

# 商業大數據分析應用與職場技能

## 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

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

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

[alan9956@gmail.com](mailto:alan9956@gmail.com)

# 大綱

- 1.商業大數據剖析、解譯與洞察力
- 2.大數據軟體實務應用
- 3.職場技能
- Q&A

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

- 姓名：李明昌 (ALAN LEE)
- 現職：中華R軟體學會 理事  
臺灣資料科學與商業應用協會 理事
- 學歷：中原大學 工業與系統工程所 博士
- 經歷：
  - 淡江大學 兼任教師
  - 育達科技大學 兼任教師
  - 佛光大學 兼任教師
  - 國立台北商業大學 兼任教師
  - 東吳大學 兼任教師
  - 育達科技大學 資訊管理系(所) 專任助理教授
  - 崇友實業 行銷企劃專員
  - 國航船務代理股份有限公司 海運市場運籌管理員
- 大專院校、資策會、工業技術研究院、國家發展委員會、中央氣象局、公平交易委員會、衛生福利部、縣市政府與日本名古屋產業大學等公民營單位演講達347場，3211小時。
- 連絡資訊：[alan9956@gmail.com](mailto: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

# 1. 商業大數據剖析、解譯與洞察力

# 認識商業大數據

---

# 資料類型

- 結構化資料 (Structured data)

- 量化資料
- 資料模型已經事先定義，彈性較少
- 資料格式較有限制
- 較適用於進行資料分析 (例: Machine Learning)
- 有限的儲存方式，例: 關聯式資料庫

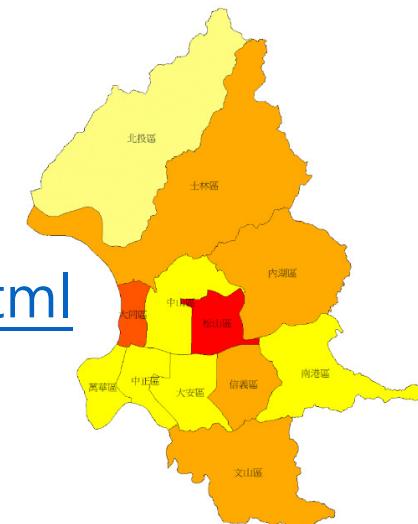
- 非結構化資料 (Unstructured data)

- 非量化資料 (網頁, 照片, 語音)
- 資料模型無事先定義，彈性較大
- 資料格式較無限制
- 較難直接使用現有套裝軟體分析(例:自然語言處理, NLP)
- 可以使用 NoSQL 儲存 (例: MongoDB, Hadoop)

- 半結構化資料 (semi-structured data)
  - 範例: JSON, XML, CSV
  - 使用元資料(metadata)註記資料特性

## 資料類型 (續)

- **時間序列資料**: 按照時間發生先後順序進行排列的資料序列。
  - 趨勢
  - 季節性/週期性
  - 殘差
- **地理資料**: 經度、緯度、行政區域、國家。
  - 主題式地圖 (Thematic map)
  - <https://rwepa.blogspot.com/2018/10/thematicmap.html>



# 資料應用

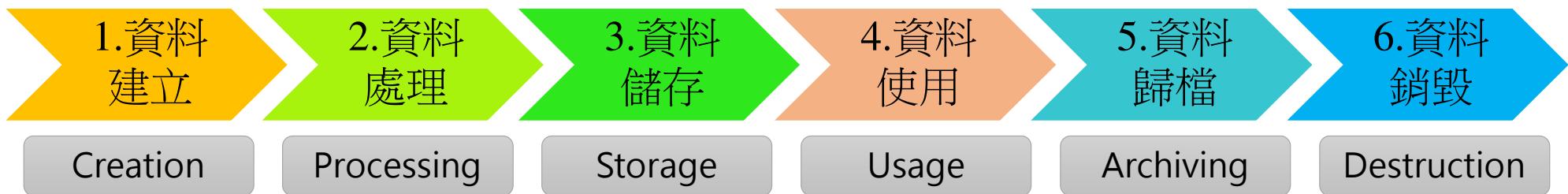
- 更深層了解客戶需求，提供更佳客戶服務 (訂定 KPI)
- 建立更有客製化的行銷活動
- 追蹤當前指標並建立新指標
- 創造更佳的產品機會和服務
- 增加營運收入
- 降低營運成本
- 創造資料附加價值



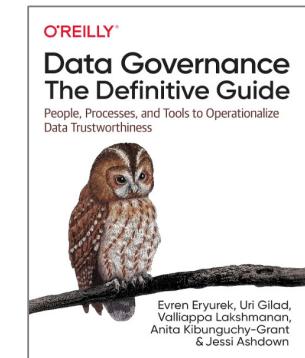
Value

# 資料生命週期管理

- 資料生命週期管理 (Data Lifecycle Management, DLM)。
- 從資料的建立到資料的銷毀, DLM包括六大步驟。
- 人員、流程、技術三大層面→ 資料治理 (Data Governance)。



Eryurek, E., Gilad, U., Lakshmanan, V., Kibunguchy-Grant, A., & Ashdown, J., *Data Governance: The Definitive Guide: People, Processes, and Tools to Operationalize Data Trustworthiness*, O'Reilly Media, 2021. [[LINK](#)]



# 1. 資料的建立

- 結構化、非結構化、半結構化資料
- 資料產生的頻率: 批次、串流
- 資料與元資料
- 資料歷程 (Lineage)
- 資料建立來源(第三方、企業內部輸入)
- 資料捕獲(POS、生產線機台、物聯網感測器)

## 2. 資料的處理

- 資料整合 (Data Integration)
  - 整合不同資料來源為單一標準資料。
- 資料清理 (Data Cleaning)
  - 辨識不正確、不完整、不相關、重複值、或格式錯誤資料，然後進行取代、修改或刪除髒資料。
- ETL運用 (Extract, Transform, Load)
  - ETL表示將資料從來源端經過抽取 ( extract ) 、轉置 ( transform ) 、載入 ( load ) 至目的端的過程。
- 資料歷程 (Lineage)
- 資料的分類(敏感性，重要性)
- 資料品質
- 手動、自動化、半自動資料的處理

### 3. 資料的儲存

- 具有保護級別的儲存設備
- 加密
- 備份
- 災難復原計畫 (Disaster Recovery Plan)

## 4. 資料的使用

- 支援企業目標與營運
- 呈現資料的價值
- 篩選、報表、視覺化、儀表板、自動生成 Excel, Word, HTML 等
- 商業智慧 (Business Intelligence, BI)
- 營運智慧 (Operational intelligence, OI)

## 5. 資料的歸檔

- 考慮不再使用、不再發佈更新的資料，最後將資料加以儲存。
- 已歸檔的資料較不用明確訂定復原流程與復原計畫。
- 資料保留策略，儲存時間長度。

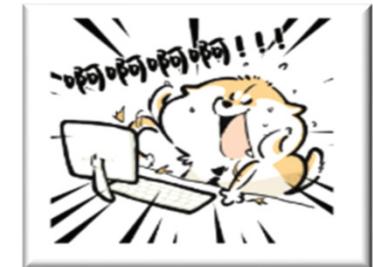
## 6. 資料的銷毀

- 刪除資料與相關副本
- 不考慮永久保存資料
- 符合相關法規
  - 金融機構防制洗錢辦法: 金融機構對國內外交易之所有必要紀錄，應至少保存五年。但法律另有較長保存期間規定者，從其規定。

# 空氣品質小時值\_新北市\_淡水站

- <https://data.gov.tw/dataset/34759>

	A	B	C	D	E	F	G	H	I
1	siteid	sitename	county	itemid	itemname	itemengname	itemunit	monitordate	concentration
2	10	淡水	新北市	33	細懸浮微粒	PM2.5	$\mu\text{g}/\text{m}^3$	2024/6/22 08:00	7
3	10	淡水	新北市	7	二氧化氮	NO2	ppb	2024/6/22 08:00	6.7
4	10	淡水	新北市	4	懸浮微粒	PM10	$\mu\text{g}/\text{m}^3$	2024/6/22 08:00	22
5	10	淡水	新北市	3	臭氧	O3	ppb	2024/6/22 08:00	30.5
6	10	淡水	新北市	2	一氧化碳	CO	ppm	2024/6/22 08:00	0.2
7	10	淡水	新北市	1	二氧化硫	SO2	ppb	2024/6/22 08:00	1.1
8	10	淡水	新北市	33	細懸浮微粒	PM2.5	$\mu\text{g}/\text{m}^3$	2024/6/22 07:00	9
9	10	淡水	新北市	7	二氧化氮	NO2	ppb	2024/6/22 07:00	8.9
10	10	淡水	新北市	4	懸浮微粒	PM10	$\mu\text{g}/\text{m}^3$	2024/6/22 07:00	28
11	10	淡水	新北市	3	臭氧	O3	ppb	2024/6/22 07:00	16.5
12	10	淡水	新北市	2	一氧化碳	CO	ppm	2024/6/22 07:00	0.28
13	10	淡水	新北市	1	二氧化硫	SO2	ppb	2024/6/22 07:00	0.6
14	10	淡水	新北市	33	細懸浮微粒	PM2.5	$\mu\text{g}/\text{m}^3$	2024/6/22 06:00	14
15	10	淡水	新北市	7	二氧化氮	NO2	ppb	2024/6/22 06:00	9.6
16	10	淡水	新北市	4	懸浮微粒	PM10	$\mu\text{g}/\text{m}^3$	2024/6/22 06:00	27
17	10	淡水	新北市	3	臭氧	O3	ppb	2024/6/22 06:00	6.5
18	10	淡水	新北市	2	一氧化碳	CO	ppm	2024/6/22 06:00	0.37
19	10	淡水	新北市	1	二氧化硫	SO2	ppb	2024/6/22 06:00	0.6
20	10	淡水	新北市	33	細懸浮微粒	PM2.5	$\mu\text{g}/\text{m}^3$	2024/6/22 05:00	7



1. 不同量測變數在同一行！
2. 日期順序怪怪的！

# 溫度資料 (固定寬度)

1	stn_id	yyyymmdd	01	02	03	04	05	06	07
2	72C440	19980101	17.9	17.9	17.9	18.0	17.7	17.8	17.7
3	72C440	19980102	16.5	16.7	16.7	16.5	16.5	16.7	16.7
4	72C440	19980103	18.4	18.4	16.9	16.0	16.0	16.1	16.4
5	72C440	19980104	18.3	17.2	16.8	16.4	16.2	16.3	18.0
6	72C440	19980105	11.9	11.8	12.1	12.0	12.1	12.3	12.8
7	72C440	19980106	14.9	15.1	14.9	14.7	14.7	14.8	14.8
8	72C440	19980107	17.0	16.8	17.0	17.1	17.7	18.0	17.9
9	72C440	19980108	17.4	17.3	17.3	17.4	17.4	18.0	17.9
10	72C440	19980109	17.8	17.1	16.7	16.5	16.4	16.3	16.7
11	72C440	19980110	16.9	17.0	18.0	18.4	18.3	18.1	18.2
12	72C440	19980111	20.2	19.5	19.8	19.6	20.1	20.6	20.5

# 離子資料 (標題列有2行)

	A	B	C	D	E	F	G	H
1	ELEMENT	H	ELEMENT	C	ELEMENT	O	ELEMENT	Mg
2	[nm]	[Atom/cm3]	[nm]	[Atom/cm3]	[nm]	[Atom/cm3]	[nm]	[Atom/cm3]
3	1.38E+00	1.52E+20	4.77E+00	7.06E+18	3.89E+00	1.75E+19	1.60E+00	2.27E+20
4	7.01E+00	6.77E+18	1.10E+01	2.08E+18	1.01E+01	1.16E+18	4.80E+00	1.78E+20
5	1.25E+01	3.02E+18	1.67E+01	1.28E+18	1.61E+01	5.12E+17	7.99E+00	1.43E+20
6	1.75E+01	2.36E+18	2.24E+01	7.56E+17	2.19E+01	2.91E+17	1.12E+01	8.85E+19
7	2.25E+01	1.98E+18	2.81E+01	5.81E+17	2.77E+01	1.98E+17	1.44E+01	4.86E+19
8	2.75E+01	1.84E+18	3.38E+01	3.78E+17	3.35E+01	1.53E+17	1.76E+01	3.16E+19
9	3.25E+01	2.27E+18	3.95E+01	3.15E+17	3.93E+01	1.27E+17	2.08E+01	2.37E+19
10	3.75E+01	1.79E+18	4.52E+01	2.71E+17	4.50E+01	1.12E+17	2.40E+01	2.06E+19
11	4.26E+01	2.11E+18	5.09E+01	2.34E+17	5.08E+01	1.02E+17	2.72E+01	1.83E+19
12	4.76E+01	2.15E+18	5.66E+01	2.06E+17	5.66E+01	9.83E+16	3.04E+01	1.65E+19
13	5.26E+01	2.21E+18	6.23E+01	1.80E+17	6.24E+01	9.11E+16	3.36E+01	1.53E+19
14	5.76E+01	2.16E+18	6.80E+01	1.55E+17	6.81E+01	8.74E+16	3.68E+01	1.52E+19

# 空氣品質指標(AQI)(歷史資料) - XML

- <https://data.gov.tw/dataset/151824>

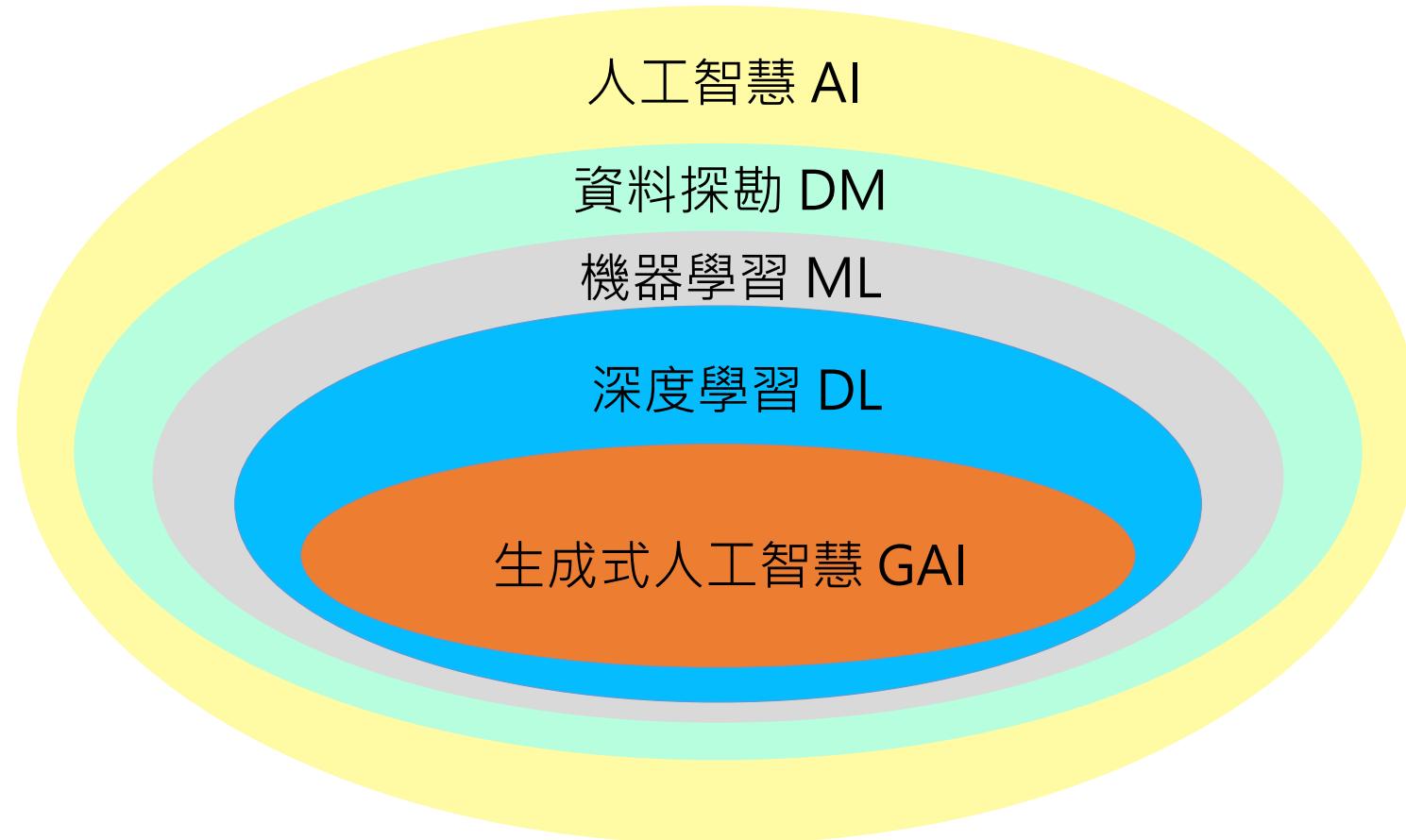
```
aqx_p_488.xml
1 <?xml version="1.0" encoding="UTF-8"?>
2 <aqx_p_488><data><sitename>屏東(枋山)</sitename><county>屏東縣</county><aqi>26
</aqi><pollutant></pollutant><status>良好</status><so2>0.1</so2><co>0.11</co><o3>32.8</o3><o3_8hr>
28.0</o3_8hr><pm10>5</pm10><pm2.5>1</pm2.5><no2>3.1</no2><nox>3.6</nox><no>0.4</no><windspeed>1
</windspeed><winddirec>85</winddirec><datacreationdate>2023-06-12 10:00
</datacreationdate><unit></unit><co_8hr>0.1</co_8hr><pm2.5_avg>1</pm2.5_avg><pm10_avg>4
</pm10_avg><so2_avg>0</so2_avg><longitude>120.651472</longitude><latitude>22.260899
</latitude><siteid>313</siteid></data><data><sitename>新北(樹林)</sitename><county>新北市
</county><aqi>38</aqi><pollutant></pollutant><status>良好</status><so2>0.3</so2><co>0.45</co><o3>
13.4</o3><o3_8hr>6.9</o3_8hr><pm10>18</pm10><pm2.5>6</pm2.5><no2>23.2</no2><nox>35.9</nox><no>12.6
</no><windspeed>0</windspeed><winddirec>152</winddirec><datacreationdate>2023-06-12 10:00
</datacreationdate><unit></unit><co_8hr>0.3</co_8hr><pm2.5_avg>7</pm2.5_avg><pm10_avg>24
</pm10_avg><so2_avg>0</so2_avg><longitude>121.38352778</longitude><latitude>24.94902778
</latitude><siteid>311</siteid></data><data><sitename>桃園(三民)</sitename><county>桃園市
</county><aqi>25</aqi><pollutant></pollutant><status>良好</status><so2>0.5</so2><co>0.44</co><o3>
22.2</o3><o3_8hr>15.2</o3_8hr><pm10>16</pm10><pm2.5>9</pm2.5><no2>15.5</no2><nox>25.2</nox><no>9.6
</no><windspeed>1</windspeed><winddirec>77</winddirec><datacreationdate>2023-06-12 10:00
</datacreationdate><unit></unit><co_8hr>0.5</co_8hr><pm2.5_avg>9</pm2.5_avg><pm10_avg>24
</pm10_avg><so2_avg>0</so2_avg><longitude>121.38352778</longitude><latitude>24.94902778
</latitude><siteid>311</siteid></data>
```

# 臺灣地區人工魚礁區範圍 - JSON

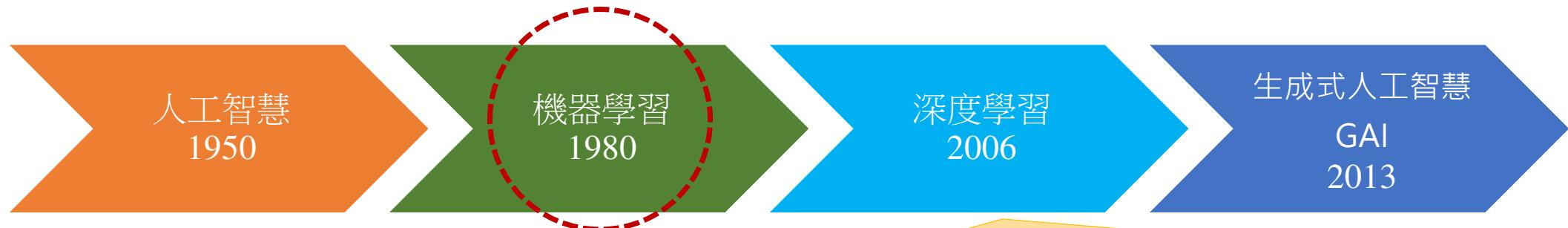
- <https://data.gov.tw/dataset/49099>

```
[  
  {  
    "field001": "基隆市",  
    "field002": "望海巷人工魚礁禁漁區",  
    "field003": "N25°08'36\"\",E121°48'18\"",  
    "field004": "26M",  
    "field005": "88/12/27",  
    "field006": "半徑0.5浬範圍內",  
    "field007": "WGS-84"  
  },  
  {  
    "field001": "基隆市",  
    "field002": "大武崙人工魚礁禁漁區",  
    "field003": "N25°10'45\"\",E121°42'40\"",  
    "field004": "20M",  
    "field005": "88/12/27",  
    "field006": "半徑0.5浬範圍內",  
    "field007": "WGS-84"  
  },  
]
```

# 人工智慧發展



# 人工智能發展



- 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 成為重要運算硬體(NVDA, \$138.99/股)。
- 2013年：變分自編碼器 (Variational Autoencoder , VAE), GAN, LSTM, RNN, Diffusion, Transformers (GPT)

# 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)
- 生成式AI (GAI)



# 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 (PCA)

## • 監督式學習 Supervised learning (執行 X-> 預測 -> 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: 使用多種演算法來獲得更佳預測結果

# 資料剖析

---

# 資料剖析 (Data Analytics)

- 資料剖析是使用工具和流程來整合和分析企業資料集。
- 資料剖析工具可以結合多個異質來源資料集。
- 資料剖析的成果是回答特定問題、發現新的問題與發現新洞察力。
- 資料剖析幫助企業做出更好的資料驅動決策。
- 技巧
  - 關係：正相關、負相關、非線性關係
  - 趨勢：上升、下降
  - 週期：日、週、月、季、年
  - 異常：品質管理、管制界限
  - 樣式(Patterns)：綜合上述所有技巧

# 資料剖析四大類型

## • 描述性分析 (Descriptive analytics)

- 使用歷史資料分析，目的是理解過去的事件。
- 資料剖析需要編譯資料、識別模式以及深入了解過去的事件或從過去的事件中學習隱藏在資料背後的潛在知識。

## • 診斷分析 (Diagnostic analytics)

- 診斷分析深入研究數據以確定特定事件背後的原因。
- 涉及識別根本原因、確定相關性並揭示不同變數之間的關係。

## • 預測分析 (Predictive analytics)

- 預測分析使用統計演算法和歷史資料來預測未來的模式或事件。
- 目標是預測未來事件並評估各種場景的機率。
- 預測分析使用線性迴歸或邏輯斯迴歸模型等演算法。

## • 規範性分析 (Prescriptive analytics)

- 規範性分析不僅可以預測未來結果，還可以推薦具體的決策行動。
- 使用最佳化和模擬技術來確定實現預期結果的最佳行動方案。

# 資料解譯

---

# 資料解譯(Data Interpretation)

- 資料視覺化: Power BI, Tableau, Python, R, SPSS, SAS, MATLAB
- Excel
  - MID, VLOOKUP函數, 樞紐分析
  - 迴歸: [https://youtu.be/i5\\_urp8XzEs](https://youtu.be/i5_urp8XzEs)
  - 下拉式選單同步視覺化: <https://youtu.be/OVA4dvkrsBM>
- 直方圖、長條圖、統計檢定 (t-test)、變異數分析 (ANOVA)
- 分類 (Classification)
- 預測 (Prediction)
- 集群 (Clustering)
- 關聯規則 (Association Rules)
- 類神經網路 (Neural Network, NN) ...

0.0375 解譯

# 洞察力

---

# 洞察力 (Insights)

- 洞察力(見解)表示對於人、事或事物的深刻理解。
- 在資料和分析的背景下，洞察力表示資料分析人員或業務使用者發現以下二項事物：
  - 新的資料模式
  - 以前不知道的變數之間關係
- 洞察力須轉換為**決策執行**。
- 依據執行結果進行**優化處理或改善**。

# 機器學習方法論 CRISP-DM 簡介

---

# 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

佔整專案時間的~80%

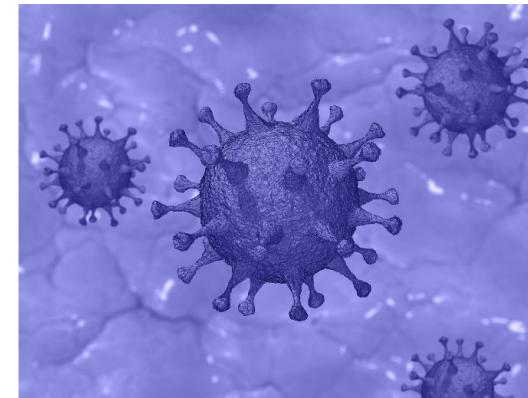
- 步驟 4：模式建立(分析、視覺化、建立模型)
- 步驟 5：評估與測試
- 步驟 6：佈署應用

步驟3 資料準備包括將資料隨機區分為：

- 訓練資料 70%
- 測試資料 30%

# 1. 商業理解

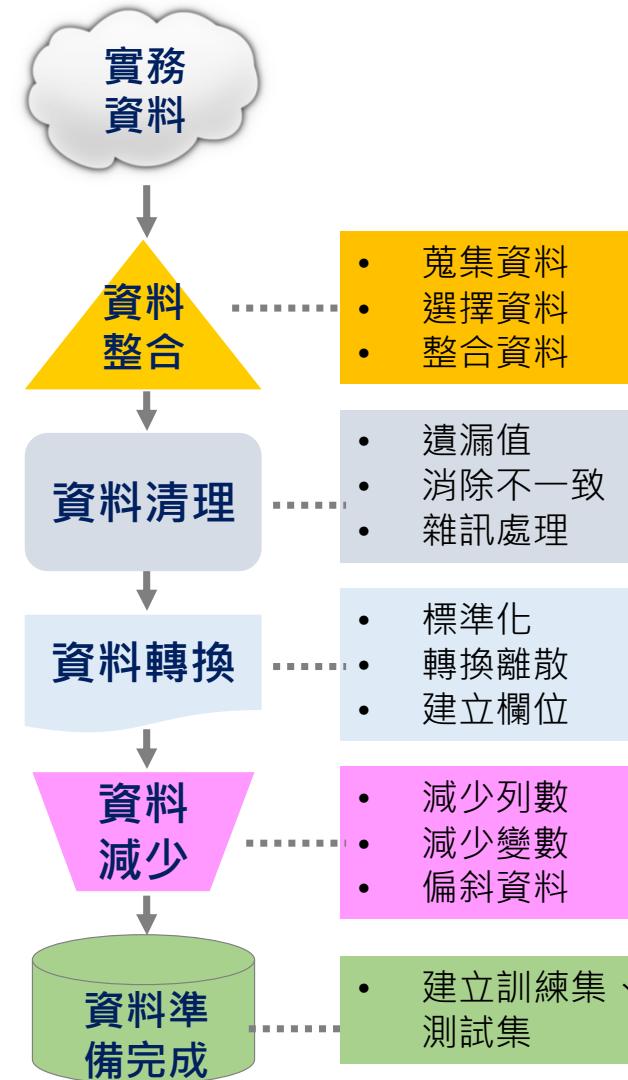
- 終極目標是要解決具體的企業問題，諸如提高購買率、找出詐欺交易、銷售預測與異常偵測等。以專業知識 (domain knowledge)進行商業理解是重要的第一步，處理重點：
  - 擬定商業目標
  - 進行當前處境評估
  - 決定資料探勘目標/成本
  - 產生專案計劃
  - 解決顧客問題
  - 思考獲利,成本,保養服務



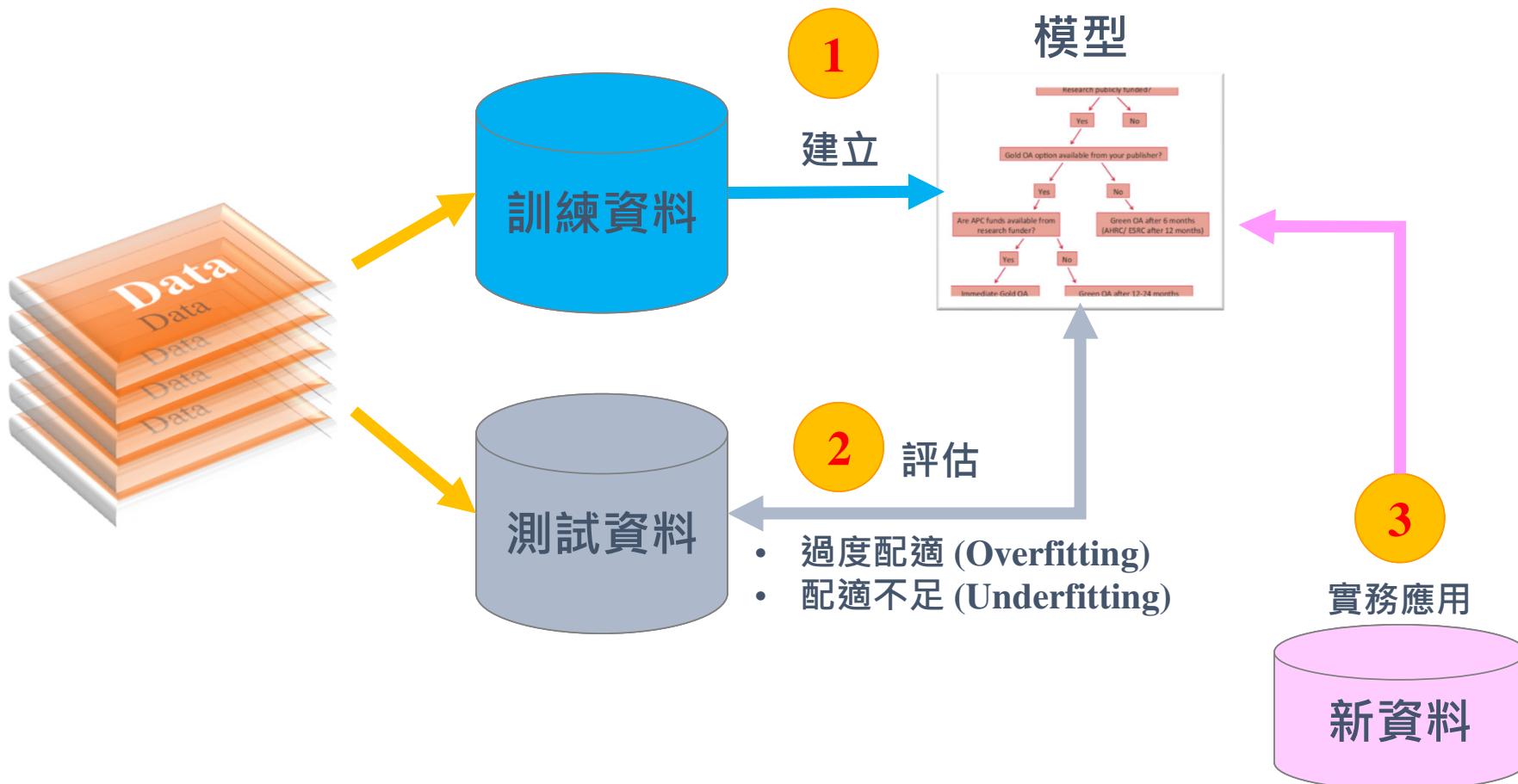
## 2. 資料理解

- 包括描述資料、探索資料、核驗資料品質
- 敘述統計分析
  - 六力分析(Python: pandas.describe 函數, R: summary函數)
- 繪圖
  - 依群組特性
  - 依時間特性
  - 新增評估欄位
- 趨勢
- 離群值 (outlier)
- 散佈圖、散佈圖矩陣
- 盒鬚圖

### 3. 資料準備



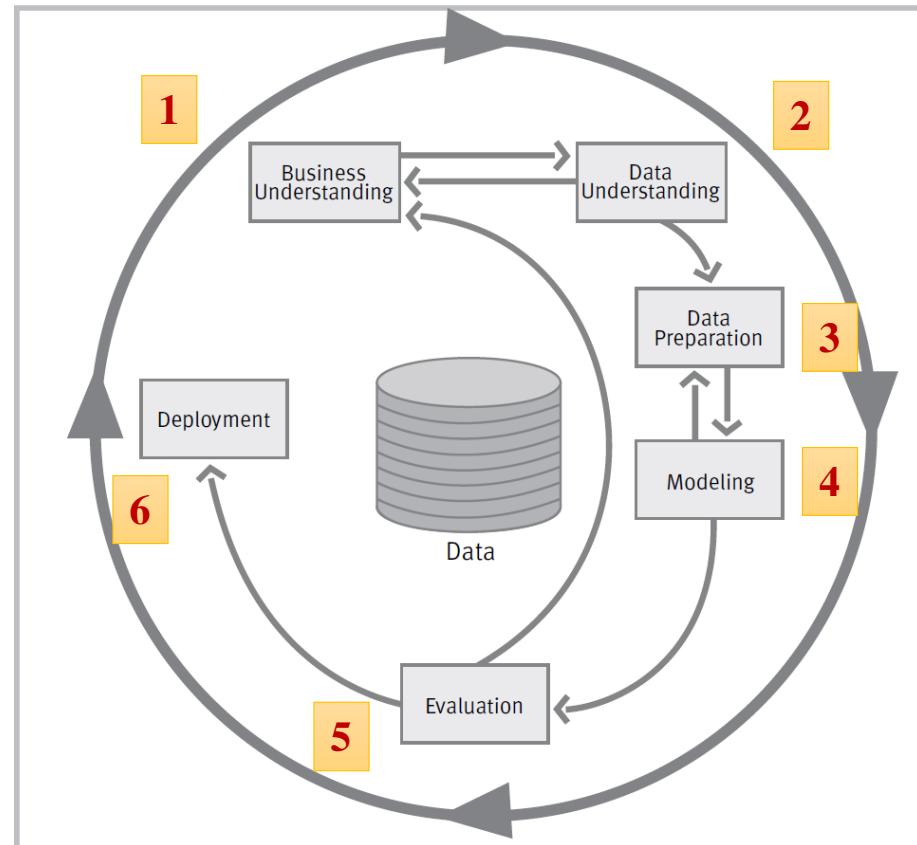
## 4. 模型建立、5. 評估與測試



## 6.佈署應用

- 資料分析專案的部署計畫(時間, 人力, 物力, 資源, \$)。
- 完成部署的服務應用，建立後續的監管與維運計畫。
- 整個專案評估報告，包含商業獲利分析、資料分析結果、部署與維運管理計畫、成本效益分析、總結整體專案與下一階段計畫。
- CRISP-DM 運用步驟
  - 使用 Power BI/Tableau/Python/R進行基礎資料分析與視覺化成果。
  - 建立機器學習預測模型 (Python: scikit-learn模組, R: tidymodels套件)。
  - 建立互動式資料分析 (Python: streamlit模組, R: shiny套件)。
  - 善用深度學習應用 (Python: tensorflow模組, R: torch套件)。

# CRISP-DM 資料探勘流程(續)



參考 [https://en.wikipedia.org/wiki/Cross\\_Industry\\_Standard\\_Process\\_for\\_Data\\_Mining](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)

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

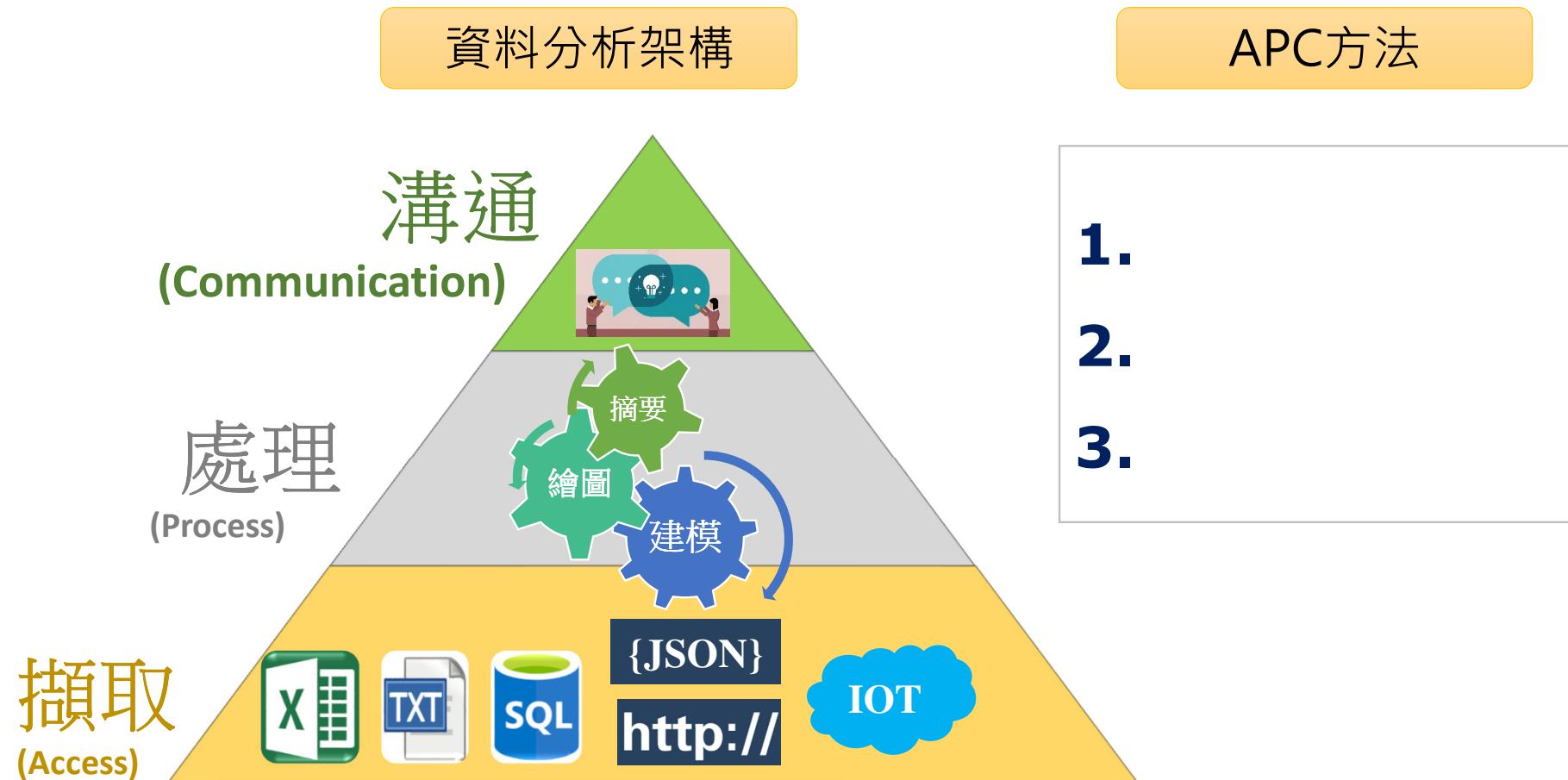
# 類別模型績效指標 - 混淆矩陣

- <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。
```

# ★★★資料分析架構→APC方法

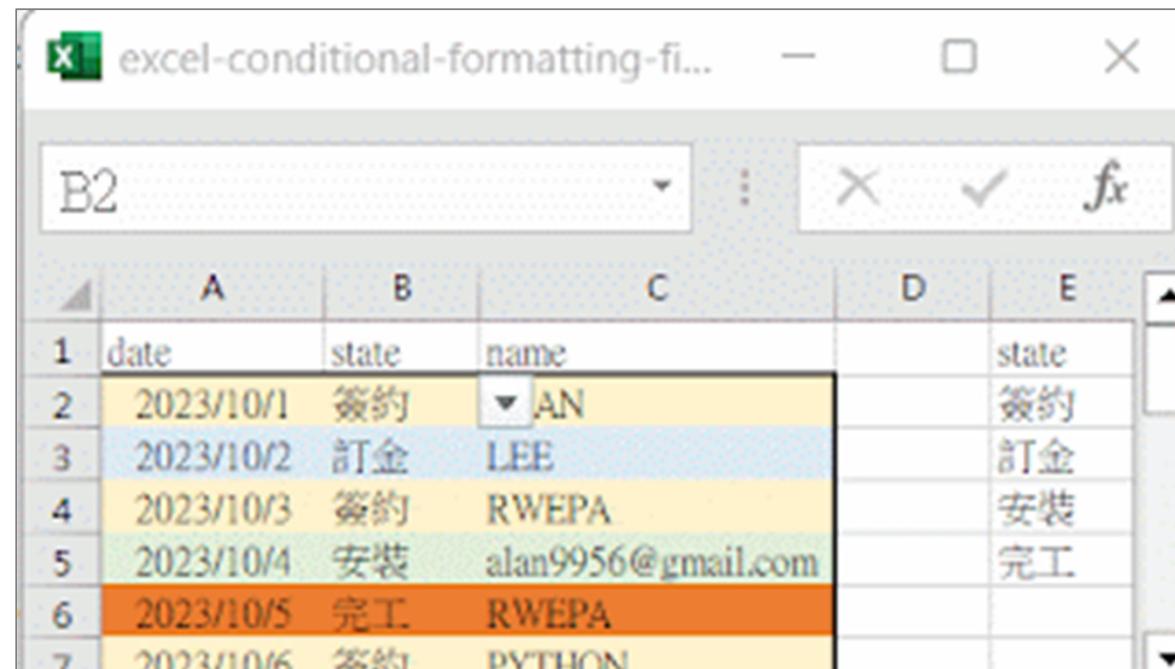


# 資料分析暨視覺化應用

---

# Excel 下拉式選單與條件式格式設定教學

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

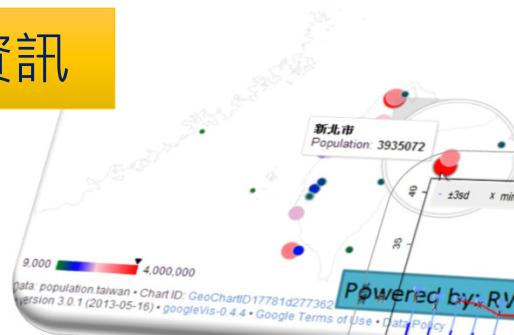


	A	B	C	D	E
1	date	state	name		state
2	2023/10/1	簽約	AN		簽約
3	2023/10/2	訂金	LEE		訂金
4	2023/10/3	簽約	RWEPA		安裝
5	2023/10/4	安裝	alan9956@gmail.com		完工
6	2023/10/5	完工	RWEPA		
7	2023/10/6	簽約	PYTHON		

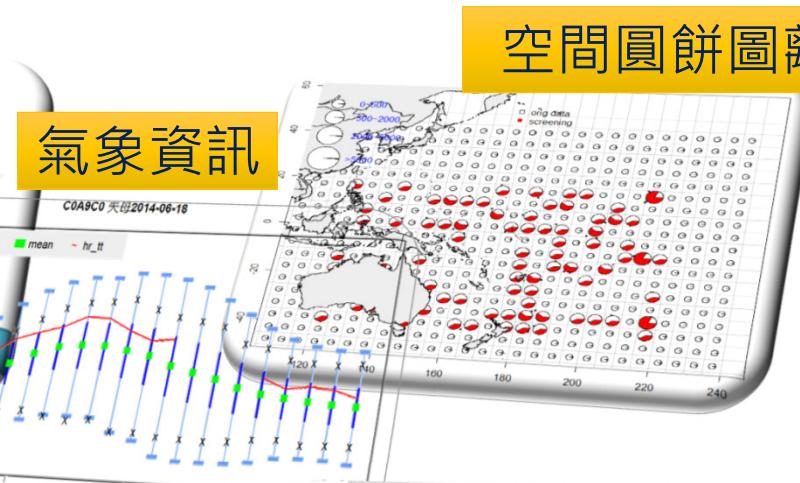
# 資料分析暨視覺化應用

Python + streamlit, R + shiny → 互動式網頁

地理資訊



氣象資訊

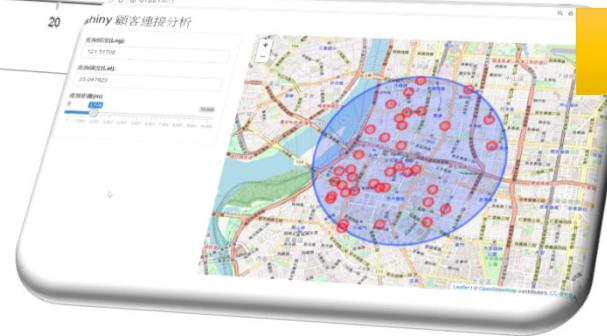


空間圓餅圖離群值分析

保險預測



顧客連結資訊



# 中央氣象局 1,600萬筆資料(14328檔案)

網頁呈現

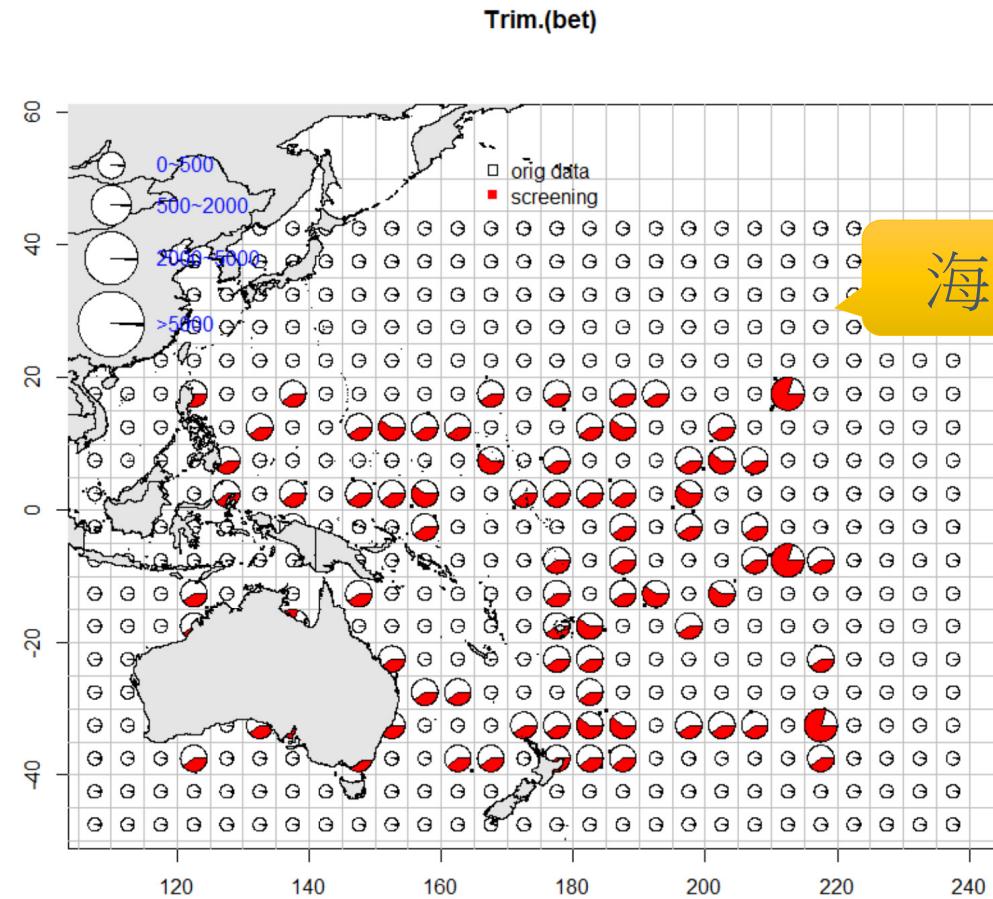


# 保險預測模型

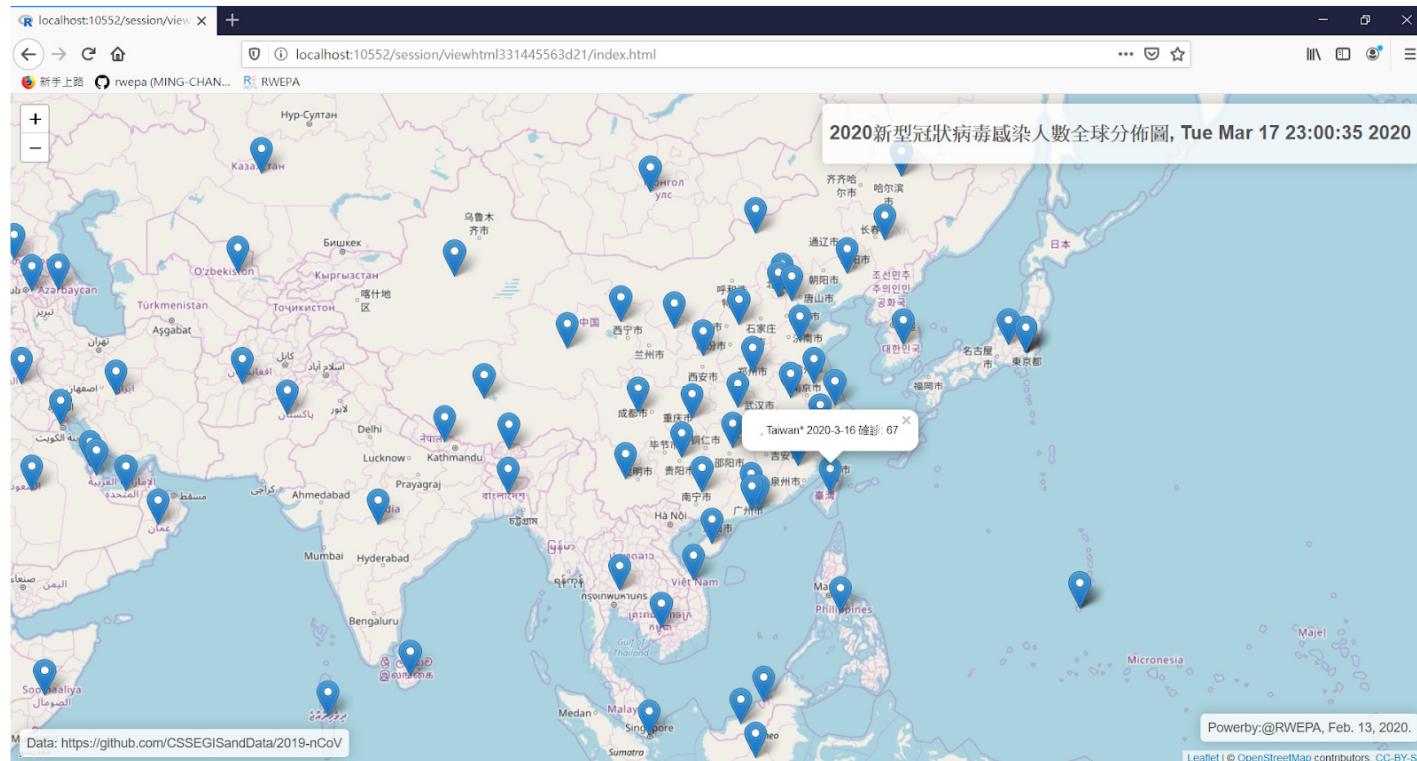
The screenshot shows the iInsurance interactive analysis platform version v.16.3.24. On the left, a yellow callout box highlights the "機率模型閾值" (Probability Model Threshold) section, which includes a slider set to 0.1 and a value input field. On the right, another yellow callout box highlights the "檢視結果" (View Results) button in the "預測資料上傳" (Prediction Data Upload) panel. A large yellow speech bubble labeled "預測結果" (Prediction Result) points to the final two columns of the table.

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

# 空間圓餅圖離群值分析



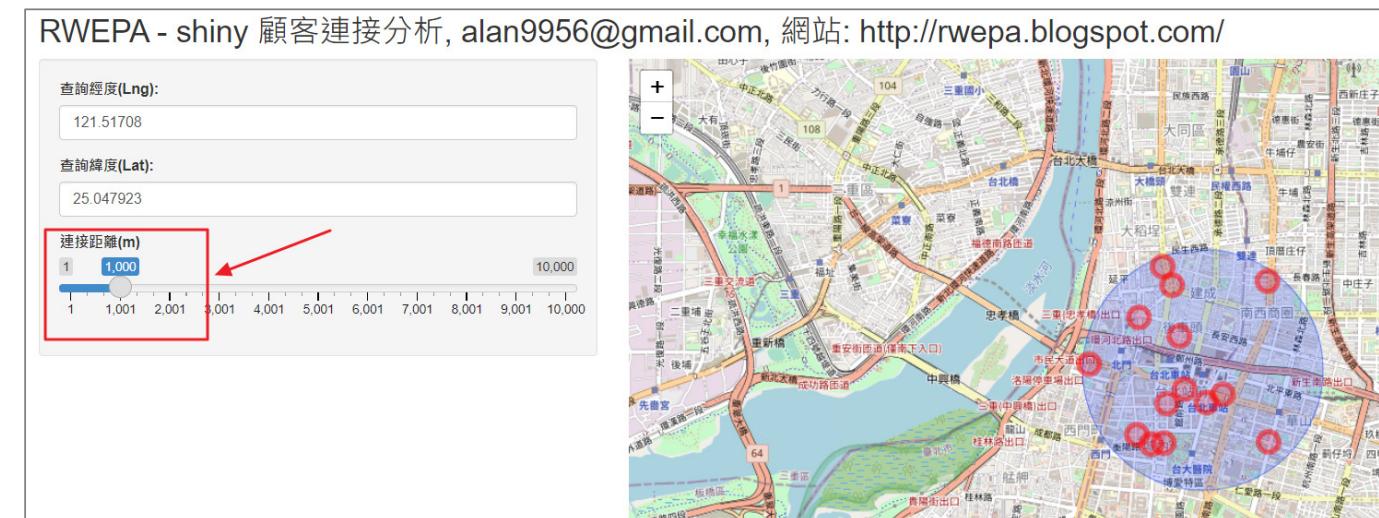
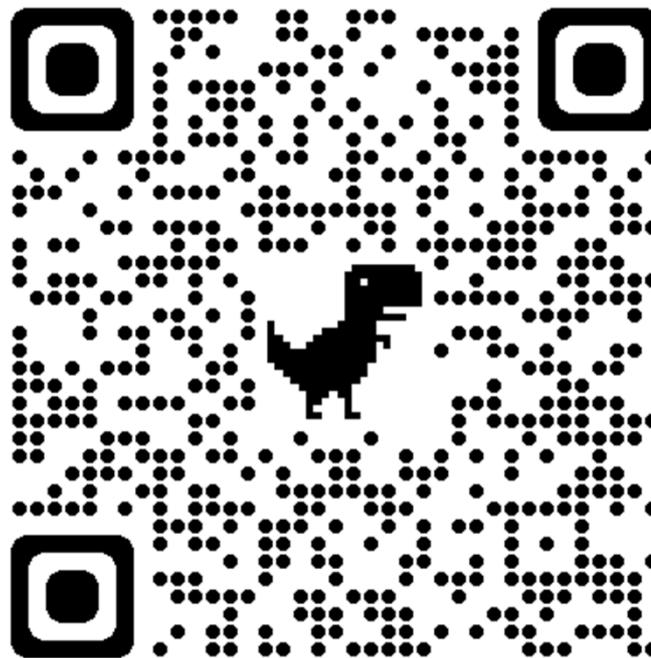
# 2020新型冠狀病毒視覺化



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

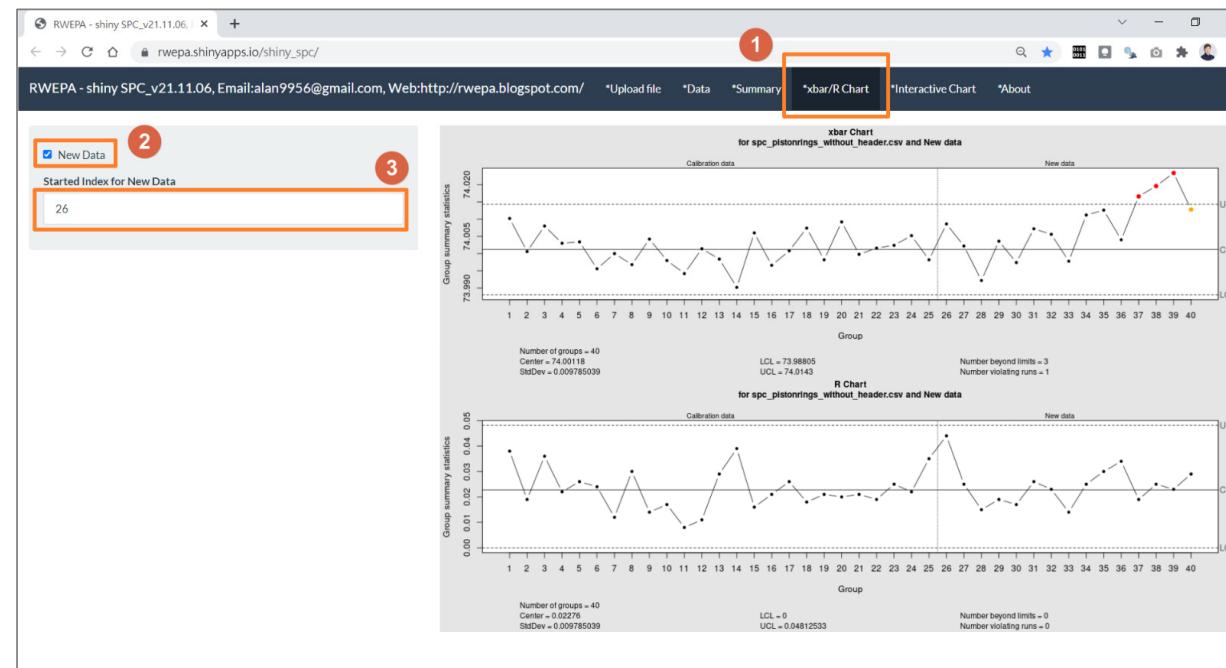
# shiny 顧客連接分析

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



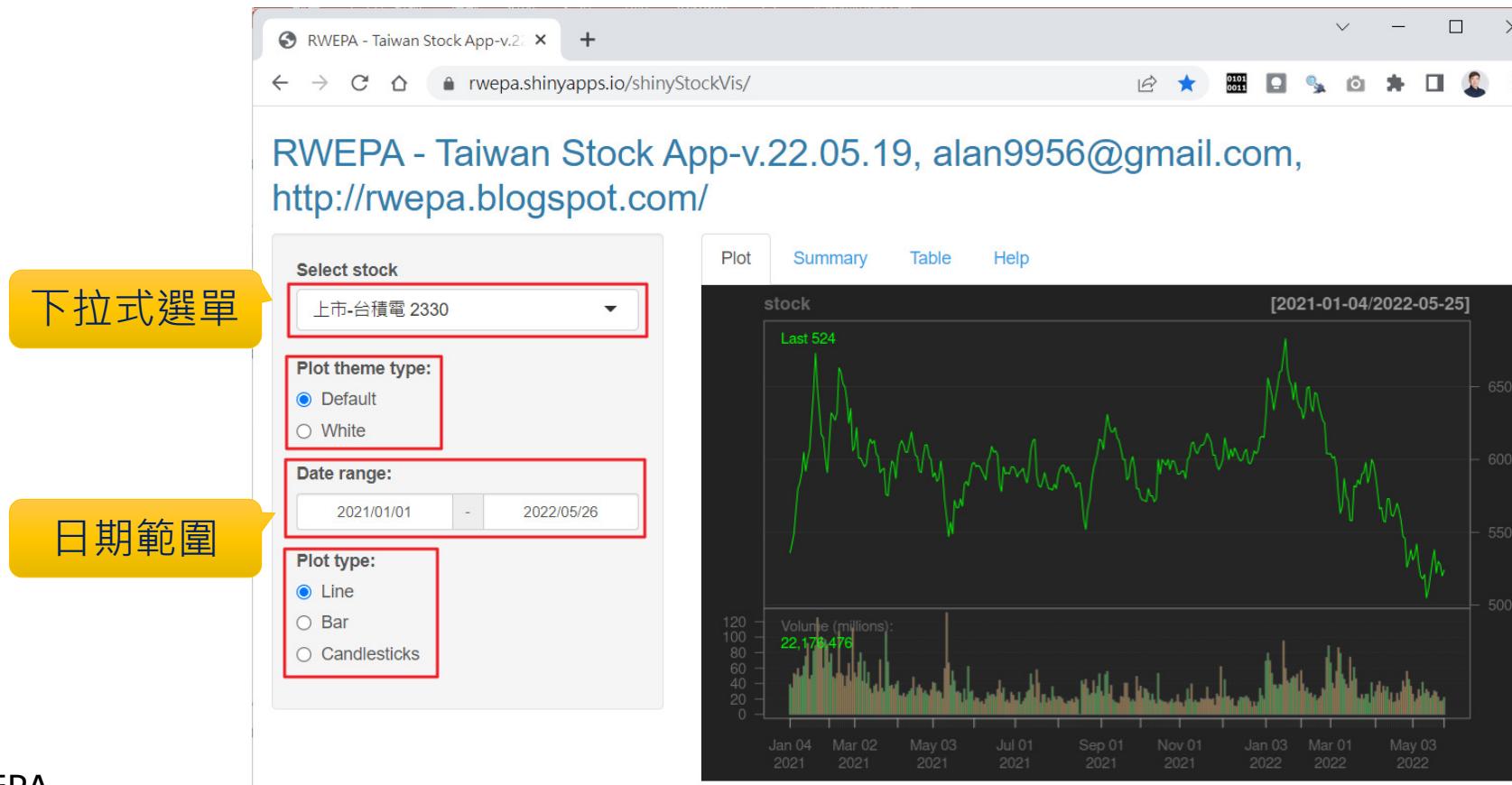
# 品質管制圖(quality control chart)應用

- 說明: <http://rwepa.blogspot.com/2021/10/r-shiny-quality-control-chart.html>
- 資料1: [https://github.com/rwepa/shiny\\_spc/blob/main/data/spc\\_wafer\\_with\\_header.csv](https://github.com/rwepa/shiny_spc/blob/main/data/spc_wafer_with_header.csv)
- 資料2: [https://github.com/rwepa/shiny\\_spc/blob/main/data/spc\\_pistonrings\\_without\\_header.csv](https://github.com/rwepa/shiny_spc/blob/main/data/spc_pistonrings_without_header.csv)
- 線上示範: [https://rwepa.shinyapps.io/shiny\\_spc/](https://rwepa.shinyapps.io/shiny_spc/)

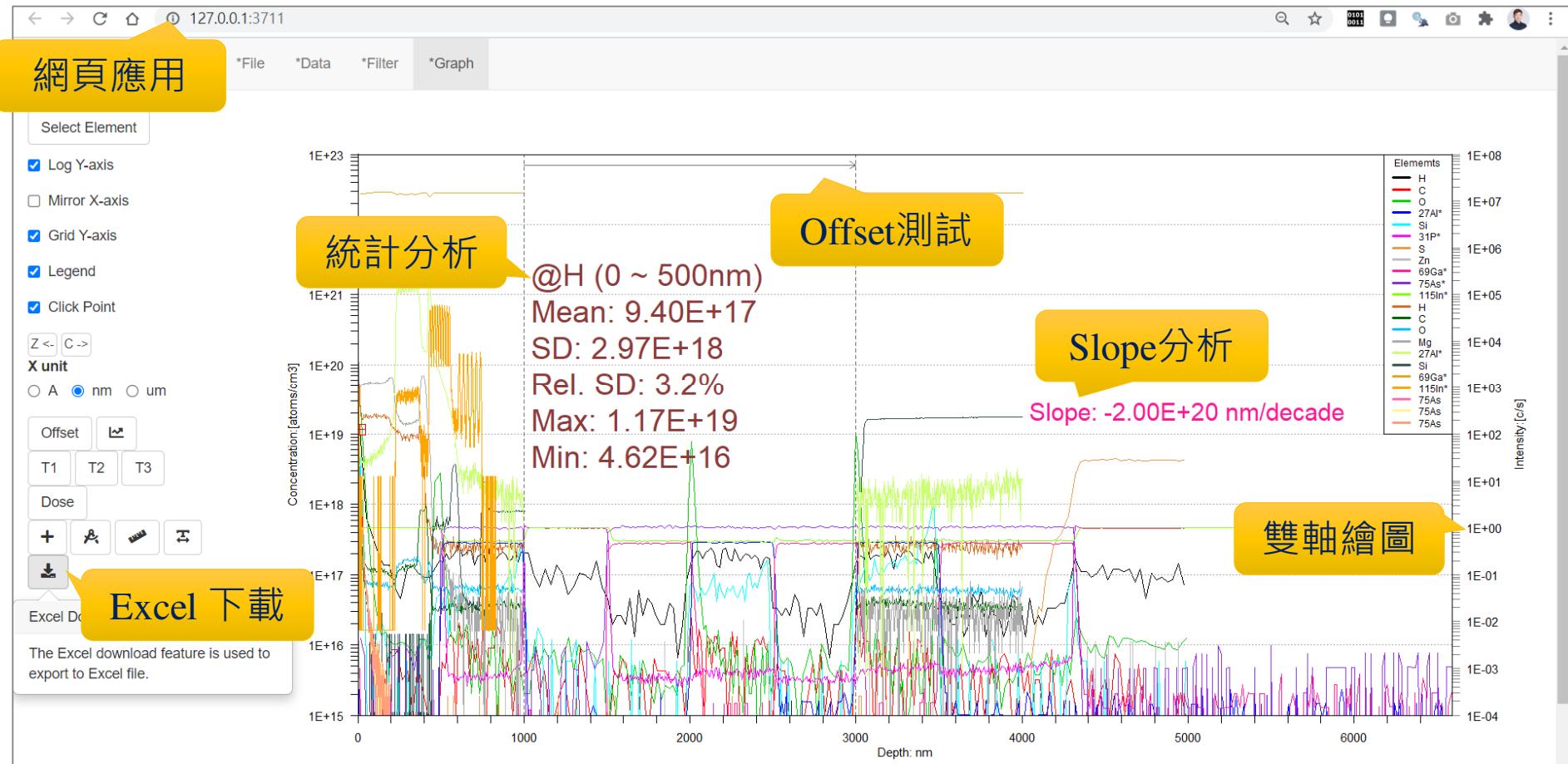


# Taiwan Stock App

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

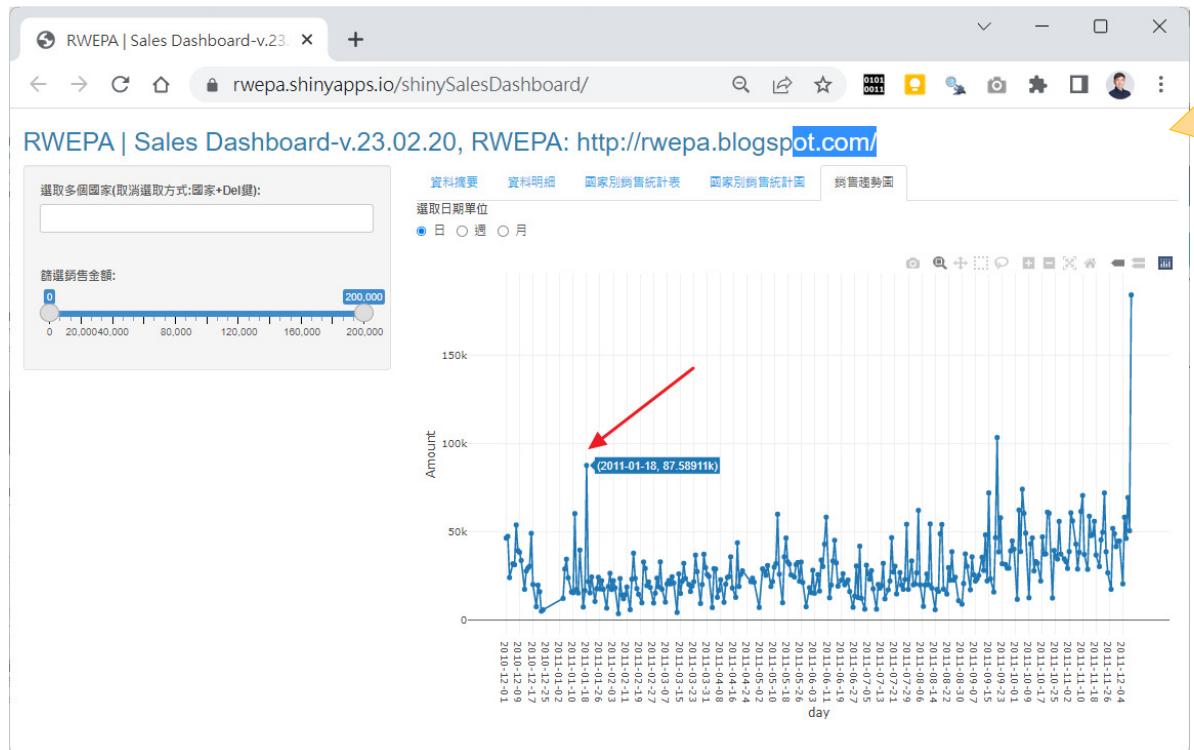


# 離子資料分析與視覺化應用



# RWEPA | shiny企業實務應用 第4集-shiny銷售儀表板

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



謝謝 ^\_ ^

訂閱 + 讚 + 開啟小鈴鐺

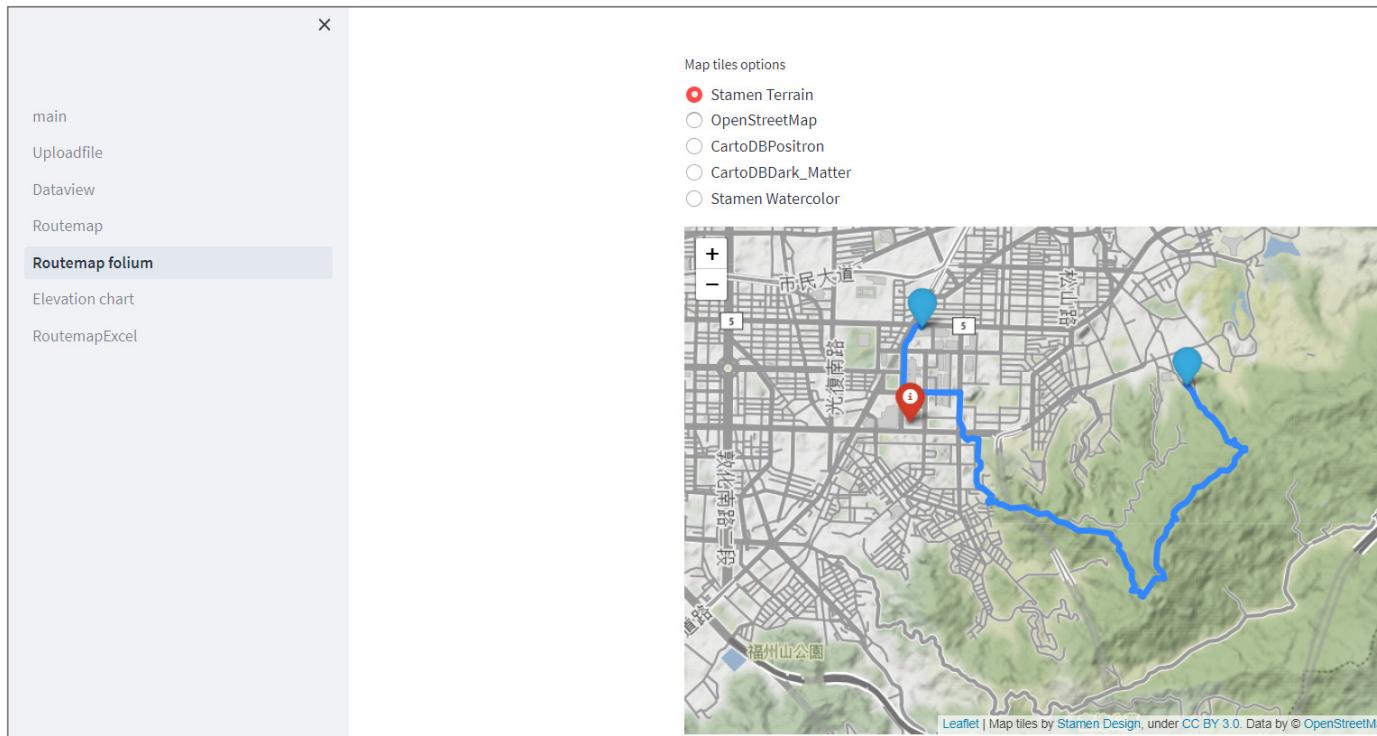
# shiny企業實務應用 第6集-小明算命師(下) - 第1季完結篇

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



# 登山路線視覺化分析平台 (Python + Streamlit)

- YouTube : [https://youtu.be/-\\_zghs2qrlg](https://youtu.be/-_zghs2qrlg)
- 系統展示 <https://rwepa-climb.streamlit.app/>

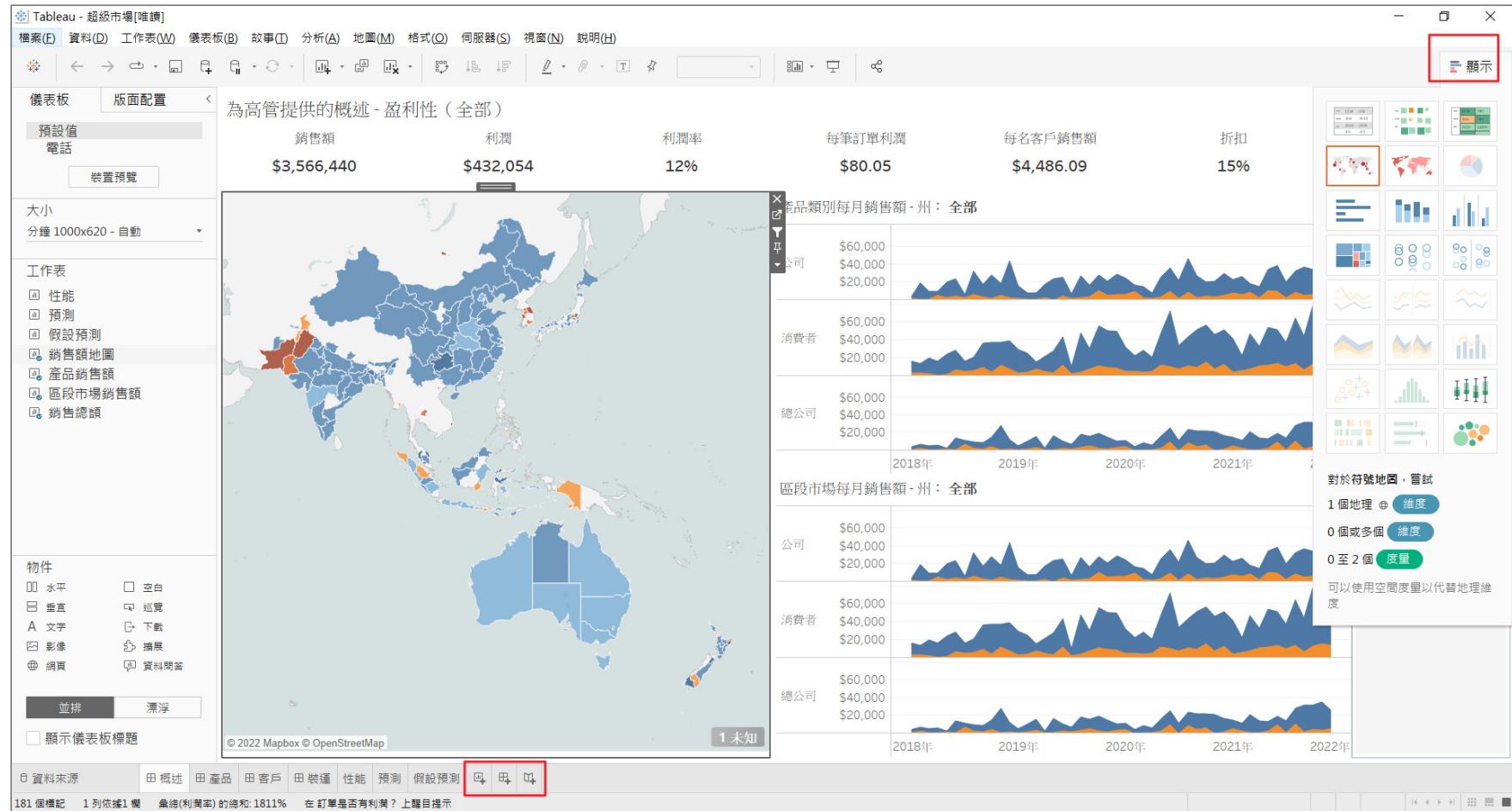


# 客戶輪廓分析 (Power BI)

The screenshot displays a Power BI Desktop interface with the following components:

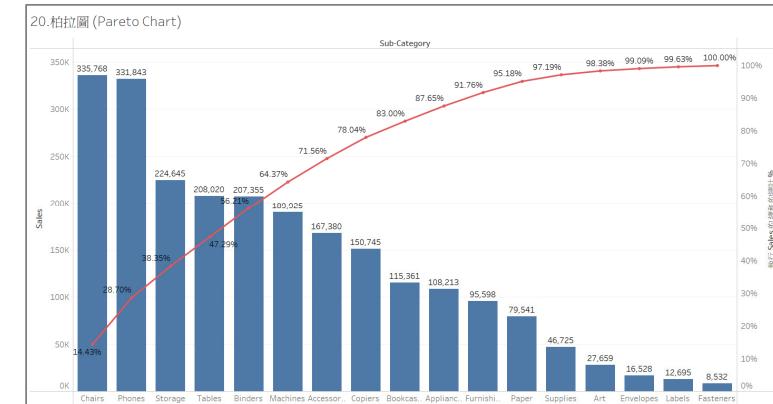
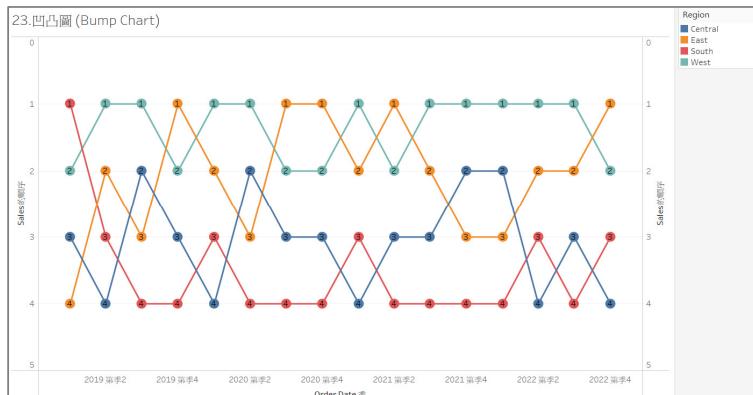
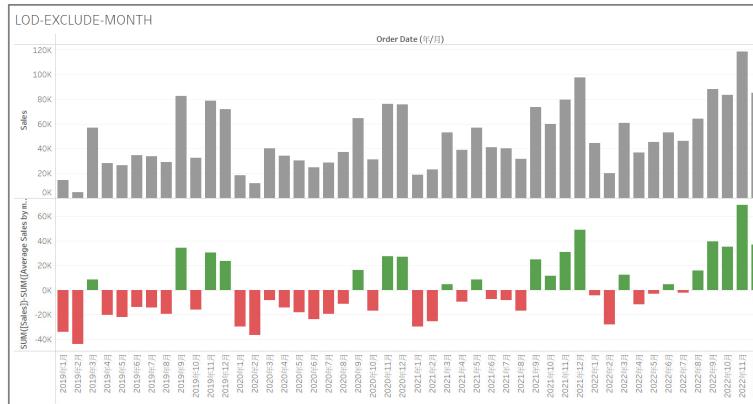
- Top Bar:** Includes tabs for 檔案 (File), 常用 (Home), 插入 (Insert), 模型化 (Modeling), 檢視 (View), 最佳化 (Optimize), and 說明 (Help). It also shows the current file name: PowerBI\_customer\_profile\_analysis - Power BI Desktop.
- Left Sidebar:** Contains sections for 剪貼簿 (Clipboard) and 資料 (Data). The Data section includes options for 取得資料 (Get Data) from various sources like Excel, 資料中樞 (Data Catalog), and SQL Server, as well as 輸入資料 (Import Data) from Dataverse and 最近使用的來源 (Recent Sources).
- Right Sidebar:** Features sections for 視覺效果 (Visualizations) and 資料 (Data). The Visualizations section shows icons for various chart types. The Data section lists tables such as DAX量值管理表 (DAX Value Management Table), 分行資料表 (Branch Table), 日期對照表 (Date Calendar), 刷卡交易資料表 (Card Transaction Table), 刷卡國別對照表 (Card Country Comparison Table), 刷卡類別對照表 (Card Category Comparison Table), 客戶資料表 (Customer Table), 理財產品代碼表 (Financial Product Code Table), 理財產品交易資料表 (Financial Product Transaction Table), 業務員資料表 (Salesperson Table), and Dynamic\_Calendar.
- Dashboard Content:**
  - Overview Section:** Displays the total number of customers (666), average age (34.8), and gender distribution (FEMALE: 378, MALE: 288). A pie chart shows 43.2% female and 56.8% male.
  - Demographic Sections:** Includes bar charts for 年齡組距 (Age Groups) and 職業 (Occupation), and horizontal bar charts for 縣市別 (County/City) and 區域別 (Region).
  - Customer Segmentation:** A dropdown menu titled "客戶等級" (Customer Grade) offers filters for 全選 (All), VIP客戶 (VIP Customer), 一般客戶 (General Customer), and 財富管理客戶 (Wealth Management Customer).

# Superstore (Tableau)



# 智慧製造應用 (Tableau)

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



25.購物籃分析 (Market Basket Analysis)

Sub-Category

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

# Python 程式設計-李明昌 免費電子書

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

## 主題: Python 程式設計-李明昌 - ipynb

檔名: Python\_Programming\_Lee\_ipynb.zip

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

下載: [https://github.com/rwepa/DataDemo/blob/master/Python\\_Programming\\_Lee\\_ipynb.zip](https://github.com/rwepa/DataDemo/blob/master/Python_Programming_Lee_ipynb.zip)



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

# R 入門資料分析與視覺化應用(7小時28分鐘)

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



課程提供教學範例的原始程式檔案與資料集 +中文字幕

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

# R 商業預測應用(8小時53分鐘)

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



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

課程提供教學範例的原始程式檔案與資料集 +中文字幕

# 如何學習？

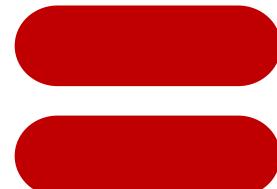
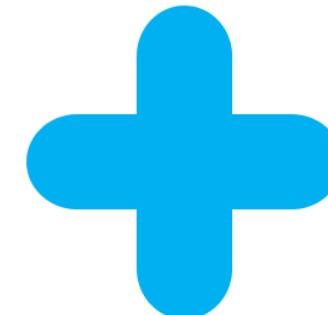
- 熟悉教材內容
- 將教材內容的資料集改為工作資料集(企業, 學術)
- 遇到問題時, 想辦法**尋找答案**
- 掌握 APC方法
- 掌握 ①摘要 ②繪圖 ③建模
- 參考網路應用文章 (進階) & 學術論文

```
尋找答案 = {"方法1": "同事,同學,朋友等",  
            "方法2": "Google",  
            "方法3": "alan9956@gmail.com"}
```

**WHY!**



# 學習目標



資料庫 (SQL)

# 資料視覺化簡介

---

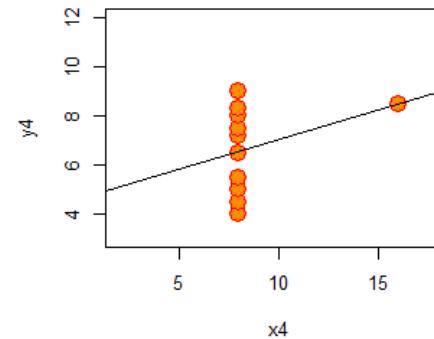
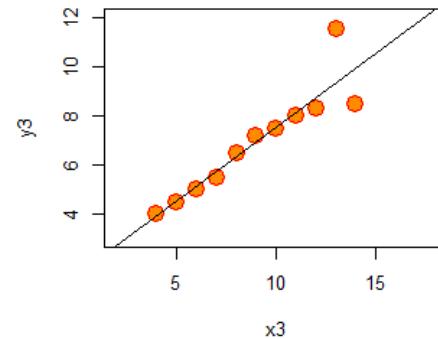
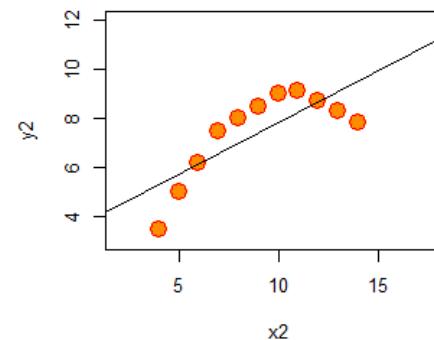
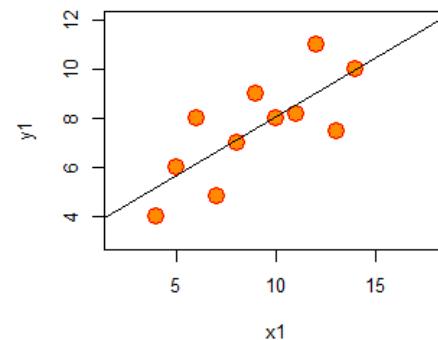
# 為什麼需要視覺化

- 人腦不善於閱讀和分析大量資料

	A	B	C	D	E	F	G
1	Country or Area	Subgroup	Year	Source	Unit	Value	Value Footnotes
2	Afghanistan	Male 5-14 yr	2002-2011	UN_Demographic a Percent		11	1
3	Afghanistan	Female 5-14 yr	2002-2011	UN_Demographic a Percent		10	1
4	Afghanistan	Total 5-14 yr	2002-2011	UN_Demographic a Percent		10	1
5	Albania	Male 5-14 yr	2002-2011	UN_Demographic a Percent		14	1
6	Albania	Female 5-14 yr	2002-2011	UN_Demographic a Percent		9	1
7	Albania	Total 5-14 yr	2002-2011	UN_Demographic a Percent		12	1
8	Algeria	Male 5-14 yr	2002-2011	UN_Demographic a Percent		6	2
9	Algeria	Female 5-14 yr	2002-2011	UN_Demographic a Percent		4	2
10	Algeria	Total 5-14 yr	2002-2011	UN_Demographic a Percent		5	2
11	Angola	Male 5-14 yr	2002-2011	UN_Demographic a Percent		22	3
12	Angola	Female 5-14 yr	2002-2011	UN_Demographic a Percent		25	3
13	Angola	Total 5-14 yr	2002-2011	UN_Demographic a Percent		24	3
14	Argentina	Male 5-14 yr	2002-2011	UN_Demographic a Percent		8	2
15	Argentina	Female 5-14 yr	2002-2011	UN_Demographic a Percent		5	2
16	Argentina	Total 5-14 yr	2002-2011	UN_Demographic a Percent		7	2
17	Armenia	Male 5-14 yr	2002-2011	UN_Demographic a Percent		5	2
18	Armenia	Female 5-14 yr	2002-2011	UN_Demographic a Percent		3	2
19	Armenia	Total 5-14 yr	2002-2011	UN_Demographic a Percent		4	2

# 為什麼需要視覺化 (續)

- 但是人腦很善於閱讀和分析圖形



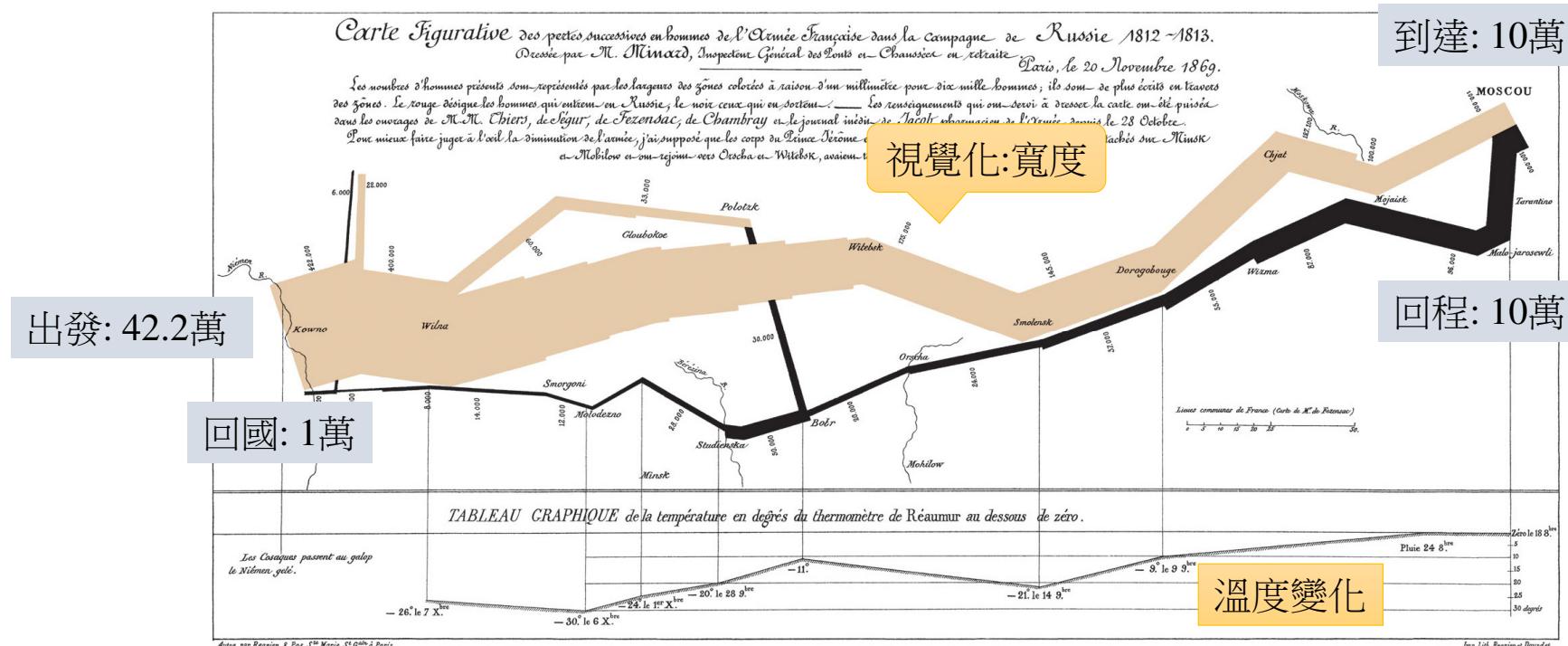
## 為什麼需要視覺化 (續)

- 人類是視覺動物，其視覺神經系統有強大的模式識別和分析能力，視覺化是啟動這套系統的途徑
- 視覺化是一種高效的資訊壓縮和展示方法，能將大量資料快速傳輸給人的大腦
- 視覺化能探索並提煉資料，並促進新的問題的提出和解決
- 探索式資料分析是資料視覺化的重要應用
  - identify properties, relationships, regularities, or patterns  
(性質, 關係, 規則, 樣式)

Patterns 樣式

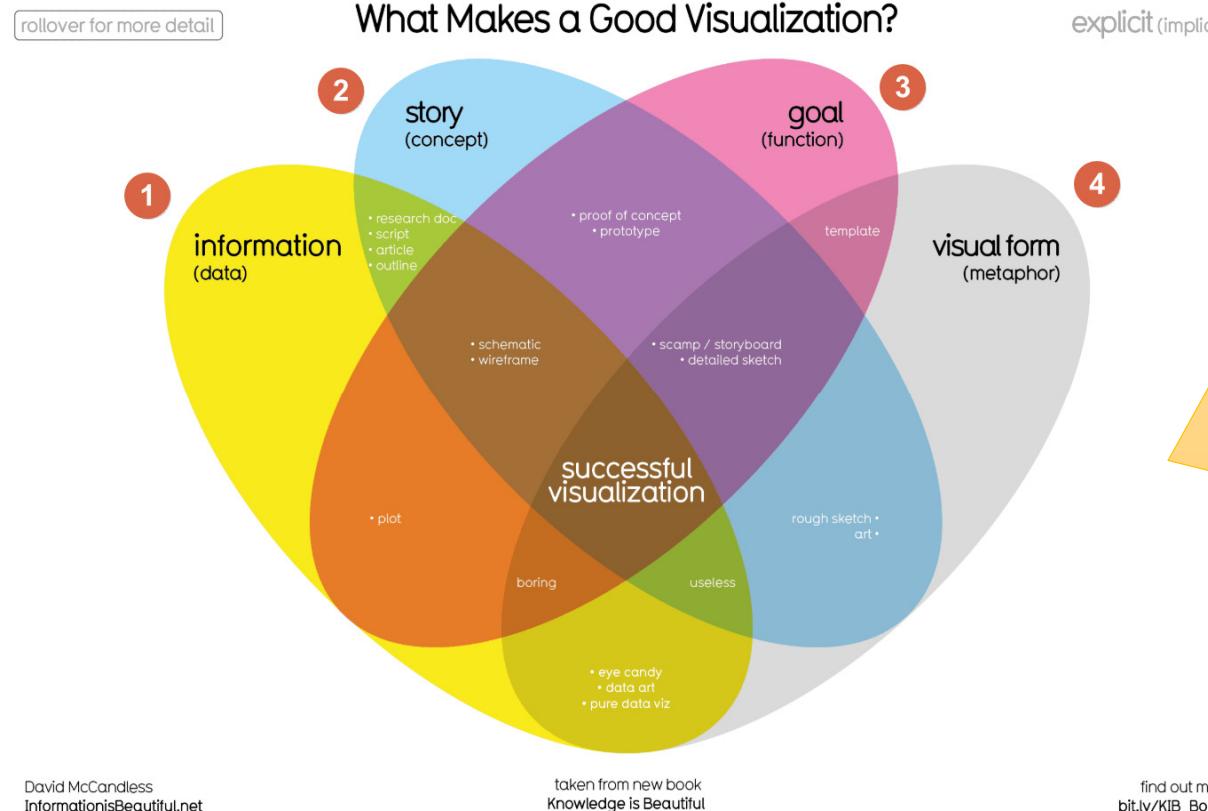
## 俄法戰爭

## 第一個著名的視覺化



- Charles Joseph Minard, a French engineer, in 1869.
- Illustrate the number of losses suffered by Napoleon's army during the disastrous march toward Moscow in 1812.6 - 1812.11

# 成功的資料視覺化-4大關鍵



1. 資訊 Information: 準確性, **真實性**, 一致性
2. 故事 Story: 資料視覺化的認同性
3. 目標 Goal: 找到重要結論
4. 視覺表現 Visual Form: 藝術的呈現

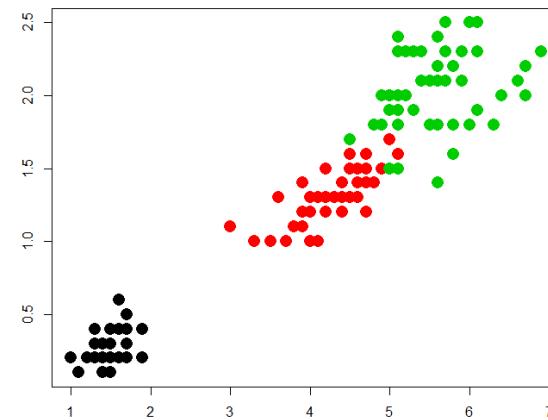
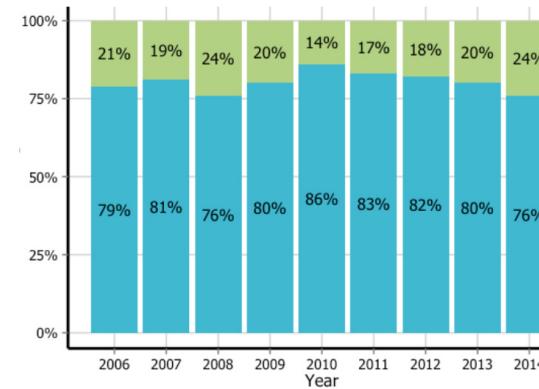
1. 參考資料 <https://informationisbeautiful.net/visualizations/what-makes-a-good-data-visualization/>
2. David McCandless , Knowledge Is Beautiful: Impossible Ideas, Invisible Patterns, Hidden Connections - Visualized, Oct 21, 2014.

# 視覺化目標

- 資料來源 (內部/外部, 免費/付費)
- 聽眾 vs. 觀眾
- 期望結果: 高層主管 vs. 基層員工
- 確認視範化目標
  - 產品銷售資料分析(產品別, 部門別, 時間別)
  - 教務研究-學生休退學預警分析(院別, 系別, 學習, 休學, 退學)
  - 醫學研究-疾病/藥品預測分析 (類別預測, 數值預測)
  - 製造分析-生產最佳化分析
  - 社會經濟資料分析 (趨勢)

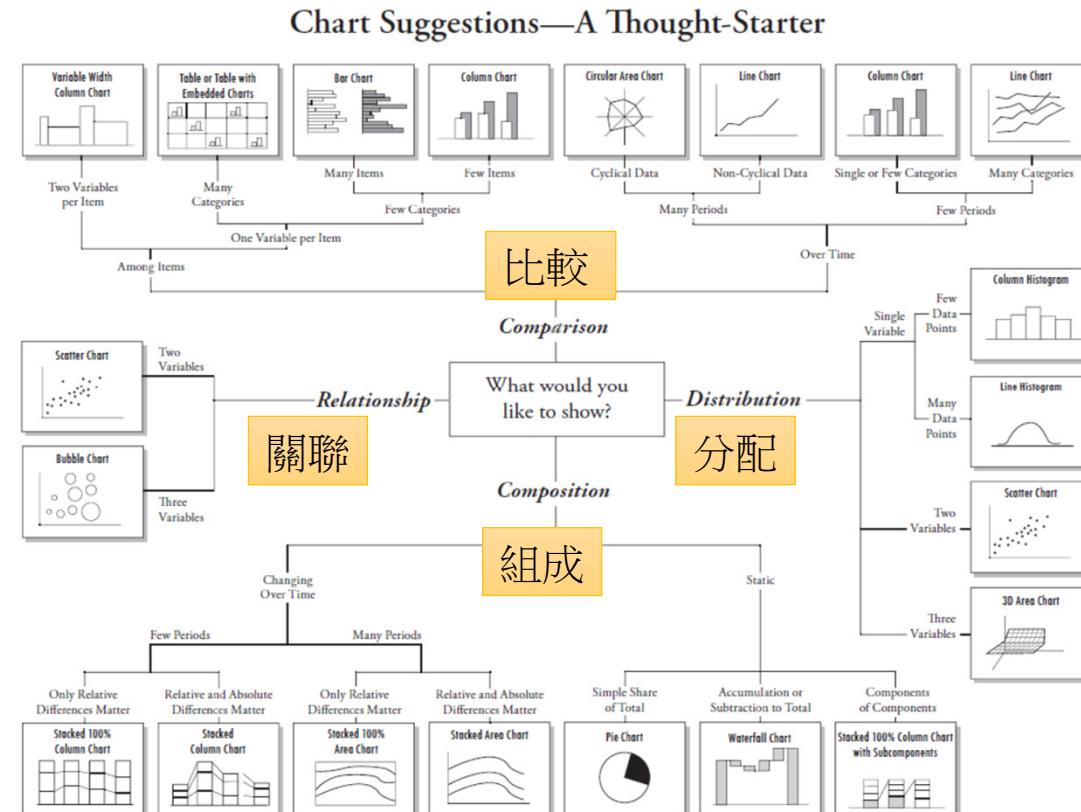
# 視覺化結果 - 四大繪圖類型

- 比較 Comparison
- 組成 Composition
- 分配 Distribution
- 關聯 Relationship



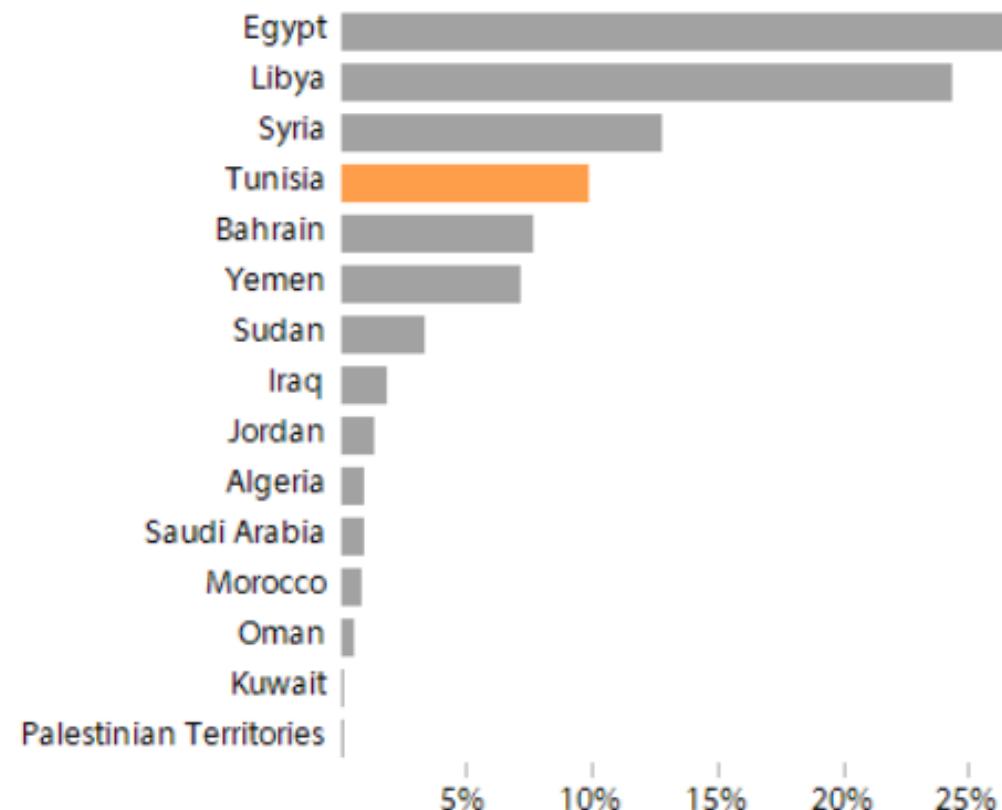
參考資料 <http://www.tatvic.com/blog/7-visualizations-learn-r/>

# 選取適當視覺化



參考資料 <https://extremepresentation.typepad.com/files/choosing-a-good-chart-09.pdf>

# 資料視覺化的各類應用



水平長條圖

- 由上至下排列
- 由大至小

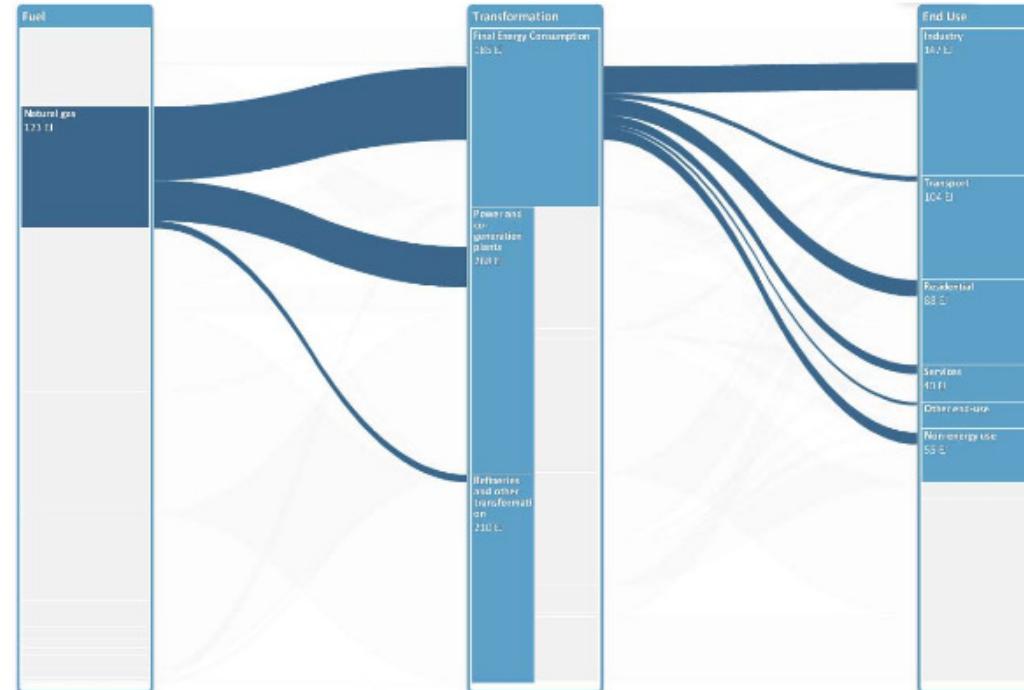
# 文字雲



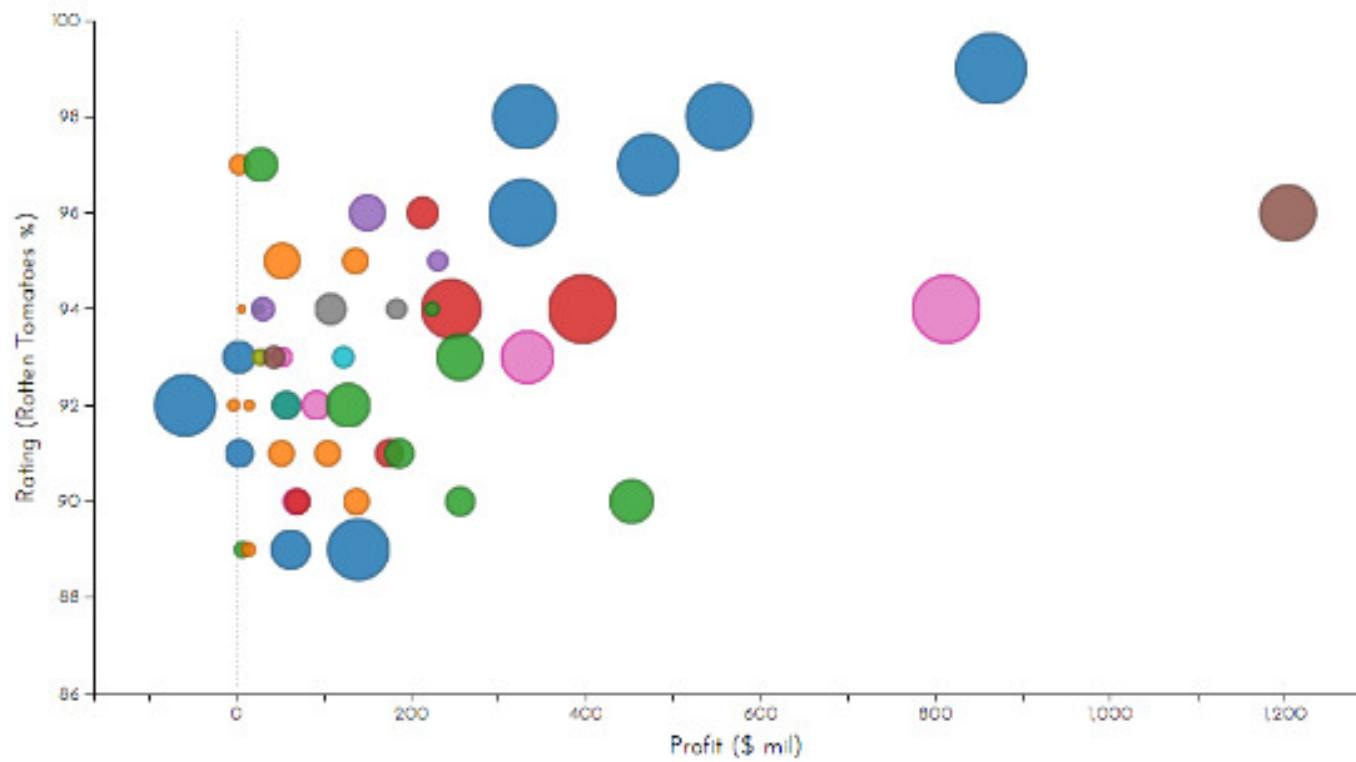
# treemap 樹狀圖(層次結構的展示)



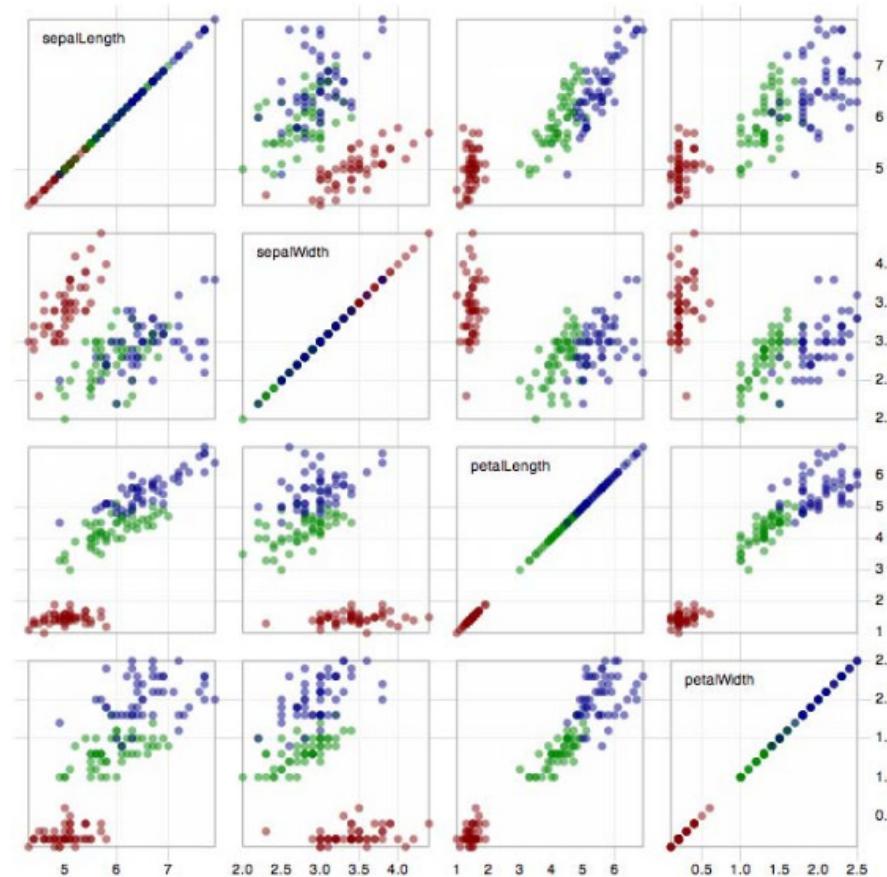
# Sankey chart 桑基圖



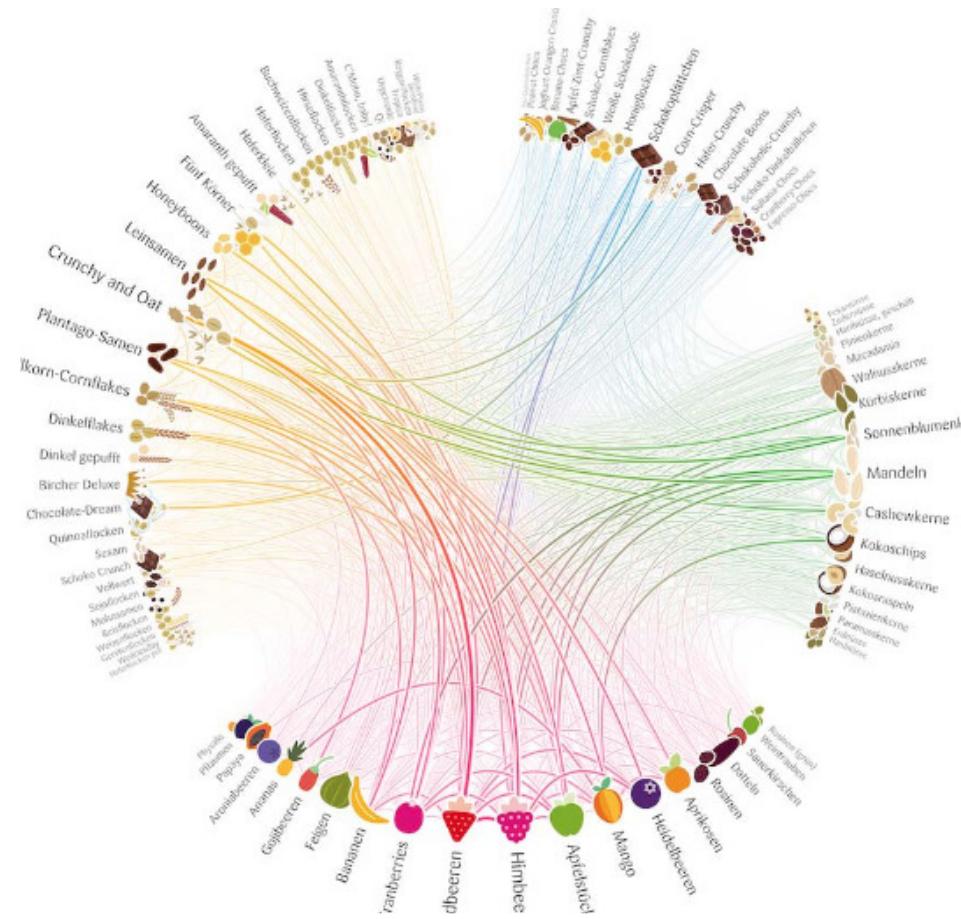
# 視覺化案例: 顯示關係的強弱



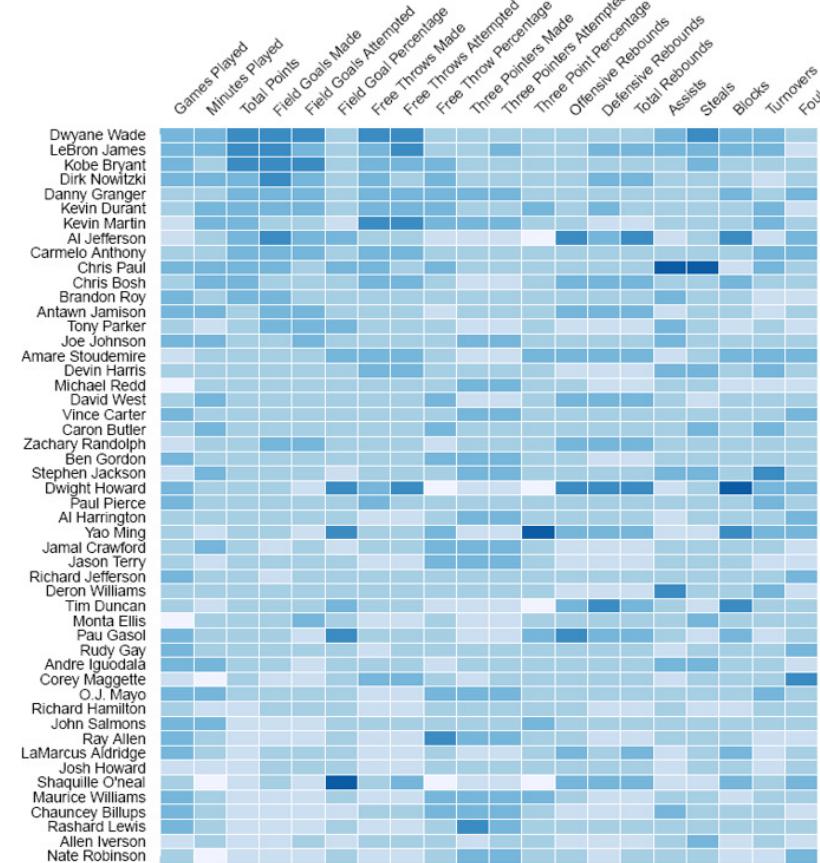
# 視覺化案例: 顯示關係的強弱(續)



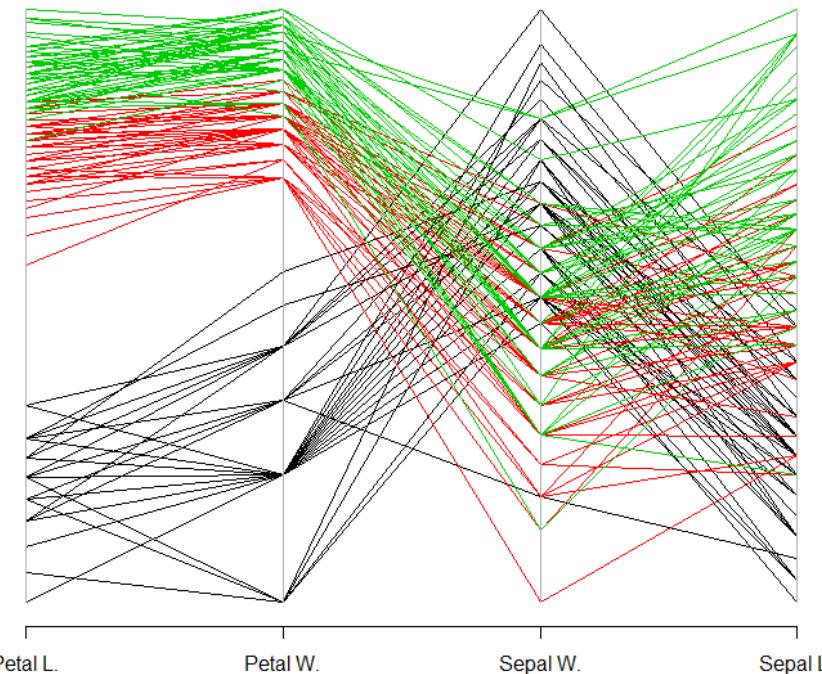
# 視覺化案例: 顯示關係的強弱(續)



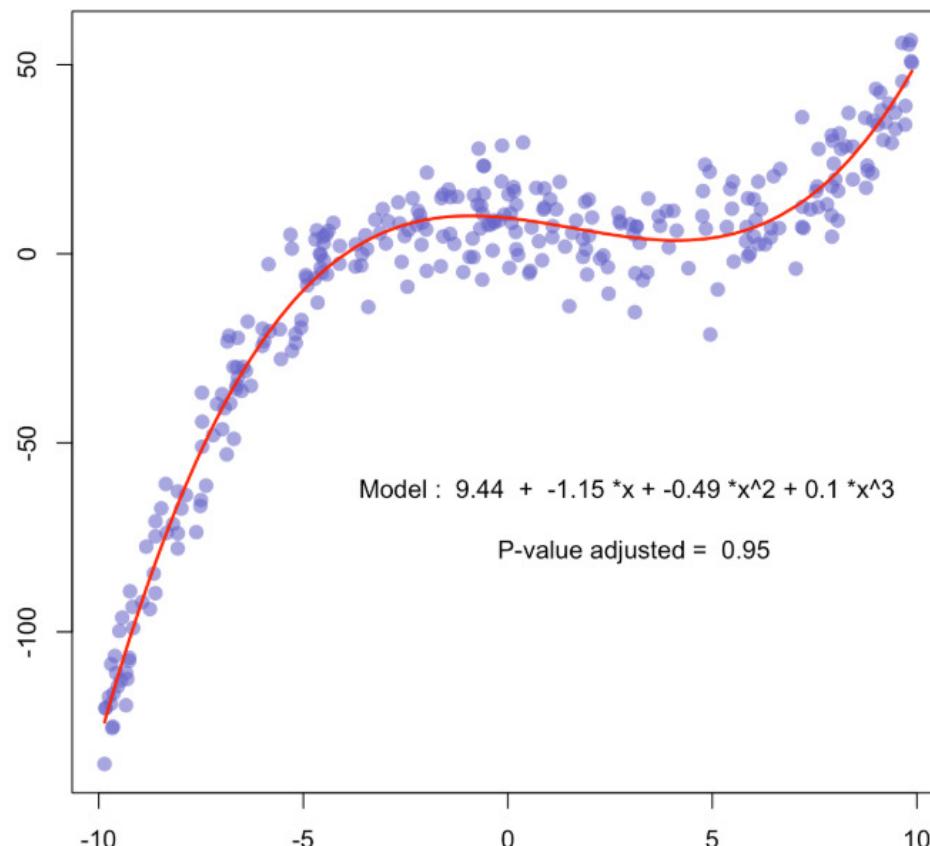
# Heatmap 熱繪圖



# 平行座標軸 Parallel coordinates



# Polynomial curve fitting



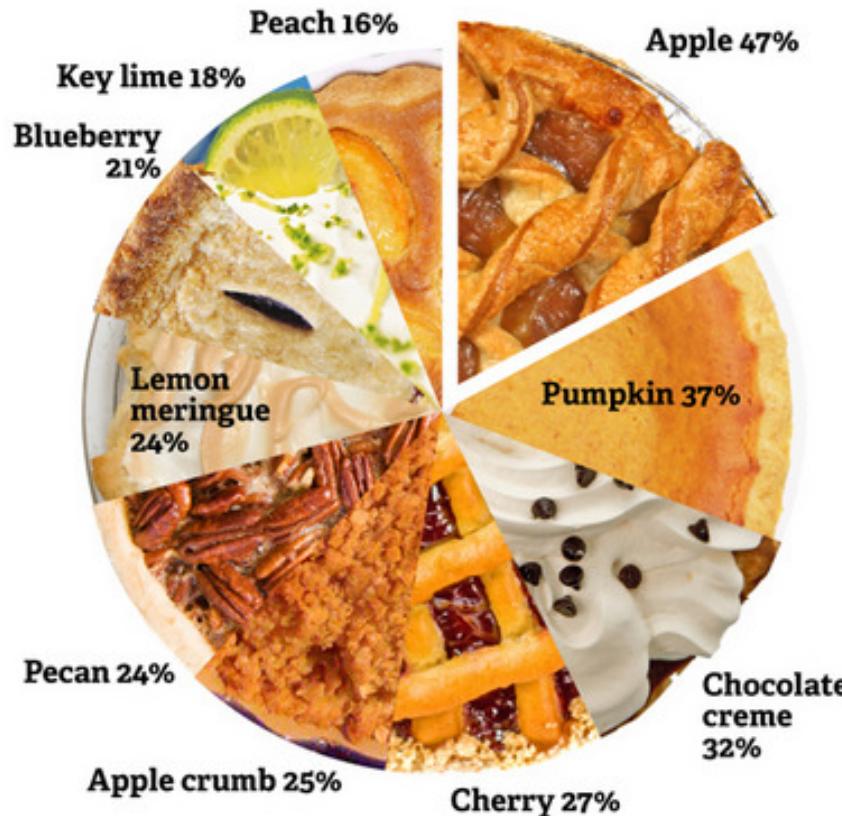
參考資料 <https://www.r-graph-gallery.com/all-graphs>

# 視覺化的陷阱

- 避免使用太複雜三維圖形
- 圓形圖處理的類別數目不可過多
- 避免使用漸變色
- 避免使用陰影

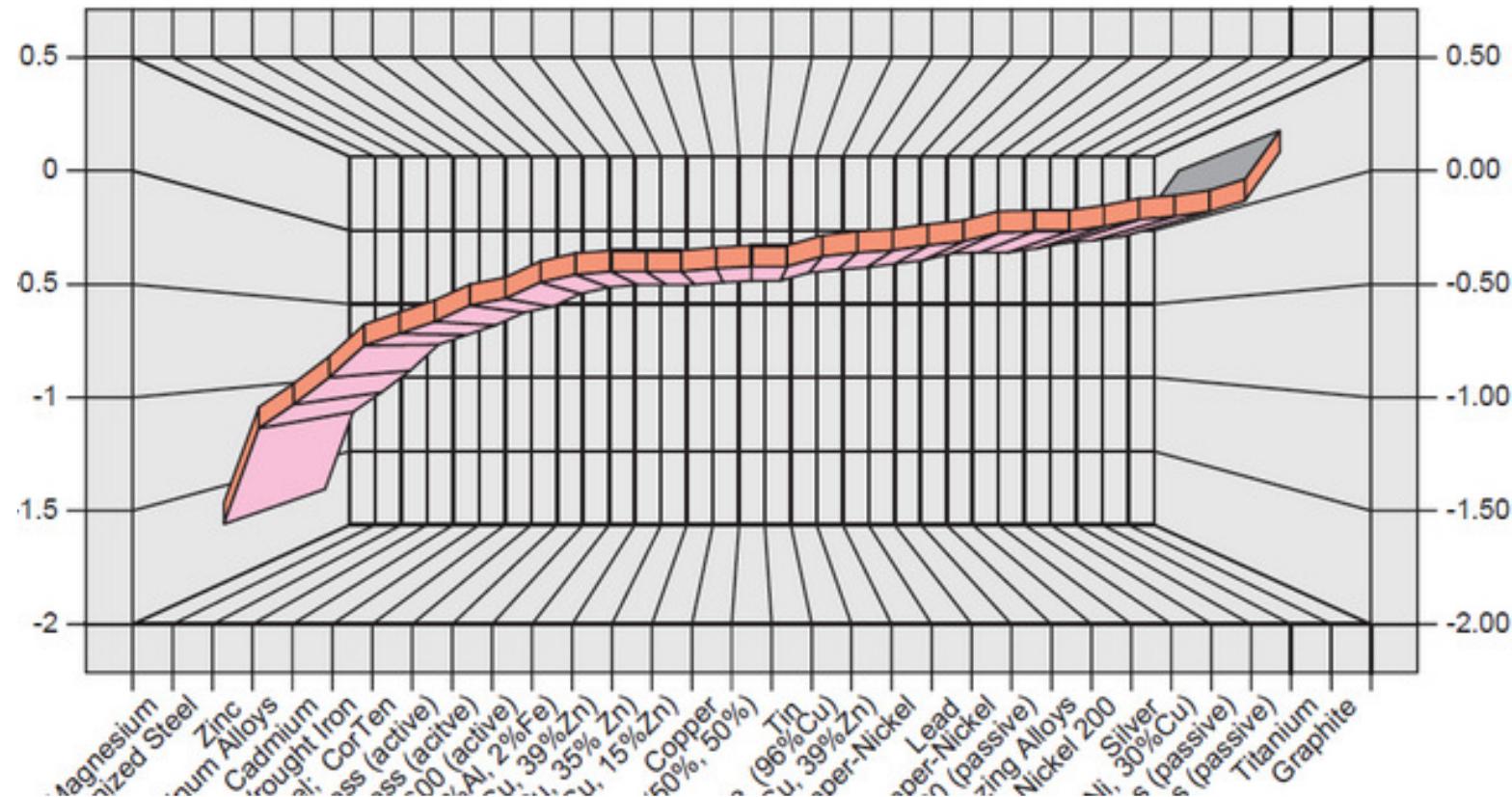
# 不佳的視覺化

What are your three most favorite types of pie?

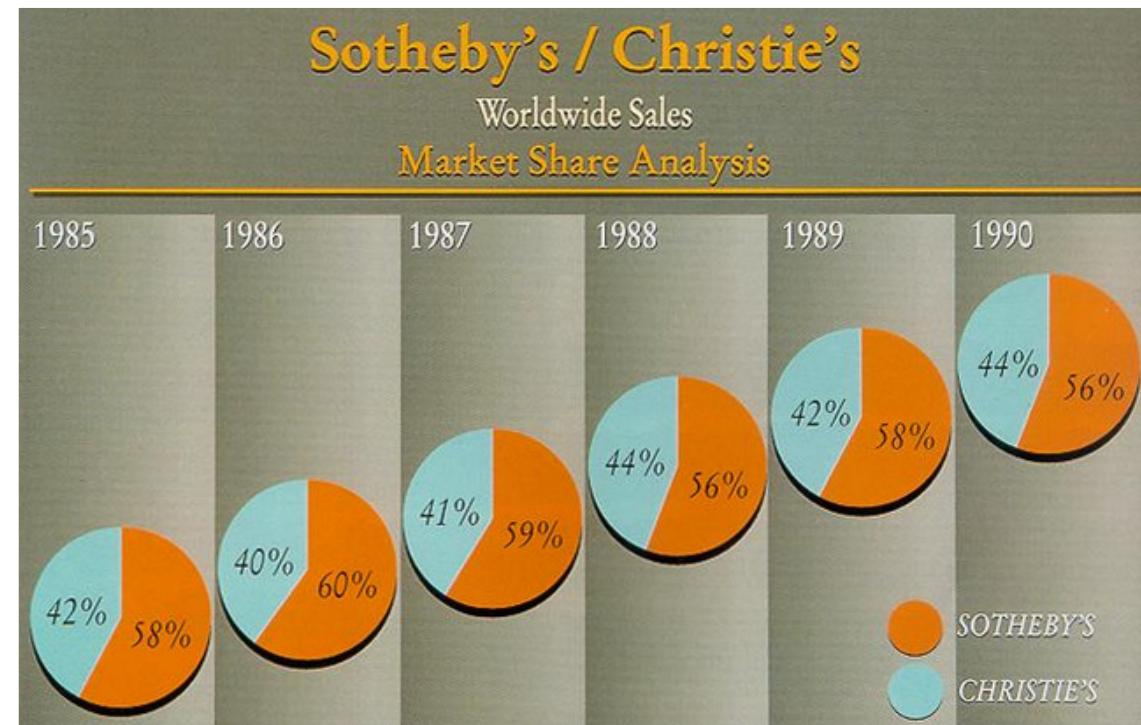


# 不佳的視覺化(續)

注意: 3D 效果偏差

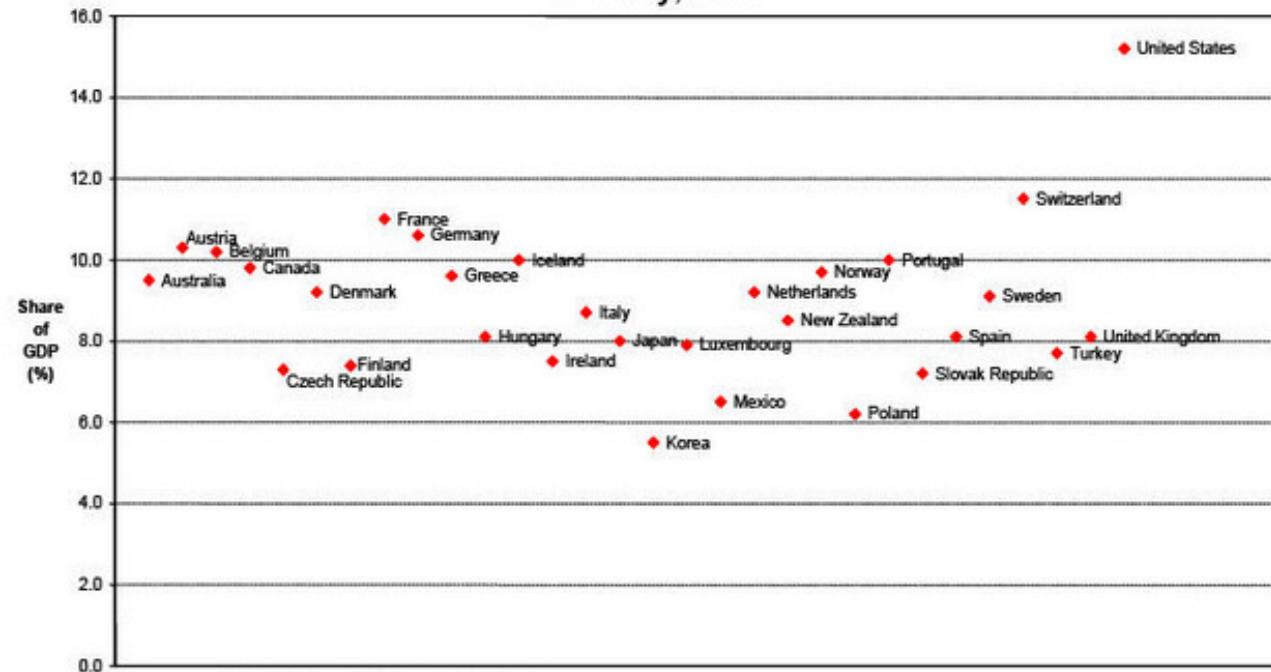


# 不佳的視覺化(續)



# 不佳的視覺化(續)

Chart 2 - Total Expenditures on Health as a Percentage Share of GDP, by OECD Country, 2004

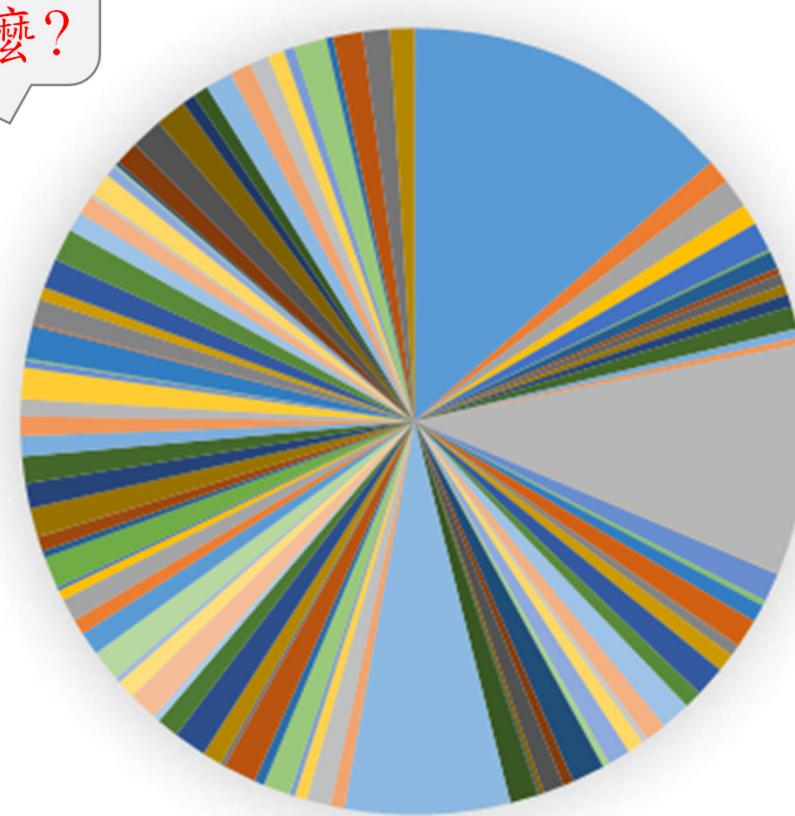


Source: OECD Health Data 2007.

Note: For the United States the 2004 data reported here do not match the 2004 data point for the United States in Chart 1 since the OECD uses a slightly different definition of "total expenditures on health" than that used in the National Health Expenditure Accounts.

## 不佳的視範化(續)

不知在  
視覺化什麼?



# 視覺化圖表

<https://datavizcatalogue.com/>

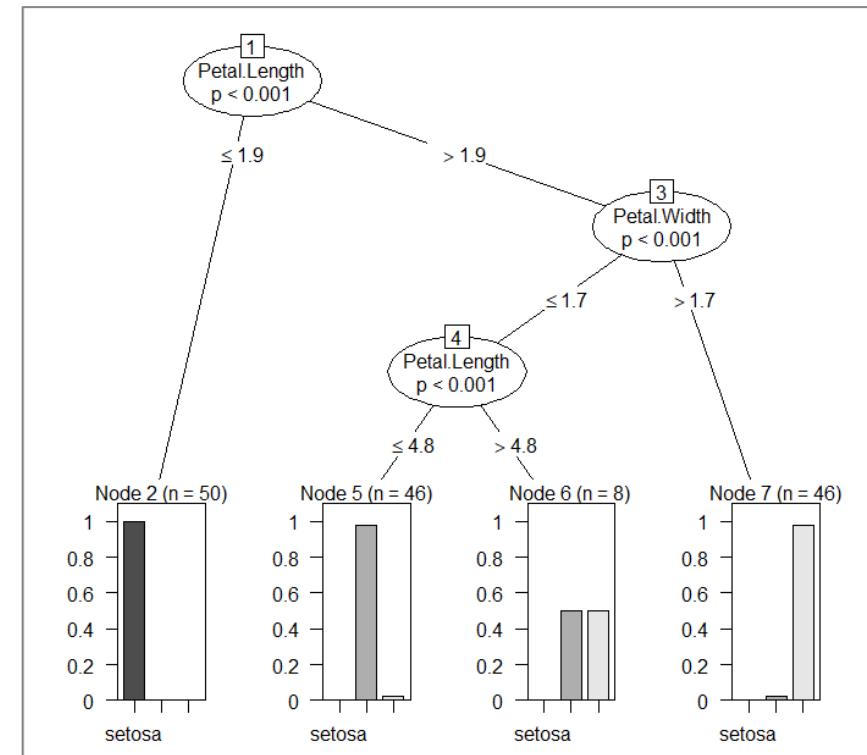


# 決策樹 (Decision tree)

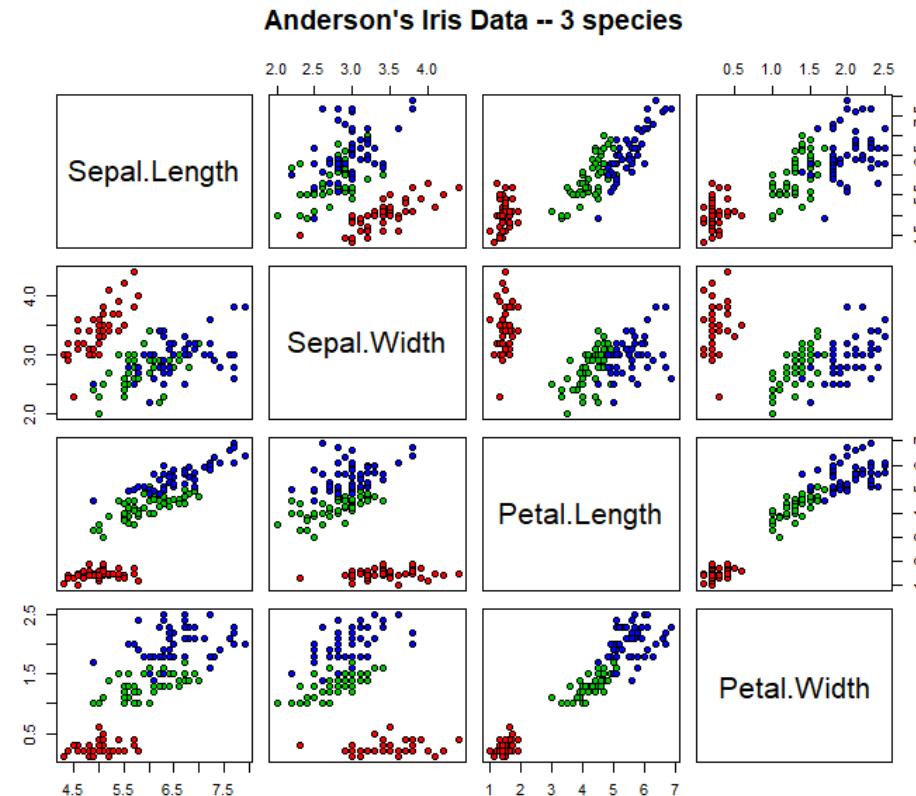
- 決策樹三大元件
  - 樹根 Root
  - 分支 Branch
  - 樹葉 Leaf

iris 資料集-決策樹

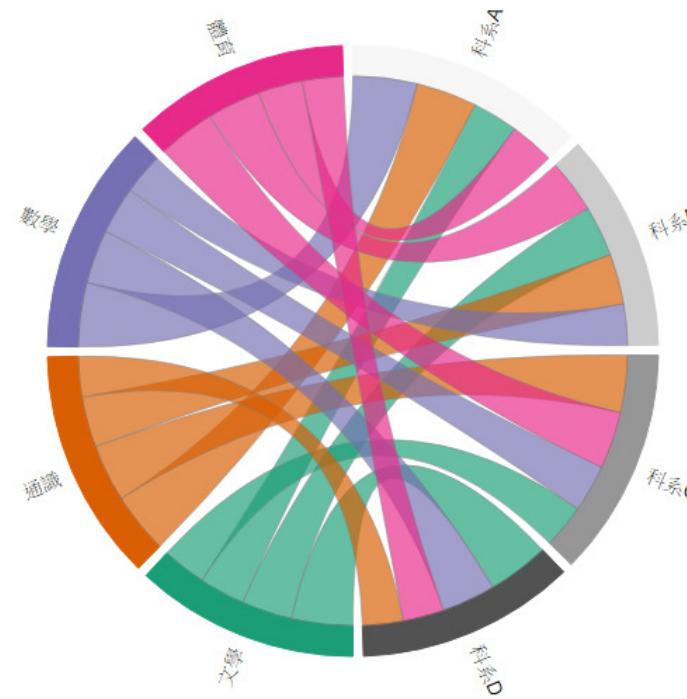
```
library(party)  
plot(ctree(Species ~ . , data=iris))
```



# 散佈圖矩陣 (Scatter plot matrix)

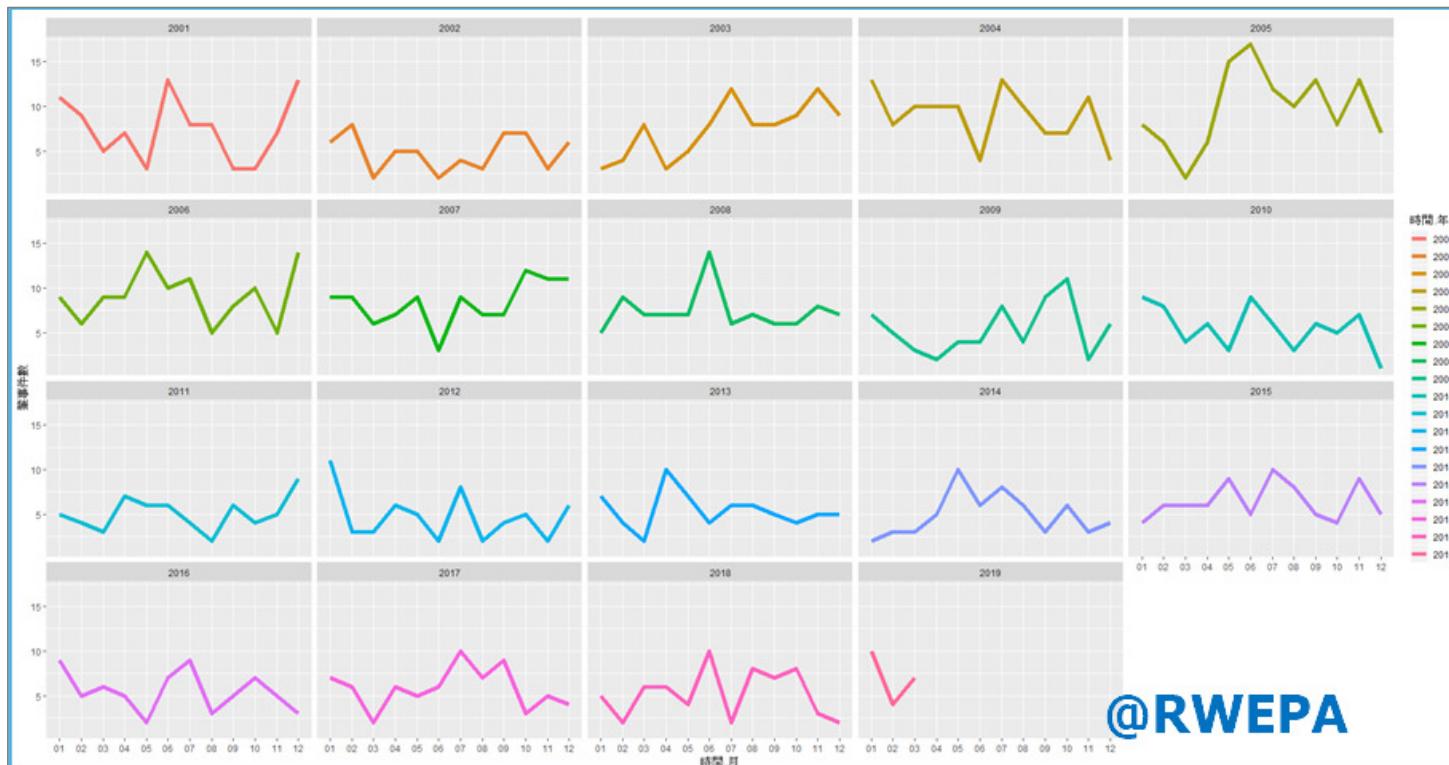


# 弦圖 (Chord diagram)



參考: <http://rwepa.blogspot.com/2019/10/chord-diagram.html>

# 時間群組走勢圖 2001~2019



@RWEPA

參考 [高速公路-交通事故]: <http://rwepa.blogspot.com/2019/05/highway.html>

## 2.大數據軟體實務應用

# Power BI資料視覺化應用

---

# Power BI 簡介與下載

- Microsoft Power BI是Microsoft開發的互動式資料視覺化軟體產品，主要應用於商業智慧。它是Microsoft Power Platform的一部分。[\[https://en.wikipedia.org/wiki/Microsoft\\_Power\\_BI\]](https://en.wikipedia.org/wiki/Microsoft_Power_BI)
- Google → **Power BI 下載**
- <https://www.microsoft.com/zh-tw/power-platform/products/power-bi/downloads>
- 選取 **【進階下載選項】**



# Power BI Desktop

- <https://www.microsoft.com/zh-tw/download/details.aspx?id=58494>



# 下載 PBIDesktopSetup\_x64.exe

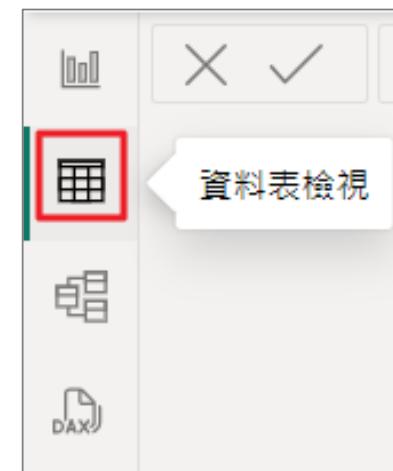
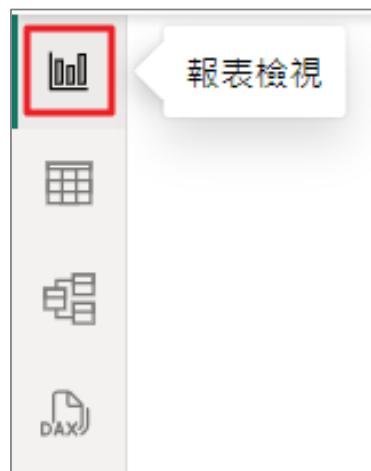
選擇您要的下載 X

<input type="checkbox"/> 檔案名稱	大小
<input type="checkbox"/> PBIDesktopSetup.exe	456.0 MB
<input checked="" type="checkbox"/> PBIDesktopSetup_x64.exe	498.8 MB

下載 大小總計 : 498.8 MB

# Power BI Desktop 四大模組概述

- Power Viewer : 報表檢視 - 提供資料視覺化
- Power Query : 資料檢視 - 提供數據工程 (資料處理)
- Power Pivot : 模型檢視 - 提供關聯式模型
- DAX Query View : DAX 查詢檢視 - 檢視及使用資料分析表示式查詢



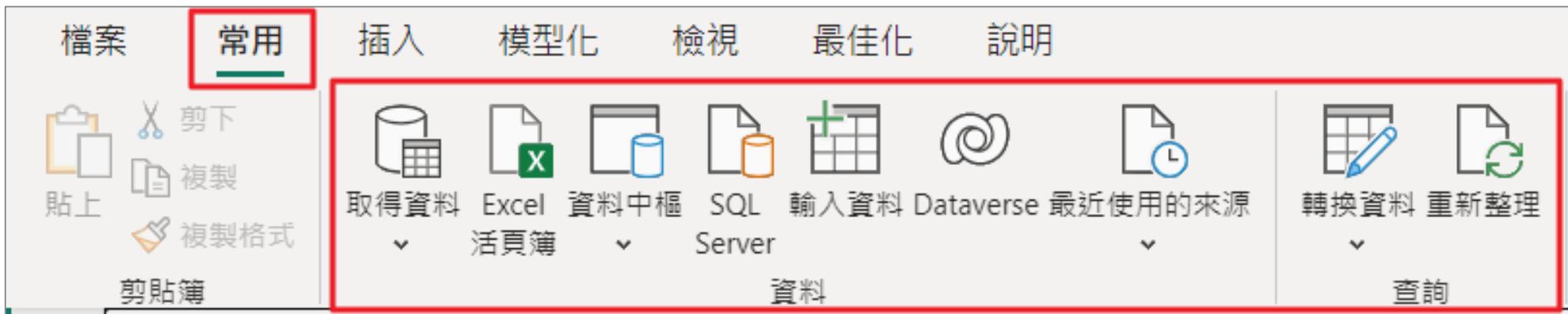
# [1]Power Viewer

- 插入\頁面、視覺效果、AI視覺效果、Power Platform、元素、走勢圖
- 視覺呈現核心，視覺效果模板提供30種以上圖表類型。
- Power BI市集裡面（Power BI visuals），提供超過200個客製化視覺效果模板檔案。性質除屬於一般商業智慧類型之外，也包含大數據分析（部分須透過R語言繪製設計）的模板檔案。



## [2]Power Query

- 常用\資料、查詢
- 蔽集資料及清理資料的重要橋樑跟工具。
- 取得異質資料來源的連接後，使用編輯查詢器進行資料清理作業，以利於Power Pivot及Power View 後續作業。
- 清理資料時，可以不用寫程式，過程還會被記錄下來，中途有錯誤時，仍可折返除錯。



## [3]Power Pivot

- 模型化\關聯性、計算、頁面重新整理、參數、安全性、問與答
- 建立資料模型的核心。利用關聯設定建立資料表之間的關係。
- 同時也是支持資料計算的主要模組。像新增資料行、量值欄位的四則運算及相關量值函數建立等，以利提供設計儀表板上的彈性。

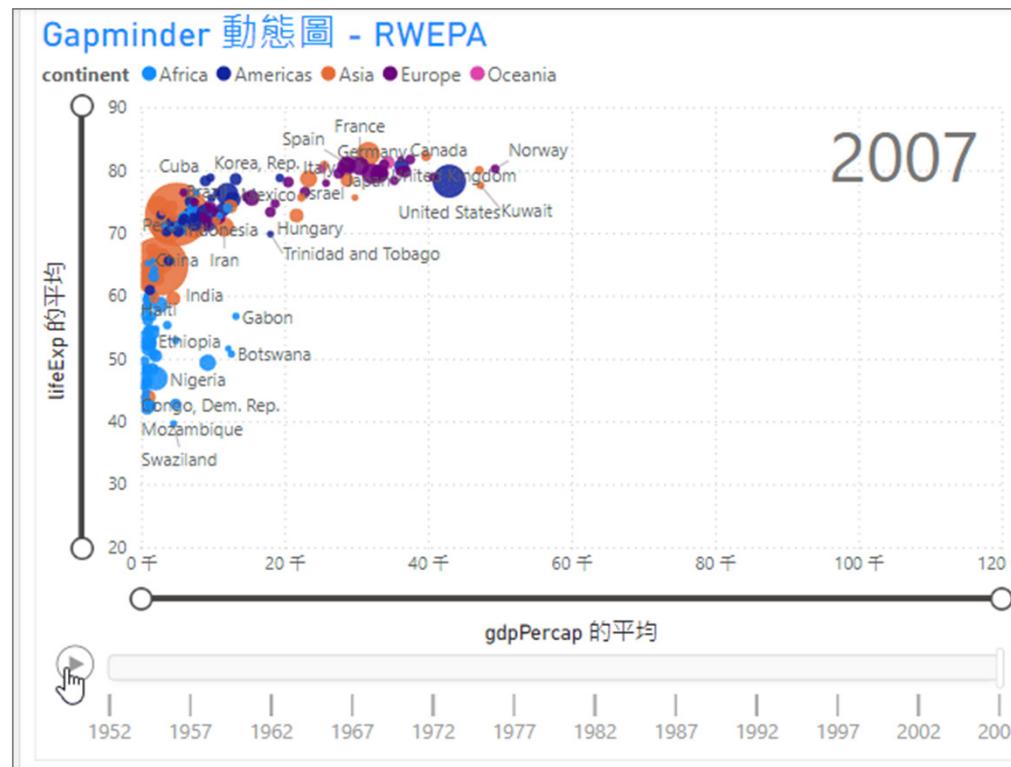


## [4]DAX Query View

- DAX ( Data Analysis Expressions ) - 資料分析語言: DAX和Excel函數寫法非常相似，在Power BI 1.0 ( Excel Power BI ) 及Power BI 2.0 ( Power BI Desktop ) 上都有使用它。
- DAX 語言的兩種類型：
  - 導出新計算欄位：類似 Excel 公式運算式，透過函數組合變化產出。
  - DAX和查詢 ( Query ) 語法陳述式類似：能傳回部分或整個資料結果。

# Power BI + 散佈圖+ 播放軸

- 下載: <https://github.com/rwepa/DataDemo/blob/master/gapminder.csv>
- [https://github.com/rwepa/DataDemo/blob/master/powerbi\\_gapminder.gif](https://github.com/rwepa/DataDemo/blob/master/powerbi_gapminder.gif)



視覺效果

組建視覺效果

資料

搜尋

financials

gapminder

1. continent
2. country
- Σ gdpPercap
- Σ lifeExp
- Σ pop
- Σ year

1. gdpPercap 的平均

2. lifeExp 的平均

Y 軸

圖例

大小

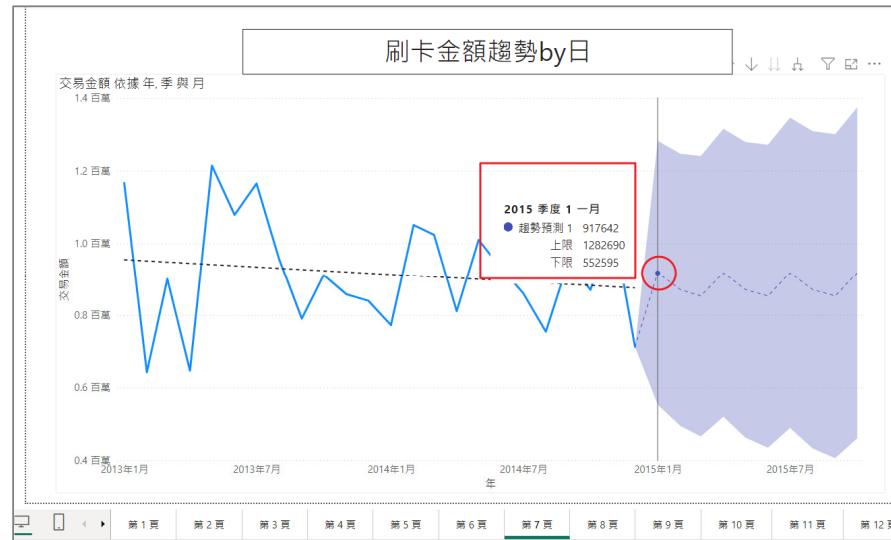
播放軸

year

# Power BI - 刷卡金額趨勢by日, RFM分析

- Power BI

- 下載: [https://github.com/rwepa/DataDemo/blob/master/PowerBI\\_powerview.pbix](https://github.com/rwepa/DataDemo/blob/master/PowerBI_powerview.pbix)
- 學習 【第 7 頁】 - 刷卡金額趨勢by日



- 學習 Power BI 進行RFM分析
- 🌸 YouTube : <https://youtu.be/Lkr9HmzLTtg>

# Tableau資料視覺化應用

---

# Tableau新時代視覺應用介紹

- Tableau Software 美國互動式資料視覺化軟體公司，由史丹佛大學電腦科學系 [Christian Chabot](#) (VizQL技術), [Chris Stolte](#)(資料庫), [Pat Hanrahan](#) (圖形引擎) 於2003年1月在加利福尼亞州山景城創立。該公司目前總部位於美國華盛頓州西雅圖市，專注於商業智慧。
- 2019年8月1日，Salesforce.com 以157億美元收購了Tableau。



Christian Chabot



Chris Stolte



Pat Hanrahan

# Tableau – 視覺化創新應用

- 圖形化操作介面, 使用滑鼠拖曳與選取
- 多樣化的資料(資料庫)連結
- 輕鬆建立視覺化圖表
- 資料過濾與資料層級剖析
- 快速分享研究結果
- API應用 (**R, Python, JavaScript, C, C++, Java, ...**)
  - 使用R統計分析功能
  - 使用R建立機器學習模型
  - 使用Python 建立深度學習模型
  - 將R/Python模型回傳至 Tableau , 進行進階視覺化應用
- 支援 Windows, macOS 作業系統

# Tableau 下載-方法1

- 產品版本 <https://www.tableau.com/zh-tw/products/desktop>

The screenshot shows the Tableau Desktop product page on the salesforce website. The top navigation bar includes links for '為什麼選擇 Tableau' (Why Choose Tableau), '產品' (Products), '解決方案' (Solutions), '資源' (Resources), and '合作夥伴' (Partners). On the right, there are buttons for '立即購買' (Buy Now) and '立即試用' (Try Now). The main content area features a heading 'Tableau Desktop' and the tagline '以業務步調取得見解' (Get insights at your pace). A large orange button labeled '開始免費試用' (Start Free Trial) is highlighted with a red box and a red arrow pointing to it. A yellow speech bubble contains the text '開始免費試用 [下載最新版本]' (Start Free Trial [Download Latest Version]). To the right, there is a visual representation of the Tableau interface showing a data visualization with a scatter plot and a 'Show Me' feature.

# Tableau 下載-方法2

- 選取下載 Tableau Desktop 網頁

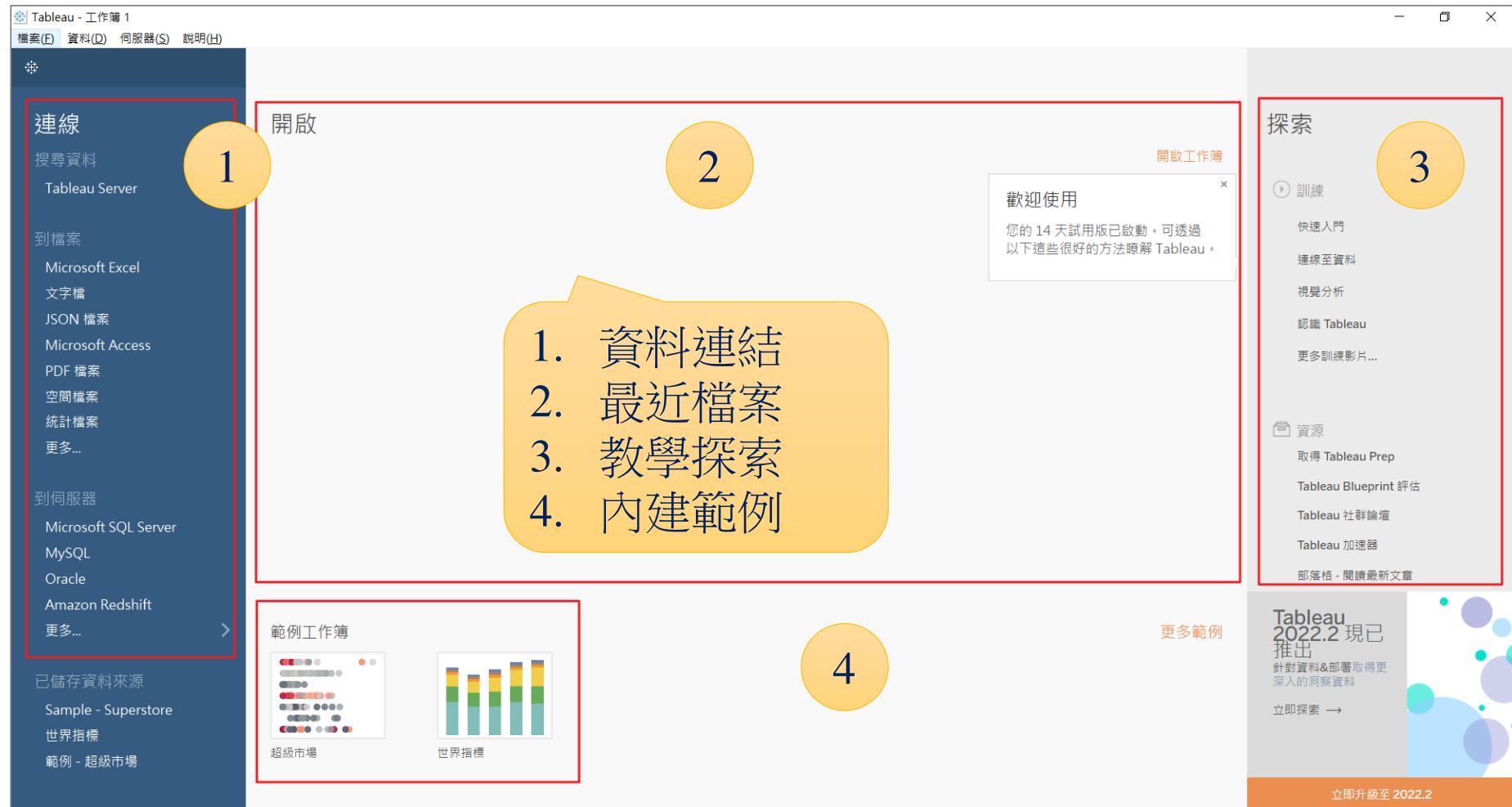
<https://www.tableau.com/zh-tw/support/releases>

The screenshot shows the 'PRODUCT RELEASES' section of the Tableau website. On the left, a sidebar lists product categories: Desktop (highlighted with a red box and circled '1'), Prep, Server, Cloud (Bridge), Advanced Management, Mobile (Blackberry), Intune, and Workspace ONE. The main content area displays a list of releases under the heading '2024.1'. A large orange callout box points to the '2023.2' release entry, which is highlighted with a red box and circled '2'. The entry for '2023.2' includes the text '已發佈 30/01/2024'. Below this, the '2023.1' release is also highlighted with a red box and circled '2', with the text '已發佈 10/01/2024'.

版本	發布日期
2024.1	
2023.3	
2023.3.6	
2023.3.5	
2023.3.4	
2023.3.3	
2023.3.2	已發佈 30/01/2024
2023.3.1	已發佈 10/01/2024

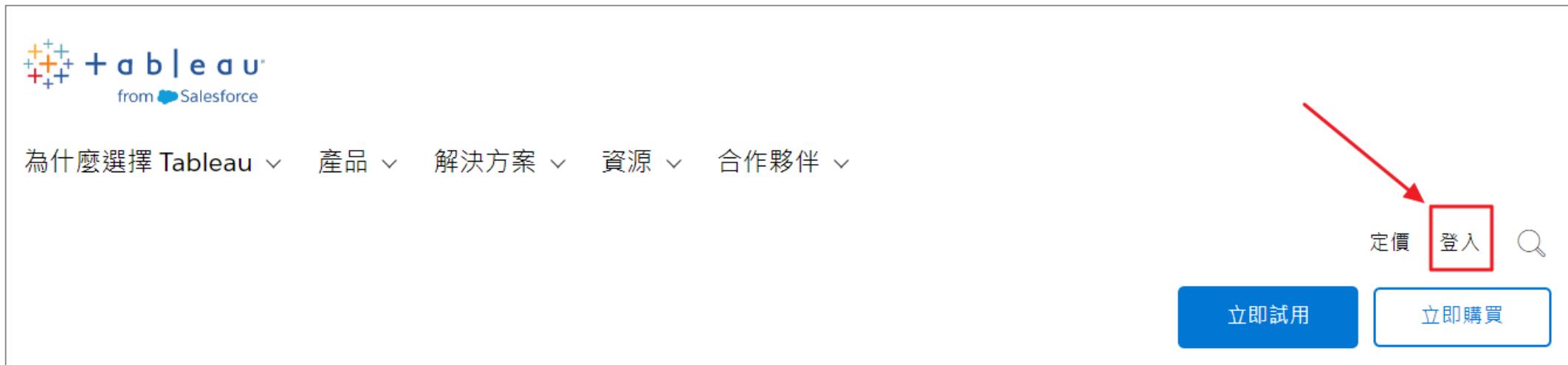
- 考量14天試用版的使用
- 安裝新版本後, 可能無法安裝舊版本
- 先選取【2023.1版本】，到期後再選取  
【2023.2版本】

# Tableau 畫面-中文版



# Tableau Public 線上版→【推薦方法，不用安裝】

- <https://www.tableau.com/zh-tw/community/public>
- 選取 [登入]



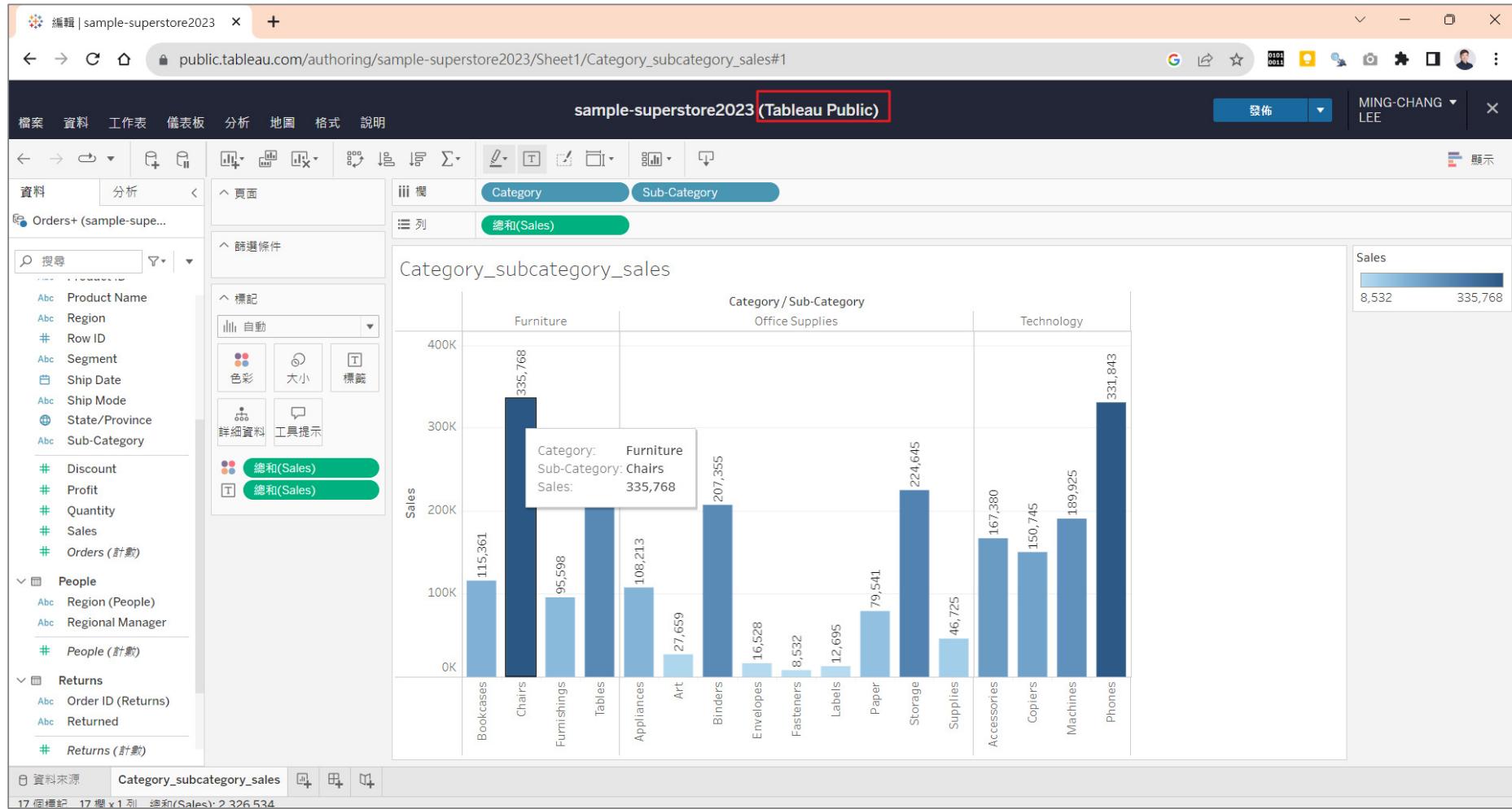
# 登入 TABLEAU PUBLIC



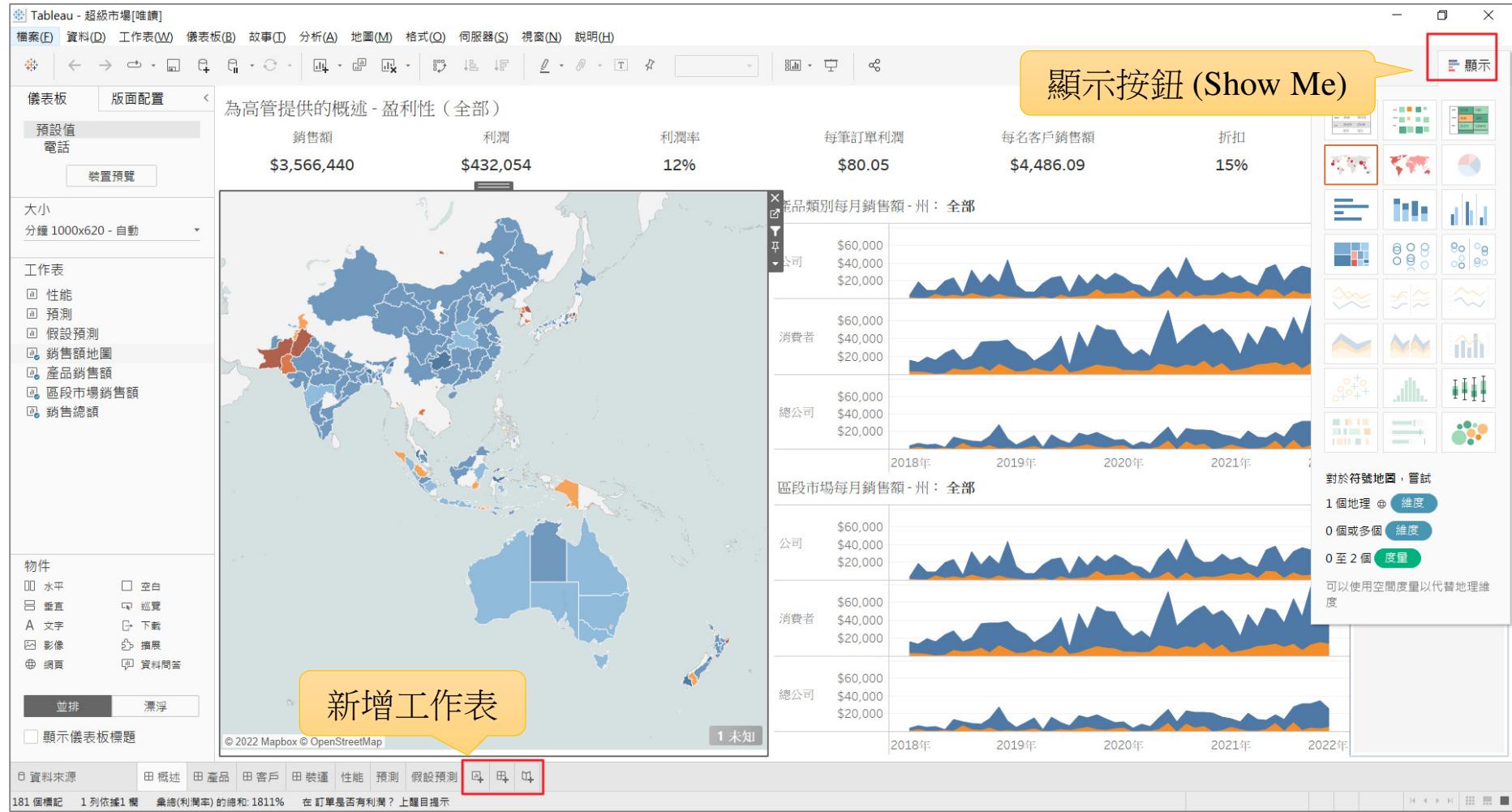
The screenshot shows the Tableau Public login form with the following fields:

- 电子邮件地址 (Email Address)
- 密码 (Password)
- 记住我 (Remember me)
- 登录 (Login) button
- Forgot Password | Create Account links
- 通过 SALESFORCE 登录 (Log in through Salesforce) button

# Tableau Public 線上版



# 內建範例 - 超級市場 (Superstore)



# 工作表 (Worksheet) 視窗

The screenshot illustrates the Tableau Worksheet view with several key components highlighted:

- 功能表列 (Function Bar):** Located at the top left, it includes navigation icons (back, forward, search), a refresh button, and a "復原 (Ctrl+Z)" (Restore) button.
- 工具列 (Tool Bar):** Located above the main workspace, it contains various toolbar icons.
- 資料 / 分析 (Data / Analysis):** A sidebar on the left containing a tree view of data sources and analysis filters.
- 分頁 (Page):** A section in the top header indicating the current page and total pages.
- 過濾 (Filter):** A section in the top header indicating the current filter applied.
- 標記 (Marks):** A panel on the left side of the workspace showing mark settings for color, size, and shape.
- 工作表視圖區 (Worksheet View Area):** The central area displaying a line chart titled "銷售額趨勢預測" (Sales Trend Forecast) showing revenue over time from 2020 to 2023. The chart includes three data series: a blue line for total sales, an orange line for product sales, and a green line for customer sales. A legend on the right identifies these series.
- 智慧圖表 (Smart View):** A panel on the right side showing a grid of various chart types and their corresponding configurations.
- 選取頁面 (Select Page):** A callout pointing to the page selection dropdown in the top header.
- 1. Worksheet 工作表  
2. Dashboard 儀表板  
3. Story 故事:** A callout pointing to the bottom navigation bar which includes links for "資料來源" (Data Source), "概述" (Overview), "產品" (Product), "客戶" (Customer), "裝運" (Shipment), "性能" (Performance), "預測" (Prediction), and "假設預測" (Assumption Prediction).

# Tableau Public

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

智慧製造Tableau運用實務

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)
- 統計品質 (Statistical Quality Control)
- 最優化 (Optimization)

李明昌博士

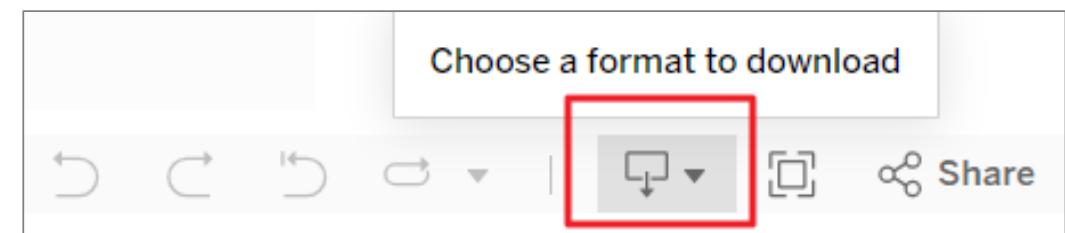
<https://www.youtube.com/c/alan9956>

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

Intelligent\_Manufacturing\_2023

MING-CHANG LEE

☆ 1    ⚡ 242



Intelligent\_Manufacturing\_2023.twbx

# Tableau Public – 31 個圖表範例

Tableau Public - Intelligent\_Manufacturing\_2023

檔案(E) 資料(D) 工作表(W) 儀表板(B) 故事(I) 分析(A) 地圖(M) 格式(O) 同伺服器(S) 視窗(N) 說明(H)

儀表板 版面配置 <

預設值 電話 裝置預覽

大小 自動

工作表

- 1.我的第一個 ...
- 水平長條圖
- 水平長條圖(平...
- 2.篩選(Filter) ...
- 篩選條件(日期...
- 彈性篩選排名
- 3.多維度的巢狀...
- 排序-手動**
- 物件
- 水平容器
- 垂直容器
- A 文字
- 擴展
- 資料問答
- Data Story
- 影像

並排 漂浮

顯示儀表板標題

26.儀表板體驗設計

20.柏拉圖 (Pareto Chart)

Sub-Category	Sales	Cumulative %
Chairs	335,768	14.43%
Phones	331,843	28.70%
Storage	224,645	38.35%
Tables	208,030	47.29%
Binders	203,355	64.37%
Machin..	189,925	71.56%
Access..	167,380	78.04%
Copiers	150,745	83.00%
Bookca..	115,361	87.65%
Applian..	108,213	91.76%
Furnish..	95,598	95.18%
Paper	79,541	97.19%
Supplies	46,725	98.38%
Art	27,659	99.09%
Envelo...	16,528	99.63%
Labels	12,695	100.00%
Fasten...	8,532	

執行 Sales 的總和的總%

主題:我的第1個Tableau 專案  
作者: RWEPA  
日期:2023.10.17  
1.柏拉圖會有互動篩選效果  
2.儀表板\動作 可以新增互動操作

排序-手動

Category Region

Category	Region	Sales
Furniture	West	250K
	Central	180K
	East	150K
	South	220K
Office Supplies	West	230K
	Central	170K
	East	210K
	South	140K
Technology	West	250K
	Central	180K
	East	150K
	South	260K

7.年增率YOY

Year	Sales 中的差異 %
2019	~5%
2020	~5%
2021	~30%
2022	~20%

Order Date

RWEPA

```
library(shiny)
shinyUI(pageWithSidebar(
  headerPanel("shiny"),
  sidebarPanel("RWEPA"),
  mainPanel(ggmap(map.taiwan)))
```

歡迎來到 RWEPA blog, 成立宗旨為提供免費2D/3D互動式繪圖, 資料視覺化, 地理視覺化, 資料, 財務分析, 多變量分析, 問卷調查, 實驗統計, 水文學等。

資料來源

22.燈號圖 (Stoplight Chart) 23.凹凸圖 (Bump Chart) 24.格子圖 (Trellis Chart) 25.銷售子類別格子圖儀表板 26.儀表板體驗設計 27.多重維度分析 28.Tableau與Map整合 31.設置動態參考線

# Python + Anaconda

---

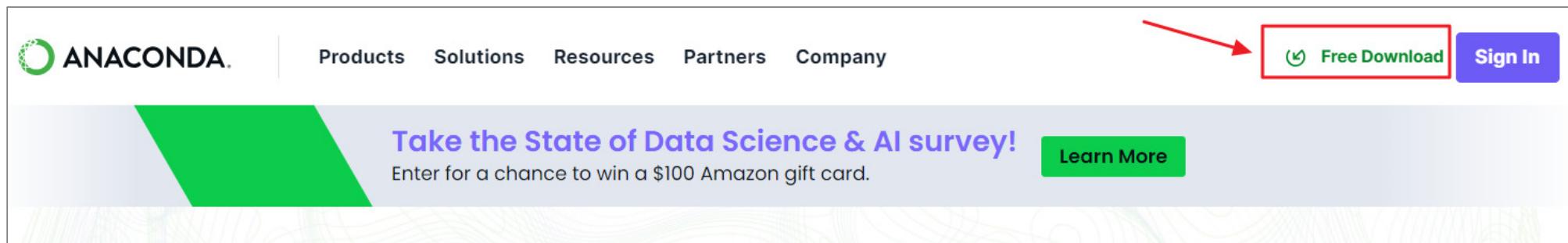
# Anaconda 特性

- Anaconda是一個免費、易於安裝/管理並支援Python語言
- 支援7500個以上的資料科學模組 (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 X和Linux



# Anaconda 下載

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



Skip registration

# Anaconda 下載 (續)



Windows

## Python 3.12

↳ 64-Bit Graphical Installer (912.3M)



Mac

## Python 3.12

↳ 64-Bit (Apple silicon) Graphical  
Installer (704.7M)

↳ 64-Bit (Apple silicon) Command  
Line Installer (707.3M)

↳ 64-Bit (Intel chip) Graphical  
Installer (734.7M)

↳ 64-Bit (Intel chip) Command Line  
Installer (731.2M)



Linux

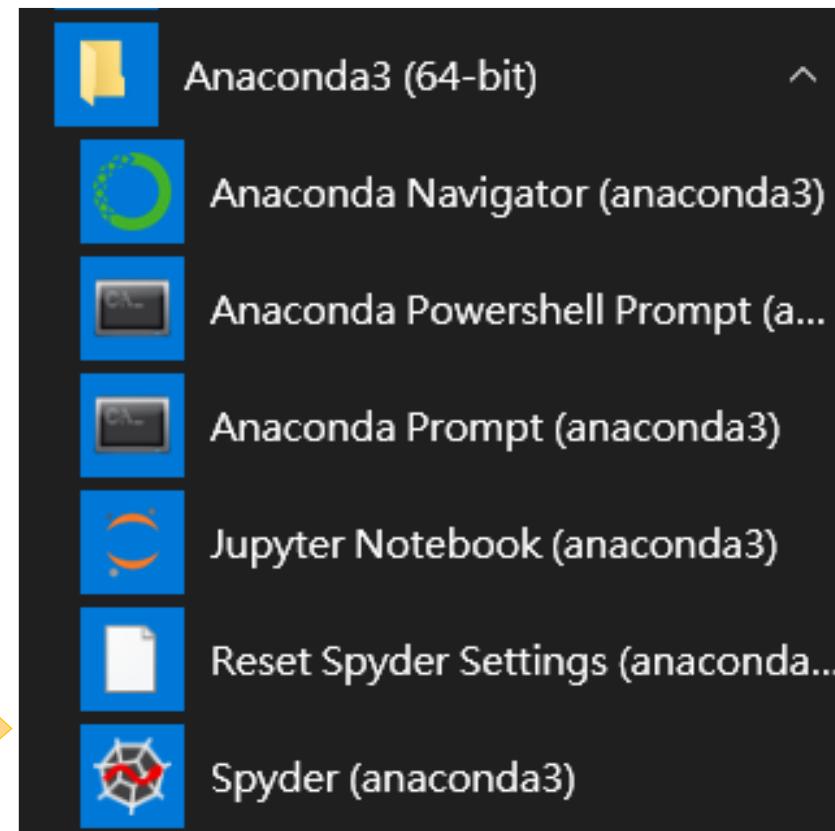
## Python 3.12

↳ 64-Bit (x86) Installer (1007.9M)

↳ 64-Bit (AWS Graviton2 / ARM64)  
Installer (800.6M)

↳ 64-bit (Linux on IBM Z & LinuxONE)  
Installer (425.8M)

# Anaconda3 (64-bit)



已安裝  
6個元件

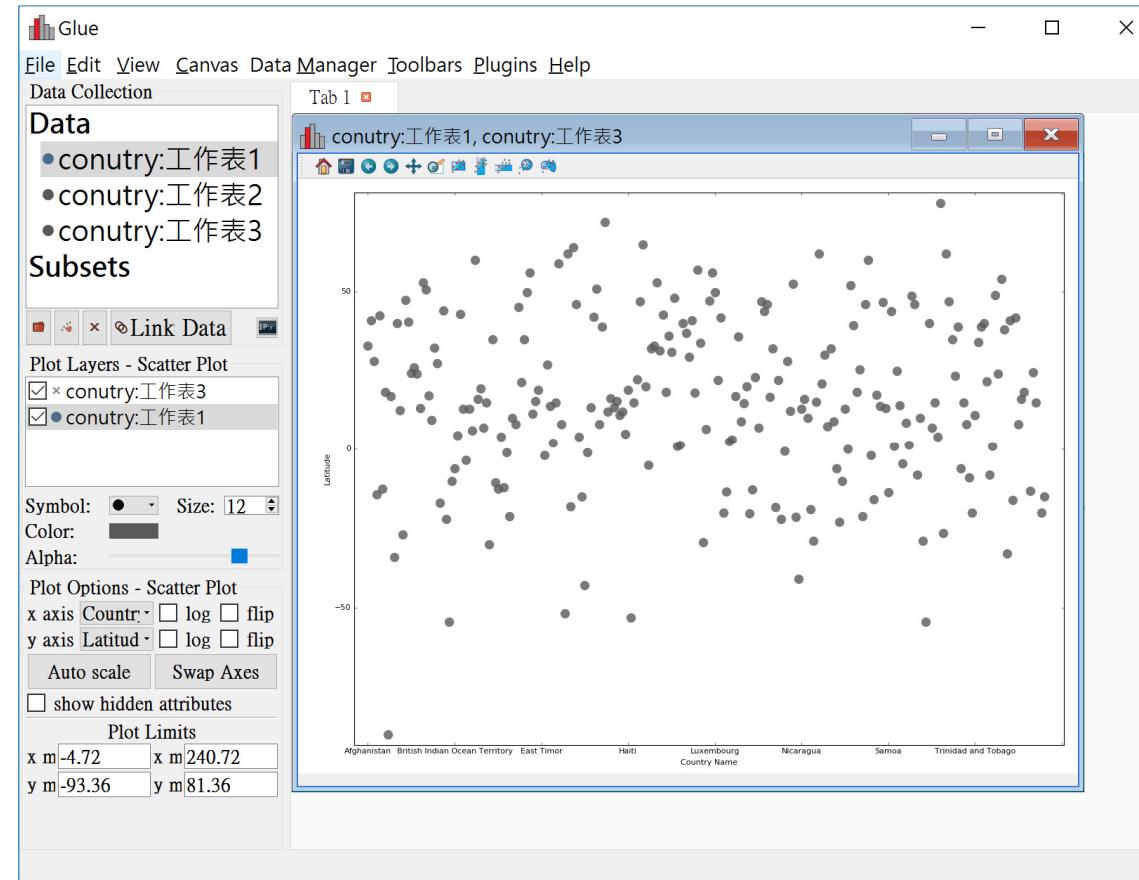
# Anaconda Navigator

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 cards:

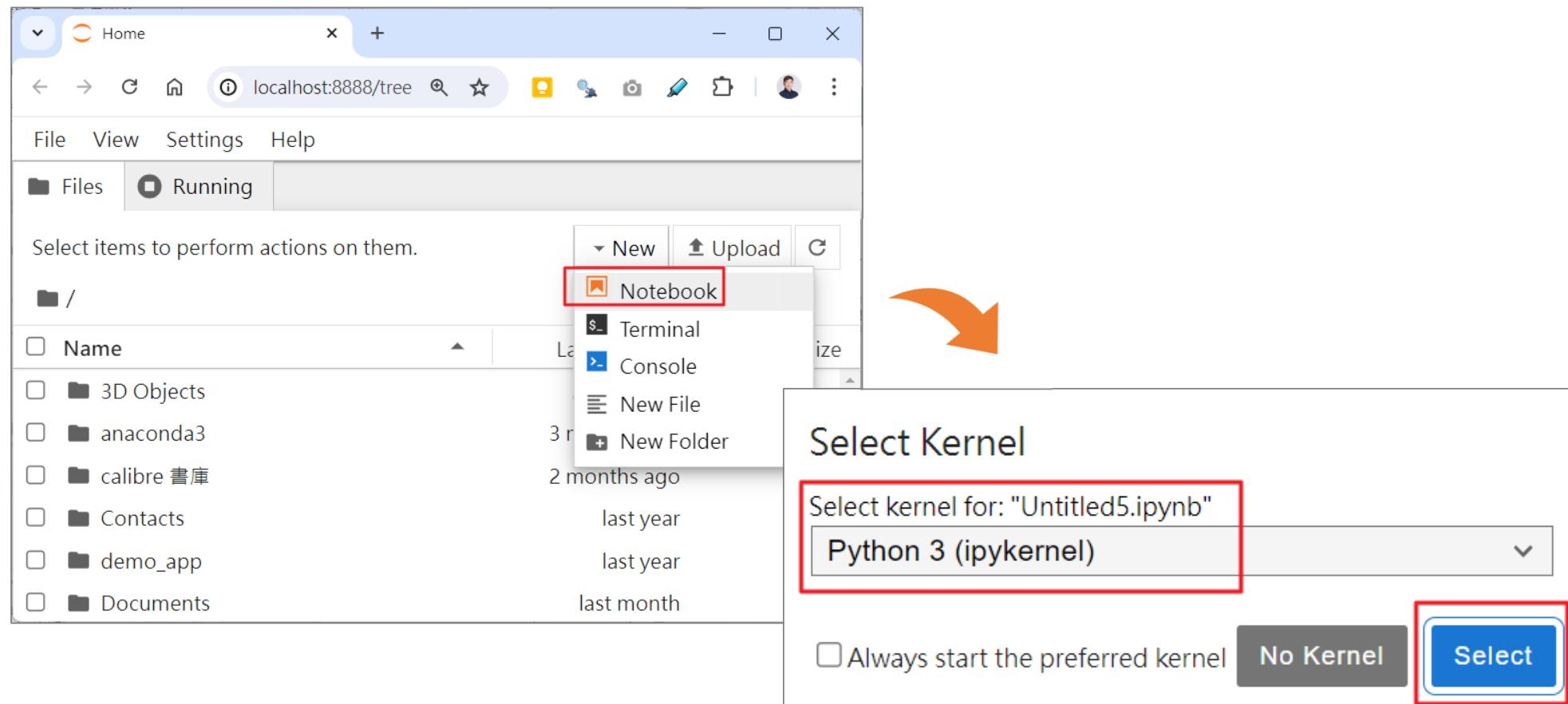
- CMD.exe Prompt (0.1.1): Run a cmd.exe terminal.
- Datalore (DL): Online Data Analysis Tool.
- IBM Watson Studio Cloud: Provides tools for data analysis and machine learning.
- JupyterLab (3.0.14): An extensible environment for computing.
- Jupyter Notebook (6.3.0): Web-based notebook environment.
- Powershell Prompt (0.0.1): Run a Powershell terminal.
- IPy (Qt Console) (5.0.3): PyQt GUI for inline figures.
- Spyder (4.2.5): Scientific Python Development Environment.
- VS Code (1.52.1): Streamlined code editor.
- Glueviz (1.0.0): Multidimensional data visualization.
- Orange 3 (3.26.0): Component-based data mining framework.
- PyCharm Professional: A full-fledged IDE.
- RStudio (1.1.456): A set of integrated tools for R.

A yellow callout bubble in the center says: "不同電腦, 安裝模組可能有差異". A yellow callout bubble at the bottom left says: "注意: 不要在此安裝 RStudio, 可能會有問題?". A yellow callout bubble on the right says: "• Launch 可直接啟動  
• Install 須安裝".

# glueviz 視覺化



# jupyter notebook



# jupyter notebook (續)

The screenshot shows a Jupyter Notebook interface with the following details:

- Title Bar:** Home, Untitled3, localhost:8888/notebooks/Untitled3.ipynb, Python 3 (ipykernel).
- Toolbar:** File, Edit, View, Run, Kernel, Trusted.
- Code Cells:**
  - [1]: `from numpy import *`
  - [2]: `random.rand(4,5)`
  - [2]: 

```
array([[0.61703923, 0.16583443, 0.45807697, 0.56591137, 0.16101254],  
       [0.2957718 , 0.13789455, 0.80815626, 0.37292718, 0.44522573],  
       [0.65767811, 0.85616686, 0.93095361, 0.58563045, 0.73270998],  
       [0.40115409, 0.71177911, 0.05690295, 0.03428265, 0.19975864]])
```
  - [ ]: (empty cell)
- Execution Buttons:** Run button (highlighted with a red box) and Cell.
- Text Annotations:** A yellow callout bubble labeled "1. 輸入程式" points to the first code cell. Another yellow callout bubble labeled "2. 執行" points to the Run button.

# jupyter notebook – 更改預設目錄



```
Anaconda Prompt - jupyter-n × + ▾  
(base) d:\>d:  
(base) d:\>jupyter-notebook  
[I 2024-06-27 20:50:50.186 ServerApp]
```

# jupyter Notebook 快速鍵

- 按 [Esc] cell旁邊為藍色：
  - 按下 **x**：刪除當前選擇的cell
  - 按下 **a**：在當前選擇的上方新增一個cell
  - 按下 **b**：在當前選擇的下方新增一個cell
  - 按下 **Shift + Enter**：執行當前的cell並且選到下一個cell
  - 按下 **Ctrl + Enter**：執行當前cell
  - 按下 **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>

### 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

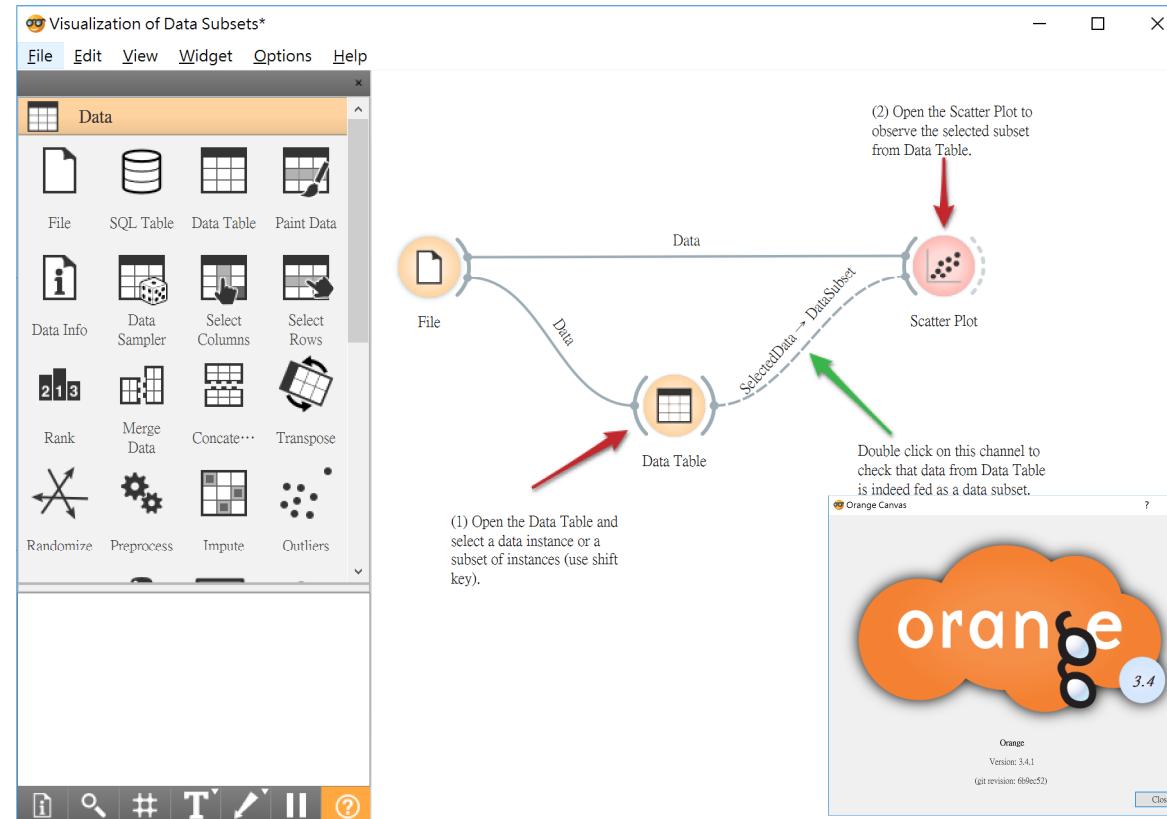
下載: [https://github.com/rwepa/DataDemo/blob/master/Python\\_Programming\\_Lee\\_ipynb.zip](https://github.com/rwepa/DataDemo/blob/master/Python_Programming_Lee_ipynb.zip)



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

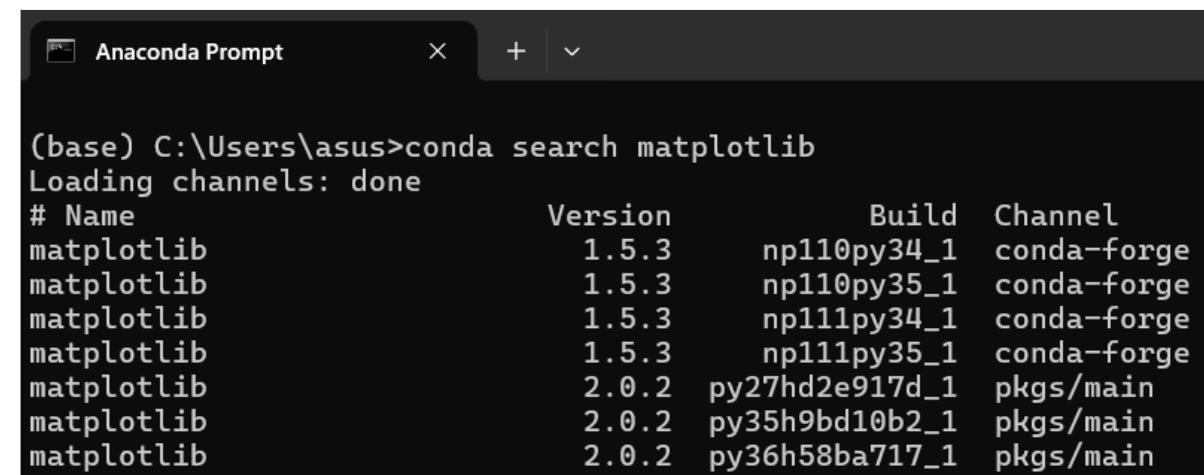
# Orange

- 使用命令提示列 開啟 Orange: python -m Orange.canvas



# Anaconda 模組管理

- 顯示已安裝模組  
    `conda list`
- 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

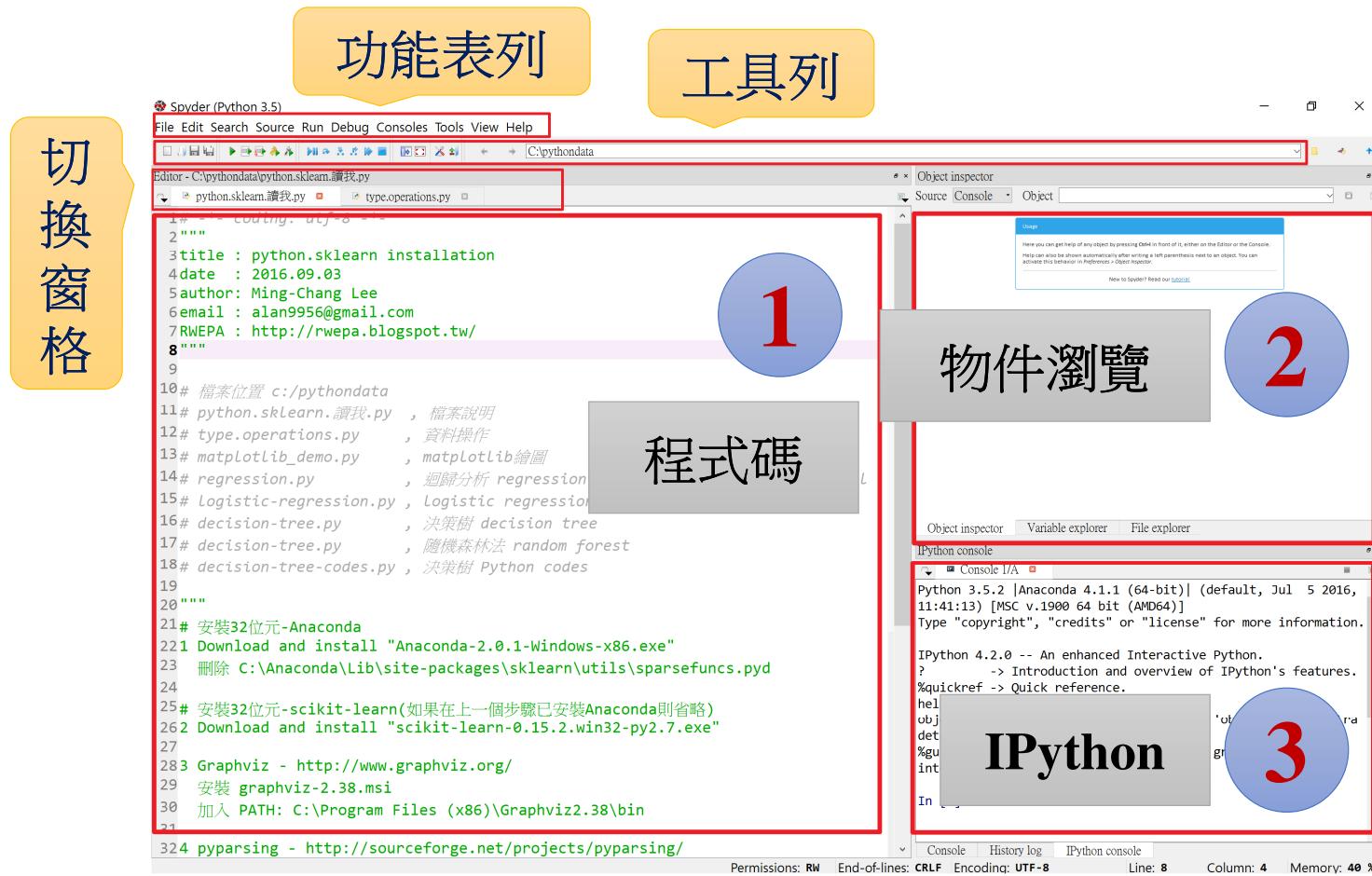


```
(base) C:\Users\asus>conda search matplotlib
Loading channels: done
# Name          Version   Build  Channel
matplotlib      1.5.3     np110py34_1  conda-forge
matplotlib      1.5.3     np110py35_1  conda-forge
matplotlib      1.5.3     np111py34_1  conda-forge
matplotlib      1.5.3     np111py35_1  conda-forge
matplotlib      2.0.2     py27hd2e917d_1  pkgs/main
matplotlib      2.0.2     py35h9bd10b2_1  pkgs/main
matplotlib      2.0.2     py36h58ba717_1  pkgs/main
```

# Spyder 軟體簡介

---

# Spyder 畫面



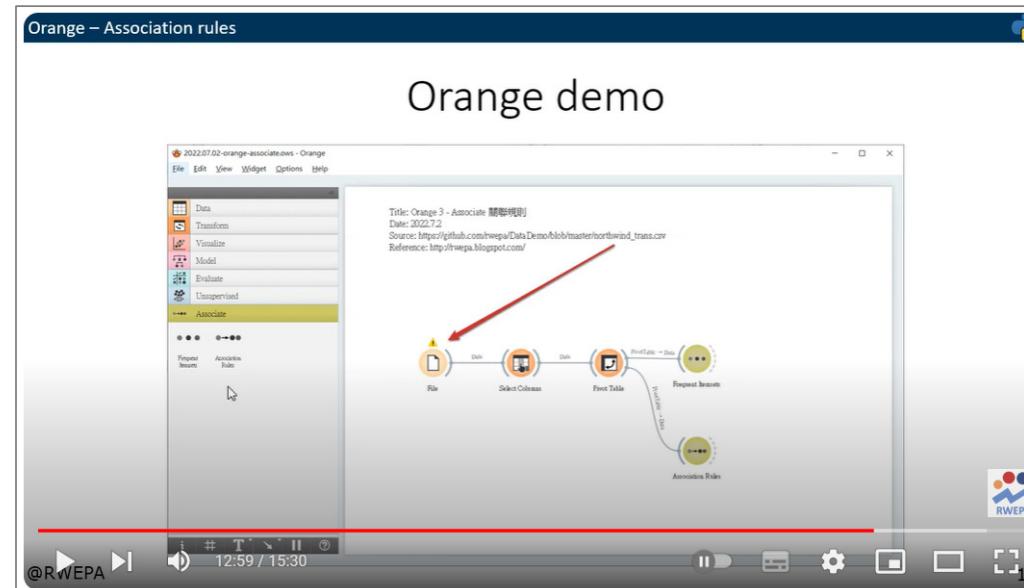
# 免費Orange視窗化資料建模

---

# Orange - Scatter Plot

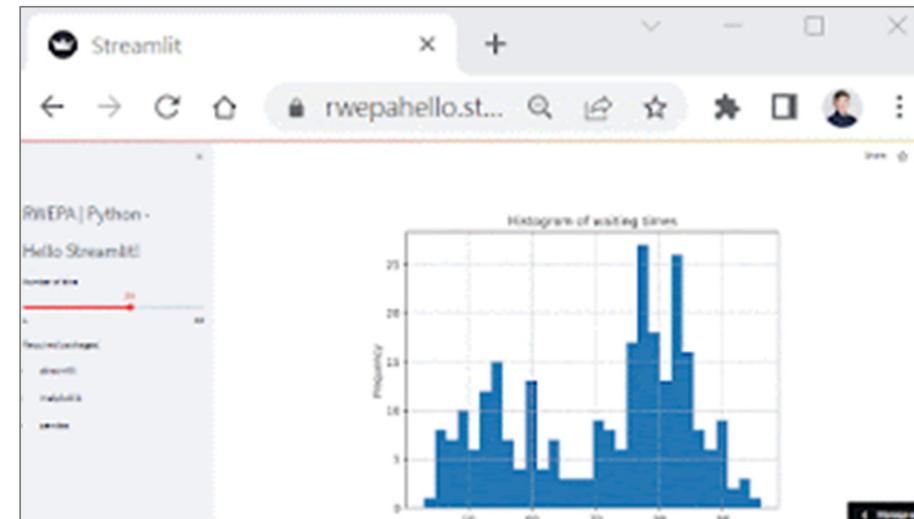
- Python Orange - 關聯規則教學

- YouTube: <https://youtu.be/rh5GxJamtNg>
- 🌸【謝謝訂閱、按讚、分享與開啟小鈴鐺💡】
- LINK: <https://rwepa.blogspot.com/2022/07/python-orange-associate-tutorial.html>
- PDF: [https://github.com/rwepa/orange3\\_associate/blob/main/2022.07.02-orange-associate.pdf](https://github.com/rwepa/orange3_associate/blob/main/2022.07.02-orange-associate.pdf)



# Streamlit 簡介

- 在 Python 環境中, Streamlit 模組提供資料分析的網頁服務框架, 內建常用的網頁操作控制項, 配合 pandas 資料分析模組, matplotlib 等視覺化模組, 可以快速建立互動式網頁服務與企業級儀表板, 其功能與 R 語言的 shiny 套件相似。
- YouTube: <https://youtu.be/FW-dl-fILvk>
- LINK: <https://rwepa.blogspot.com/2023/01/python-streamlit-dashboard.html>
- Cloud 佈署成果: <https://rwepahello.streamlit.app/>



# 登山路線視覺化分析平台(Streamlit)

- 🌸 YouTube (包括中文字幕) : [https://youtu.be/-\\_zghs2qrlg](https://youtu.be/-_zghs2qrlg)
- LINK: <https://rwepa.blogspot.com/2023/08/visualization-climbing-routes-with.html>
- 🌸 系統展示: <https://rwepa-climb.streamlit.app/>



# R + RStudio

---

# R官方網頁



The screenshot shows the official website for The R Project for Statistical Computing. On the left, there's a navigation menu with links like [Home], Download, CRAN, R Project, About R, Logo, Contributors, What's New?, and Reporting. A yellow speech bubble highlights the 'Download' link. To the right of the menu, there's a large 'R' logo and some descriptive text. The main title 'The R Project for Statistical Computing' is centered above a 'Getting Started' section.

[Home]

Download

CRAN

R Project

About R

Logo

Contributors

What's New?

Reporting

下 載

繪 圖

# The R Project for Statistical Computing

## Getting Started

## 統計計算

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

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

# R-下載

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

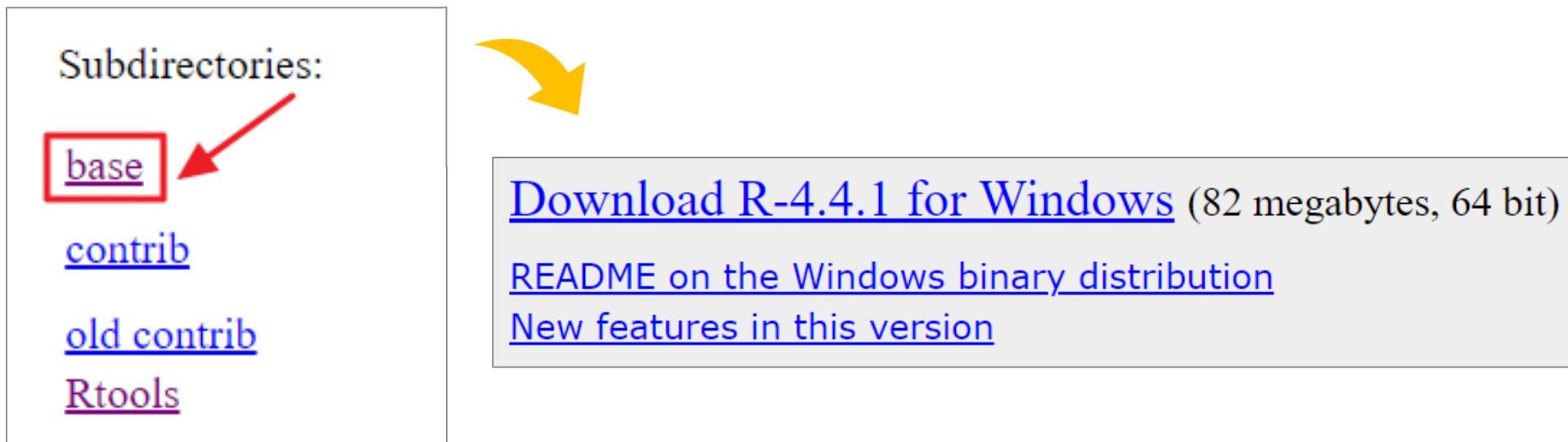


- [Download R for Linux \(Debian, Fedora/Redhat, Ubuntu\)](#)
- [Download R for macOS](#)
- [Download R for Windows](#)



## R-下載 (續)

- 選取 base → 下載 [R-4.4.1-win.exe]



- R安裝路徑: 保留原路徑,不要修改
- 安裝參考說明, 2006  
[https://github.com/rwepa/DataDemo/blob/master/windows\\_intall\\_R.pdf](https://github.com/rwepa/DataDemo/blob/master/windows_intall_R.pdf)

# R執行畫面

The screenshot shows the RGui (64-bit) interface. The R Console window displays the R version information and a warning about the license. The R Graphics window titled 'R Graphics: Device 2 (ACTIVE)' shows a plot of 100 random uniform numbers (runif(100)). Below the plot, the command `> plot(runif(100), type="l")` is visible.

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

**demo(graphics)**

**demo(persp)**

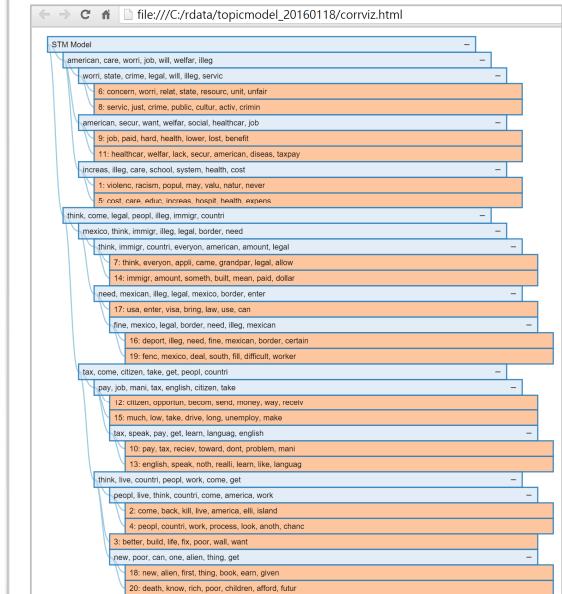
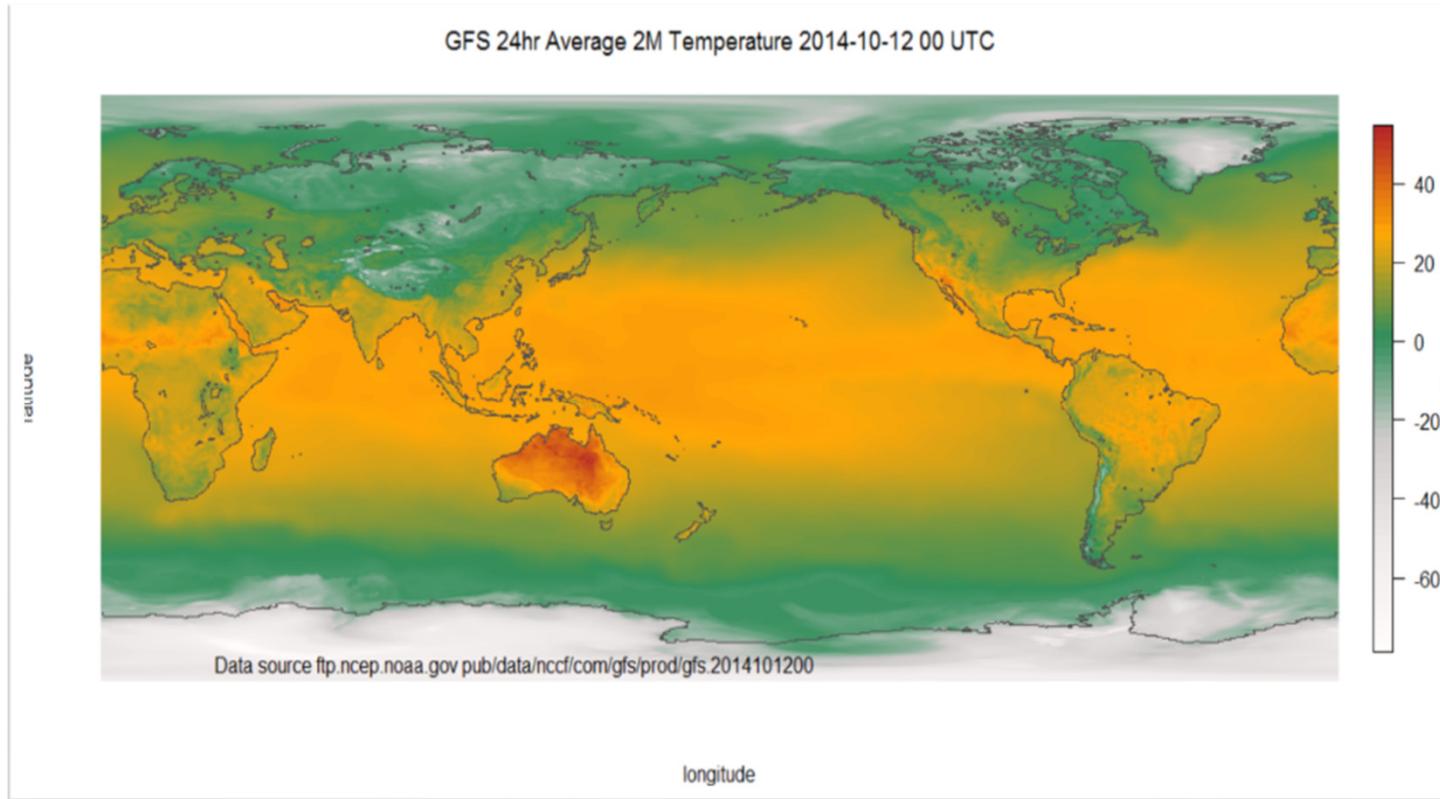
**大小寫須一致**

**參考文獻:  
citation()**

**安裝R, 登入名稱:**

1. 不要使用空格
2. 不要使用中文字型

# Rstudio 整合式開發環境



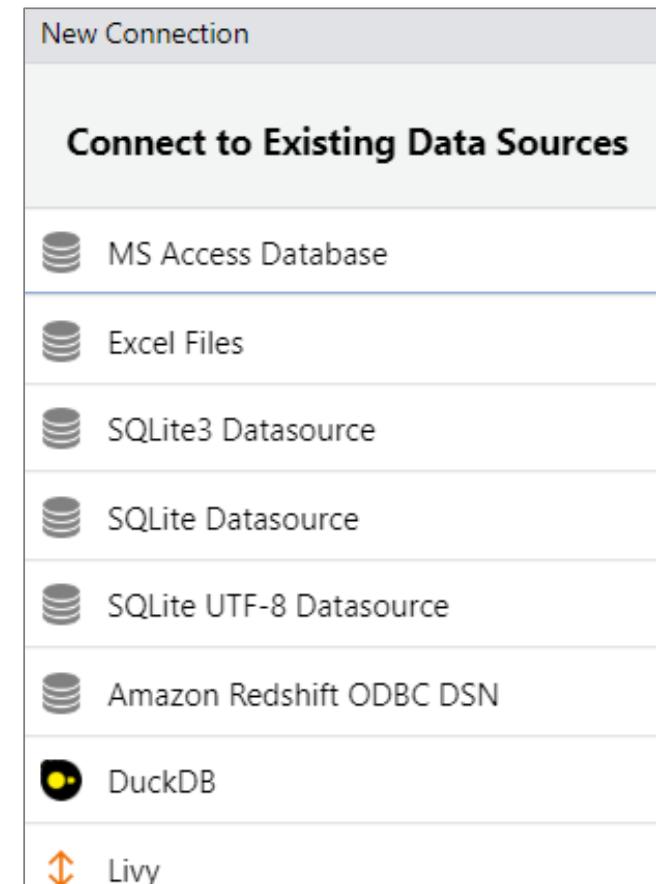
主題模型

視覺化應用

(全球2M氣溫圖)

# RStudio - 特性

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



# RStudio 下載

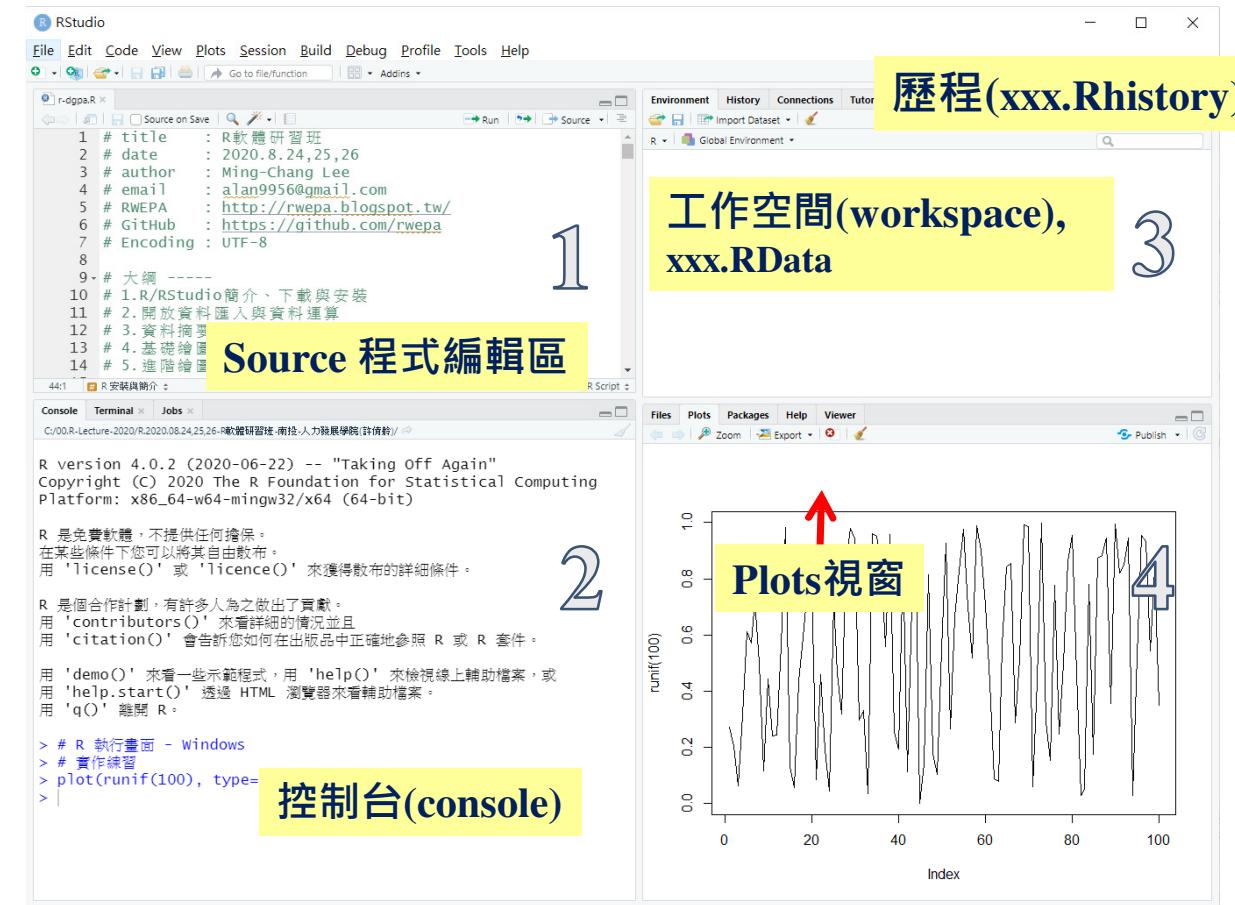
• ~~<http://www.rstudio.com/>~~



<https://posit.co/>

The screenshot shows the posit.co homepage. At the top, there's a navigation bar with links: PRODUCTS (highlighted with a red box), SOLUTIONS, LEARN & SUPPORT, EXPLORE MORE, PRICING, and a search icon. A prominent blue button labeled 'DOWNLOAD RSTUDIO' is also highlighted with a red box and has a red arrow pointing towards it. Below the navigation, the text 'Deployment made easy' is displayed in a large, bold font. Underneath, a paragraph explains the deployment capabilities: 'Deploy all of your work, including Shiny, Streamlit, and Dash applications. Models. Quarto documents. Jupyter Notebooks. Reports. Dashboards. Even APIs. With customizable access controls and authentication options that make IT happy.' To the right of the text, there's a photograph of a person working at a computer, with the monitor displaying a dashboard of data visualizations.

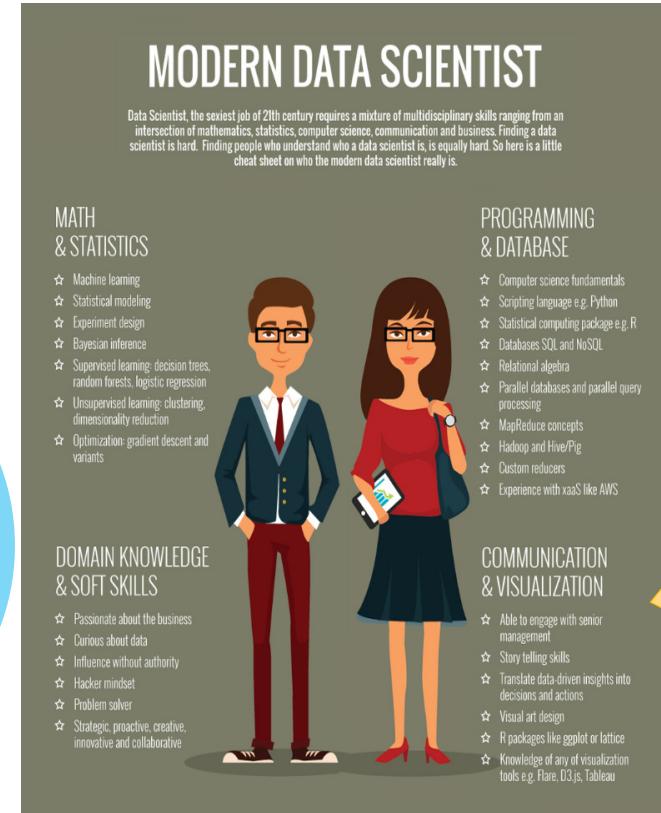
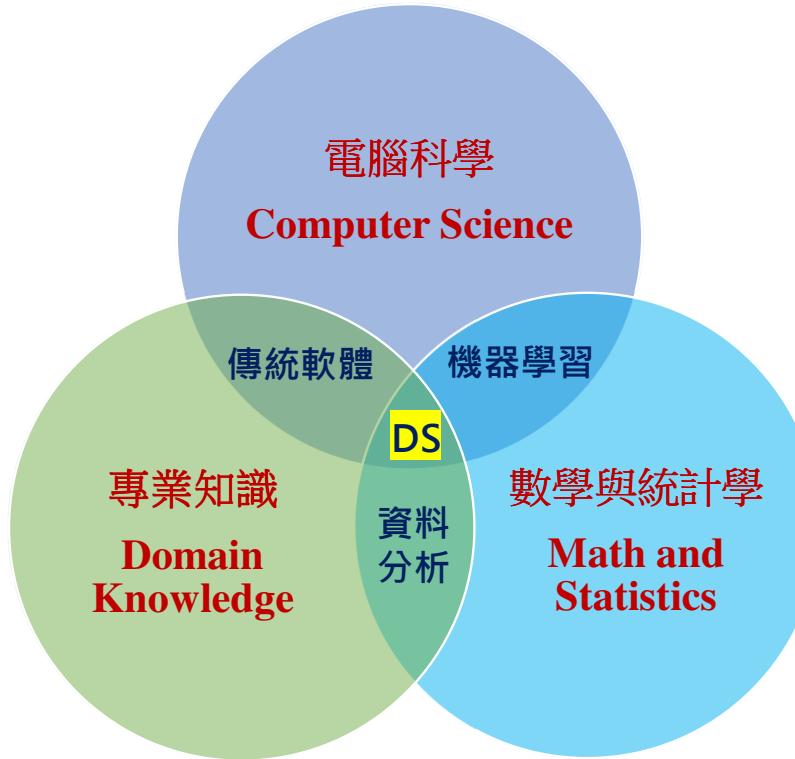
# R/RStudio環境的基礎觀念



Ctrl + Shift + F10: 重新啟動R

### 3. 職場技能

# 現代資料科學家 (Modern Data Scientist) 四項技能



4.溝通與視覺化

參考: <https://towardsdatascience.com/a-long-term-data-science-roadmap-which-wont-help-you-become-an-expert-in-only-several-months-4436733e63ff>

# 開放資料

- <https://rwepa.blogspot.com/2019/09/dataset.html>
- DataDemo 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/>



# 職場技能

- 畢業學分數：128學分 → 找出最有興趣的課程與方向
- 強化統計, 數學基礎課程
- 加強機器學習, 人工智慧等課程之學習
- Power BI, Tableau, Python, R 之學習
- 善用 Open Data 建立資料分析專案 → 置於 GitHub
- 多看看 RWEPA 網站免費資源 + YouTube 影片
  - <https://www.youtube.com/@alan9956>

# 謝謝您的聆聽

## Q & A



李明昌

*alan9956@gmail.com*

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