

第2章 R語言簡介與基本運算功能

- R/Python/SQL程式設計與應用
(R/Python/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

主題

- 1. 2024/12/26(四) 第1章 教務管理規定、數據分析概論
- 2. 2025/01/02(四) 第2章 R語言簡介與基本運算功能 
- 3. 2025/01/07(二) 第3章 各種物件資料的運算與處理
- 4. 2025/01/22(三) 第4章 數據的整合、清理與轉換
- 5. 2025/02/04(二) 第4章 數據的整合、清理與轉換【補充篇】MySQL資料庫應用

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

大綱

- 2.1 R 簡介與安裝
- 2.2 RStudio 簡介與安裝
- 2.3 R基本概念
- 2.4 輔助說明
- 2.5 套件(Package)
- 2.6 開放資料分析



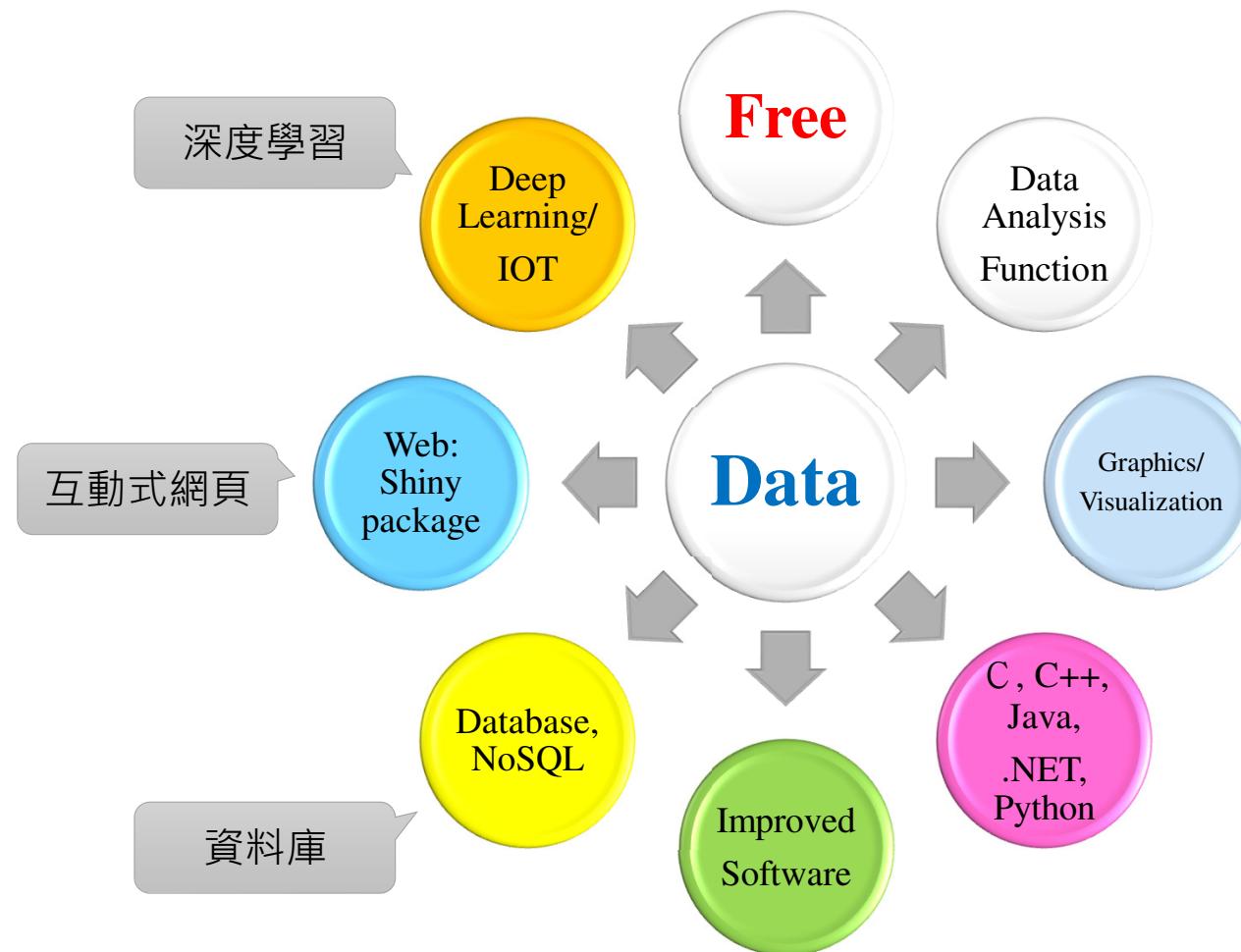
2.1 R 簡介與安裝

認識 R

- 1976 - 貝爾實驗室 John Chambers, Rick Becker, and Allan Wilks 研發 S 語言 - 統計分析程式語言。
- 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
 - 2004年10月 - R 2.0.0
 - 2013年4月 - R 3.0.0
 - 2024年10月 - R 4.4.2



R-八大功能



R官方網頁

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

The R Project for Statistical Computing

Getting Started

統計計算

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

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

..

R-下載

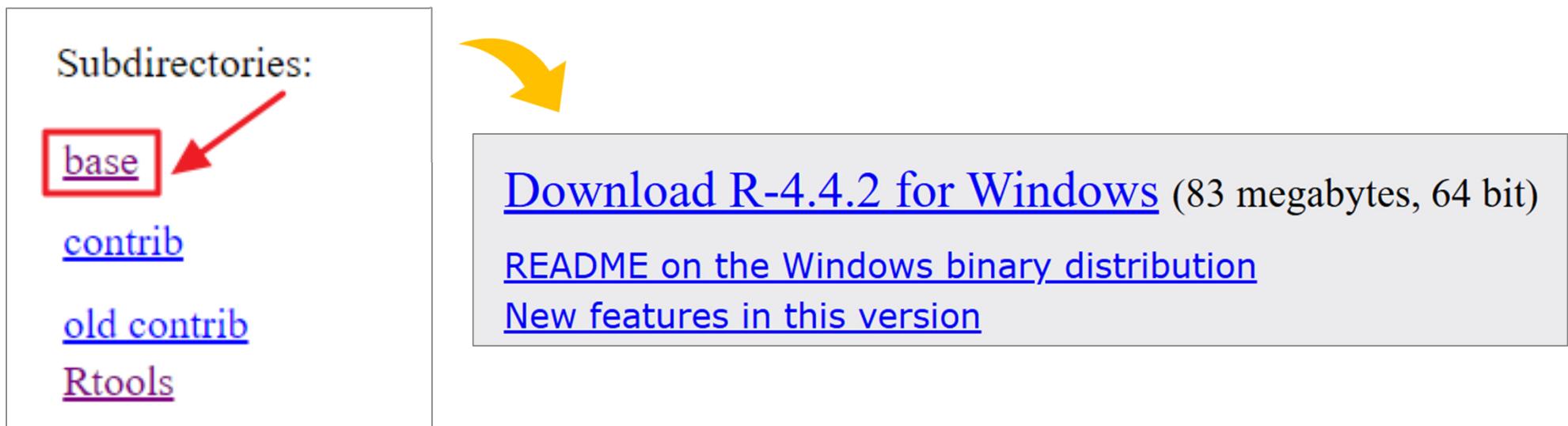
- 官網: <http://www.r-project.org/>
- 選取左側 Download \ CRAN
- 選取Taiwan CRAN: <https://cran.csie.ntu.edu.tw/>
- 選取 Download R for Windows

- [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-4.4.2-win.exe]



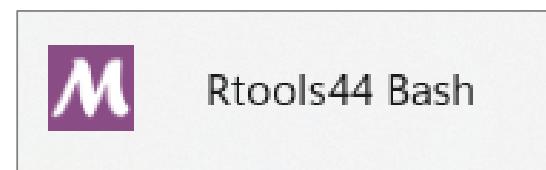
- R安裝路徑: 保留原路徑,不要修改
- 安裝參考說明, 2006
https://github.com/rwepa/DataDemo/blob/master/windows_intall_R.pdf

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

- 安裝完成：程式集 \ Rtoolsxx Bash



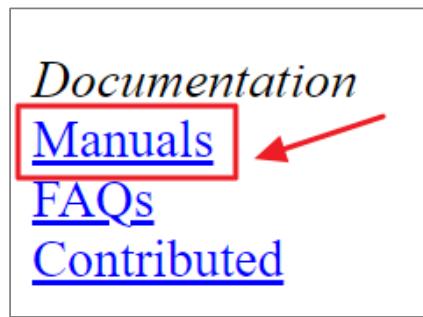
✓ 安裝 R

✓ 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)		

contributed documentation
(貢獻文件, 免費啦)

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

R Manuals (續)

Contributed Documentation

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

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

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

English Documents

Documents with more than 100 pages:

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

好書!

R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

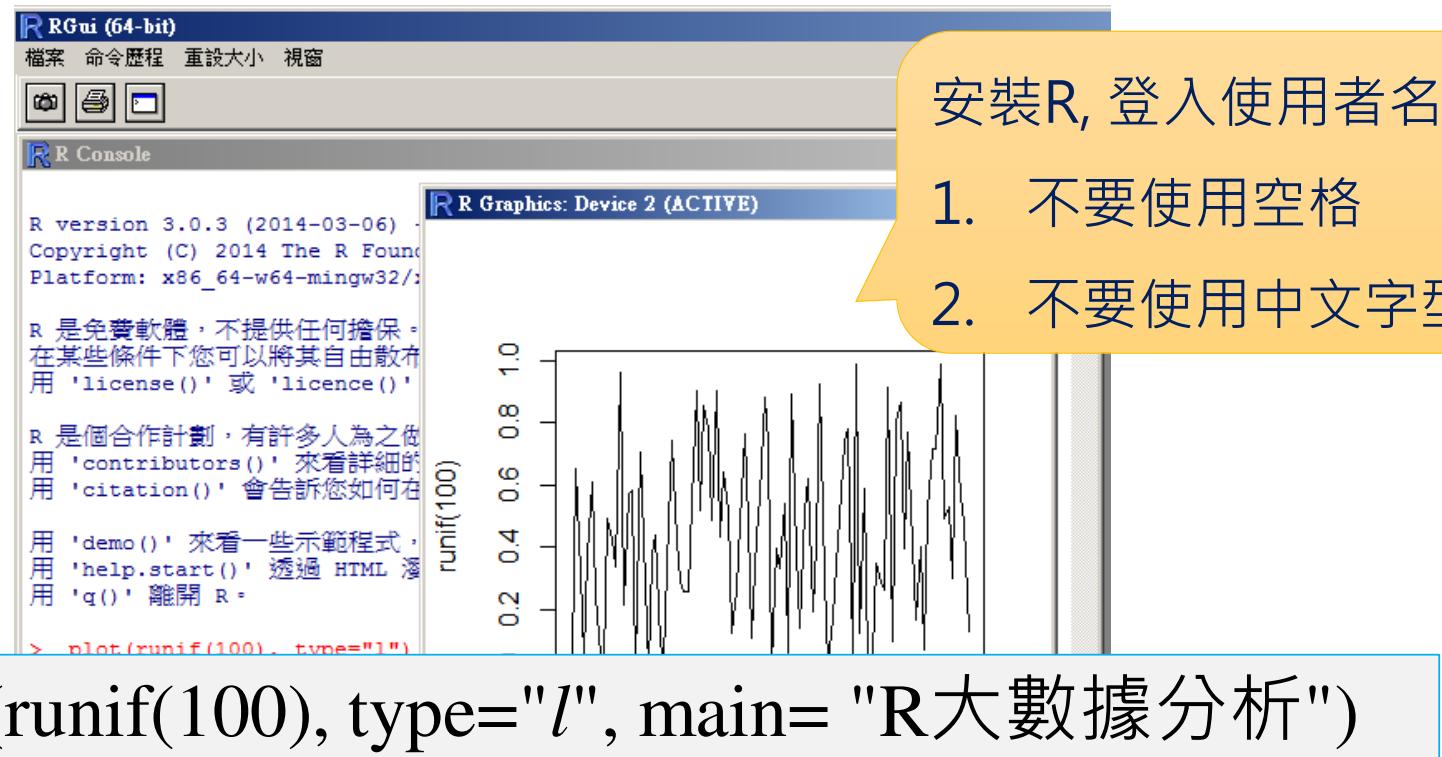
正確引用

R, ~~https://www.R-project.org/~~, 2024

錯誤引用

實作
練習

R 執行畫面



```
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 準
用 'q()' 離開 R。

> plot(runif(100), type="l")
R Graphics: Device 2 (ACTIVE)
```

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

demo(graphics)

demo(persp)

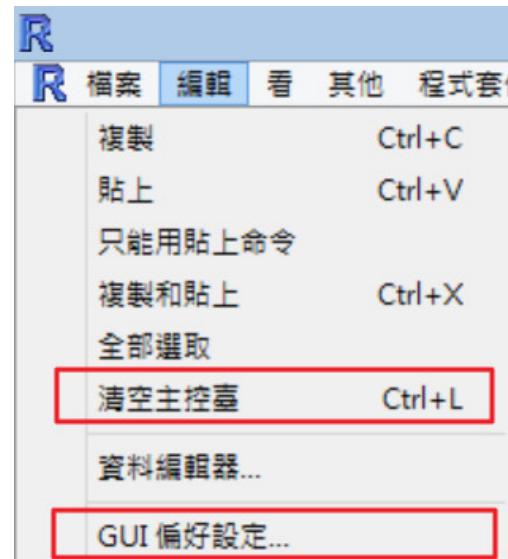
大小寫
須一致參考文獻:
citation()

R 功能表

檔案



編輯



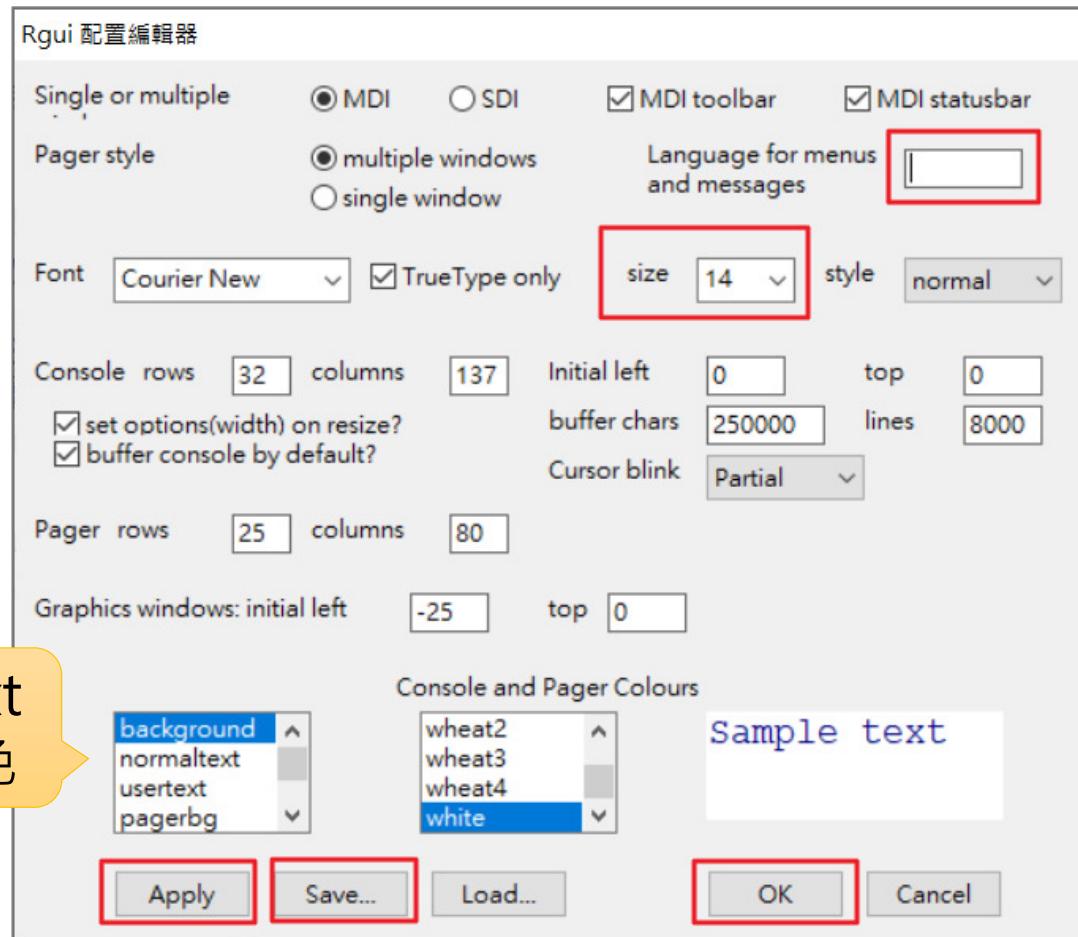
輔助



現行目錄 getwd()

儲存控制台-文字檔

編輯 \ GUI 偏好設定

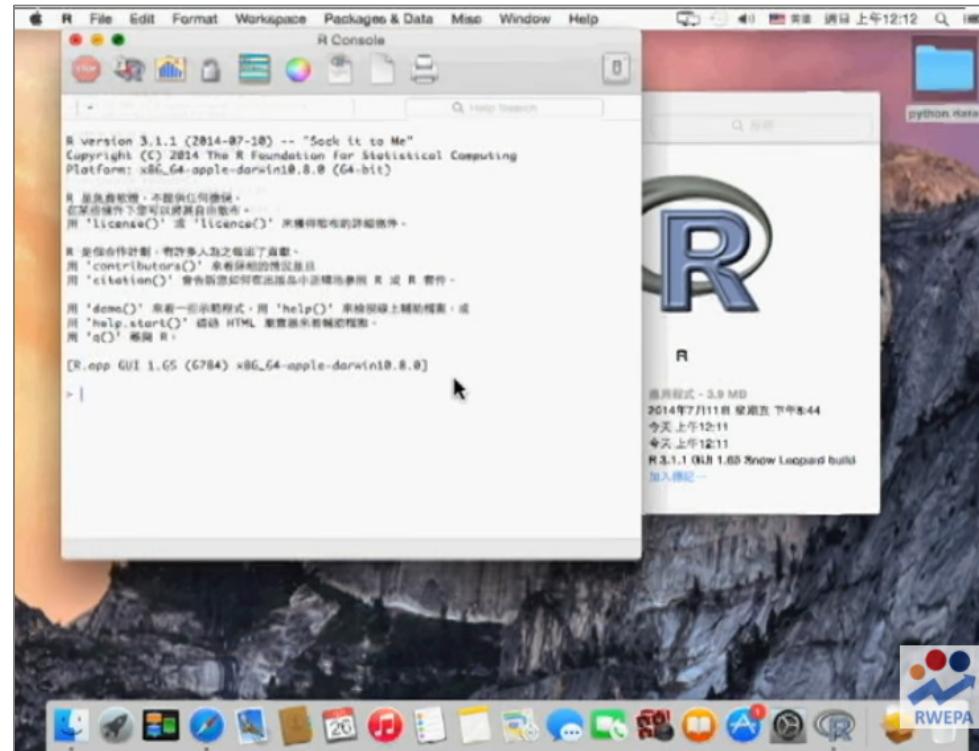


normaltext
程式碼顏色

- Language: en 英文
- size: 字型大小

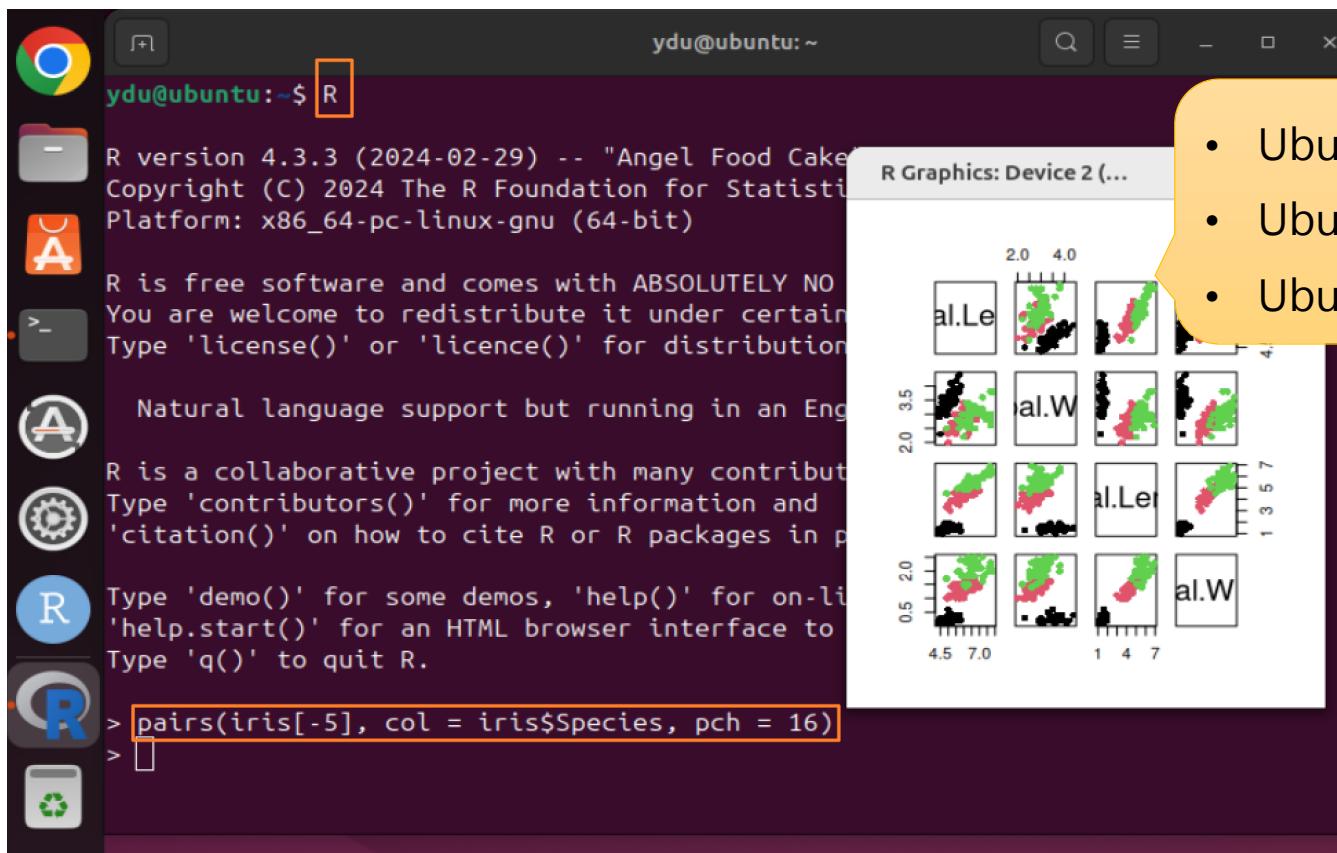
R for Mac

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



R for Ubuntu

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



A screenshot of an Ubuntu desktop environment. On the left, there's a dock with various icons, including a browser, file manager, terminal, and RStudio. The terminal window in the center shows the R command-line interface. The user has run the command `pairs(iris[-5], col = iris$Species, pch = 16)`, which generates a 4x4 matrix of scatter plots comparing the first four features of the Iris dataset. The plots are color-coded by species: setosa (black), versicolor (green), and virginica (red). A yellow callout bubble on the right side of the screen lists three bullet points:

- Ubuntu 中文輸入法
- Ubuntu 22.04.2 LTS 安裝 R-4.3.0
- Ubuntu 22.04.2 LTS 安裝 RStudio

新增R檔案練習



實作
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.

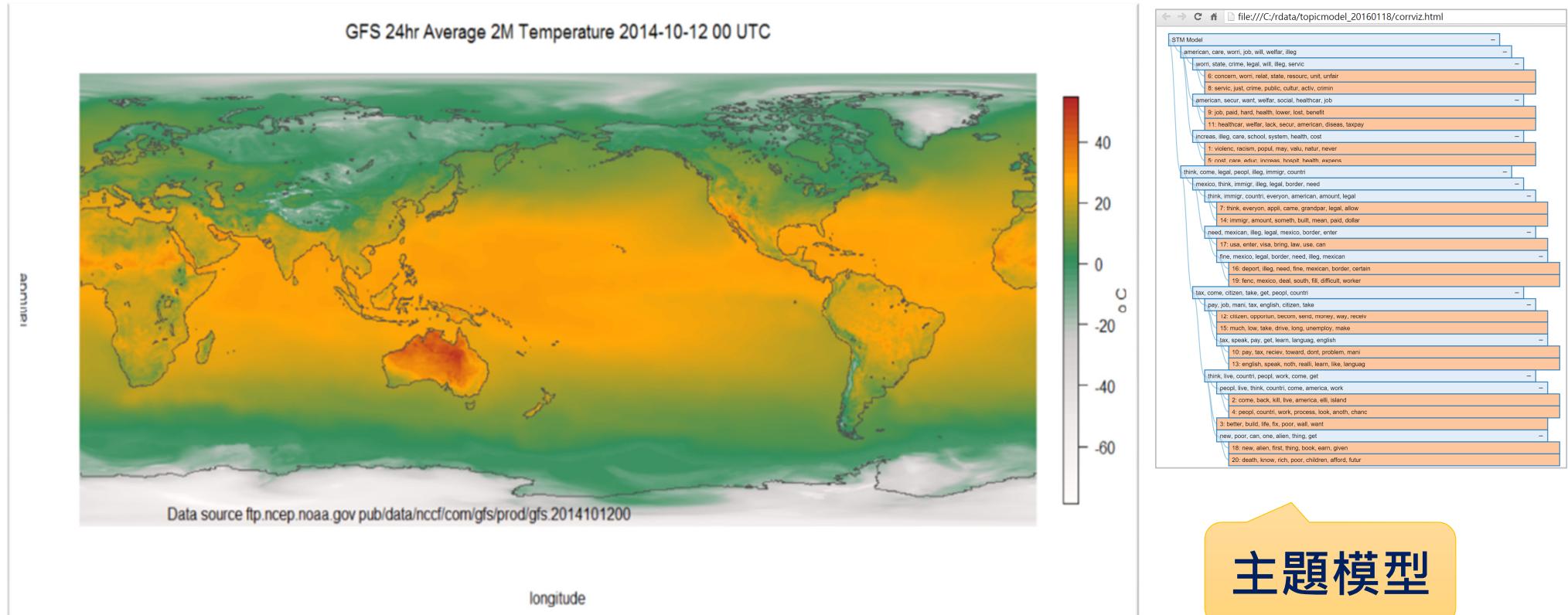
- 步驟1: 選取程式碼
- 步驟2: 按  Run line or selection 或 Ctrl + R
- 步驟3: 按 File \ Save \ MyFirstProject.R

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

2.2 RStudio 簡介與安裝

整合式開發環境 - RStudio



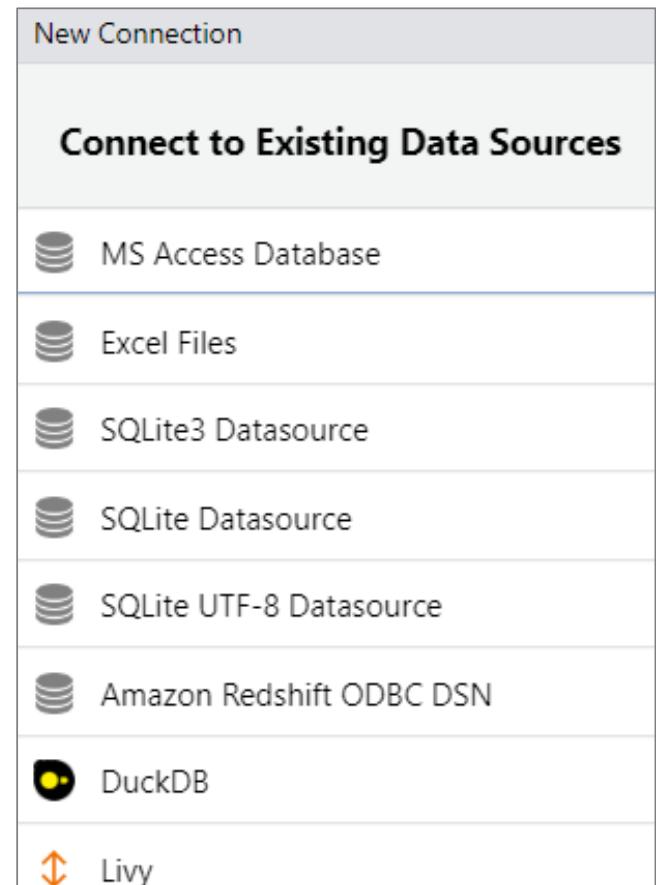
主題模型

視覺化應用

(全球2M氣溫圖)

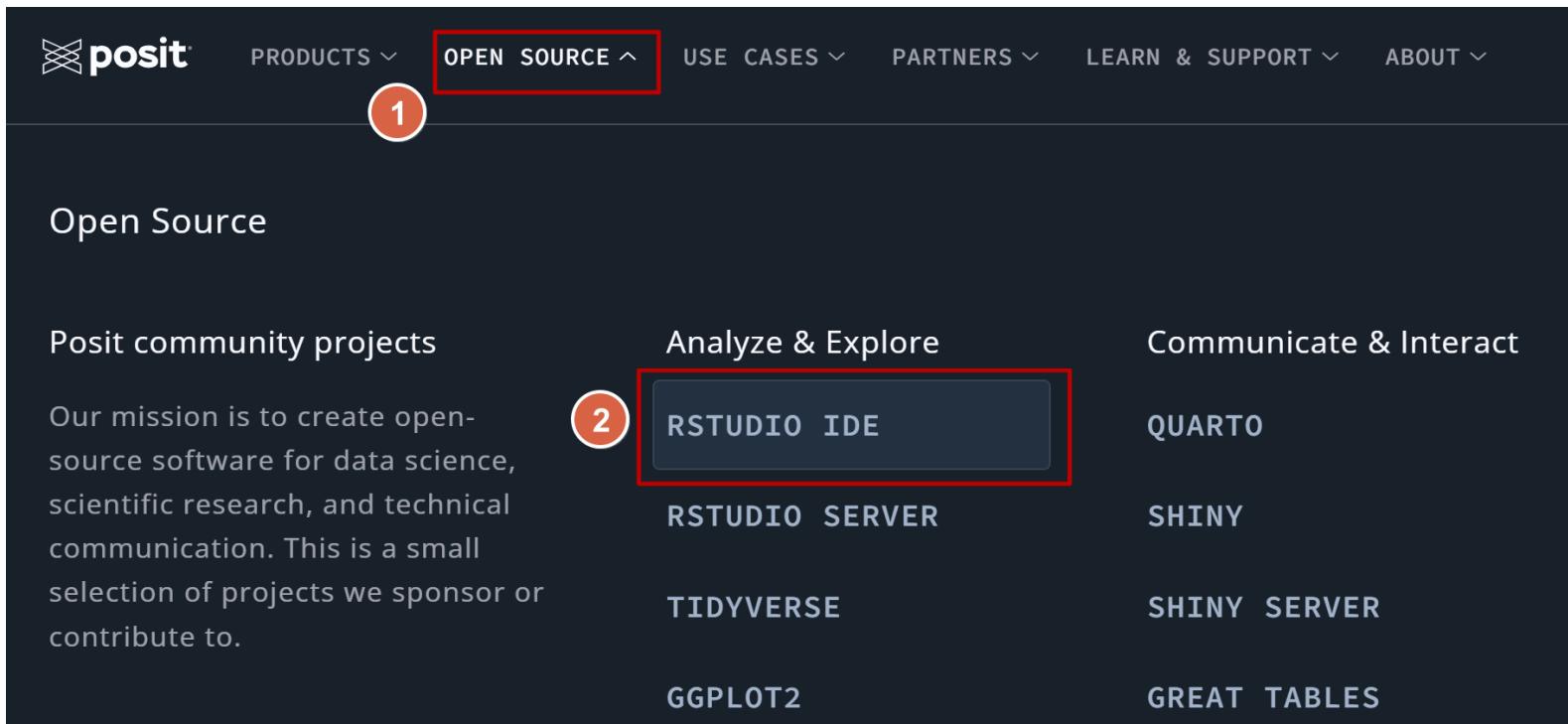
RStudio - 特性

- 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/ website's "Open Source" section. At the top, there is a navigation bar with links: PRODUCTS ▾, OPEN SOURCE ▾ (which is highlighted with a red box and circled with a red number 1), USE CASES ▾, PARTNERS ▾, LEARN & SUPPORT ▾, and ABOUT ▾.

The main content area is titled "Open Source". It features several sections:

- Posit community projects**: A brief description followed by a paragraph about their mission to create open-source software for data science, scientific research, and technical communication.
- Analyze & Explore**: A section containing "RSTUDIO IDE" (which is highlighted with a red box and circled with a red number 2), "RSTUDIO SERVER", "TIDYVERSE", and "GGPLOT2".
- Communicate & Interact**: A section containing "QUARTO", "SHINY", "SHINY SERVER", and "GREAT TABLES".

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

伺服器版本



RStudio-2024.12.0-467.exe下載

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

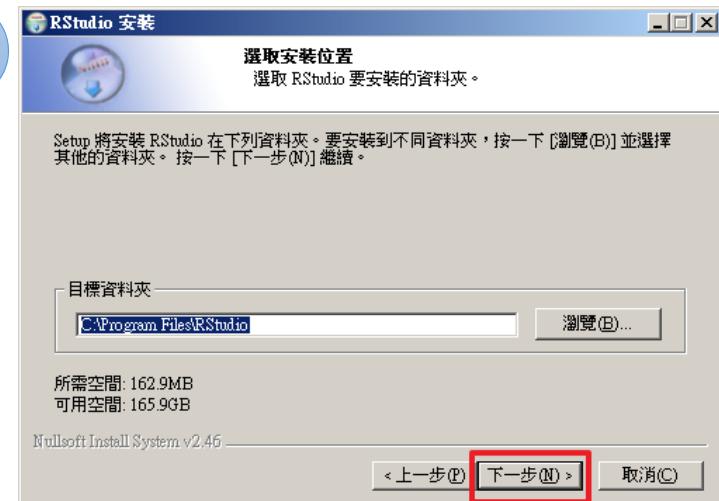
OS	Download	Size
Windows 10/11	RSTUDIO-2024.12.0-467.EXE ↓	265.27 MB
macOS 13+	RSTUDIO-2024.12.0-467.DMG ↓	617.71 MB
Ubuntu 20/Debian 11	RSTUDIO-2024.12.0-467-AMD64.DEB ↓	203.15 MB
Ubuntu 22/Debian 12	RSTUDIO-2024.12.0-467-AMD64.DEB ↓	203.15 MB
Ubuntu 24	RSTUDIO-2024.12.0-467-AMD64.DEB ↓	203.15 MB

RStudio 安裝

1



2



3

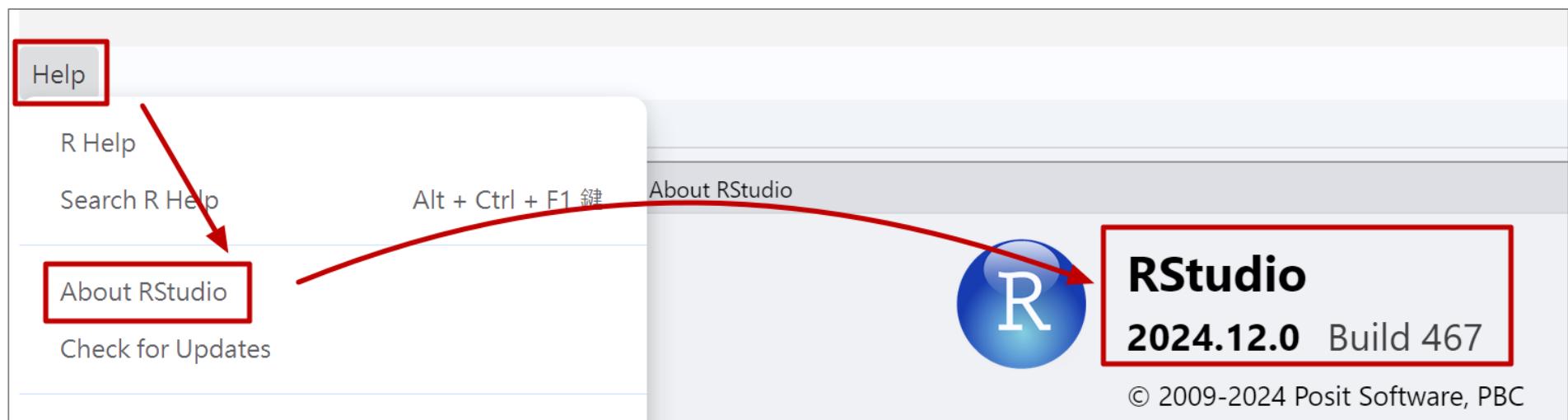


4



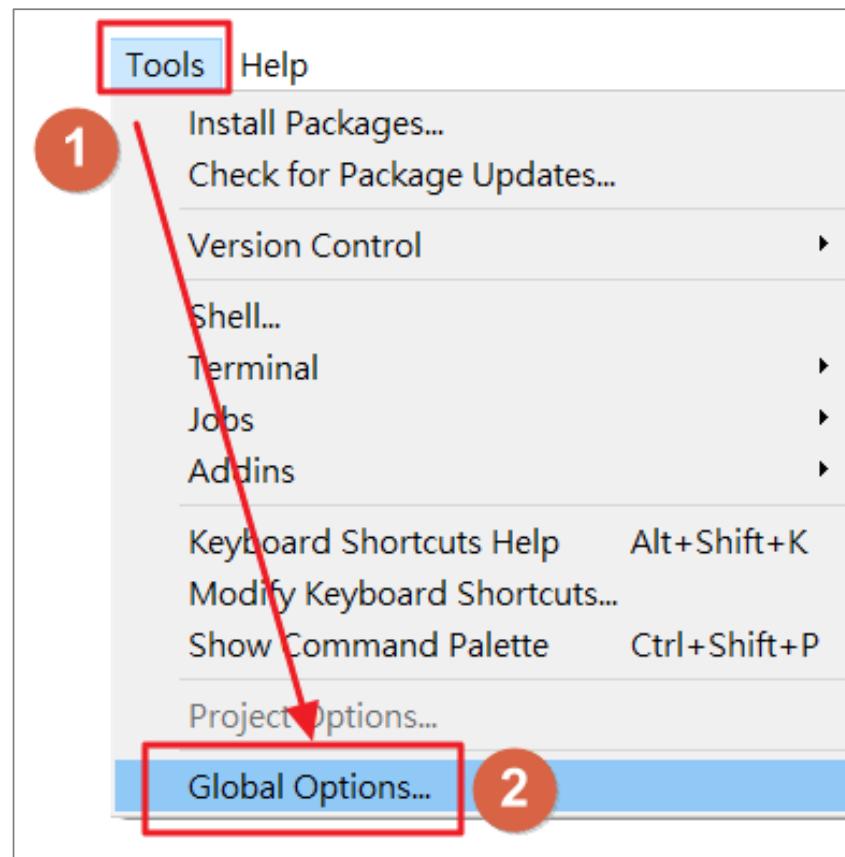
RStudio 版本訊息

- Help \ About RStudio

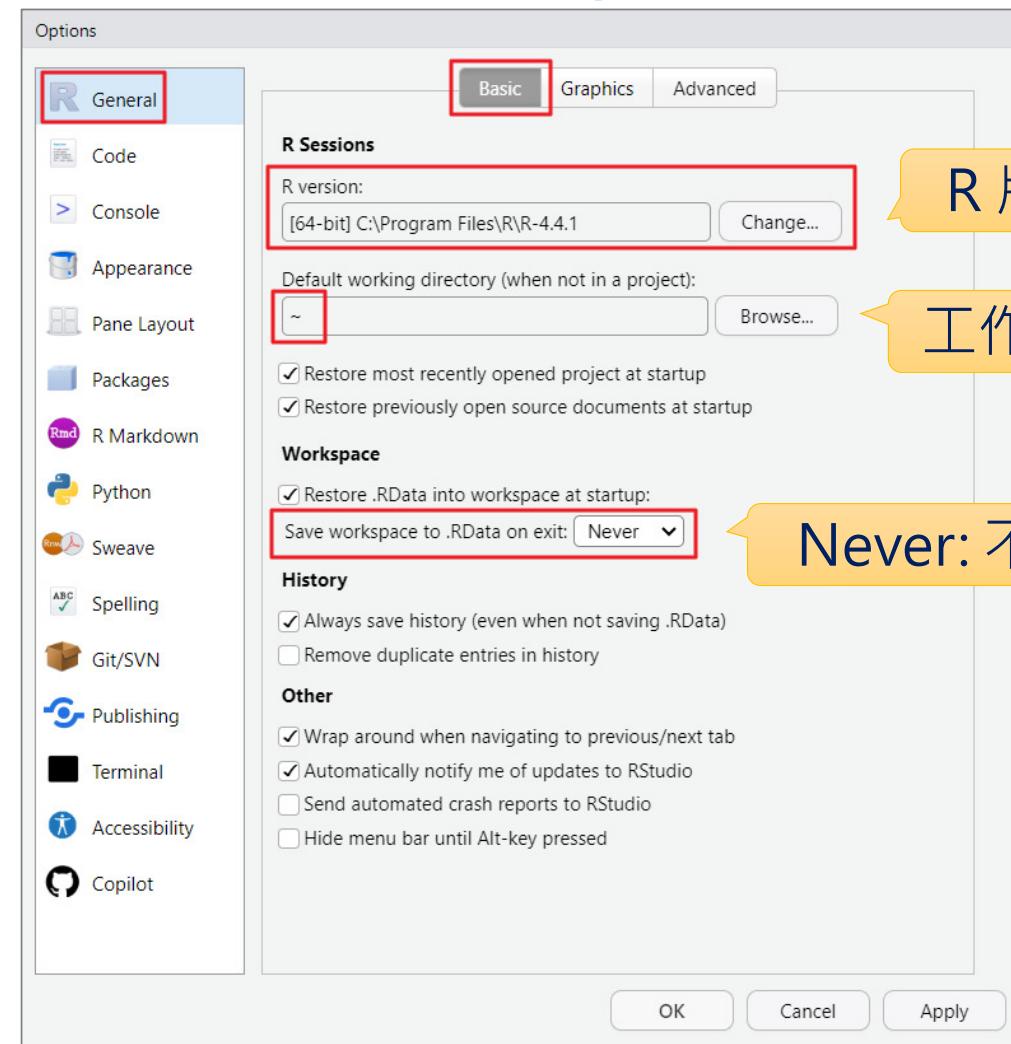


RStudio-選項設定

- Tools \ Global Options



General \ Basic

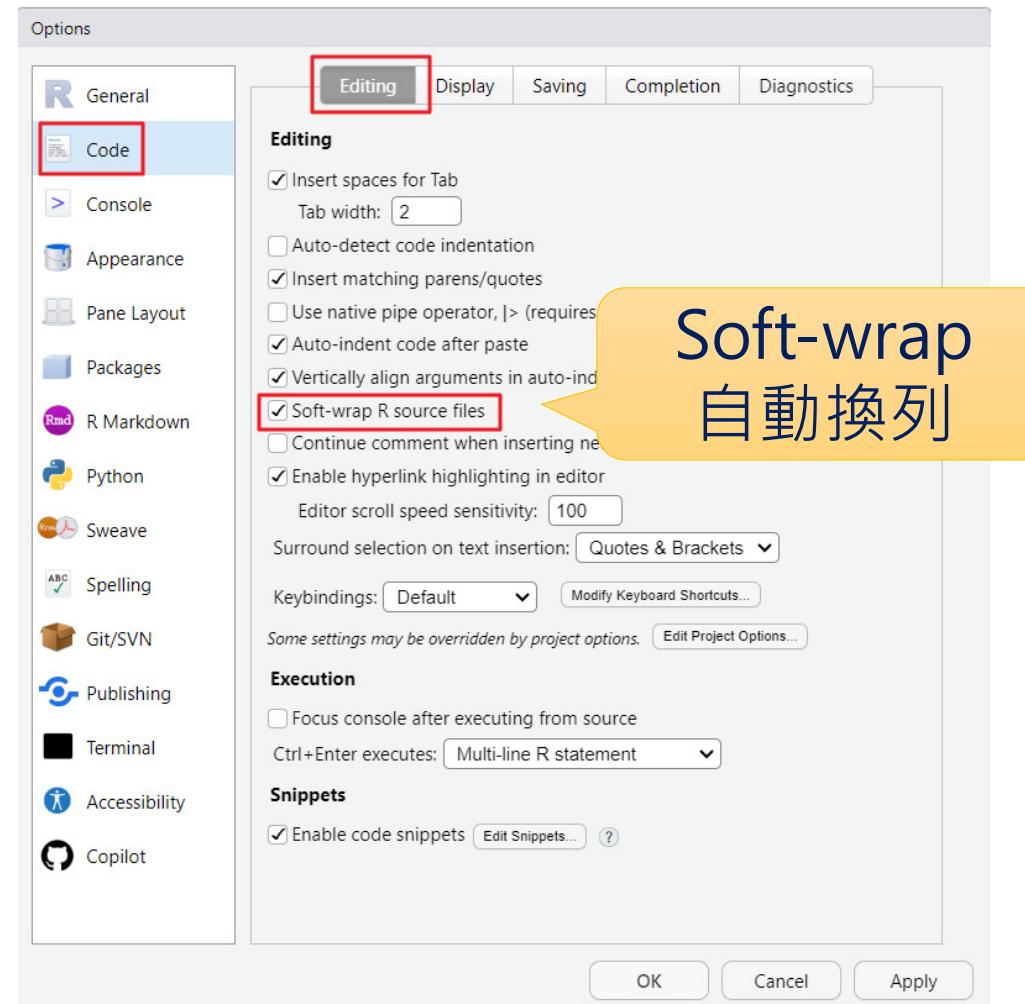


R 版本

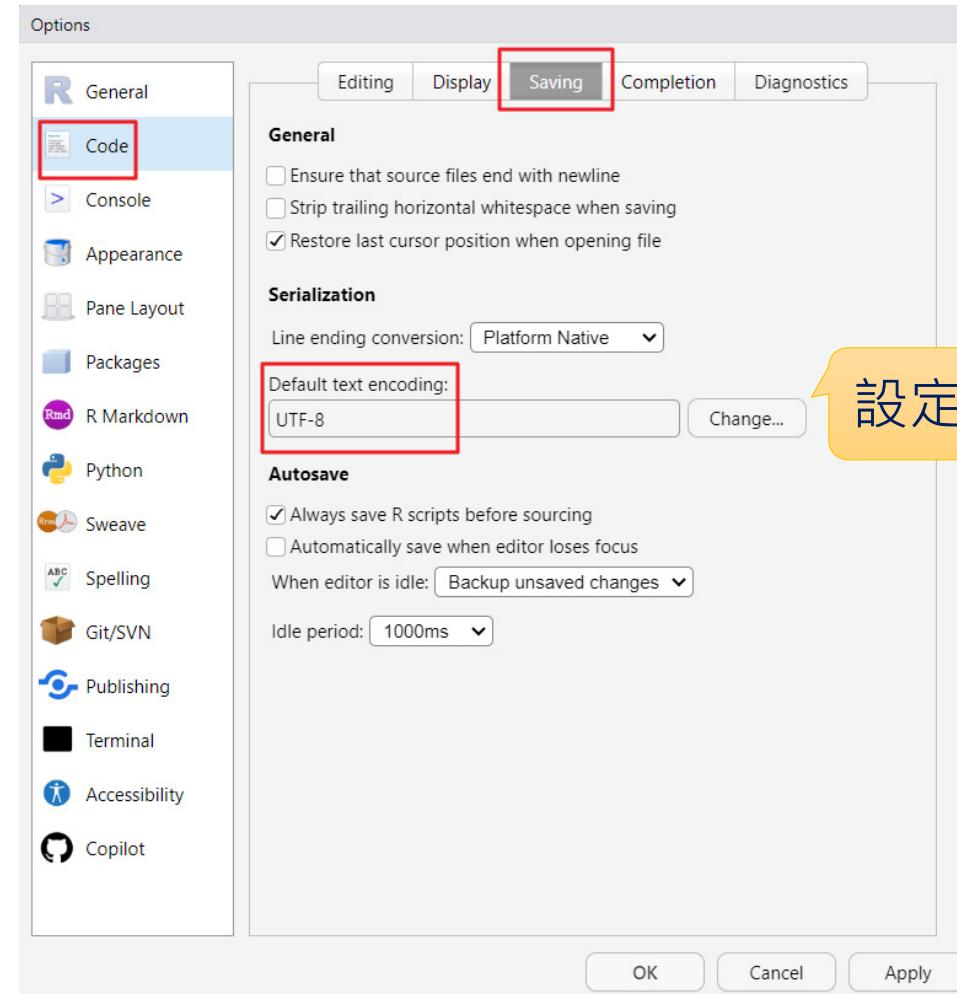
工作目錄 C:/rdata

Never: 不用儲存RData

Code \ Editing



Code \ Saving

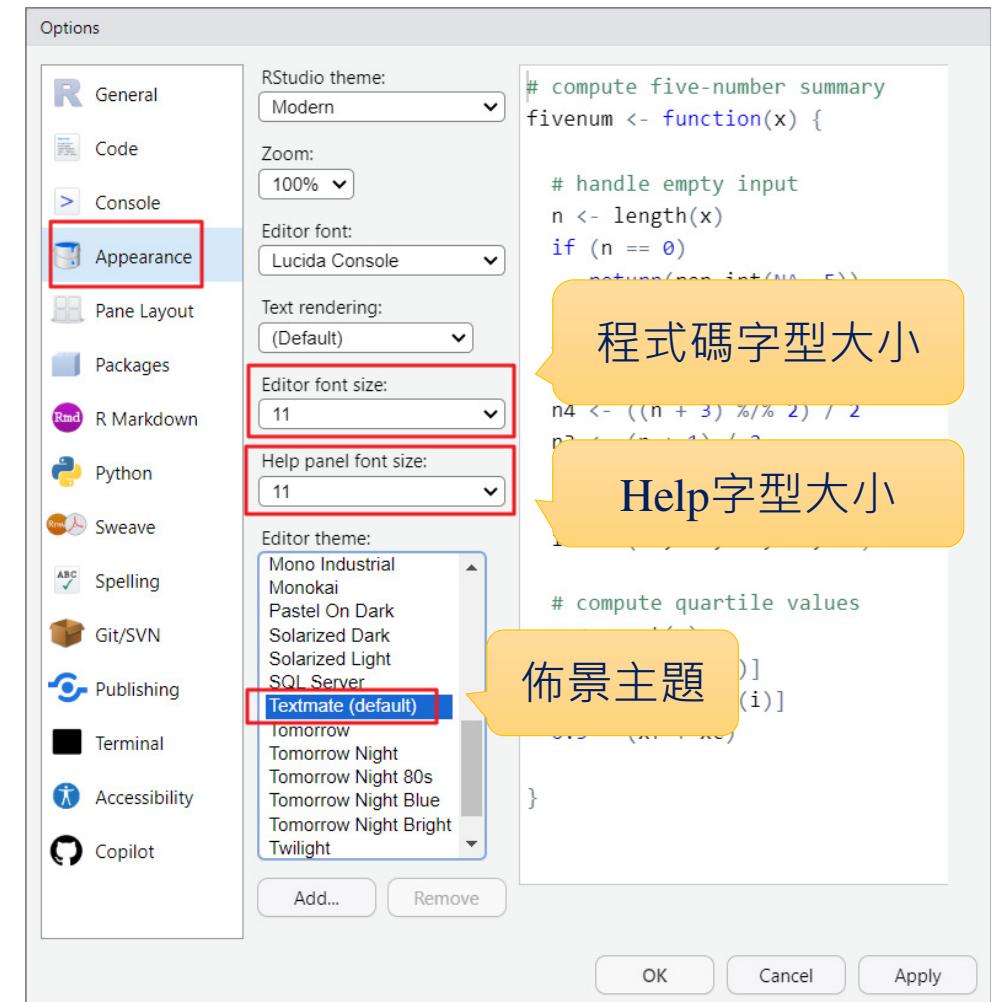


設定程式編碼 UTF-8

RStudio-選項設定(續)

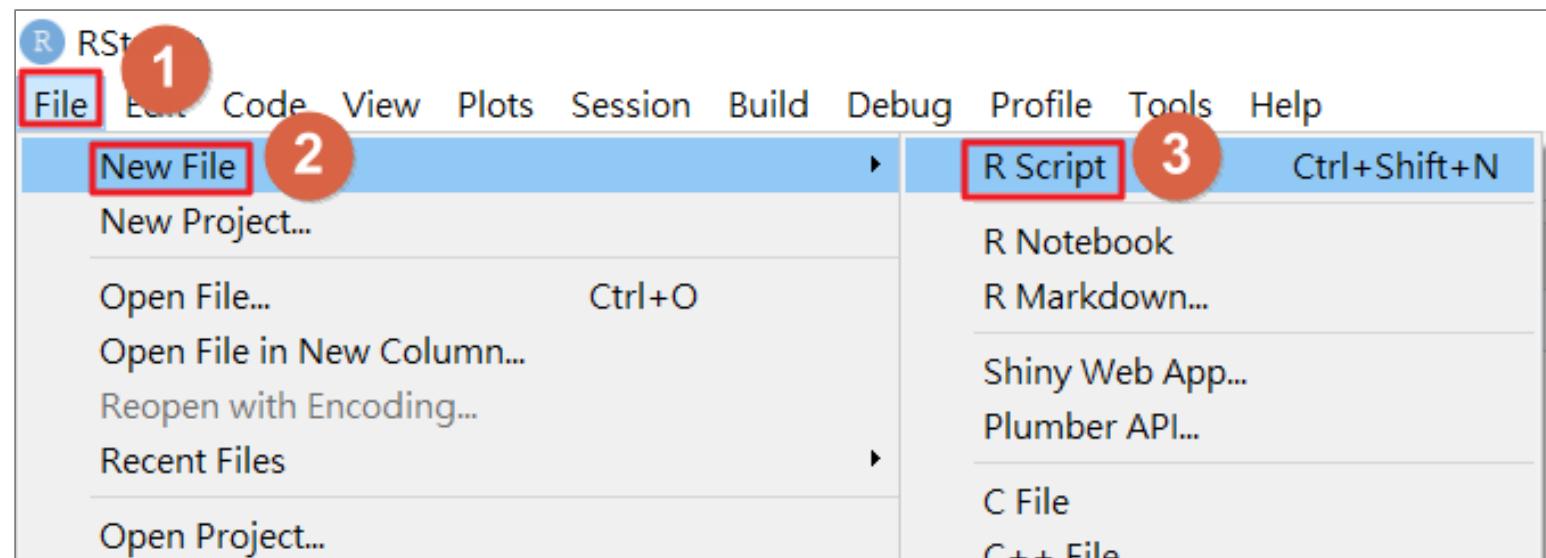
- Appearance \ Editor font size
- Appearance \ Help panel font size
- Editor theme \ Textmate

設定完成，
可能須重新啟動RStudio



新增檔案

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





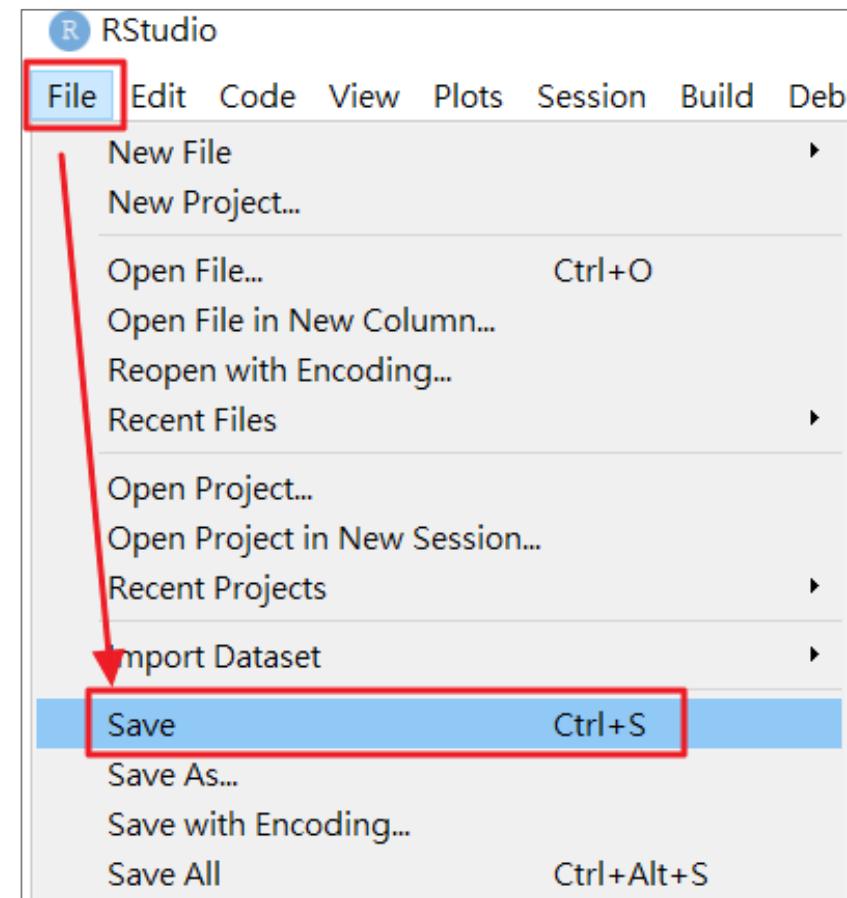
儲存檔案

參考「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





RStudio 程式碼執行

- 滑鼠置於程式碼最後(或選取整個程式碼) \ 按 Run (或 CTRL + Enter)

The screenshot shows the RStudio interface. In the top-left pane, there is a code editor window titled "Untitled1" containing the following R code:

```
1 plot(runif(100), type="l", main= "R大數據分析")
2
```

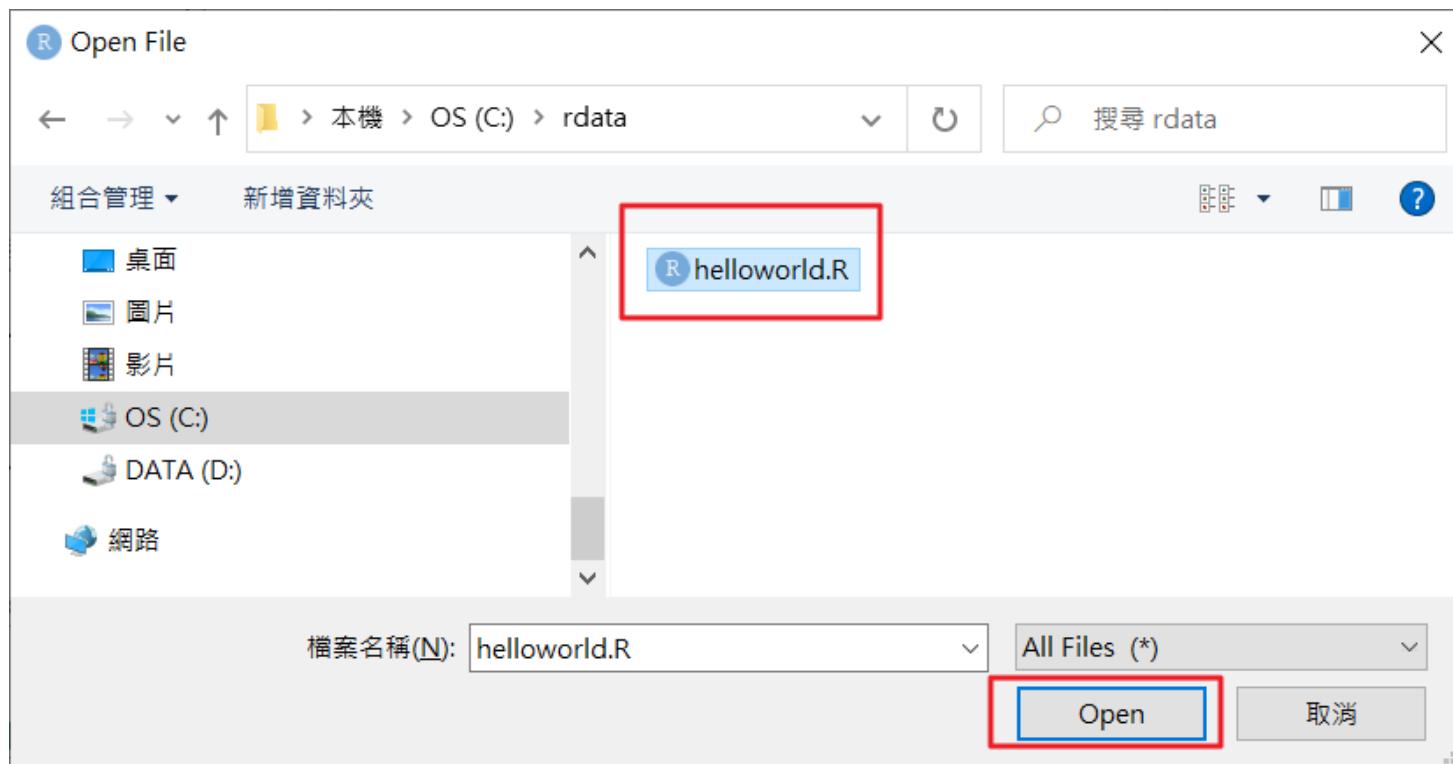
Two red circles with numbers indicate specific actions: circle 1 points to the closing parenthesis ')' at the end of the main argument, and circle 2 points to the "Run" button in the toolbar above the code editor.

In the bottom-left pane, the "Console" tab is active, showing R version 4.4.2 and some introductory text about R help functions like citation(), demo(), help(), and help.start().

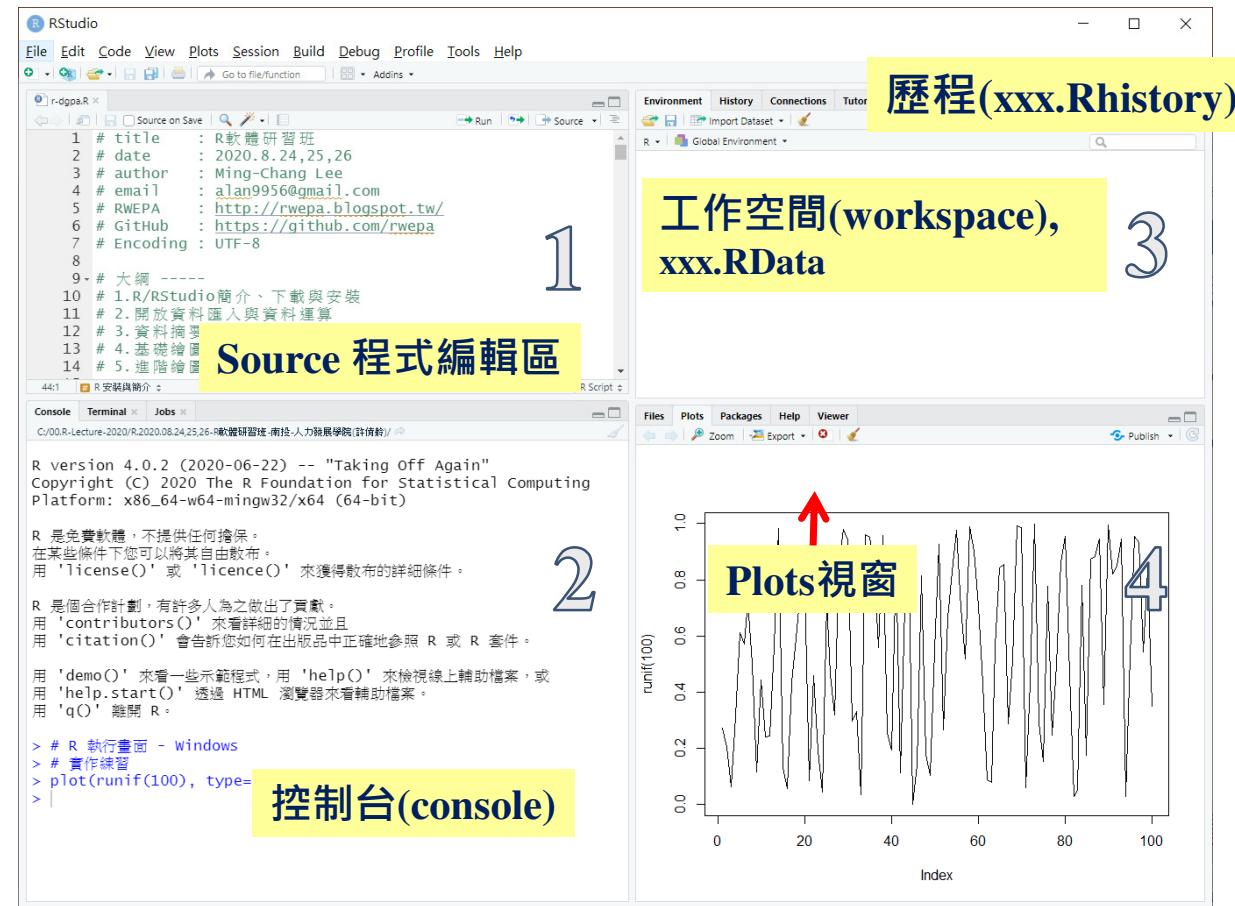
In the bottom-right pane, a plot titled "R大數據分析" is displayed. The plot shows a line graph of 100 random uniform numbers (runif(100)) plotted against an index from 0 to 100. The y-axis ranges from 0.0 to 1.0, and the x-axis is labeled "Index".

開啟檔案

- File \ Open File \ helloworld.R



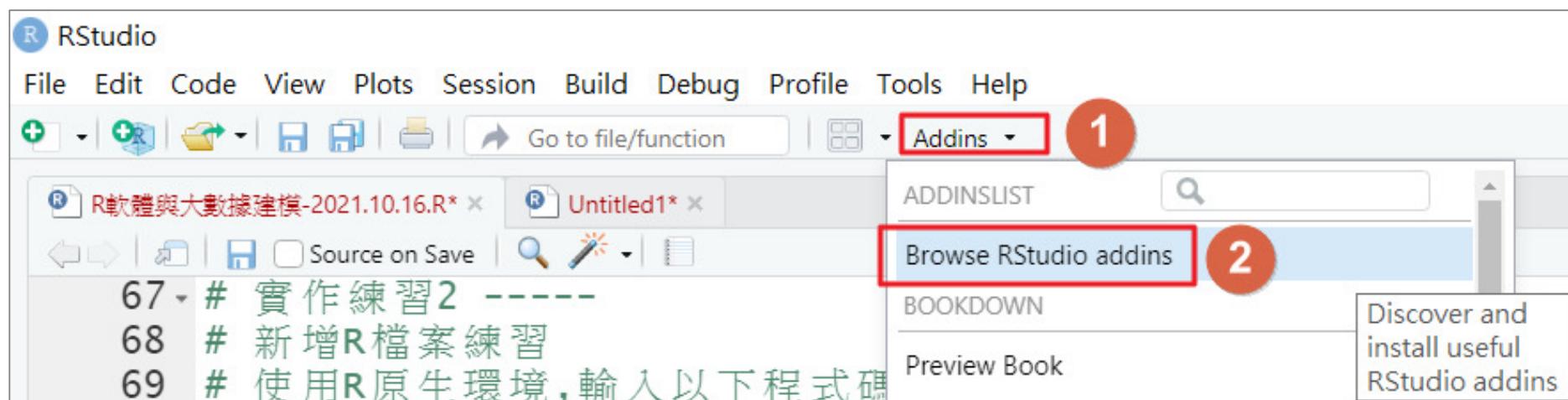
R/RStudio環境的基礎觀念



Ctrl + Shift + F10: 重新啟動R

RStudio Addins (外掛功能)

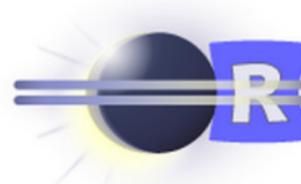
- `install.packages("addinslist")`
- Addins \ Browse RStudio addins



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 4.11.0
An Eclipse based IDE (integrated development environment) plug-in for R.
 - <https://projects.eclipse.org/projects/science.statet>
- Jupyter-notebook
- Visual Studio Code
 - <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)

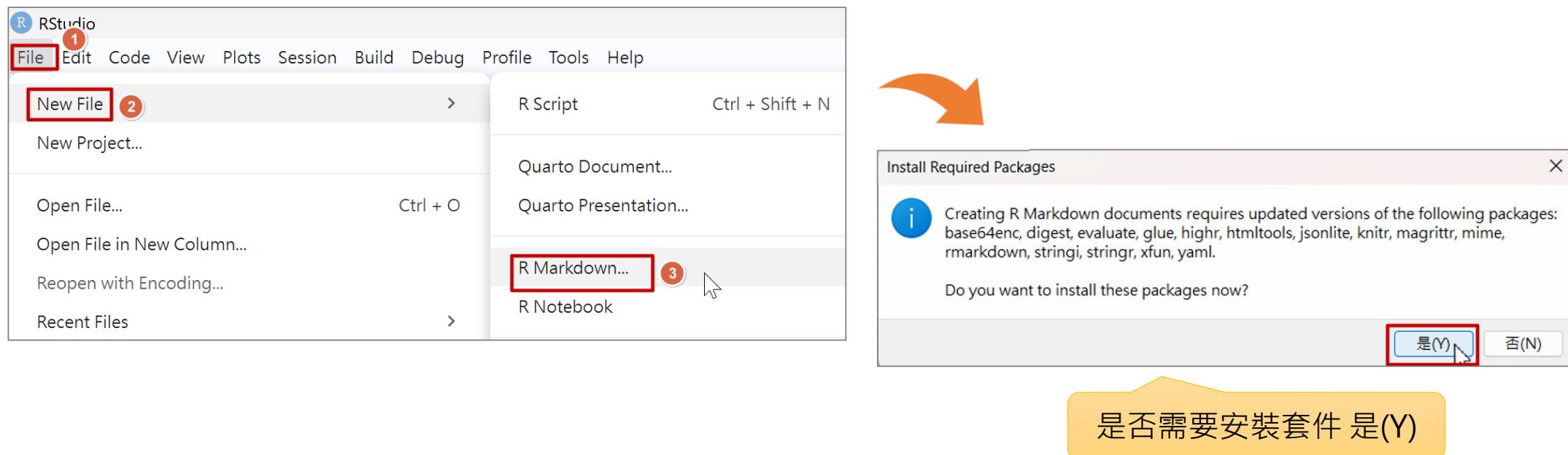
- 章節功能可以快速切換程式碼

R Markdown

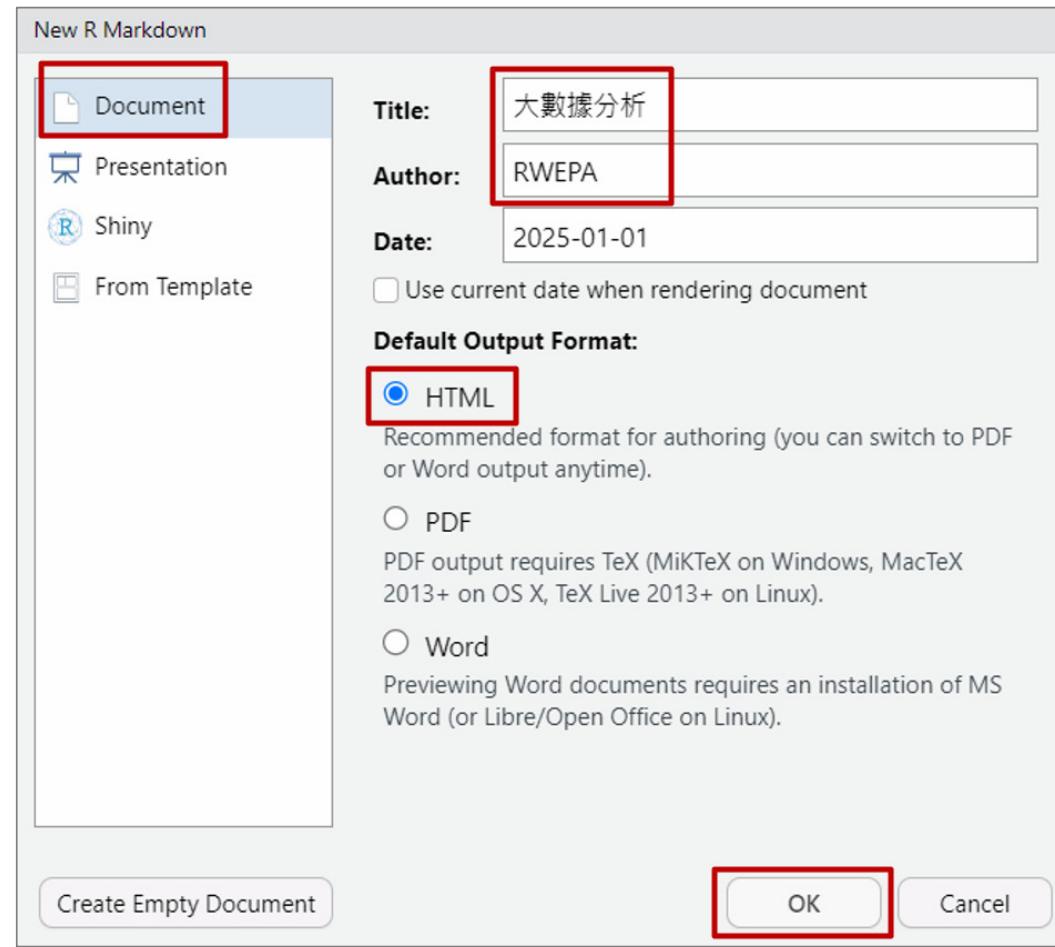
(R 標記語言)

R Markdown 檔案

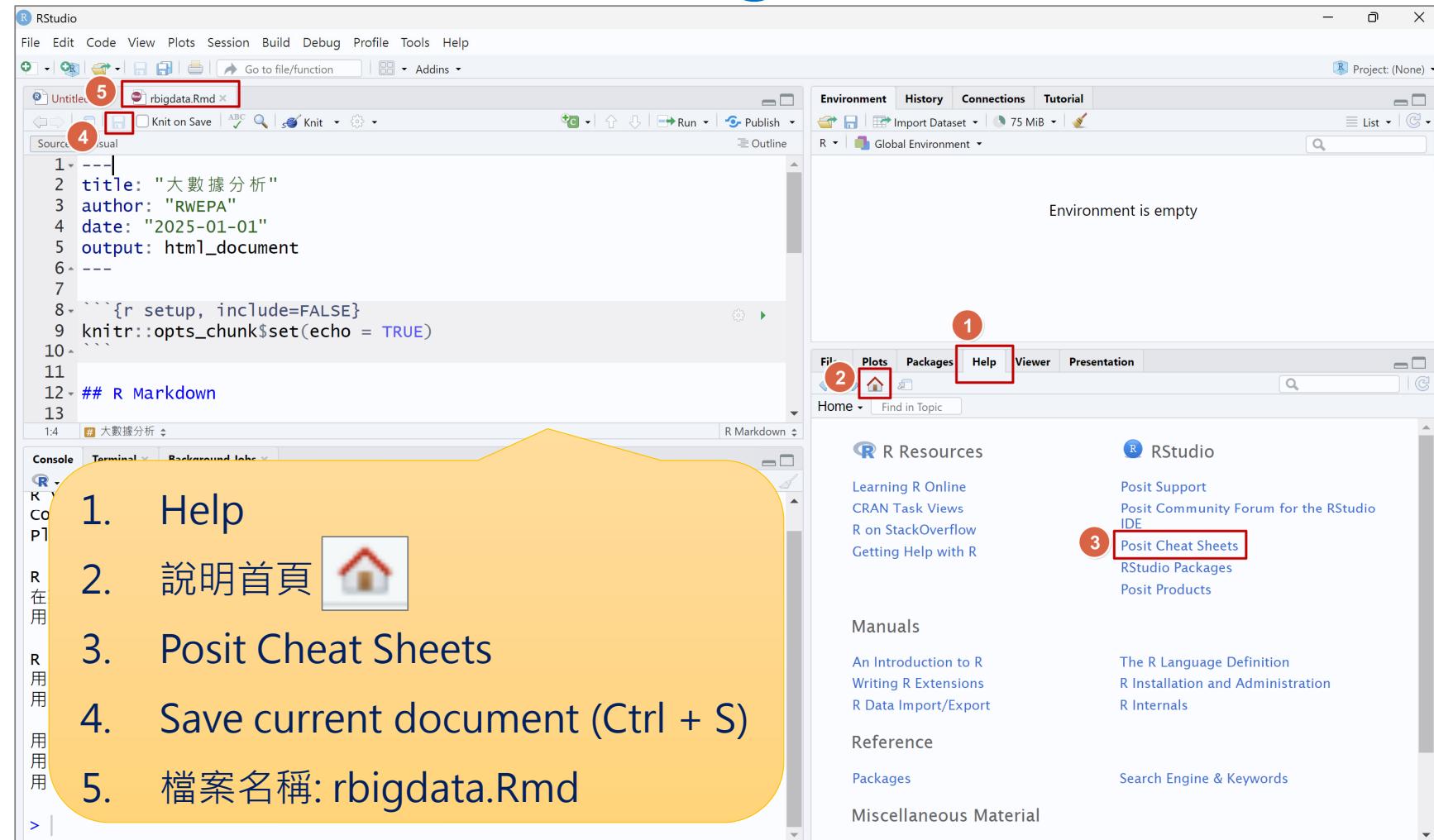
- File \ New File \ R Markdown \ 是否需要安裝套件? \ [Yes]



R Markdown 檔案 (續)



完成檔 – rbigdata.Rmd



1. Help

2. 說明首頁 

3. Posit Cheat Sheets

4. Save current document (Ctrl + S)

5. 檔案名稱: rbigdata.Rmd

1. Environment

2. Plots

3. Posit Cheat Sheets

4. Knit on Save

5. rbigdata.Rmd

```
1 ---  
2 title: "大數據分析"  
3 author: "RWEPA"  
4 date: "2025-01-01"  
5 output: html_document  
6 ---  
7  
8 ```{r setup, include=FALSE}  
9 knitr::opts_chunk$set(echo = TRUE)  
10```  
11  
12 ## R Markdown  
13
```

Environment is empty

R Resources

An Introduction to R

Writing R Extensions

R Data Import/Export

Manuals

The R Language Definition

R Installation and Administration

R Internals

Reference

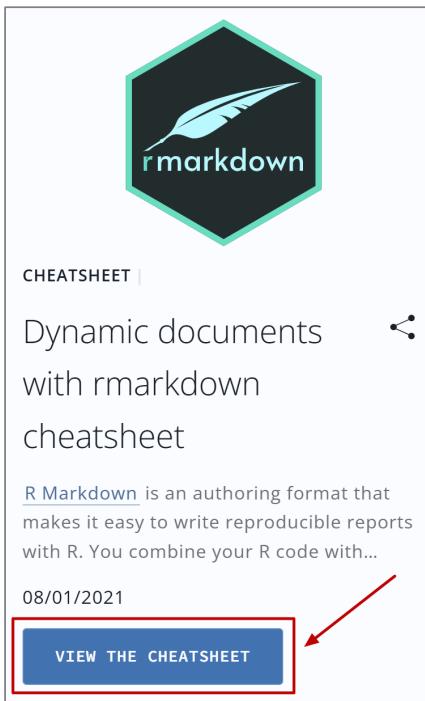
Packages

Miscellaneous Material

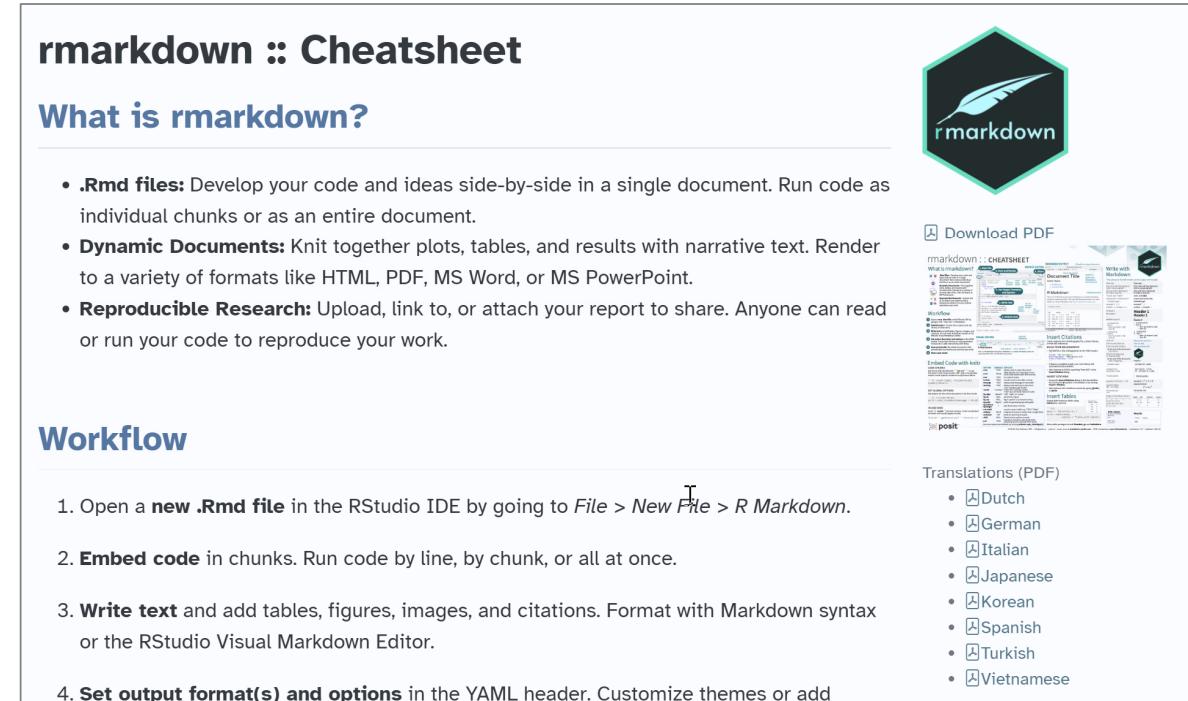
Search Engine & Keywords

Posit Cheatsheets 線上說明

- <https://posit.co/resources/cheatsheets/>
- <https://rstudio.github.io/cheatsheets/html/rmarkdown.html>



The image shows the cover of the rmarkdown cheatsheet. It features a dark hexagonal logo with a white feather icon and the word "rmarkdown". Below the logo, the text "CHEATSHEET" is written vertically. The main title "Dynamic documents with rmarkdown cheatsheet" is displayed in a large, sans-serif font. A small "Share" icon is located to the right of the title. At the bottom, there is a short description: "R Markdown is an authoring format that makes it easy to write reproducible reports with R. You combine your R code with...". The date "08/01/2021" is at the bottom left, and a blue button labeled "VIEW THE CHEATSHEET" is at the bottom right. A red arrow points from the "VIEW THE CHEATSHEET" button to the "VIEW THE CHEATSHEET" link on the RStudio page.



The image shows the "rmarkdown :: Cheatsheet" page from RStudio. The header features the "rmarkdown" logo. The main section is titled "What is rmarkdown?". It lists three key features:

- **.Rmd files:** Develop your code and ideas side-by-side in a single document. Run code as individual chunks or as an entire document.
- **Dynamic Documents:** Knit together plots, tables, and results with narrative text. Render to a variety of formats like HTML, PDF, MS Word, or MS PowerPoint.
- **Reproducible Research:** Upload, link to, or attach your report to share. Anyone can read or run your code to reproduce your work.

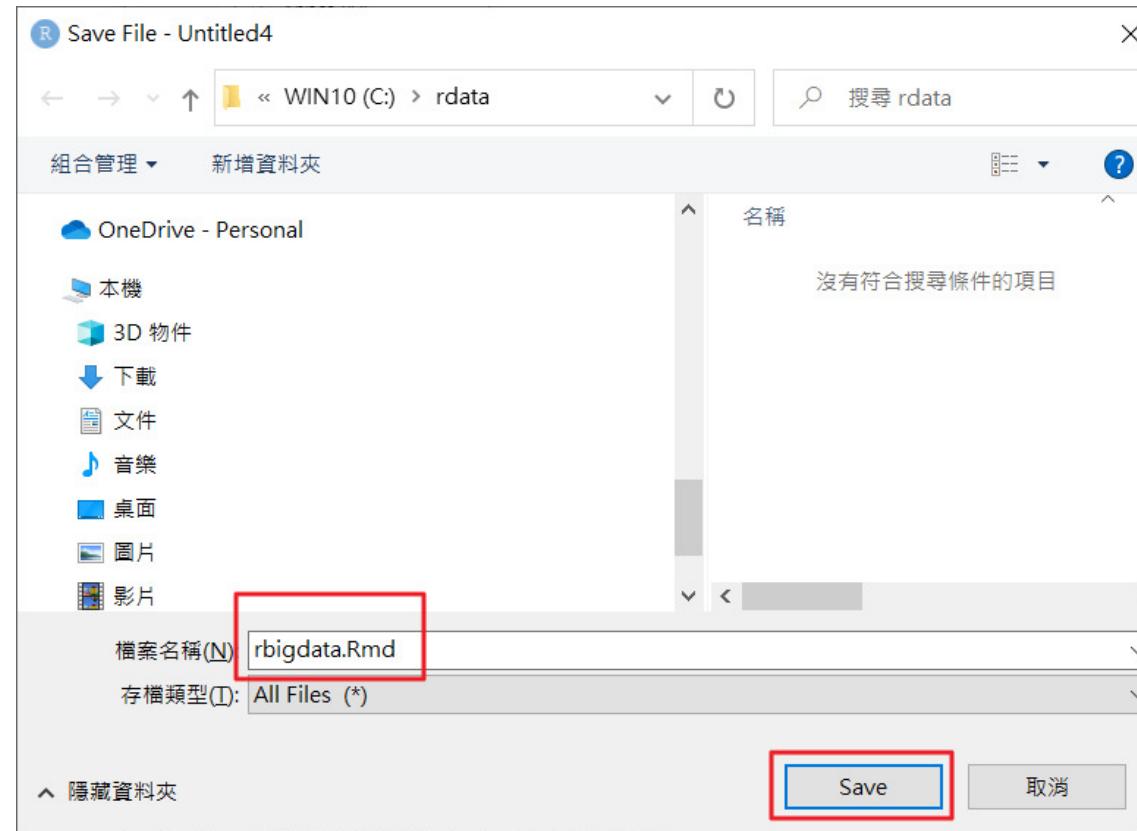
Below this is a "Workflow" section with four numbered steps:

1. Open a **new .Rmd file** in the RStudio IDE by going to *File > New File > R Markdown*.
2. **Embed code** in chunks. Run code by line, by chunk, or all at once.
3. **Write text** and add tables, figures, images, and citations. Format with Markdown syntax or the RStudio Visual Markdown Editor.
4. **Set output format(s) and options** in the YAML header. Customize themes or add

To the right of the workflow, there is a "Download PDF" button and a preview of the PDF document. Below that is a "Translations (PDF)" section with links to Dutch, German, Italian, Japanese, Korean, Spanish, Turkish, and Vietnamese versions.

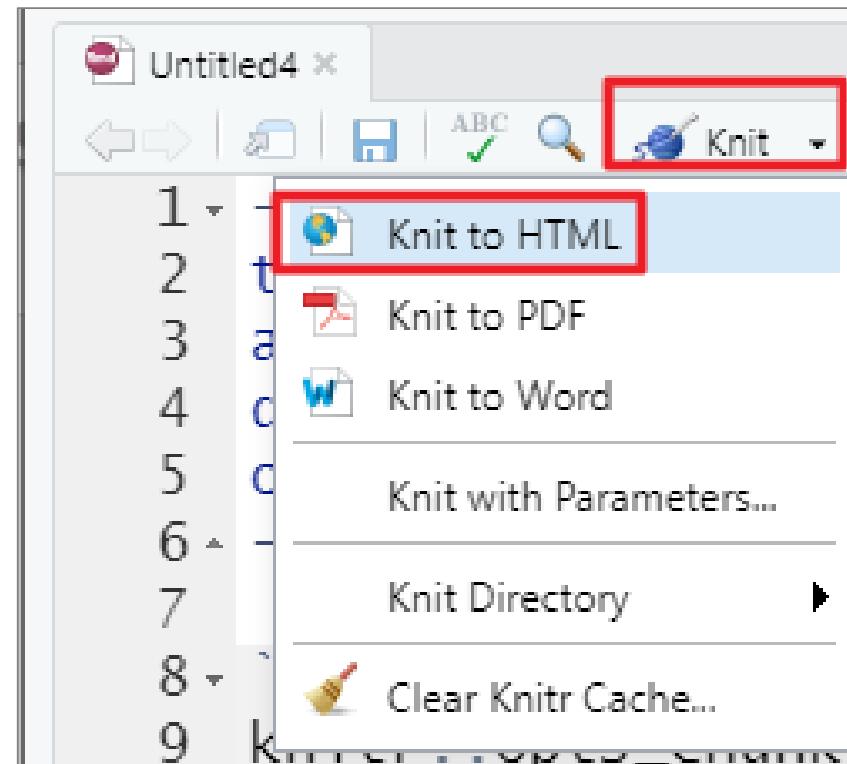
RStudio - Markdown (續)

- File \ Save \ rbigdata.Rmd

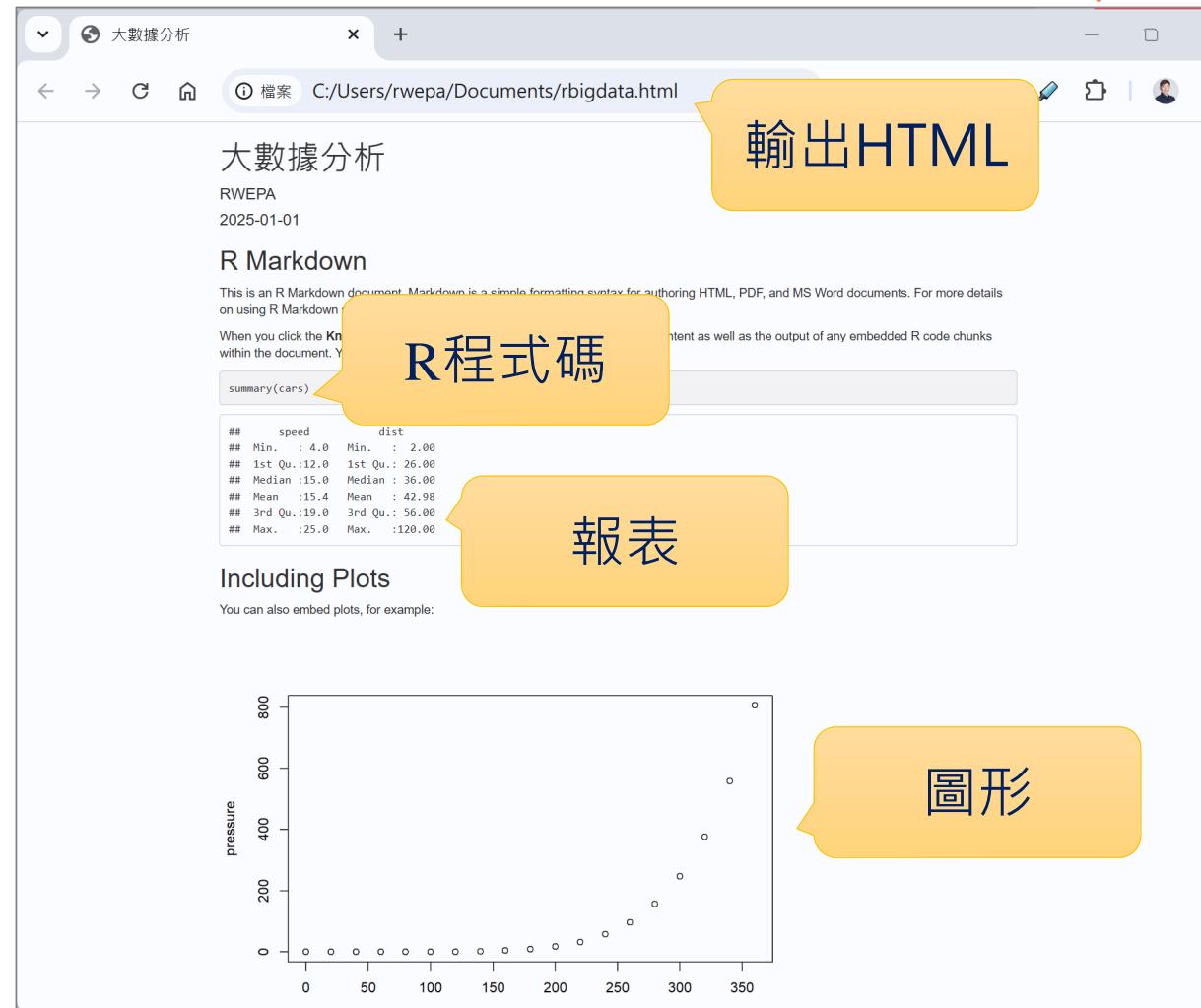


Rmd 檔案轉換

- Knit HTML
- Knit PDF
- Knit Word



Markdown : HTML ✓



大數據分析

RWEPA
2025-01-01

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.

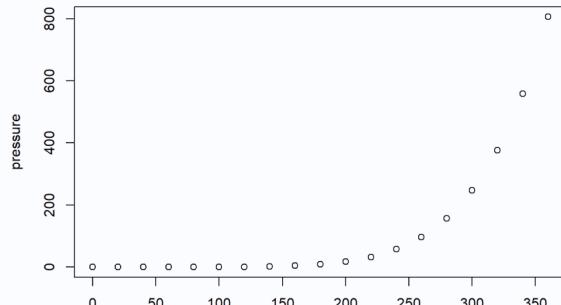
When you click the Knit button, a document is generated that contains both the content as well as the output of any embedded R code chunks.

```
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

Including Plots

You can also embed plots, for example:



A scatter plot showing the relationship between speed (x-axis, ranging from 0 to 350) and pressure (y-axis, ranging from 0 to 800). The data points show a positive correlation, with most points clustered at lower speeds and pressures, and a few outliers at higher speeds and pressures.

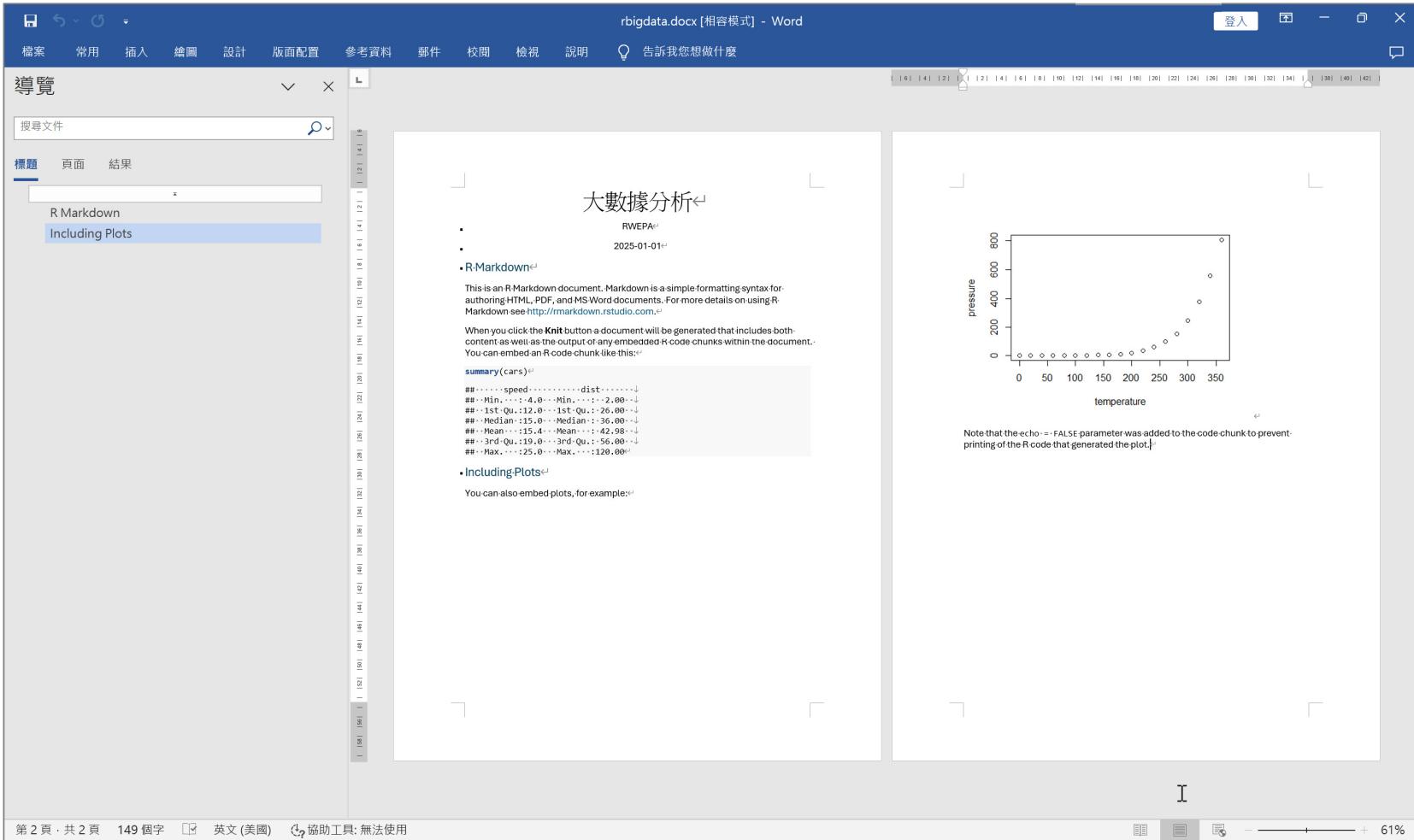
輸出HTML

R程式碼

報表

圓形

Markdown : Word ✓



The screenshot shows a Microsoft Word document titled "rbigdata.docx". The document contains an R Markdown section with the title "大數據分析" and a subtitle "R Markdown". It includes a timestamp "2025-01-01" and a summary of the "cars" dataset:

```
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 ..
```

The document also features an embedded scatter plot showing pressure versus temperature, with data points ranging from approximately (50, 40) to (350, 800).

At the bottom of the Word window, there is a note: "Note that the echo := FALSE parameter was added to the code chunk to prevent printing of the R code that generated the plot."

Page navigation at the bottom indicates "第 2 頁 · 共 2 頁 149 個字" and "61%".

Rmd 轉換 PDF 錯誤畫面

```
Console Terminal × Render × Background Jobs ×
~/rbigdata.Rmd
output file: rbigdata.knit.md

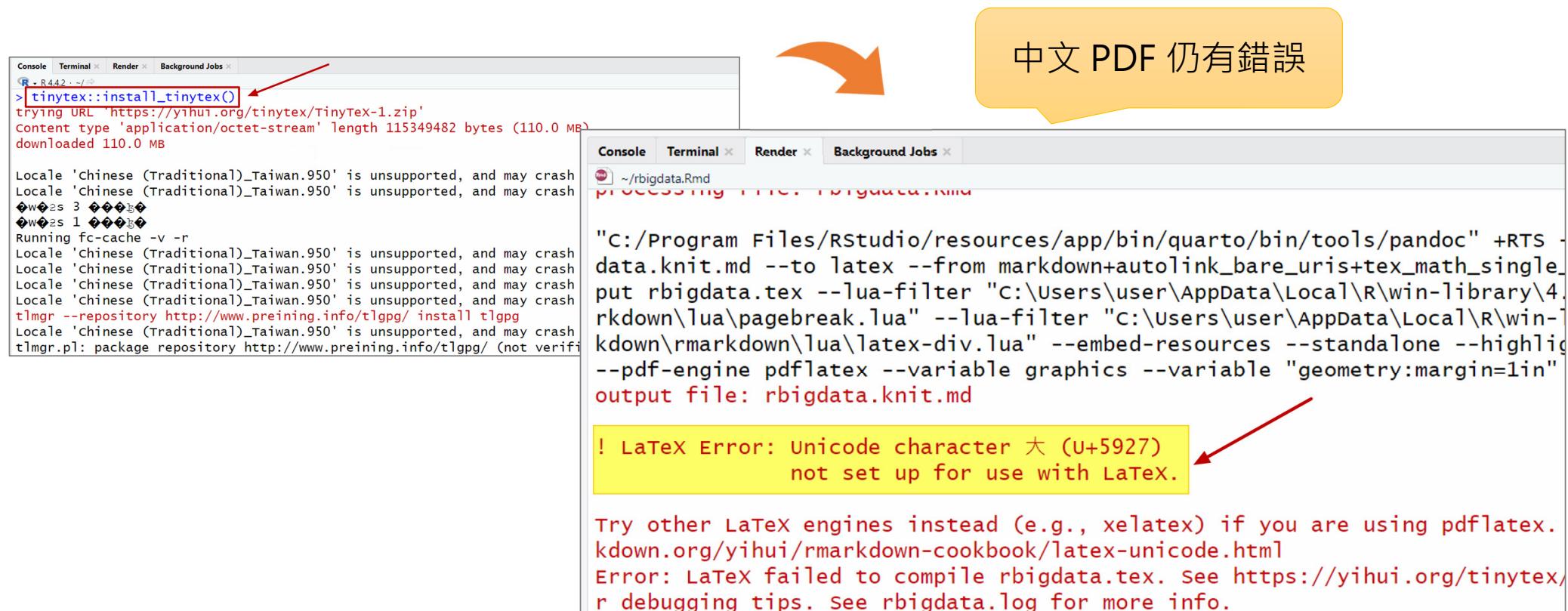
"C:/Program Files/RStudio/resources/app/bin/quarto/bin/tools/pandoc" +RTS -K512m -RTS rbigdata.knit.md --to latex --from markdown+autolink_bare_uris+tex_math_single_backslash --output rbigdata.tex --lua-filter "C:\Users\user\AppData\Local\R\win-library\4.4\rmarkdown\rmarkdown\lua\pagebreak.lua" --lua-filter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

PDF 中文錯誤 – 方法1 (tinytex)

- `tinytex::install_tinytex()`

中文 PDF 仍有錯誤



The screenshot shows two RStudio sessions. The top session shows the command `> tinytex::install_tinytex()` being run in the Console, which successfully downloads the TinyTeX package. The bottom session shows an Rmd file being rendered, but it fails with a LaTeX error message: "! LaTeX Error: Unicode character 大 (U+5927) not set up for use with LaTeX." This indicates that while the tinytex package was installed, it did not resolve the specific LaTeX engine configuration needed to handle the character correctly.

```

Console Terminal × Render × Background Jobs ×
R - R4.2.2 ~/rbigdata.Rmd
> tinytex::install_tinytex()
trying URL 'https://yihui.org/tinytex/TinyTeX-1.zip'
Content type 'application/octet-stream' length 115349482 bytes (110.0 MB)
downloaded 110.0 MB

Locale 'Chinese (Traditional)-Taiwan.950' is unsupported, and may crash
Locale 'Chinese (Traditional)-Taiwan.950' is unsupported, and may crash
  ♫ w2as 3 ♫ ♫ b ♫
  ♫ w2as 1 ♫ ♫ b ♫
Running fc-cache -v -r
Locale 'Chinese (Traditional)-Taiwan.950' is unsupported, and may crash
  tlmgr --repository http://www.preining.info/tlpgp/ install tlpgp
Locale 'Chinese (Traditional)-Taiwan.950' is unsupported, and may crash
  tlmgr.pl: package repository http://www.preining.info/tlpgp/ (not verified)

Console Terminal × Render × Background Jobs ×
~/rbigdata.Rmd
PROCESSING rbigdata.Rmd

"C:/Program Files/RStudio/resources/app/bin/quarto/bin/tools/pandoc" +RTS -data.knit.md --to latex --from markdown+autolink_bare_uris+tex_math_singleput rbigdata.tex --lua-filter "c:\users\user\AppData\Local\R\win-library\4.2\rkdown\lua\pagebreak.lua" --lua-filter "c:\users\user\AppData\Local\R\win-kdown\rmarkdown\lua\latex-div.lua" --embed-resources --standalone --highlighter pdflatex --variable graphics --variable "geometry:margin=1in" --pdf-engine pdflatex --variable "geometry:margin=1in" output file: rbigdata.knit.md

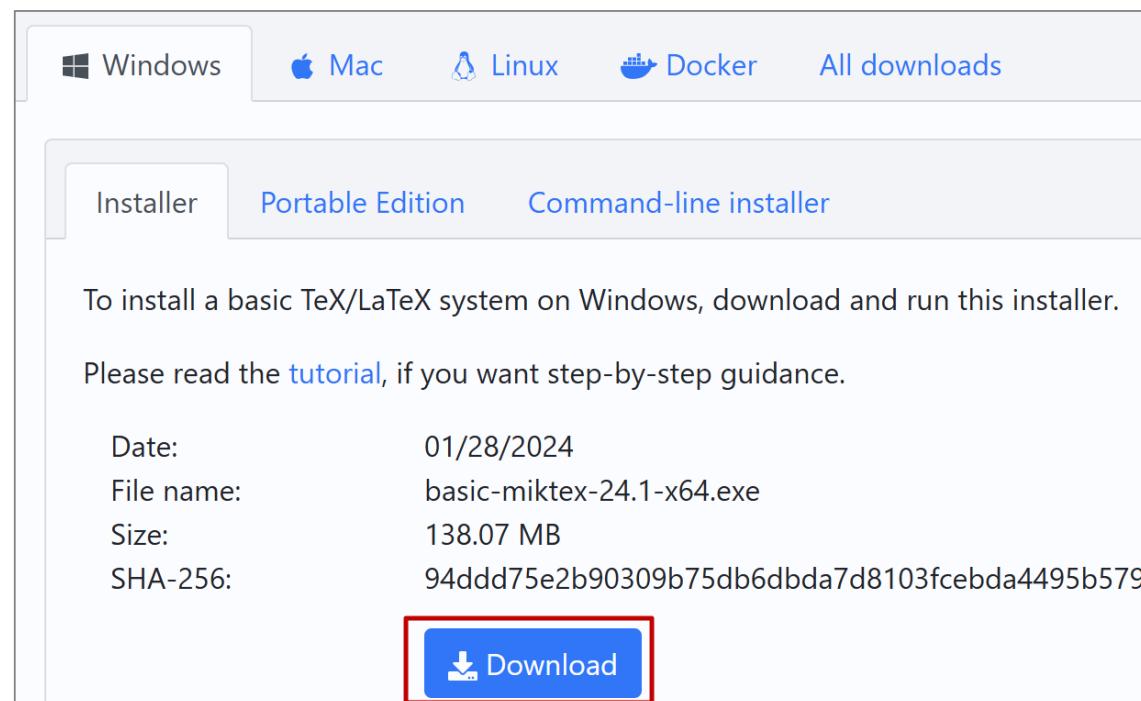
! LaTeX Error: Unicode character 大 (U+5927)
not set up for use with LaTeX.

Try other LaTeX engines instead (e.g., xelatex) if you are using pdflatex.
kdown.org/yihui/rmarkdown-cookbook/latex-unicode.html
Error: LaTeX failed to compile rbigdata.tex. See https://yihui.org/tinytex/r debugging tips. See rbigdata.log for more info.

```

PDF 中文錯誤 – 方法2 (Miktex)

- 下載 Miktex: <https://miktex.org/download>
- basic-miktex-24.1-x64.exe (138.07MB)



設定較繁瑣, 暫不使用

PDF 中文錯誤 – 方法3 推薦此法

- RWEPA | R Markdown 解決轉換中文字型PDF
- YouTube: <https://youtu.be/6Wc75BH02iE>
- LINK: <http://rwepa.blogspot.com/2024/09/rmarkdown-chinese-font-pdf.html>
- Slide: https://github.com/rwepa/DataDemo/blob/master/rmarkdown_chinese_font.pdf

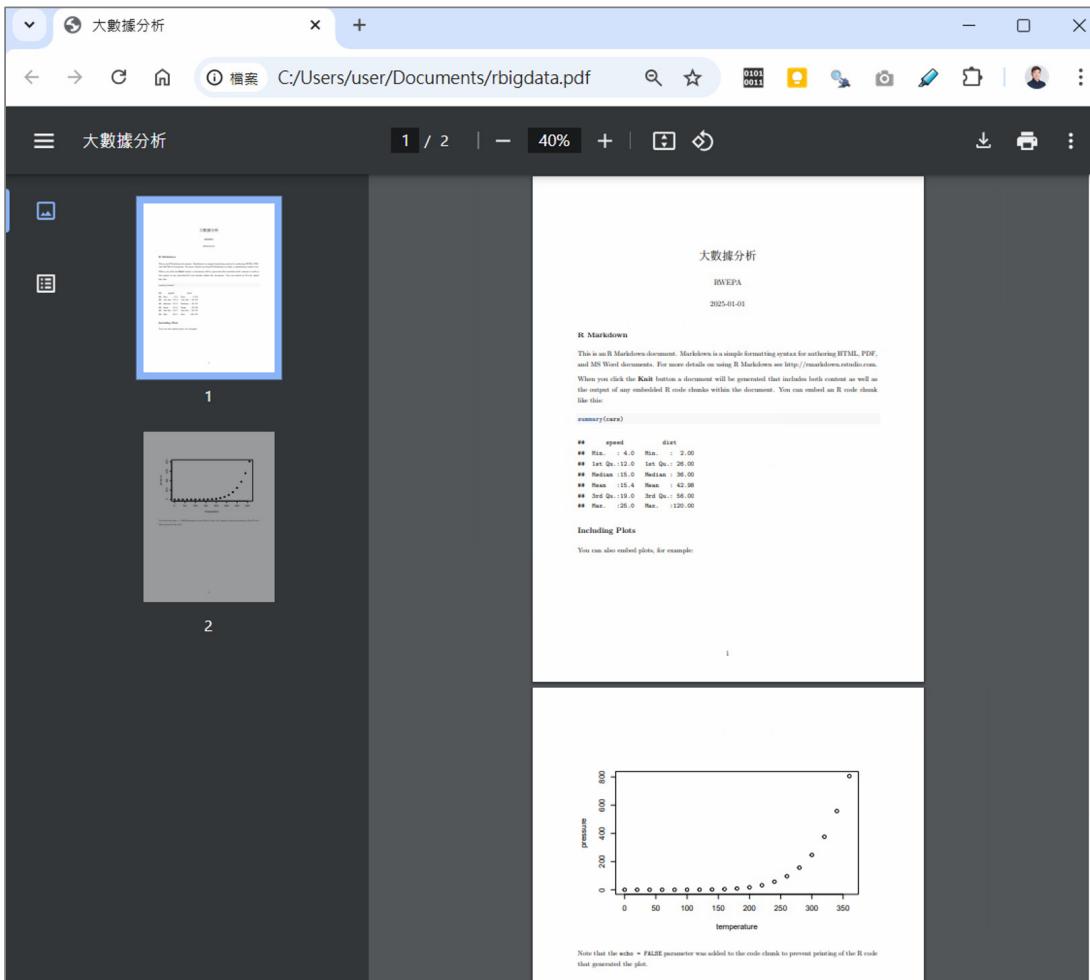
```
1 ---  
2 title: "大數據分析"  
3 author: "RWEPA"  
4 date: "2025-01-01"  
5 header-includes:  
6 - \usepackage{ctex}  
7 output:  
8   pdf_document: default  
9   html_document: default  
10  word_document: default  
11 ---
```

header-includes:
- \usepackage{ctex}

安裝三種字型

實作
練習

PDF 中文字型 - 方法3 完成版



The screenshot shows a PDF viewer window displaying a document titled "大數據分析". The document contains R Markdown code and two plots. The R Markdown code includes a summary of the "cars" dataset and a scatter plot of "pressure" vs "temperature".

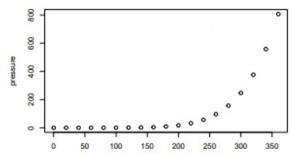
R Markdown

```
summary(cars)
```

```
##   speed   dist
## Min. : 4.0  Min. :  2.00
## 1st Qu.:12.0  1st Qu.: 26.00
## Median :19.0  Median : 36.00
## Mean   :19.4  Mean   : 43.98
## 3rd Qu.:29.0  3rd Qu.: 56.00
## Max. :50.0  Max. :120.00
```

Including Plots

You can also embed plots, for example:

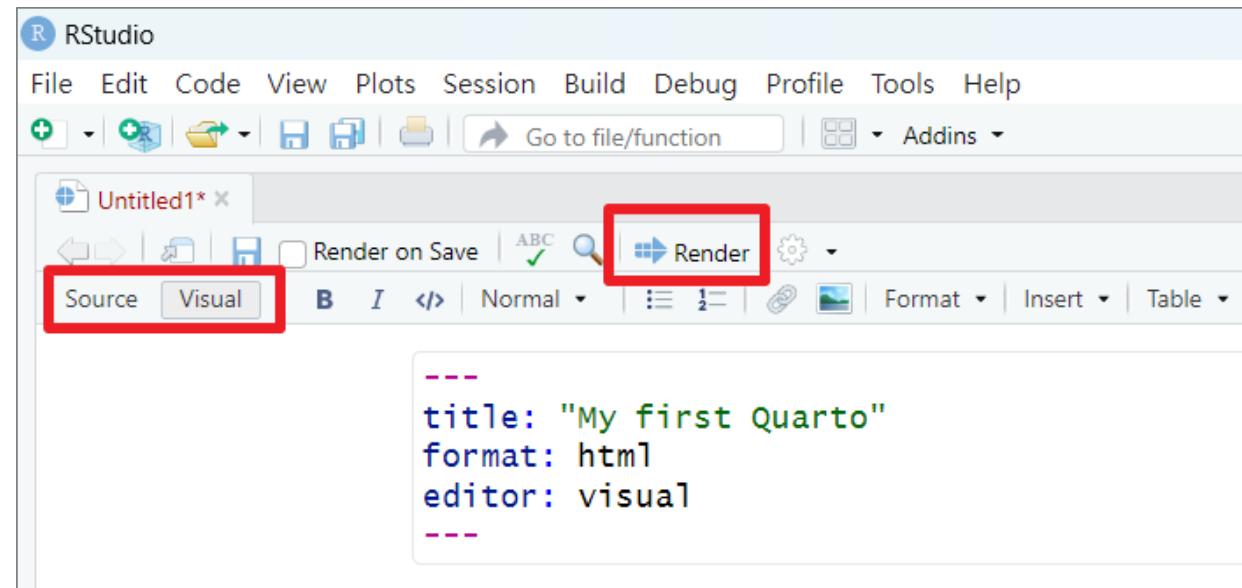


A scatter plot showing the relationship between "pressure" (Y-axis, 0 to 600) and "temperature" (X-axis, 0 to 350). The data points show a positive correlation, with most points clustered below 250 degrees Celsius and above 400 bars, and a few outliers at higher temperatures.

- 查詢輸出 rbigdata.pdf
- getwd()

Quarto, Since 2022

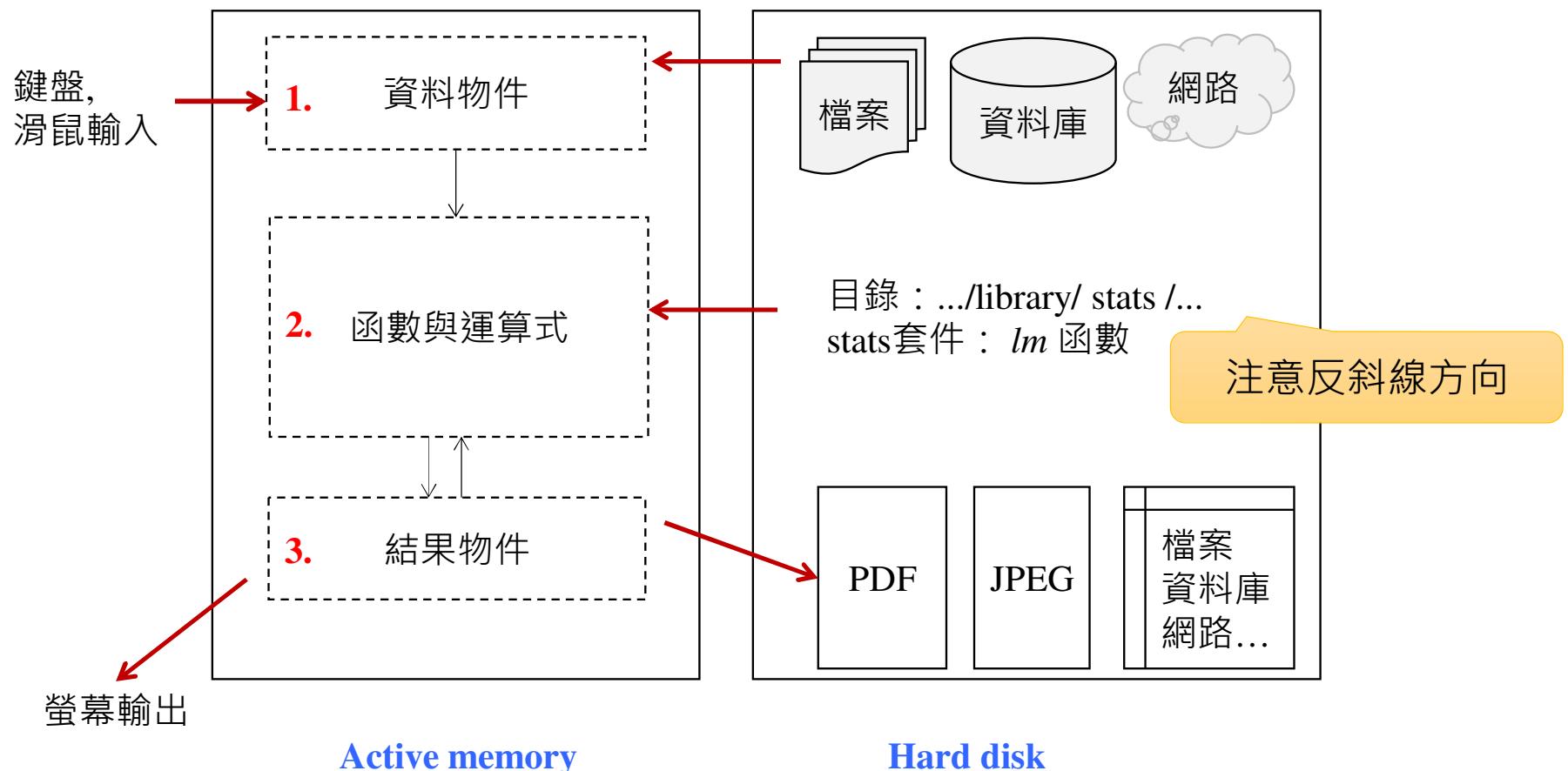
- An open-source scientific and technical publishing system
- File \ New File \ Quarto Document... (R + Python + ... 程式碼)



參考: <https://quarto.org/>, 2022.7.28.

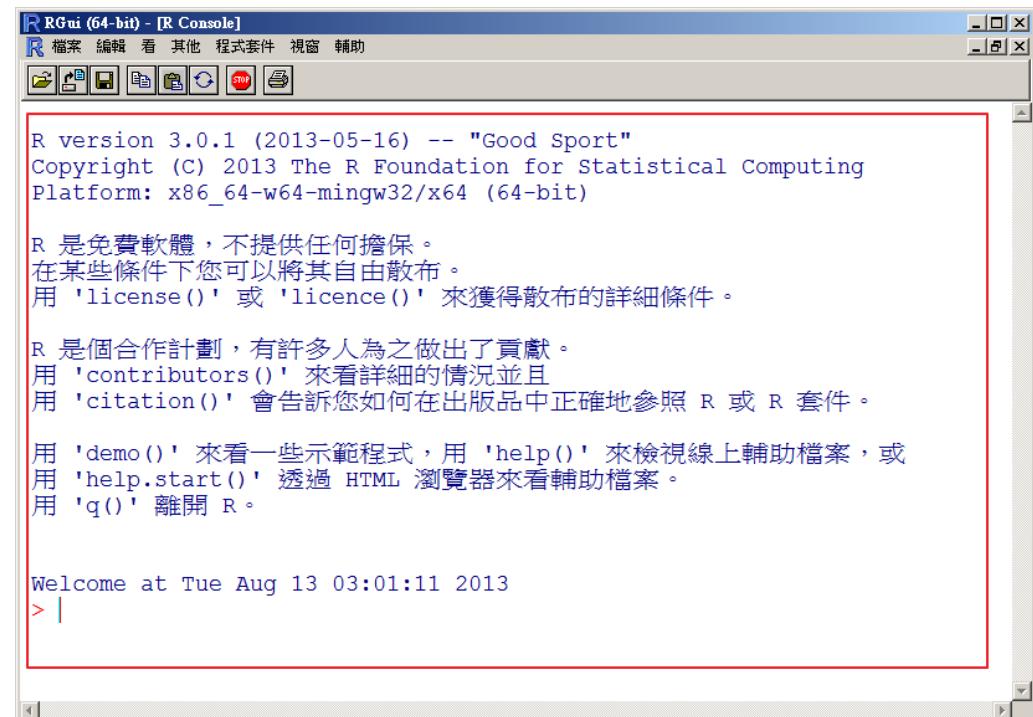
2.3 R基本概念

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 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)。
- 建議物件名稱中不要有 “.” (句點) 符號。
- 物件名稱起始字元不可為數字。
- 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 

2.4 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

type="n"

plot.default {graphics}

2.套件

1

2

The Default Scatterplot Function

3

R Documentation

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

7.參數

2.5 套件(Package)

- 類似 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"
>
```



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

R套件 - 46類別

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

Contributed Packages

(2024.12.28)

Available Packages

Currently, the CRAN package repository features **21860** 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, **46 views** are available.

46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: 2024.11.27 - 46個套件類別

CRAN Task View:

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

CRAN (Taiwan):

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

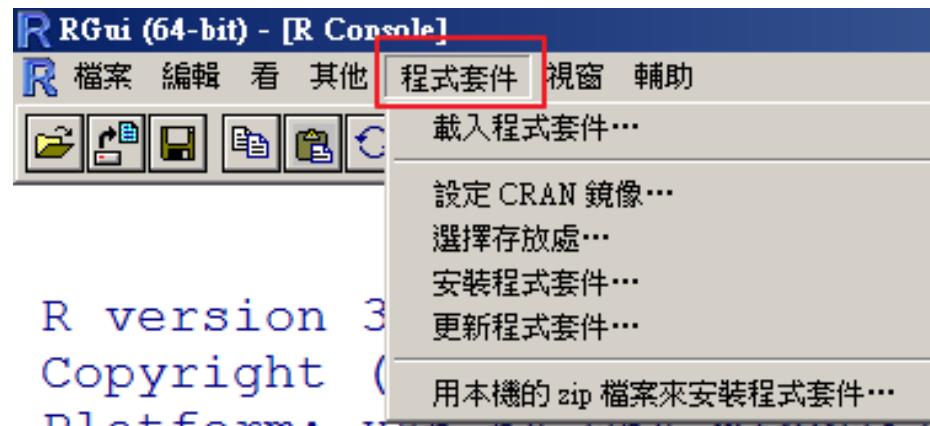
選取 CRAN 網站左側 [Packages]，套件區分成以下類別，中文說明如下：

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

```
> sessionInfo()
R version 4.4.2 (2024-10-31 ucrt)
Platform: x86_64-w64-mingw32/x64
Running under: Windows 11 x64 (build 26100)

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=
[5] LC_TIME=Chinese (Traditional)_Taiwan.utf8

time zone: Asia/Taipei
tzcode source: internal

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

套件安裝目錄

- .Library

```
>  
> # 預設套件安裝目錄  
> .Library  
[1] "C:/PROGRA~1/R/R-44~1.2/library"  
>
```

- .libPaths()

- 可能全部安裝在 R \ library

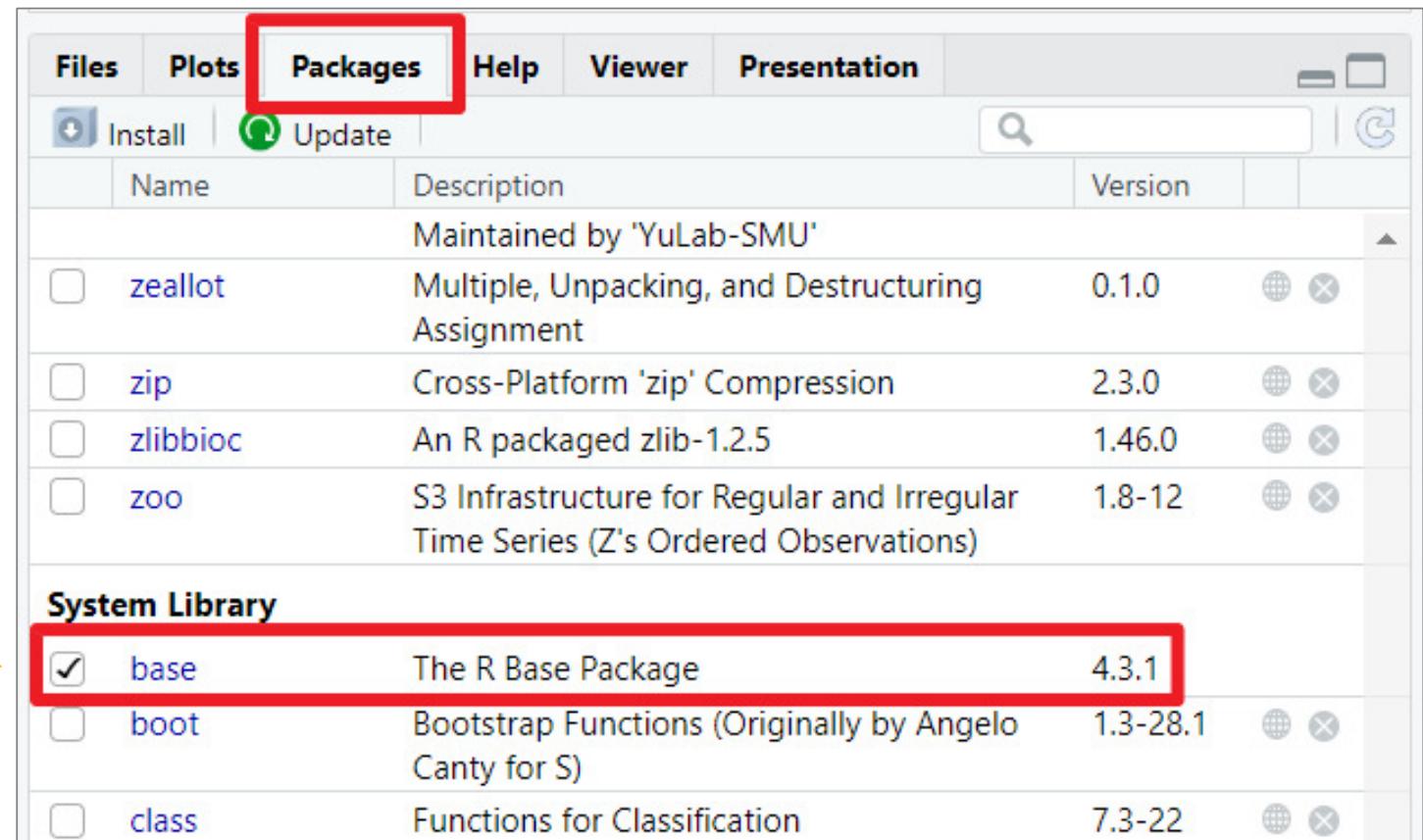
```
> # 套件安裝目錄  
> # 可能全部安裝在 R\library  
> .libPaths()  
[1] "C:/Users/rwepa/AppData/Local/R/win-library/4.4"  
[2] "C:/Program Files/R/R-4.4.2/library"
```

已安裝套件

```
> # 已安裝套件
> myinstalled <- installed.packages()
> class(myinstalled) # "matrix" "array"
[1] "matrix" "array"
> dim(myinstalled)
[1] 1512   16
> mypackage <- myinstalled[, 1] # matrix[列, 行]
> mypackage[1:10]
      abind          actuar          ada
      "abind"        "actuar"       "ada"
      adabag    additivityTests    ade4
      "adabag"  "additivityTests"  "ade4"
      ade4TkGUI      adegenet    adegraphics
      "ade4TkGUI"    "adegenet"  "adegraphics"
      adehabitatLT
      "adehabitatLT"
>
```

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

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, which is checked and highlighted with a red border.

Name	Description	Version
zeallot	Maintained by 'YuLab-SMU' Multiple, Unpacking, and Destructuring Assignment	0.1.0
zip	Cross-Platform 'zip' Compression	2.3.0
zlibbioc	An R packaged zlib-1.2.5	1.46.0
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



實作
練習

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

練習 class()

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

2.6 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

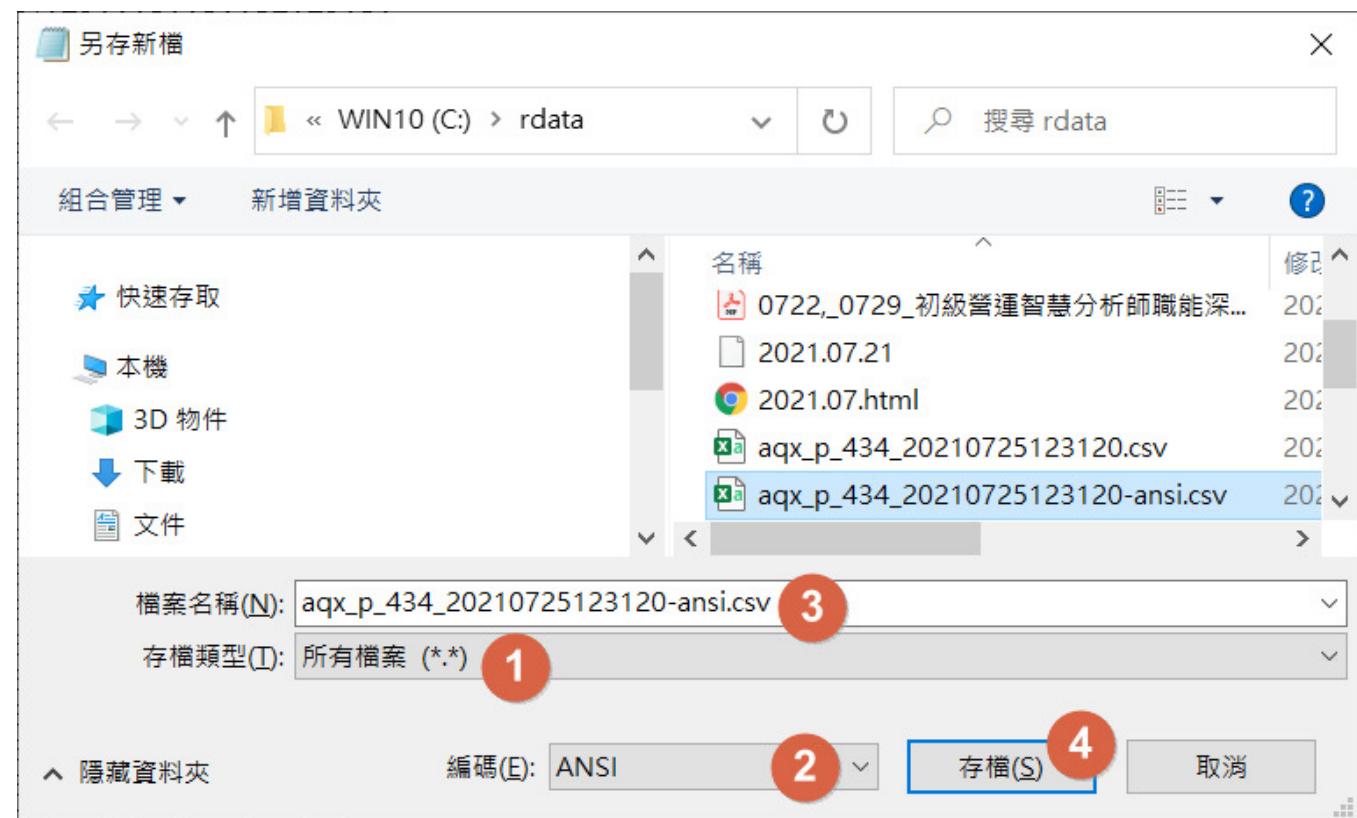
R-v3.x早期版本匯入 UTF-8 文字檔會有錯誤

```

> myfile <- "aqx_p_434_20210725123120.csv"
>
> aq <- read.table(myfile, header=TRUE, sep=",") # error!
      1      2      3
Error in scan(file = file, what = what, sep = sep, quote = quote, dec = dec, :
  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
```

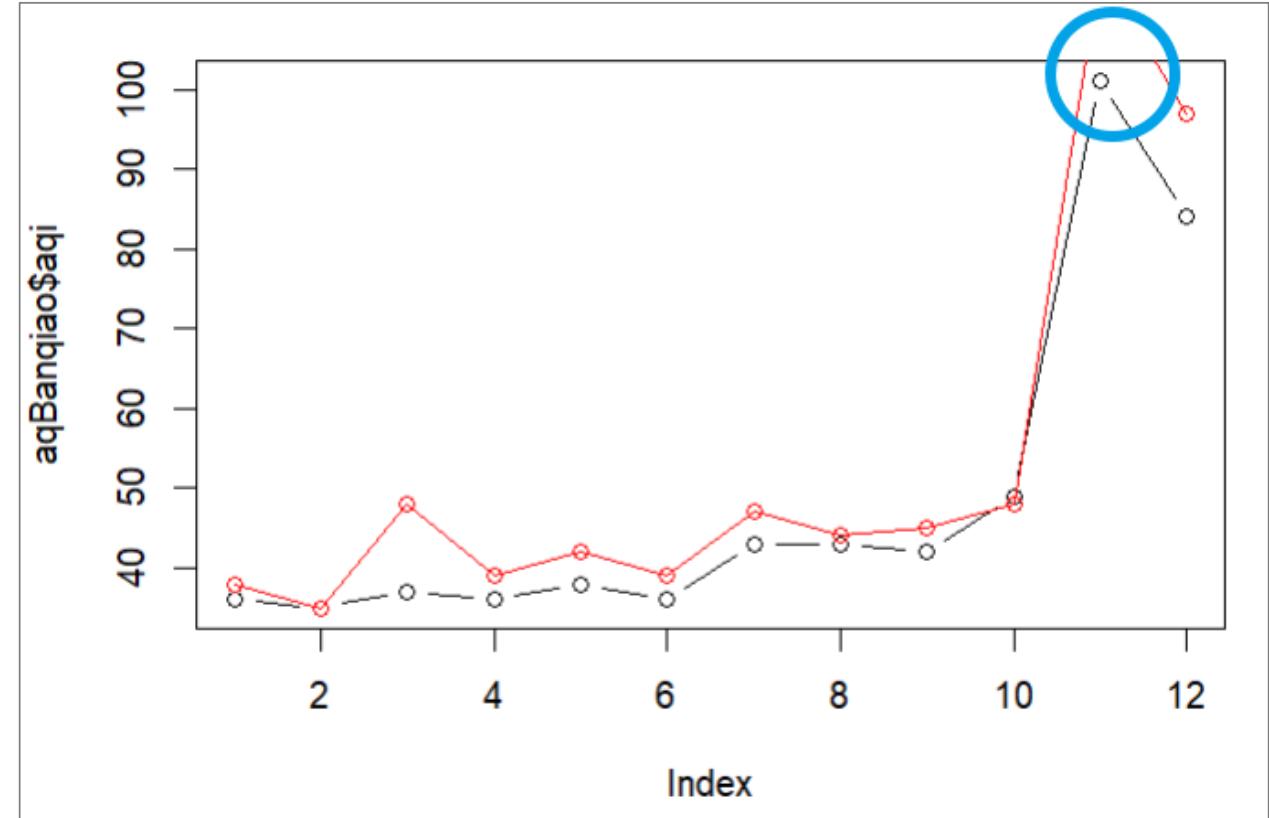
資料篩選與排序

`==` 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

步驟 5. 汇出資料

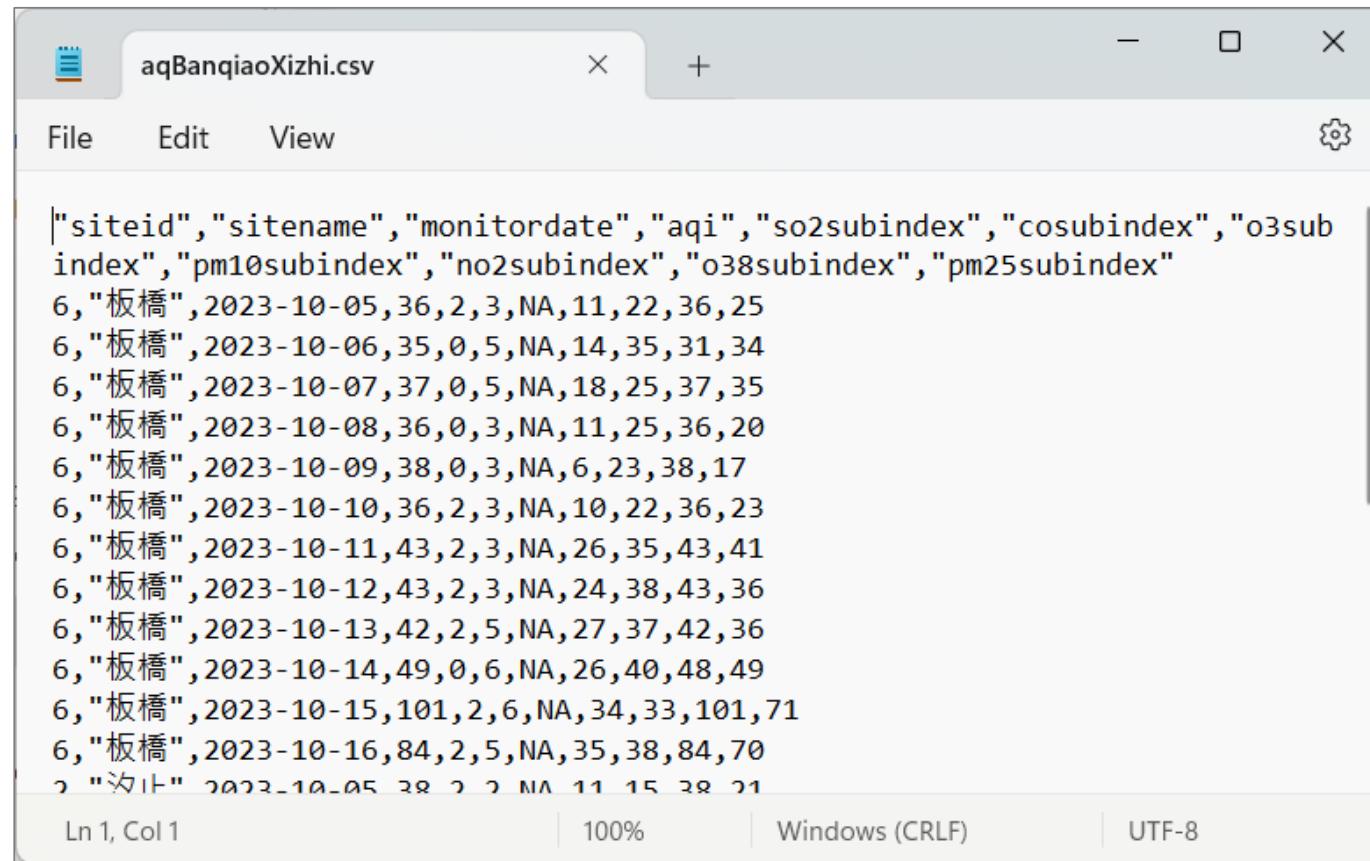
rbind: 列合併

```
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 'File origin' dropdown is set to '65001: Unicode (UTF-8)' (highlighted with a red box). The 'Text' section displays the first four rows of the CSV 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 (\geq 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 (\geq 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

- 匯入 SPSS {foreign}
- foreign 套件可讀取以下檔案格式:
 - 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

謝謝您的聆聽

Q & A



李明昌

alan9956@gmail.com

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