

Python醫療大數據應用

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

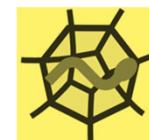
<http://rwepa.blogspot.com/>

大綱

1. 醫療大數據應用



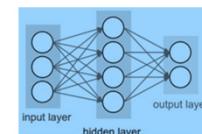
2. Python/AAnaconda環境介紹



3. 全民健保資料庫



4. 深度學習的發展



5. 皮馬印第安人糖尿病預測分析



6.Q & A



1. 醫療大數據應用

(Big Data Applications In Healthcare)



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

- 姓名：李明昌 (ALAN LEE)
- 現職：中華R軟體學會 常務理事
臺灣資料科學與商業應用協會 常務理事
- 學歷：中原大學 工業與系統工程所 博士
- 研究：
 - 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)
- 大專院校、資策會、工業技術研究院、國家發展委員會、中央氣象局、公平交易委員會、各縣市政府與日本名古屋產業大學等公民營單位演講達320餘場，2900小時以上。
- 連絡資訊：alan9956@gmail.com



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

何謂大數據

- 大數據(巨量資料, Big Data) 指的是所涉及的資料量規模巨大到無法透過人工，在**合理時間內**達到擷取、管理、處理、並整理成為**人類所能解讀的形式**的資訊

```
1546107 World,WLD,"Presence·of·peace·keepers  
1546108 Samoa,WSM,"Presence·of·peace·keepers  
1546109 "Yemen,·Rep.",YEM,"Presence·of·peace  
1546110 South·Africa,ZAF,"Presence·of·peace·  
1546111 "Congo,·Dem.·Rep.",ZAR,"Presence·of·  
1546112 Zambia,ZMB,"Presence·of·peace·keeper
```

維基: https://en.wikipedia.org/wiki/Big_data

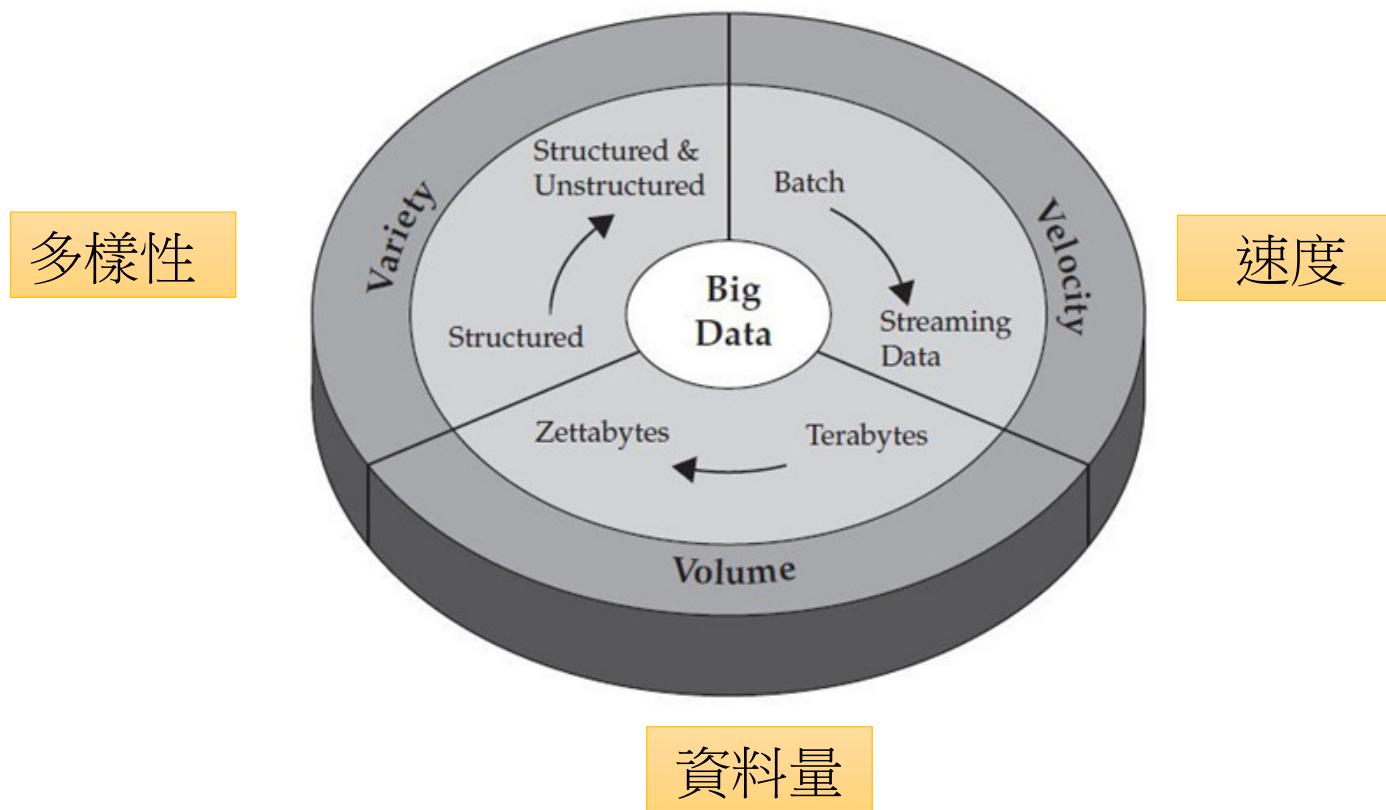
154萬筆資料

巨量資料的3V定義

- 2001年Gartner 公司的分析師 Douglas Laney 提出3V
(三個關鍵挑戰)
 - 資料量 (**Volume**) : 資料總量很大
 - 速度 (**Velocity**) : 資料產生的速度快
 - 多樣性 (**Variety**) : 資料種類繁多

參考: Lancy, D., 3D Data Management: Controlling Data Volume, Velocity, and Variety, Gartner, 2001.

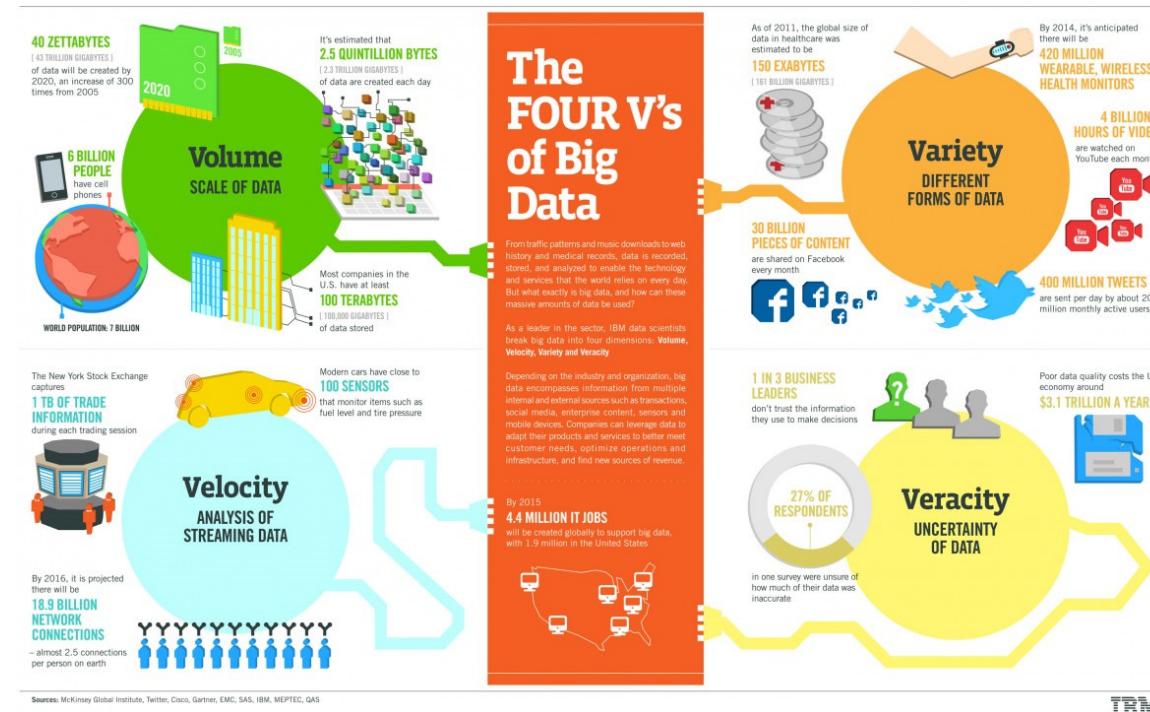
IBM-3V



參考: IBM, Understanding Big Data, 2012.

IBM的4V定義

- IBM 的 Big Data 4V: 新增 Veracity (真實性)



參考: <http://www.ibmbigdatahub.com/infographic/four-vs-big-data>

IBM - Veracity

**1 IN 3 BUSINESS
LEADERS**

don't trust the information
they use to make decisions

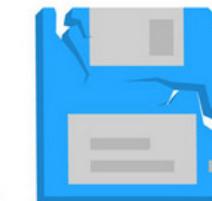


**27% OF
RESPONDENTS**

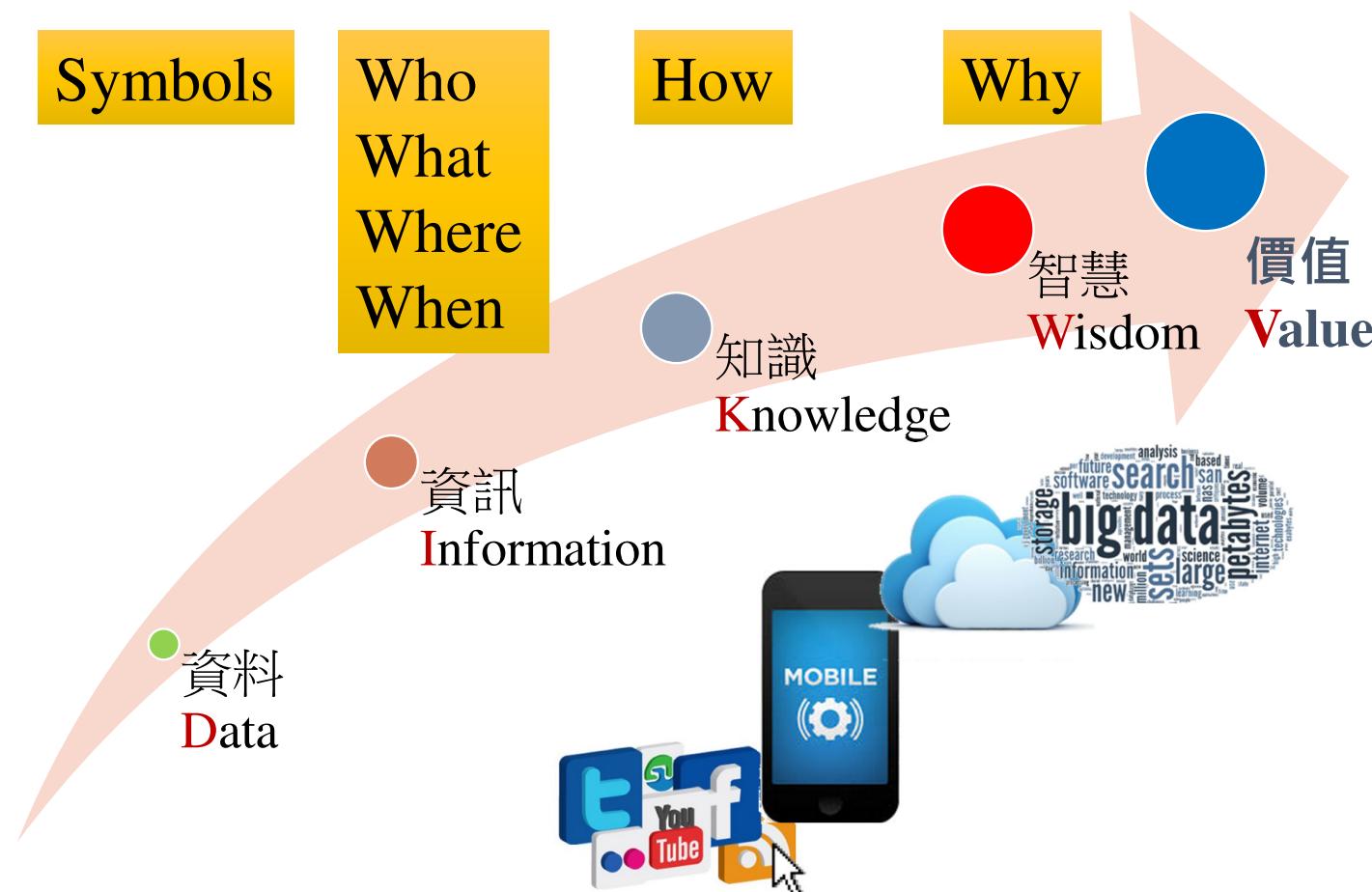
in one survey were unsure of
how much of their data was
inaccurate

Veracity
**UNCERTAINTY
OF DATA**

Poor data quality costs the US
economy around
\$3.1 TRILLION A YEAR



科技驅動與資料加值

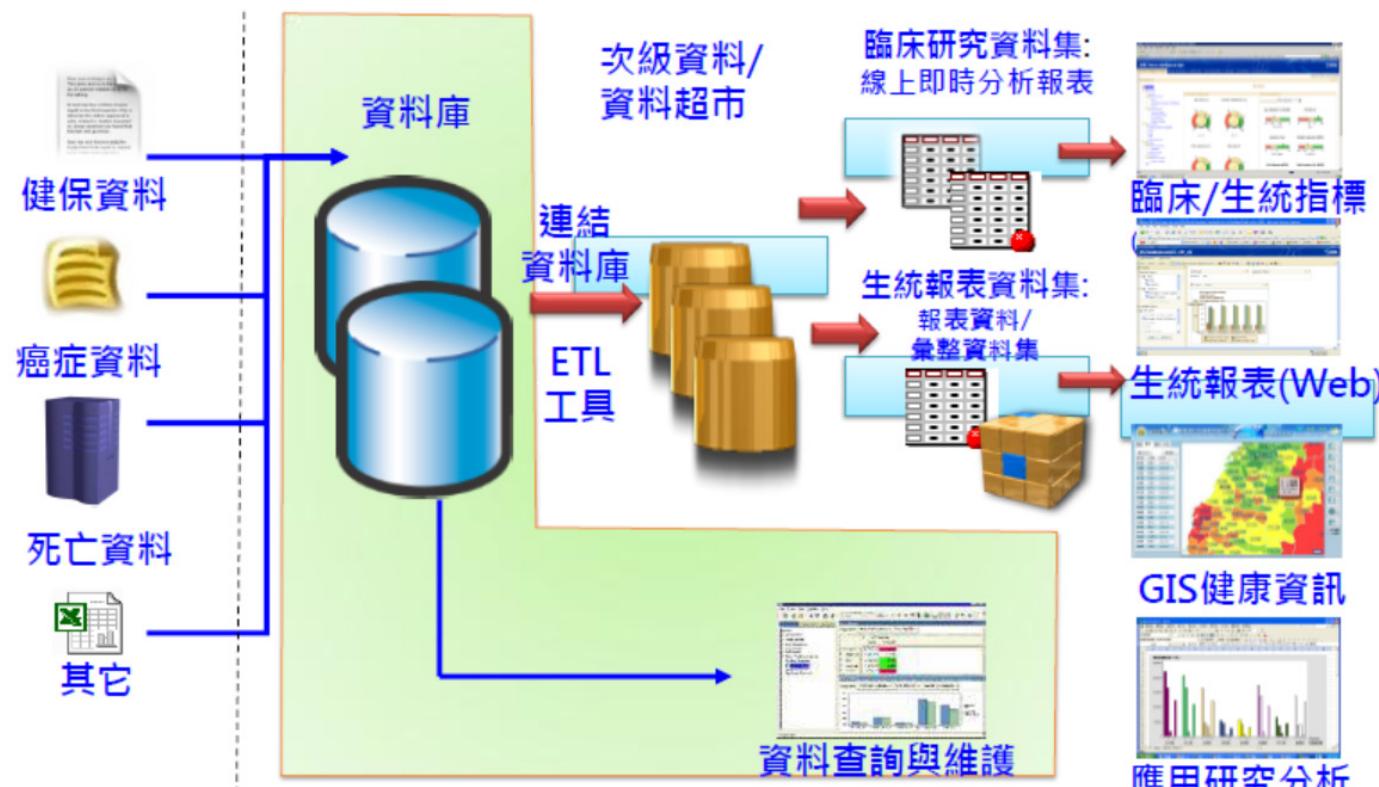


加值服務趨勢

- 加值服務(Value-Added Service, VAS)
 - 將某項非核心技術、產品或服務利用新方式加以修正改善，以創造更高的**價值**。
 - VAS廣泛應用於各產業及研究領域，以通訊及網路產業為首的例子分別為Web2.0應用及行動上網等(維基百科)。
- 行政院科技會報辦公室於101年1月18日
 - 政府資料加值(open data)推動策略會議
 - GIS、健康醫療、行動服務

健康加值應用平台

INPUT → PROCESS → OUTPUT



政府公開資料加值推動策略會議:

<http://www.bost.ey.gov.tw/cp.aspx?n=E48F3D33FA666886>

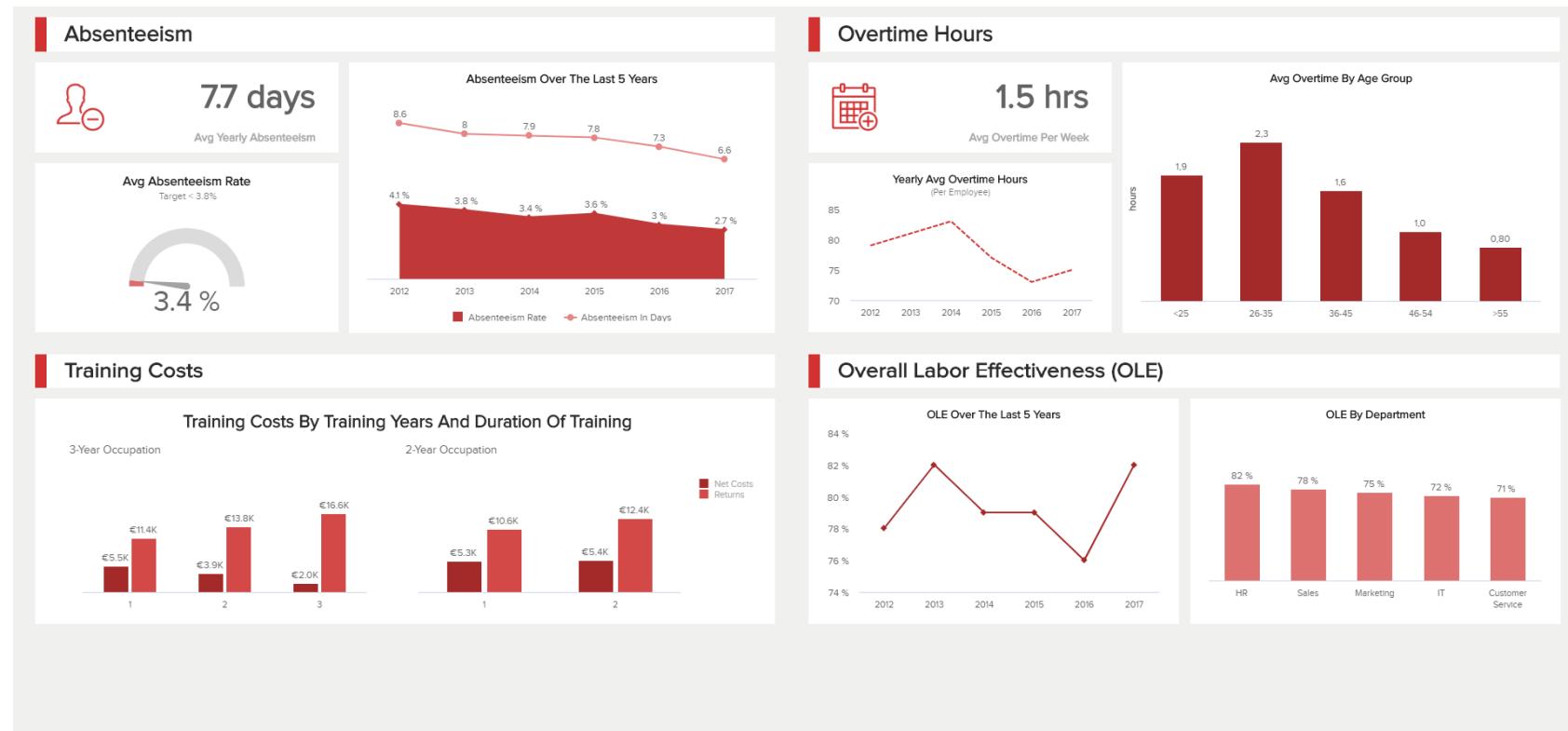
醫療大數據分析應用

- 醫療數據電子化並降低成本
- 全民健保資料
 - 預防流行病
 - 疾病治癒
- 醫療資料儀表板
- 醫療 + 人工智慧
 - 醫學影像分析
 - 肺部腫瘤預測
- 論文研究



醫療人力資源儀表板

- <https://www.datapine.com/dashboard-examples-and-templates/human-resources>



MRI線上分析系統

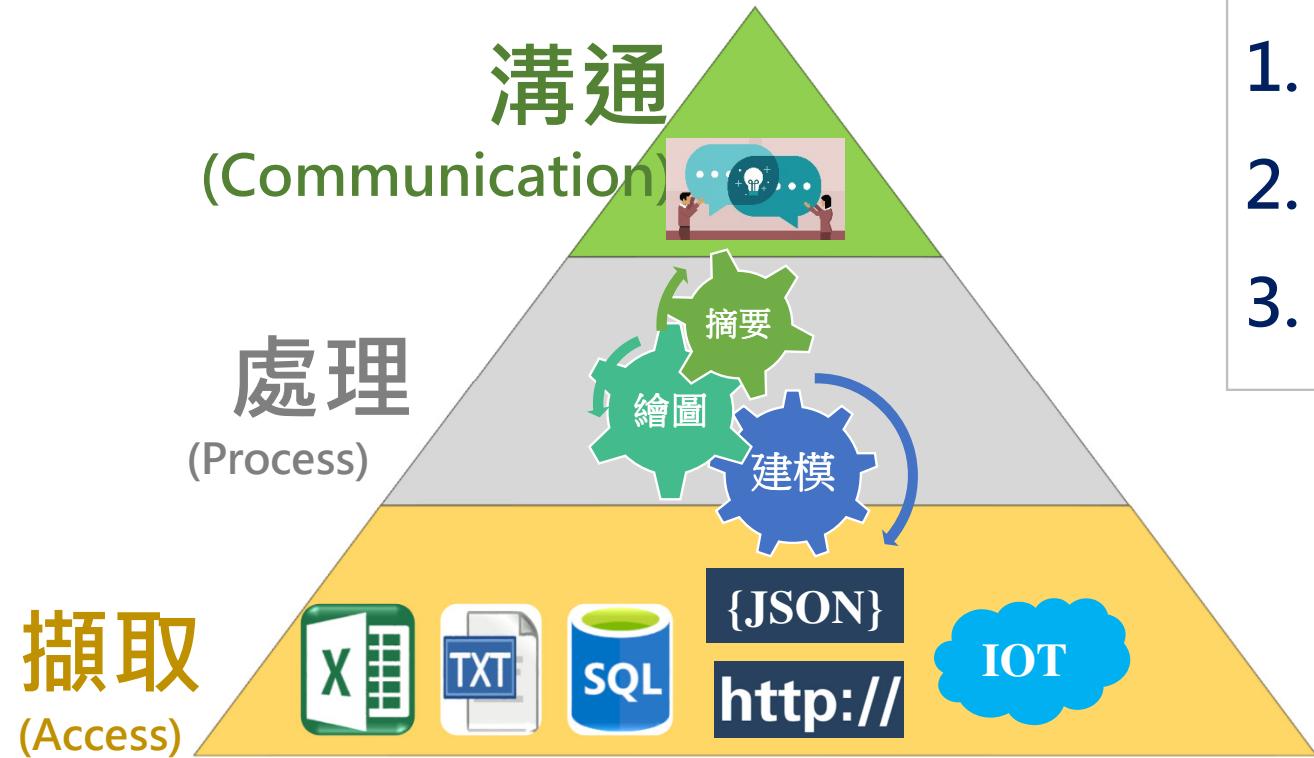
- <https://haozhu233.shinyapps.io/shinyMRI-contest/>





資料分析的心法

★★★資料分析架構→APC方法

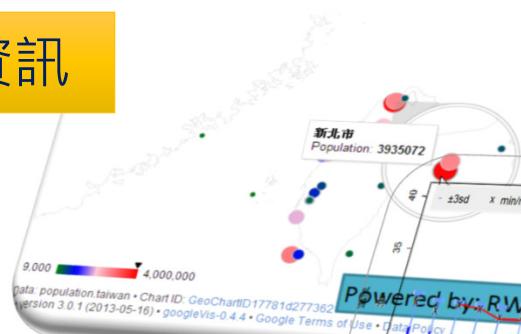


- 1.
- 2.
- 3.

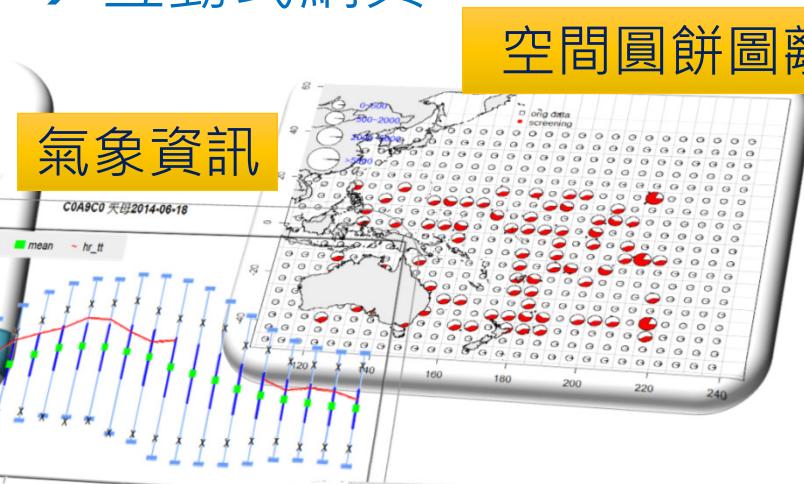
資料分析與視覺化應用

R + shiny → 互動式網頁

地理資訊



氣象資訊

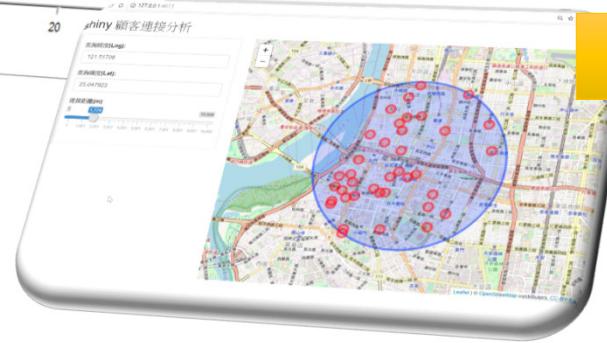


空間圓餅圖離群值分析

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

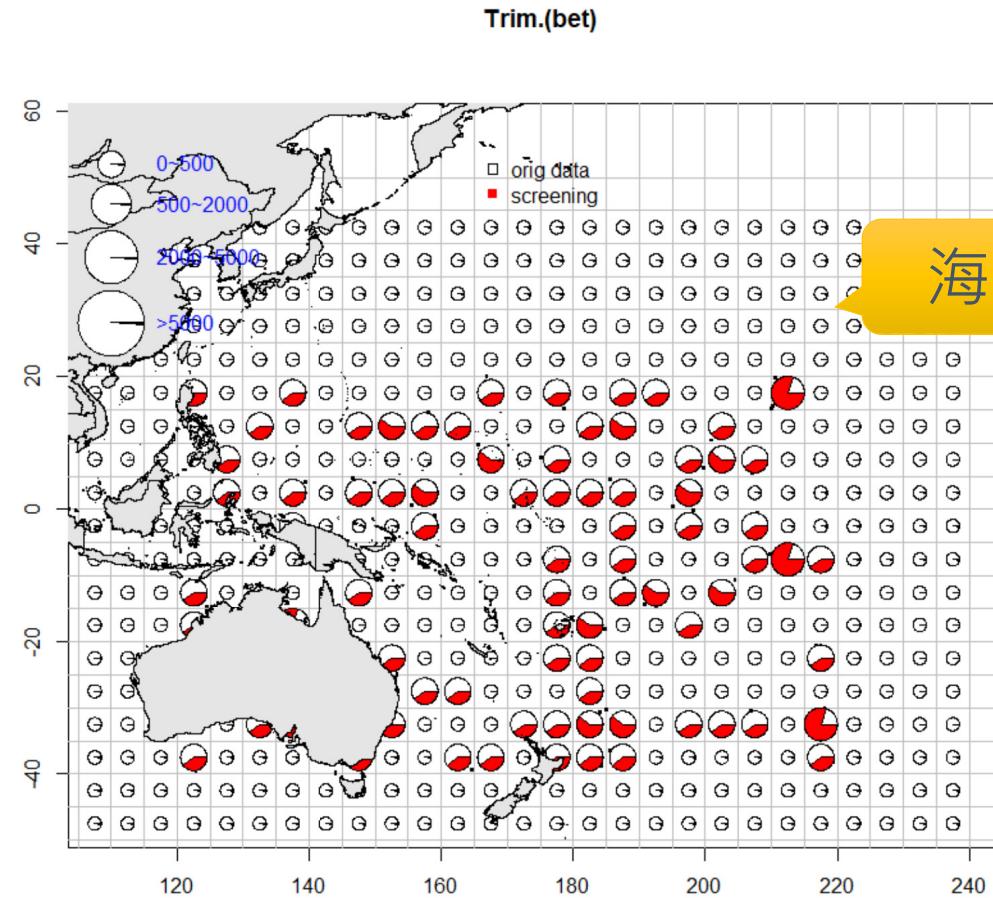


保險預測模型

The screenshot shows the iinsurance interactive analysis platform version v.16.3.24. The top navigation bar includes options like '檔案上傳', '資料處理', '統計圖表', '模型評估', and '預測模型'. A yellow callout bubble labeled '機率模型閥值調整' points to a slider control for setting the '機率模型閥值' (Probability Model Threshold) from 0.01 to 1, currently set at 0.1. Another yellow callout bubble labeled '預測結果' points to a table titled '檢視結果' (View Results) which displays 10 entries of predicted outcomes. The table columns include: 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), 車齡組合 (Age Group Combination), 車齡 (Age), 預測機率 (Prediction Probability), and 理賠 (Claim). The table shows various entries with predicted probabilities ranging from 0.0694 to 0.1866 and claim outcomes '有' (Yes) or '無' (No).

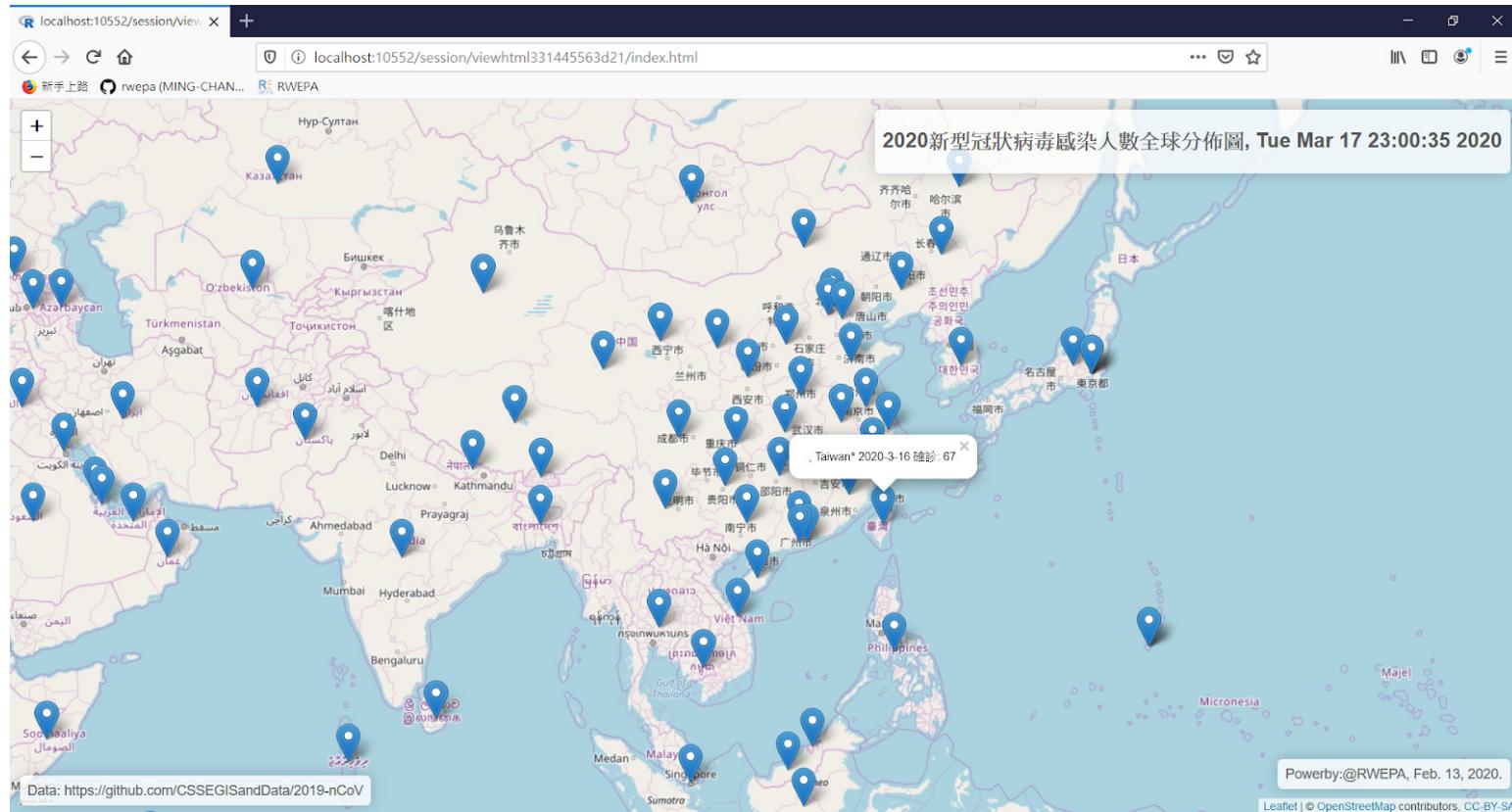
性別	女性	車輛種類	私家車	曝露風險	曝露風險對數	無索償折扣	被保險人年齡	私家車一車齡 0	私家車二車齡 1	私家車三車齡 2	車齡組合	車齡 0_1_2	預測機率	理賠		
M	0	A	1	0.9144422	-0.08944106	50	4	1	0	0	1	0	2	0.1069	有	
M	0	A	1	0.8158795	-0.20348856	20	4	0	0	1	1	2	2	0.1441	有	
3	M	0	A	1	0.8377823	-0.17699695	50	3	0	0	1	1	2	2	0.1866	有
4	M	0	A	1	0.4325804	-0.83798702	50	6	0	1	0	1	1	2	0.0944	無
5	M	0	A	1	0.7173169	-0.33223755	50	4	0	0	1	1	2	2	0.1218	有
6	M	0	A	1	0.8377823	-0.17699695	50	4	0	0	1	1	2	2	0.1495	有
7	M	0	A	1	0.8487337	-0.16400975	50	5	0	0	1	1	2	2	0.1422	有
8	F	1	A	1	0.8268309	-0.19015503	10	3	0	0	1	1	2	2	0.1733	有
9	M	0	A	1	0.7145791	-0.33606164	0	5	1	0	0	1	0	2	0.0694	無
10	M	0	A	1	0.3340178	-1.09656101	0	3	0	0	1	1	2	2	0.0783	無

空間圓餅圖離群值分析



2020新型冠狀病毒視覺化

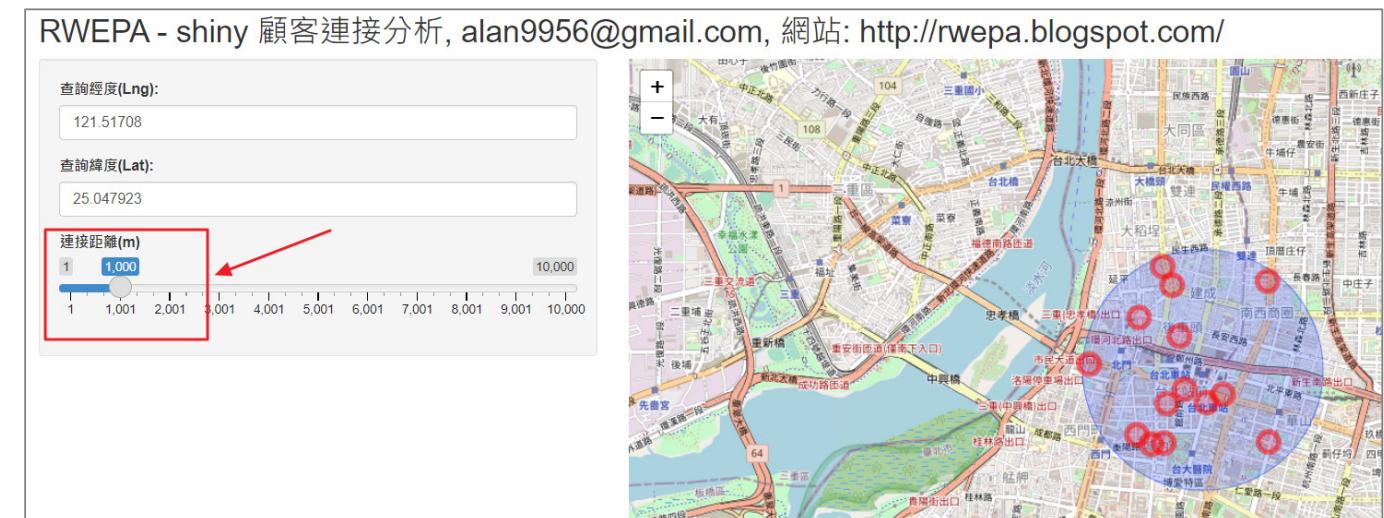
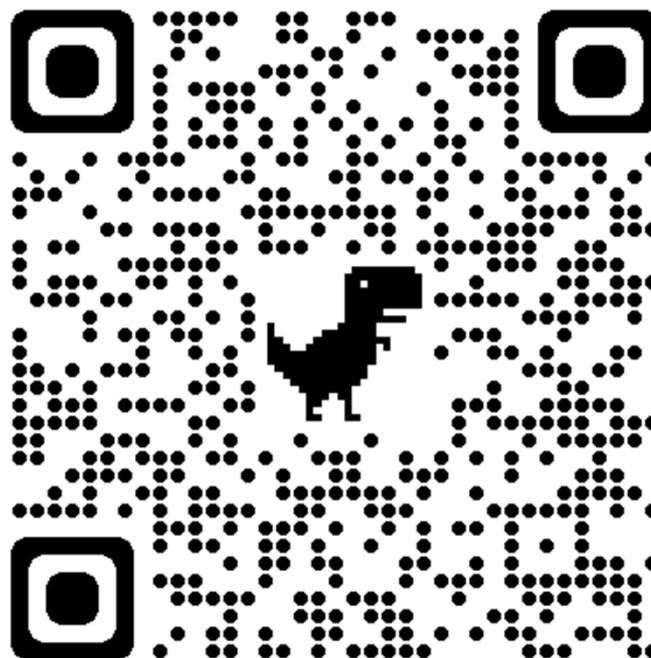
- <http://rwepa.blogspot.com/2020/02/2019nCoV.html>





shiny 顧客連接分析

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



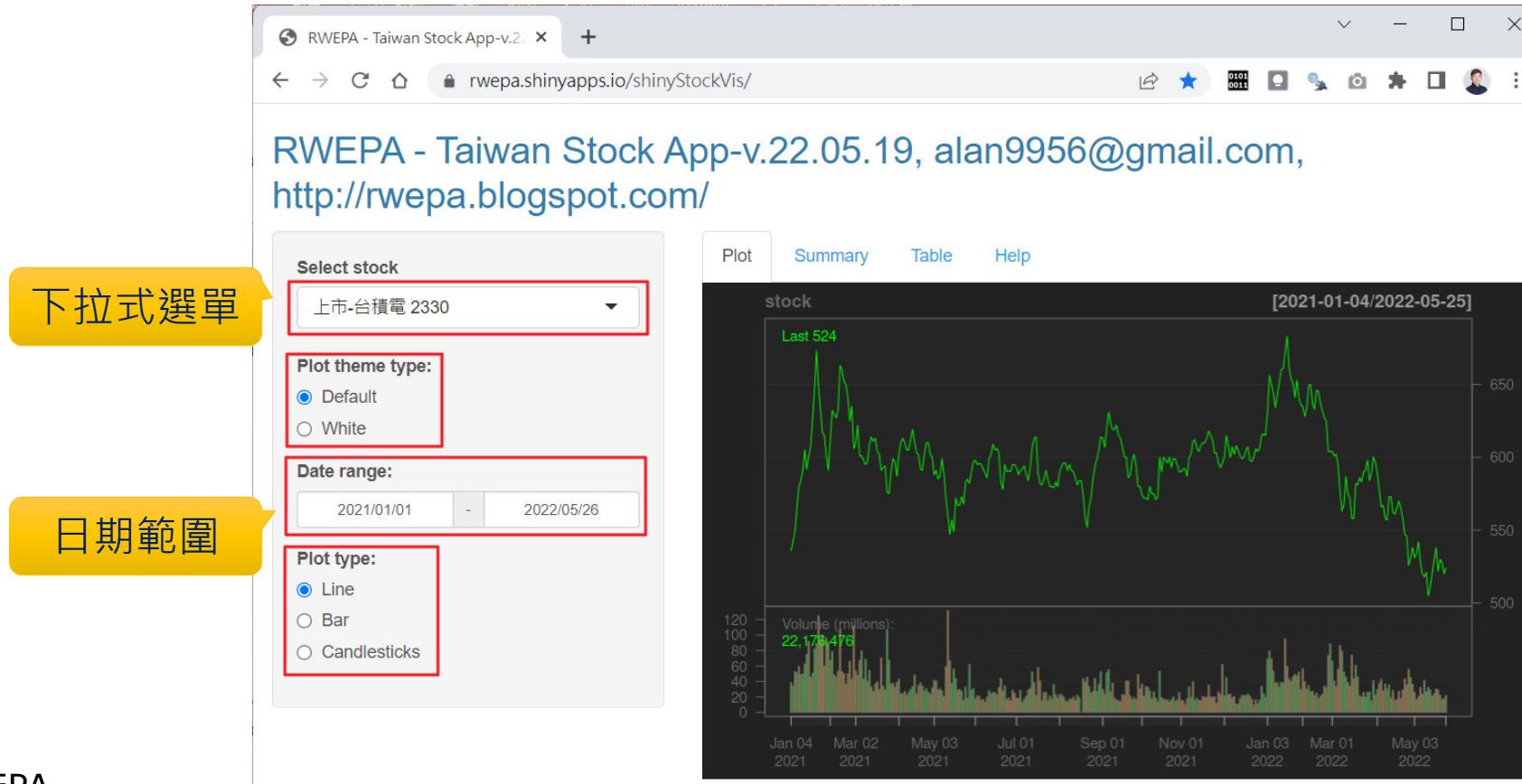
shiny 品質管制圖(quality control chart)應用

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

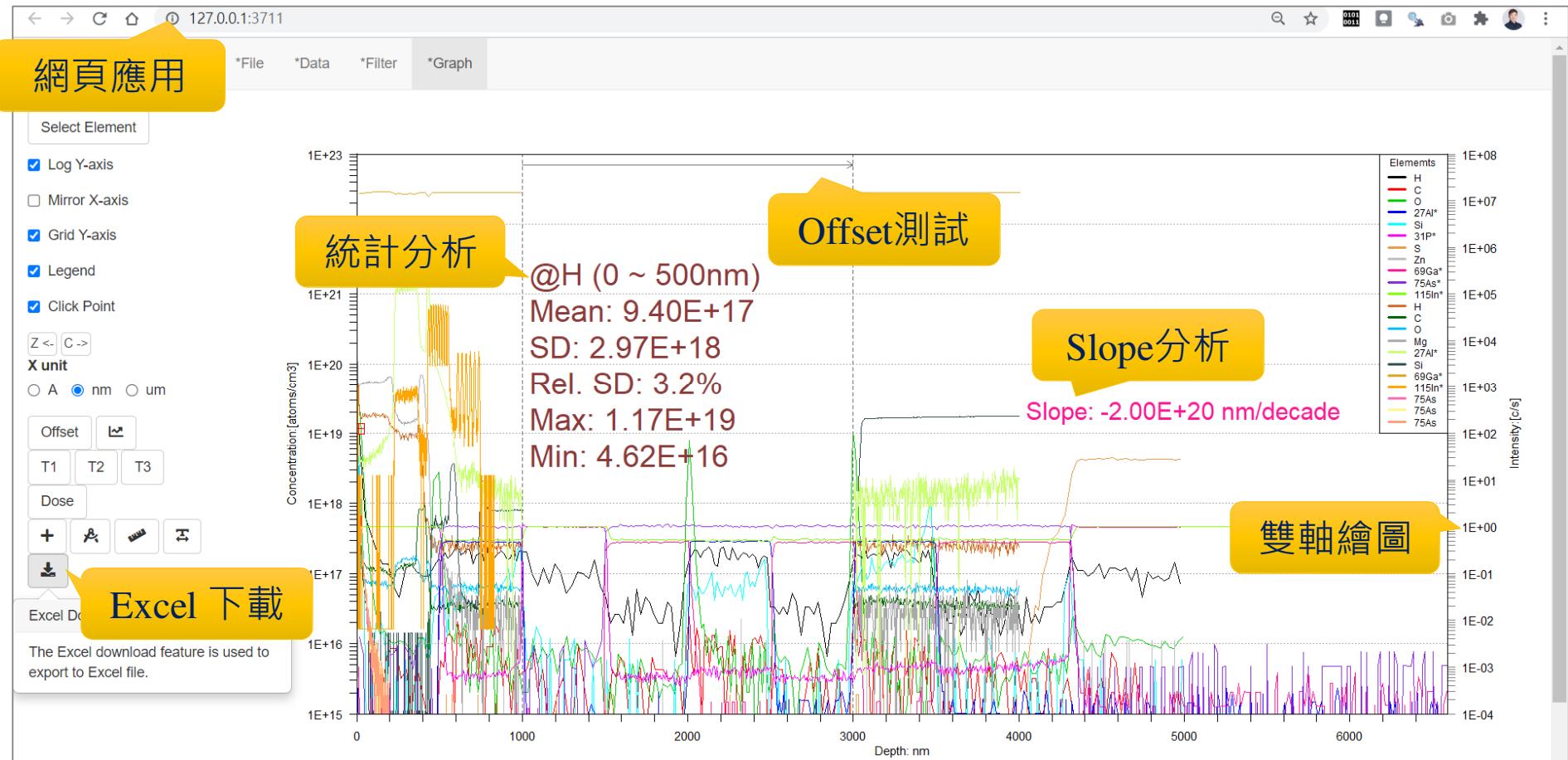


Taiwan Stock App

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



離子資料分析與視覺化應用



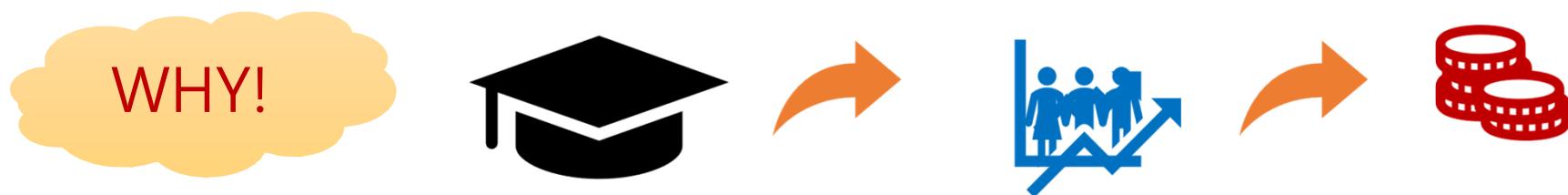
Python 使用 shiny 模組教學

- <https://youtu.be/s2fgEAa6lq0>
- <http://rwepa.blogspot.com/2022/10/shiny-for-python.html>



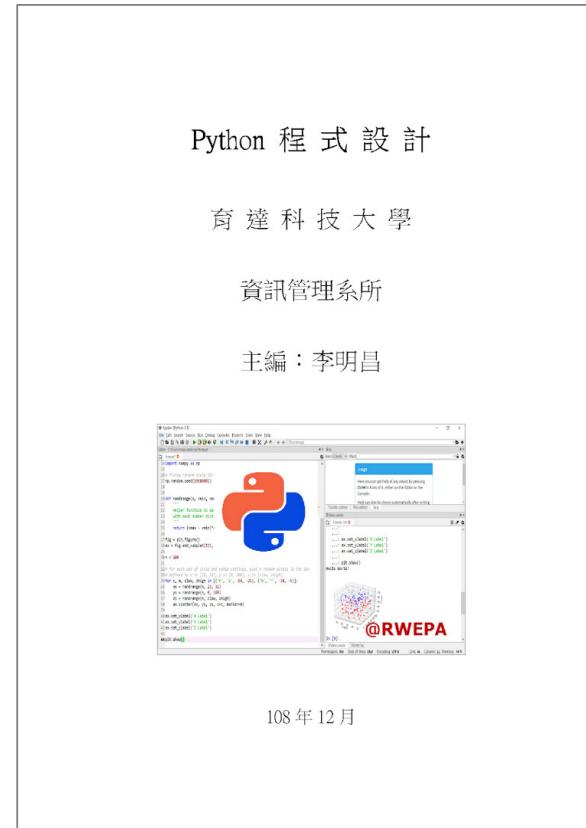
如何學習 Python?

- 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

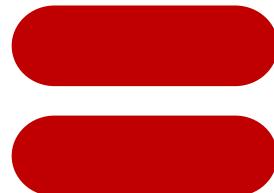
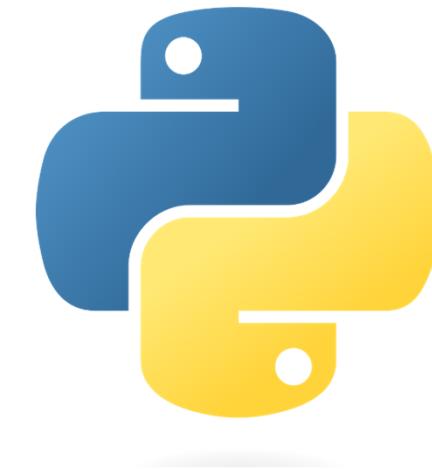
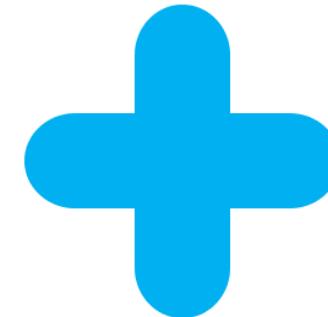


Python 程式設計-李明昌 免費電子書 - PDF 分享, 220頁

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



學習目標



R 入門資料分析與視覺化應用(7小時28分鐘+字幕)

- <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+字幕)

- <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



2.Python/Anaconda環境介紹



Python 簡介與安裝



Python 簡介

- 吉多·范羅蘇姆 (Guido van Rossum) 在1989年的聖誕節期間研發 Python 語言。
 - https://en.wikipedia.org/wiki/Guido_van_Rossum
 - Python 3.11.0 - 2022年11月13日
- 特性：
 - 跨平台
 - 開放性
 - 易讀性
 - 動態語言
 - 直譯語言
 - 豐富套件(模組)
 - 其他語言結合, 例: Cython 編譯成執行檔(.exe)



Python 下載

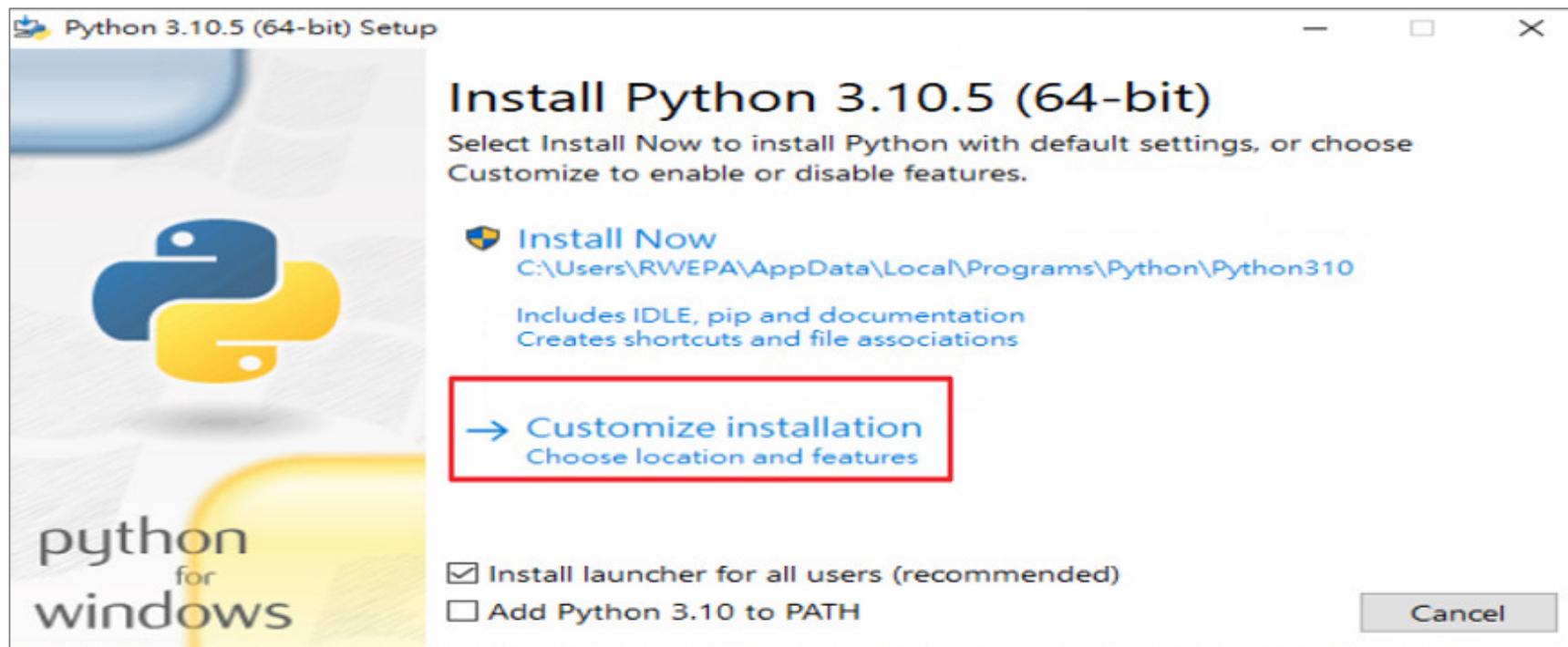
- <https://www.python.org/>

The screenshot shows the Python.org homepage with a dark blue header. The navigation bar includes links for Python, PSF, Docs, PyPI, Jobs, and Community. The main content area features the Python logo and a search bar with a magnifying glass icon. A sidebar on the left contains a snippet of Python code and a list of download options: All releases, Source code, Windows, macOS, Other Platforms, License, and Alternative Implementations. The central column is titled "Download for Windows" and features a prominent button for "Python 3.11.0". Below this, there is a note about compatibility with Windows 7 or earlier, followed by a link to view the full list of downloads.

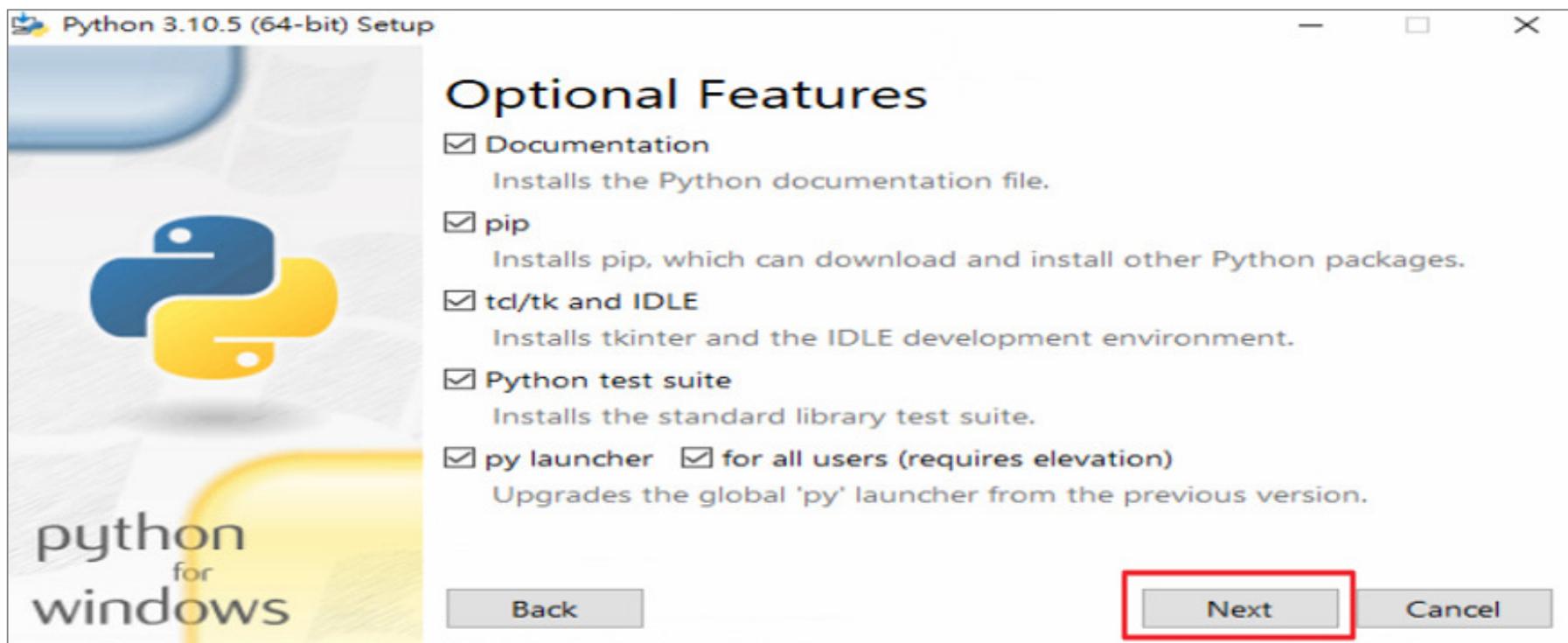
Python is a programming language that lets you work quickly
and integrate systems more effectively. [» Learn More](#)

Python 安裝

- python-3.10.5-amd64.exe (最新版 python-3.11.0-amd64.exe)

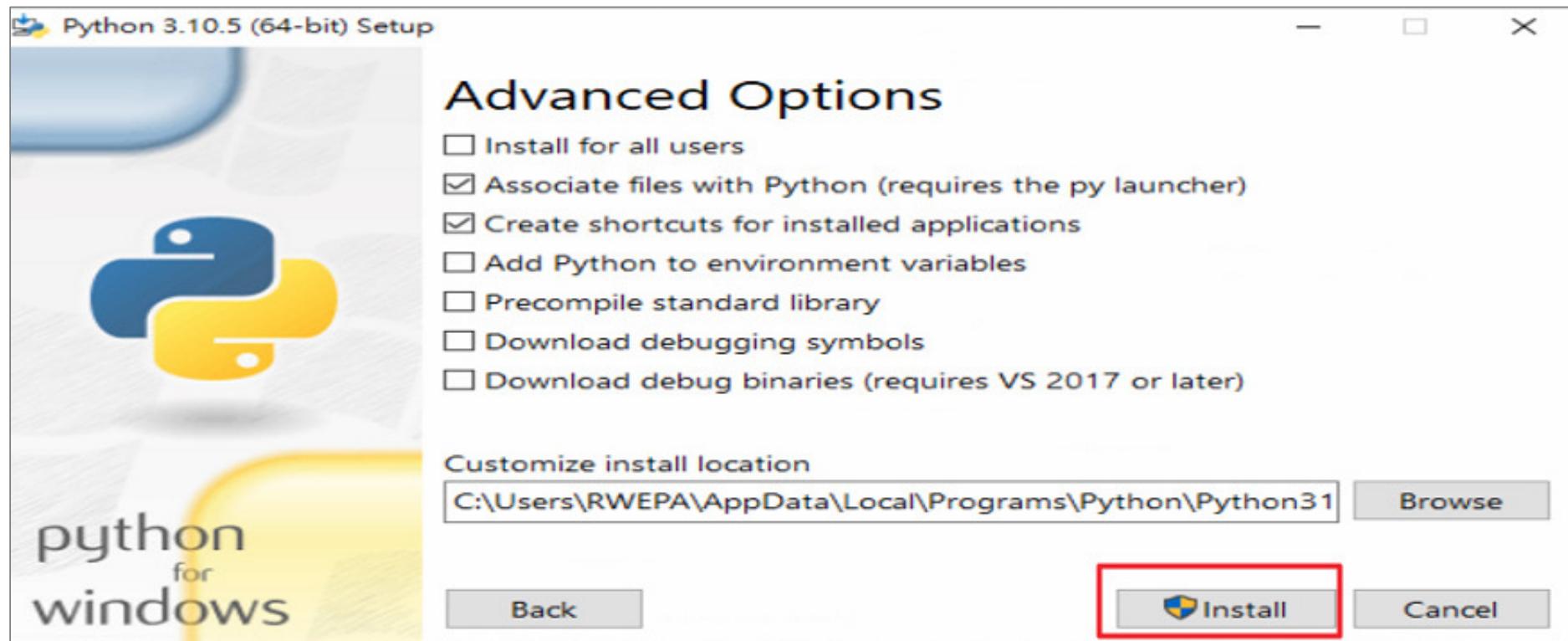


Python 安裝 (續)



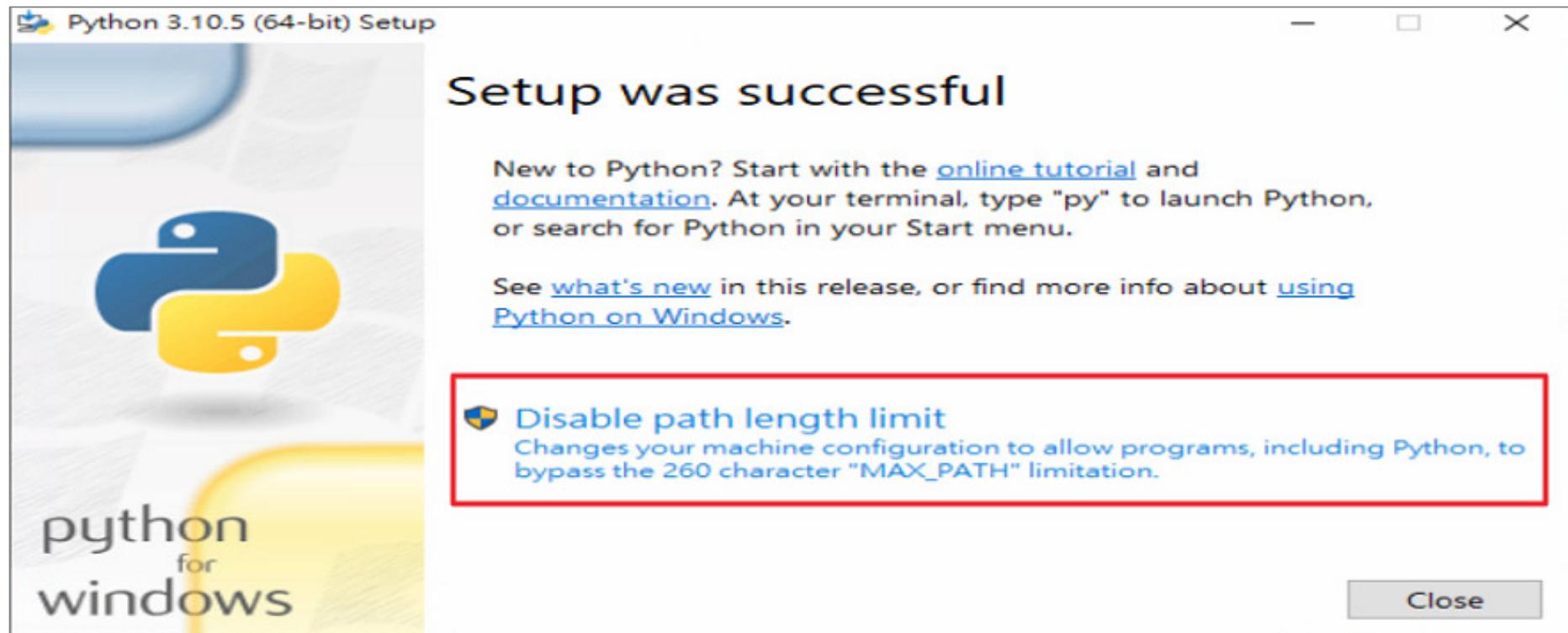


Python 安裝 (續)

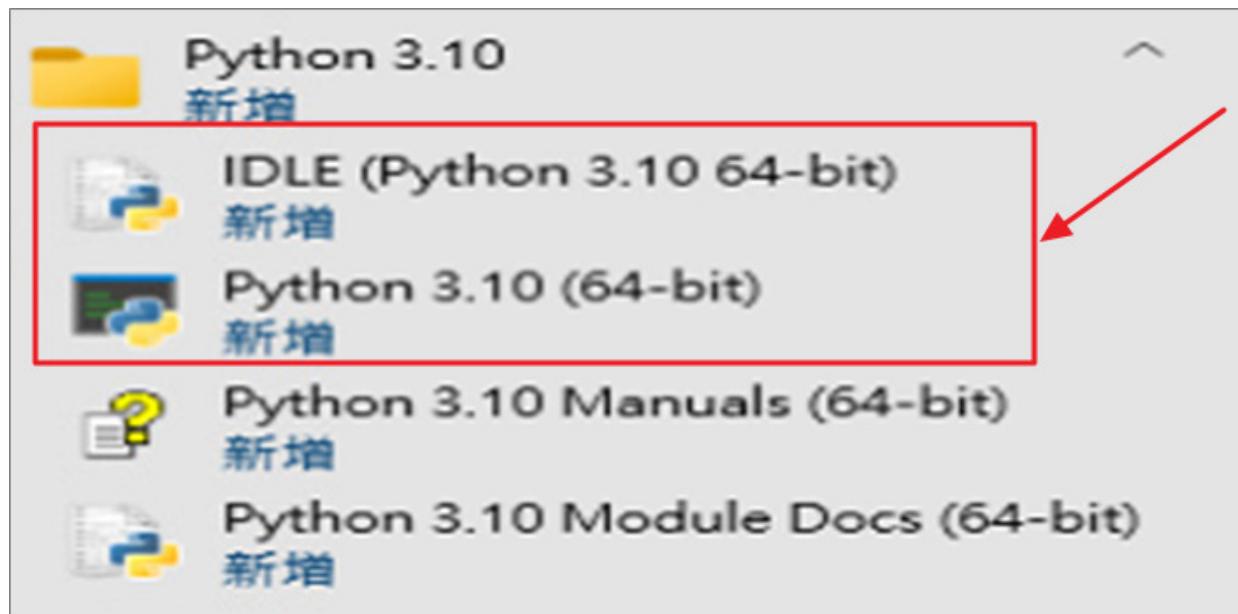




Python 安裝 – 取消字元長度限制



Python 安裝完成



安裝完成有2個執行模式：

1. IDLE
2. 命令提示字元



Python 執行 - 命令提示字元

The screenshot shows a Windows Command Prompt window titled "命令提示字元". The window displays the following text:

```
Microsoft Windows [版本 10.0.19042.928]
(c) Microsoft Corporation. 著作權所有，並保留一切權利。

C:\Users\rwepa>python
Python 3.8.5 (default, Sep  3 2020, 21:29:08) [MSC v.1916 64 bit (AMD64)] :: Anaconda, Inc. on win32

Warning:
This Python interpreter is in a conda environment, but the environment has
not been activated. Libraries may fail to load. To activate this environment
please see https://conda.io/activation

Type "help", "copyright", "credits" or "license" for more information.
>>> quit()

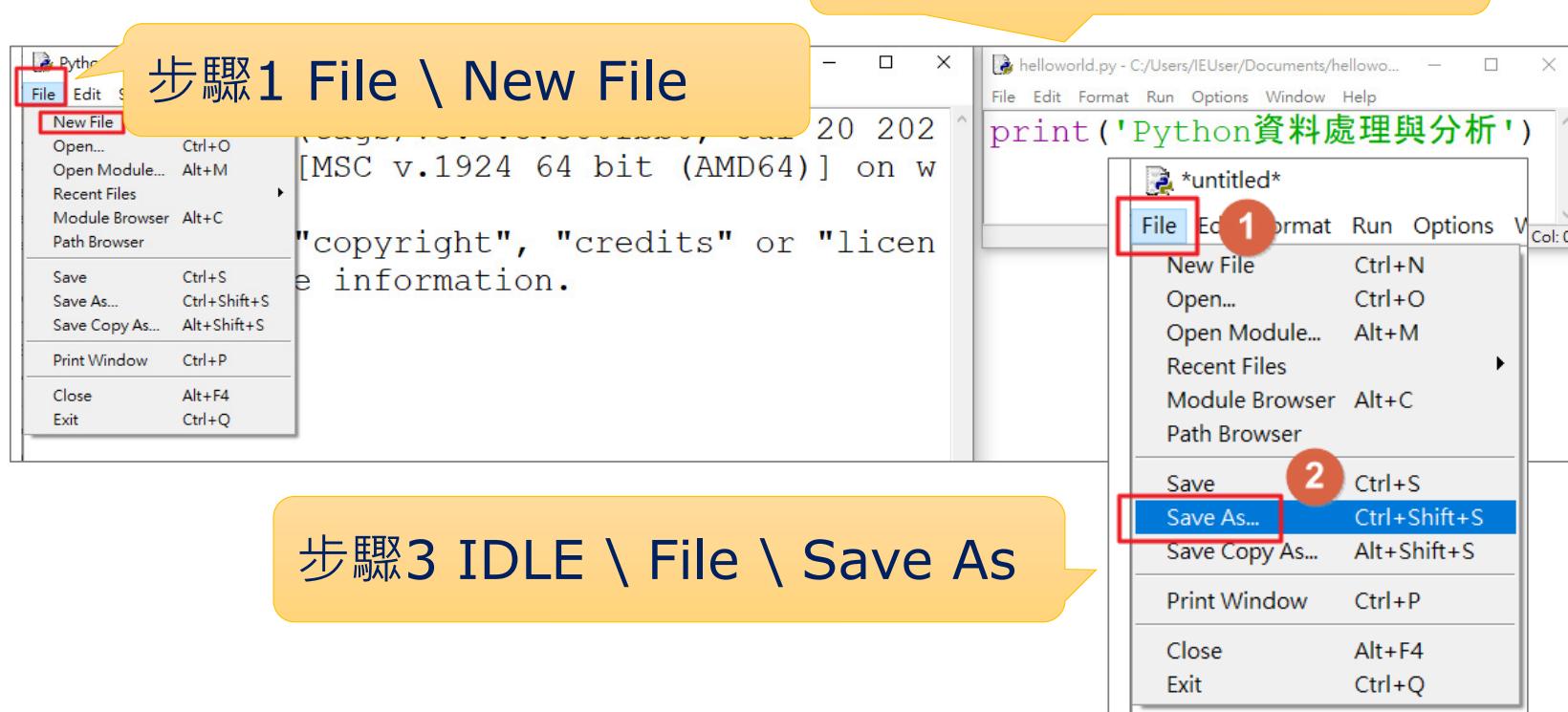
C:\Users\rwepa>
```

- python
- 1+2
- quit()

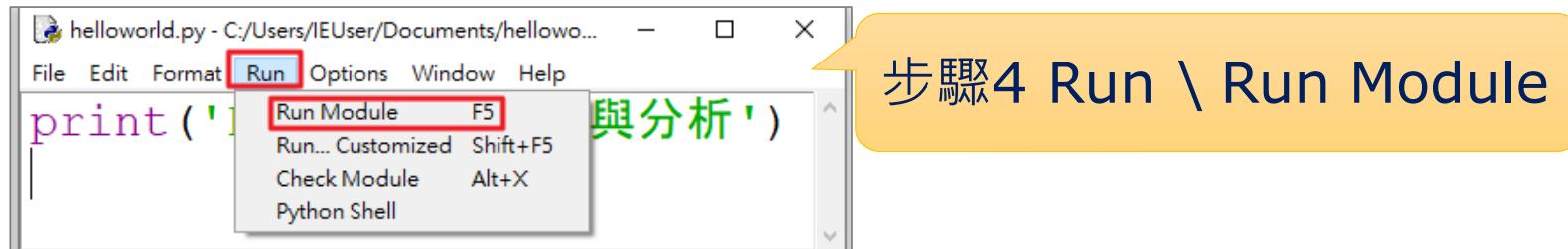


Python 執行 - IDLE模式

- IDLE 可先將檔案編輯與儲存



Python 執行 - IDLE模式 (續)

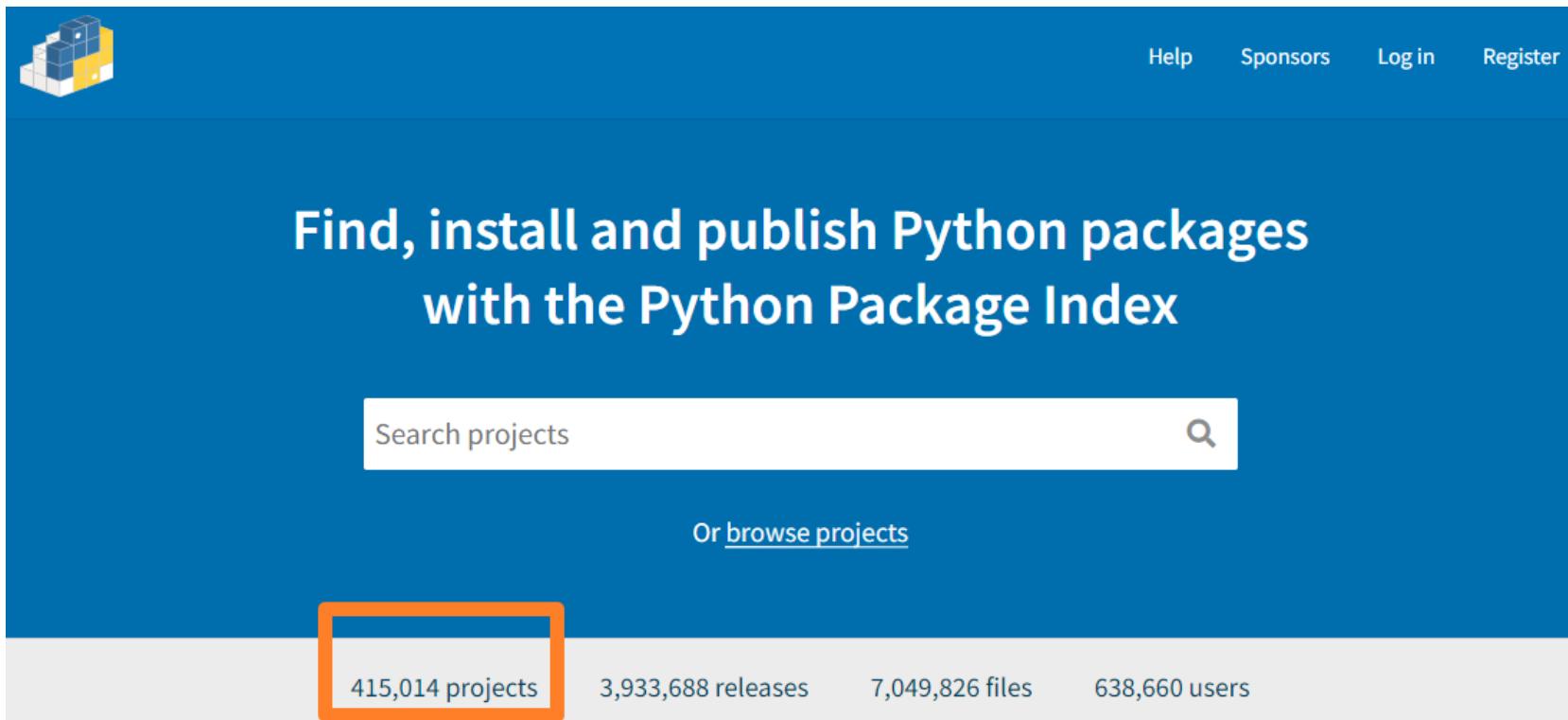


步驟4 Run \ Run Module

A screenshot of the Python IDLE terminal window. The title bar says "Python". The window shows the Python startup message: "Python 3.8.5 (tags/v3.8.5:580fbbf, Jul 20 2020, 15:57:54) [MSC v.1924 64 bit (AMD64)] on win32". It then displays the command-line interaction: ">>> = RESTART: C:/Users/IEUser/Documents/helloworld.py". A yellow callout bubble points to this line, which is highlighted with a red box. Below it, the text "Python資料處理與分析" is displayed in blue. The bottom right corner of the window shows "Ln: 7 Col: 4".

PyPI (Python Package Index)

- <https://pypi.org/>
- 41萬專案總表





已安裝模組(pip)

The image shows two command-line windows. The top window is titled '命令提示字元' and shows the output of the 'pip list' command on Windows. The bottom window is titled '命令提示字元 - python -c help('os')' and shows the help documentation for the 'os' module in Python.

Top Window (命令提示字元):

```
Microsoft Windows [版本 10.0.19043.1055]
(c) Microsoft Corporation. 著作權所有，並保有所有權利。
C:\Users\88697>pip list
Package           Version
-----
absl-py          0.12.0
appdirs          1.4.4
astunparse       1.6.3
cachetools      4.2.1
certifi          2020.6.22
chardet          4.0.0
defusedxml      0.5.0
distlib          0.3.1
engineering-numpy 1.1.0
```

Bottom Window (命令提示字元 - python -c help('os')):

```
C:\Users\88697>python -c help('os')
Help on module os:

NAME
    os - OS routines for NT or Posix depending on what system we're on.

MODULE REFERENCE
    https://docs.python.org/3.8/library/os

    The following documentation is automatically generated from the Python
    source files. It may be incomplete, incorrect or include features that
    are considered implementation detail and may vary between -- More --
```

Callout Bubbles:

- **pip list**
- 模組說明
python -c help('模組')
- 按 Q 關閉說明



Python IDE

- Anaconda, 包括 Spyder:
 - <https://www.anaconda.com/>
- PyCharm:
 - <https://www.jetbrains.com/pycharm/>
- WinPython:
 - <http://winpython.github.io/>
- RStudio - Terminal 視窗
 - <https://www.rstudio.com/products/rstudio/>
- Visual Studio Code
 - <https://code.visualstudio.com/docs/python/python-tutorial>
- Google:
 -  Google Colaboratory

示範軟體 (Spyder)

IDE 整合開發環境-
Integrated
Development
Environment

RStudio – 執行 Python

The screenshot shows the RStudio interface with a Python script named '01.python.讀我.py' open in the code editor. The script contains several lines of Python code, including imports for os and matplotlib, and operations like changing the current working directory. A yellow callout bubble points to the terminal window, labeled '步驟1 先在 Terminal 視窗輸入 python'. In the terminal, the user has run 'python' and is now in a Python 3.7.1 environment. Another yellow callout bubble points to the terminal input area, labeled '步驟2 執行 Ctrl + Alt + Enter'. The RStudio environment pane shows an 'Environment' tab with an 'Empty Environment' message. The file browser pane shows various Python files and their details. A third yellow callout bubble points to the terminal output, labeled '步驟3 顯示結果', which shows the list of files from the script's print statement.

```
5 author: Ming-Chang Lee
6 email : alan9956@gmail.com
7 RWEPA : http://rwepa.blogspot.tw/
8 """
9
10 # 檔案資料夾 c:/pythondata
11 # 01.python.讀我.py      , 檔案說明
12 # 02.type.operations.py , 資料操作
13 # 03.matplotlib.demo.py , matplotlib繪圖
14 # 04.understandingdata.py , 資料理解 pandas
15 # 05.regression.py      , 迴歸分析,線性模型
16 # 06.logistic.regression.py , logistic regression-分類
17 # 07.decision.tree.py   , 決策樹 decision tree & 隨機森林法 random forest
18 # 08.decision.tree-codes.py , 決策樹 Python 程式碼
19 # 09.associationrule.py , 關聯規則
20
21 # 切換工作目錄
22 import os # 載入 os 套件
23 os.getcwd() # 讀取工作目錄
24 os.chdir("C:/pythondata") # 變更工作目錄
25 os.getcwd()
26 os.listdir(os.getcwd()) # 顯示檔案清單
27
28 # 顯示模組提供之函數
2633
```

Console Terminal x

Terminal 1 (busy) - python

Microsoft Windows [版本 10.0.17134.590]
(c) 2018 Microsoft Corporation. 著作權所有。

```
C:\rdata\python
Python 3.7.1 (default, Dec 10 2018, 22:54:23) [MSC v.1915 64 bit (AMD64)] :: Anaconda custom (64-bit) on win32
Type "help", "copyright", "credits" or "license" for more information.
>>> import os # 載入 os 套件
>>> os.getcwd() # 讀取工作目錄
'C:\\rdata'
>>> os.chdir("C:/pythondata") # 變更工作目錄
>>> os.getcwd()
'C:\\pythondata'
>>> os.listdir(os.getcwd()) # 顯示檔案清單
['01.python.讀我.py', '02.type.operations.py', '03.matiplotlib.demo.py', '04.understandingdata.py', '05.regression.py', '06.logistic.regression.py', '07.decision.tree.py', '08.decision.tree-codes.py', '09.associationrule.py', '2017年10月每小時網路流量.png', 'data', 'random.plot.pdf', 'random.plot.png', 'titanic-RWEPA.png']
>>> 
```

Environment History Connections

Global Environment

Environment is empty

Files Plots Packages Help Viewer

	Size	Modified
01.python.讀我.py	1.4 KB	Feb 18, 2019, 5
02.type.operations.py	5.6 KB	Dec 21, 2018, 5
03.matplotlib.demo.py	4.7 KB	Dec 21, 2018, 5
04.understandingdata.py	2.3 KB	Dec 21, 2018, 5
05.regression.py	2.8 KB	Dec 21, 2018, 5
06.logistic.regression.py	1.4 KB	Aug 13, 2018, 1
07.decision.tree.py	6.7 KB	Dec 21, 2018, 5
08.decision.tree-codes.py	3.7 KB	Dec 21, 2018, 5
09.associationrule.py	2.1 KB	Dec 21, 2018, 5
2017年10月每小時網路流量.png	133.8 KB	Aug 19, 2018
data	15.7 KB	Aug 19, 2018
random.plot.pdf	15.7 KB	Aug 19, 2018



Anaconda 簡介與安裝





Anaconda 特性

- Anaconda是一個免費、易於安裝/管理並支援Python語言
- 支援 8000個以上的資料科學與機器學習模組 (package)
- 支援模組的下載, 安裝, 更新
- 支援 Spyder (支援 Python IDE)
- 支援 jupyter-notebook
- 支援 Windows、Mac OS和Linux

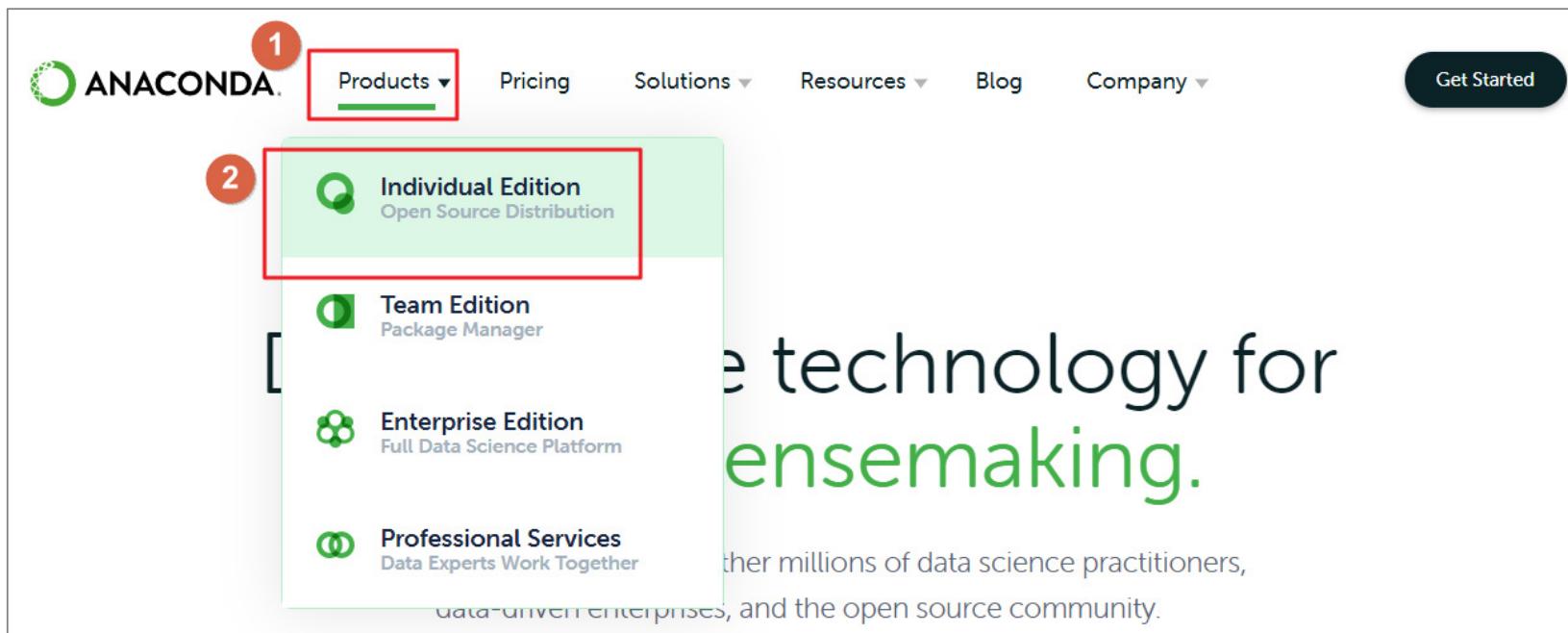
Anaconda 特性 (續)



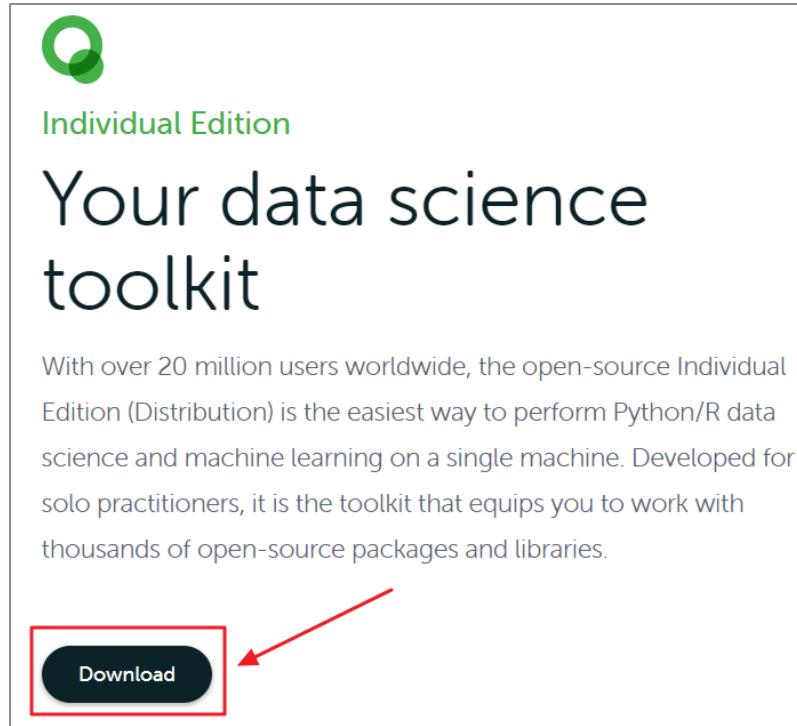
source: <https://www.anaconda.com/distribution/>

Anaconda 下載

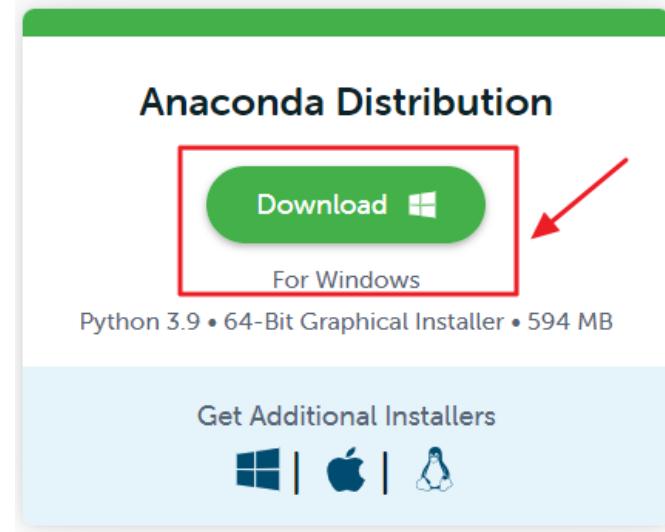
- <https://www.anaconda.com/>



Anaconda 下載 (續)



The screenshot shows the landing page for the Anaconda Individual Edition. It features a large green 'Q' logo and the text 'Individual Edition'. Below this, the heading 'Your data science toolkit' is displayed in a large, bold, dark font. A descriptive paragraph follows, stating: 'With over 20 million users worldwide, the open-source Individual Edition (Distribution) is the easiest way to perform Python/R data science and machine learning on a single machine. Developed for solo practitioners, it is the toolkit that equips you to work with thousands of open-source packages and libraries.' At the bottom, there is a 'Download' button with a red border and a red arrow pointing to it from the left.

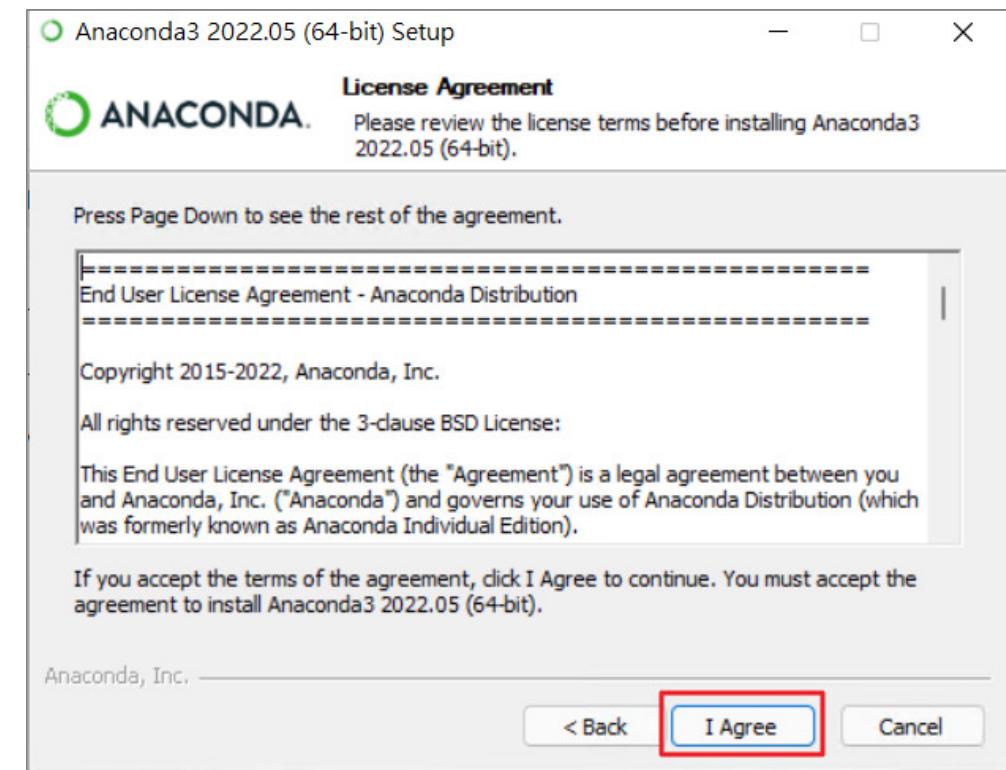
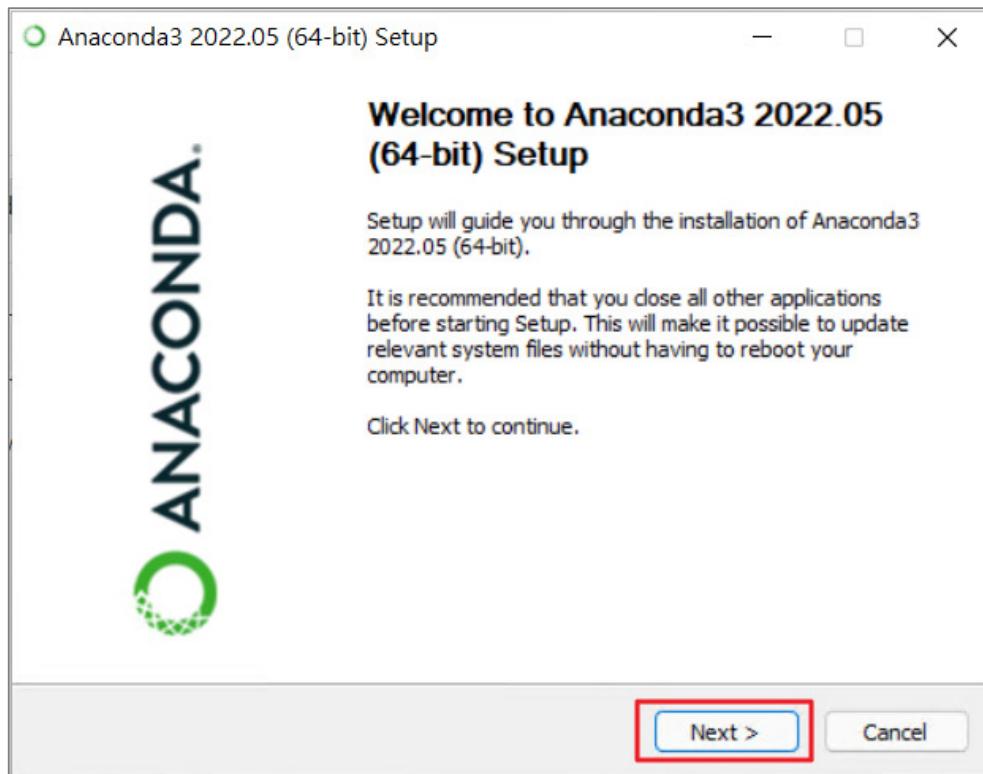


The screenshot shows the download page for the Anaconda Distribution. The title 'Anaconda Distribution' is at the top. Below it is a large green 'Download' button with a white Windows icon, which is also highlighted with a red box and a red arrow. The text 'For Windows' is underneath the button. Further down, it says 'Python 3.9 • 64-Bit Graphical Installer • 594 MB'. Below this, there is a section titled 'Get Additional Installers' with icons for Windows, Apple, and Linux.

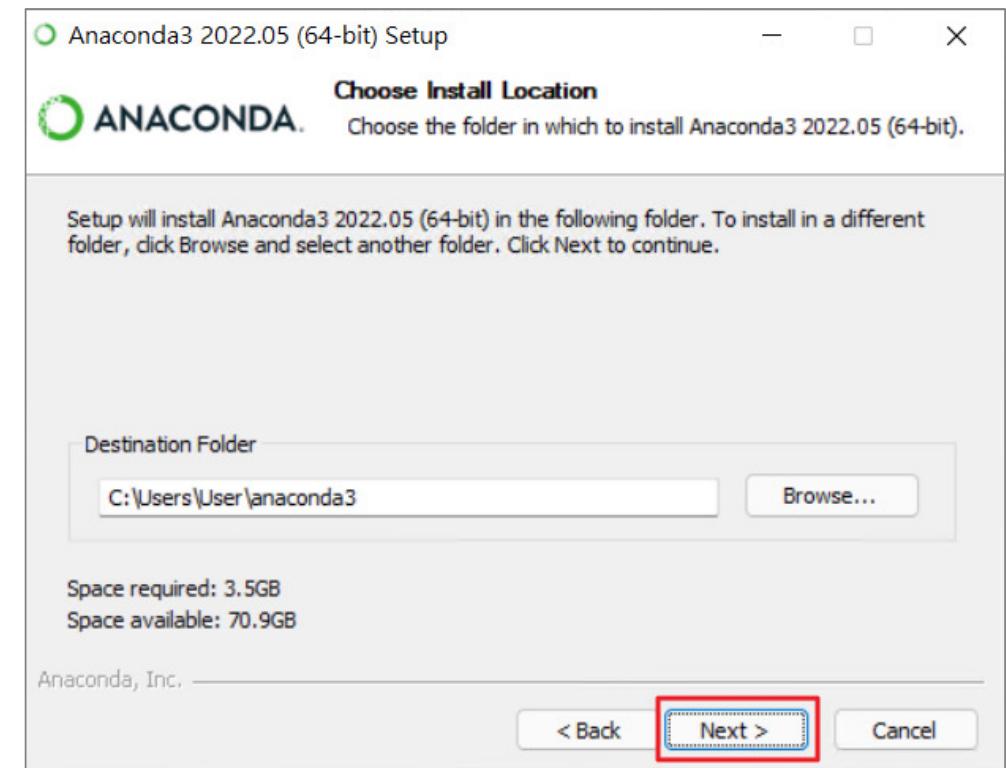
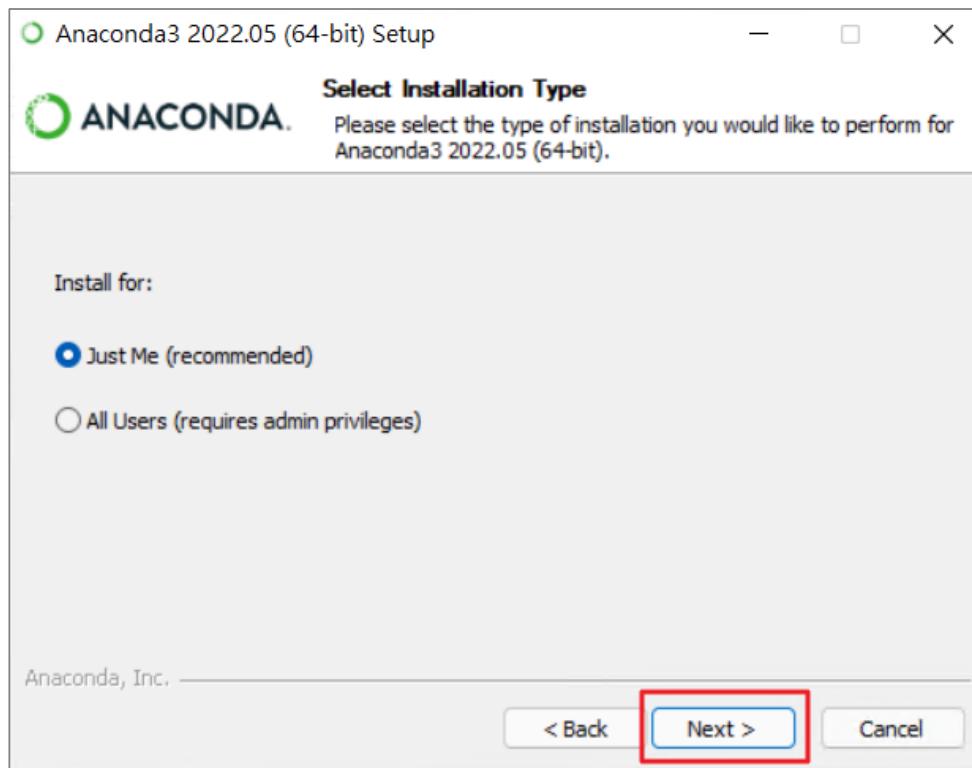
Anaconda3-2022.05-Windows-x86_64.exe

下載 64位元,
594 MB

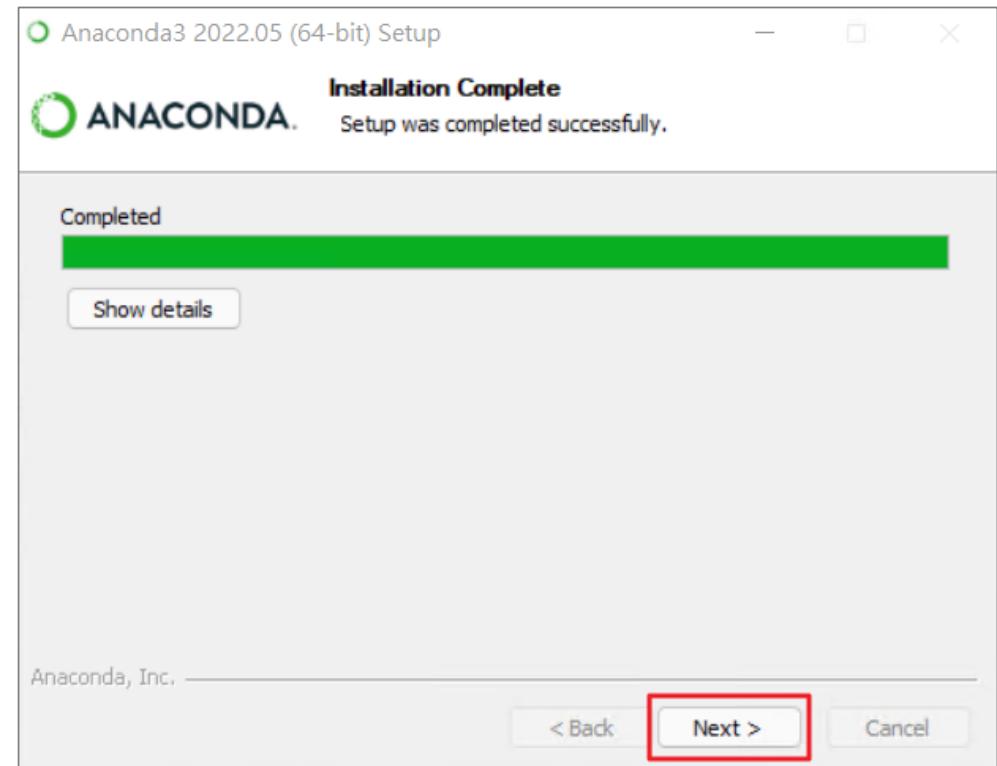
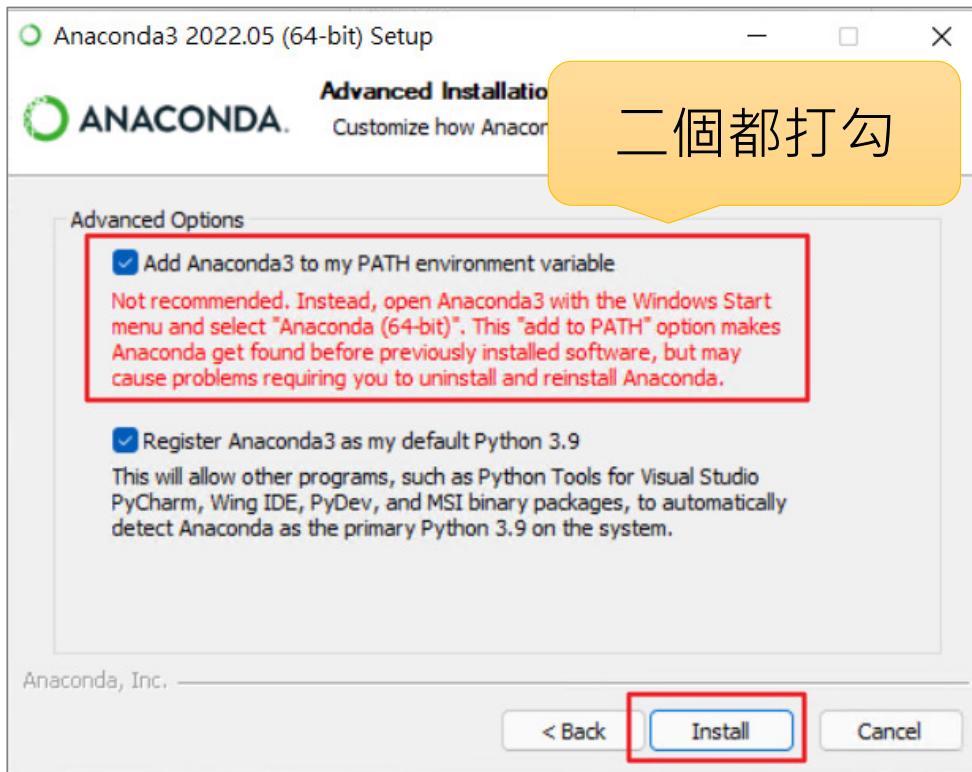
Anaconda 安裝



Anaconda 安裝 (續)



Anaconda 安裝 (續)





Anaconda3 (64-bit)



已安裝
6個元件

撰寫 Python 程式



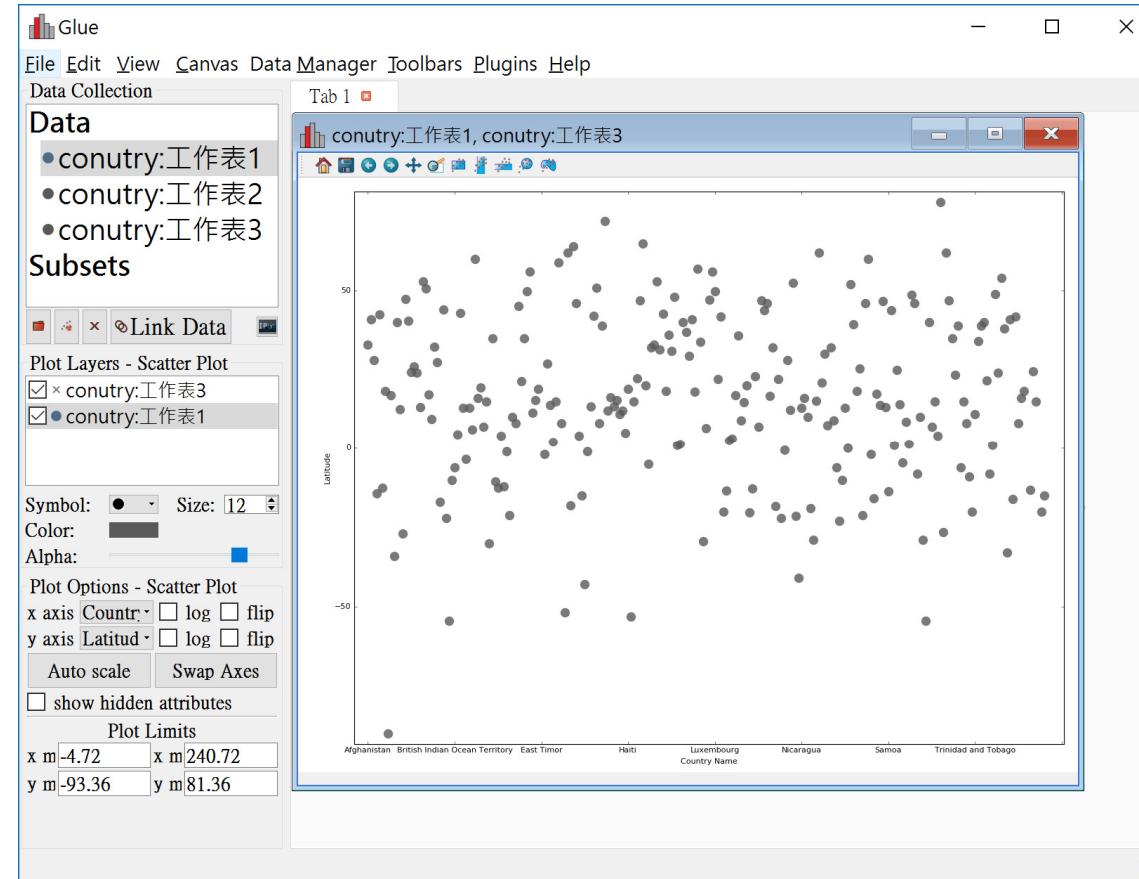
Anaconda Navigator

The screenshot shows the Anaconda Navigator interface. On the left is a sidebar with links to Home, Environments, Learning, Community, and Anaconda Nucleus (with a 'Join Now' button). The main area displays a grid of application icons, each with a 'Launch' or 'Install' button. A yellow callout bubble contains the text: '不同電腦, 安裝模組可能有差異' (Different computers, module installation may differ). Another yellow callout bubble at the bottom right contains the text: '注意: 不要在此安裝 RStudio, 可能會有問題?' (Note: Do not install RStudio here, there may be problems?).

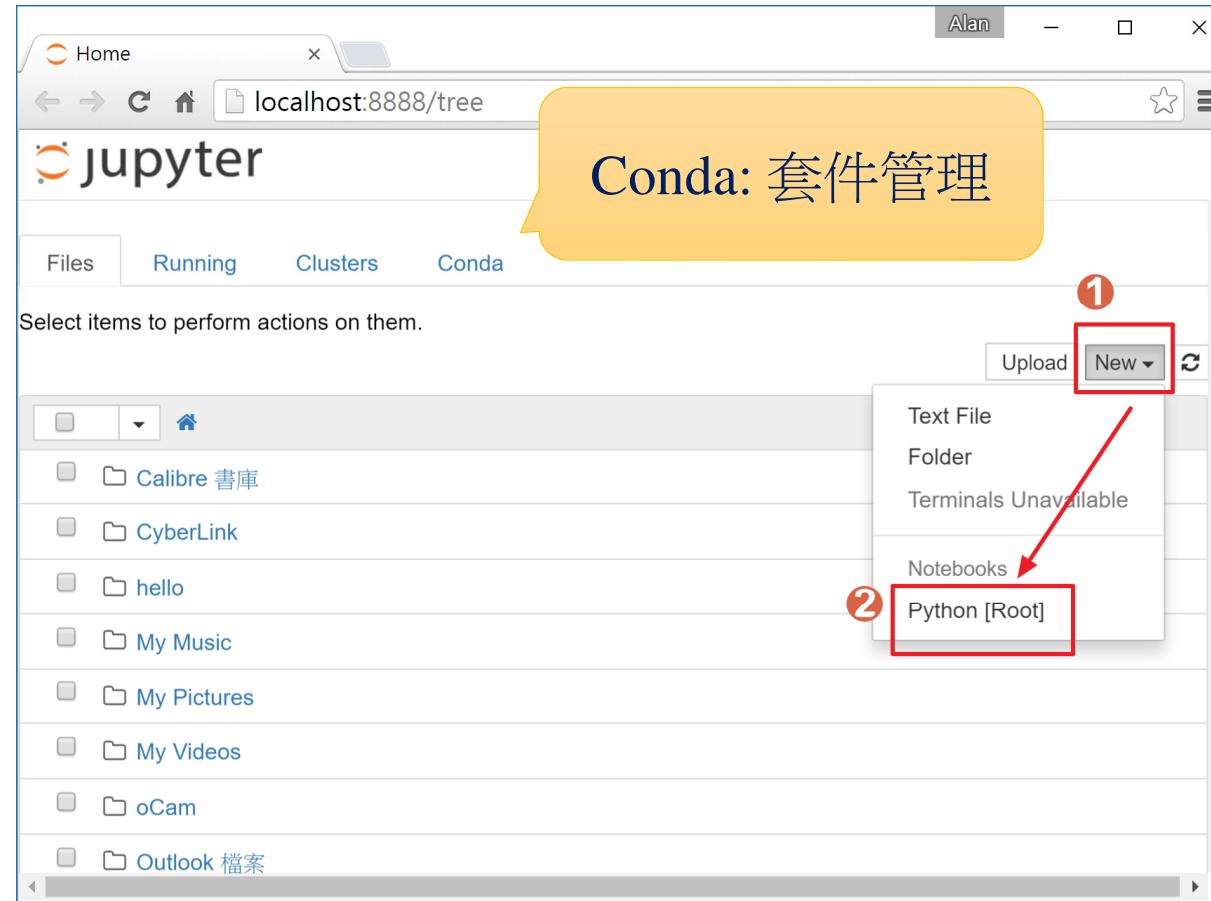
Application	Description	Action
CMD.exe Prompt	Run a cmd.exe terminal with your current environment from Navigator activated.	Launch
Datalore	Online Data Analysis Tool with smart coding assistance by JetBrains. Edit and run your Python notebooks in the cloud and share them with your team.	Launch
IBM Watson Studio Cloud	IBM Watson Studio Cloud provides you the tools to analyze and visualize data, to cleanse and shape data, to create and train machine learning models. Prepare data and build models, using open source data science tools or visual modeling.	Launch
JupyterLab	An extensible environment for interactive and reproducible computing, based on the Jupyter Notebook and Architecture.	Launch
Jupyter Notebook	Web-based, interactive computing notebook environment. Edit and run human-readable docs while describing the data analysis.	Launch
Powershell Prompt	Run a Powershell terminal with your current environment from Navigator activated	Launch
IPyConsole	PyQt GUI that supports inline figures, proper multiline editing with syntax highlighting, graphical calltips, and more.	Launch
Spyder	Scientific Python Development Environment. Powerful Python IDE with advanced editing, interactive testing, debugging and introspection features.	Launch
VS Code	Streamlined code editor with support for development operations like debugging, task running and version control.	Launch
Glueviz	Multidimensional data visualization across files. Explore relationships within and among related datasets.	Install
Orange 3	Component based data mining framework. Data visualization and data analysis for novice and expert. Interactive workflows with a large toolbox.	Install
PyCharm Professional	A full-fledged IDE by JetBrains for both Scientific and Web Python development. Supports HTML, JS, and SQL.	Install
RStudio	A set of integrated tools designed to help you be more productive with R. Includes R essentials and notebooks.	Install

- Launch 可直接啟動
- Install 須安裝

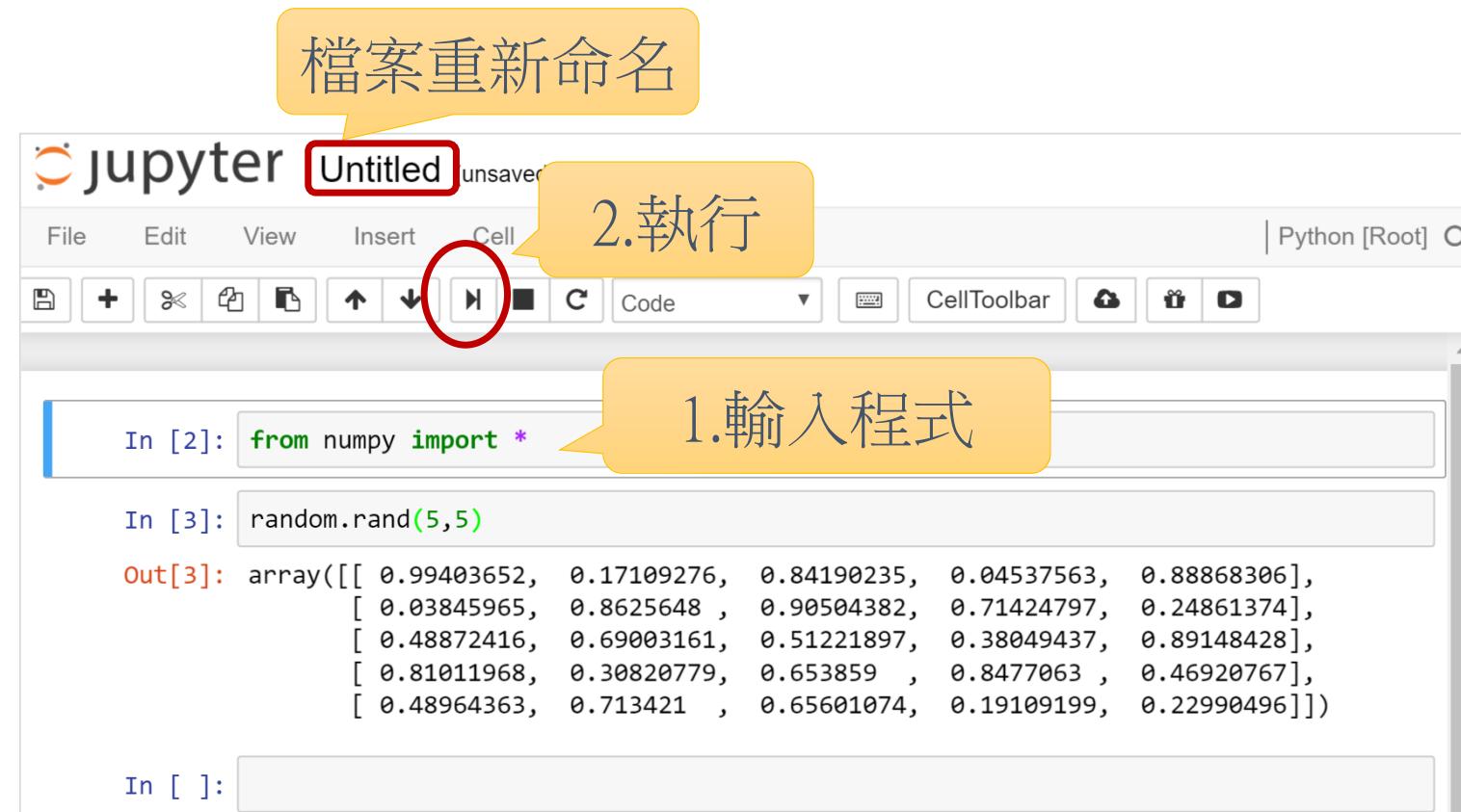
glueviz 視覺化



jupyter-notebook



jupyter notebook (續)



Jupyter-notebook 更改預設目錄

- cd C:\
- jupyter-notebook

或使用 cd \



The screenshot shows the Jupyter Notebook interface. On the left, there's a file tree with several folders like '00.business', '00.data', '00.ir.topicmodel', and '00.R-ebooks-R'. At the bottom, it shows 'localhost:8888/tree#'. On the right, a context menu is open with the following options:

- Upload
- New ▾ (highlighted with a red box and circled with a red number 1)
- Name ▾ (highlighted with a red box and circled with a red number 2)
- Julia 1.2.0
- Python 3 (highlighted with a red box)
- Other:
 - Text File
 - Folder
 - Terminal

Jupyter-notebook 快速鍵

- 按 [Esc] cell旁邊為藍色：
 - 按 **x**：刪除當前選擇的cell
 - 按 **a**：在當前選擇的上方新增一個cell
 - 按 **b**：在當前選擇的下方新增一個cell
 - 按 **Shift + Enter**：執行當前的cell並且選到下一個cell
 - 按 **Ctrl + Enter**：執行當前cell
 - 按 **M**：轉成markerdown模式，可以看到紅色框框內容從code變成markerdown



開啟 → Python程式設計-李明昌.ipynb, 瀏覽加入數學式主題

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

實作練習

主題: Python 程式設計-李明昌 - ipynb

檔名: Python_Programming_Lee_ipynb.zip

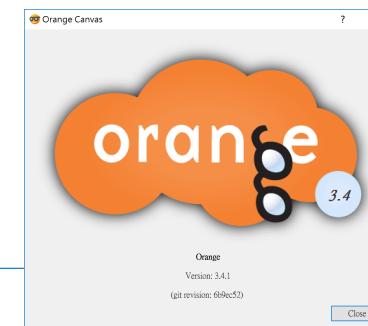
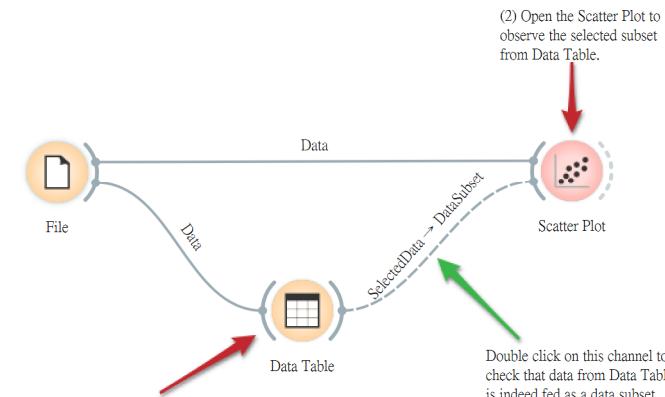
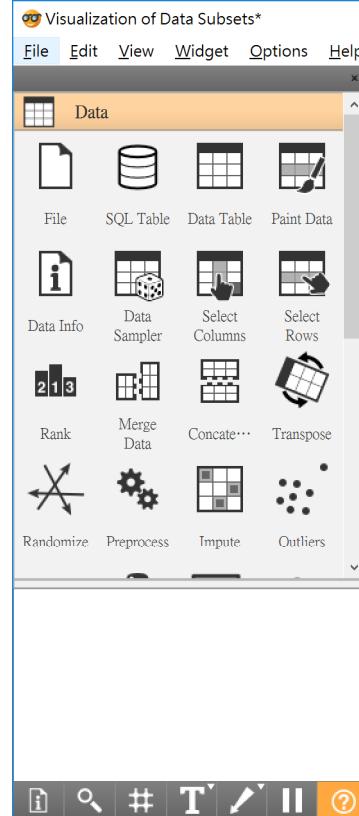
包括 Python 程式設計-李明昌電子書的原始 ipynb 檔案, 圖檔, 部分資料集

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

Orange



安裝 Orange

conda install -c conda-forge orange3

使用命令提示列 開啟 Orange

python -m Orange.canvas

Orange - Scatter Plot



Anaconda 模組管理

- 顯示已安裝模組
conda list

```
命令提示字元
Microsoft Windows [版本 10.0.19043.1055]
(c) Microsoft Corporation. 著作權所有，並保留一切權利。

C:\Users\88697>conda list
# packages in environment at C:\Users\88697\anaconda3:
#
# Name           Version      Build  Channel
_ipyw_jlab_nb_ext_conf    0.1.0      py38_0
alabaster          0.7.12     pyhd3eb1b0_0
anaconda          2021.05     py38_0
anaconda-client     1.7.2      py38_0
anaconda-navigator  2.0.3      py38_0
anaconda-project    0.9.1     pyhd3eb1b0_1
anyio              2.2.0     py38haa95532_2
appdirs             1.4.4      py_0
argh                0.26.2     py38_0
argon2-cffi        20.1.0    py38h2bbff1b_1
asn1crypto          1.4.0      py_0
astroid             2.5       py38haa95532_1
```



Anaconda 模組管理(續)

- 尋找官網套件
`conda search matplotlib`
- 安裝模組
`conda install 模組名稱`
- 更新模組
`conda update 模組名稱`

conda虛擬環境

- <https://github.com/rwepa/DataDemo/blob/master/iPAS-python-program.py#L2667>

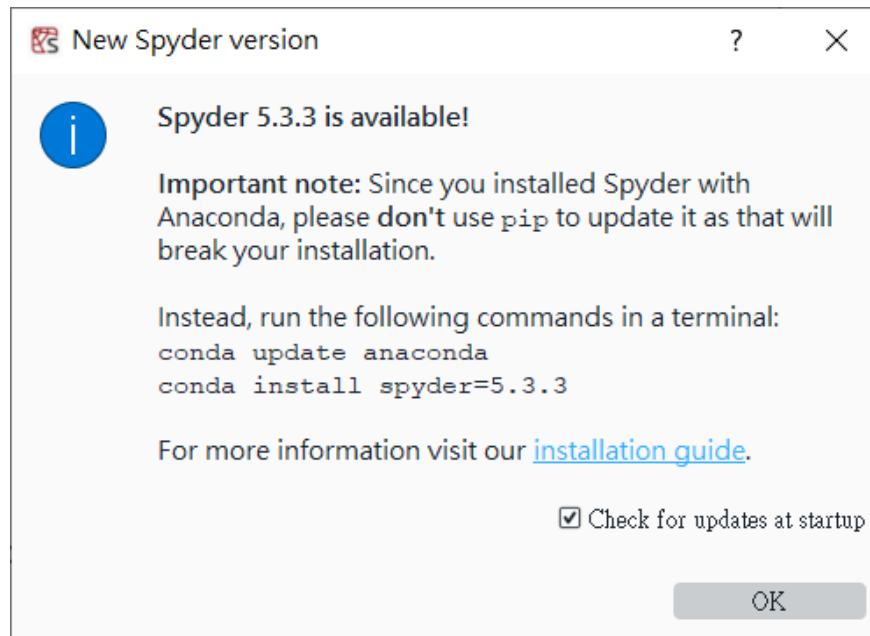
```
# 檢視所有虛擬環境清單  
conda env list  
# 啟用 myenv 虛擬環境  
conda activate myenv  
# 關閉虛擬環境  
conda deactivate  
# 建立 myenv 虛擬環境  
# --name 也可以使用 -n  
# --file 也可以使用 -f  
conda create --name myenv
```

```
# 建立特定 python 版本的虛擬環境  
conda create -n myenv python=3.9  
# 建立特定 scipy 模組版本的虛擬環境  
conda create -n myenv scipy=1.9.0  
# 建立特定 python 版本與特定 scipy 模組版本的虛擬環境  
conda create -n myenv python=3.9  
scipy=1.9.0 astroid babel
```



Spyder 軟體簡介

Spyder 更新

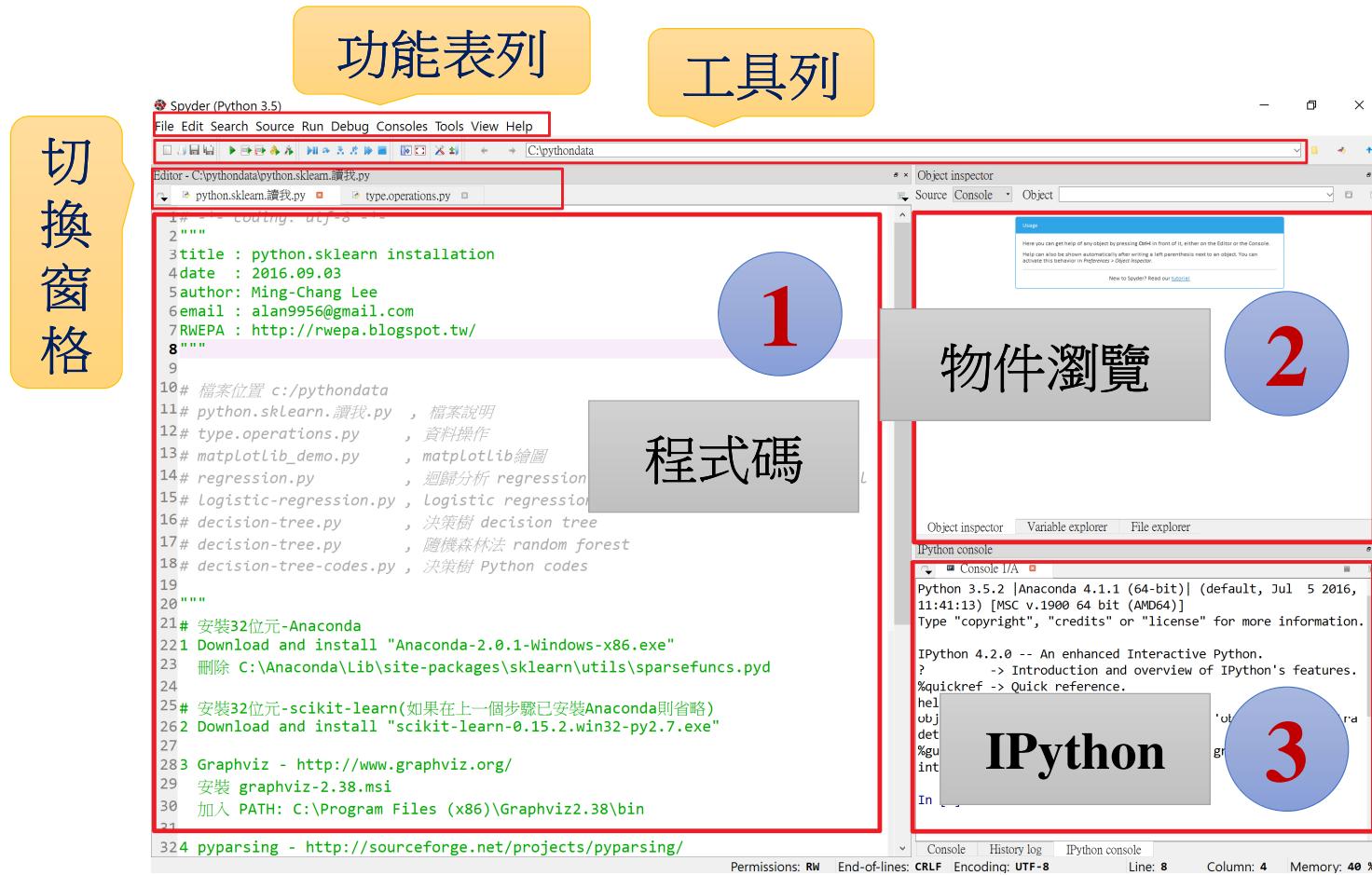


- Anaconda 整體更新
- Spyder 安裝新版

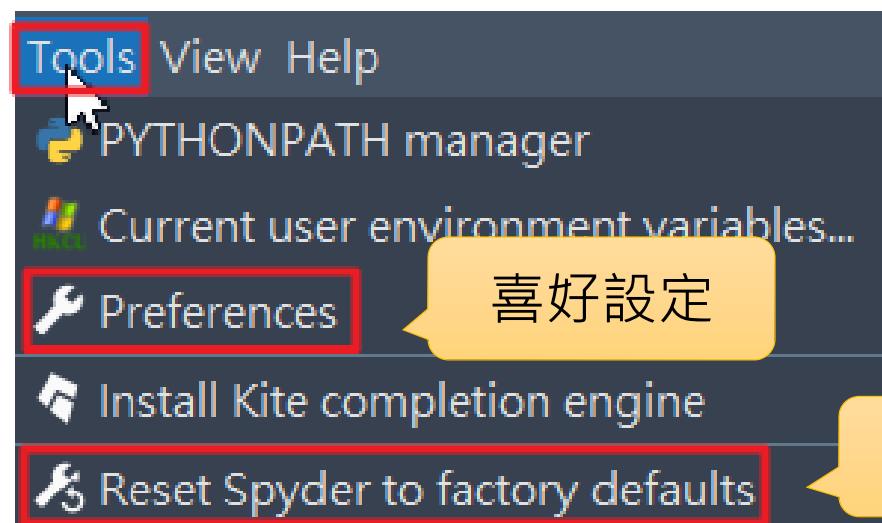
`conda update anaconda`
`conda install spyder=5.3.3`



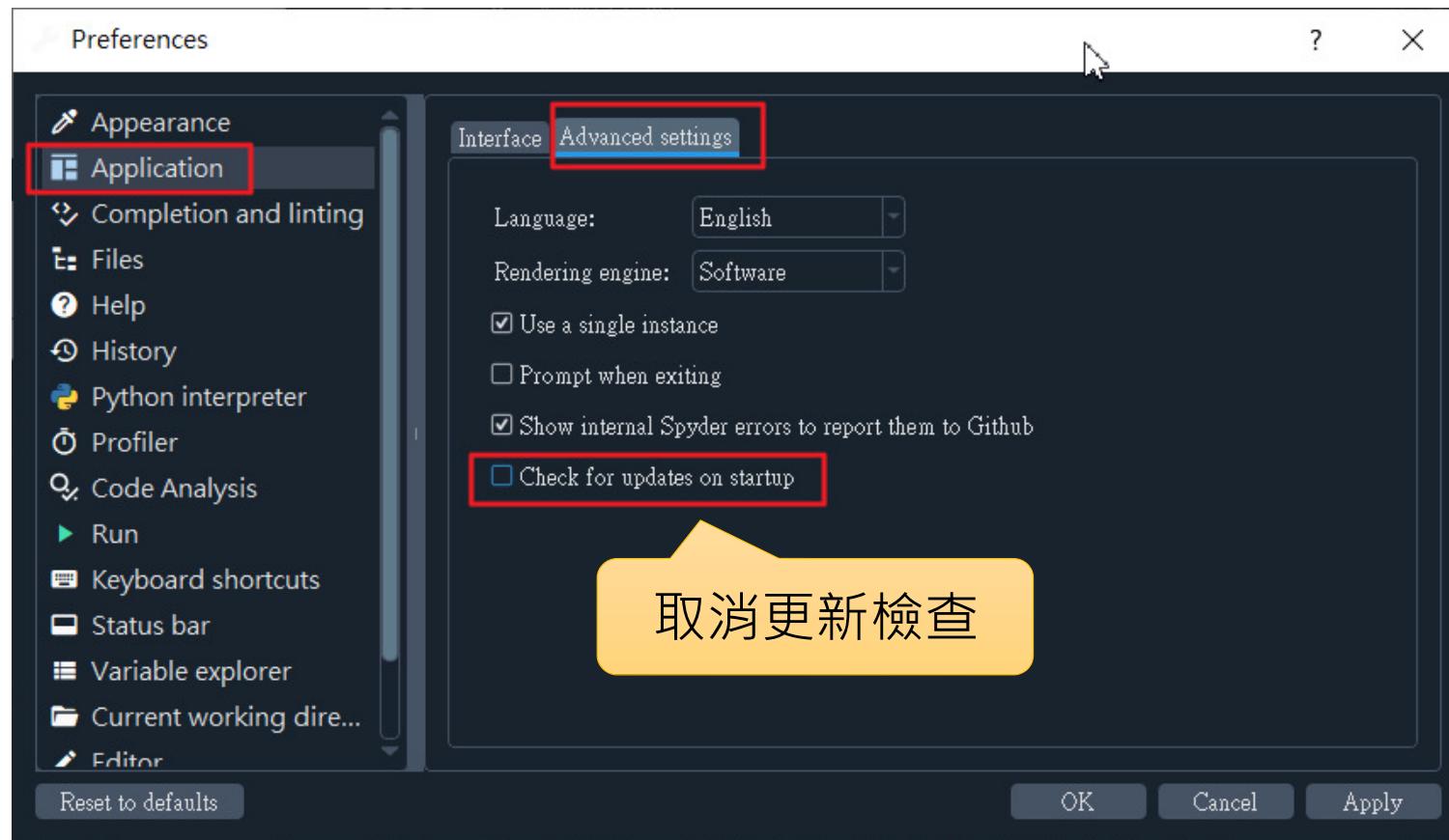
Spyder 畫面



喜好設定 Tools\Preferences



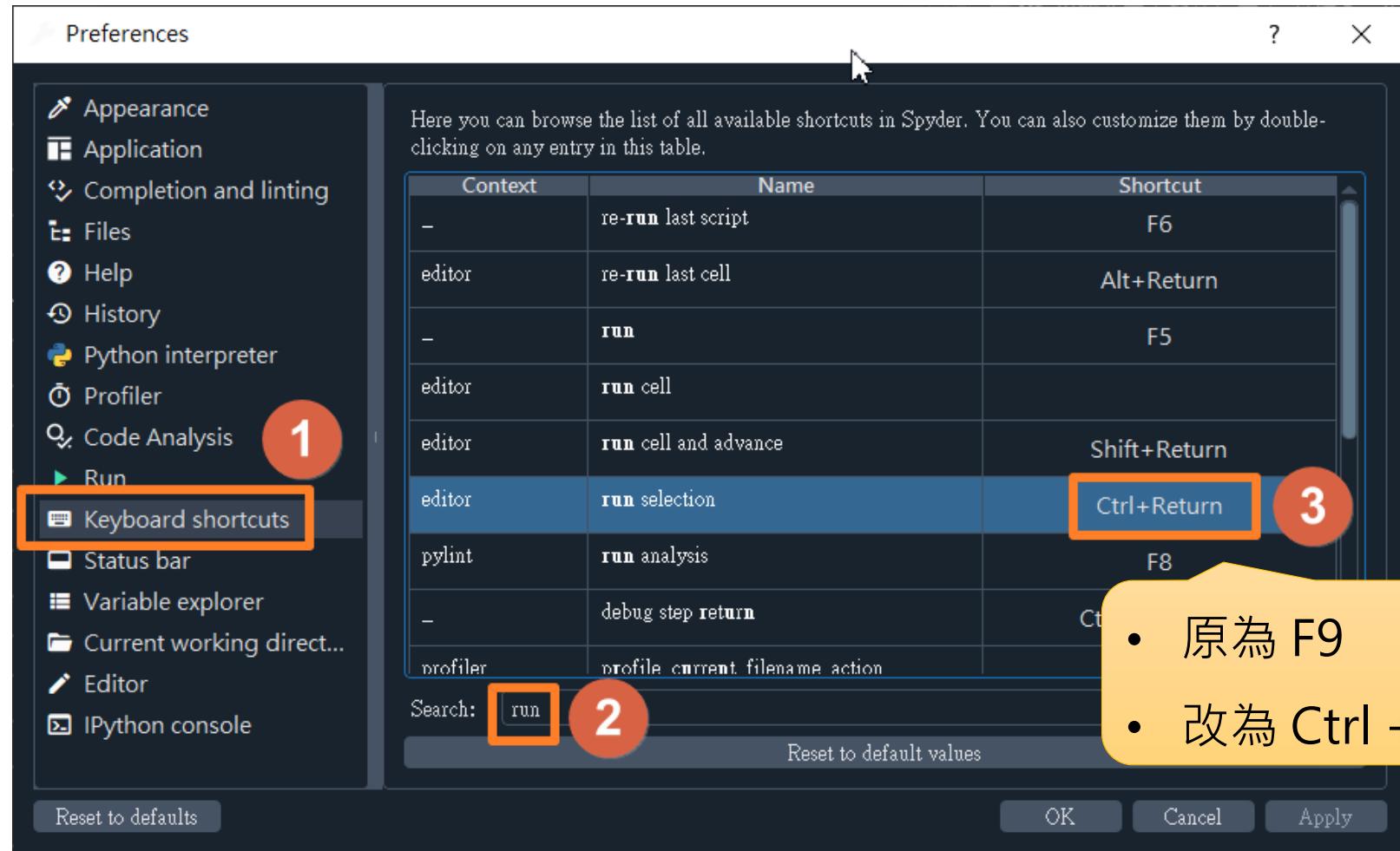
Preferences \ Application \ Advanced setting



Preferences \ Appearance



Preferences \ Keyboard shortcuts \ run selection



Spyder 好用的快速鍵

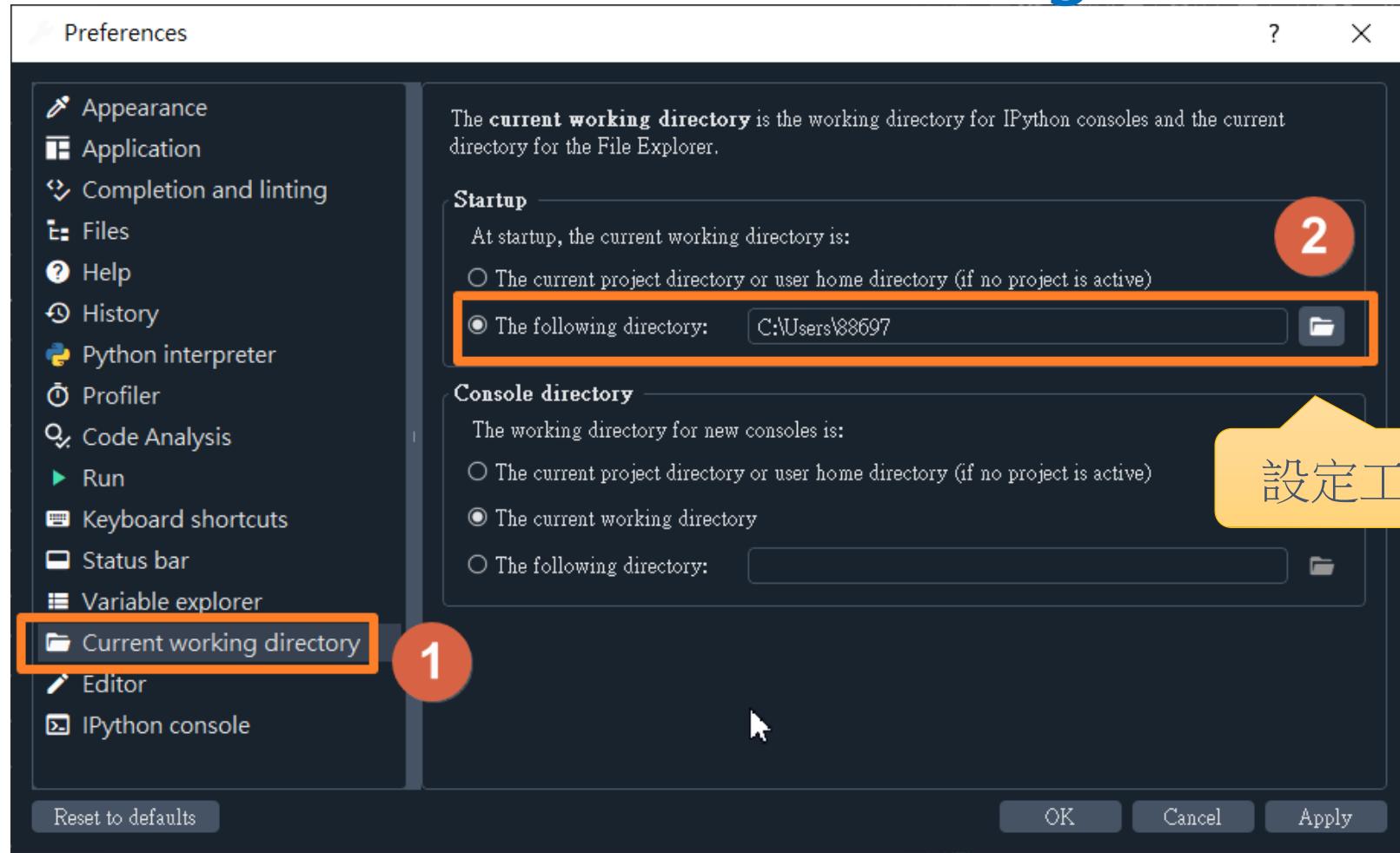
- 執行選取程式碼 **Ctrl + Enter (原為 F9)**
- 註解切換 **Ctrl + 1**
- 尋找 **Ctrl + F**
- 取代 **Ctrl + R**
- 切換至編輯視窗 **Ctrl + Shift + E**
- 切換至 Ipython Console **Ctrl + Shift + I**
- 檔案切換 **Ctrl + P**
- **重新啟動 Ipython Console** **Ctrl + .**
- 檢視放大 **Ctrl + "+"**
- 檢視縮小 **Ctrl + "-"**
- 重新啟動 Spyder **Alt + Shift + R**

Ipython Console 視窗

- 清空: **Ctrl + L**
- 清空: **%clear**

 Split vertically	(上下垂直分割)	Ctrl+{
 Split horizontally	(左右水平分割)	Ctrl+_
 Close this panel		Alt+Shift+W

Preferences \ Current working directory



Preferences \ Editor \ Display \ Wrap lines



實作練習

The screenshot shows the PyCharm Preferences dialog with the 'Editor' category selected (highlighted by a red box with number 1). Within the 'Display' tab (also highlighted by a red box with number 2), the 'Wrap lines' checkbox is highlighted with an orange border. A yellow callout bubble contains the text '自動換列' (Automatic Line Wrap). At the bottom right of the dialog, a red crown icon is displayed above the text '恭喜您, 開啟人生 Python 學習之旅 ^_^'.

Preferences

Display Source code Run code Advanced settings

Show tab bar

Show selector for classes and functions

Show indent guides

Show code folding

Show line numbers

Show blank spaces

Highlight current line

Highlight current cell

Wrap lines

Scroll past the end

Highlight occurrences after 1500 ms

OK Cancel Apply

1

2

自動換列

恭喜您, 開啟人生
Python 學習之旅 ^_^



Python 命令提示字元

- 建立 C:\mydata\helloworld.py

```
C:\mydata\helloworld.py
helloworld.py x
1 # -*- coding: utf-8 -*-
"""
2
3 Created on Tue Jun 22 23:06:49 2021
4
5 @author: rwepa
"""
6
7
8 print("大數據Python應用")
```

輸入： python helloworld.py

```
命令提示字元
C:\mydata>python --version
Python 3.8.8

C:\mydata>dir
磁碟區 C 中的磁碟是 WIN10
磁碟區序號: E428-7C96

C:\mydata 的目錄

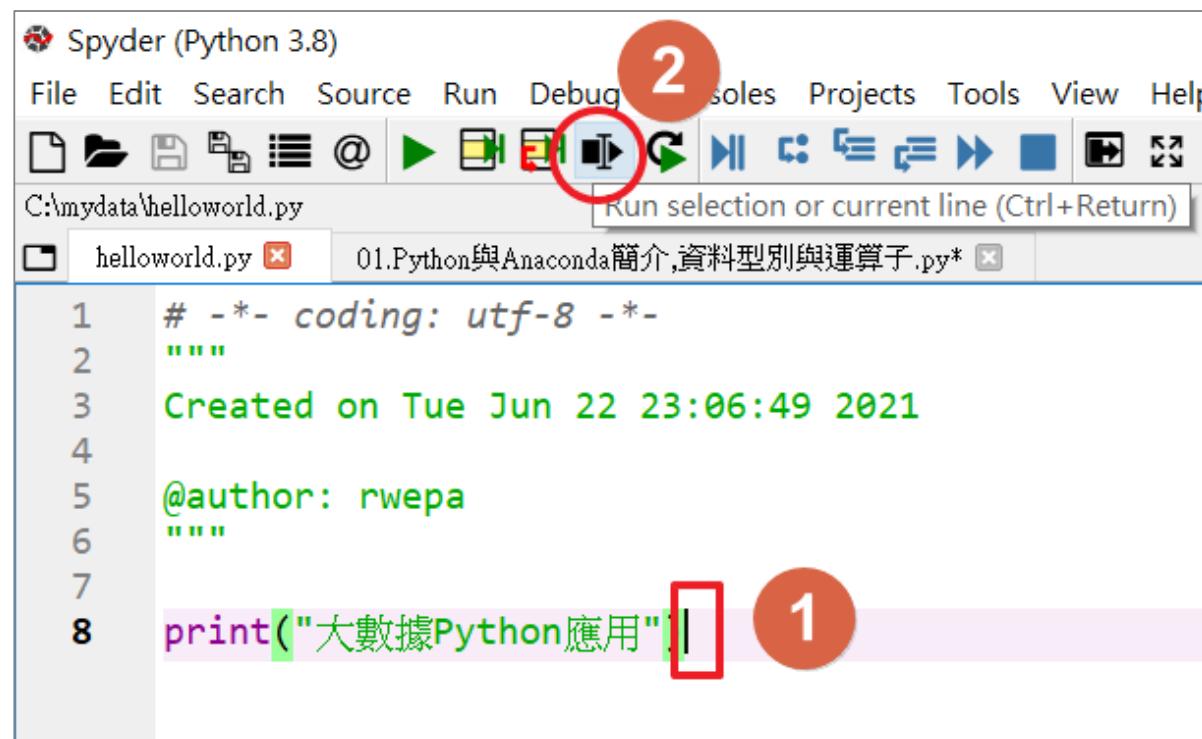
2021/06/22 下午 11:10    <DIR> .
2021/06/22 下午 11:10    <DIR> ..
2021/06/22 下午 11:08          1 個檔案   122 位元組
                           2 個目錄  62,354,751,488 位元組可用
                           .           122 位元組可用

C:\mydata>python helloworld.py
大數據Python應用

C:\mydata>
```

Python 使用 Spyder

- 選取資料列
- 按 [Run selection or current line]





3.全民健保資料庫

匯入、串聯、分割、篩選與轉換



全民健康保險研究資料庫

The screenshot shows a web browser displaying the NHIRD website (https://nhird.nhri.org.tw/date_01.html). The page title is "加值資料庫內容說明". The main content area is titled "說明" (Description) and includes sections for "重要通知" (Important Notice), "加值資料庫內容說明 [English]" (Value-added Database Content Description [English]), and "資料來源" (Data Sources). A red arrow points from the "說明" link in the sidebar to the "說明" section in the main content area. A yellow callout box at the bottom contains the URL <https://nhird.nhri.org.tw/>.

說明

NEW 重要通知
自102/9/1將就罕見疾病、後天免疫症候群二項及102/12/1起增加精神疾病一項資料申請案，實施再次模糊處理，詳情請參考**注意事項說明**。

加值資料庫內容說明 [[English](#)]

- 基本資料檔
- 系統抽樣檔
- 特定主題分檔
- 承保抽樣歸人檔
- 糖尿病病人抽樣歸人檔
- NEW 肺癌病人歸人檔
- 注意事項

資料來源

每年中健保署將前一年的健保資料選取可供研究使用的檔案匯出，將身分欄位加密後，交由國衛院製作成「全民健康保險研究資料庫」及各加值資料檔案。健保署提供的檔案分為下列基本資料檔與原始資料檔兩大類，每個檔案資料欄位名稱和資料描述，在「譯碼簿」中有詳細的說明。

<https://nhird.nhri.org.tw/>

- 醫事機構基本資料檔 (HOSB)
- 醫事機構副檔資料檔 (HOSX)
- 專科醫師證書主檔 (DOC)



資料來源

院所

醫師

病患

- 基本資料檔：
 - 1. 醫事機構病床主檔 (BED)
 - 2. 醫事機構診療科別明細檔 (DETA)
 - 3. 醫事機構基本資料檔 (HOSB)
 - 4. 醫事機構副檔資料檔 (HOSX)
 - 5. 專科醫師證書主檔 (DOC)
 - 6. 醫事人員基本資料檔 (PER)
 - 7. 重大傷病證明明細檔 (HV)
 - 8. 醫事機構服務項目檔 (HOX)
 - 9. 藥品主檔 (DRUG)
 - 10. 承保資料檔 (ID)
 - 11. 評鑑資料檔(HOSP_GRAD)
 - 12. 醫事機構類別明細檔(HOSTDTL)
 - 13. 執業資料紀錄檔(LIC)

- 原始資料檔：

- 1. 住院費用申請總表主檔 (DT)
- 2. 門診費用申請總表主檔 (CT)
- 3. 住院醫療費用清單明細檔 (DD)
- 4. 住院醫療費用醫令清單明細檔 (DO)
- 5. 門診處方及治療明細檔 (CD)
- 6. 門診處方醫令明細檔 (OO)
- 7. 特約藥局處方及調劑明細檔 (GD)
- 8. 特約藥局處方醫令檔 (GO)
- 9. 承保資料檔 (ID)
- 10. 物理治療所調劑檔 (GDD)
- 11. 物理治療所醫令檔 (GOO)

費用





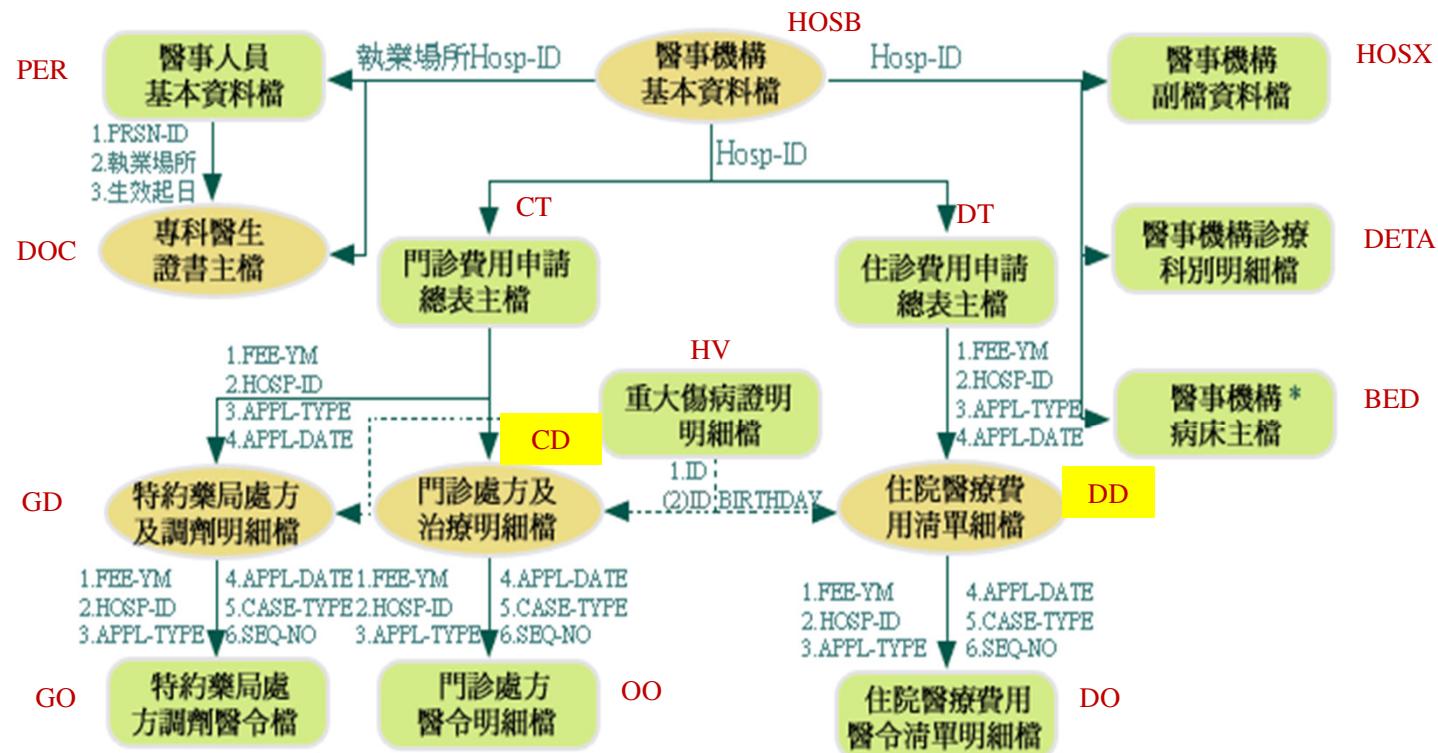
譯碼簿

The screenshot shows a web browser displaying the National Health Insurance Research Database (NHIRD) website at https://nhird.nhri.org.tw/date_02.html. The page title is "譯碼簿". The left sidebar has a red box around the "譯碼簿" link under the "說明" category. A red arrow points from this link to the main content area. The main content area lists five items:

1. 資料描述(完整資料下載 [內](#))
2. 代碼說明(完整資料下載 [內](#))
3. 各檔案間串檔變項說明，請按此 ([GIF 檔，37 KB](#))
4. 練習用虛擬資料檔(97/08/29上線)
5. 健保其他相關業務說明，請參考[中央健康保險署網站](#)。

各檔案間串檔變項說明

各檔案間串檔變項說明



註: *須注意生效起訖日期
 (2)可由ID+BIRTHDAY串檔

→ 各檔案間由所註明變項串檔可獲得對應資訊
 → 各檔案間可由所註明變項串檔,但未必獲得對應資料



全民健保處方及治療明細檔_西醫住院

全民健保處方及治療明細檔—西醫住院
(HEALTH-02 : H_NHI_IPDTE)

模擬資料庫使用手冊

一、 檔案內容說明

檔案代號	H_NHI_IPDTE	資料筆數	14,297 筆(以 2014 為例)		
中文檔名	全民健保處方及治療明細檔—西醫住院	英文檔名	Inpatient Expenditures by Admissions		
檔案大小	7.25MB(以 2014 為例)	欄位數	80		
屬性	年度歷史檔	週期	年	評碼簿 更新日期	2016/09/25
資料描述	1. 原則上一次或一段期間住院一筆資料。 2. 包含病人當次住院之診斷、醫療機構及費用等相關資料。 3. 資料年度包含 2014 年				
注意事項	1. 病人住院期間較長時(通常超過一個月以上)，按月申報住院資料，因此要計算一次住院必須將切帳資料歸戶處理。 2. 補報資料應審慎處理，計算費用時，補報資料應合併計算；但計算醫療利用(人次或人日)時，若為整筆補報應計入，若為部分費用補報則不應重覆計算。				
主鍵與比對欄位	1. 身分證字號(ID)、醫師身分證字號(PRSN_ID)及醫療機構代碼(HOSP_ID)經加密處理。 2. 本檔的唯一鍵值為：FEE_YM + APPL_TYPE + APPL_DATE + CASE_TYPE + SEQ_NO + HOSP_ID，此為與住院醫令檔(H_NHI_IPDTO)對照之鍵值。 3. 歸同次住院的鍵值：ID + HOSP_ID + IN_DATE。 4. 個人歸戶及個人資料對照的鍵值：ID。 5. 醫療機構歸戶及醫療機構資料對照的鍵值：HOSP_ID。				



全民健保處方及治療明細檔_西醫住院 - 唯一鍵值

二、 檔案格式及資料描述

全民健保處方及治療明細檔—西醫住院 (H_NHI_IPDTE)					
序號	中文欄位名稱	英文欄位名稱	型態	長度	資料描述
1	身分證字號	ID	Char	12	經加密處理。身分證字號或外籍居留證號碼(如無居留證號碼為護照號碼)。
2	醫師身分證字號	PRSN_ID	Char	12	經加密處理，醫師或原處方醫師之身分證字號或外籍居留證號。
3	醫療機構代號	HOSP_ID	Char	9	經加密處理，為辨別醫療機構的鍵值。
4	費用年月	FEE_YM	Char	6	YYYYMM。
5	申報類別	APPL_TYPE	Char	1	1：送核 2：補報 4：申復送核 5：申復補報。
6	申報日期	APPL_DATE	Char	8	經加密處理。
7	案件分類	CASE_TYPE	Char	2	詳案件分類 (CASE_TYPE) 譯碼說明。
8	流水號	SEQ_NO	Char	10	
9	給付類別	GAVE_KIND	Char	1	詳見欄位給付類別(GAVE_KIND) 譯碼說明

pd.read_sas {pandas}

- 下載 https://github.com/rwepa/DataDemo/blob/master/h_nhi_ipdte103.sas7bdat

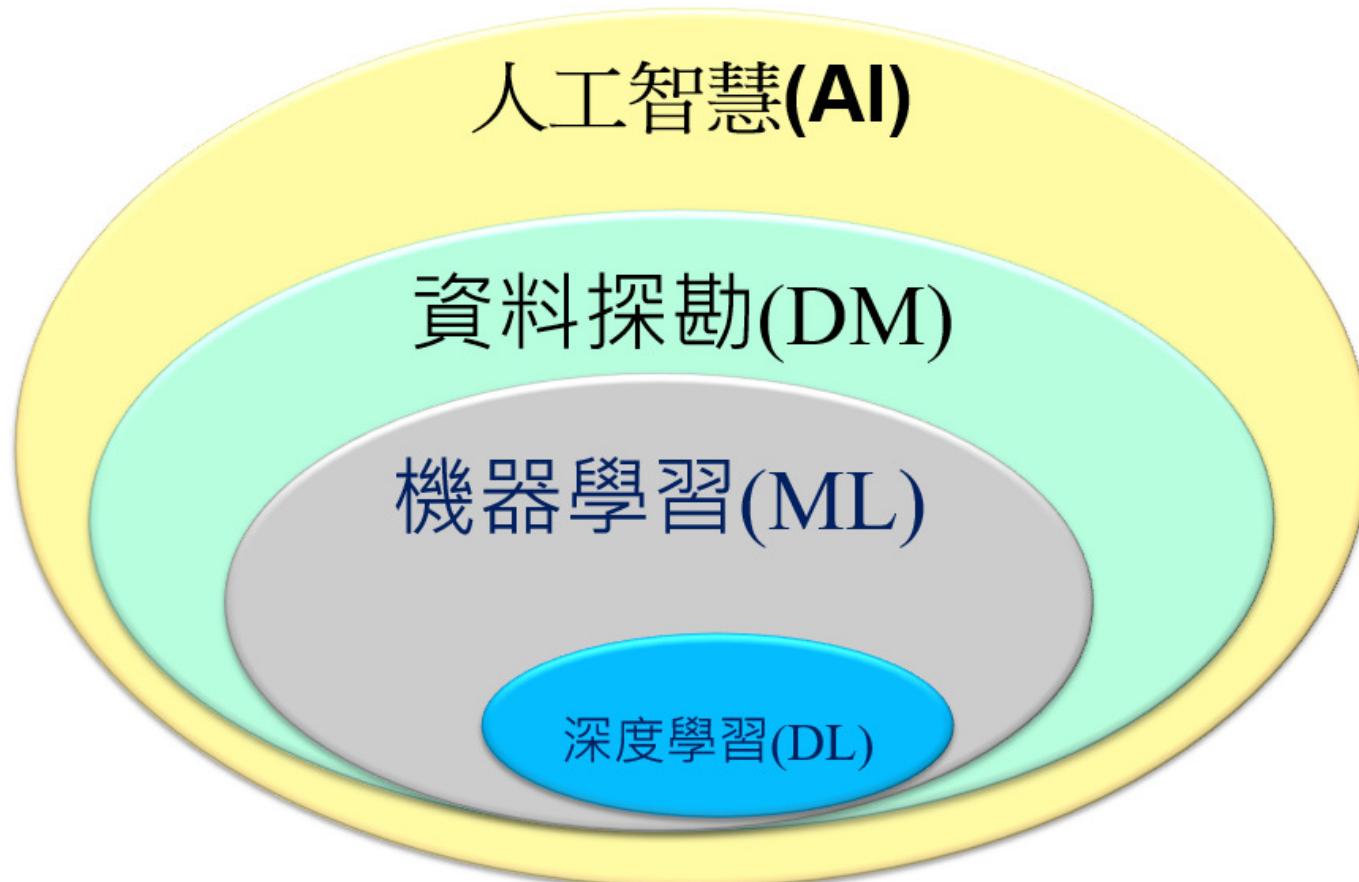
```
In: import pandas as pd  
  
In: urls="C:/rdata/h_nhi_ipdte103.sas7bdat"  
  
In: df=pd.read_sas(urls)  
  
In: df  
Out:  
      ID          PRSN_ID ... CHILD_MARK TW_DRGS_SUIT_MARK  
0    b')])))#*+;[*<' b'%%%**#=_=]~~' ...     NaN      b'1'  
1    b')])))+-^$]$[/' b'~%~@&[>#*^#_ ...     NaN      b'C'  
2    b')])))/_~^(&_ b'*]*>=)&}:+}~' ...     NaN      b'0'  
3    b')])))~$<_&#>' b']]]:+$/_$};: ...     NaN      b'1'  
4    b')])))<@%/-@/\\" b'+%+[ -"-*%]<<' ...     NaN      b'C'  
...  
14292 b'...  
14293 b'...  
14294 b'...  
14295 b'...  
14296 b'(%(("<^#!{# b'%%%+}@<-*=~/' ...     NaN      b'C'  
  
[14297 rows x 80 columns]
```

資料已經去識別化



4.深度學習的發展

人工智慧簡介



人工智慧發展史



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



人工智慧技術

人工智慧技術

- 人工智慧 (artificial intelligence, AI) 亦稱人工智能、機器智慧，指由人類製造出來的機器所表現出來的智慧。
- 通常人工智慧是指透過普通電腦程式來呈現人類智慧的技術。該詞也指出研究這樣的智慧系統是否能夠實現，以及如何實現。同時，透過醫學、神經科學、機器人學及統計學等的進步，常態預測則認為人類的很多職業也逐漸被其取代。

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



AI 第一次潮流 (1/3)

- 第一次潮流在1956-1987，「人工智慧」名詞第一次出現在1956年的美國達特矛斯 (Dartmouth) 會議上。當時由約翰·麥卡錫(John McCarthy)等十位學者出席會議並且秉持著「人類般思考的機器稱為人工智慧」的信念，針對初期的人工智慧程式和各種相關的理論進行討論。
- 成果：
 - 感知機 (深度學習的雛形) 被提出
 - 用機器證明「數學原理」定義，也提出歸結原理
 - 發展出模型識別程序，並編製可分辨積木構造的程序
 - 編制通用的問題解決程序 (General Problem Solver，GPS)
 - 研製成功專家系統 DENDRAL
 - 研發人工智慧語言 (List Processing，LISP) ←

約翰麥卡錫 (John McCarthy) 1967

Professor John McCarthy - Stanford University ~1967

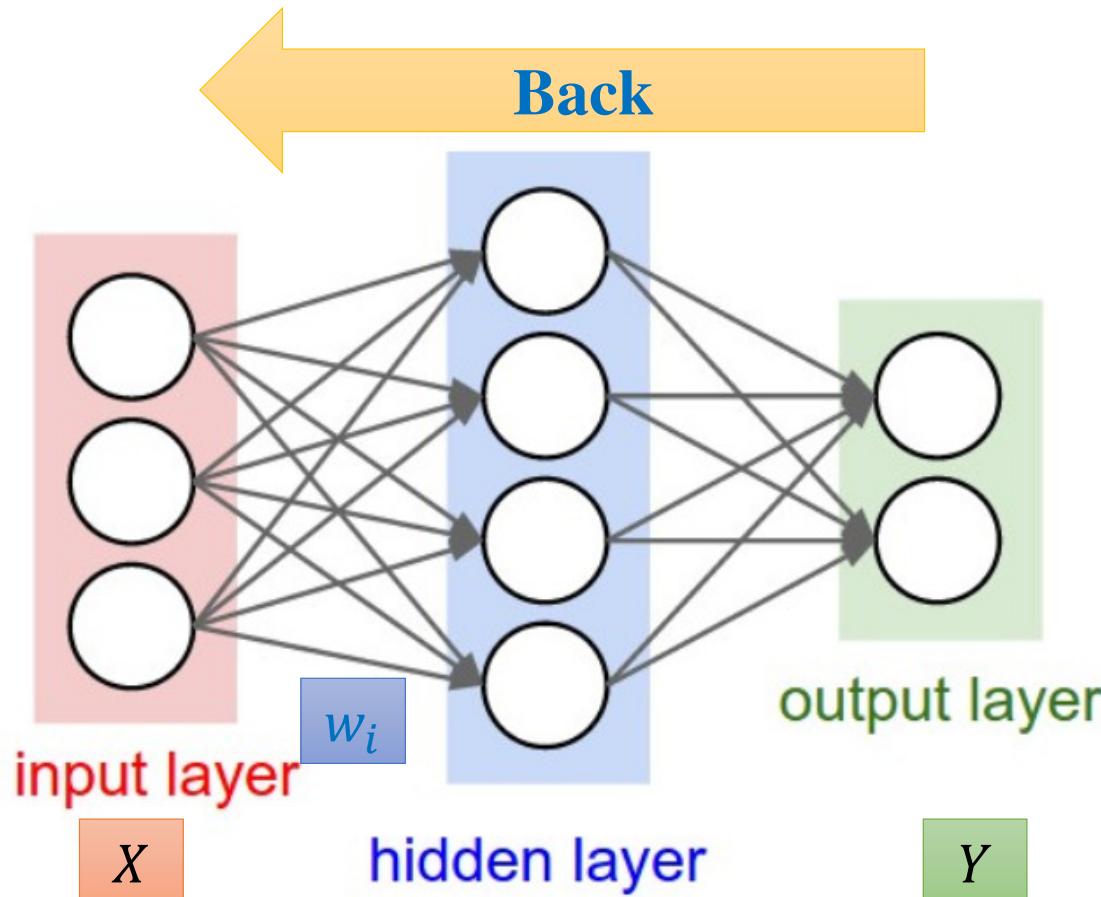


This picture was originally downloaded from exhibits.stanford.edu. The author is unknown. This picture was altered to enhance the clarity of its content and to add comments. This picture is used for research, teaching and education and it is a transformative use. We consider that the use of this picture in this context represents fair use. Please, do not try to take down this picture before considering whether our conduct constitutes fair use!

AI 第二次潮流 (2/3)

- 第二次潮流在1987-1993，以規則庫建立的專家系統。
- 成果：
 - 專家系統的誕生
 - 發現「智慧」需要建立在分類的知識多種處理方式
 - 反向傳播演算法（Backpropagation，BP演算法）實現神經網路訓練
 - 研究人員首次提出：機器為了獲得真正的智慧，它必要有感知、生存、與世界**交互的能力**，對事物的推理能力比抽象能力更重要，這也推動未來自然語言、機器視覺的發展。

Backpropagation，BP演算法



AI 第三次潮流 (3/3)

- 第三次潮流在1993-2022，深度學習概念和急遽發展的時代。
- 成果：
 - 1997年，IBM的計算機系統Deep Blue戰勝國際象棋世界冠軍Garry Kasparov，這也成為重要的里程碑。
 - 2005年，Stanford開發的機器人在沙漠上自動行駛210公里，贏得DARPA挑戰大賽頭獎。
 - 2006年，Geoffrey Hinton提出多層神經網絡的深度學習算法、Eric Schmidt在搜索引擎大會提出「雲端計算」概念。
 - 2014年，微軟亞洲研究院發佈人工智慧聊天機器人和語音助手Cortana。
 - 2016年，Google公司的人工智慧程式「AlphaGO」，與韓國棋手李世乭在圍棋上正面交鋒。AlphaGo 所使用的深度學習技術引起全球關注。
 - 2017年，AlphaGo團隊提出AlphaGo Zero 以自我學習方式，在中國烏鎮圍棋峰會，挑戰排名第一的圍棋冠軍柯潔，以3比0獲勝。

AlphaGo Zero



參考: <https://qbi.uq.edu.au/blog/2017/10/google-alphago-zero-masters-game-three-days>



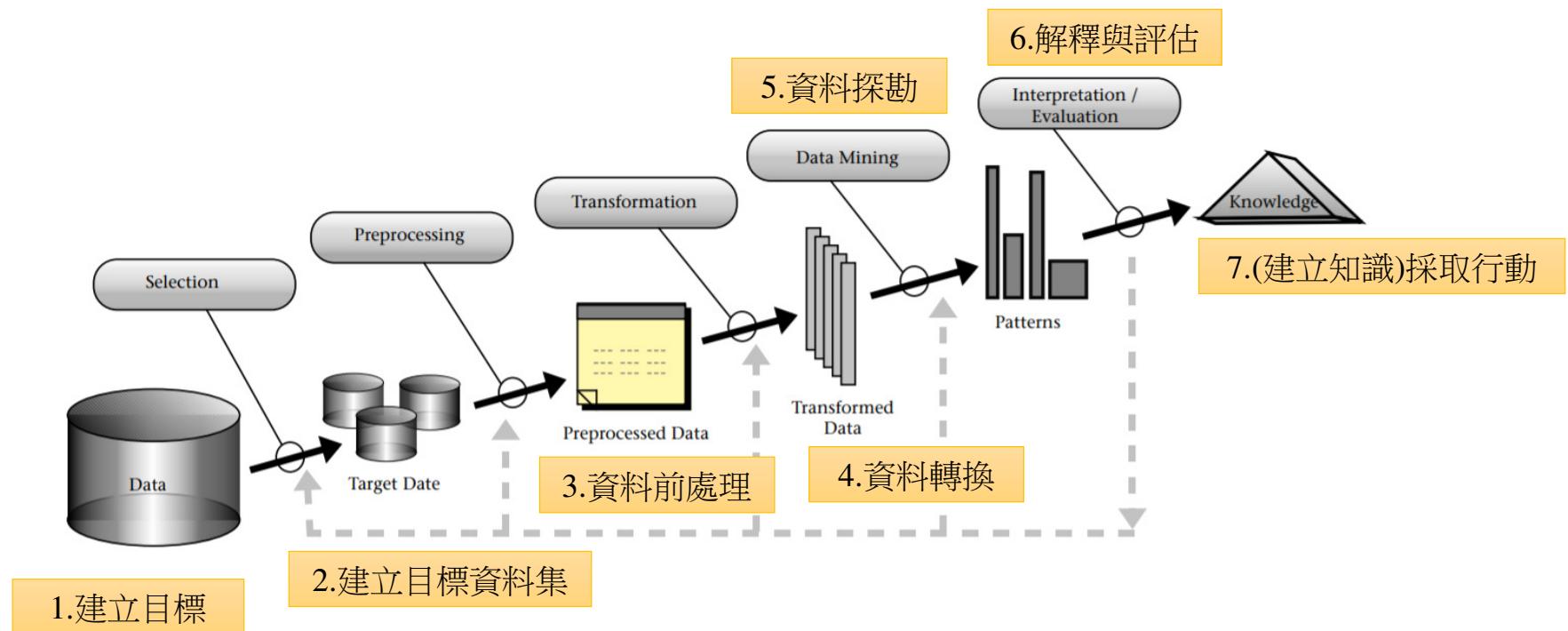
機器學習

KDD → 機器學習

資料庫中的知識發掘

(Knowledge Discovery in Database, KDD, 1996)

- Usama Fayyad, Gregory Piatetsky-Shapiro, and Padhraic Smyth , *Knowledge Discovery and Data Mining: Towards a Unifying Framework*, KDD'96: Proceedings of the Second International Conference on Knowledge Discovery and Data Mining, August 1996 Pages 82–88.
<https://www.kdnuggets.com/gpspubs/aimag-kdd-overview-1996-Fayyad.pdf>



機器學習 Machine learning

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



監督式學習 vs. 非監督式學習

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



機器學習方法論 CRISP-DM 簡介

資料探勘生命週期 - CRISP-DM

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

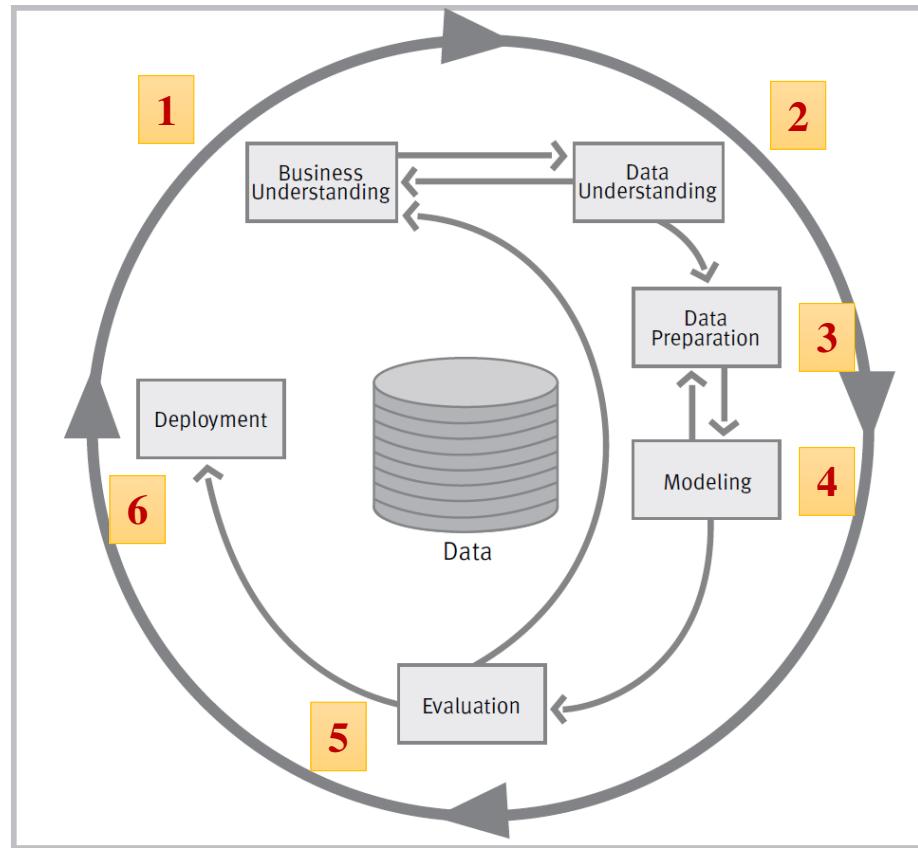
CRISP-DM 資料探勘流程(續)

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

佔整專案時間的~80%

- 訓練資料70%
- 測試資料30%

CRISP-DM 資料探勘流程(續)



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

數值模型績效指標

- 不可直接使用誤差的算術平均!

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

- 均方誤差 (Mean Squared Error, MSE)

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

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

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

- 平均絕對誤差 (Mean Absolute Error, MAE)

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

類別模型績效指標 - 混淆矩陣

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

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

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



深度學習

卷積神經網路

- 卷積神經網路 (Convolutional Neural Network, CNN) 是一種前饋式(feed-forward)神經網路，它的人工神經元對於大型圖像處理有出色表現。
- 卷積神經網絡由輸入層、隱藏層和輸出層組成。
- 卷積神經網路由一個或多個卷積層和頂端的全連通層（對應經典的神經網路）組成，同時也包括關聯權重和池化層 (pooling layer) 。

卷積神經網路 (續)

- 在前饋神經網絡中，任何中間層都被稱為隱藏層，因為它們的輸入和輸出被激活函數和最終卷積屏蔽。
- 在卷積神經網絡中，隱藏層包括執行卷積的層。
- 這一結構使得卷積神經網路能夠利用輸入資料的二維結構。與其他深度學習結構相比，卷積神經網路在圖像和語音辨識方面能夠給出更好的結果。這一模型也可以使用反向傳播演算法進行訓練。相比較其他深度、前饋神經網路，卷積神經網路需要考量的參數更少，使之成為一種頗具吸引力的深度學習結構。

卷積 (滑動+內積)

1	1	1	0	0
0	1	1	1	0
0	0	1	1	1
0	0	1	1	0
0	1	1	0	0

Image

1	0	1
0	1	0
1	0	1

3x3矩陣

$$\begin{aligned}1 \times 1 + 1 \times 0 + 1 \times 1 &= 1 + 0 + 1 = 2 \\0 \times 0 + 1 \times 1 + 1 \times 0 &= 0 + 1 + 0 = 1 \\0 \times 1 + 0 \times 0 + 1 \times 1 &= 0 + 0 + 1 = 1 \\2 + 1 + 1 &= 4\end{aligned}$$

卷積 (續)

1 <small>x1</small>	1 <small>x0</small>	1 <small>x1</small>	0	0
0 <small>x0</small>	1 <small>x1</small>	1 <small>x0</small>	1	0
0 <small>x1</small>	0 <small>x0</small>	1 <small>x1</small>	1	1
0	0	1	1	0
0	1	1	0	0

Image

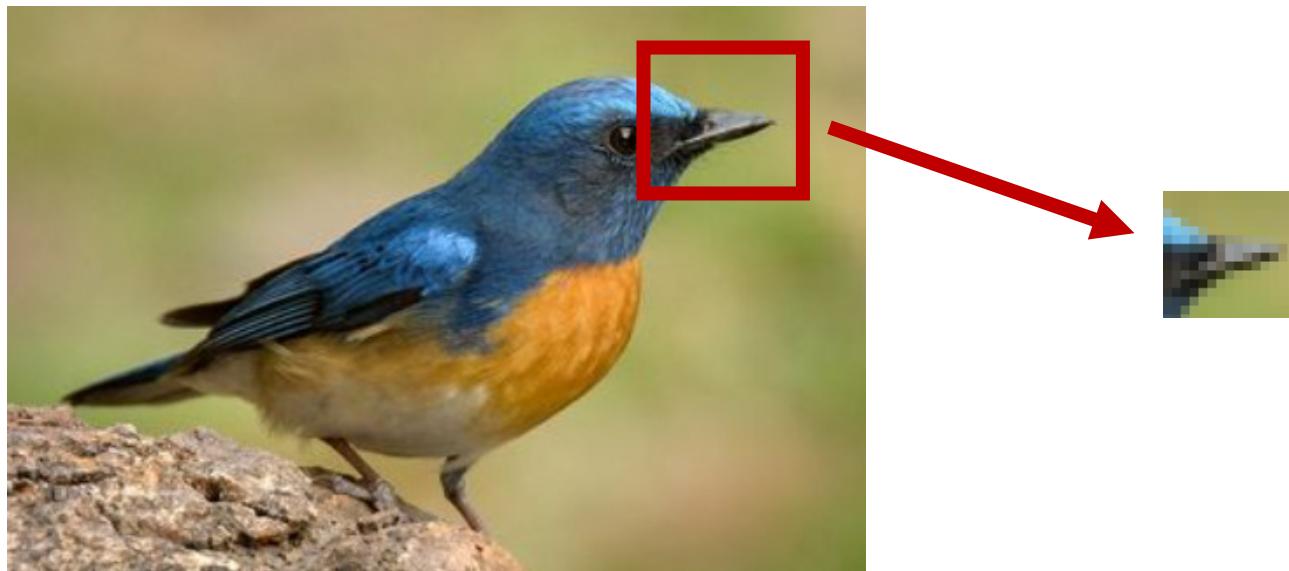
4		

Convolved
Feature

目標是找出最佳的
Convolved Feature
(卷積特徵)- filter

卷積層的特色

- 保留圖片中的空間結構，並從此結構中萃取出特徵



池化層 (Pooling Layer)

- 池化層的主要概念是，當我們在做圖片的特徵萃取的過程中，圖形的縮放應該不會影響到我們的目的，經由這樣的 scaling 我們也可以再一次的減少神經網路的參數。

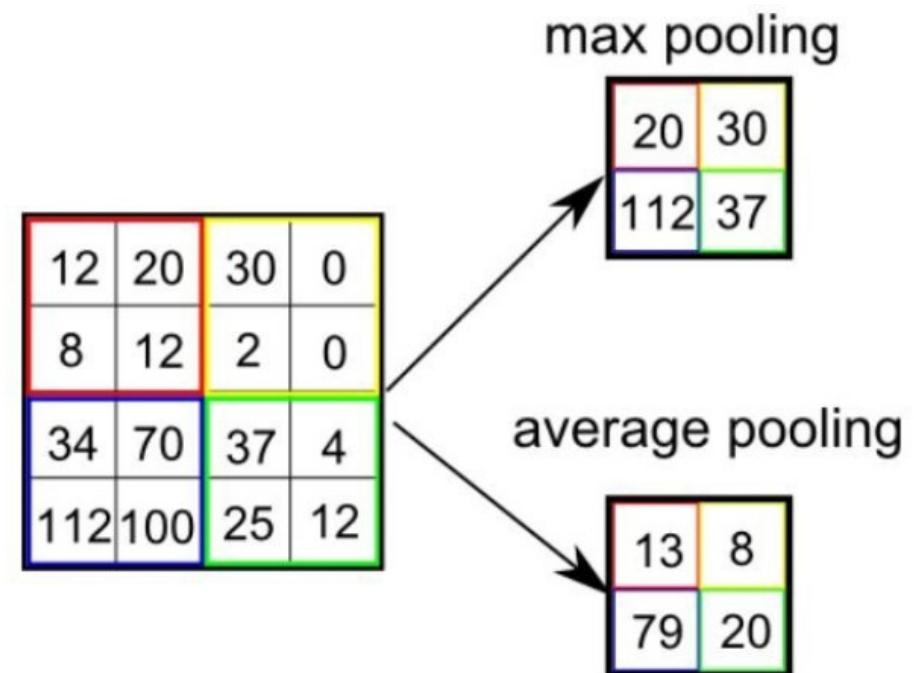


subsampling



Pooling Layer 池化層

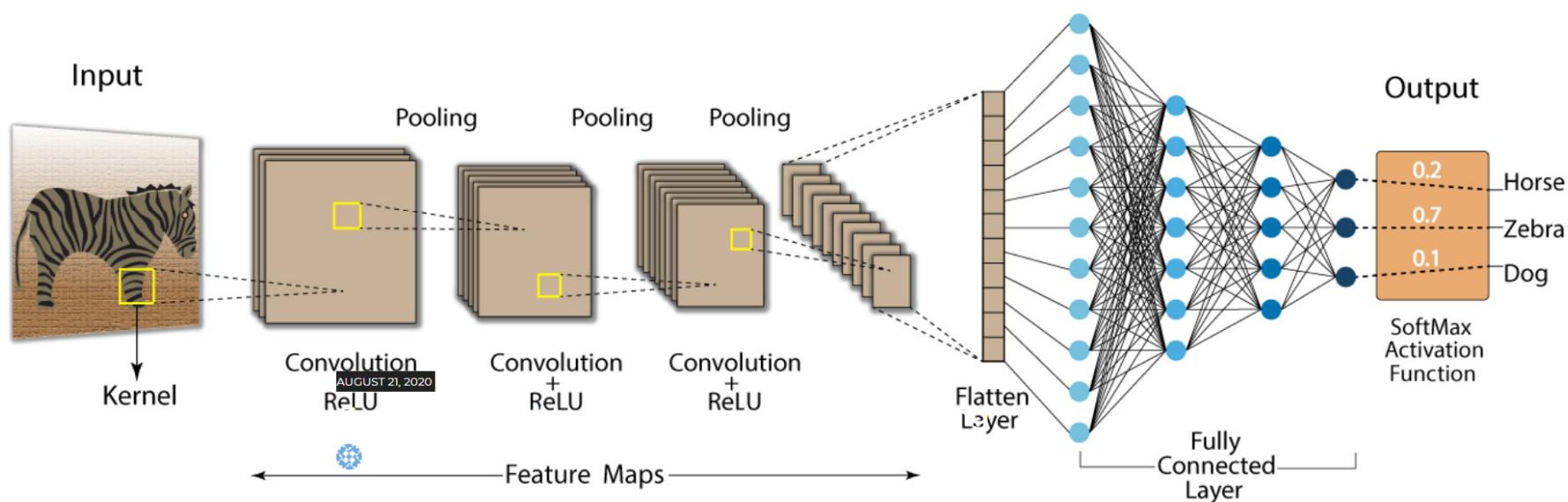
- 圖片的特徵萃取的過程中，圖形的縮放應該不會影響到我們的目的，經由這樣的 scaling 我們也可以減少神經網路的參數。
- 常用的 pooling 方式有 Max pooling 與 Average pooling
- 目前主要使用 Max pooling



全連接層

- 這邊的全連接層跟我們進行手寫辨識的方式一樣，說穿了就是一個分類器，把我們經過數個卷積、池化後的結果進行分類。

Convolution Neural Network (CNN)



<https://developersbreach.com/convolution-neural-network-deep-learning/>

深度學習CNN - MNIST範例

- 參考 <https://github.com/rwepa/DataDemo/blob/master/iPAS-python-program.py#L2150>

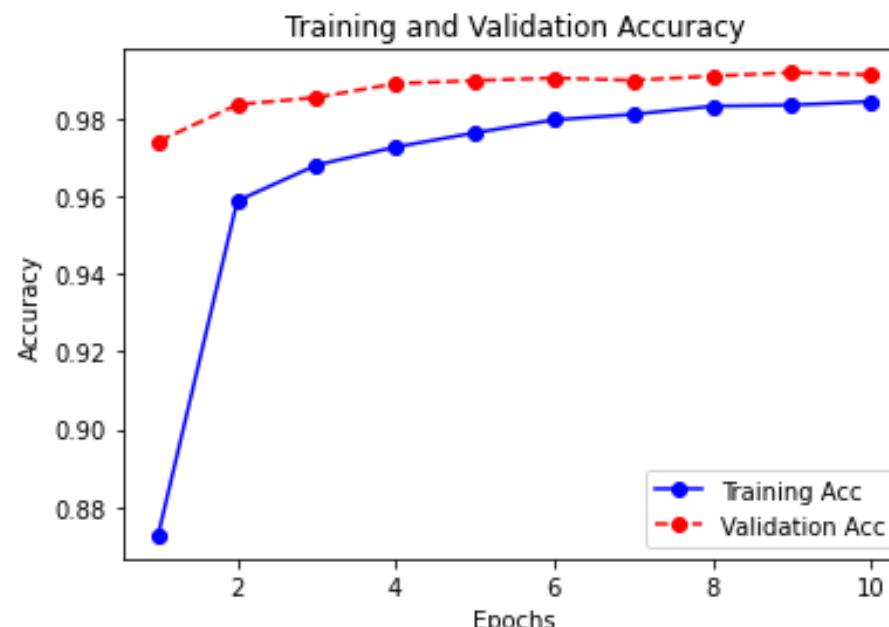
Testing ...

1875/1875 [=====] - 14s 7ms/step - loss: 0.0184 - accuracy: 0.9943

訓練資料集的準確度 = 0.99

313/313 [=====] - 3s 8ms/step - loss: 0.0237 - accuracy: 0.9919

測試資料集的準確度 = 0.99

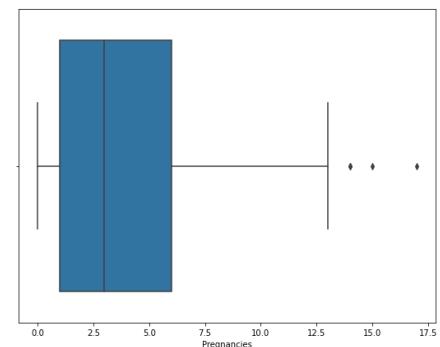
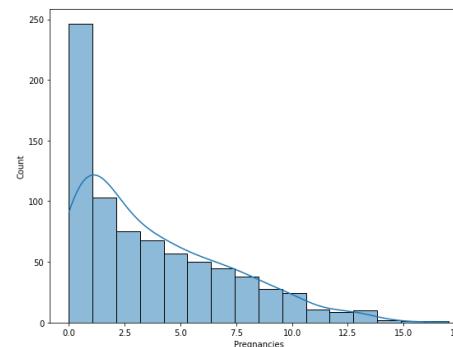




5.皮馬印第安人糖尿病預測分析

皮馬印第安人糖尿病預測分析

- <https://github.com/rwepa/DataDemo/blob/master/diabetes.csv>
- <https://github.com/rwepa/DataDemo/blob/master/iPAS-python-program.py#L2774>
- 資料筆數: 768
- 欄位個數: 9
- 欄位名稱:
 1. Pregnancies 懷孕次數
 2. Glucose 血漿葡萄糖濃度
 3. BloodPressure 舒張壓
 4. SkinThickness 三頭肌皮脂厚度
 5. Insulin 胰島素濃度
 6. BMI 身體質量指數
 7. DiabetesPedigreeFunction 糖尿病函數,依家族糖尿病史而計算個人患有糖尿病的風險值
 8. Age 年齡
 9. Outcome {1:有糖尿病, 0:無糖尿病} --> 反應變數



Python 模組

模組	功能	
Numpy	大型多維度陣列,矩陣	
Scipy	最佳化,線性代數,積分,內插法,FFT,信號和圖像處理	
Pandas	使用 DataFrame 進行資料分析,繪圖	
Matplotlib	靜態, 動畫圖形視覺化	
Statsmodels	統計模型	
Scikit-learn	機器學習	
Tensorflow	深度學習	
Biopython	生物計算	
Scanpy	單細胞分析	



參考資料

- RWEPA
 - <http://rwepa.blogspot.com/>
- Python免費電子書PDF
 - <http://rwepa.blogspot.com/2020/02/pythonprogramminglee.html>
- R與實驗設計應用系列
 - <https://youtu.be/IJupX5c3hi4>
- 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>

反應曲面法模型

- 一階模型 (First-order model)
 $y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_k x_k + \varepsilon$
- 二階模型 (Second-order model)
 $y = \beta_0 + \sum_{i=1}^k \beta_i x_i + \sum_{i=1}^k \sum_{j=1, j \neq i}^k \beta_{ij} x_i x_j + \varepsilon$

謝謝您的聆聽

Q & A



李明昌

alan9956@gmail.com

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