

# 產業人才巨量資料分析- 台灣策略與績效管理學會雲端修煉講座

## 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](mailto:alan9956@gmail.com)  
<http://rwepa.blogspot.com/>

# 大綱

- 個人簡介
- 產業人才巨量資料分析需求
- 分析工具 – Excel, **R**, Python, Julia
- 巨量資料分析的心法
- R 網頁服務案例
- Q & A

台灣策略與績效管理學會

TAIWAN STRATEGY AND PERFORMANCE  
MANAGEMENT ASSOCIATION



雲端修煉講座

## 產業人才巨量資料分析

時間 : 2021/8/26(四) 晚上20:30-21:30

地點 : Webex 線上教室

E-MAIL: [alan9956@gmail.com](mailto:alan9956@gmail.com)

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

### 主講人 :

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

李明昌博士



■李明昌博士擅長於R/Python/Julia/SQL等語言程式設計與應用、資料視覺化、機器學習、統計品管與最佳化等研究方法與應用。不但在國內外知名期刊發表其研究成果，更多次擔任International Journal of Production Research、Transactions on Fuzzy Systems、Journal of the Operational Research Society、International Journal of Advanced Manufacturing Technology、Journal of Cleaner Production等SCI期刊評審委員。

# 個人簡介

# @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
- 經歷：
  - 育達科技大學 資訊管理系(所) 專任助理教授
  - 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
  - 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
- 國內外各大專院校、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營單位演講達280餘場，2500小時以上。
- 連絡資訊：[alan9956@gmail.com](mailto:alan9956@gmail.com)



## @RWEPA



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

2021年7月17日 星期六

## R data analysis, visualization, machine learning course

近日公佈二門課程，歡迎各位R友參考。

### 課程1：R入門資料分析與視覺化應用

時間：7小時28分鐘，附中文字幕

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

RWEPA 搜尋此網誌 (例: task)

搜尋

- GitHub DataDemo
- 關於作者
- R與實驗設計應用影片(6)
- ★★★R入門資料分析與視覺化(付費,中文字幕)
- ★★★R商業預測與應用(付費,中文字幕)
- iPAS-R-tutorial
- iPAS-Python-tutorial
- R教學-基礎篇/程式碼(免費)
- Python程式設計PDF(免費)
- ★R 4.1.1-Wndows下載
- ★RStudio-1.4.1717下載
- R-bloggers
- Python-blogger
- Bioconductor



# 產業人才巨量資料分析需求

# 產業人才巨量資料分析流程

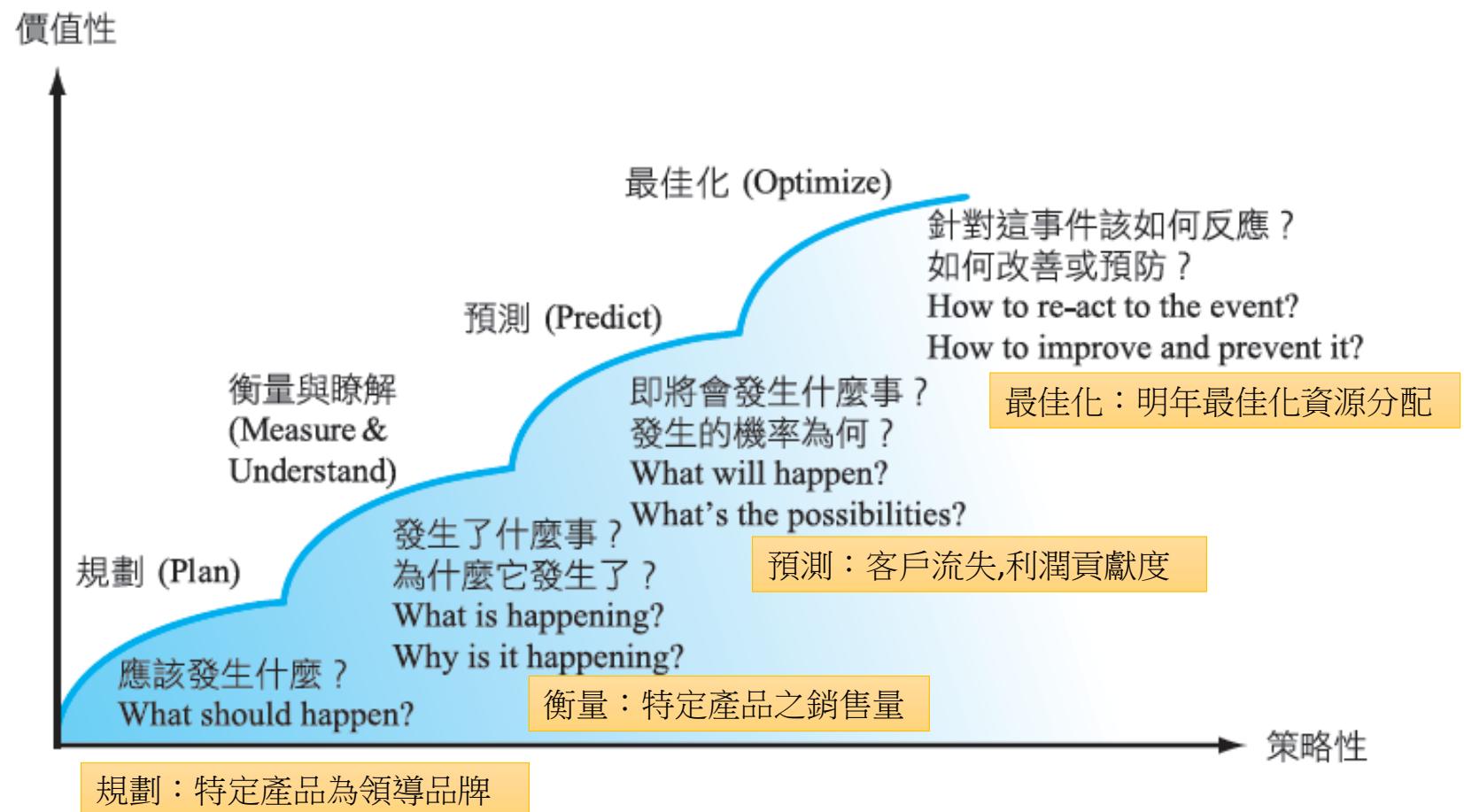


台灣策略與績效管理學會

TAIWAN STRATEGY AND PERFORMANCE  
MANAGEMENT ASSOCIATION



# 資料分析應用層級



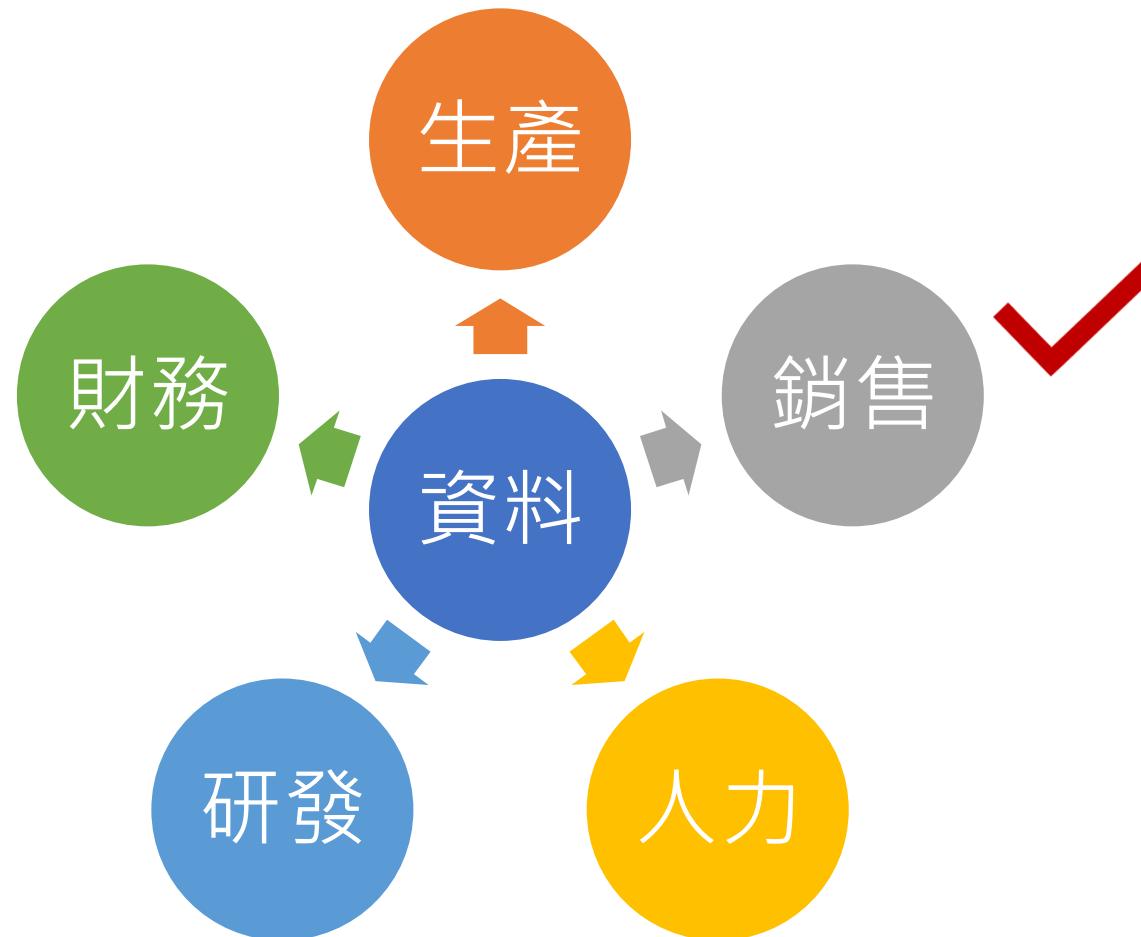
# 巨量資料分析技能



- 數學,統計
- 專業知識,軟體技術
- 程式,資料庫
- 溝通,視覺化

參考 <https://www.proschoolonline.com/blog/data-science-skills>

# 巨量資料分析產業應用



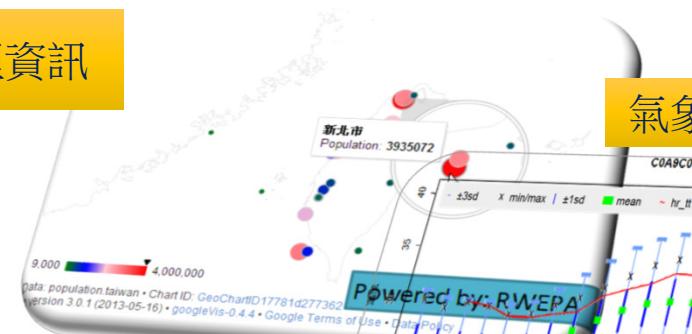
# 銷售應用-RFM分析

- 在動態顧客行為方面，RFM分析（Recency, Frequency, Monetary analysis）呈現忠誠度和滿意度相關的行為變數。
- 使用資料挖礦的集群(Clustering)技術：運用動態顧客資料庫做為分析，其結果比靜態的人口統計變數（例：僅考慮男性，女性）來區隔市場更具有價值性。
  - **消費時間 (Recency)**：該顧客最後下訂單的時間為何？分析最近購買的顧客是否傾向再度購買？
  - **消費頻率 (Frequency)**：該顧客下了多少次的訂單？分析較常購買的顧客是否會較易回應？
  - **消費金額 (Monetary)**：該顧客總共花費金額？分析金額較多的顧客未來更會消費的可能性？

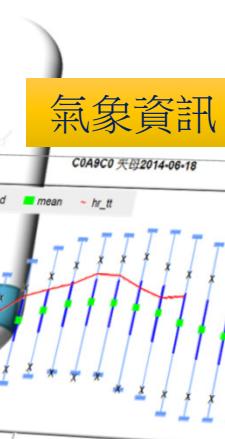
# 資料視覺化應用

R + shiny → 互動式網頁

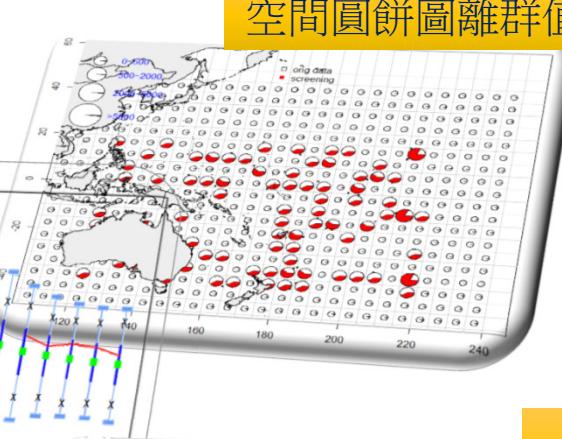
地理資訊



氣象資訊



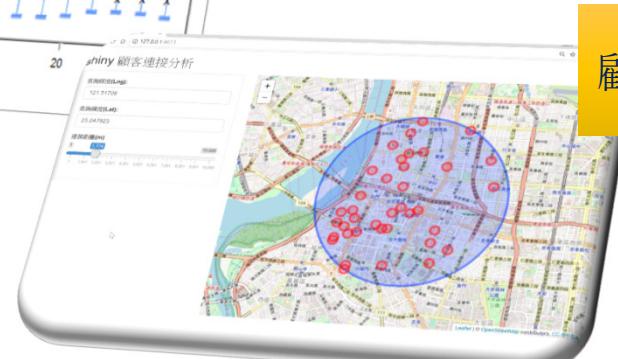
空間圓餅圖離群值分析



保險預測



顧客連結資訊



網頁呈現

## 中央氣象局 1,600 萬筆資料

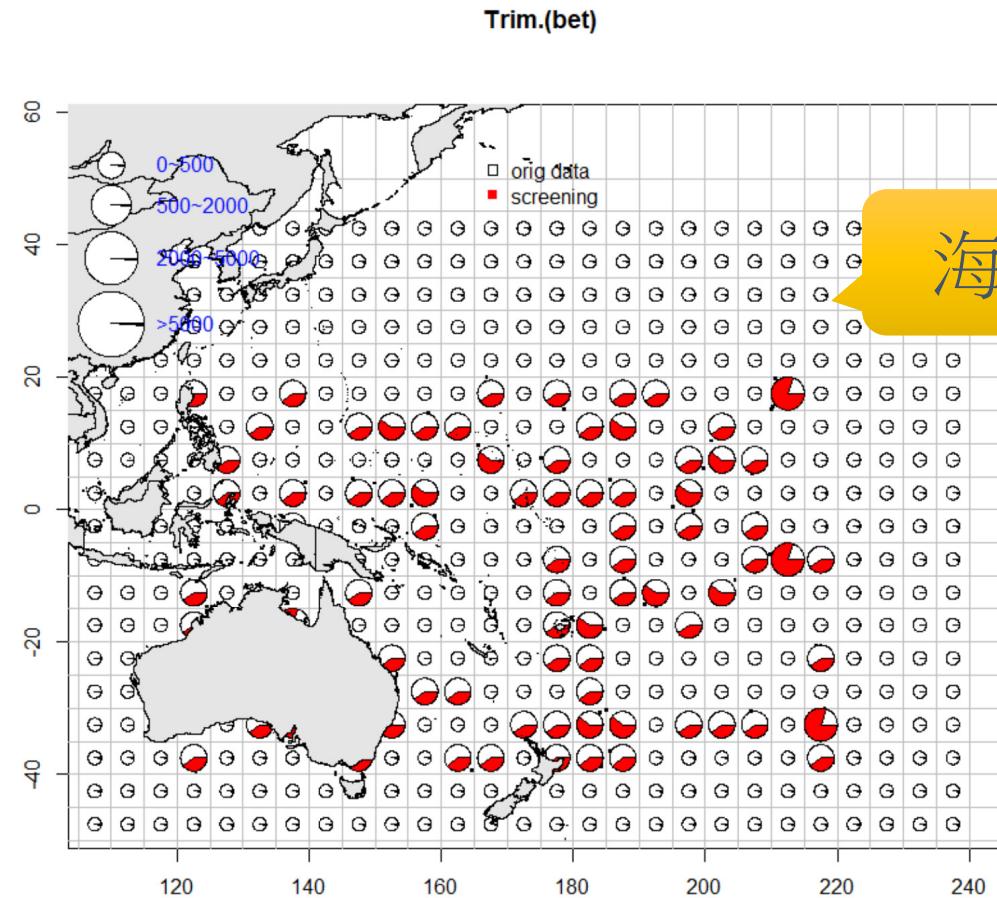
14,328 個檔案，18 個資料夾



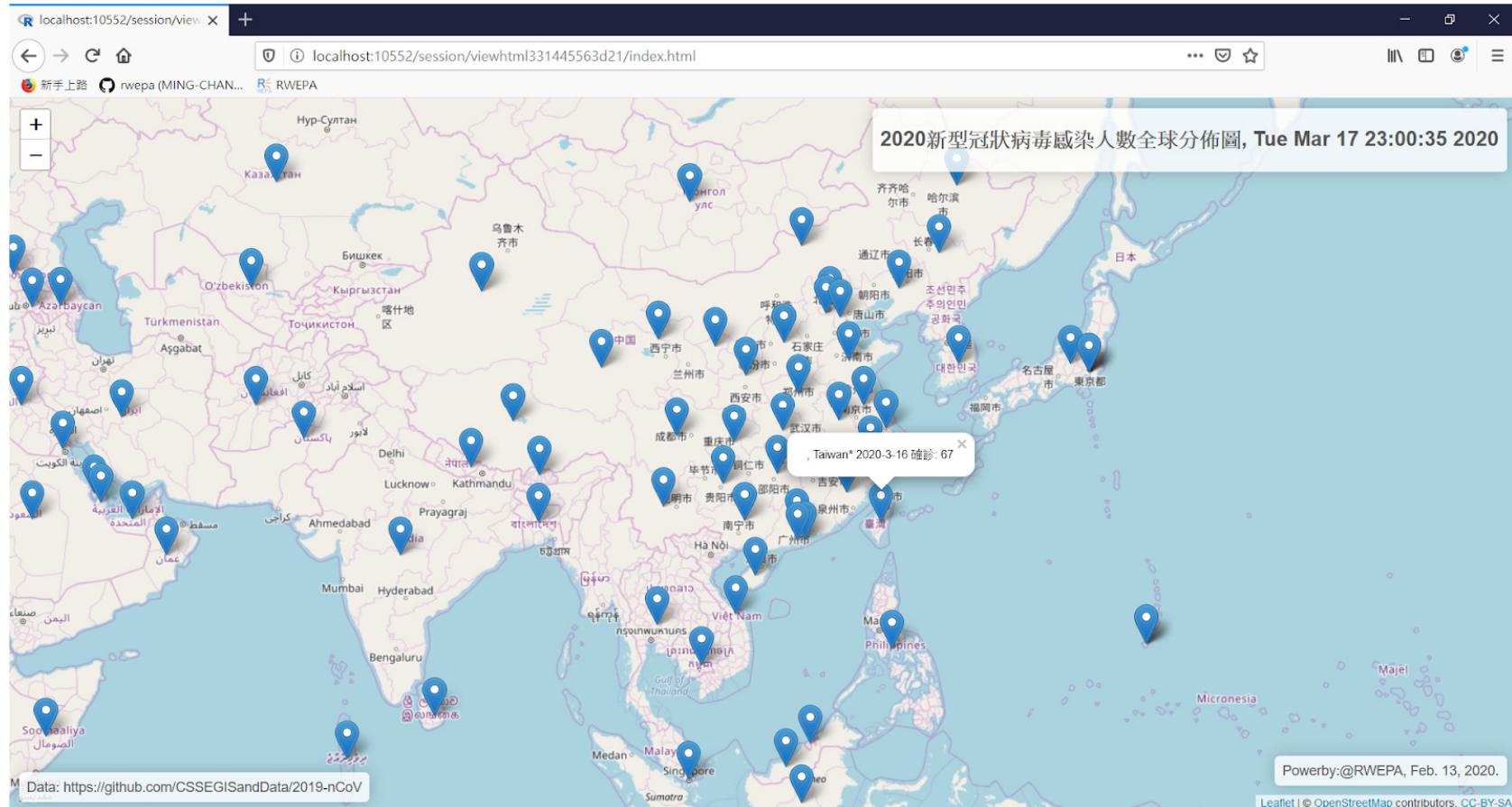
客製化選單

R統計運算

# 空間圓餅圖離群值分析

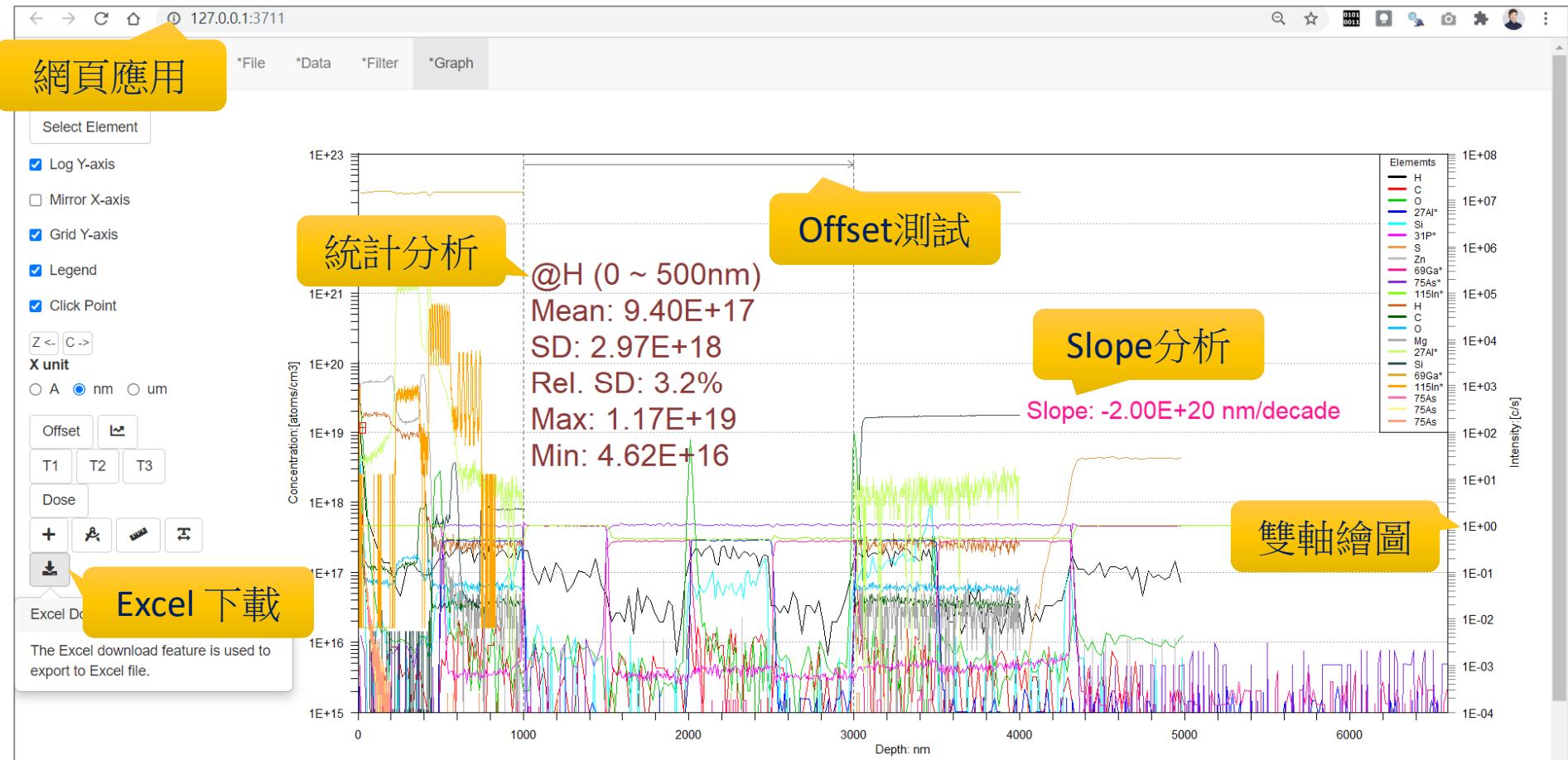


# 2020新型冠狀病毒視覺化應用



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

# 離子資料分析與視覺化應用



# 分析工具 – Excel, R, Python, Julia

# 分析工具 – Excel

- Microsoft Excel 2019: 104萬餘筆資料限制
- 排序, 樞紐分析, 圖表, 合併彙算, 篩選, VBA程式
- 限定思考模式！

A	B	C	D	E	F	G	
1	WEEK_END_DATE	STORE_NUM	UPC	UNITS	VISITS	HHS	SPEND
1048572	14-Jan-09	367	1111009477	13	13	13	18.07
1048573	14-Jan-09	367	1111009497	20	18	18	27.8
1048574	14-Jan-09	367	1111009507	14	14	14	19.32
1048575	14-Jan-09	367	1111035398	4	3	3	14
1048576	14-Jan-09	367	1111038078	3	3	3	7.5

1,048,576筆資料限制

# 大數據分析免費工具



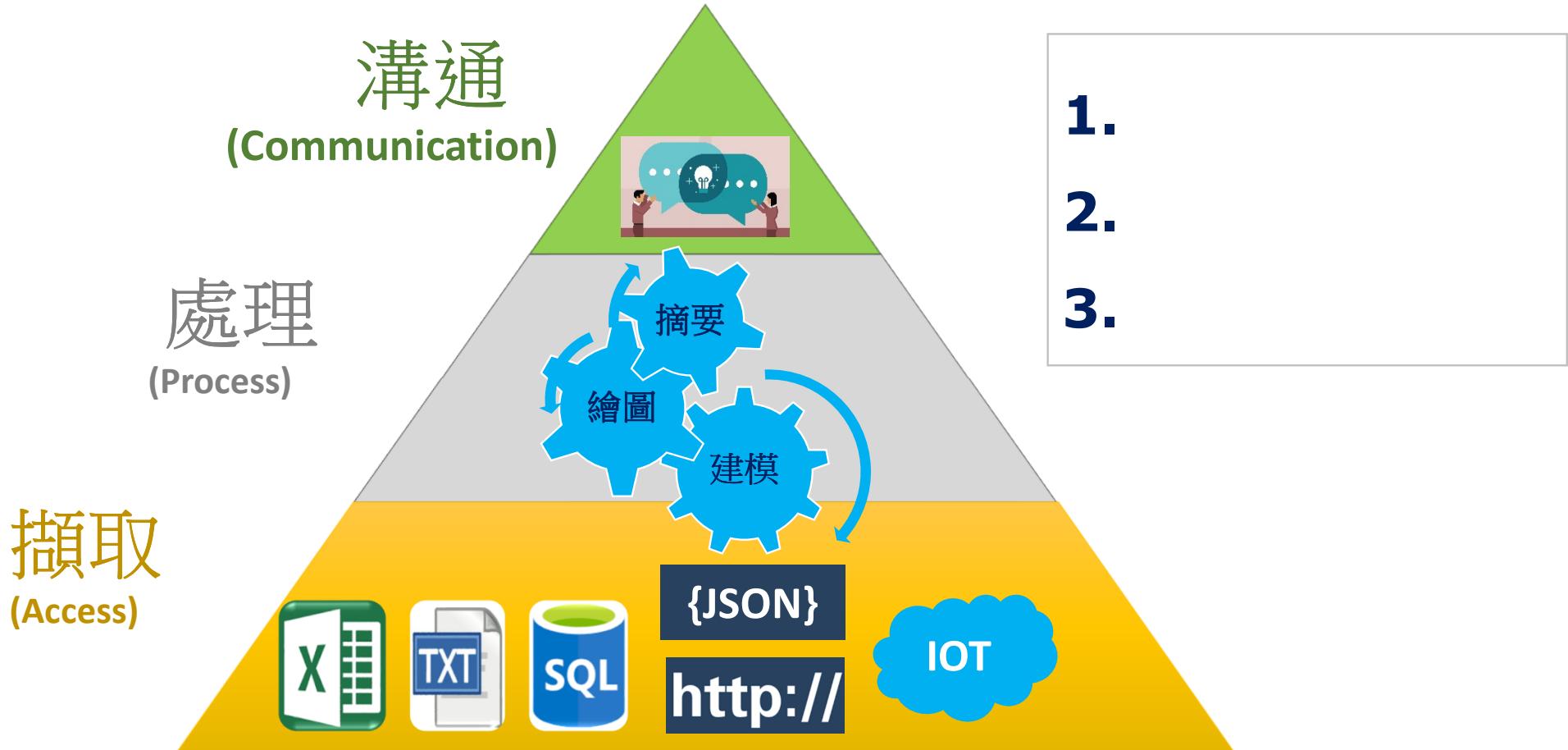
軟體	Python	R	Julia
Released	1991	2000	2012
用途	程式語言 系統結合	統計,繪圖,視覺化 程式語言	科學計算 程式語言
版本	自由軟體 物件導向	自由軟體 物件導向	自由軟體 物件導向
附加功能	免費模組	免費套件	免費模組
使用者	工科+ 商管	商管+ 工科	商管+ 工科



# 巨量資料分析的心法



# 巨量資料分析心法 – APC method



# 如何有效學習 R

---



# 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>

所有教學範例的  
原始程式檔案與  
資料集



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

# R 商業預測應用 (8小時53分鐘)

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

所有教學範例的  
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

# R 學習歷程



- 熟悉教材內容
- 將教材內容的資料集改為職場資料集
- 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



# R 簡介

---

# 認識R

- 1976 - 貝爾實驗室 John Chambers, Rick Becker, and Allan Wilks研發S語言。
- 1993 - Ross Ihaka and Robert Gentleman, University of Auckland, New Zealand 研發R語言。
  - R是一種基於S語言所發展出具備統計分析、繪圖與資料視覺化的程式語言。
- 1997年 - R的核心開發團隊 (R development core team) 成立，專責R原始碼的修改與編寫。
  - 2000年2月 - R 1.0.0
  - 2013年3月 - R 2.15.3
  - 2021年8月 - R 4.1.1



# R-八大功能



# R應用領域





# R 網頁服務案例

# 汽車產險預測模型 – demo (1/3)



The screenshot shows the user interface of the iInsurance interactive analysis platform version v.16.3.24. At the top, there is a navigation bar with links for '檔案上傳', '資料處理', '統計圖表', '模型評估', and '預測模型'. Below the navigation bar, a file upload dialog box is open, prompting the user to select an Excel file for insurance claims. The file 'SingaporeAuto.xlsx' has been selected, and the 'Upload complete' button is visible. A green callout bubble to the right of the upload area displays the message '上傳完成訊息' (Upload completed message). In the bottom right corner, there is a progress bar indicating the status of the data processing task, which is at 89% completion. An orange callout bubble above the progress bar displays the message '上傳資料進度顯示' (Upload data progress display).

iInsurance互動式分析平台\_v.16.3.24

•檔案上傳 •資料處理 •統計圖表 •模型評估 •預測模型

選取一個保險理賠Excel檔案

Browse... SingaporeAuto.xlsx

Upload complete

上傳完成訊息

上傳 Excel 資料檔

上傳資料進度顯示

資料處理 ... 89%

# 汽車產險預測模型 – demo (2/3)

iInsurance互動式分析平台\_v.16.3.24 • 檔案上傳 • 資料處理 ▾ • 統計圖表 ▾ • 模型評估 ▾ • 預測模型 ▾

Show 10 entries

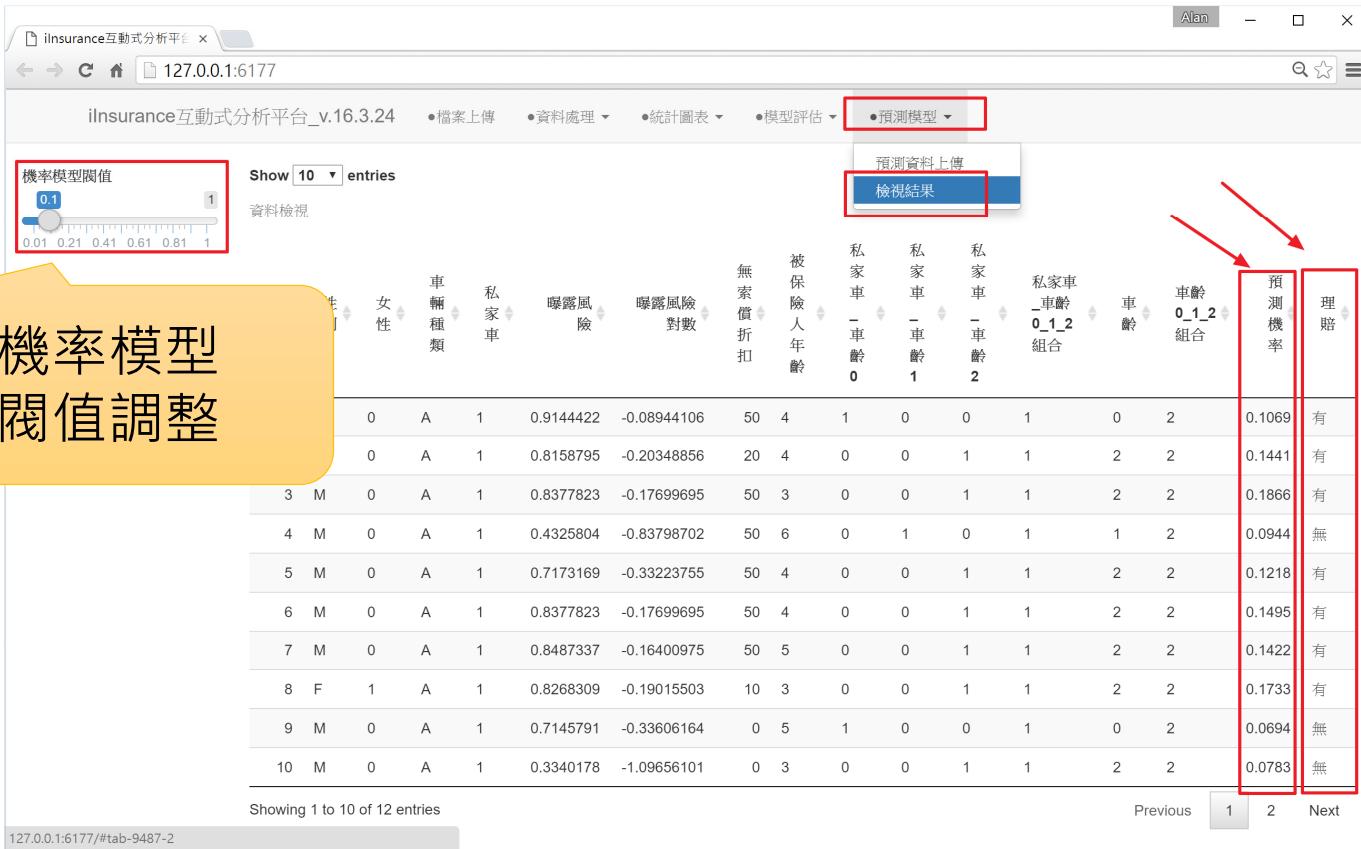
資料檢視

編號	性別	女性	車輛種類	私家車	理賠次數	曝露風險	曝露風險對數	無索償折扣	被保險人年齡	私家車_車齡0	私家車_車齡1	私家車_車齡2	私家車_車齡0_1_2組合	車齡	車齡0_1_2組合
1	U	0	T	0	0	0.668035592	-0.403413825	30	0	0	0	0	0	0	2
2	U	0	T	0	0	0.566735113	-0.567863257	30	0	0	0	0	0	0	2
3	U	0	T	0	0	0.503764545	-0.685646293	30	0	0	0	0	0	0	2
4	U	0	T	0	0	0.914442163	-0.089441058	20	0	0	0	0	0	0	2
5	U	0	T	0	0	0.536618754	-0.622467391	20	0	0	0	0	0	0	2
6	U	0	T	0	0	0.752908967	-0.283810953	20	0	0	0	0	0	0	2
7	U	0	M	0	0	0.670773443	-0.39932384	20	0	0	0	0	0	6	6
8	U	0	M	0	0	0.837782341	-0.176996949	20	0	0	0	0	0	6	6
9	U	0	M	0	0	0.167008898	-1.789708186	20	0	0	0	0	0	6	6
10	U	0	M	0	0	0.503764545	-0.685646293	20	0	0	0	0	0	6	6

Showing 1 to 10 of 7,483 entries

Previous 1 2 3 4 5 ... 749 Next

# 汽車產險預測模型 – demo (2/3)



The screenshot shows the iInsurance Interactive Analysis Platform interface. A yellow callout bubble on the left points to a slider labeled "機率模型閥值" (Probability Model Threshold) set at 0.1. A blue callout bubble on the right points to the "檢視結果" (View Results) button in the top navigation bar. Red boxes highlight the "預測模型" (Prediction Model) tab, the "機率模型閥值" slider, the "檢視結果" button, and the "預測機率" (Prediction Probability) and "理賠" (Claim) columns in the results table.

機率模型閥值調整

預測結果

序號	性別	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
0	A	1	0.9144422	-0.08944106	50	4	1	0	0	1	0	2	0.1069	有
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	0.1866	有
4	M	0	A	1	0.4325804	-0.83798702	50	6	0	1	0	1	0.0944	無
5	M	0	A	1	0.7173169	-0.33223755	50	4	0	0	1	1	0.1218	有
6	M	0	A	1	0.8377823	-0.17699695	50	4	0	0	1	1	0.1495	有
7	M	0	A	1	0.8487337	-0.16400975	50	5	0	0	1	1	0.1422	有
8	F	1	A	1	0.8268309	-0.19015503	10	3	0	0	1	1	0.1733	有
9	M	0	A	1	0.7145791	-0.33606164	0	5	1	0	0	1	0.0694	無
10	M	0	A	1	0.3340178	-1.09656101	0	3	0	0	1	1	0.0783	無

# THANK YOU!

Q & A

MING-CHANG LEE

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

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