覺化

- 繪圖 \ 散佈圖矩陣



產生報表

- [R 的Markdown文件] \ 產生報表

R語法檔 R的Markdown文件

```
cor(marketing[,c("facebook","newspaper","sales","youtube")], use="complete")
```
散佈圖矩陣: ~facebook+newspaper+sales+youtube
```{r}
scatterplotMatrix(~facebook+newspaper+sales+youtube, regLine=FALSE, smooth=FALSE, diagonal=list(method="density"),
  data=marketing)
```

```

Output  產生報表



# html 輸出

Replace with Main Title

ASUS  
2024-04-02

```
> marketing <- read.table("C:/Users/asus/Downloads/marketing.csv", header=TRUE, stringsAsFactors=TRUE, sep=",",
+ na.strings="NA", dec=".," , strip.white=TRUE)
```

資料集摘要: marketing

```
> summary(marketing)
```

|          | youtube | facebook      | newspaper     | sales         |
|----------|---------|---------------|---------------|---------------|
| Min. :   | 0.84    | Min. : 0.00   | Min. : 0.35   | Min. : 3.92   |
| 1st Qu.: | 89.25   | 1st Qu.:11.04 | 1st Qu.:15.30 | 1st Qu.:12.45 |
| Median:  | 179.70  | Median:27.00  | Median:39.90  | Median:15.48  |
| Mean :   | 176.45  | Mean :27.82   | Mean :38.65   | Mean :16.83   |
| 3rd Qu.: | 262.54  | 3rd Qu.:41.68 | 3rd Qu.:54.12 | 3rd Qu.:20.88 |
| Max. :   | 355.68  | Max. :59.52   | Max. :135.80  | Max. :32.49   |
| Na's :   | 1       |               |               |               |

計算遺漏的觀察值個數: marketing

```
> sum(is.na(x)) # NA counts
```

|         | youtube | facebook | newspaper | sales |
|---------|---------|----------|-----------|-------|
| youtube | 0       | 3        | 0         | 0     |

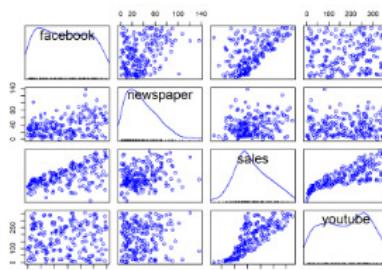
相關性矩陣: marketing

```
> cor(marketing[,c("facebook","newspaper","sales","youtube")], use="complete")
```

|           | facebook   | newspaper  | sales      | youtube    |
|-----------|------------|------------|------------|------------|
| facebook  | 1.00000000 | 0.55104800 | 0.56265556 | 0.04172030 |
| newspaper | 0.55104800 | 1.00000000 | 0.55104800 | 0.51108555 |
| sales     | 0.56265556 | 0.55104800 | 1.00000000 | 0.71108555 |
| youtube   | 0.04172030 | 0.51108555 | 0.71108555 | 1.00000000 |

散佈圖矩阵: ~facebook+newspaper+sales+youtube

```
> scatterplotMatrix(~facebook+newspaper+sales+youtube, regline=FALSE, smooth=FALSE, diagonal=list(method="density"),
+ data=marketing)
```



# docx 輸出

# 謝謝您的聆聽

## Q & A



李明昌

*alan9956@gmail.com*

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