

1. R/RStudio簡介

- 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

大綱

- 1. 2024/09/07 R/RStudio簡介
 - 1.1 R/RStudio下載、使用與套件簡介
 - 1.2 開放資料實務應用
- 2. 2024/09/14 資料物件與迴圈
 - 2.1 資料物件簡介
 - 2.2 流程控制與for迴圈
- 3. 2024/09/21 資料摘要分析
 - 3.1 apply、lapply與樞紐分析
 - 3.2 長寬資料轉換
- 4. 2024/09/28 資料視覺化分析
 - 4.1 使用graphics套件建立統計圖形
 - 4.2 使用ggplot2套件建立視覺化分析
- 5. 2024/10/05 預測模型整合案例
 - 5.1 銷售與人力資源預測模型
 - 5.2 shiny互動式App佈署應用

1.1 R/RStudio下載、使用與套件簡介

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347場，3211小時。
- 連絡資訊：alan9956@gmail.com

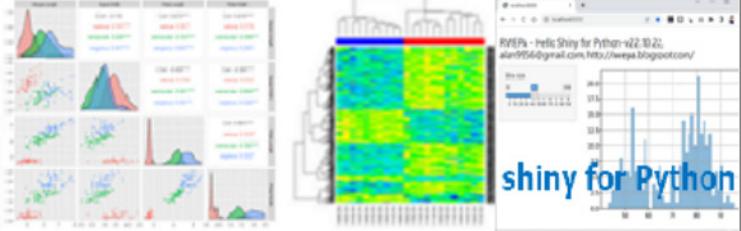


- iPAS 巨量資料分析師 證照推廣
- iPAS 營運智慧分析師 證照推廣

RWEPA



```
library(shiny)
shinyUI(pageWithSidebar(
  headerPanel("shiny"),
  sidebarPanel(c1 <- makeCluster(4),
    map.taiwan <- get_map(location = "Taipei, Taiwan",
      maptype = "satellite"))
  library(foreach)
  foreach(i = 1:4, .combine = cbind) %doSNOW%
    plot(c1[[i]]$x, c1[[i]]$y, col = c1[[i]]$z,
      main = paste("Plot", i))
  ))
  library(ggmap)
  ggmap(map.taiwan)
```



歡迎來到 RWEPA blog, 成立宗旨為提供免費R(Python)等資訊。R(Python)包括大量套件可應用於不同領域, 例如:
2D/3D互動式繪圖, 資料視覺化, 地理視覺化, 資料探勘, 機器學習, 深度學習, 線性與非線性最佳化問題, 時間序列, 空間資料, 財務分析, 多變量分析, 問卷調查, 實驗設計, 統計製程管制, 存活分析, 臨床實驗分析, 社會網絡分析, 生物資訊, 醫學統計, 水文學, 天文學等。

2024年7月11日 星期四

Service : alan9956@gmail.com

Stable Diffusion Model

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, 師資協同教學.

RWEPA | Stable Diffusion Model 本地端安裝-AI生成式繪圖

RWEPA 搜尋此網誌 (例: task)

1 2

@RWEPA

6

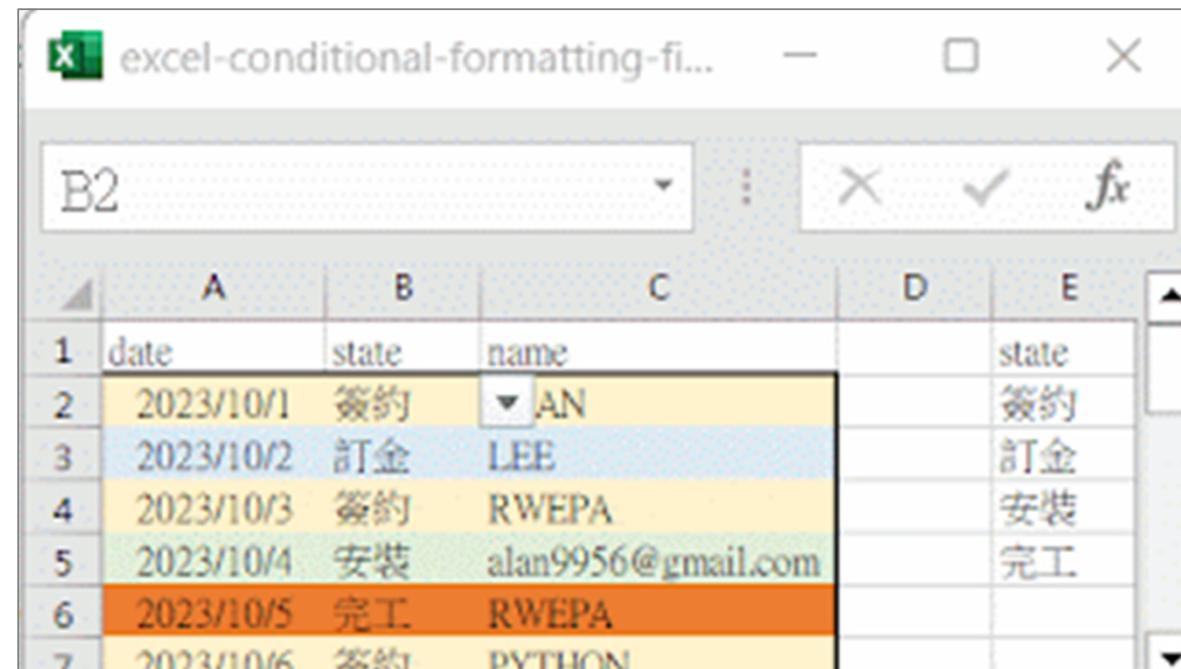
資料分析暨視覺化應用

★★★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>

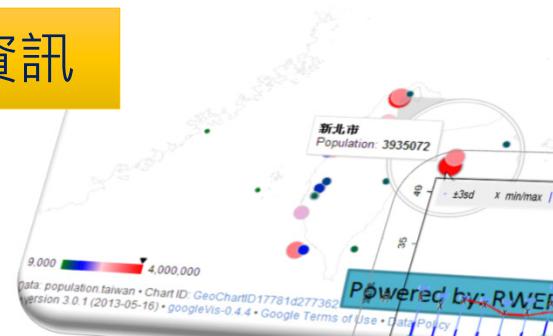


| | A | B | C | D | E |
|---|-----------|-------|--------------------|-------|---|
| 1 | date | state | name | state | |
| 2 | 2023/10/1 | 簽約 | AN | 簽約 | |
| 3 | 2023/10/2 | 訂金 | LEE | 訂金 | |
| 4 | 2023/10/3 | 簽約 | RWEPA | 安裝 | |
| 5 | 2023/10/4 | 安裝 | alan9956@gmail.com | 完工 | |
| 6 | 2023/10/5 | 完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

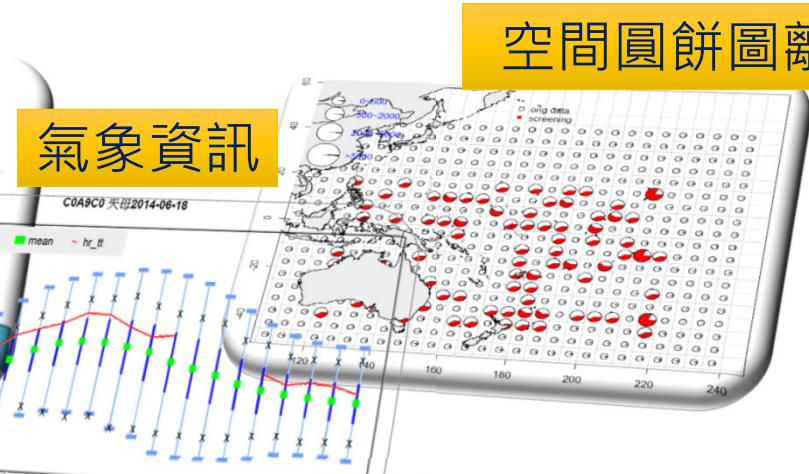
Excel 限制?

R+Shiny, Python+Streamlit 互動式平台

地理資訊



氣象資訊

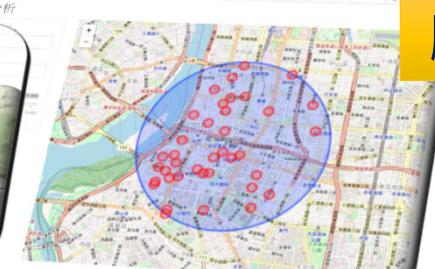


空間圓餅圖離群值分析

保險預測



登山路線視覺化



顧客連結資訊



中央氣象局 1,600萬筆資料(14,328檔案)

網頁呈現

The screenshot displays the '气象资料视觉化暨互动式分析平台_qc_v0.16.5.1' interface. Key features shown include:

- 一般 (General) / 自訂 (Custom):** Buttons for general and custom settings.
- 14,328 個檔案, 18 個資料夾 (14,328 files, 18 folders):** Summary statistics.
- 選取測站 (Select Station):** Set to 466900 淡水.
- 選取月份 (1-12) (Select Month 1-12):** Set to 6.
- 選取小時 (1-24) (Select Hour 1-24):** Set to 1.
- 選取年之範圍 (Select Year Range):** Set from 1998 to 2014.
- 選取QC閾值 (Select QC Threshold):** Set to 5.
- QC_UKref/hr_tt_06mm_01hr_466900.txt:** Scatter plot titled 'QC_UKref/hr_tt_06mm_01hr_466900.txt'. The x-axis is '推估值-UK' (Estimated Value-UK) ranging from 20 to 28. The y-axis is '實際值-TT' (Actual Value-TT) ranging from 20 to 24. A red 'X' marks a data point at approximately (20.5, 23.5). A dashed red line represents the regression fit.
- 學生化殘差直方圖 (Studentized Residual Histogram):** Histogram titled '學生化殘差直方圖' (Studentized Residual Histogram). The x-axis is '學生化殘差' (Studentized Residual) ranging from -4 to 6. The y-axis is '次數' (Frequency) ranging from 0 to 200. The distribution is roughly symmetric.
- 客製化選單 (Customized Menu):** A yellow box highlighting the 'Reg_QC圖/表' (Reg_QC Chart/Table) button in the top navigation bar.
- R統計運算 (R Statistical Calculation):** A yellow box highlighting the bottom section showing statistical results for station 466900 on March 6, 2003, at hour 1.

| 觀測站 | 日期 | 時 | 觀測值TT | 推估值UK | 學生化殘差 | CMT | SD | Outliers |
|--------|----------|---|-------|-------|-------|--------|-------|----------|
| 466900 | 20030601 | 1 | 23.8 | 20.6 | 5.163 | 23.751 | 2.159 | 0.0227 |

Show 10 entries

Showing 1 to 1 of 1 entries

Previous 1 Next

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. The top navigation bar includes links for document upload, data processing, statistical charts, model evaluation, and prediction models. A red box highlights the 'Prediction Model' dropdown menu. Below it, a slider labeled 'Probability Model Threshold' is set to 0.1, with a red box around its value and the slider itself.

A yellow speech bubble on the right side of the screen contains the text '預測結果' (Prediction Result). Red arrows point from this text towards the rightmost columns of the data table, which are labeled '預測機率' (Prediction Probability) and '理賠' (Claim).

The main data table displays 10 entries (out of 12 total) with various columns including編號 (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 (Age), 車齡組合 (Age Combination), 預測機率 (Prediction Probability), and 理賠 (Claim). The last two columns are highlighted with red boxes.

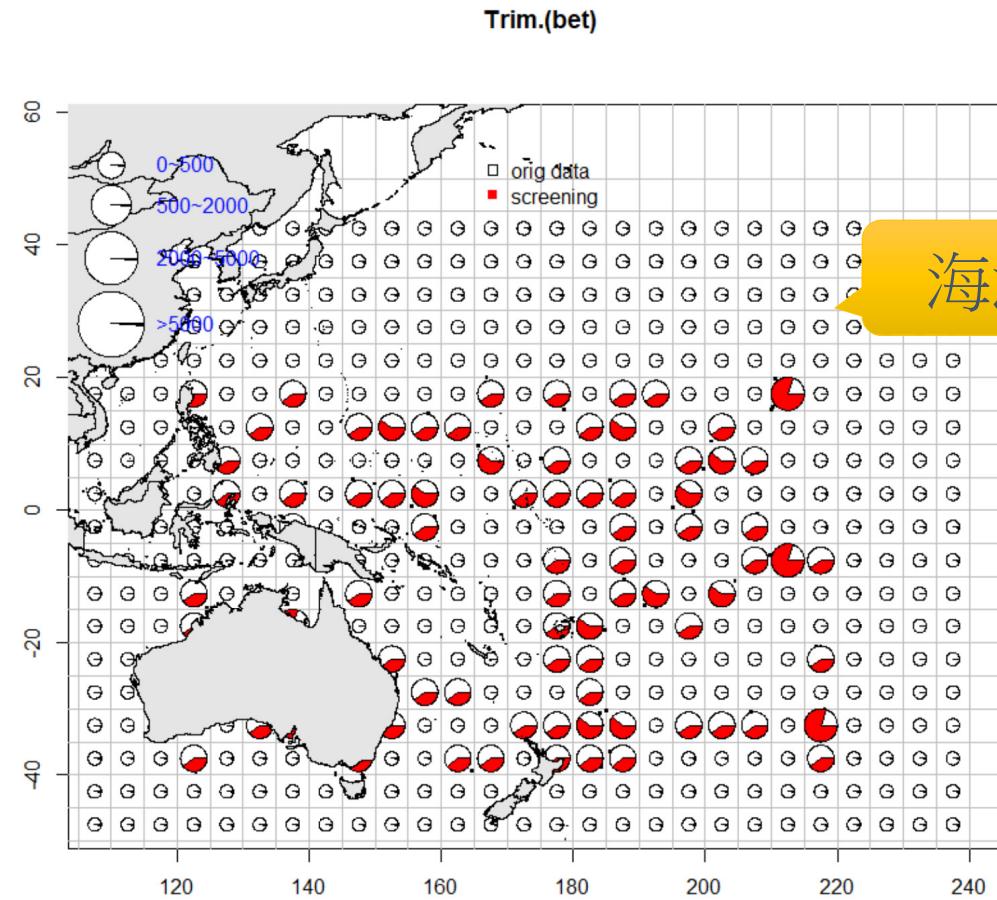
| 編號 | 性別 | 女性 | 車輛種類 | 私家車 | 曝露風險 | 曝露風險對數 | 無索償折扣 | 被保險人年齡 | 私家車一車齡 | 私家車二車齡 | 私家車三車齡 | 私家車車齡組合 | 車齡 | 車齡組合 | 預測機率 | 理賠 |
|----|----|----|------|-----|-----------|-------------|-------|--------|--------|--------|--------|---------|----|------|--------|----|
| 1 | M | 0 | A | 1 | 0.9144422 | -0.08944106 | 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

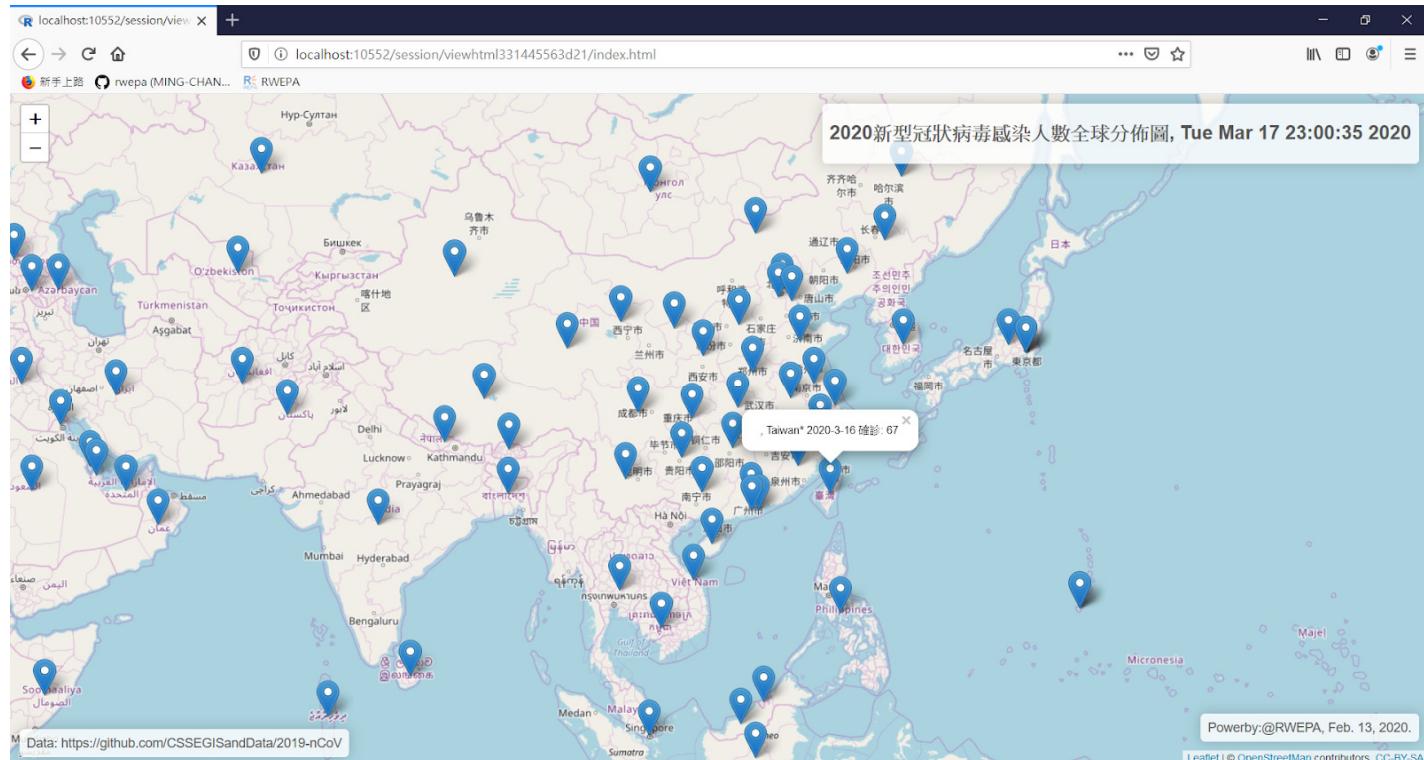
Previous 1 2 Next

127.0.0.1:6177/#tab-9487-2

空間圓餅圖離群值分析



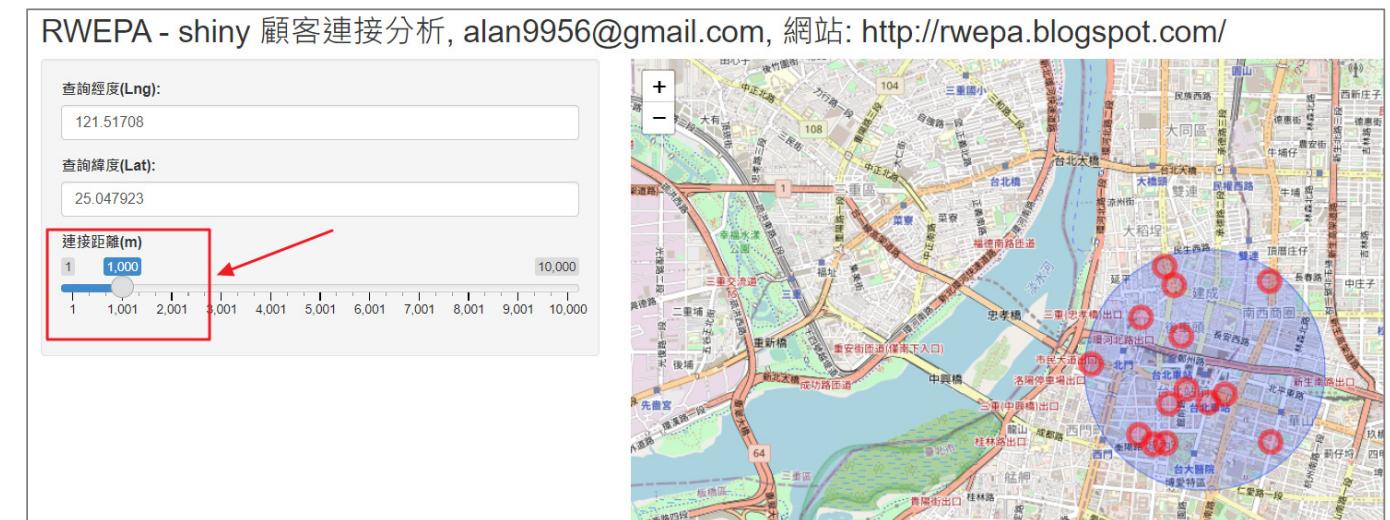
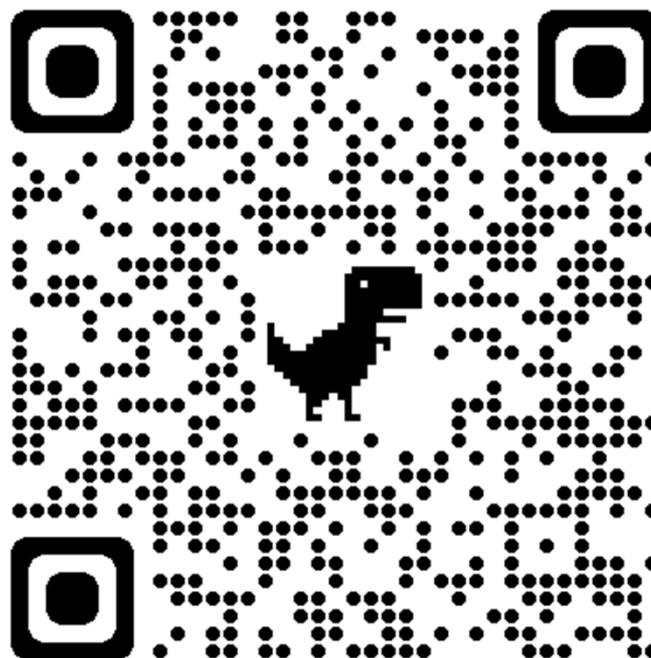
2020新型冠狀病毒視覺化



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

shiny 顧客連接分析

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



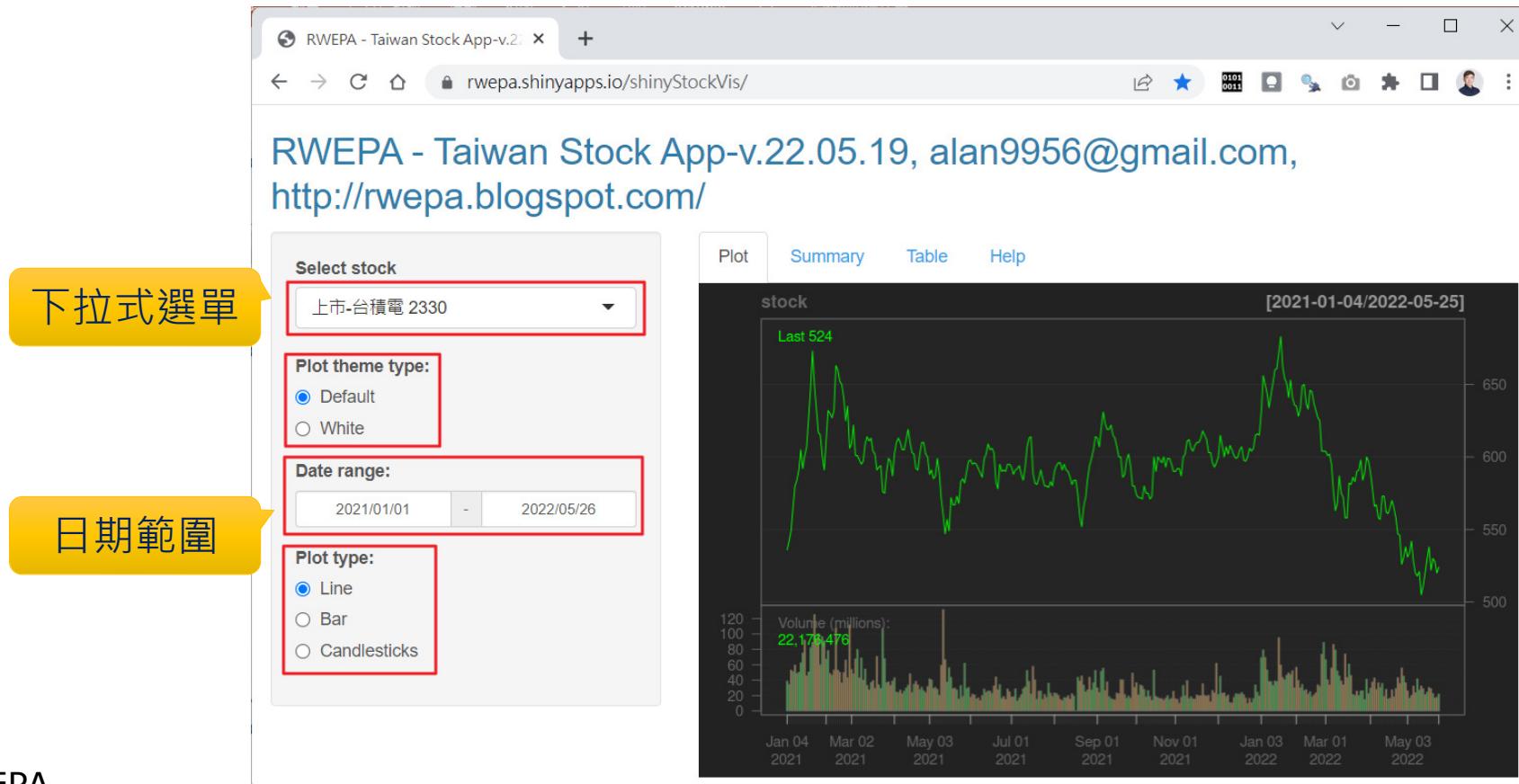
品質管制圖(quality control chart)應用

- 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/>

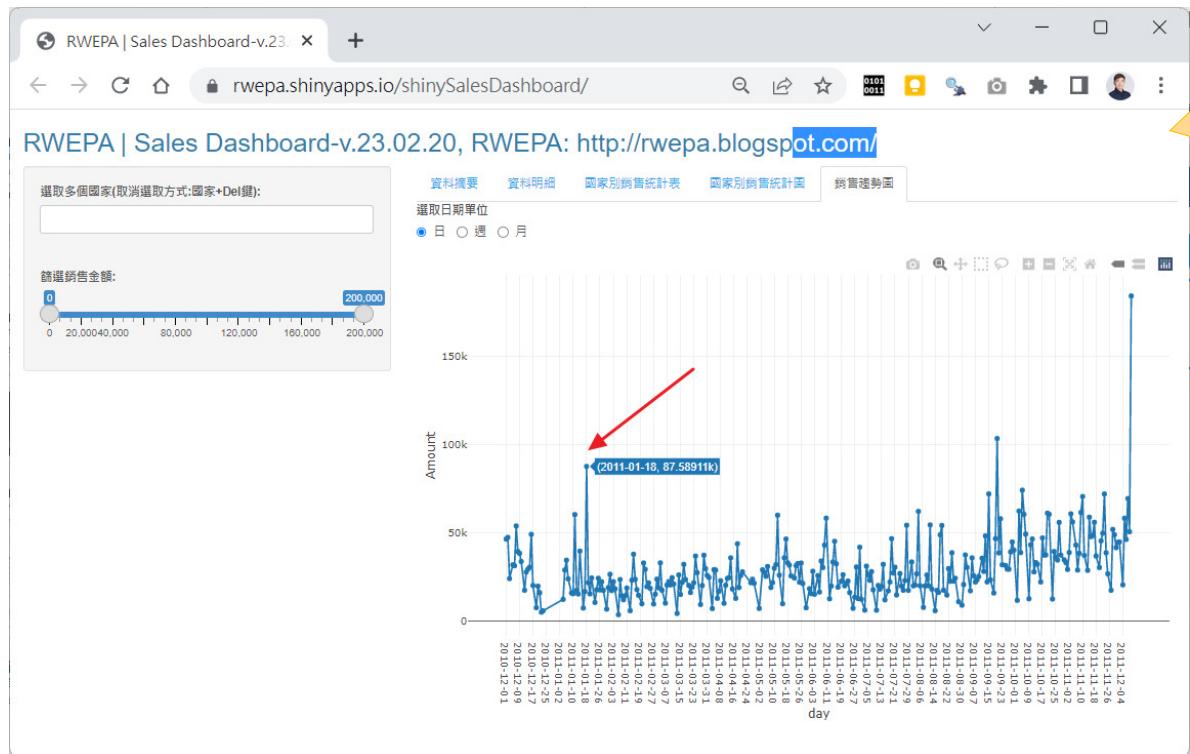


離子資料分析與視覺化應用



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

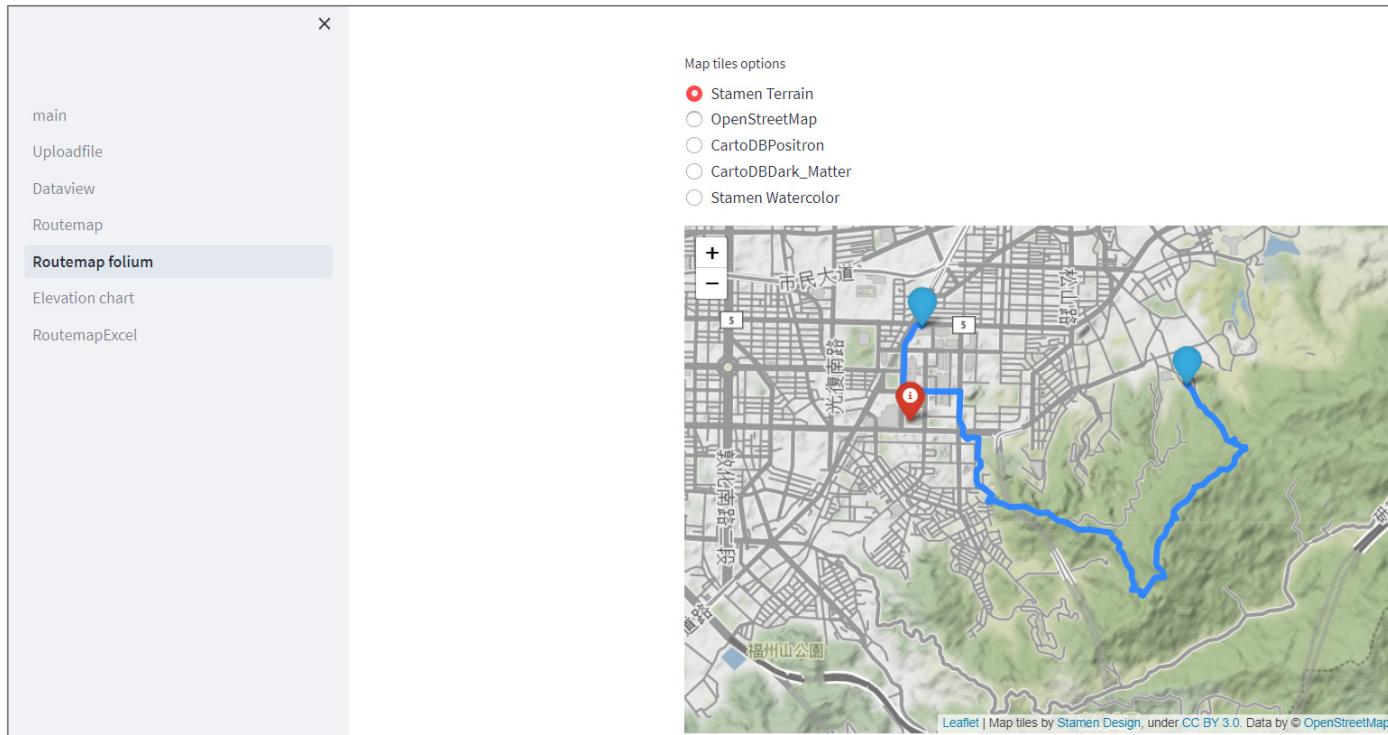
shiny企業實務應用 第6集-小明算命師(下) - 第1季完結篇

- 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

登入

檔案 常用 插入 模型化 檢視 最佳化 說明

剪貼簿 貼上 剪下 複製 複製格式 取得資料 Excel 資料中種 SQL 輸入資料 Dataverse 最近使用的來源 資料

查詢 轉換資料 重新整理 新增視覺效果 文字方塊 更多視覺效果 插入 計算 新增量值 快速量值 敏感度 發行 共用

客戶輪廓Overview

客戶數 平均年齡

666 34.8

FEMALE MALE

288 (43.2%) 378 (56.8%)

客戶等級 全選 VIP客戶 一般客戶 財富管理客戶

年齡組距 01.25歲以下 02.25-35歲 03.35-50歲 04.50-60歲 05.超過60歲

職業 其他未... 製造/營造業 資訊科技業 金融相關... 餐飲服務/... 批發零售業 軍/警/公/教 醫療相關... 運輸倉儲業 專業人員 服務業 宗教、慈... 政府官員... 教育業

21% 17% 13% 11% 11% 10% 9% 5% 2% 1% 0% 0% 0%

縣市別 新北市 台北市 桃園市 台中市 高雄市 台南市 彰化縣 基隆市 新竹市 新竹縣 苗栗縣 南投縣 花蓮縣 雲林縣

29% 23% 11% 8% 8% 5% 4% 3% 2% 2% 1% 1% 1% 1%

區域別 新店區 內湖區 中和區 松山區 大安區 三重區 中山區 板橋區 新莊區 桃園區 東區 信義區 中壢區 中正區

4% 4% 3% 3% 3% 3% 3% 3% 3% 3% 3% 2% 2%

視覺效果 組建視覺效果

資料 搜尋

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

頁面 9 之 1 84%

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

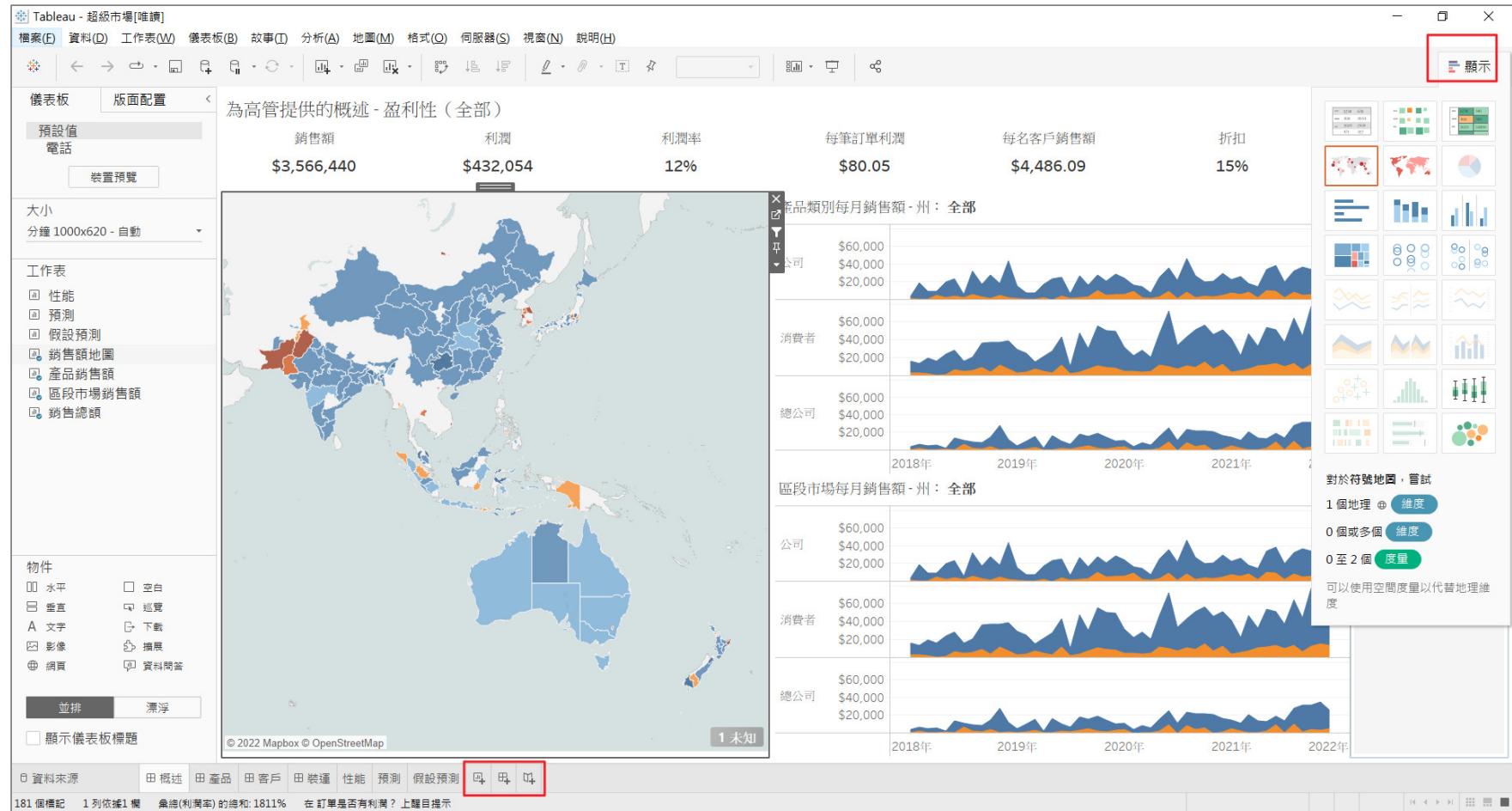
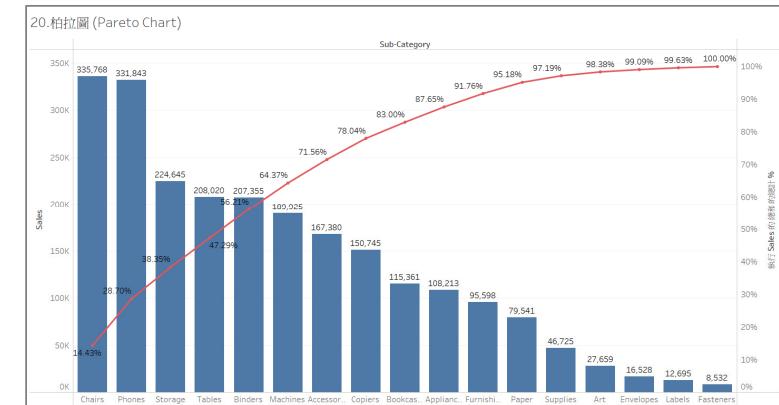
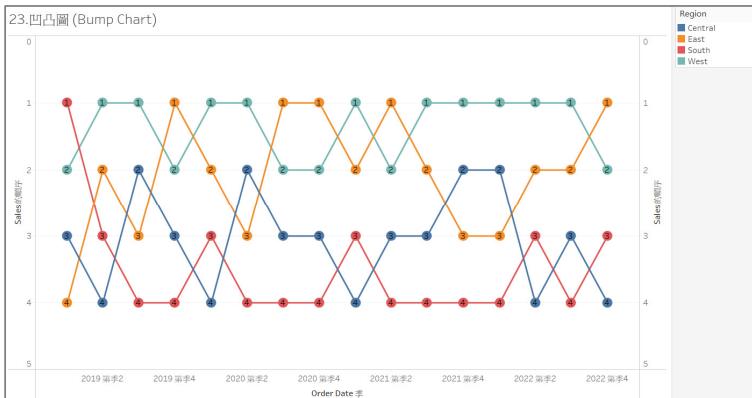
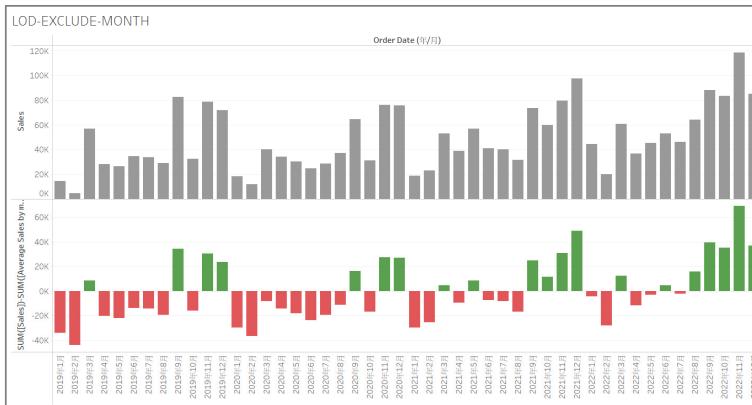


Tableau - 智慧製造應用

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



25.購物籃分析 (Market Basket Analysis)

Sub-Category

| Sub-Cat.. | Access.. | Applian.. | Art | Binders | Bookcas.. | Chairs | Copiers | Enviro.. | Fasten.. | Furnishi.. | Machine.. | Paper | Phones | Storage | Supplies | Tables | | |
|-------------|----------|-----------|-----|---------|-----------|--------|---------|----------|----------|------------|-----------|-------|--------|---------|----------|--------|-----|----|
| Accessories | 718 | 60 | 89 | 161 | 29 | 64 | 6 | 32 | 27 | 115 | 47 | 22 | 19 | 110 | 78 | 56 | 24 | 32 |
| Appliances | 60 | 459 | 63 | 135 | 13 | 49 | 9 | 21 | 23 | 82 | 21 | 9 | 110 | 78 | 56 | 24 | 32 | |
| Art | 89 | 63 | 756 | 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 | |
| Enviro.. | 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 | |
| Furnishi.. | 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 | |
| Machines | 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 R?

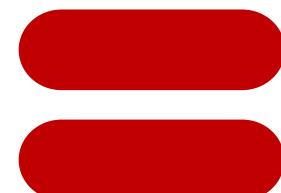
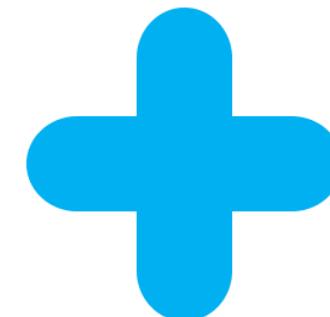
- 熟悉教材內容
- 將教材內容的資料集改為工作資料集(企業, 學術)
- 遇到問題時, 想辦法**尋找答案**
- 掌握 APC方法
- 掌握 ①摘要 ②繪圖 ③建模
- 參考網路應用文章 (進階) & 學術論文

```
> 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"
```

WHY!



學習目標



R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
 - 2013年4月 - R 3.0.0
 - 2024年6月 - R 4.4.1



Google
台灣

facebook

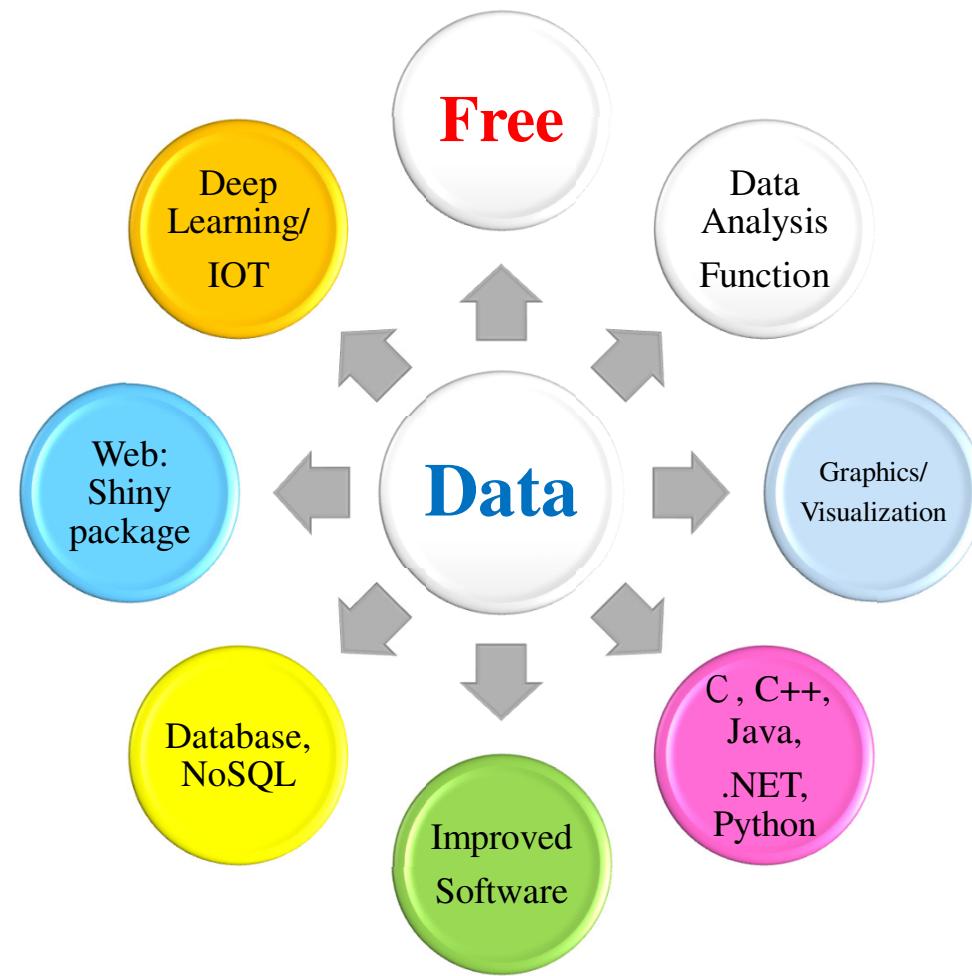
sas[®]



IBM
SPSS

+ a b | e a u[®]

R-八大功能



R安裝

R官方網頁

[\[Home\]](#)[Download](#)[CRAN](#)**下載**[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

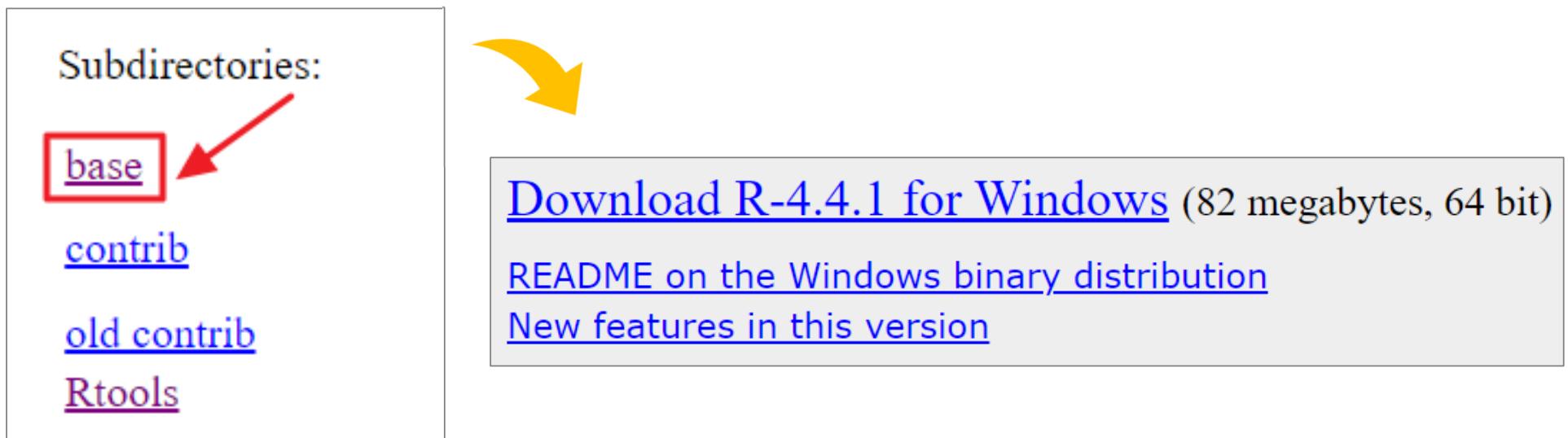


- [Download R for Linux](#) ([Debian](#), [Fedora/Redhat](#), [Ubuntu](#))
- [Download R for macOS](#)
- [Download R for Windows](#)



R-下載 (續)

- 選取 base → 下載 [R-4.4.1-win.exe]



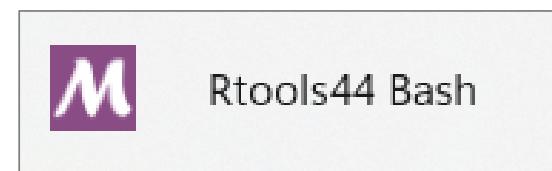
- R安裝路徑: 保留原路徑,不要修改
- 安裝參考說明, 2006
https://github.com/rwepa/DataDemo/blob/master/windows_intall_R.pdf

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

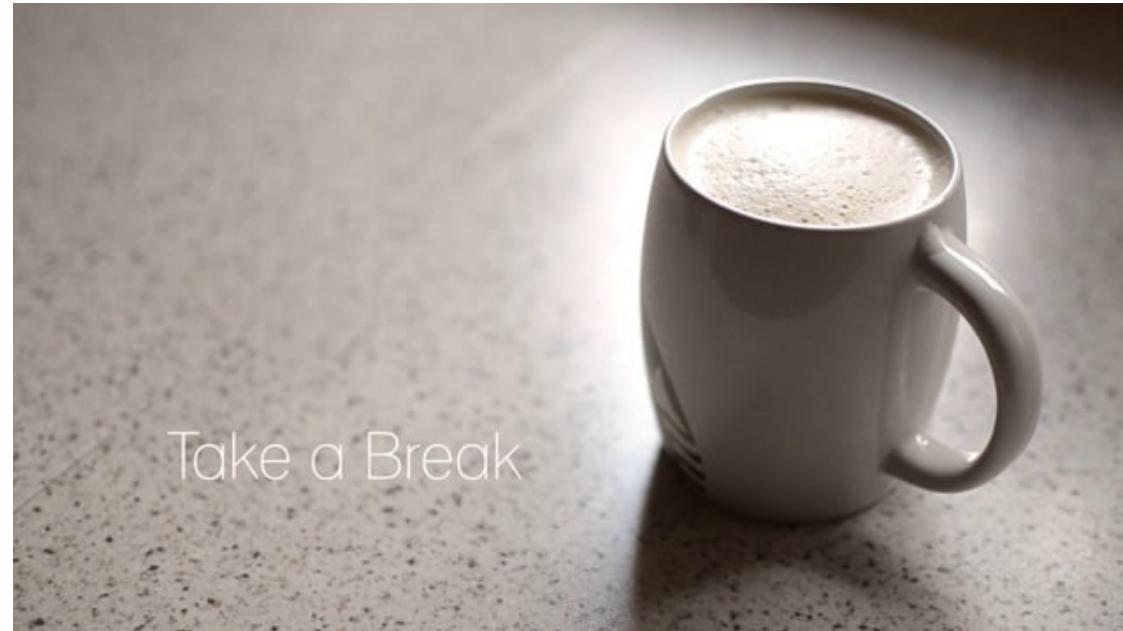
- 安裝完成：程式集 \ Rtoolsxx Bash



Rtools44 Bash

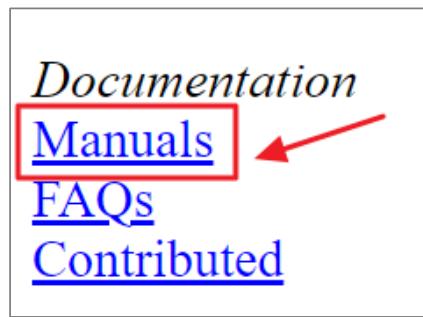
✓ 安裝 R

✓ 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). All books 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!

LINK: <https://cran.csie.ntu.edu.tw/doc/contrib/usingR.pdf>

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

TRY:
`runif(100000)`

```
R Gui (64-bit)
檔案 命令歷程 重設大小 視窗
R Console
R version 3.0.3 (2014-03-06) -- "Continuous Integration"
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布  

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  

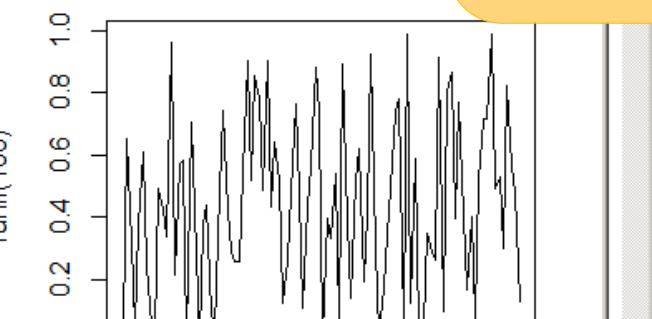
用 'citation()' 會告訴您如何在  

用 'demo()' 來看一些示範程式，  

用 'help.start()' 透過 HTML 滾動  

用 'q()' 離開 R。

> plot(runif(100), type="l")
```



- 安裝R, 登入名稱:
1. 不要使用空格
 2. 不要使用中文字型

`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



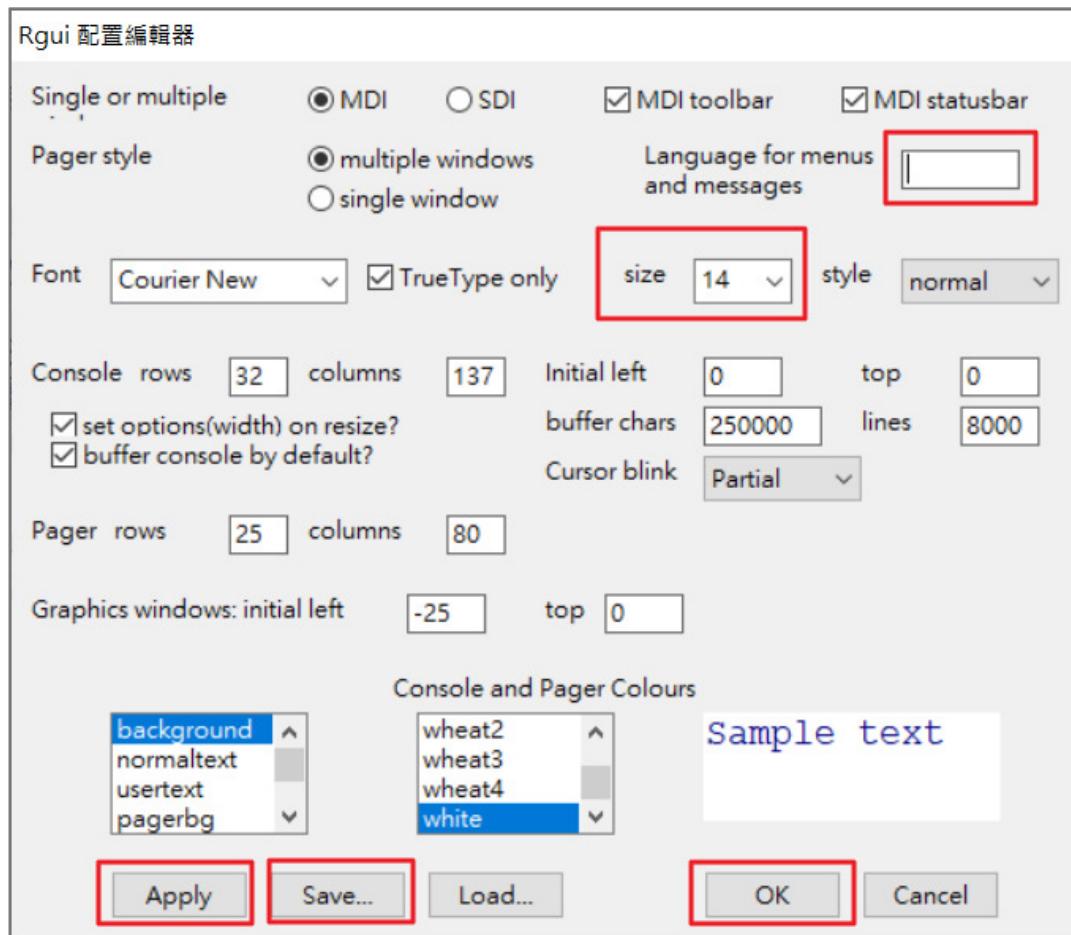
現行目錄 getwd()

輔助



儲存控制台-文字檔

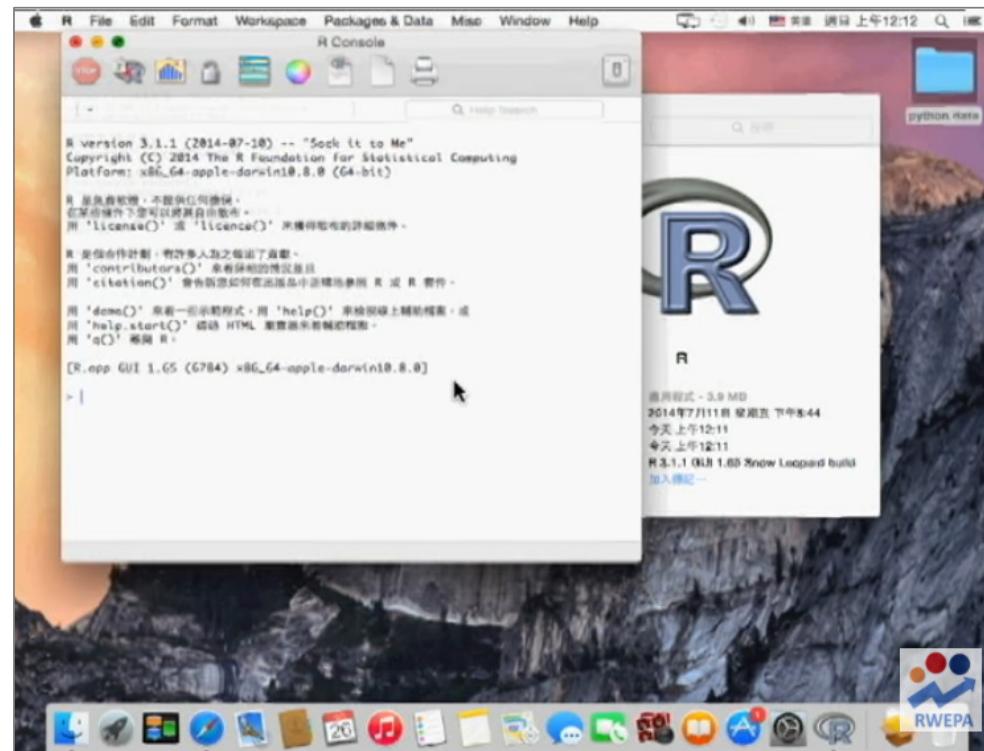
編輯 \ GUI 偏好設定



- Language: en 英文
- size: 字型大小

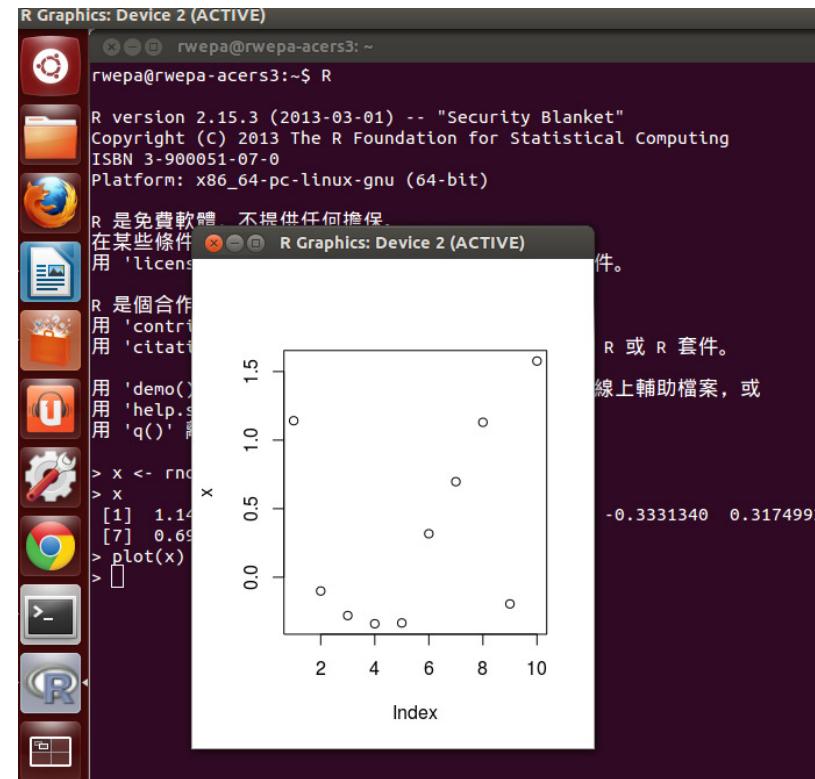
R for Mac

- <https://youtu.be/72MYRBNo5Bk>
- Xcode, Fortran compiler: <https://cran.r-project.org/bin/macosx/tools/>



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. In the center is the R Console window showing the command history and the resulting text output. On the right is the R Graphics window displaying a 4x4 scatter plot matrix for the Iris dataset, titled "RWEPA-iris資料集散佈圖矩陣".

• 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 \ 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] > pa
+ pc
+ co
+ main="RWEPA-iris資料集散佈圖矩陣"
> |
```

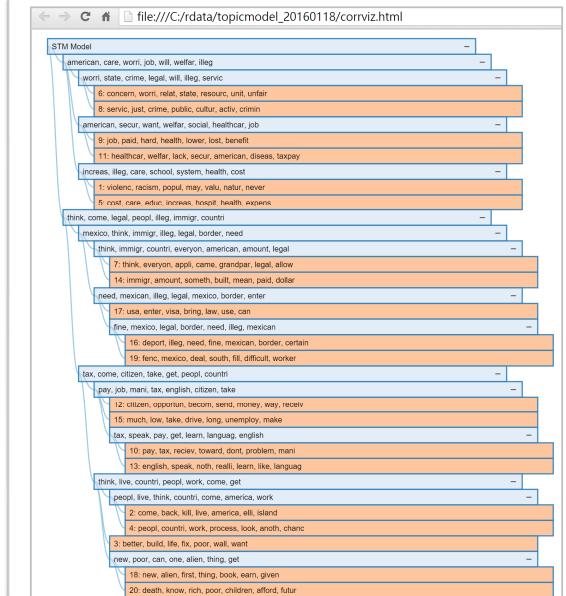
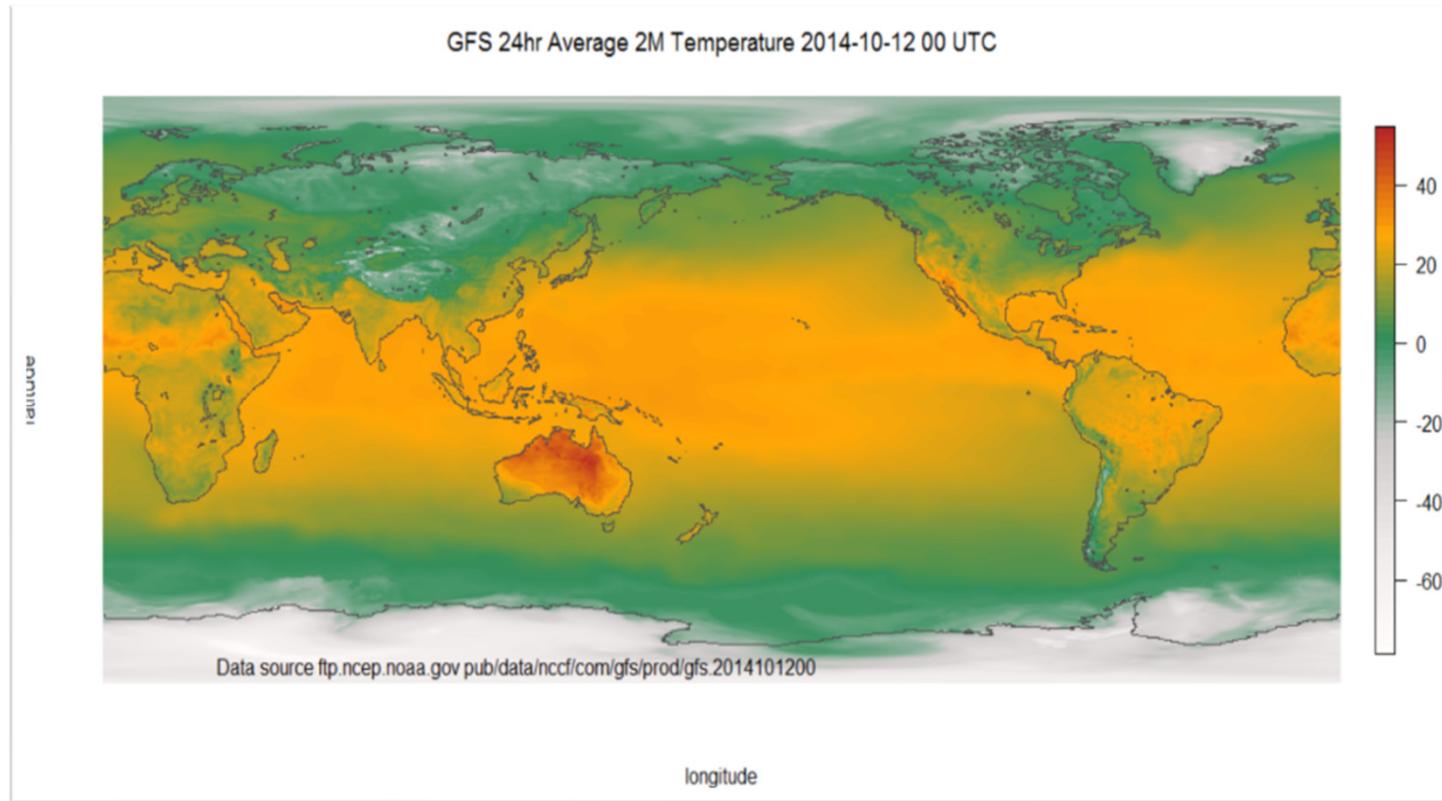
R 是
是在某
用，
R 是
用，
用，
用。

RWEPA-iris資料集散佈圖矩陣

| | 2.0 | 3.0 | 4.0 | 0.5 | 1.5 | 2.5 |
|--------------|---------------------|---------------------|---------------------|---------------------|---------------------|---------------------|
| Sepal.Length | ● ● ● ● ● ● ● ● ● ● | ● ● ● ● ● ● ● ● ● ● | ● ● ● ● ● ● ● ● ● ● | ● ● ● ● ● ● ● ● ● ● | ● ● ● ● ● ● ● ● ● ● | ● ● ● ● ● ● ● ● ● ● |
| Sepal.Width | ● ● ● ● ● ● ● ● ● ● | ● ● ● ● ● ● ● ● ● ● | ● ● ● ● ● ● ● ● ● ● | ● ● ● ● ● ● ● ● ● ● | ● ● ● ● ● ● ● ● ● ● | ● ● ● ● ● ● ● ● ● ● |
| Petal.Length | ● ● ● ● ● ● ● ● ● ● | ● ● ● ● ● ● ● ● ● ● | ● ● ● ● ● ● ● ● ● ● | ● ● ● ● ● ● ● ● ● ● | ● ● ● ● ● ● ● ● ● ● | ● ● ● ● ● ● ● ● ● ● |
| Petal.Width | ● ● ● ● ● ● ● ● ● ● | ● ● ● ● ● ● ● ● ● ● | ● ● ● ● ● ● ● ● ● ● | ● ● ● ● ● ● ● ● ● ● | ● ● ● ● ● ● ● ● ● ● | ● ● ● ● ● ● ● ● ● ● |

RStudio 簡介與安裝

整合式開發環境 - RStudio



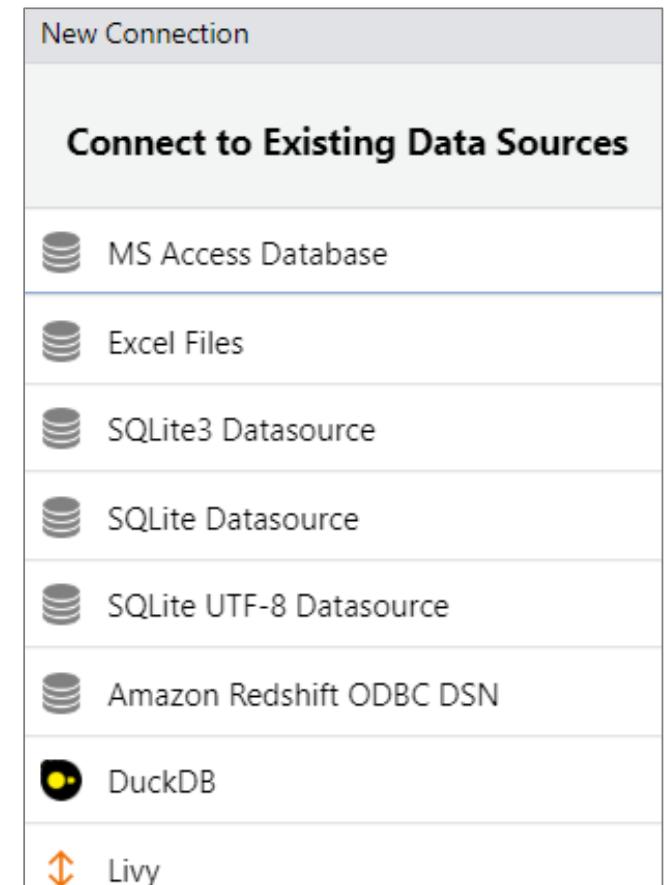
主題模型

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's 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" (also highlighted with a red box). A red arrow points from the crossed-out URL in the list above to this button. Below the navigation bar, the main content area features a large heading "Deployment made easy" and a descriptive paragraph about deploying various types of applications. On the right side of the page, there's a photograph of a person working at a computer, displaying a Shiny dashboard with penguin data analysis.

Deployment
made easy

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.

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

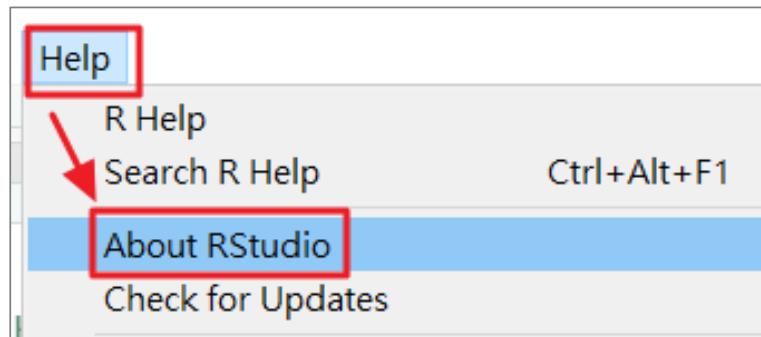
[DOWNLOAD RSTUDIO DESKTOP FOR WINDOWS](#)

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

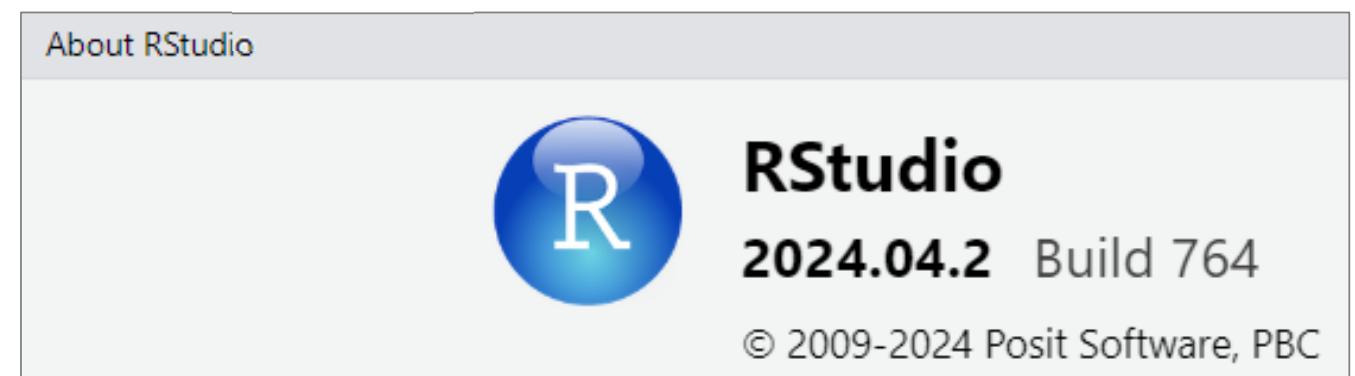
RStudio 安裝



RStudio 版本訊息

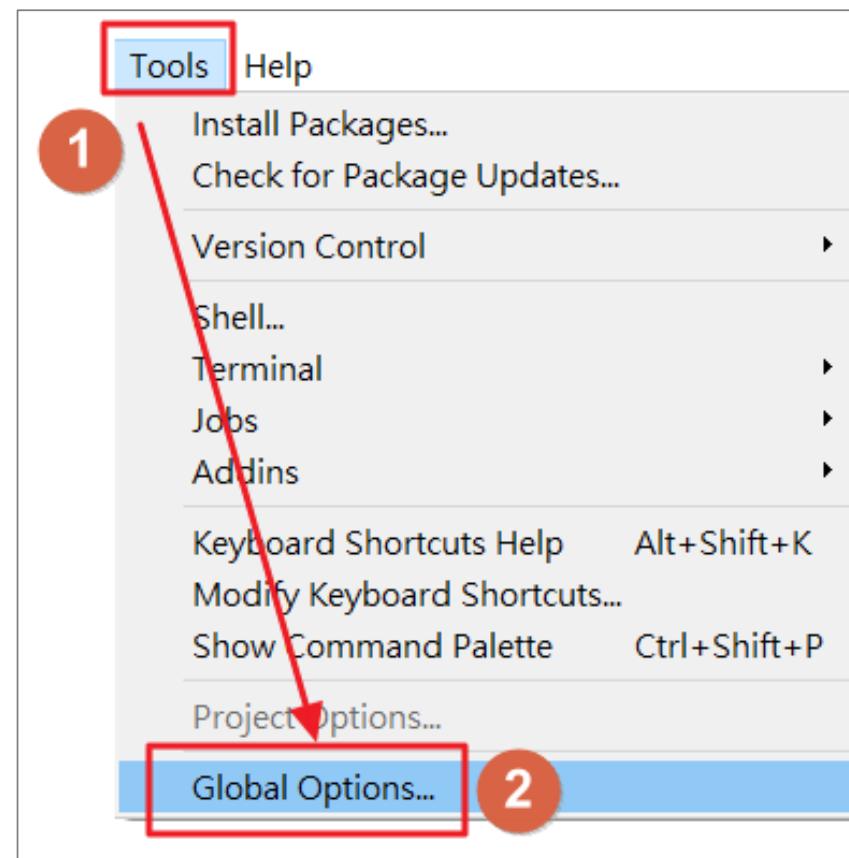


Help \ About RStudio

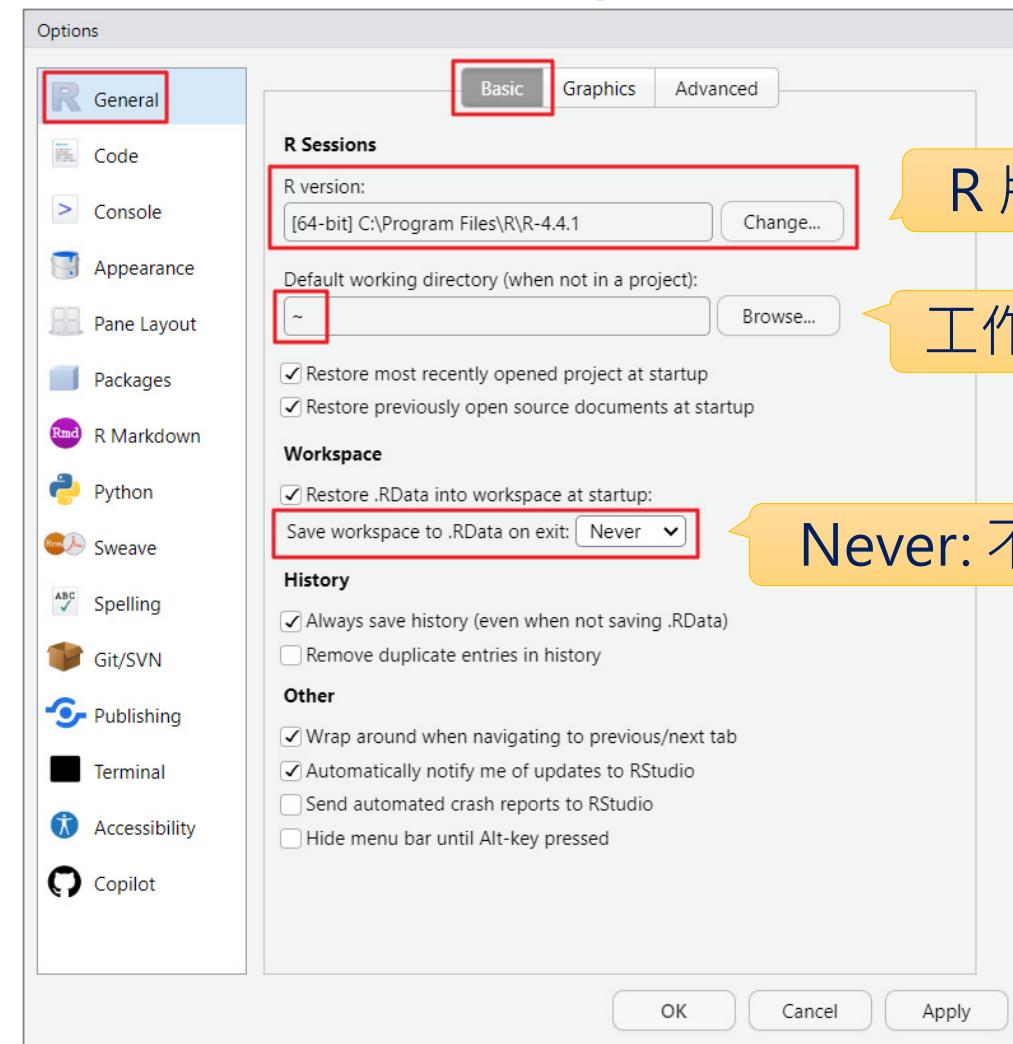


RStudio-選項設定

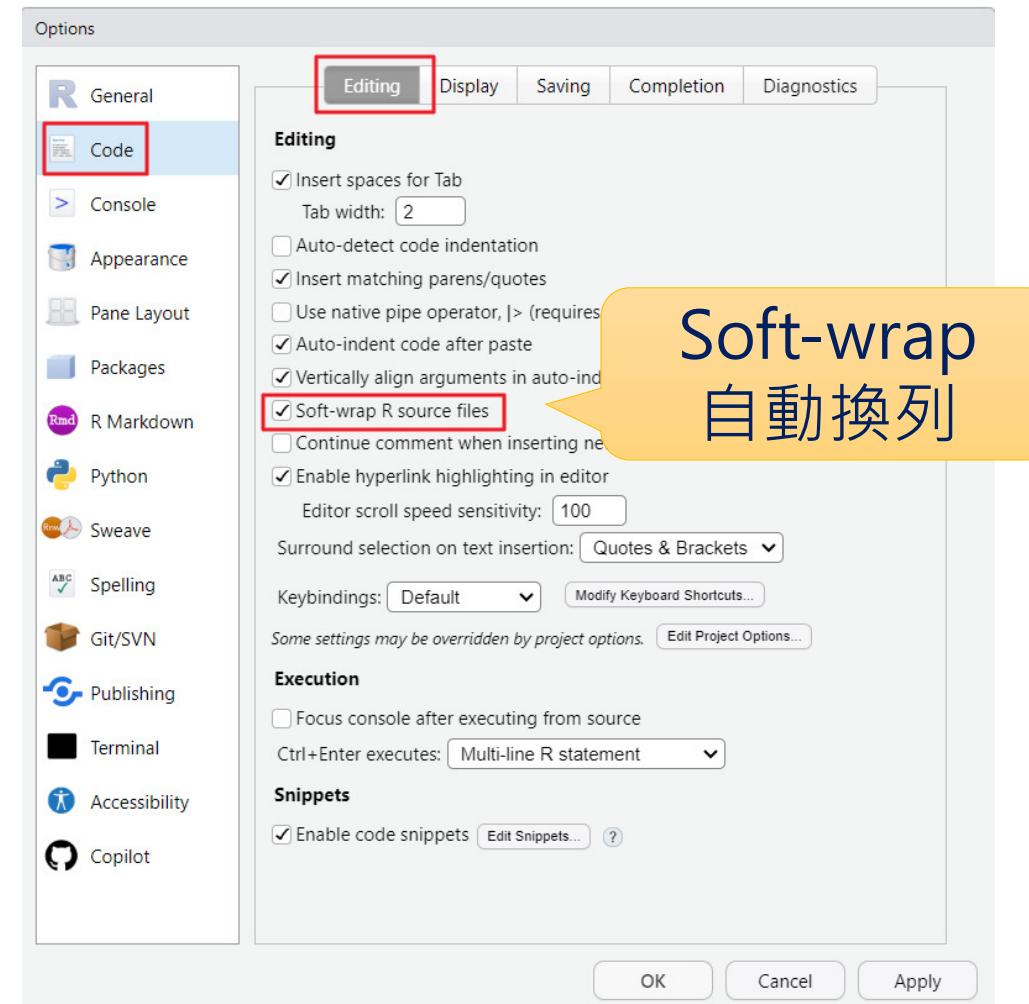
- Tools \ Global Options



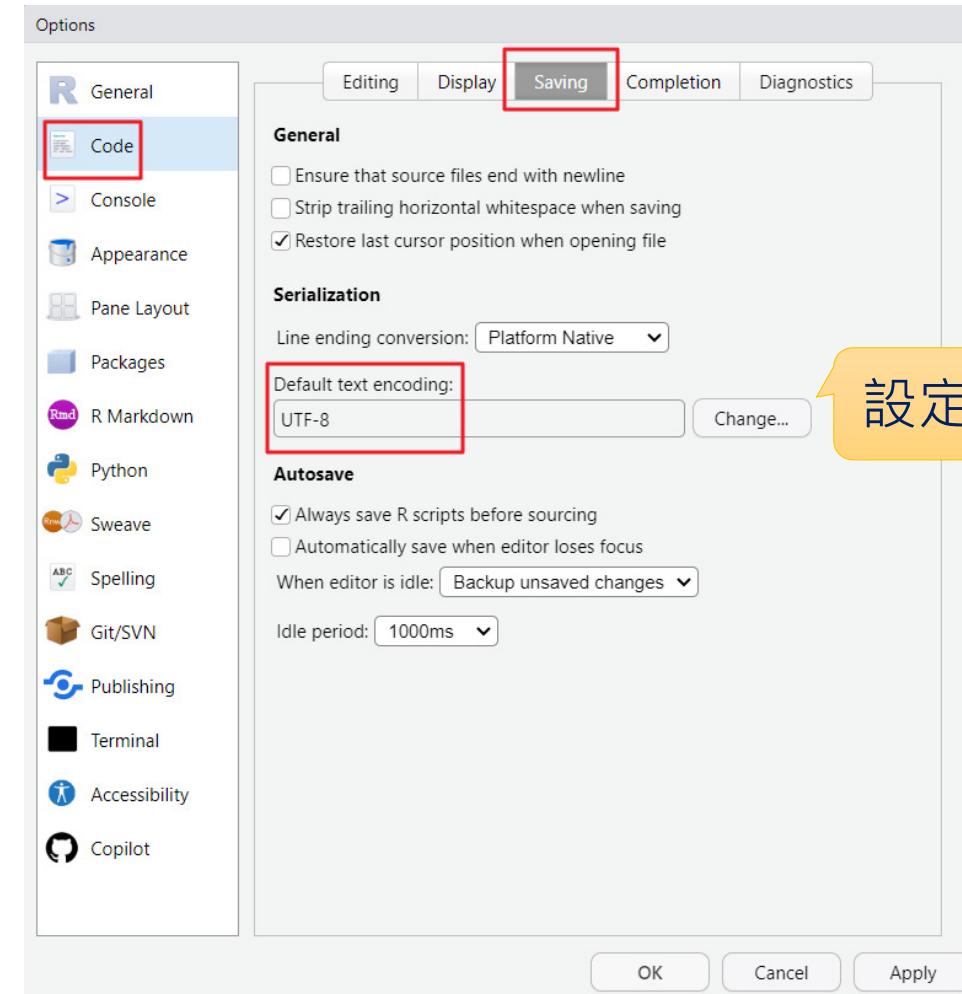
General \ Basic



Code \ Editing



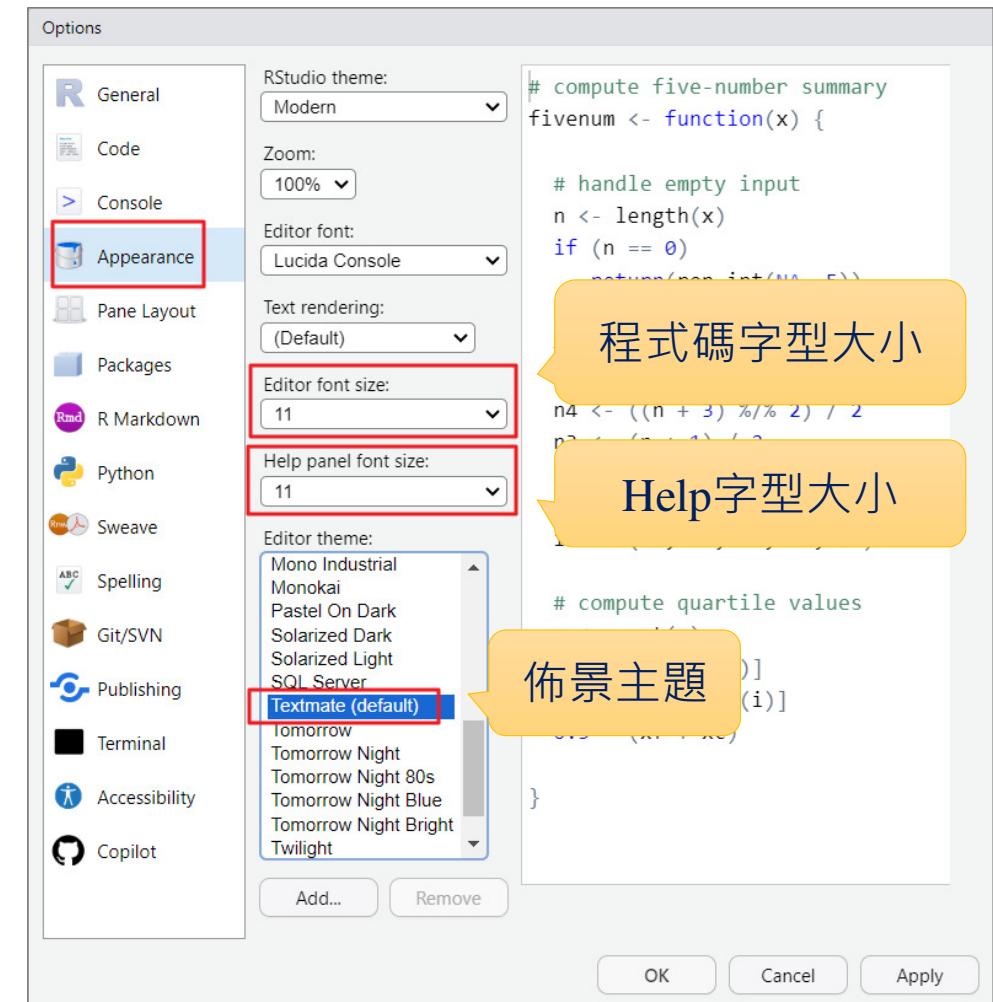
Code \ Saving



RStudio-選項設定(續)

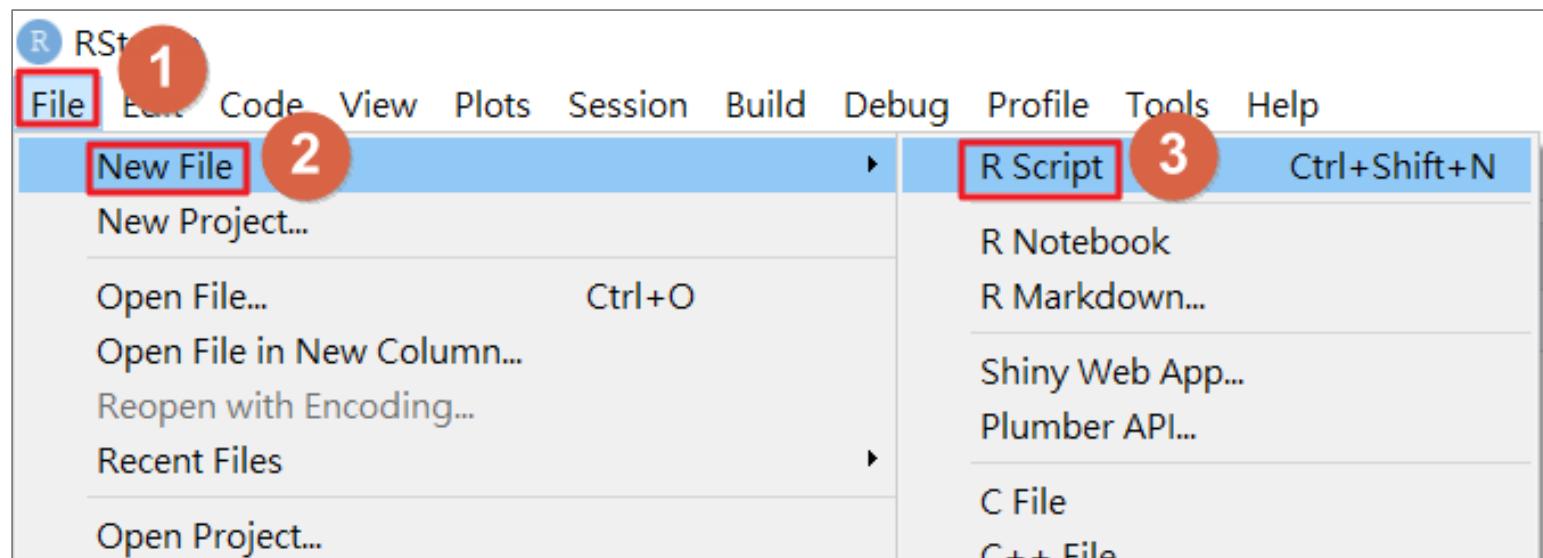
- Appearance \ Editor font size
- Appearance \ Help panel font size
- Editor theme \ TextMate

設定完成，
可能須重新啟動RStudio



新增檔案

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



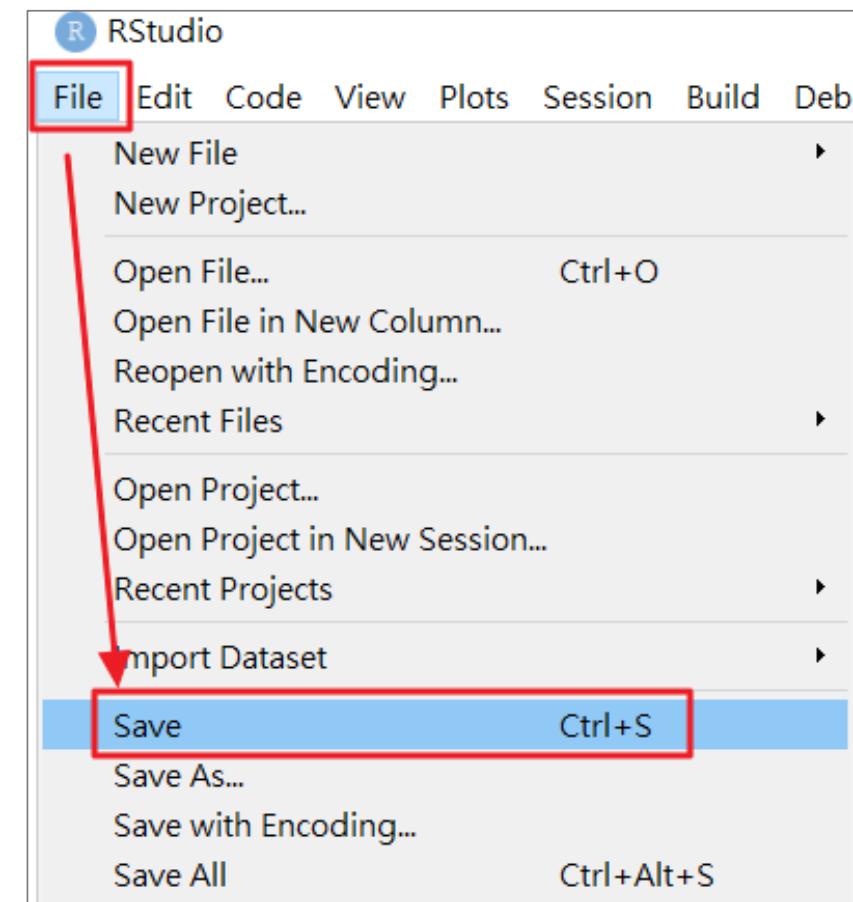
儲存檔案

輸入「新增R檔案練習」

實作
練習

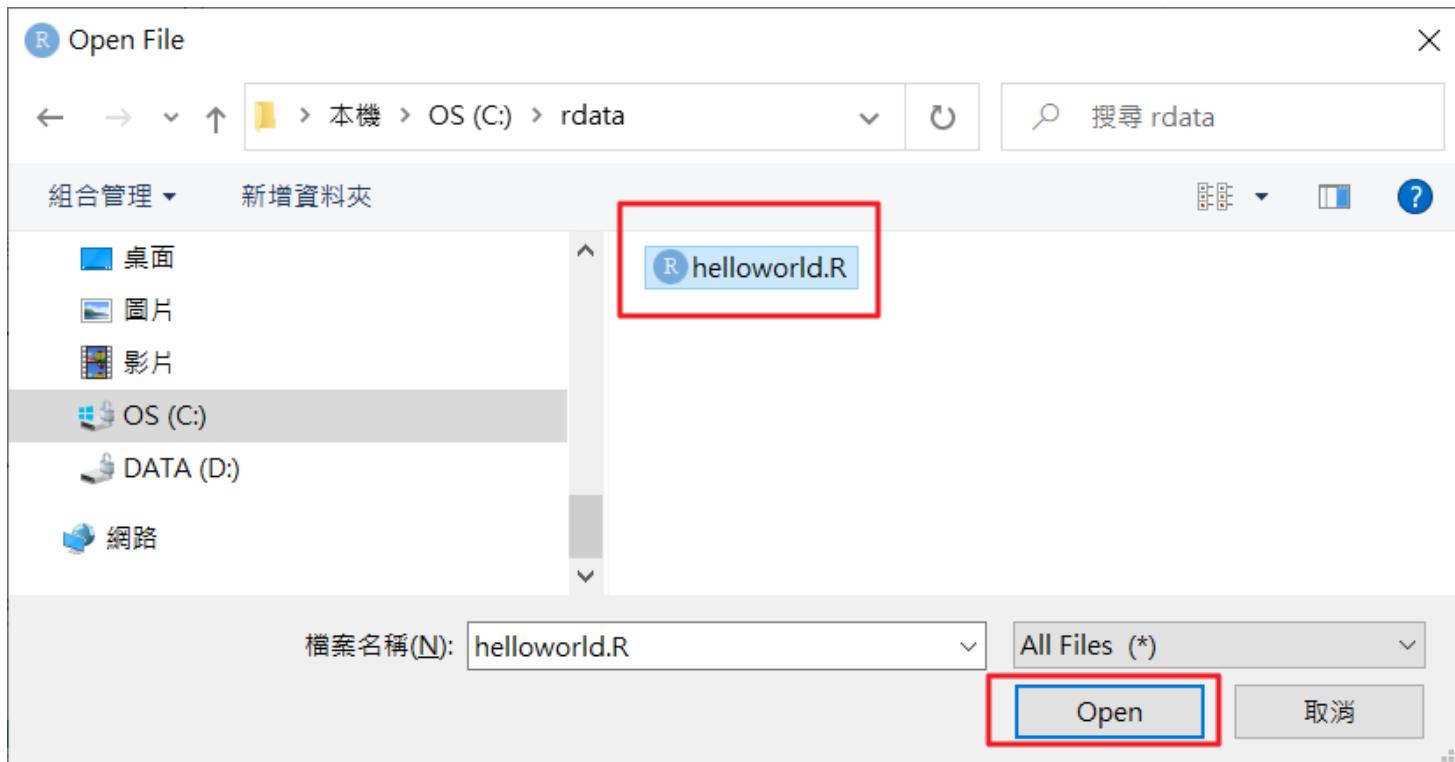
```
1 plot(runif(10), type="b", main= "R大數據分析")
2 x <- rnorm(10)
3 x
4 pairs(iris[-5],
5         pch=16,
6         col=iris$Species,
7         main="RWEPA-iris資料集散佈圖矩陣")
```

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

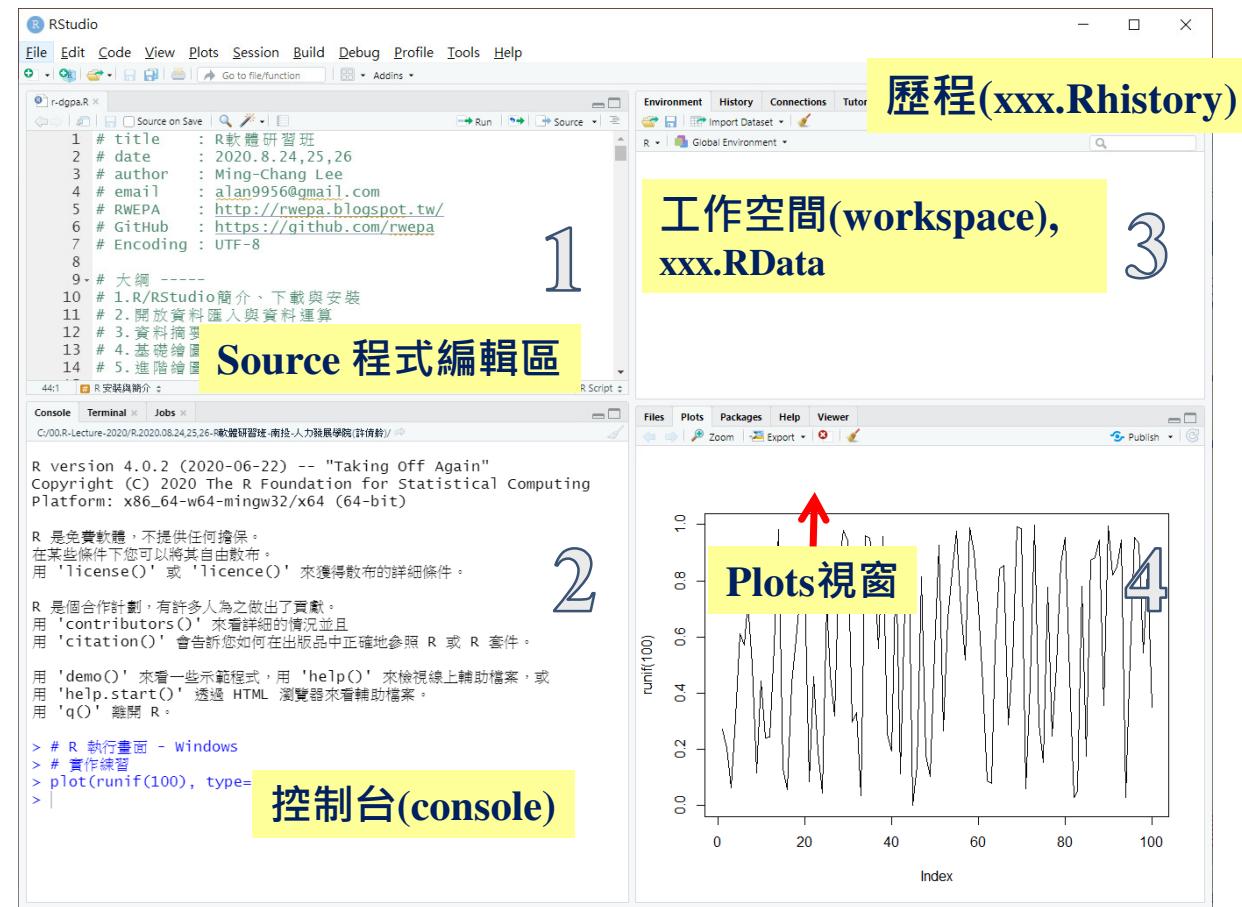


開啟檔案

- File \ Open File \ helloworld.R



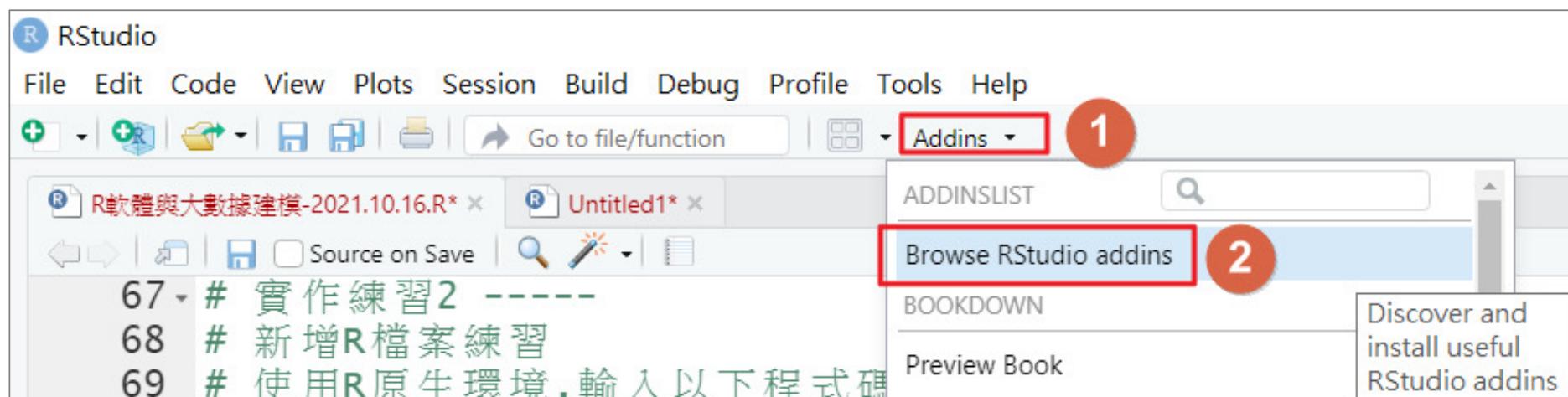
R/RStudio環境的基礎觀念



Ctrl + Shift + F10: 重新啟動R

RStudio Addins (外掛功能)

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



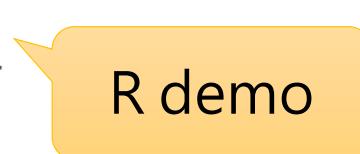
參考: <https://docs.posit.co/ide/user/ide/guide/productivity/add-ins.html>

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 + 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>
- Visual Studio Code + R
 - <https://vscode.dev.org.tw/docs/languages/r>
- Anaconda + Jupyter Notebook
 -  



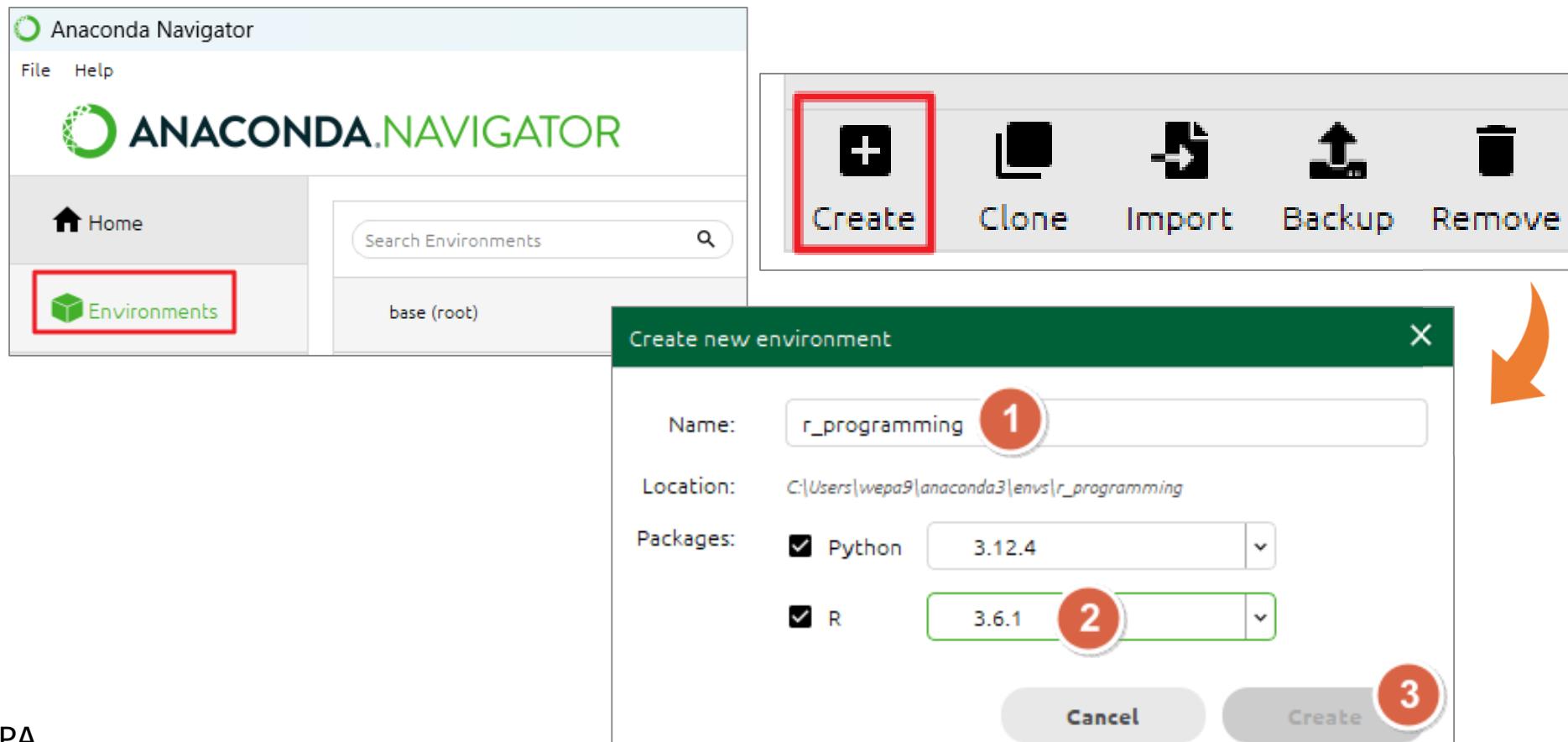
Jupyter Notebook + R

(方法1 使用Anaconda Navigator)

- 僅支援 R-3.6.1
- 先安裝 Anaconda: <https://www.anaconda.com/download>
- <https://docs.anaconda.com/navigator/tutorials/r-lang/>

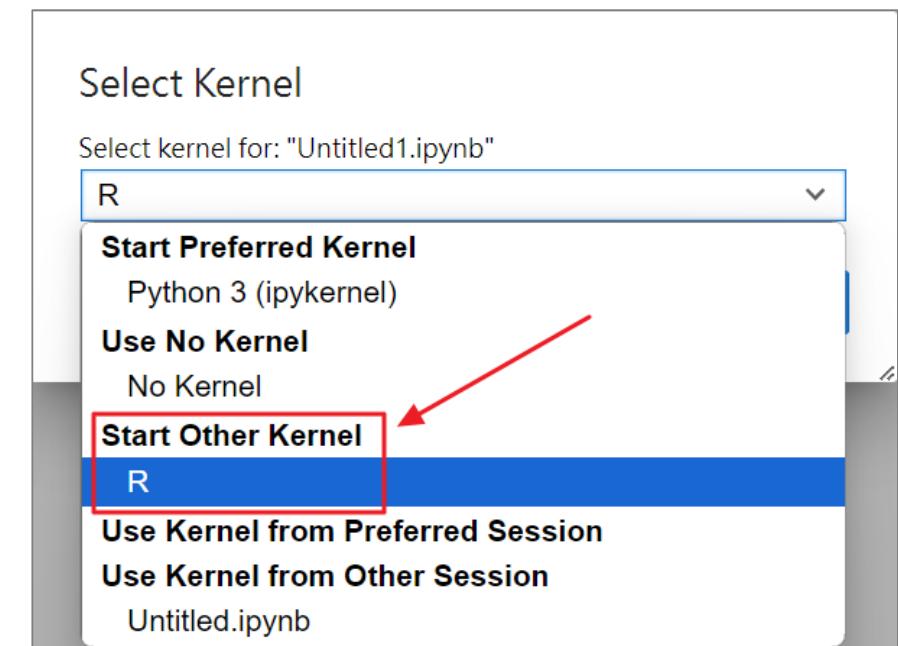
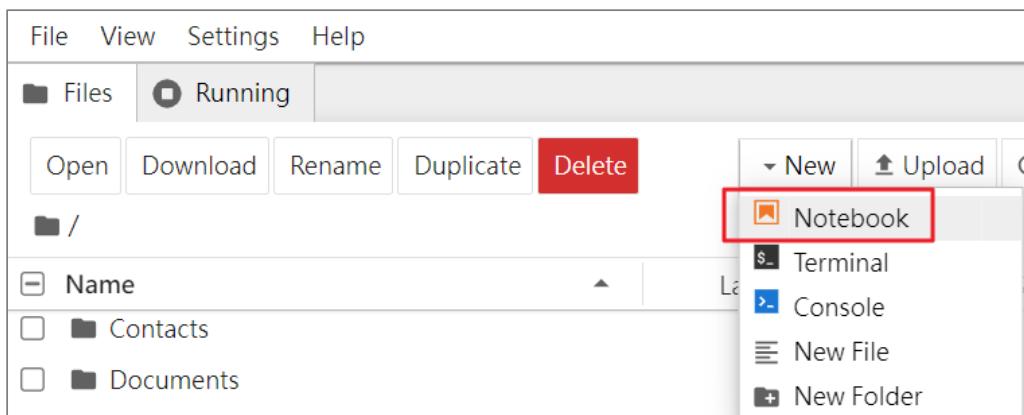
Anaconda Navigator

- Environments \ Create \ 輸入完成後 按 [Create]

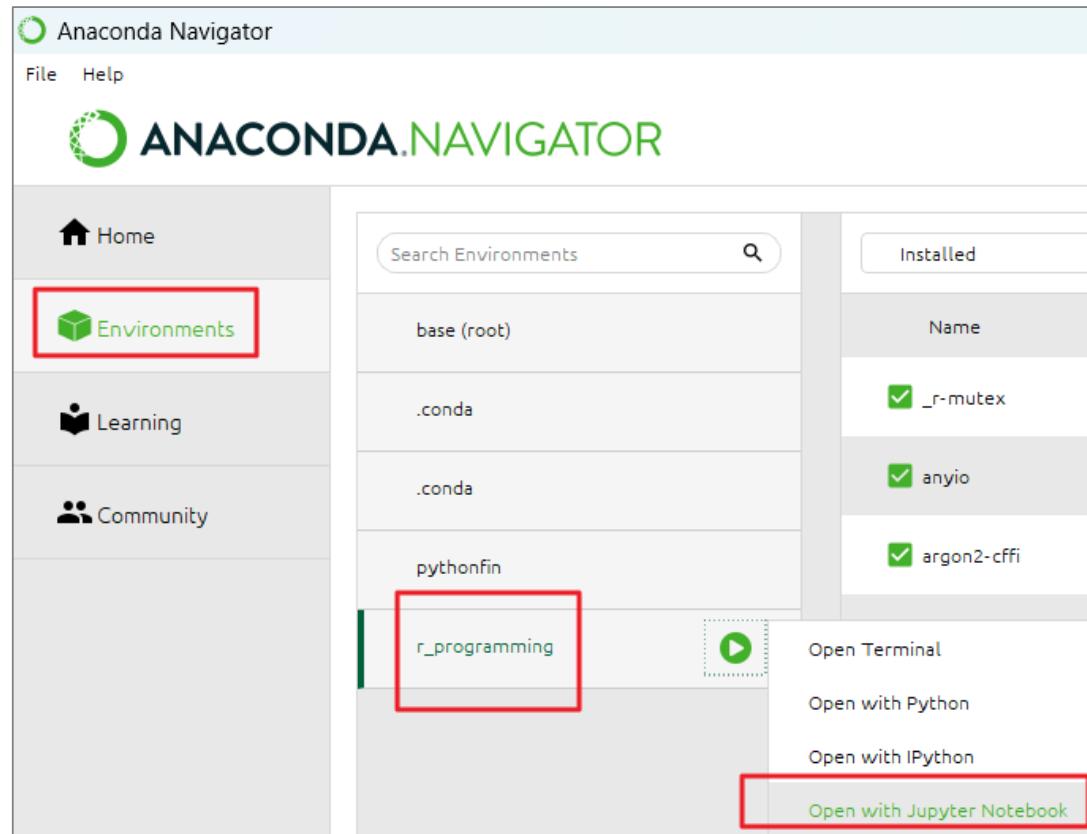


Jupyter Notebook

- New \ Notebook
- Select Kernel \ R



Open with Jupyter Notebook



An orange arrow points from the "Open with Jupyter Notebook" button in the previous screenshot to this Jupyter Notebook interface. The notebook cell [1] displays the following R session information:

```
[1]: sessionInfo()
R version 3.6.1 (2019-07-05)
Platform: x86_64-w64-mingw32/x64 (64-bit)
Running under: Windows 10 x64 (build 22631)

Matrix products: default

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

Jupyter Notebook + R

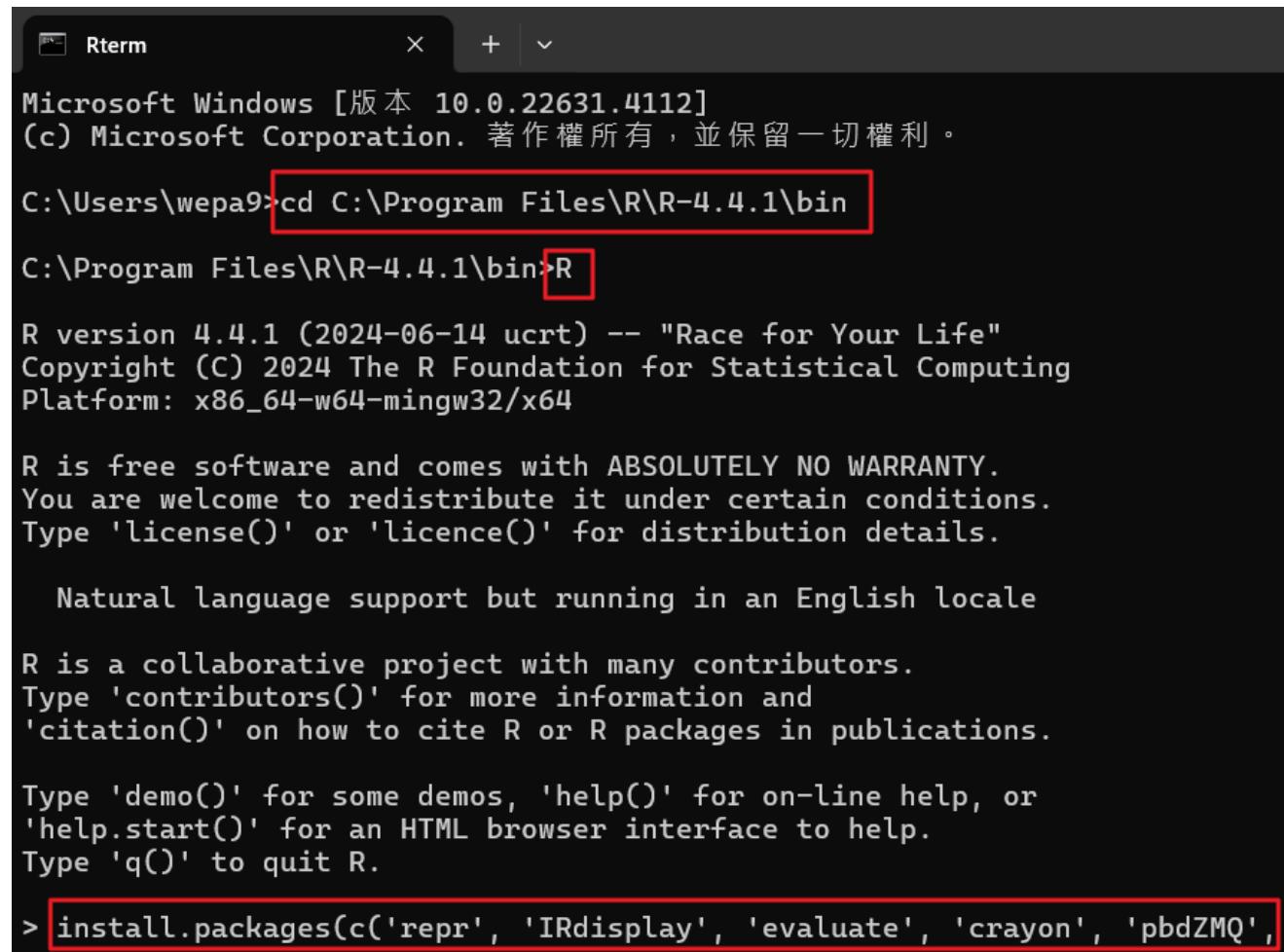
(方法2 使用R原生環境)

- 支援最新版 R-4.4.1
- <https://www.datacamp.com/blog/jupyter-and-r-markdown-notebooks-with-r>

安裝 IRkernel 三步驟

- # 步驟1 開啟R終端機視窗
- 開啟命令提示字元
- cd C:\Program Files\R\R-4.4.1\bin
- R
- # 步驟2 安裝套件
- install.packages(c('repr', 'IRdisplay', 'evaluate', 'crayon', 'pbdZMQ', 'devtools', 'uuid', 'digest'))
- devtools::install_github('IRkernel/IRkernel')
- IRkernel::installspec()
- # 步驟3 關閉R環境
- q() \ 按 n
- exit

R終端機視窗



Rterm

Microsoft Windows [版本 10.0.22631.4112]
(c) Microsoft Corporation. 著作權所有，並保留一切權利。

```
C:\Users\wepa9>cd C:\Program Files\R\R-4.4.1\bin
```

```
C:\Program Files\R\R-4.4.1\bin>R
```

R version 4.4.1 (2024-06-14 ucrt) -- "Race for Your Life"
Copyright (C) 2024 The R Foundation for Statistical Computing
Platform: x86_64-w64-mingw32/x64

R is free software and comes with ABSOLUTELY NO WARRANTY.
You are welcome to redistribute it under certain conditions.
Type 'license()' or 'licence()' for distribution details.

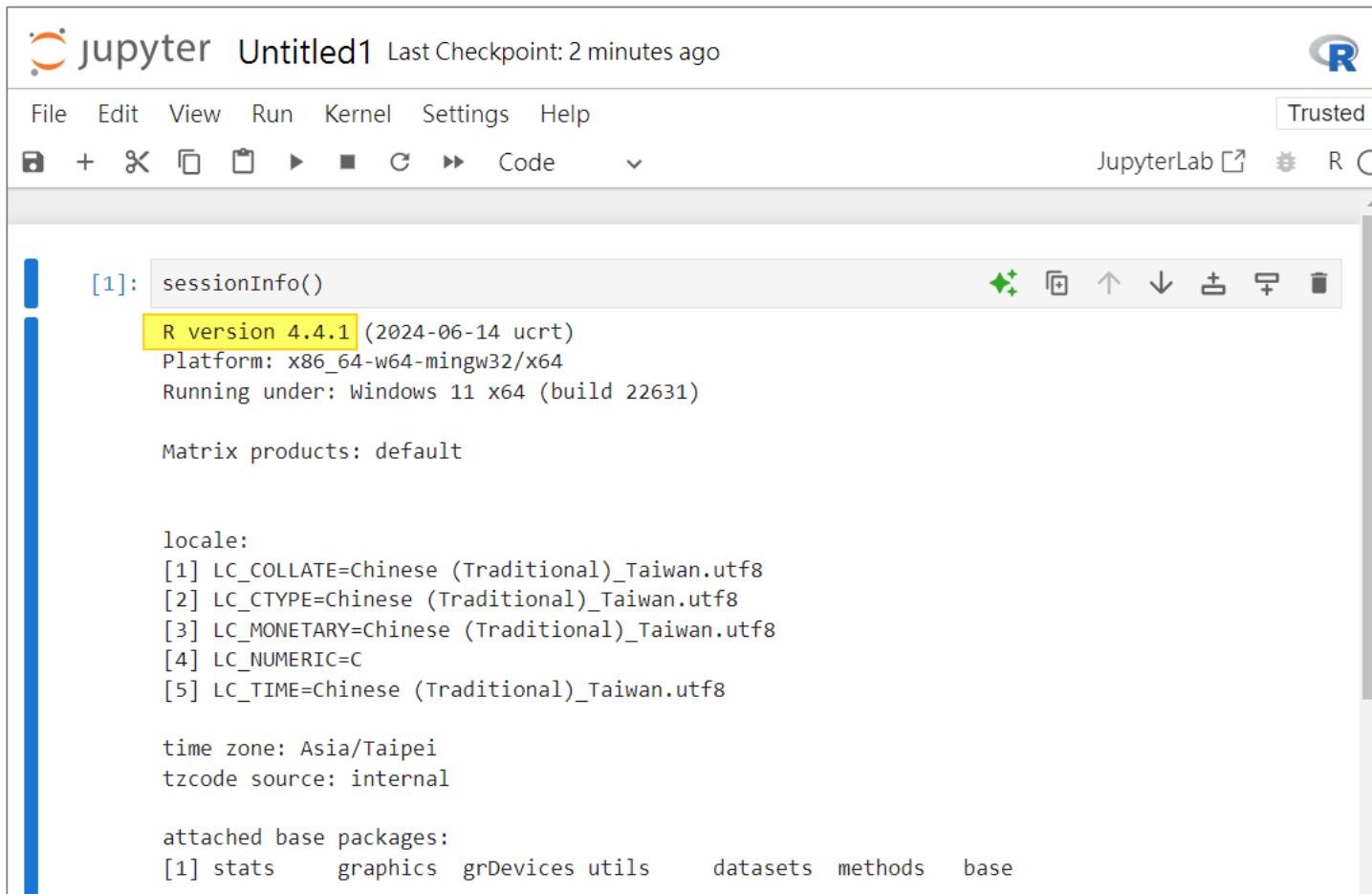
Natural language support but running in an English locale

R is a collaborative project with many contributors.
Type 'contributors()' for more information and
'citation()' on how to cite R or R packages in publications.

Type 'demo()' for some demos, 'help()' for on-line help, or
'help.start()' for an HTML browser interface to help.
Type 'q()' to quit R.

```
> install.packages(c('repr', 'IRdisplay', 'evaluate', 'crayon', 'pbZIPQ',
```

Jupyter Notebook – R demo



jupyter Untitled1 Last Checkpoint: 2 minutes ago

File Edit View Run Kernel Settings Help Trusted

Code JupyterLab R

```
[1]: 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
```

Jupyter Notebook – R demo (續)

File Edit View Run Kernel Settings Help Trusted

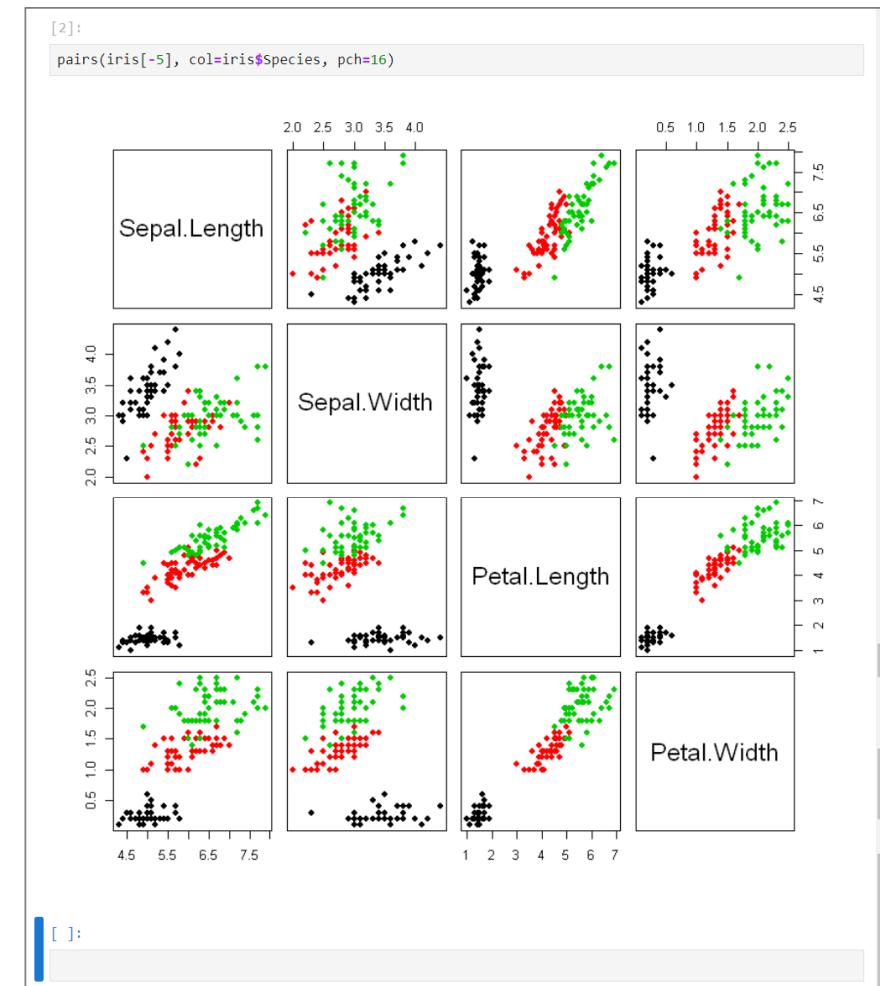
JupyterLab R

Title: Jupyter Notebook + R

Author: Ming-Chang, Lee

```
[1]:  
head(iris, n=3)
```

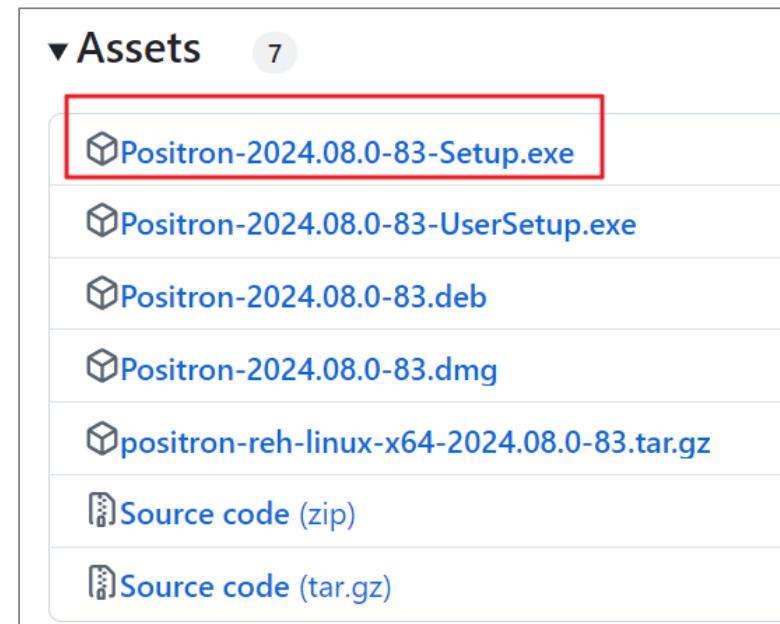
| Sepal.Length | Sepal.Width | Petal.Length | Petal.Width | Species |
|--------------|-------------|--------------|-------------|---------|
| 5.1 | 3.5 | 1.4 | 0.2 | setosa |
| 4.9 | 3.0 | 1.4 | 0.2 | setosa |
| 4.7 | 3.2 | 1.3 | 0.2 | setosa |



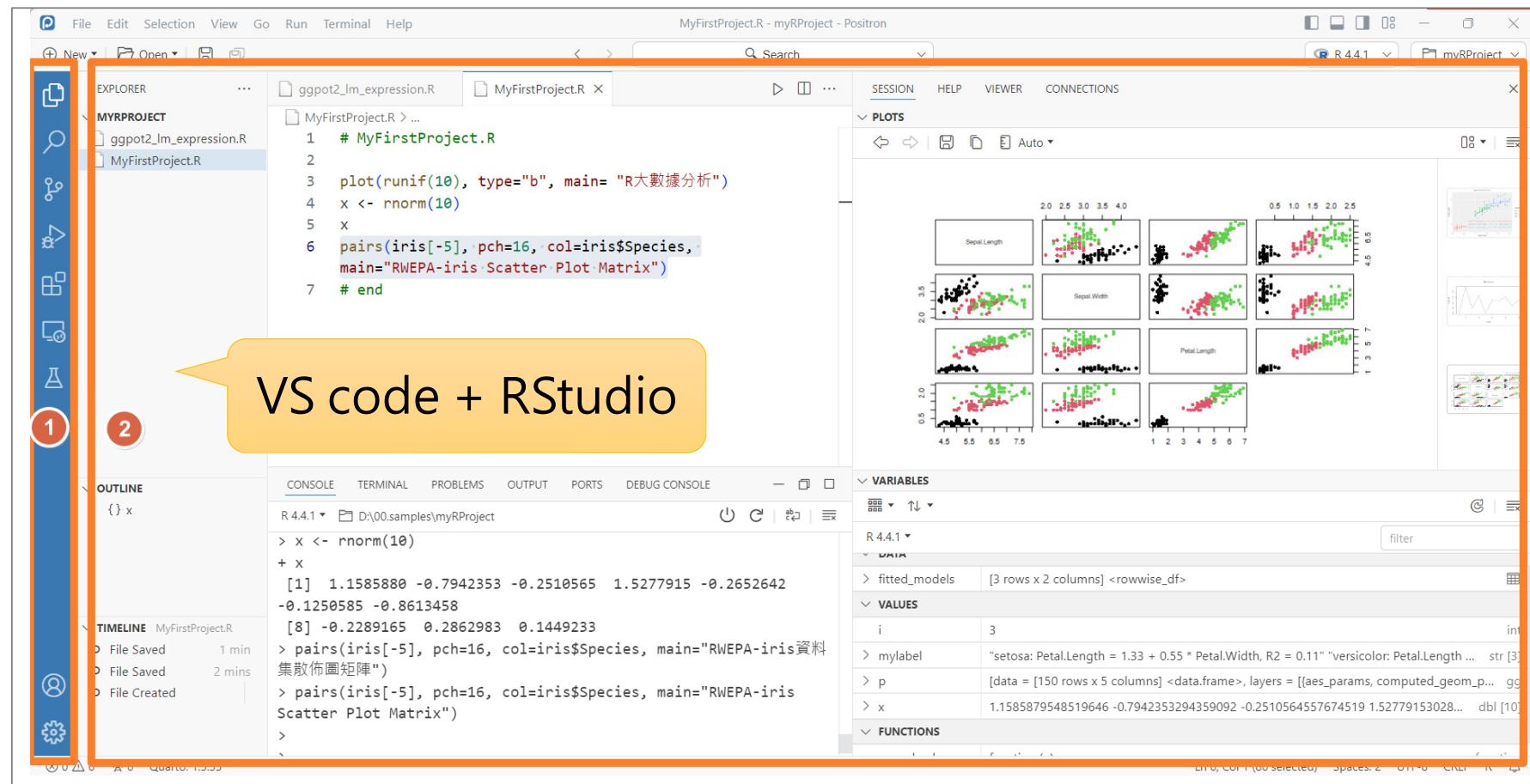
Positron (整合 R, Python, Jupyter)

Positron IDE, Since 2024.6

- 首頁: <https://github.com/posit-dev/positron>
- 下載: <https://github.com/posit-dev/positron/releases>



Positron+R 執行畫面



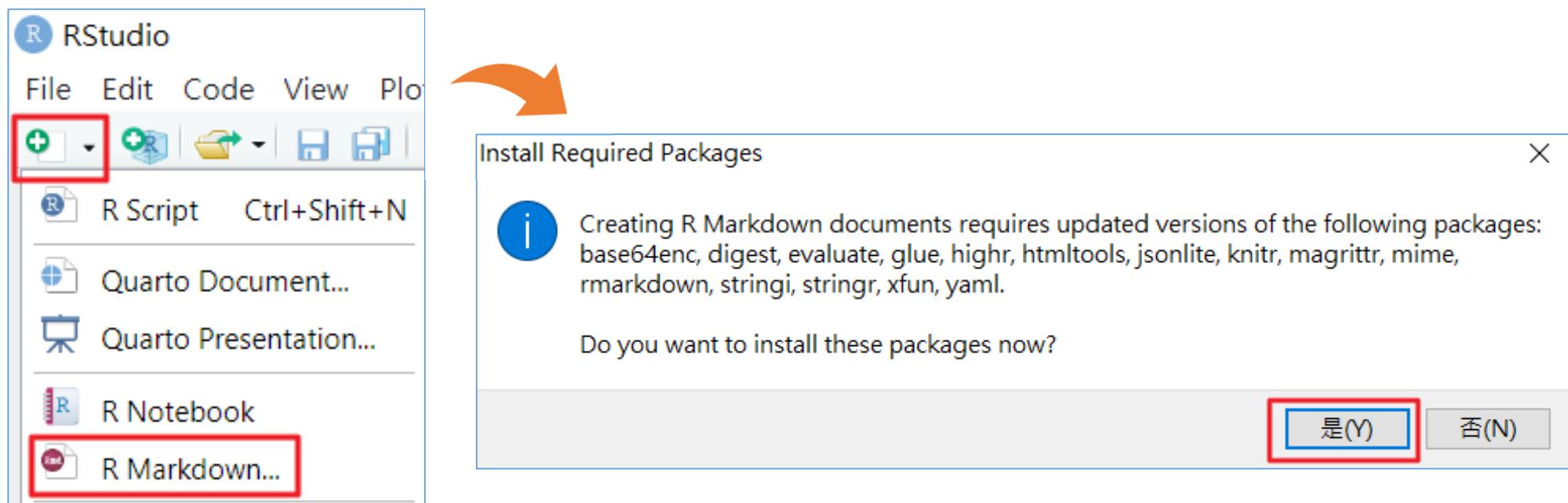


R Markdown

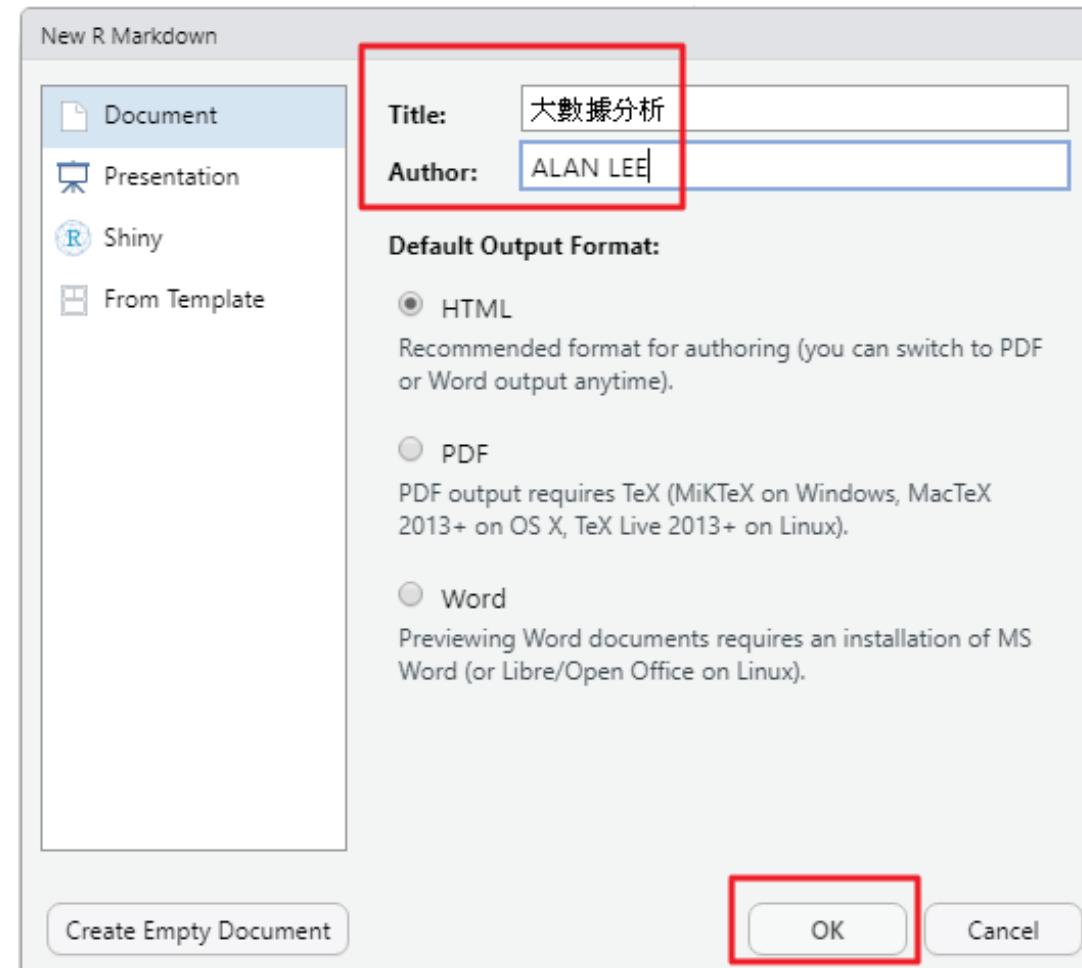
(R 標記語言)

新增 R Markdown 檔案

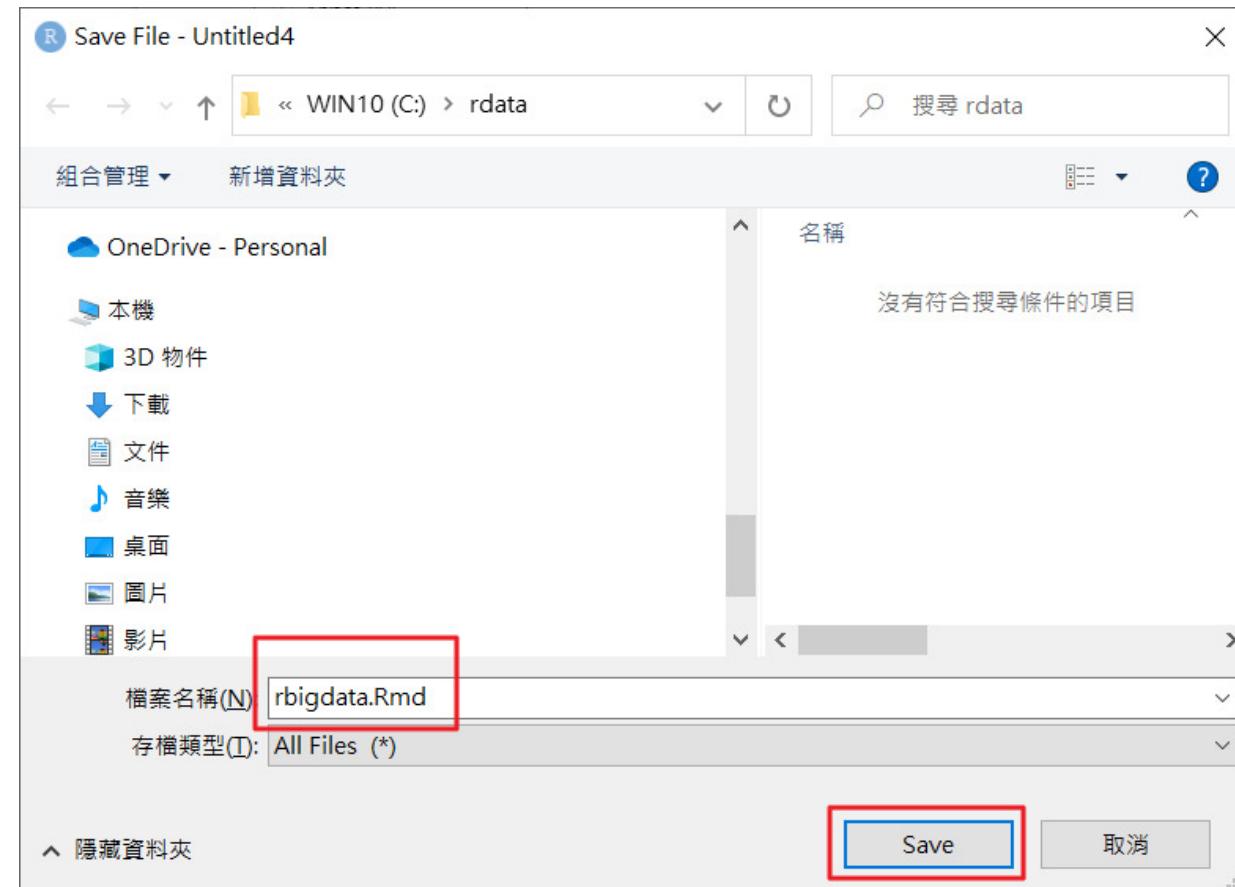
-  New File \ R Markdown
- Install Required Packages \ 是(Y)



New R Markdown



儲存 rbigdata.Rmd



The screenshot shows the RStudio interface with several help-related features highlighted:

- 1. Help Tab:** The "Help" tab in the top navigation bar is highlighted with a red box and a red circle containing the number 1.
- 2. Help Icon:** The "Help" icon in the bottom navigation bar (next to "Plots" and "Packages") is highlighted with a red box and a red circle containing the number 2.
- 3. Posit Cheat Sheets:** In the bottom right corner, there is a sidebar titled "RStudio" with a yellow callout pointing to the "Posit Cheat Sheets" link. This link is highlighted with a red box and a red circle containing the number 3.

RStudio Help Resources:

- Posit Support
- Posit Community Forum for the RStudio IDE
- Posit Cheat Sheets
- RSudio Packages
- Posit Products

RStudio Environment View:

Environment is empty

RStudio Console Output:

```
R version 4.4.2  
e"  
Copyright (C) 2023 The R Foundation for Statistical Computing  
Platform: x86_64-w64-mingw32/x64 (64-bit)  
  
R is free software.  
Y.  
You are welcome to redistribute it.  
Type 'license' or 'copyright' for more information.
```

RStudio Project View:

rbigdata.Rmd

Source Visual

```
1 ---  
2 title: "大數據分析"  
3 author: "ALAN LEE"  
4 date: "2024-09-04"  
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,
```

Console Terminal × Background Jobs ×

R 4.4.1 · ~/

Project: (None)

Environment History Connections Tutorial

Import Dataset 203 MiB

Global Environment

Outline

Knit on Save ABC Knit

File Edit Code View Plots Session Build Debug Profile Tools Help

Files Plots Packages Help Viewer Presentation

Find in Topic

Resources

News Overflow Help with R

Posit Support

Posit Community Forum for the RStudio IDE

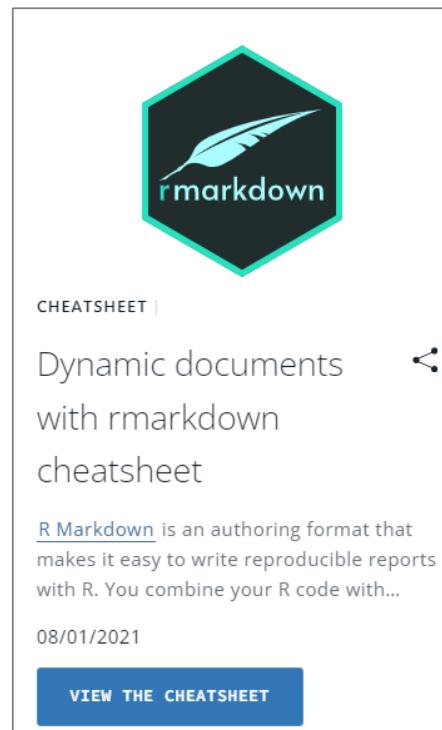
Posit Cheat Sheets

RSudio Packages

Posit Products

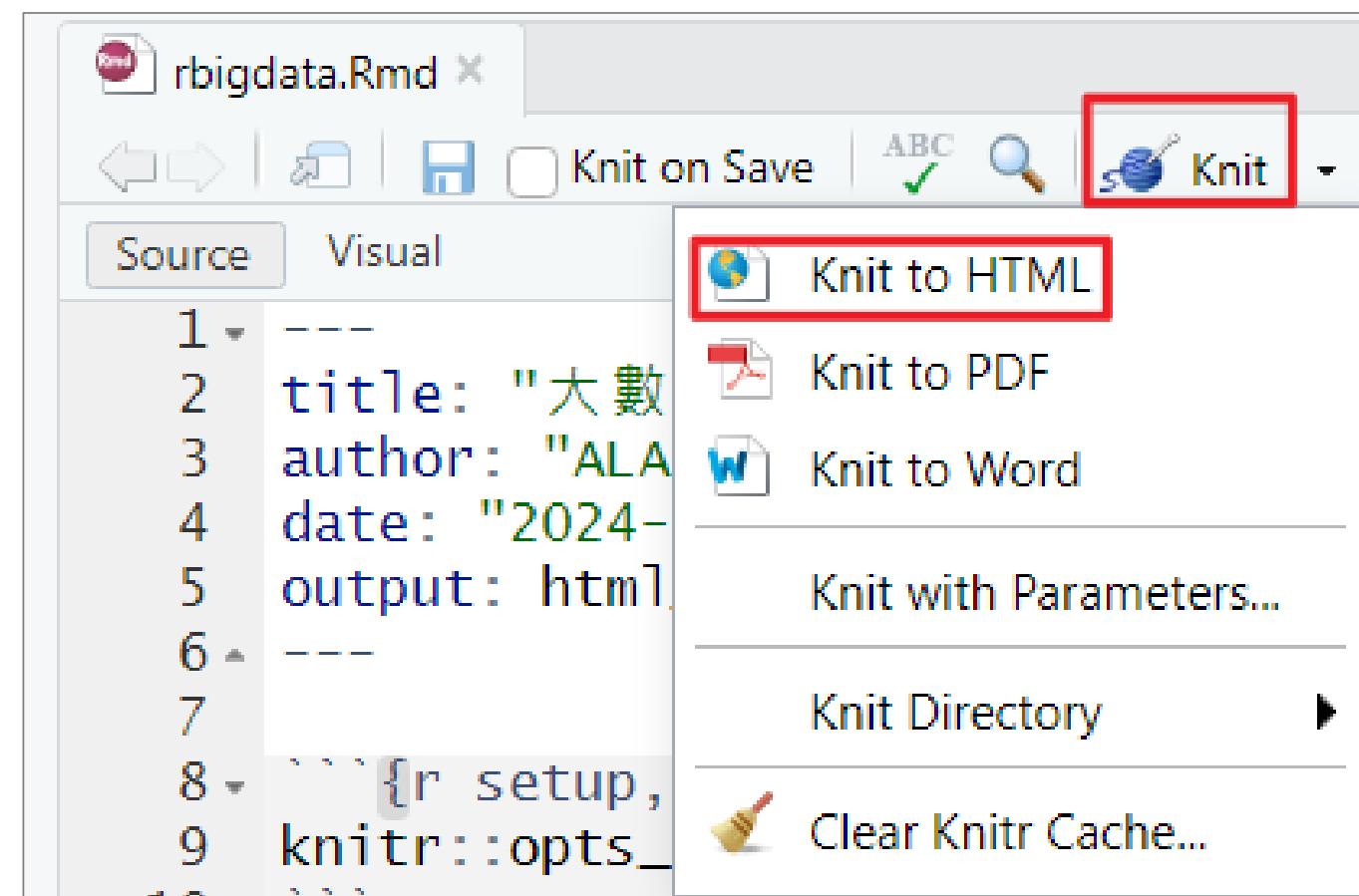
R Markdown Cheatsheet 線上說明

- <https://www.rstudio.com/resources/cheatsheets/>
- <https://rstudio.github.io/cheatsheets/html/rmarkdown.html>



R Markdown (續)

- Knit to HTML
- Knit to PDF
- Knit to Word



實作
練習

R Markdown: HTML

① 檔案 C:/rdata/rbigdata.html 輸出HTML

1 標題
大數據分析
ALAN LEE
2024-09-04

2 文字
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. :25.0 Max. :120.00

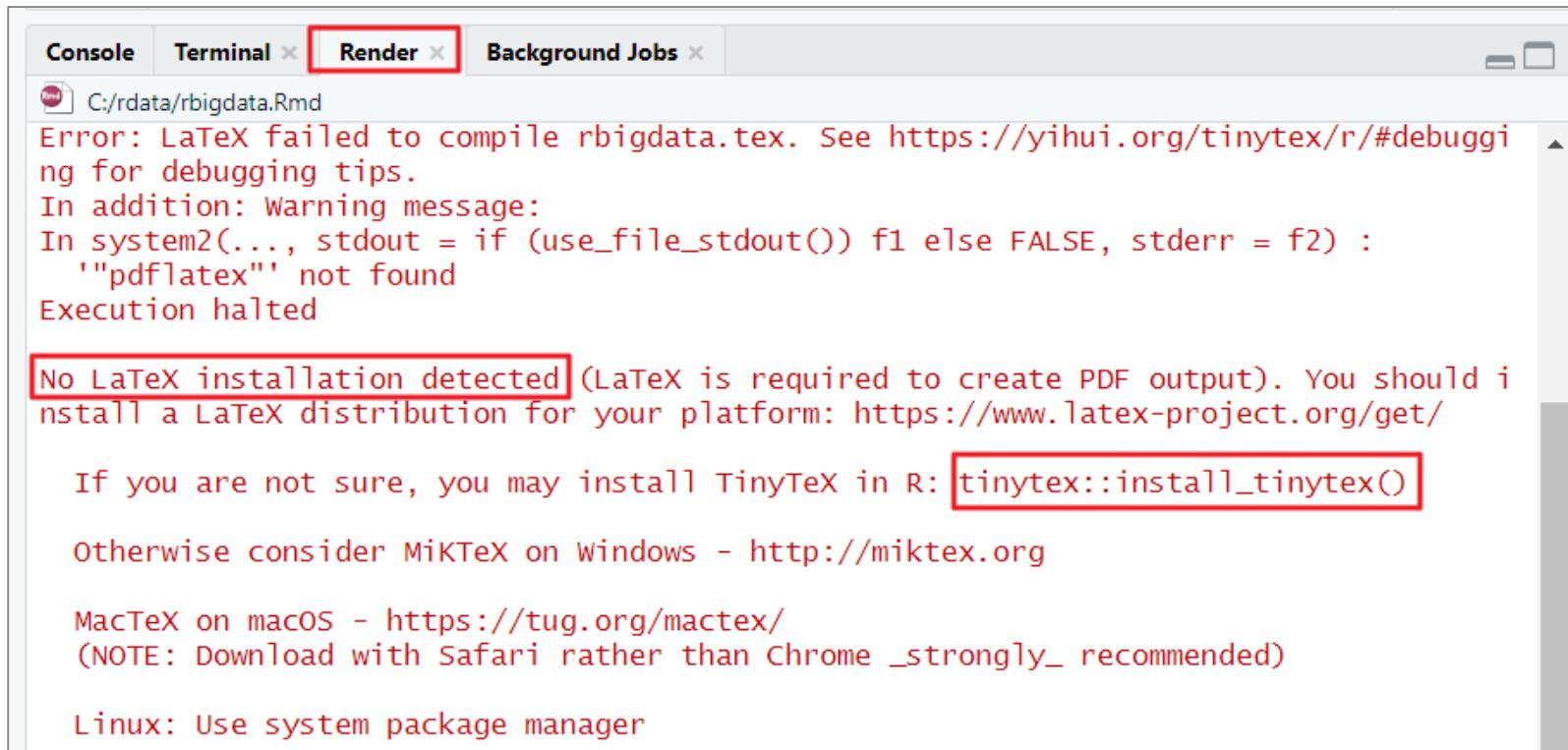
3 R 程式碼與執行結果
Including Plots
You can also embed plots, for example:

Note that the echo = FALSE parameter was added to the code chunk to prevent printing of the R code that generated the plot.

熟悉 .rmd 架構

R Markdown: PDF

- Error: 沒有安裝 LaTeX



The screenshot shows the RStudio interface with the 'Render' tab selected. A red box highlights the 'Render' tab in the top navigation bar. The main pane displays the following error message:

```
C:/rdata/rbigdata.Rmd
Error: LaTeX failed to compile rbigdata.tex. See https://yihui.org/tinytex/r/#debugging for debugging tips.
In addition: Warning message:
In system2(..., stdout = if (use_file_stdout()) f1 else FALSE, stderr = f2) :
  '"pdflatex"' not found
Execution halted

No LaTeX installation detected (LaTeX is required to create PDF output). You should install a LaTeX distribution for your platform: https://www.latex-project.org/get/

If you are not sure, you may install TinyTeX in R: tinytex::install\_tinytex\(\)

Otherwise consider MiKTeX on Windows - http://miktex.org

MacTeX on macOS - https://tug.org/mactex/
(NOTE: Download with Safari rather than Chrome strongly recommended)

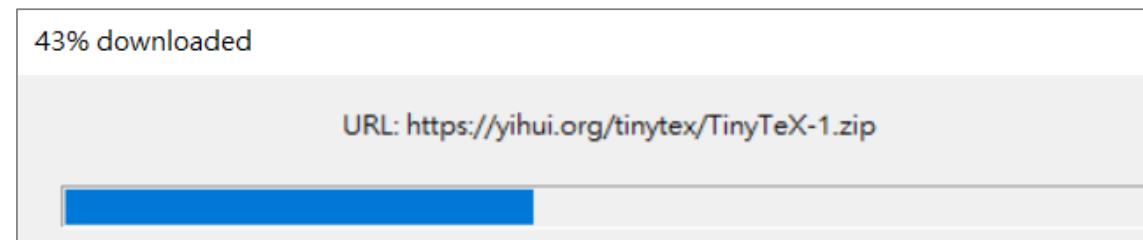
Linux: Use system package manager
```

The 'No LaTeX installation detected' message and the 'tinytex::install_tinytex()' link are also highlighted with red boxes.

R Markdown: PDF – 解決方法

- 方法1: `tinytex::install_tinytex()` ← 優先使用方法

```
> tinytex::install_tinytex()
trying URL 'https://yihui.org/tinytex/TinyTeX-1.zip'
Content type 'application/octet-stream' length 137692855 bytes (131.3 MB)
```



- 方法2: 下載 Miktex: <https://miktex.org/download>
 - basic-miktex-21.2-x64.exe (243.83MB)

Execution halted

processing file: rbigdata.Rmd

```
"C:/Program Files/RStudio/resources/app/bin/quarto/bin/tools/pandoc" +RTS -K512m -RTS r  
bigdata.knit.md --to latex --from markdown+autolink_bare_uris+tex_math_single_backslash  
--output rbigdata.tex --lua-filter "C:\Users\IEUser\AppData\Local\R\win-library\4.4\rma  
rmarkdown\rmarkdown\lua\pagebreak.lua" --lua-filter "C:\Users\IEUser\AppData\Local\R\win-l  
ibrary\4.4\rmarkdown\rmarkdown\lua\latex-div.lua" --embed-resources --standalone --high  
light-style tango --pdf-engine pdflatex --variable graphics --variable "geometry:margin  
=1in"
```

output file: rbigdata.knit.md

```
! LaTeX Error: Unicode character å¤§ (U+5927)  
not set up for use with LaTeX.
```

中文編碼錯誤

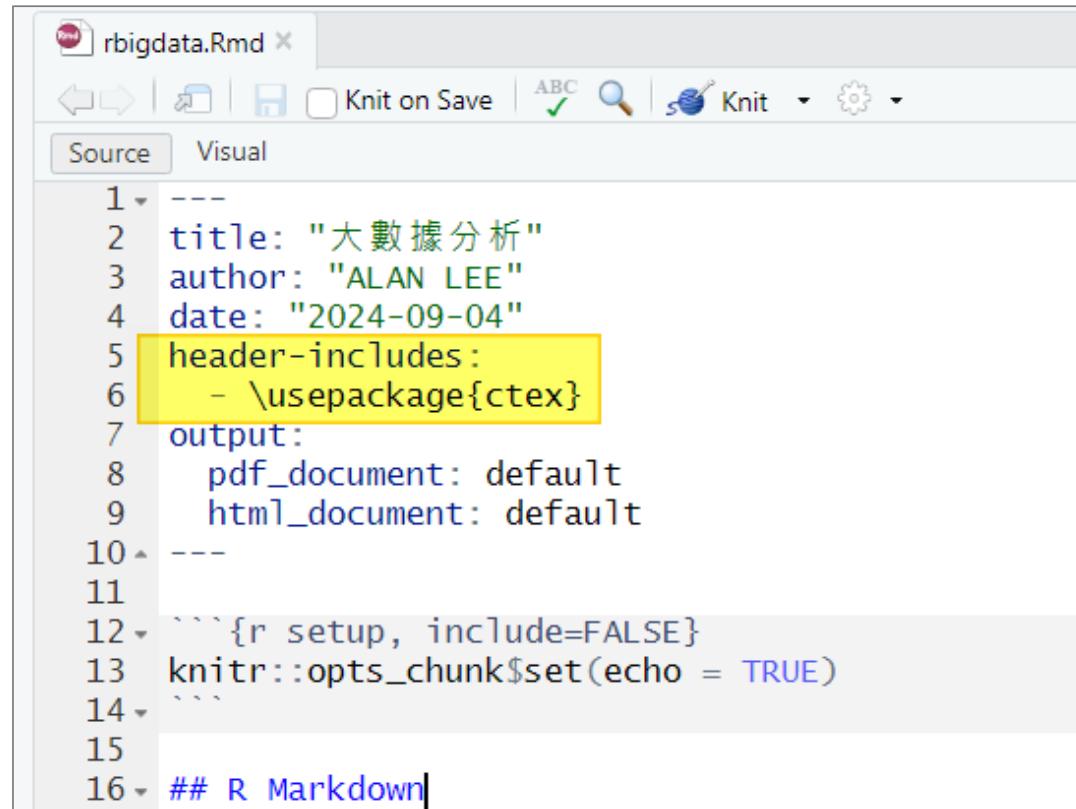
Try other LaTeX engines instead (e.g., xelatex) if you are using pdflatex. See <https://bookdown.org/yihui/rmarkdown-cookbook/latex-unicode.html>

Error: LaTeX failed to compile rbigdata.tex. See <https://yihui.org/tinytex/r/#debugging> for debugging tips. See rbigdata.log for more info.

Execution halted

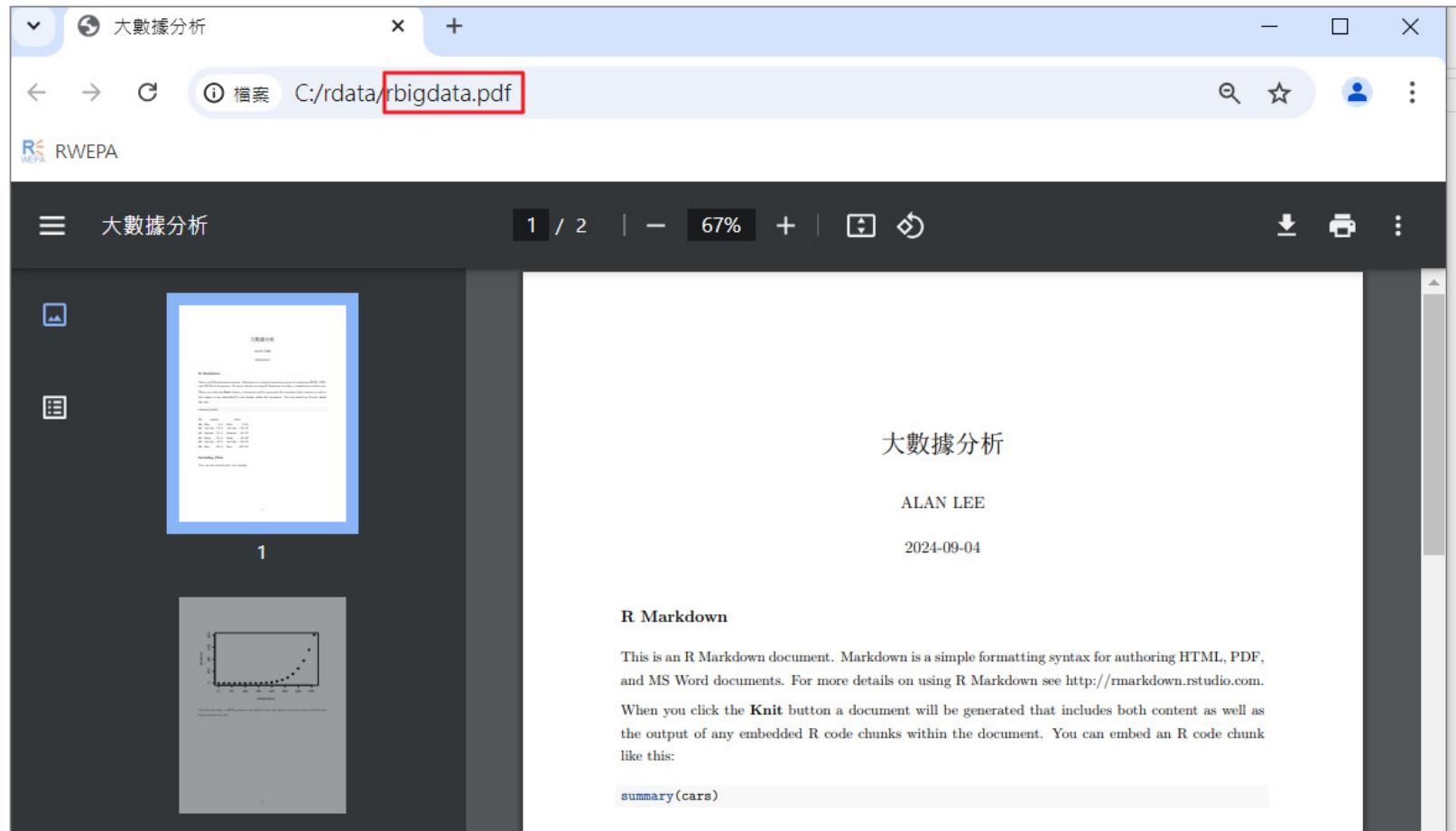
R Markdown: PDF – 中文設定方法

- https://github.com/rwepa/ipas_bda/blob/main/ipas-r-program.R#L1717

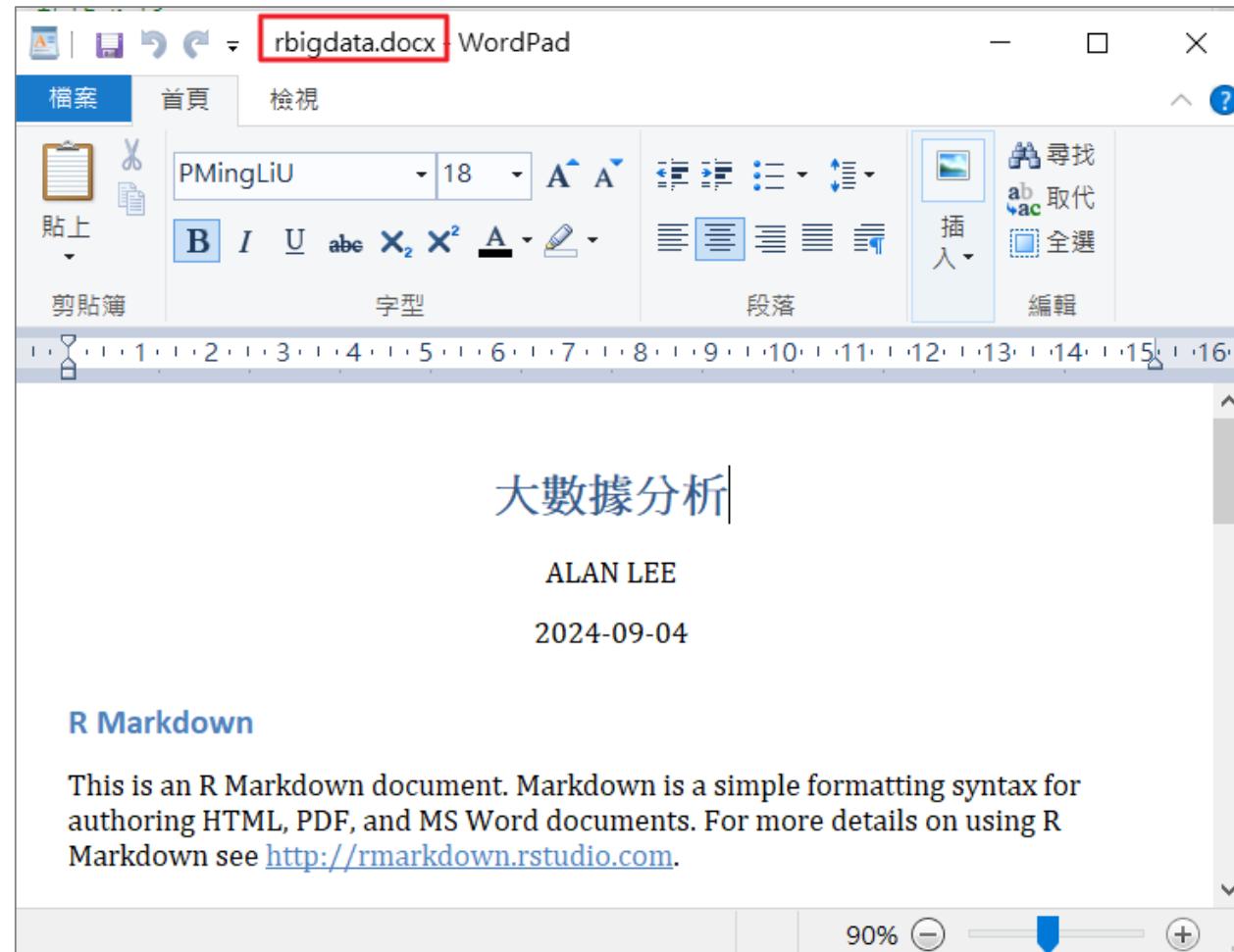


```
rbigdata.Rmd x
Source Visual
1 ---  
2 title: "大數據分析"  
3 author: "ALAN LEE"  
4 date: "2024-09-04"  
5 header-includes:  
6   - \usepackage{ctex}  
7 output:  
8   pdf_document: default  
9   html_document: default  
10 ---  
11  
12 ```{r setup, include=FALSE}  
13 knitr::opts_chunk$set(echo = TRUE)  
14 ```  
15  
16 ## R Markdown|
```

R Markdown: PDF – 完成版



R Markdown: Word – 完成版



Quarto

(R Markdown 進階版 – for R and Python)

Quarto, Since 2022.7

- Quarto: RMarkdown 的下一代產品.
- 下載: <https://quarto.org/>

The screenshot shows the Quarto website homepage. At the top, there's a navigation bar with links for Overview, Get Started, Guide, Extensions, Reference, Gallery, Blog, and Help. The main heading is "Welcome to Quarto" followed by the subtitle "An open-source scientific and technical publishing system". Below this, there's a bulleted list of features:

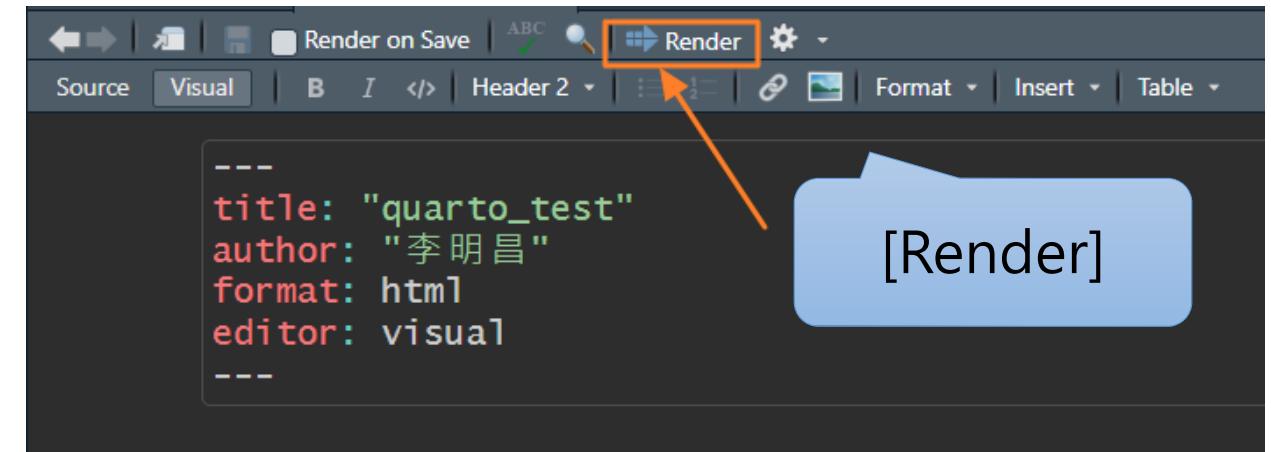
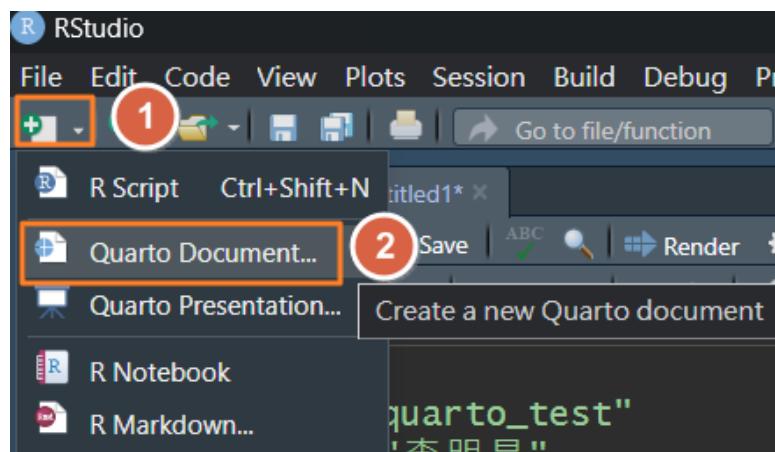
- Author using Jupyter notebooks or with plain text markdown in your favorite editor.
- Create dynamic content with Python, R, Julia, and Observable.
- Publish reproducible, production quality articles, presentations, dashboards, websites, blogs, and books in HTML, PDF, MS Word, ePub, and more.
- Share knowledge and insights organization-wide by publishing to Posit Connect, Confluence, or other publishing systems.
- Write using Pandoc markdown, including equations, citations, crossrefs, figure panels, callouts, advanced layout, and more.

At the bottom, there's a tagline: "Analyze. Share. Reproduce. You have a story to tell with data—tell it with Quarto." Two buttons are at the very bottom: "Get Started" and "Guide". To the right of the main content area, there's a sidebar titled "Publishing Options" with icons for Jupyter, X (likely referring to R), and R. A red box highlights the "Jupyter" icon. To the right of this sidebar is a list of publishing options:

- Websites / Blogs
- Dashboards
- Articles
- Presentations
- Books
- Knowledge Repos

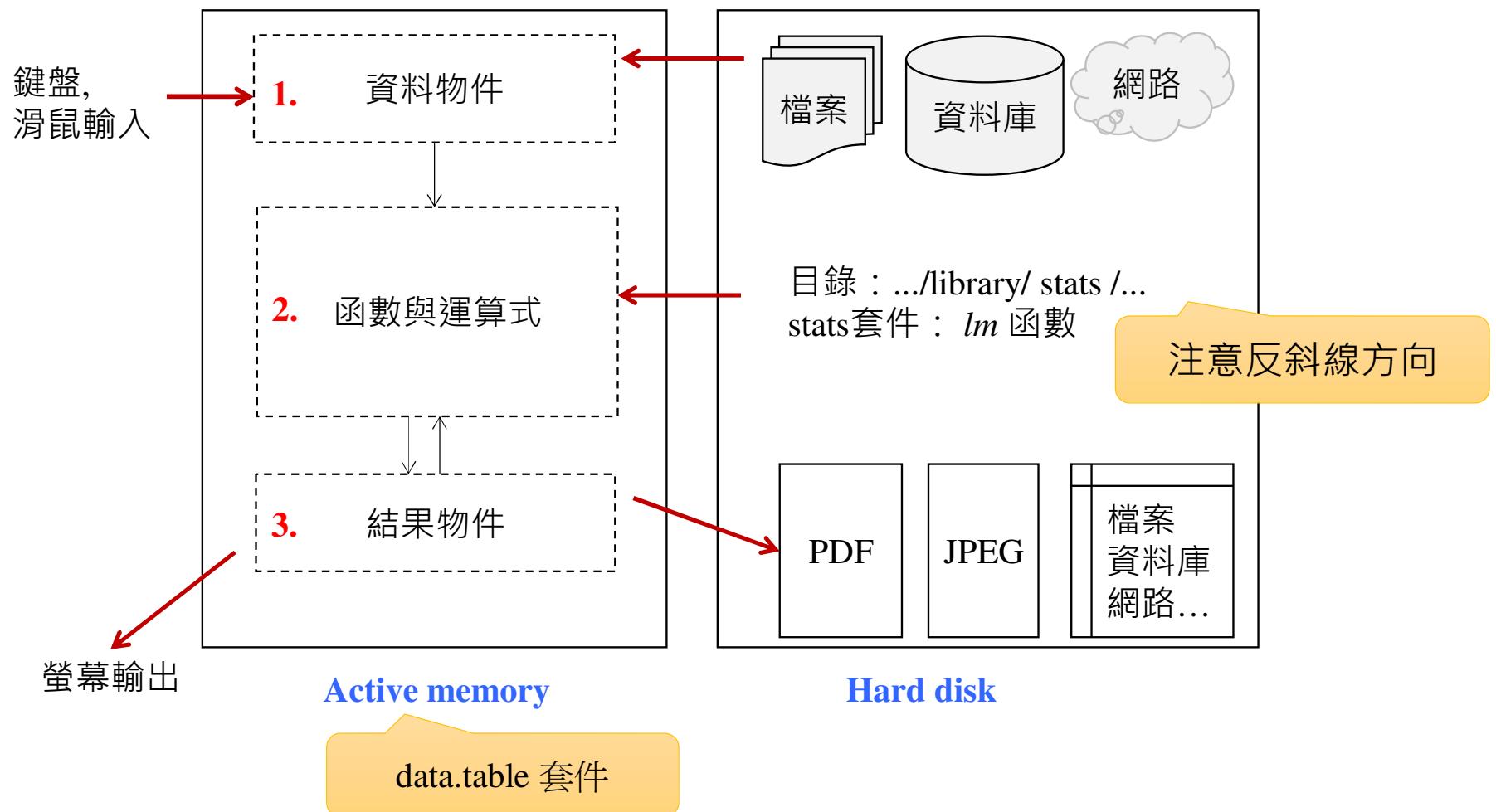
Quarto

- An open-source scientific and technical publishing system
- File \ New File \ Quarto Document...
- File \ Save As \ XXX.qmd → Render 建立 HTML , PDF, Word, ...



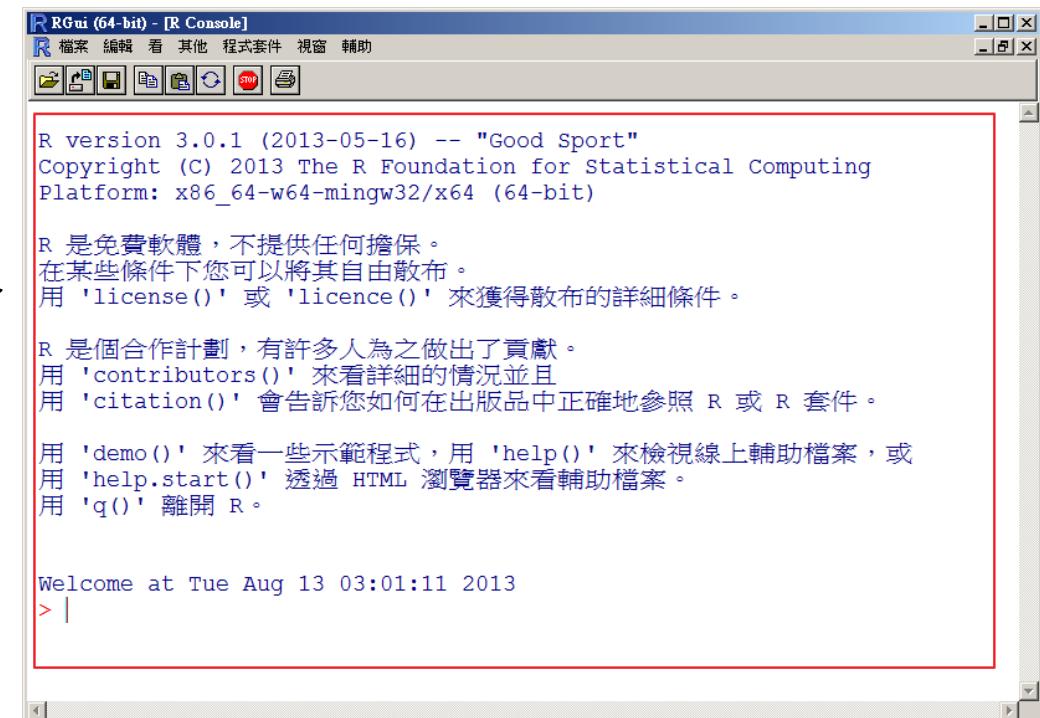
R 基礎操作

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



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)
- \$ 符號 資料物件\$變數名稱
- 結果[] 顯示編號

Python: 資料物件.變數名稱

```
> iris$Sepal.Length
[1] 5.1 4.9 4.7 4.6 5.0 5.4 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
[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
[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
[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 ...
[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
```

- 取出第145個值
iris\$Sepal.Length[145]

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), 建議少用句點。
- 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), 下底線 _ 皆可。
- 物件名稱中間不可有空格。

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



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

type="n"

plot.default {graphics}

2.套件

1

2

The Default Scatterplot Function

3

R Documentation

3.R文件

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,  
      ...)
```

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.

4.簡單說明

5

5.詳細說明

6

6.方法

7

7.參數

套件 Package

(Python: 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"
>
```

在输出结果中，"package:graphics" 和 "package:datasets" 两个项被红色方框高亮显示，并且有一个指向 "package:datasets" 的红色箭头。

"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 (2024.9.5)

Available Packages

Currently, the CRAN package repository features 21240 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

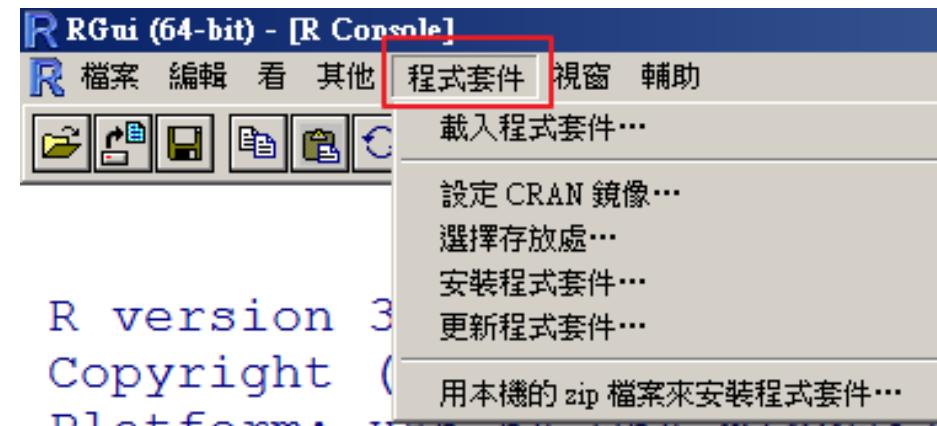
CRAN Task View:
<https://cran.csie.ntu.edu.tw/web/views/>

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"
```

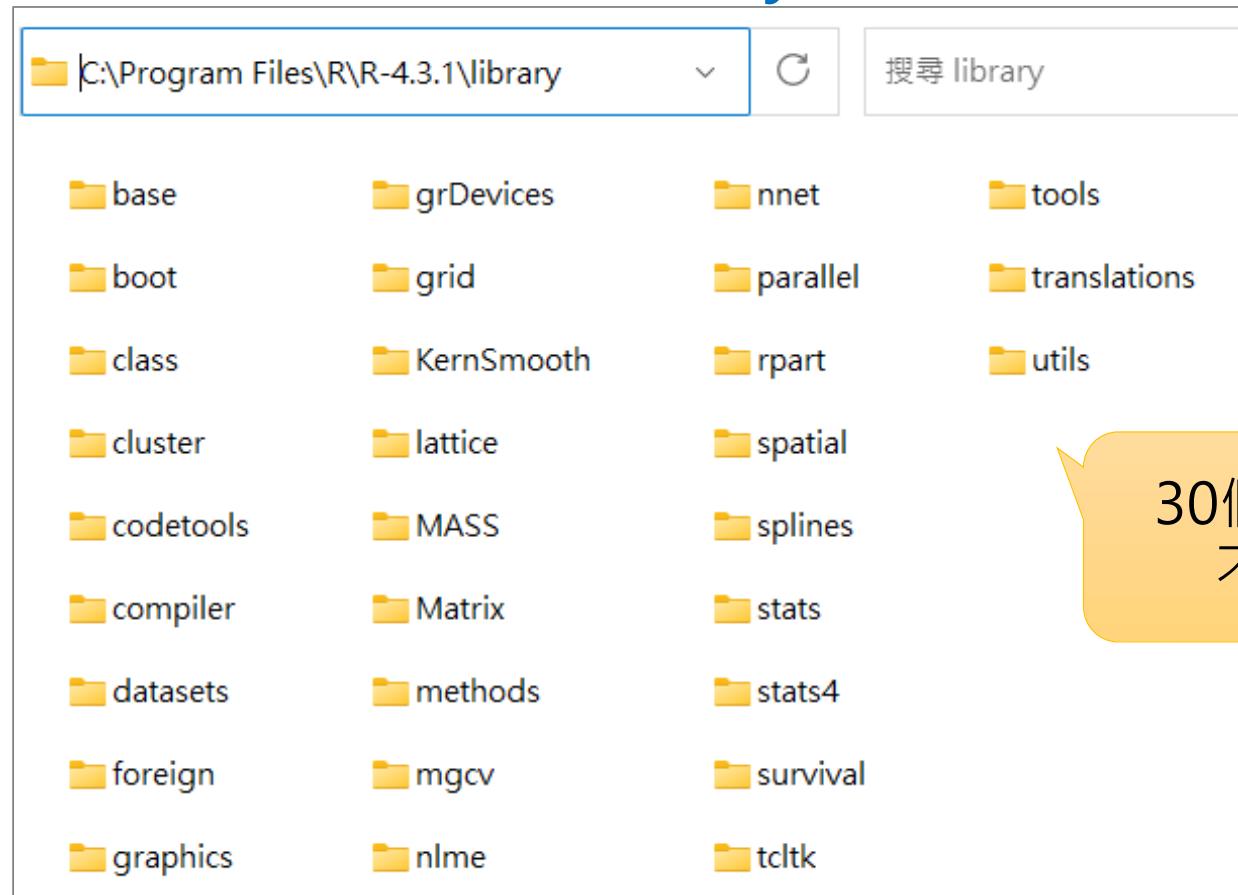
已安裝套件

```
> # 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"
                                         "airGR"    "airGRteaching
                                         "airGR"   "airGRteaching"
```

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

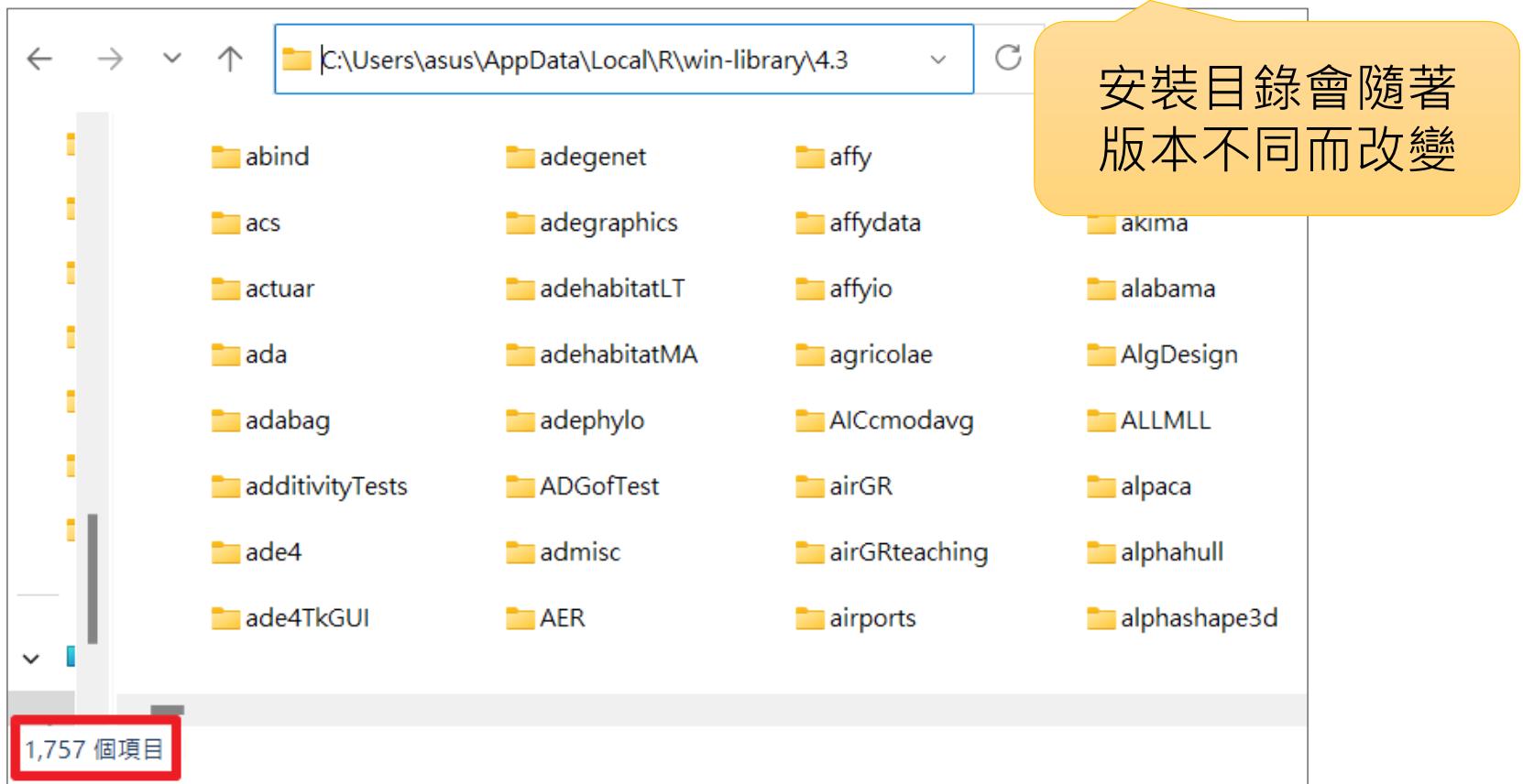
套件安裝目錄1

- 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 |
|--|--|----------|---------|
| Maintained by 'YuLab-SMU' | | | |
| <input type="checkbox"/> zeallot | Multiple, Unpacking, and Destructuring Assignment | 0.1.0 | |
| <input type="checkbox"/> zip | Cross-Platform 'zip' Compression | 2.3.0 | |
| <input type="checkbox"/> zlibbioc | An R packaged zlib-1.2.5 | 1.46.0 | |
| <input type="checkbox"/> zoo | S3 Infrastructure for Regular and Irregular Time Series (Z's Ordered Observations) | 1.8-12 | |
| 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

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(1/3)` 判定是否為有限的
- `is.infinite(Inf)` 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

1.2 開放資料實務應用

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"          先建立 C:\rdata 資料夾  
> setwd(workpath) 2  
  
>  
> # 已更改變為 "c:/rdata" 工作目錄  
> getwd()  
[1] "c:/rdata" 3
```

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

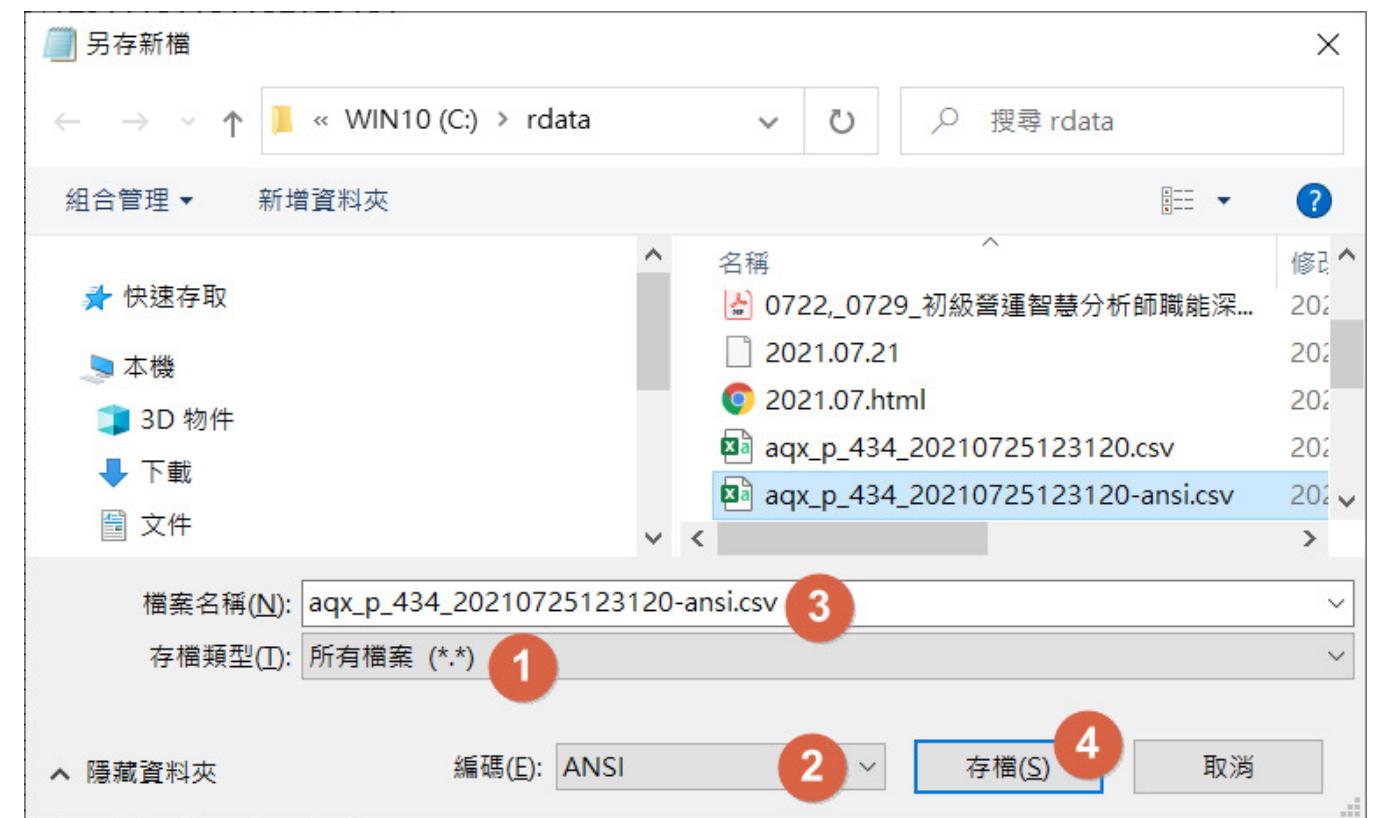
R-v3.x早期版本匯入 UTF-8 文字檔會有錯誤

```

> myfile <- "aqx_p_434_20210725123120.csv"
>
1   > aq <- read.table(myfile, header=TRUE, sep=",") # error!
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    懶批<9c><92> 2021-07-24  28          2         1        NA
2       19 閩\u0080<e9>\u200d,2021-07-24          35          2         1        NA       25
3       20    搞喲\u200d 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-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 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
```

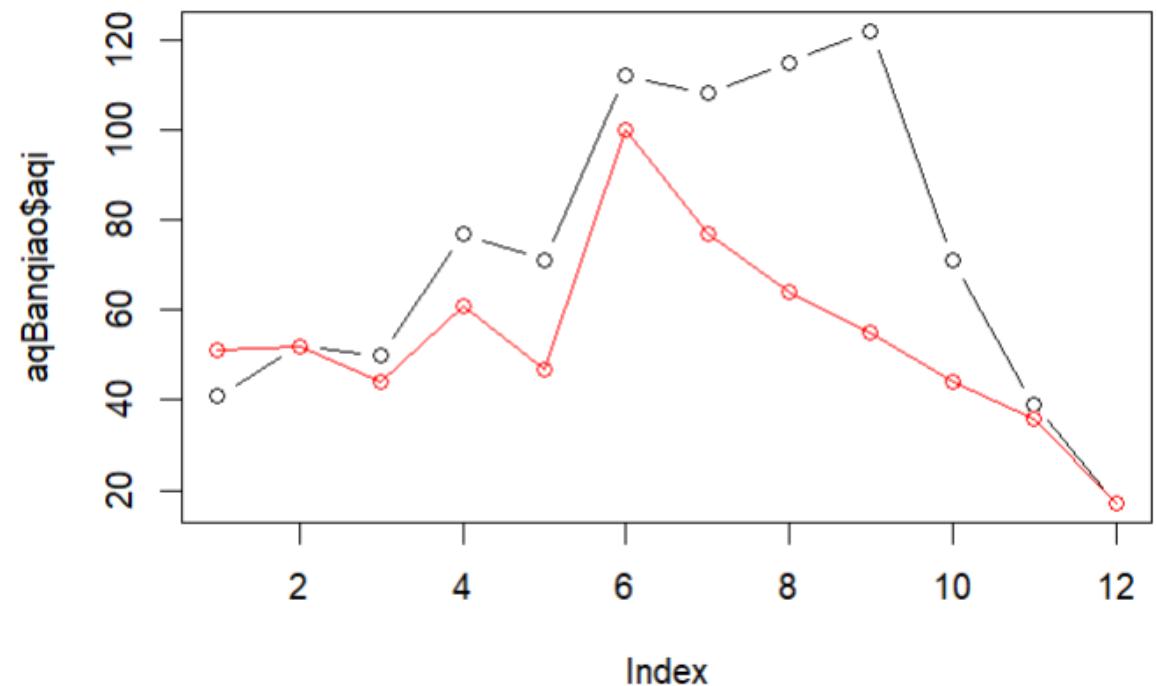
資料篩選與排序

`==` 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")
```



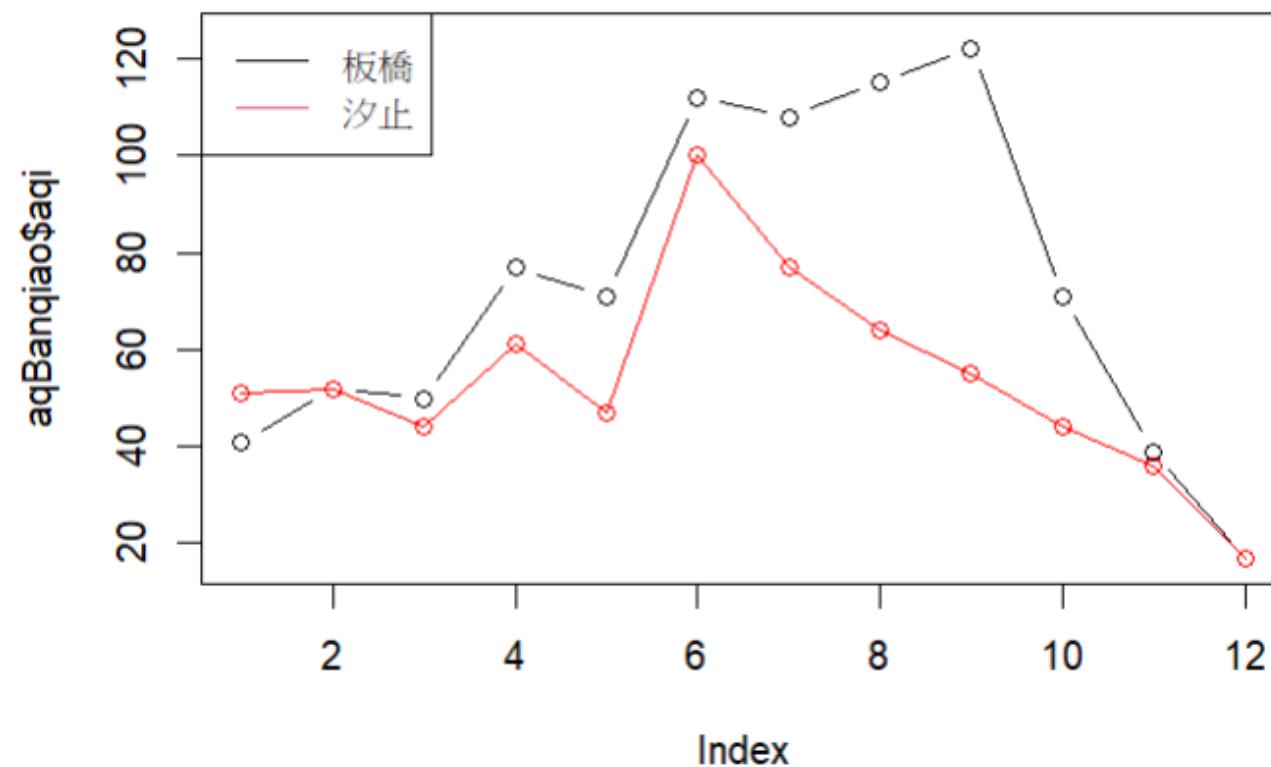
優化趨勢圖

```
> # 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

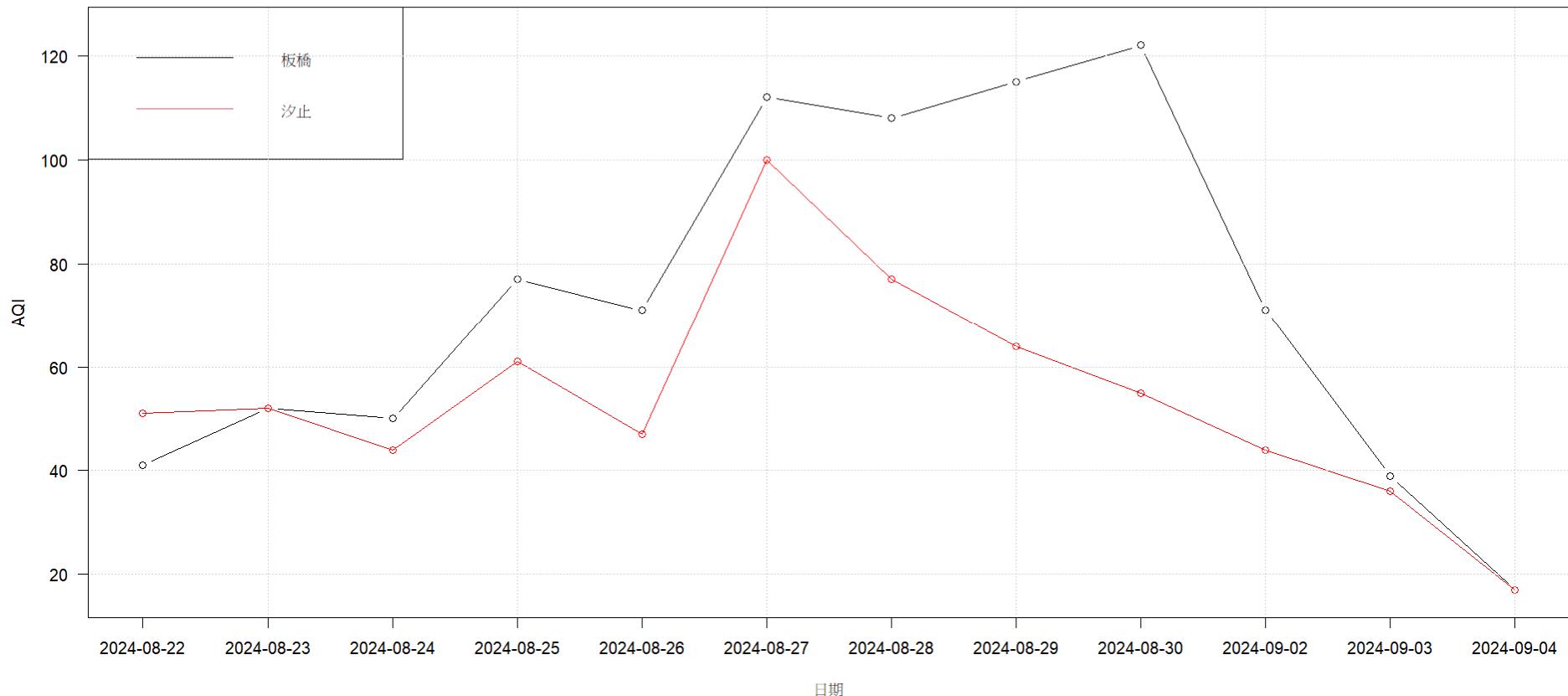
優化趨勢圖

2024-09-04 AQI 板橋vs.汐止



優化趨勢圖-revised

2024-09-04 AQI 板橋vs.汐止



步驟 5. 汇出資料

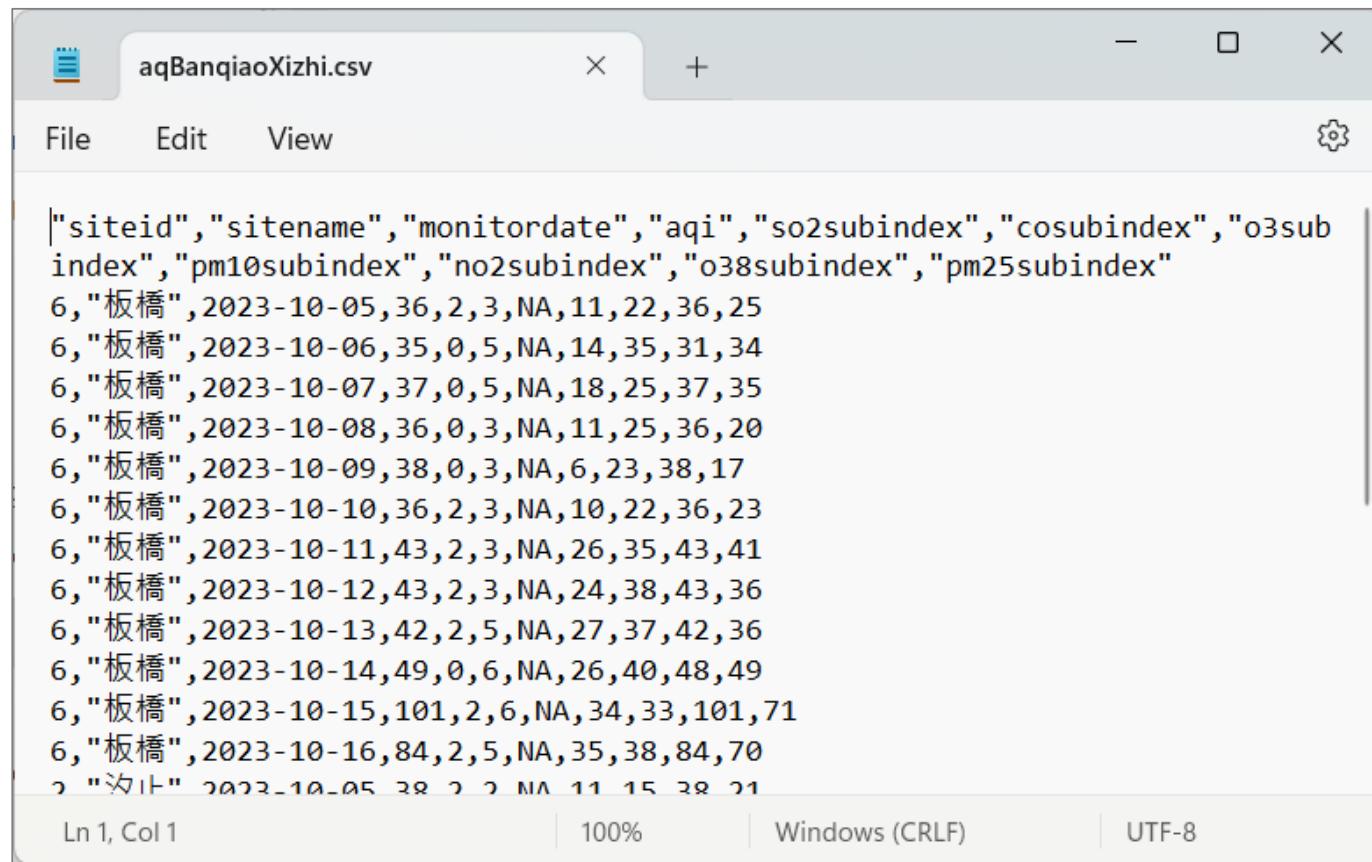
rbind: 列合併

```
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



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 | ?蹵? | 2023/10/5 | 36 | 2 | 3 | NA | | 11 | 22 | 36 |
| 3 | 6 | ?蹵? | 2023/10/6 | 35 | 0 | 5 | NA | | 14 | 35 | 31 |
| 4 | 6 | ?蹵? | 2023/10/7 | 37 | 0 | 5 | NA | | 18 | 25 | 37 |
| 5 | 6 | ?蹵? | 2023/10/8 | 36 | 0 | 3 | NA | | 11 | 25 | 36 |
| 6 | 6 | ?蹵? | 2023/10/9 | 38 | 0 | 3 | NA | | 6 | 23 | 38 |
| 7 | 6 | ?蹵? | 2023/10/10 | 36 | 2 | 3 | NA | | 10 | 22 | 36 |
| 8 | 6 | ?蹵? | 2023/10/11 | 43 | 2 | 3 | NA | | 26 | 35 | 43 |
| 9 | 6 | ?蹵? | 2023/10/12 | 43 | 2 | 3 | NA | | 24 | 38 | 43 |
| 10 | 6 | ?蹵? | 2023/10/13 | 42 | 2 | 5 | NA | | 27 | 37 | 42 |
| 11 | 6 | ?蹵? | 2023/10/14 | 49 | 0 | 6 | NA | | 26 | 40 | 48 |
| 12 | 6 | ?蹵? | 2023/10/15 | 101 | 2 | 6 | NA | | 34 | 33 | 101 |
| 13 | 6 | ?蹵? | 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 'Import Data' dialog box for the file 'aqBanqiaoXizhi.csv'. The '檔案原點' (File Origin) section has '65001: Unicode (UTF-8)' selected (highlighted with a red box). The '分隔符號' (Separator) is set to '逗號' (Comma). The '資料類型偵測' (Data Type Detection) section is set to '依據前 200 個列' (Based on first 200 columns). The preview area shows the following data:

| 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

R 連結資料庫

R 連結資料庫

- RStudio 連結MySQL
 -  YouTube (包括中文字幕) : <https://youtu.be/jsdM-y1nspQ>
 - <https://rwepa.blogspot.com/2024/05/using-rstudio-to-connect-mysql.html>
- 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套件: OCI Based Oracle Database Interface for R
- PostgreSQL
 - RPostgreSQL套件: R Interface to the PostgreSQL Database System
- SQLite
 - RSQLite套件: SQLite Interface for R

參考資料

- RWEPA
 - <http://rwepa.blogspot.com/>
- iPAS-R-tutorial
 - https://github.com/rwepa/ipas_bda/blob/main/ipas-r-program.R
- 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/>