

資料倉儲導論-第2章 免費R語言,RStudio簡介,下載與安裝

大數據分析

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



李明昌博士

alan9956@gmail.com

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

大綱

2.1 資料科學的心法

2.2 R語言簡介

2.3 R語言下載與安裝

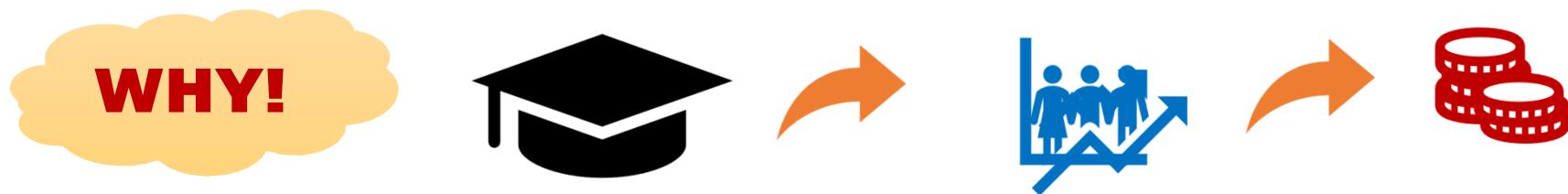
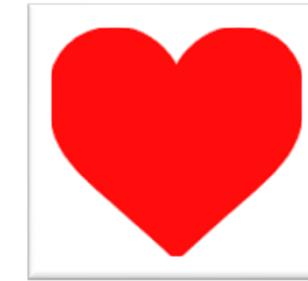
2.4 RStudio簡介

2.5 RStudio下載與安裝

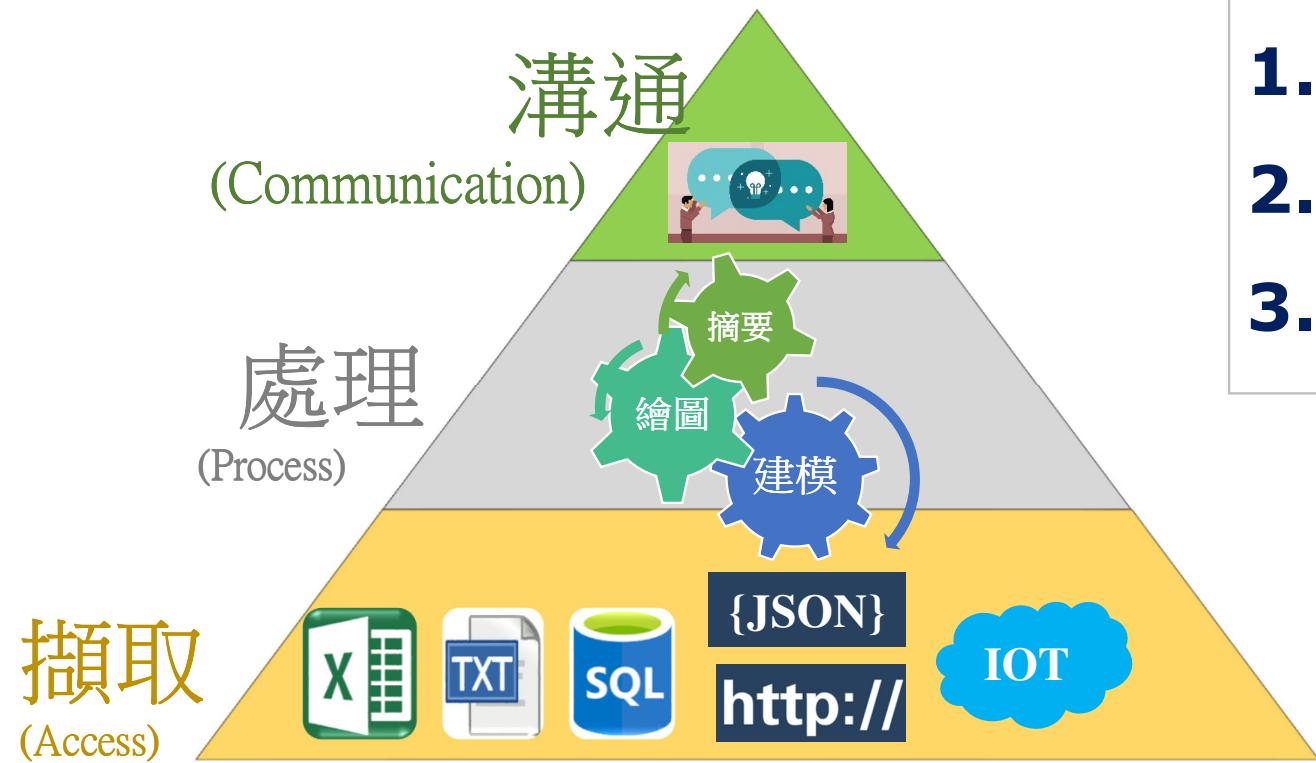
2.1 資料科學的心法

資料科學的心法

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



★★★資料分析架構→APC方法



- 1.
- 2.
- 3.

R 入門資料分析與視覺化應用(7小時28分鐘)

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

課程提供教學範例的原始程式檔案與資料集



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

R 商業預測應用(8小時53分鐘)

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



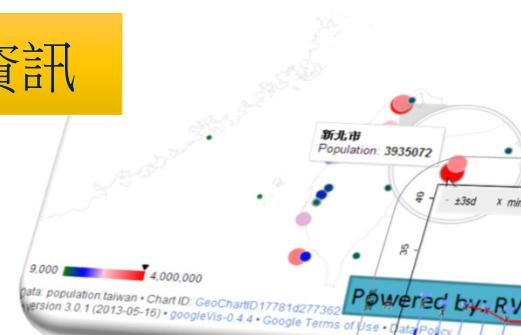
課程提供教學範例的原始程式檔案與資料集

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

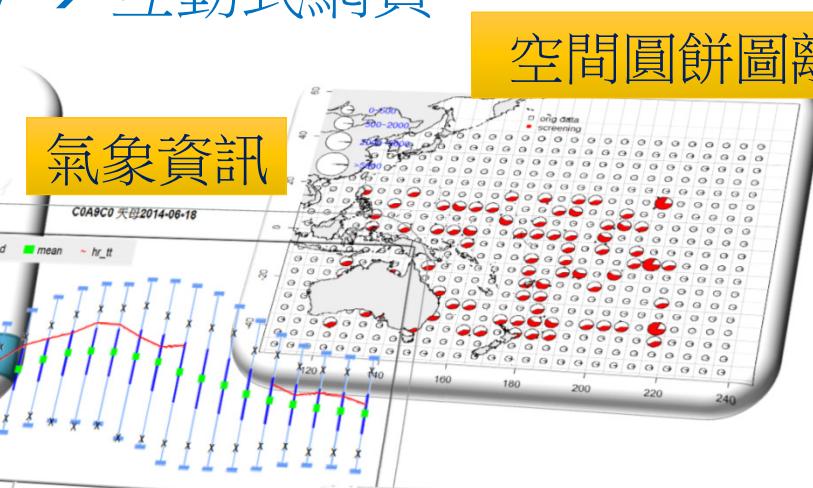
資料分析/視覺化應用

R + shiny → 互動式網頁

地理資訊



氣象資訊



空間圓餅圖離群值分析

保險預測



顧客連結資訊



中央氣象局 1,600萬筆資料

網頁呈現



客製化選單

R統計運算

保險預測模型

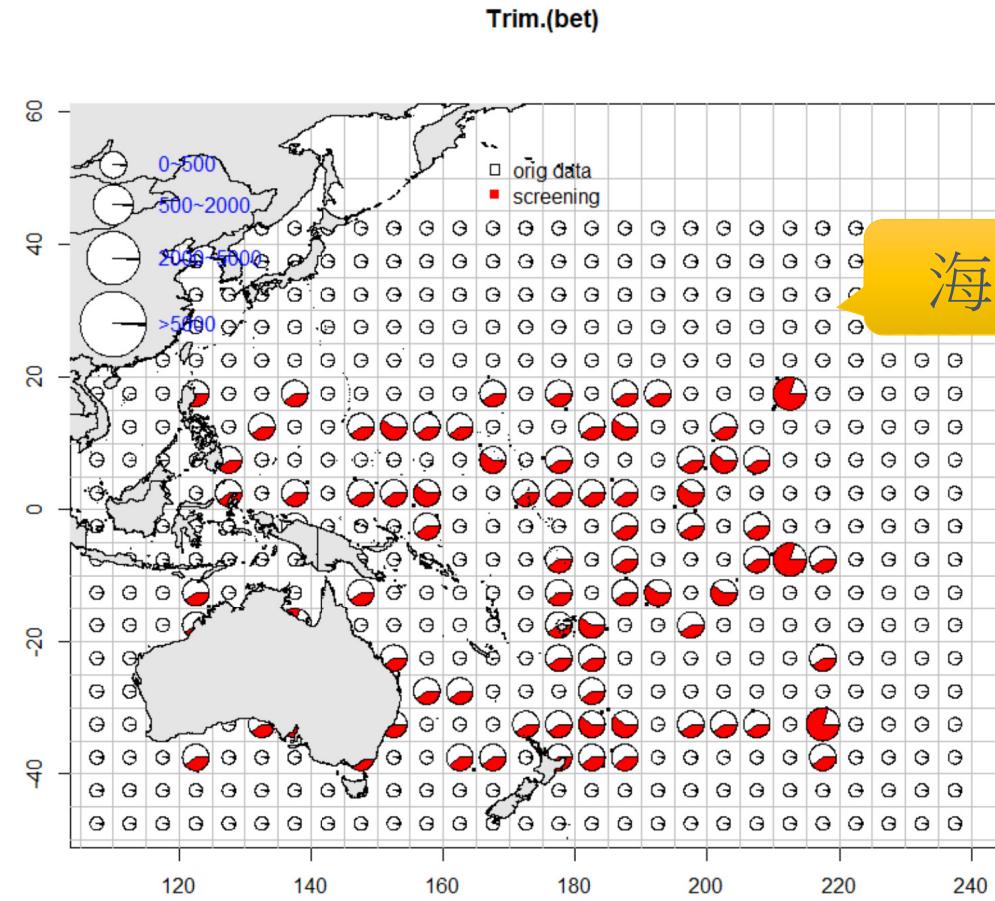
The screenshot shows a web-based data analysis interface for the iInsurance platform. The top navigation bar includes links for '檔案上傳', '資料處理', '統計圖表', '模型評估', and '預測模型'. A red box highlights the '預測模型' dropdown. Below it, a yellow callout box labeled '機率模型閾值調整' points to a slider control labeled '機率模型閾值' with a value of '0.1'. Another red box highlights the '檢視結果' button in the '預測資料上傳' section. A large yellow callout box labeled '預測結果' points to the rightmost columns of the main table, which are labeled '預測機率' and '理賠'. The table displays 10 entries of predicted results, each with various demographic and vehicle information fields.

性別	女性	車輛種類	私家車	曝露風險		無索償折扣	被保險人年齡	私家車 一車齡 0	私家車 一車齡 1	私家車 一車齡 2	私家車 -車齡 0_1_2 組合	車齡 組合	車齡 0_1_2	預測機率	理賠	
				曝露風險對數	對數											
M	0	A	1	0.9144422	-0.08944106	50	4	1	0	0	1	0	2	0.1069	有	
M	0	A	1	0.8158795	-0.20348856	20	4	0	0	1	1	2	2	0.1441	有	
3	M	0	A	1	0.8377823	-0.17699695	50	3	0	0	1	1	2	2	0.1866	有
4	M	0	A	1	0.4325804	-0.83798702	50	6	0	1	0	1	1	2	0.0944	無
5	M	0	A	1	0.7173169	-0.33223755	50	4	0	0	1	1	2	2	0.1218	有
6	M	0	A	1	0.8377823	-0.17699695	50	4	0	0	1	1	2	2	0.1495	有
7	M	0	A	1	0.8487337	-0.16400975	50	5	0	0	1	1	2	2	0.1422	有
8	F	1	A	1	0.8268309	-0.19015503	10	3	0	0	1	1	2	2	0.1733	有
9	M	0	A	1	0.7145791	-0.33606164	0	5	1	0	0	1	0	2	0.0694	無
10	M	0	A	1	0.3340178	-1.09656101	0	3	0	0	1	1	2	2	0.0783	無

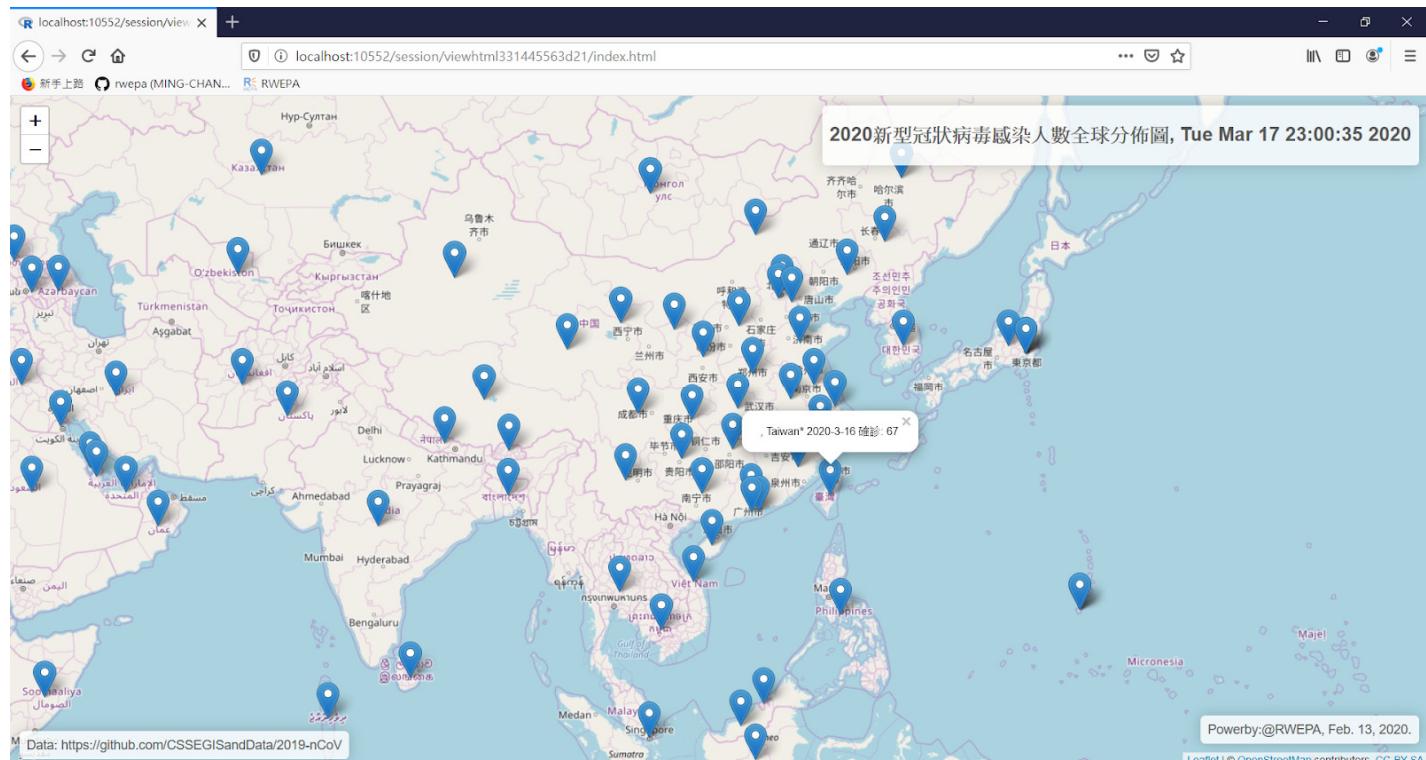
Showing 1 to 10 of 12 entries

127.0.0.1:6177/#tab-9487-2

空間圓餅圖離群值分析



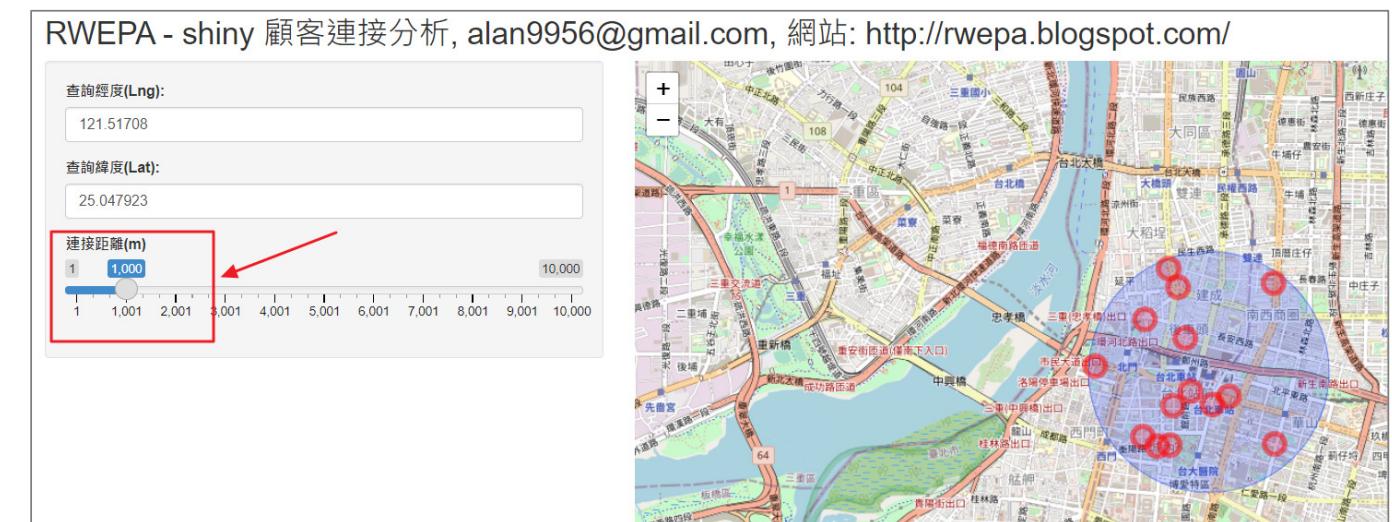
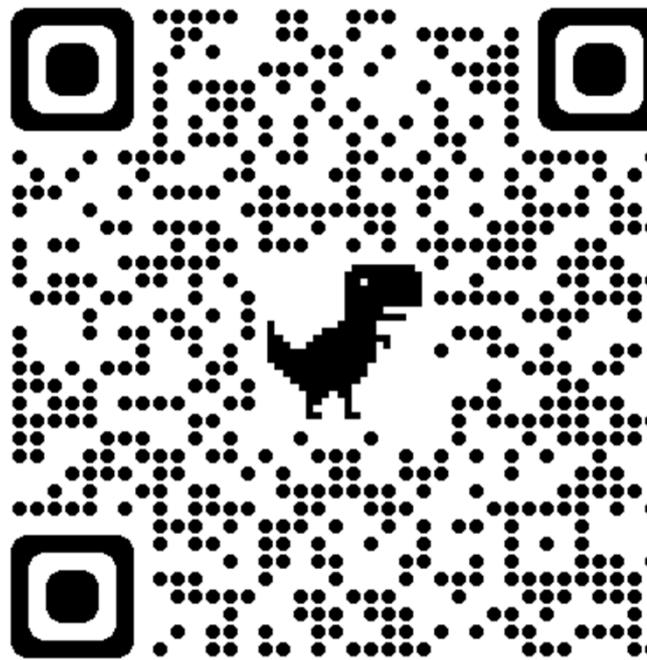
2020新型冠狀病毒視覺化



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

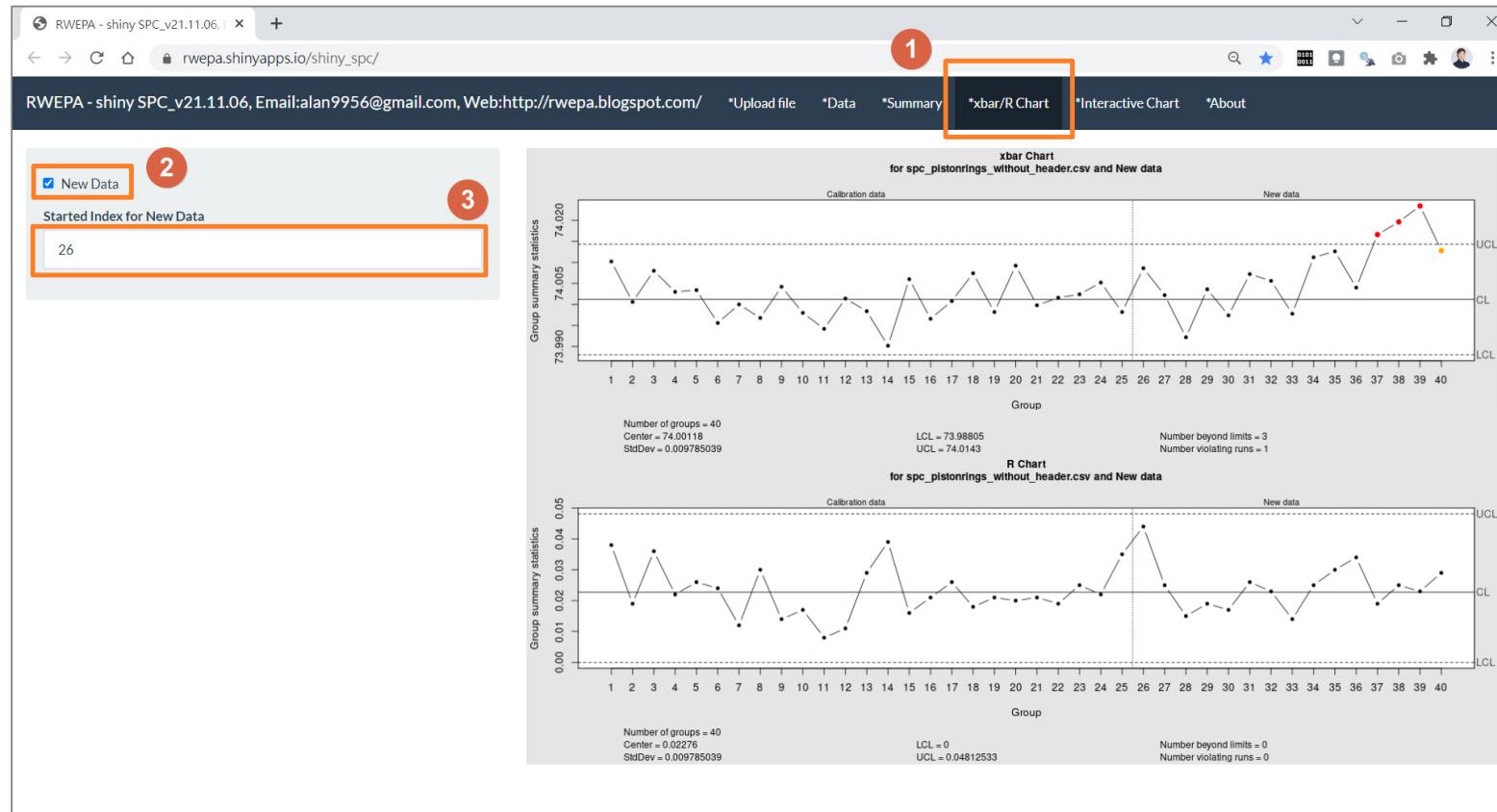
shiny 顧客連接分析

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



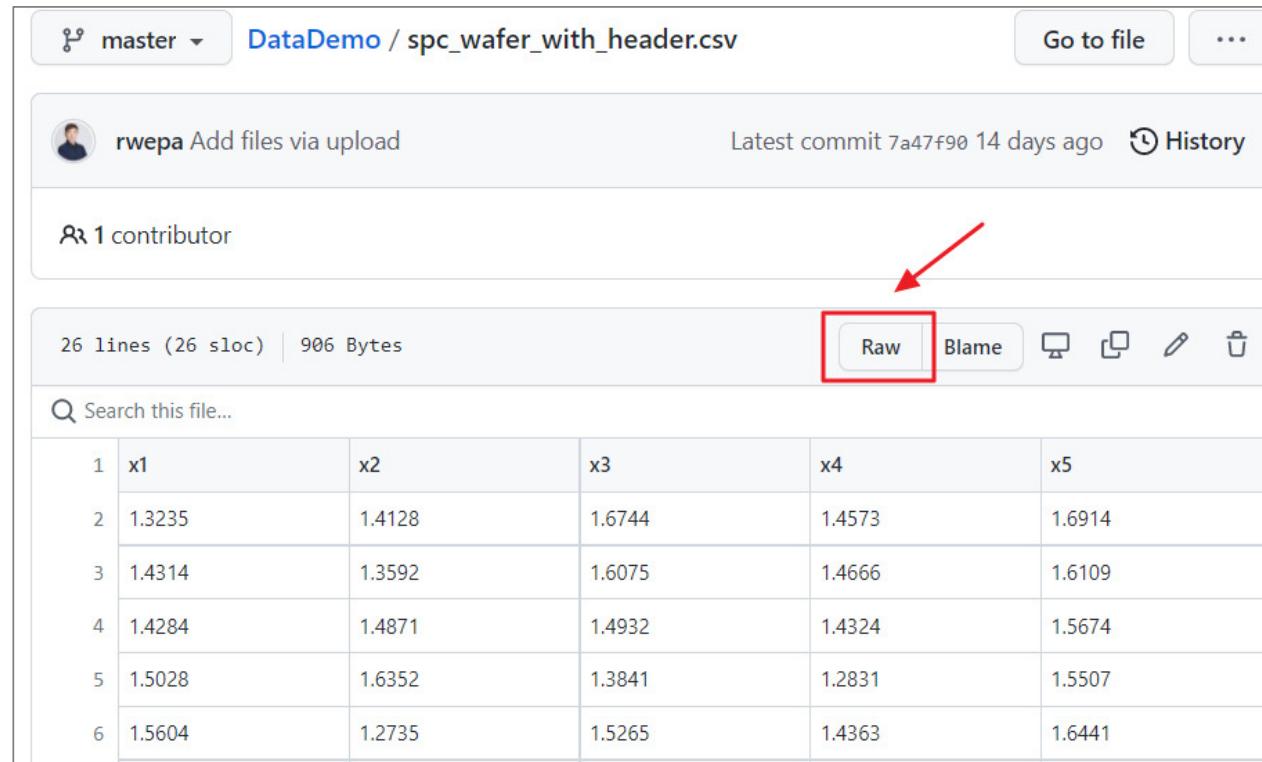
品質管制圖(quality control chart)應用

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



品質管制圖應用 (續)

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

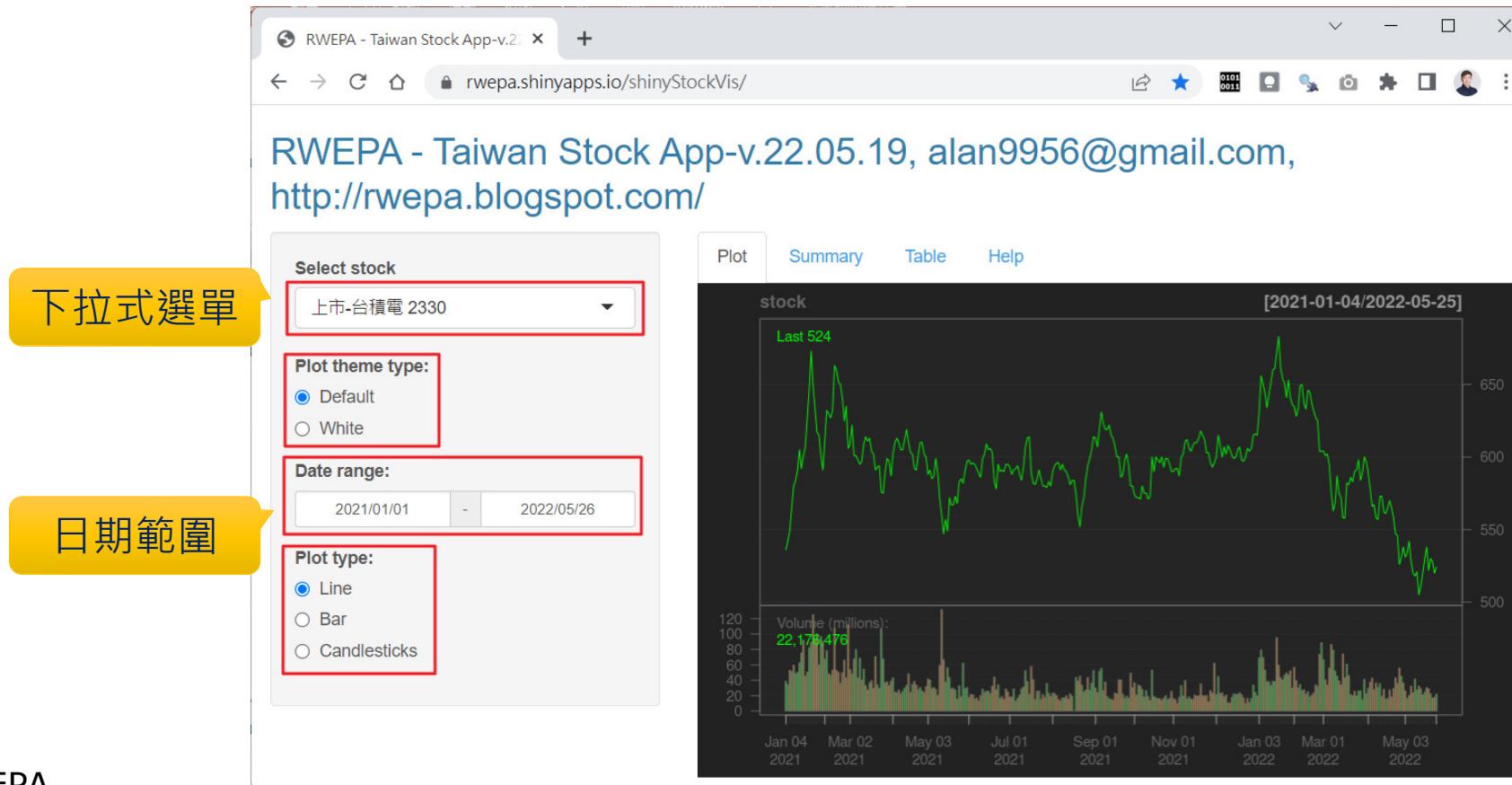


The screenshot shows a GitHub repository page for 'DataDemo'. The file 'spc_wafer_with_header.csv' is displayed. The GitHub interface includes a header with 'master', a file path 'DataDemo / spc_wafer_with_header.csv', and buttons for 'Go to file' and '...'. Below this is a commit history section showing a single commit by 'rwepa' with the message 'Add files via upload' and a date of '14 days ago'. A red arrow points to the 'Raw' button in the toolbar below the file statistics. The statistics show '26 lines (26 sloc) | 906 Bytes'. The table below contains six rows of data:

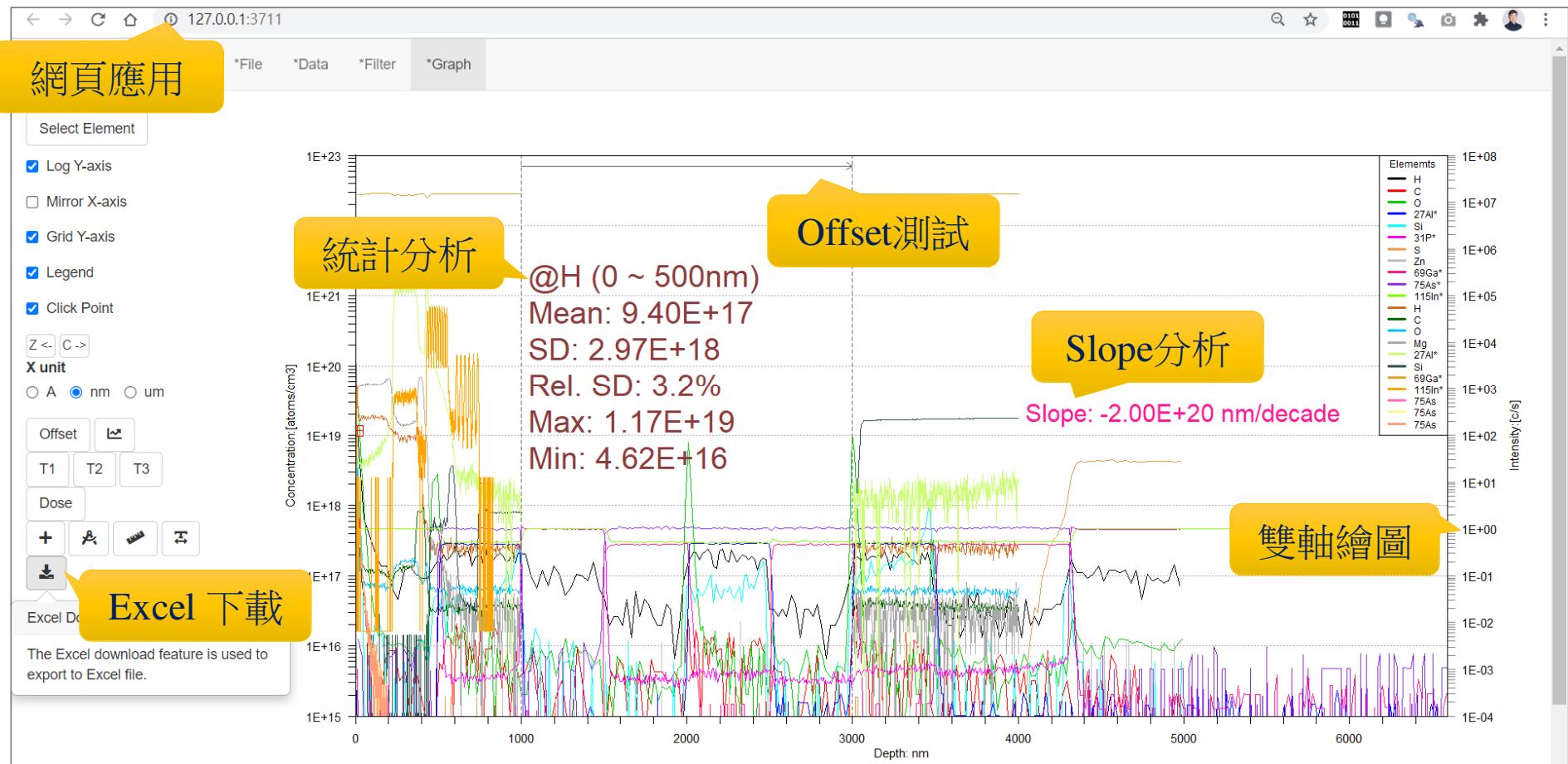
	x1	x2	x3	x4	x5
2	1.3235	1.4128	1.6744	1.4573	1.6914
3	1.4314	1.3592	1.6075	1.4666	1.6109
4	1.4284	1.4871	1.4932	1.4324	1.5674
5	1.5028	1.6352	1.3841	1.2831	1.5507
6	1.5604	1.2735	1.5265	1.4363	1.6441

Taiwan Stock App

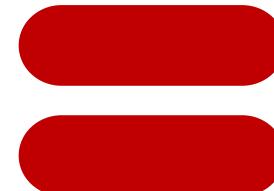
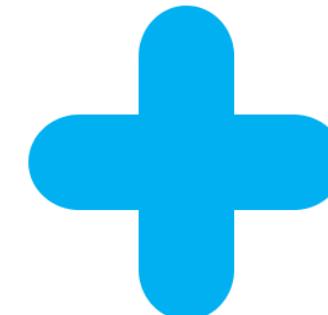
- <https://rwepa.shinyapps.io/shinyStockVis/>



離子資料分析與視覺化應用

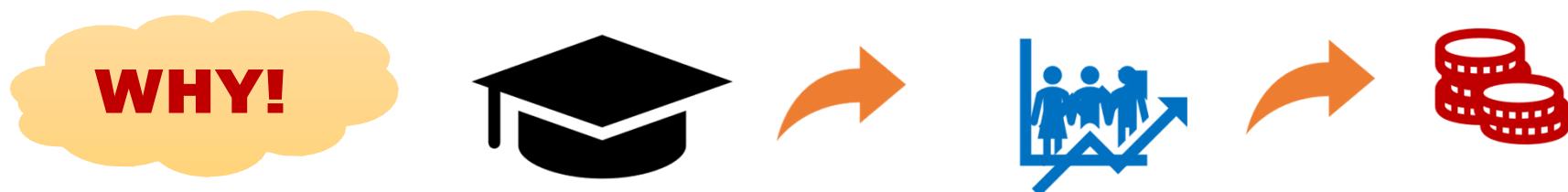


學習目標



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
- 遇到問題時，想辦法尋找答案
- 掌握 APC方法
- 掌握 摘要, 繪圖, 建模
- 參考網路應用文章 (進階) & 學術論文



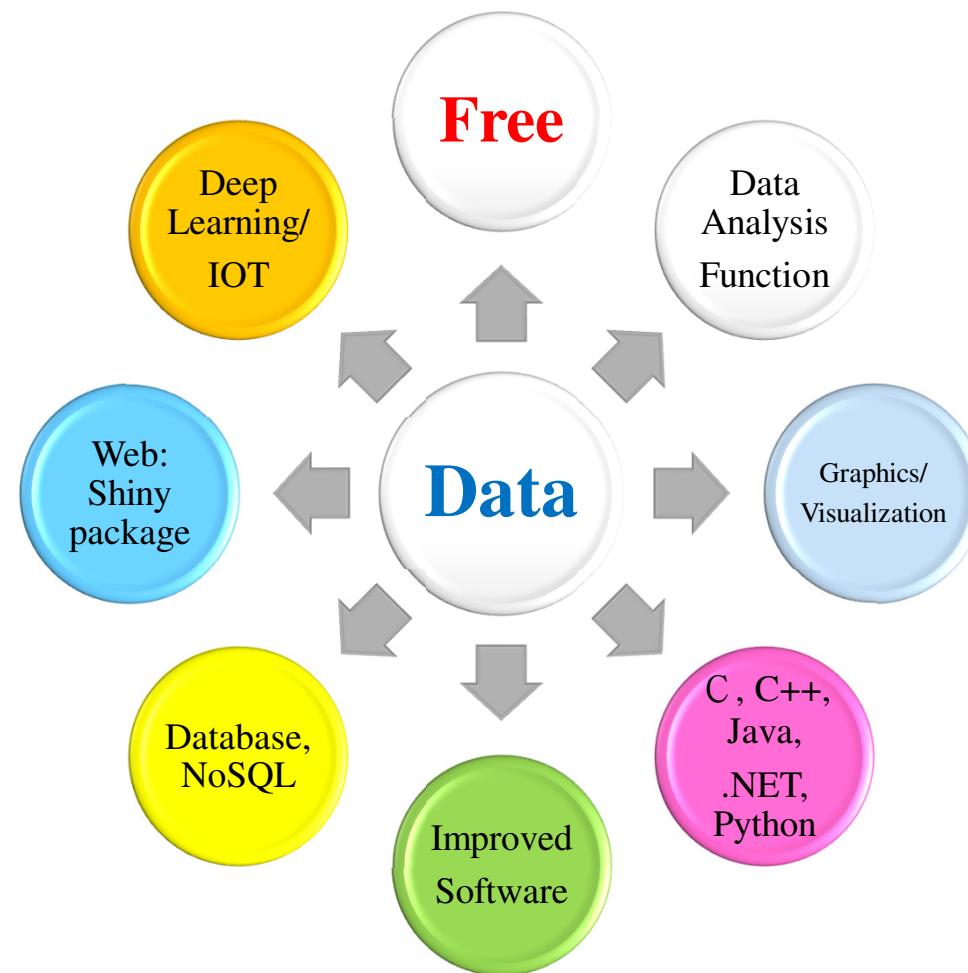
2.2 R語言簡介

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
 - 2022年6月 – R 4.2.1



R-八大功能



2.3 R語言下載與安裝

R官方網頁



The screenshot shows the official website for The R Project for Statistical Computing. At the top left is the R logo. Below it is a navigation menu with the following items: [Home], Download, CRAN, 繪圖 (highlighted in red), R Project, About R, Logo, Contributors, What's New?, and Reporting. A yellow speech bubble contains the word "下載".

The R Project for Statistical Computing

Getting Started

統計計算

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

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

..

R-下載

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

Taiwan

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

- 選取 Download R for Windows

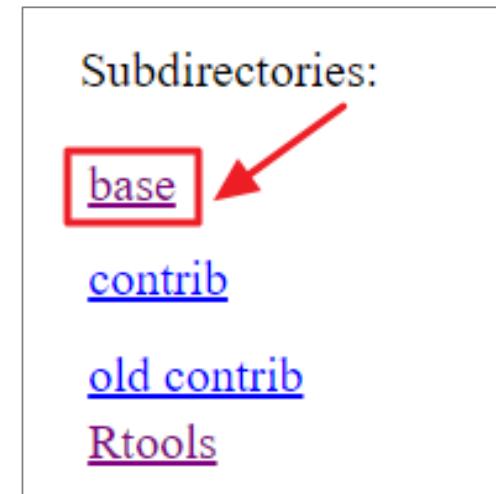


- [Download R for Linux](#)
- [Download R for \(Mac\) OS X](#)
- [Download R for Windows](#)



R-下載 (續)

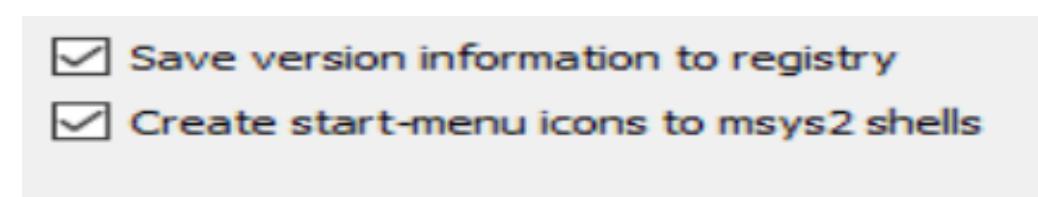
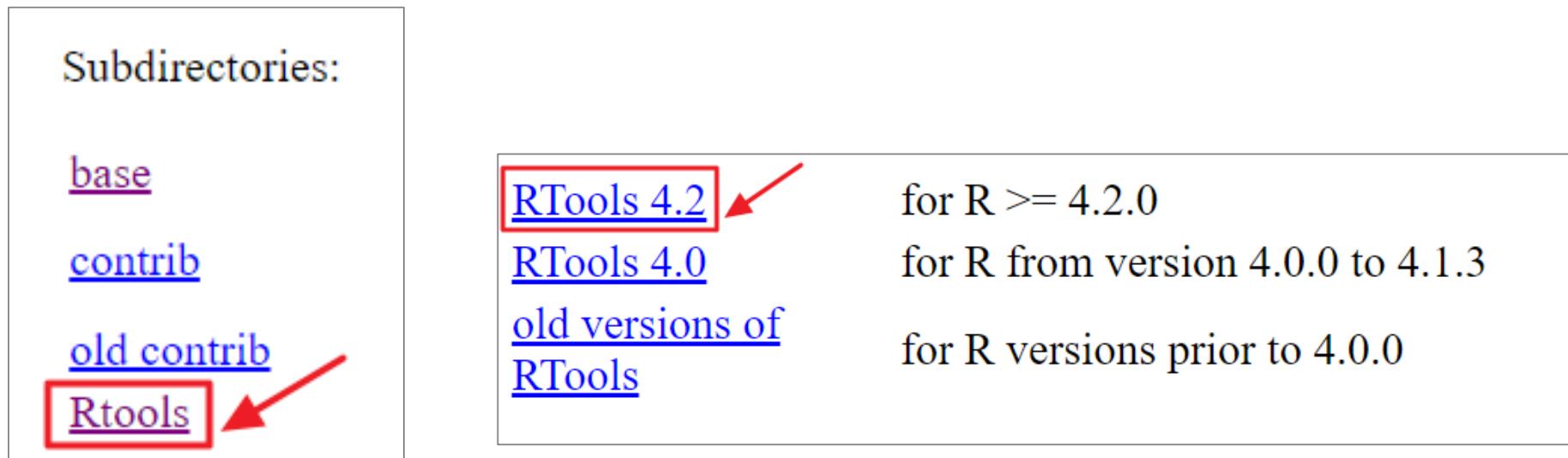
- 選取 base → 下載 [R-4.2.1-win.exe]



- 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: 一定要保留原路徑 C:\rtools42
- <https://cloud.r-project.org/bin/windows/Rtools/rtools42/files/rtools42-5253-5107-signed.exe>



2個都打勾

✓ 安裝 R

✓ 安裝 Rtools



R Manuals (使用手冊)

The R Manuals

edited by the R Development Core Team.

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

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

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

contributed documentation
(貢獻文件, 免費啦)

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

R Manuals (續)

Contributed Documentation

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

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

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

English Documents

Documents with more than 100 pages:

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



好書!



實作練習

R 執行畫面

The screenshot shows the R GUI interface. The R Console window displays the R version information and a sample plot command. The R Graphics window shows a highly oscillatory line plot generated by the command.

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

```
demo(graphics)
```

```
demo(persp)
```

大小寫須一致

安裝R, 登入名稱:

1. 不要使用空格
2. 不要使用中文字型

R 功能表

檔案



編輯



現行目錄 getwd()

輔助



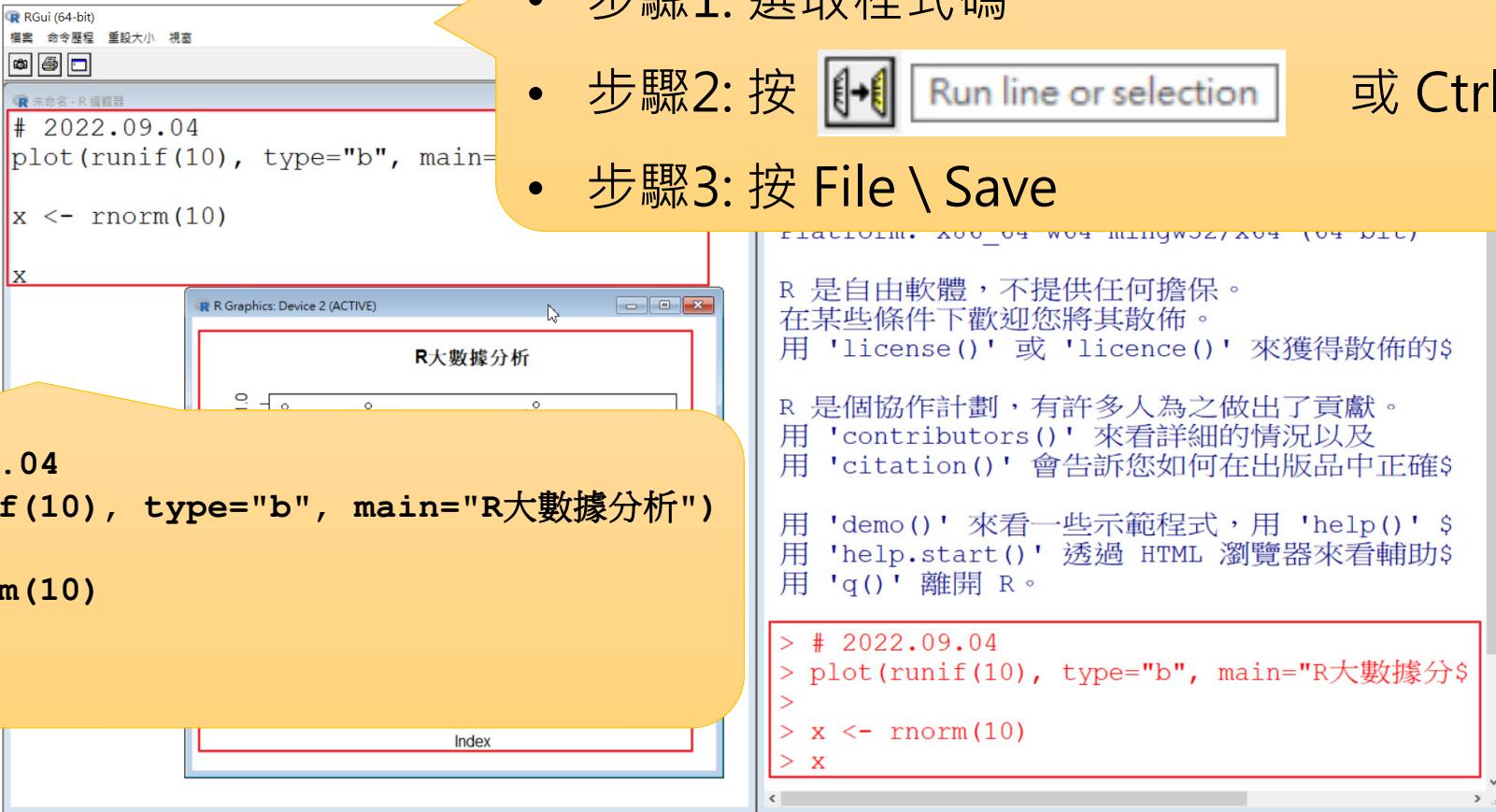
儲存控制台-文字檔



實作練習

新增R檔案練習

- 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



2022.09.04
plot(runif(10), type="b", main="R大數據分析")

x <- rnorm(10)

x

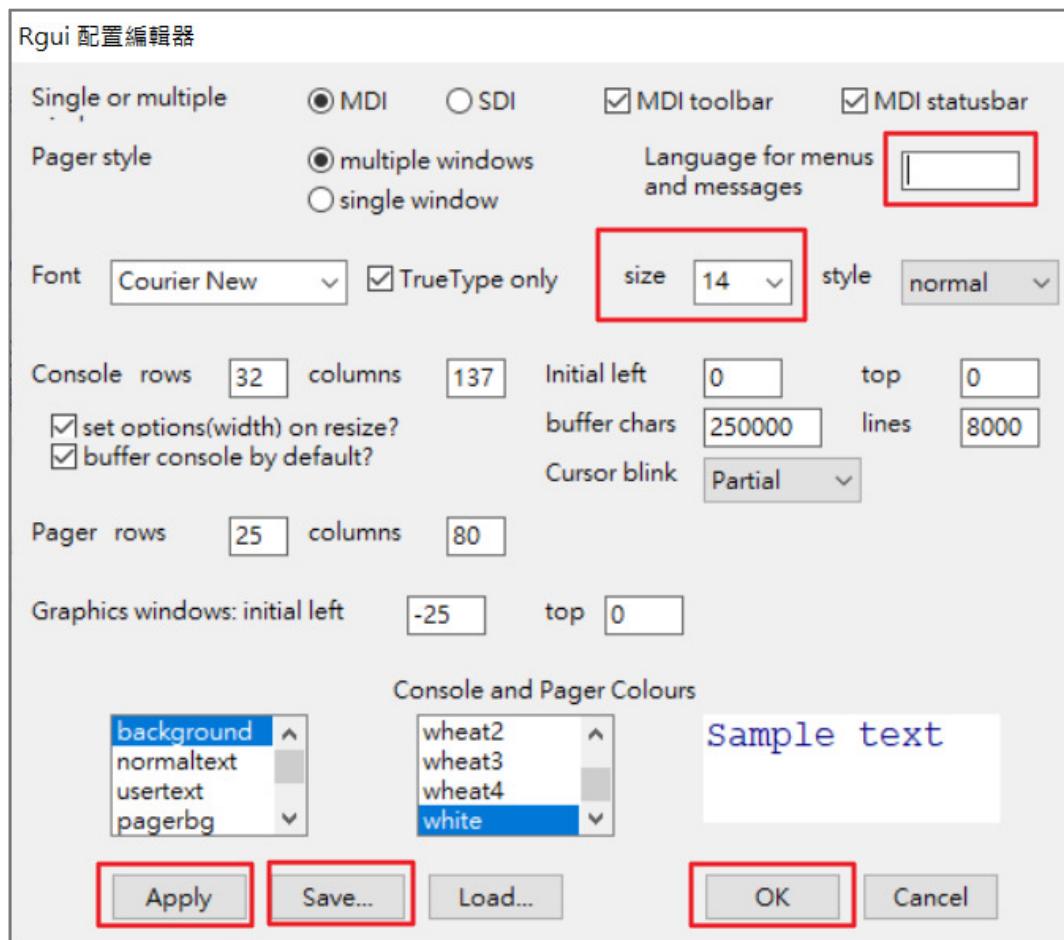
R 是自由軟體，不提供任何擔保。
在某些條件下歡迎您將其散佈。
用 'license()' 或 'licence()' 來獲得散佈的\$

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

用 'demo()' 來看一些示範程式，用 'help()' \$
用 'help.start()' 透過 HTML 瀏覽器來看輔助\$
用 'q()' 離開 R。

```
> # 2022.09.04
> plot(runif(10), type="b", main="R大數據分析")
>
> x <- rnorm(10)
> x
```

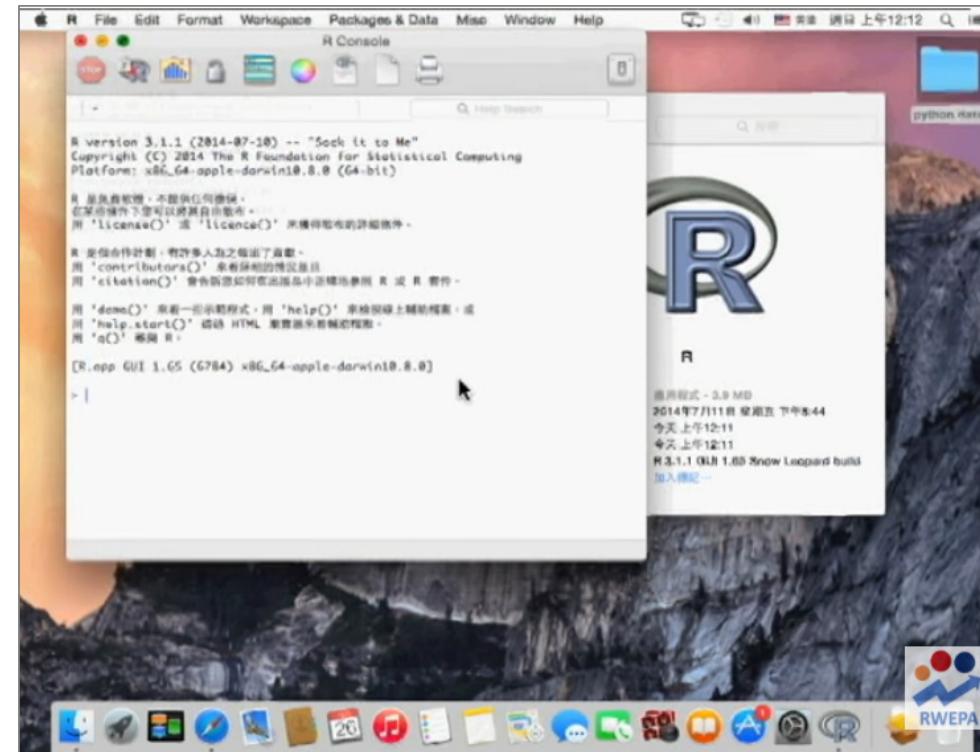
編輯 \ GUI 偏好設定



- Language: en 英文
- size: 字型大小

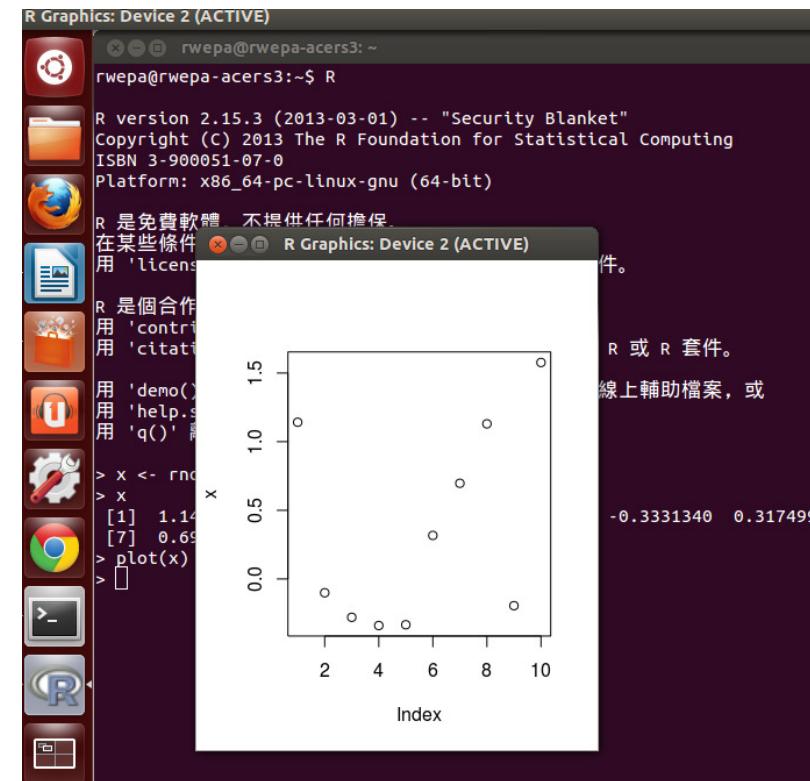
R for Mac

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



R for Ubuntu

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



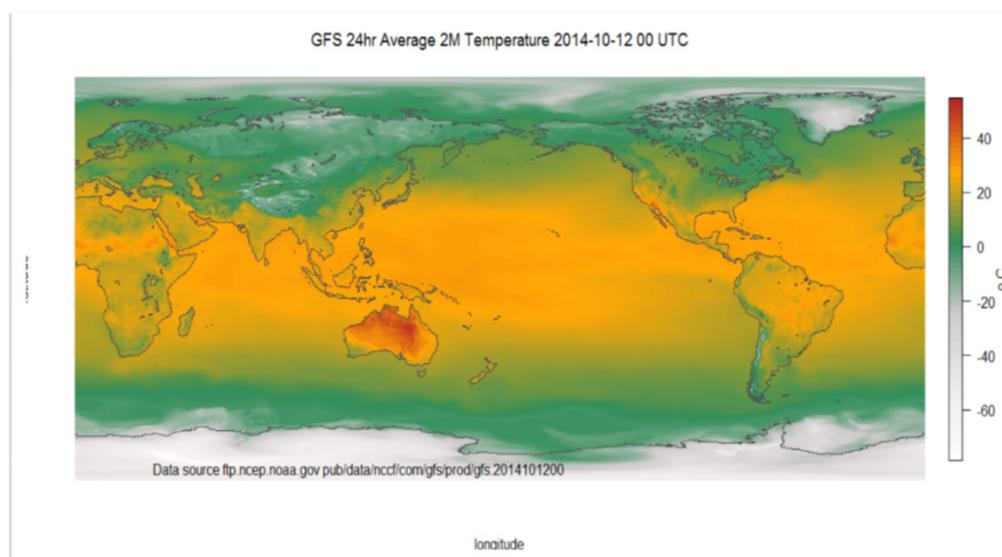
2.4 RStudio簡介

- 更名為 posit
<https://posit.co/>



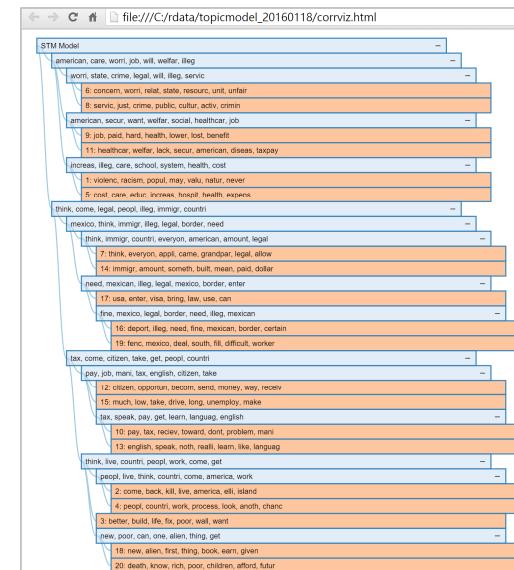
整合式開發環境 - RStudio

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



視覺化應用

(全球2M氣溫圖)



主題模型

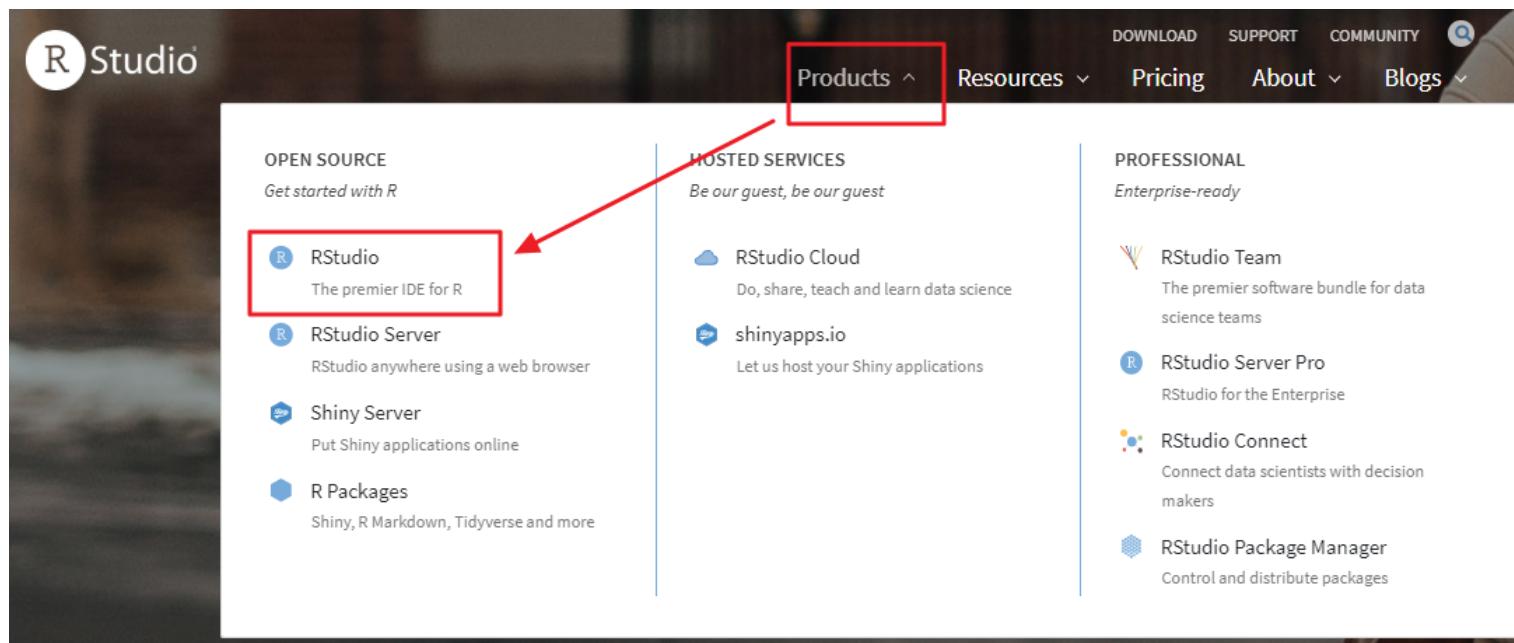
RStudio – 特性

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

2.5 RStudio下載與安裝

RStudio 下載

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



RStudio 下載 (續)

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

單機版 /year
Named Users
伺服器版本

[DOWNLOAD](#)
[BUY](#)
[Learn more](#)

[DOWNLOAD](#)
[BUY](#)
[Learn more](#)



免費版





RStudio 下載 (續)

RStudio Desktop 2022.07.1+554 - [Release Notes ↗](#)

- 1.** Install R. RStudio requires R 3.3.0+ [↗](#).
- 2.** Download RStudio Desktop. Recommended for your system:



DOWNLOAD RSTUDIO FOR WINDOWS

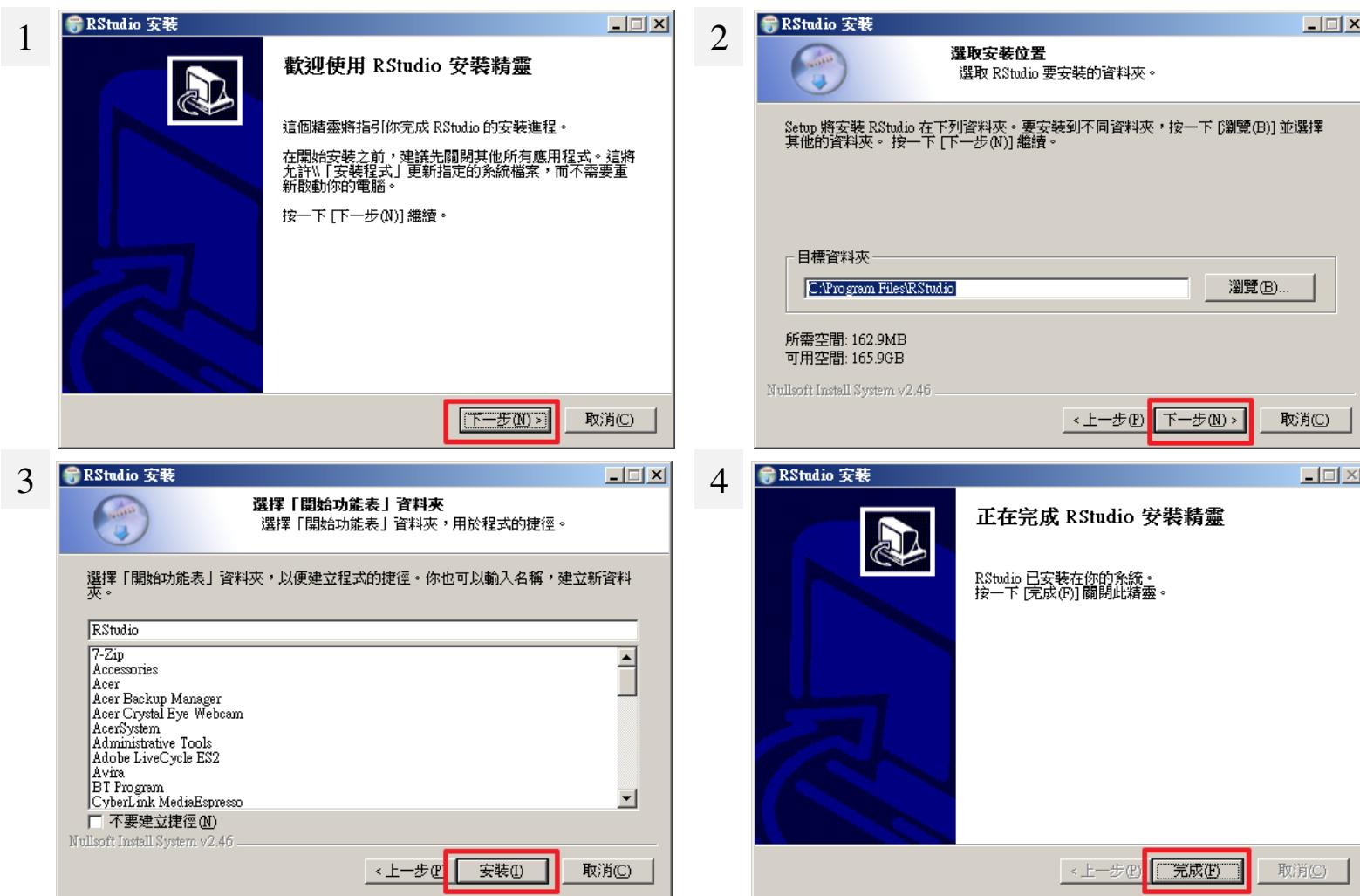
2022.07.1+554 | 190.14MB

Requires Windows 10/11 (64-bit)

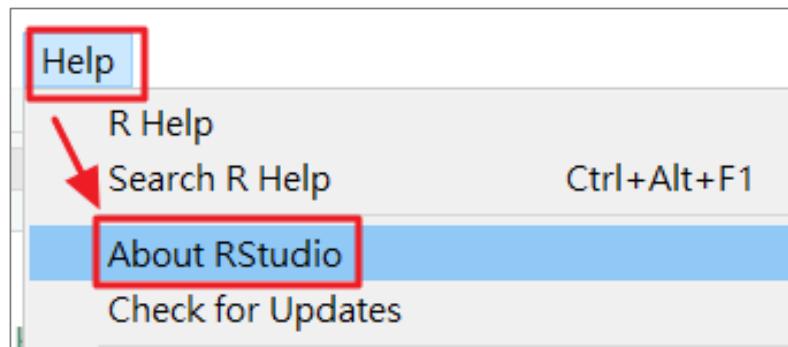


RStudio-2022.07.1-554.exe
(181 MB)

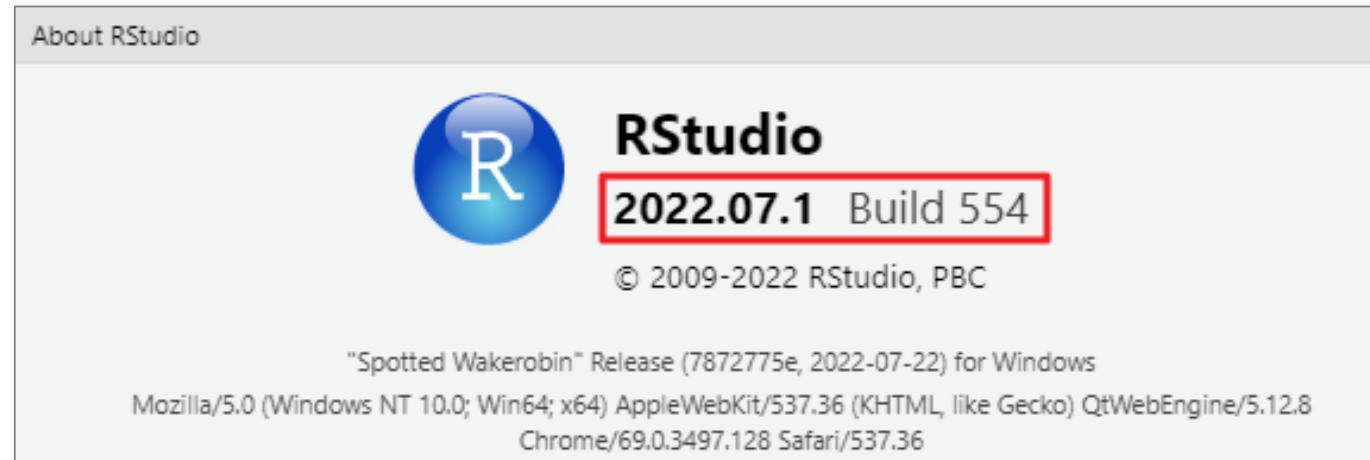
RStudio 安裝



RStudio 版本訊息

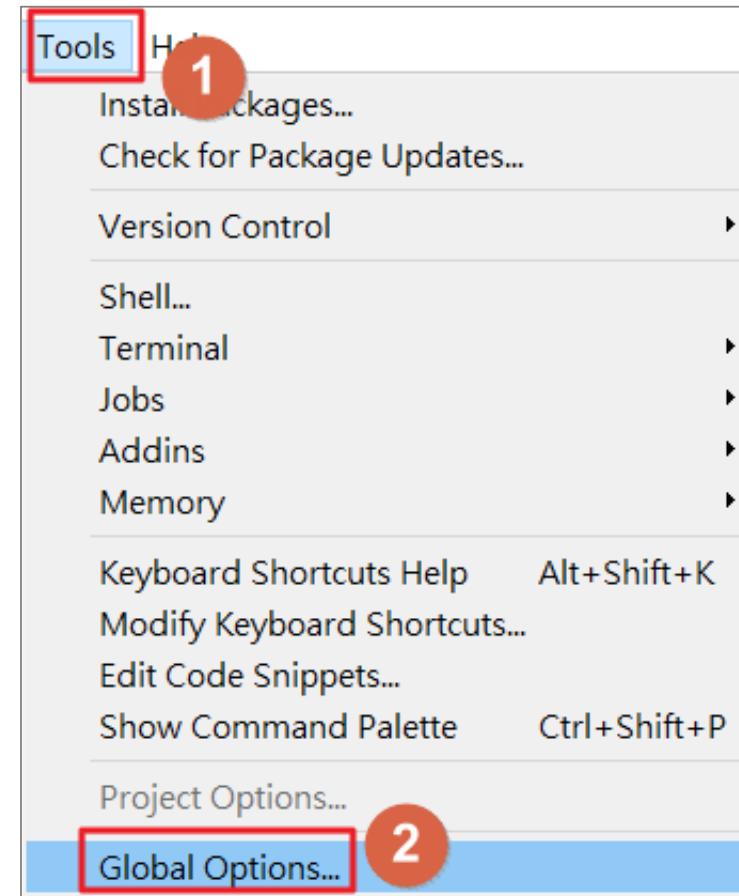


Help \ About RStudio

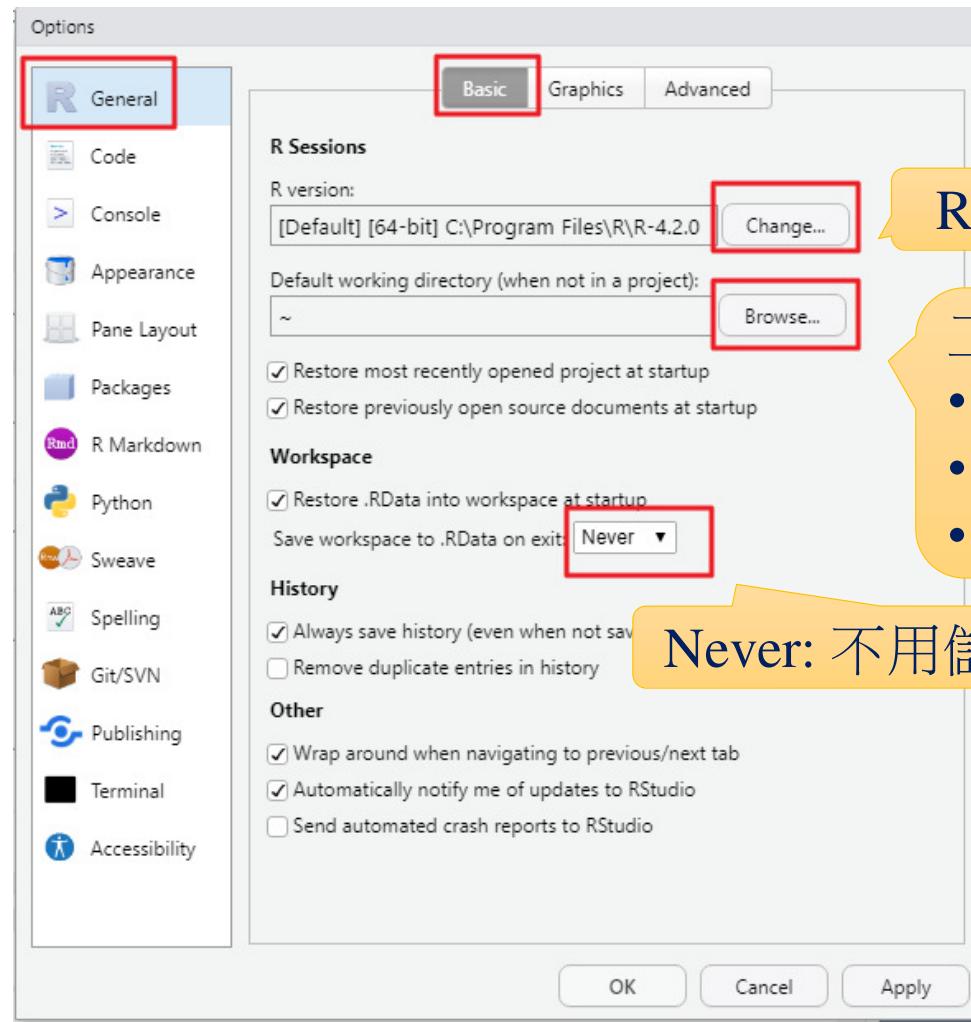


RStudio 選項設定

- Tools \ Global Options



General \ Basic



Windows 10.

```
> getwd()
[1] "C:/Users/asus/OneDrive/文件"
```

R 版本

工作目錄 ~

- Windows: 我的文件
- Mac: 使用者名稱
- Ubuntu: 使用者名稱

Never: 不用儲存RData

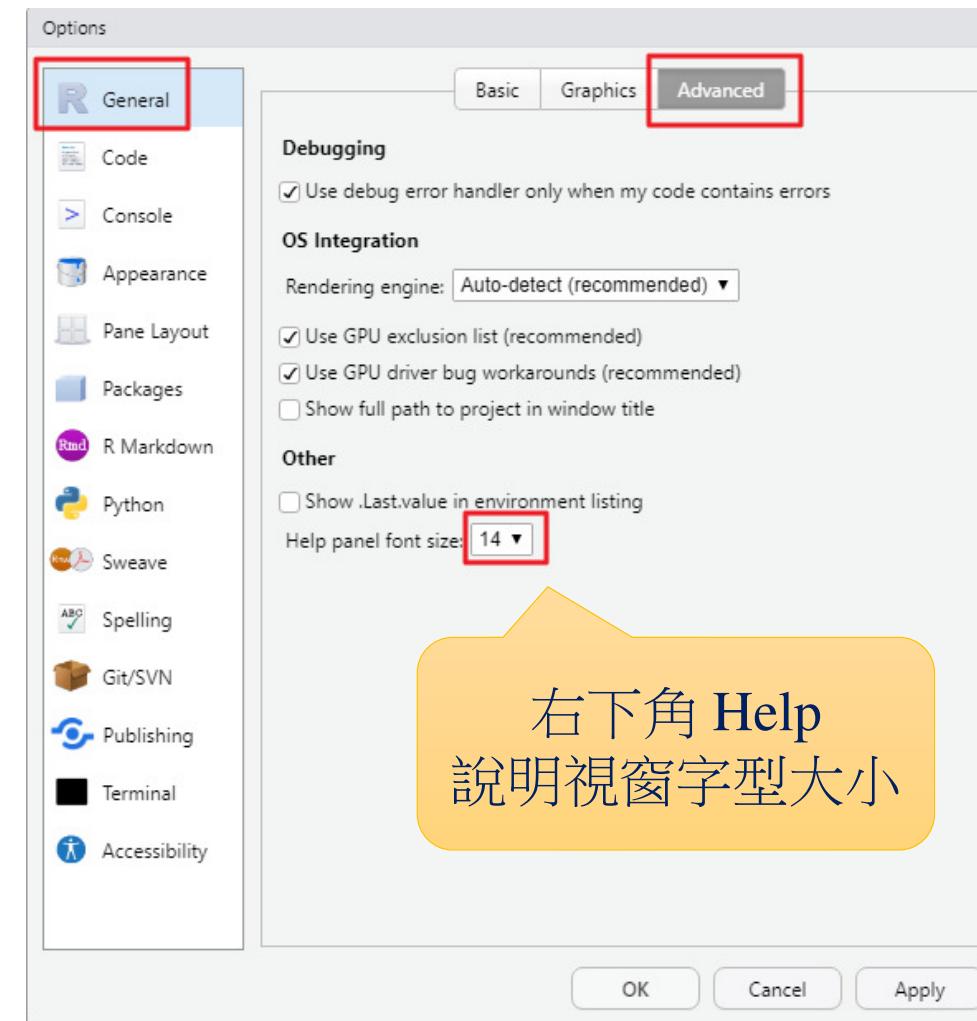
macOS Catalina

```
> getwd()
[1] "/Users/rwepa"
```

Ububtu 20.04

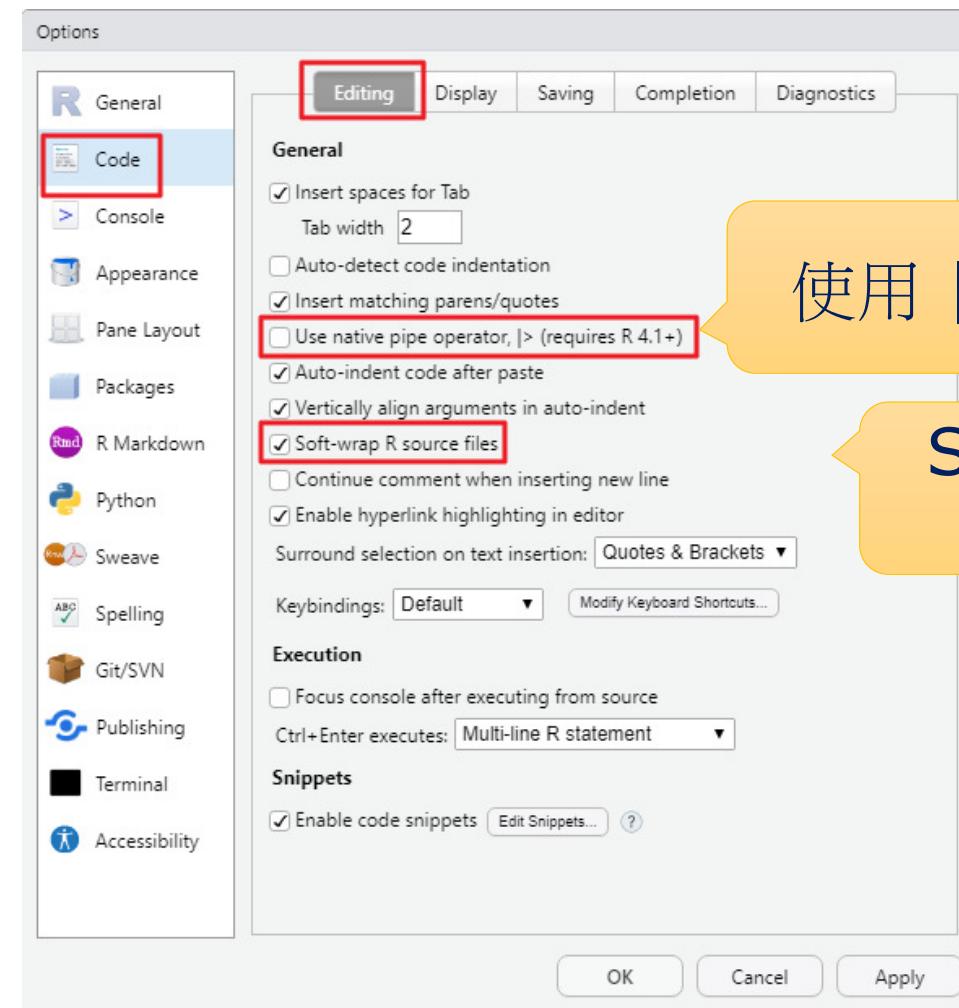
```
> getwd()
[1] "/home/rwepa"
```

General \ Advanced



右下角 Help
說明視窗字型大小

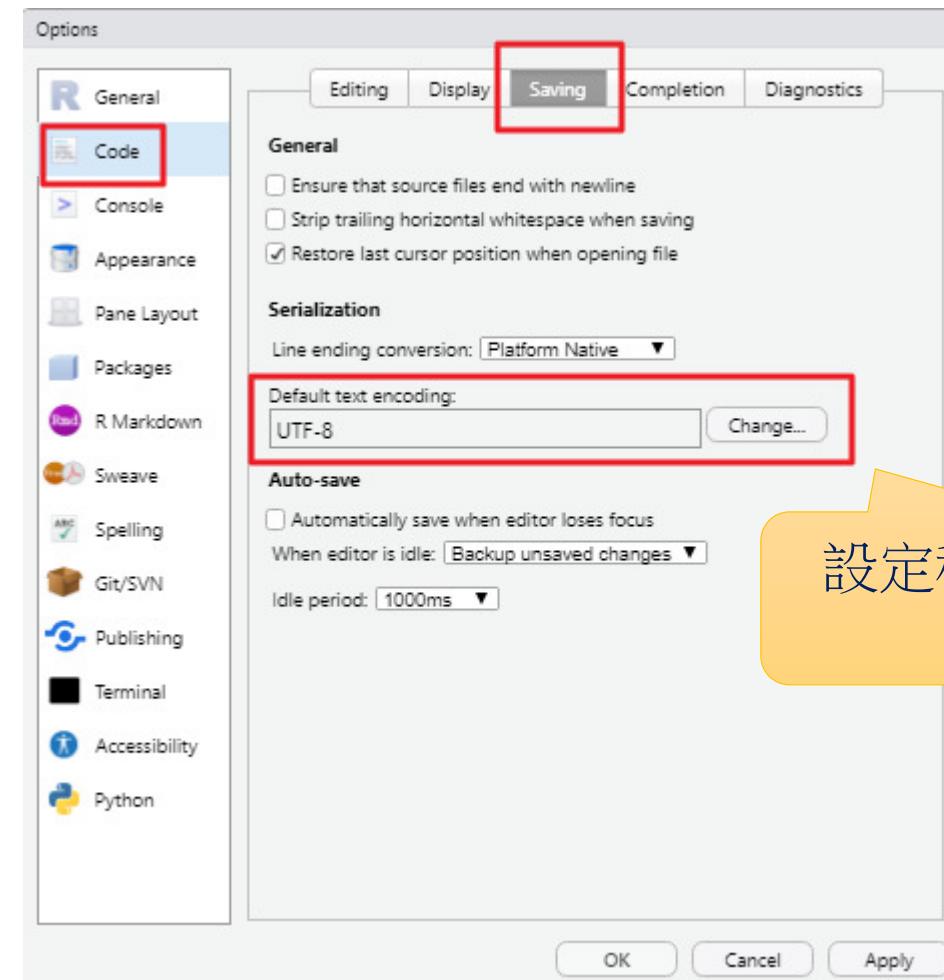
Code \ Editing



使用 |> 管線操作

Soft-wrap
自動換列

Code \ Saving

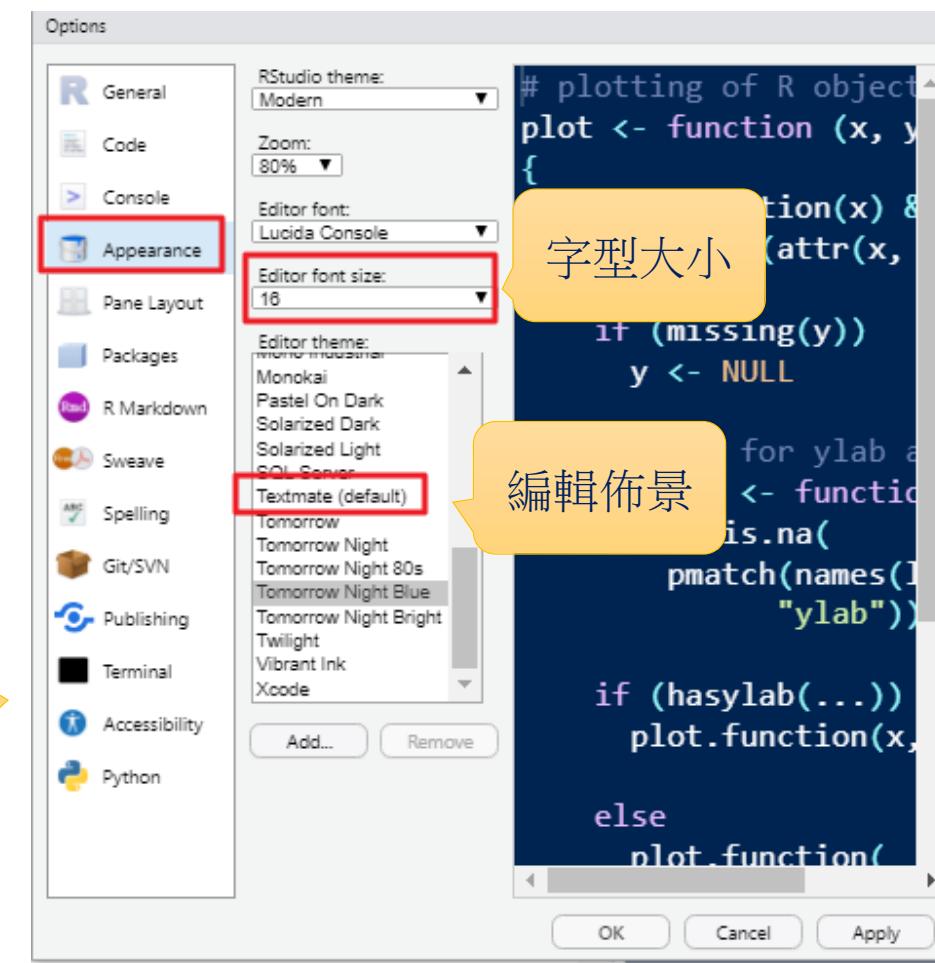


設定程式編碼 預設值
UTF-8

RStudio-選項設定(續)

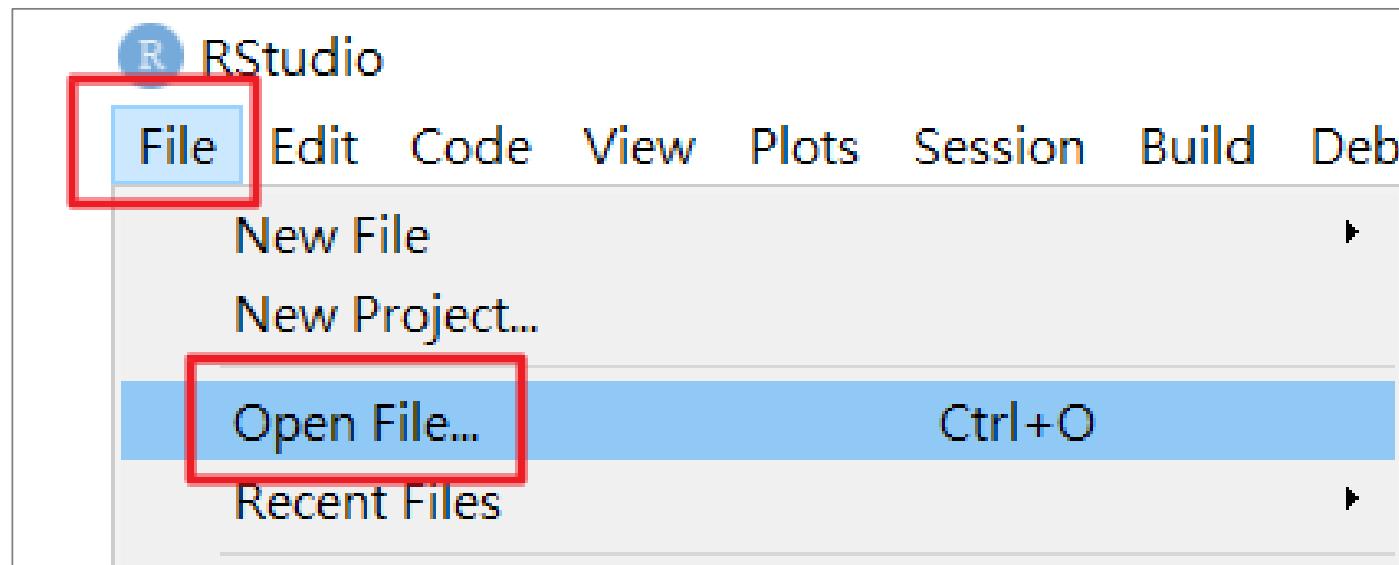
- Appearance \ Editor theme
 - 預設值:
TextMate

設定完成，須重新啟動RStudio



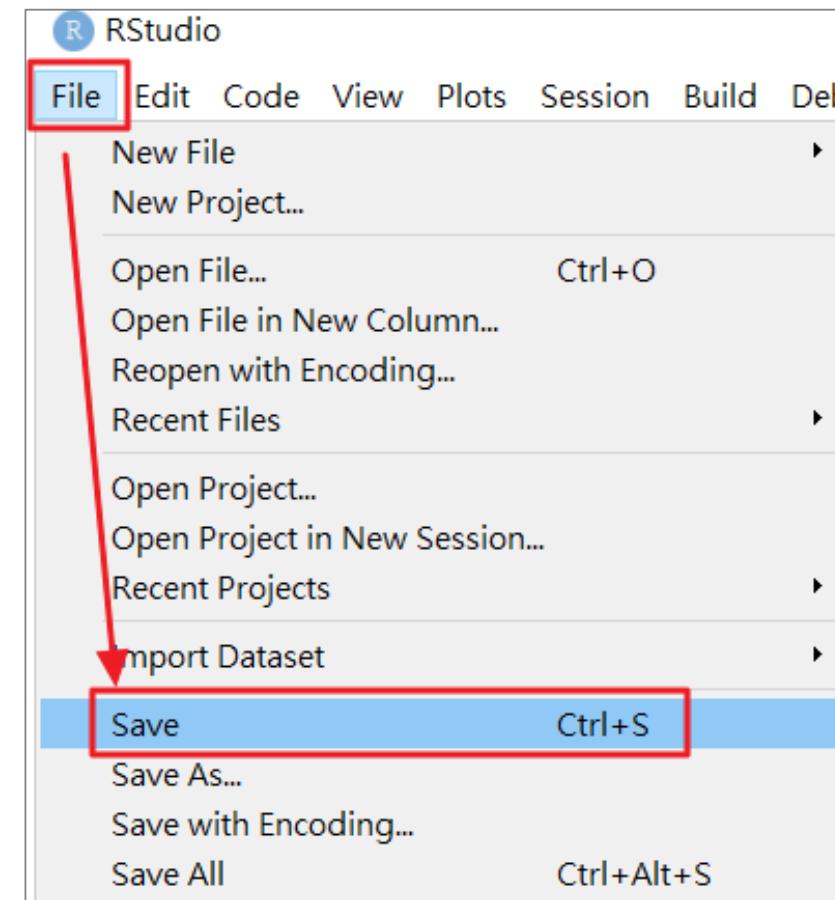
開啟檔案

- File \ Open File

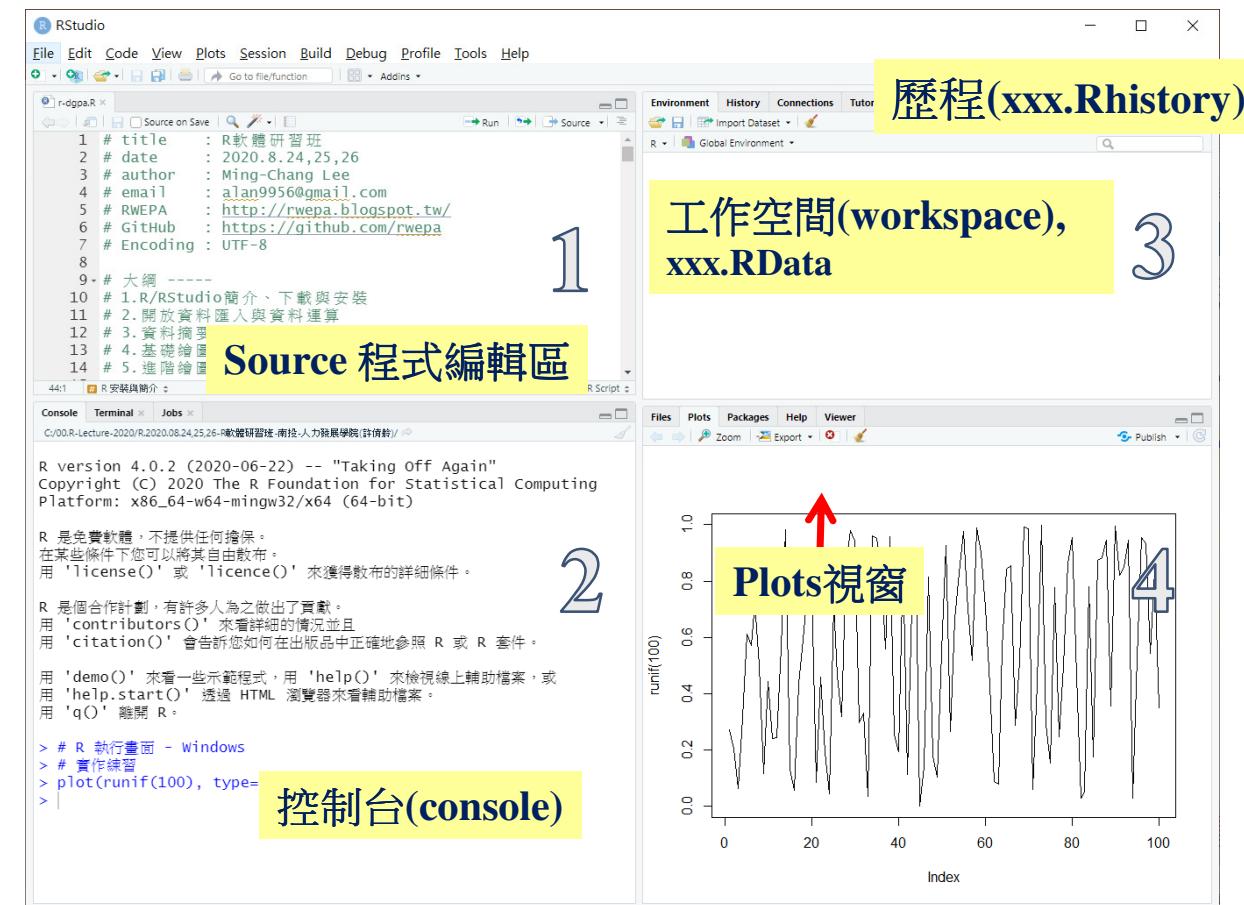


儲存檔案

- File \ Save
(CTRL + S)



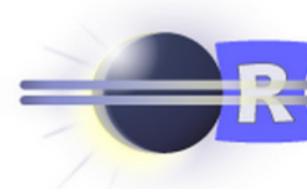
R/RStudio環境的基礎觀念



Ctrl + Shift + F10: 重新啟動R

R + Editor

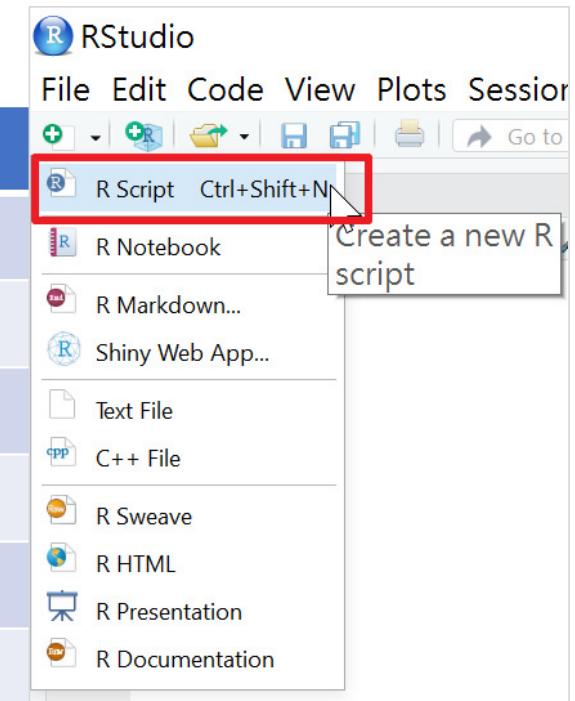
- R – 原生環境
- RStudio – IDE 整合介面
- Eclipse
 - StatET 4.x.0:
An Eclipse based IDE (integrated development environment) plug-in for R.
 - <https://projects.eclipse.org/projects/science.statet>
- 如何安裝 Visual Studio R 工具
 -



R Tools for Visual Studio

RStudio 快速鍵

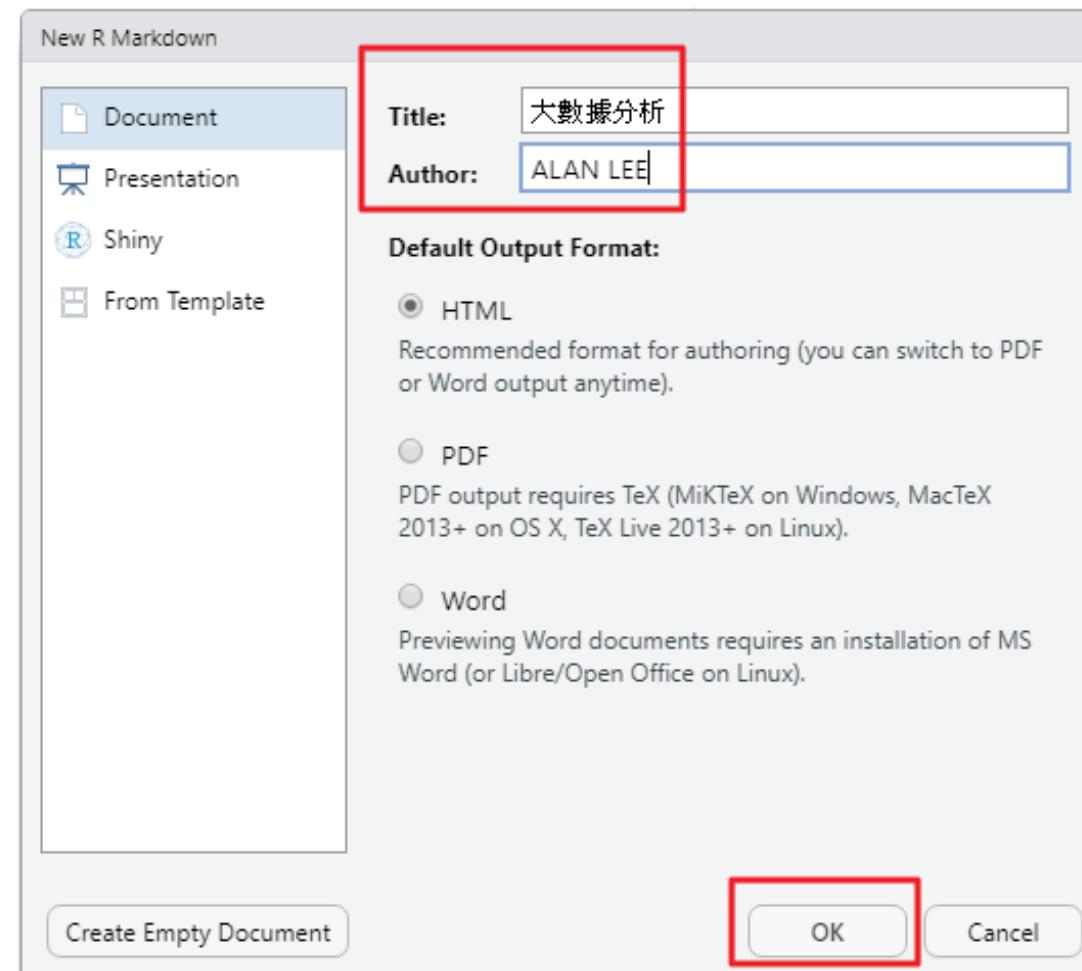
快速鍵	功能
Ctrl + Shift + N	建立新的R程式
Ctrl + S	儲存檔案
Ctrl + Shift + R	建立章節 (-----)
Alt + -	指派符號
Ctrl + Shift + C	註解
Ctrl + Enter	執行程式
Ctrl + Shift + F10	重新啟動R
Alt + Shift + K	快速鍵總表 (Esc 退出)



R Markdown

(R 標記語言)

RStudio - Markdown



RStudio - Markdown (續)

The screenshot shows the RStudio interface with a Markdown file open in the left pane. The code includes R Markdown syntax like `title: "大數據分析"`. The right pane shows an empty environment. A yellow callout bubble on the left contains three numbered steps:

1. Help
- 2.
3. RStudio Cheat Sheets

Red numbers 1, 2, and 3 point to specific elements in the RStudio interface:

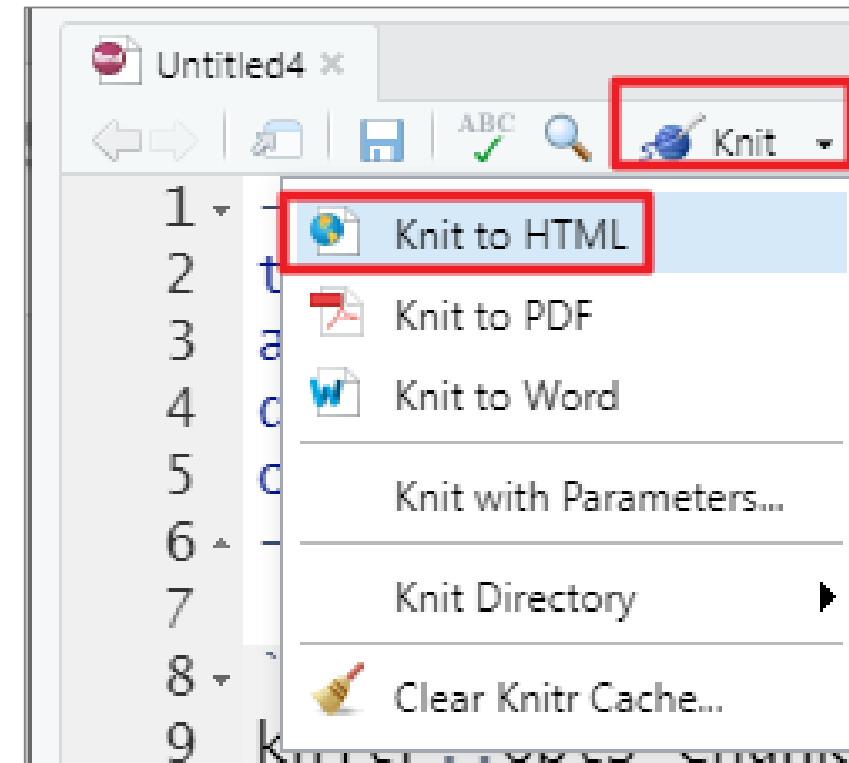
- 1 points to the Help tab in the top navigation bar.
- 2 points to the Help icon in the top navigation bar.
- 3 points to the RStudio Cheat Sheets link in the Help sidebar.

R Markdown Cheatsheet 線上說明

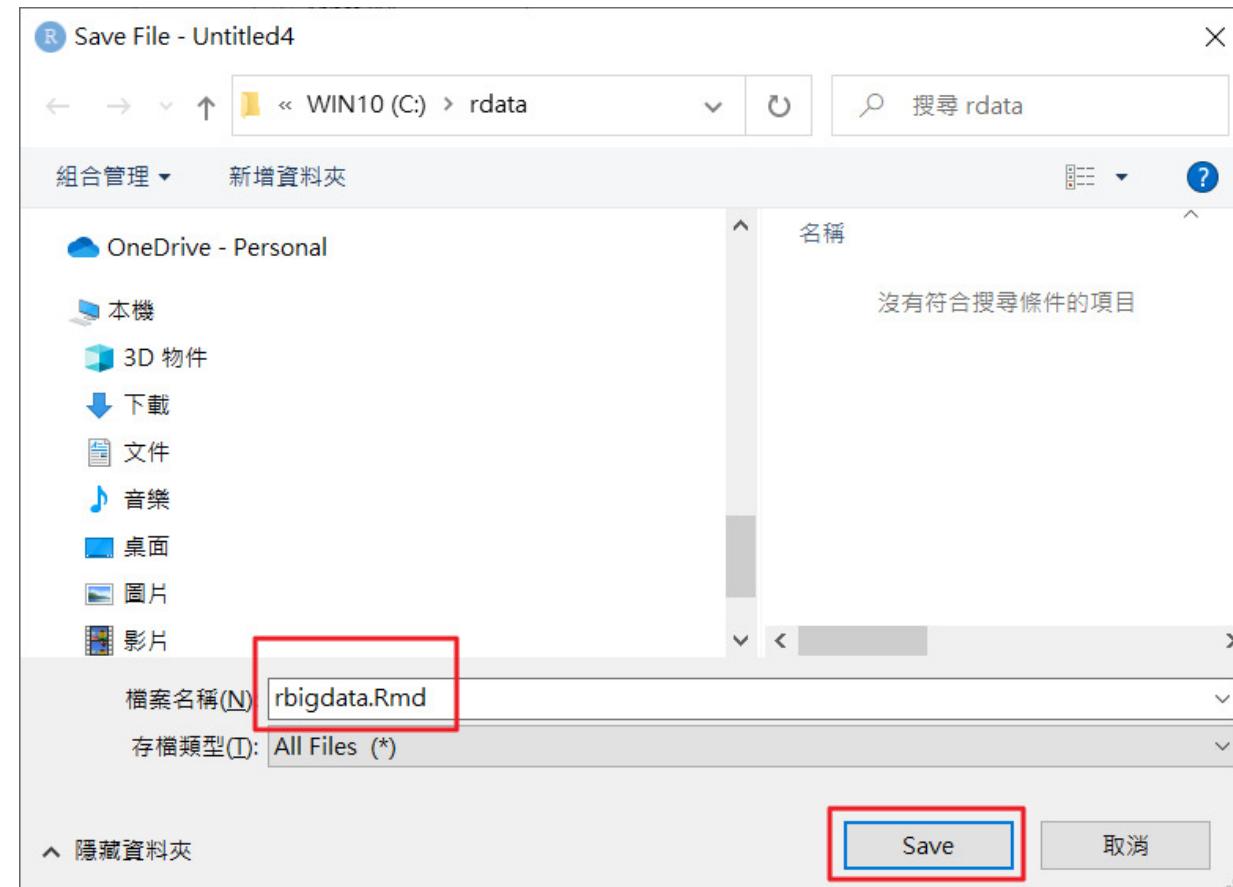
- <https://www.rstudio.com/resources/cheatsheets/>

RStudio - Markdown (續)

- Knit HTML
- Knit PDF
- Knit Word



RStudio - Markdown (續)



RStudio - Markdown : HTML

The screenshot shows an RStudio window displaying an R Markdown document. The title is "大數據分析" by ALAN LEE, dated 2020/7/6. The main content is titled "R Markdown" and includes a code chunk with the command `summary(cars)`. The resulting output is a table of car summary statistics:

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

Below this, there is a section titled "Including Plots" with a note about embedding plots. A scatter plot of "pressure" vs "speed" is shown, with data points forming a positive correlation. Three yellow callout boxes highlight different features: one points to the "summary(cars)" code, another to the table output, and a third to the scatter plot.

輸出HTML

R程式碼

報表

圓形

```
## #> 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
```

pressure

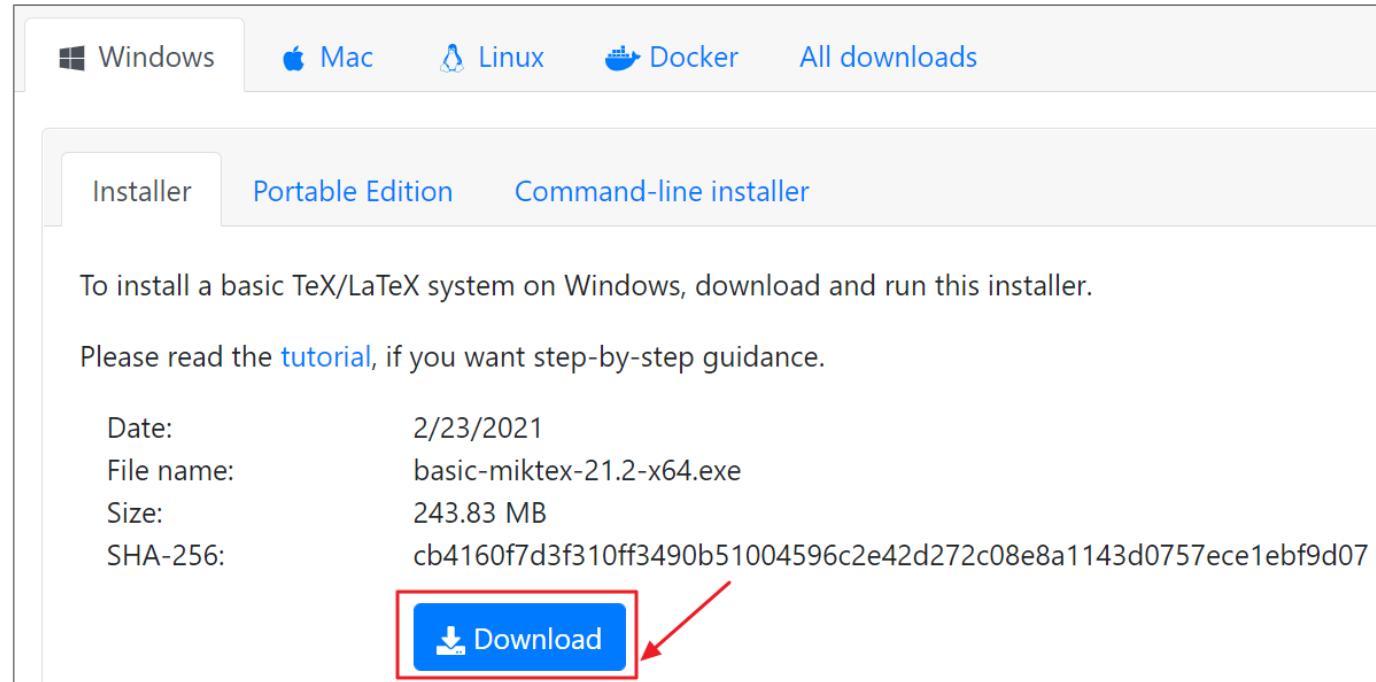
speed

0 50 100 150 200 250 300 350

0 200 400 600 800

RStudio - Markdown : PDF

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



Knit to PDF



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

big data
alan lee
2021/3/29

R Markdown

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

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

```
summary(cars)
```

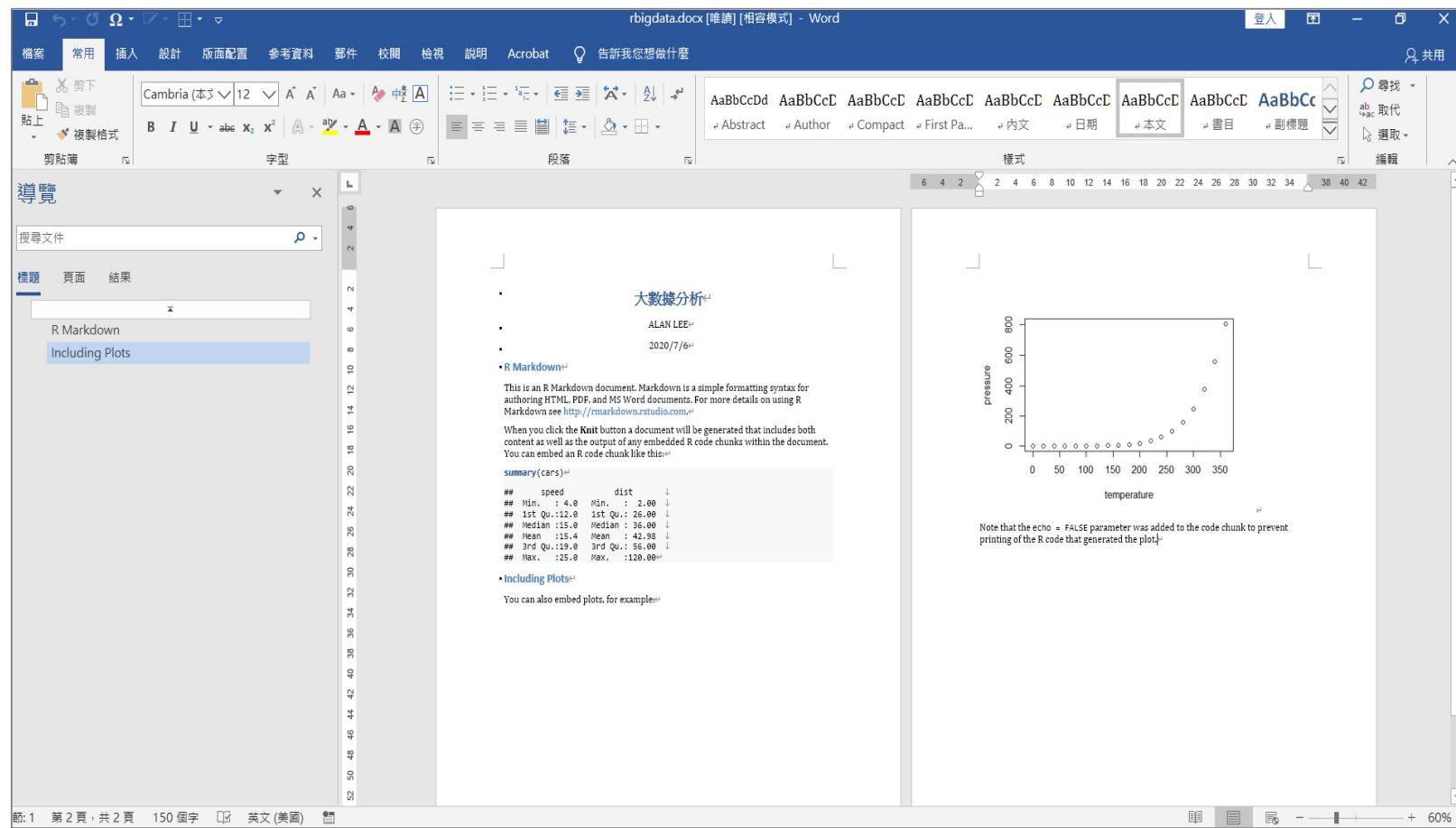
speed dist
Min. : 4.0 Min. : 2.00
1st Qu.:12.0 1st Qu.: 26.00
Median :15.0 Median : 36.00
Mean :15.4 Mean : 42.98
3rd Qu.:19.0 3rd Qu.:56.00
Max. :28.0 Max. :120.00

Including Plots

You can also embed plots, for example:

中文可能有問題?

RStudio - Markdown 轉換為 Word



參考資料

- RWEPA
 - <http://rwepa.blogspot.com/>
- R 入門資料分析與視覺化應用
 - <https://mastertalks.tw/products/r?ref=MCLEE>
- R 商業預測與應用
 - <https://mastertalks.tw/products/r-2?ref=MCLEE>

謝謝您的聆聽

Q & A



李明昌

EMAIL: alan9956@gmail.com

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