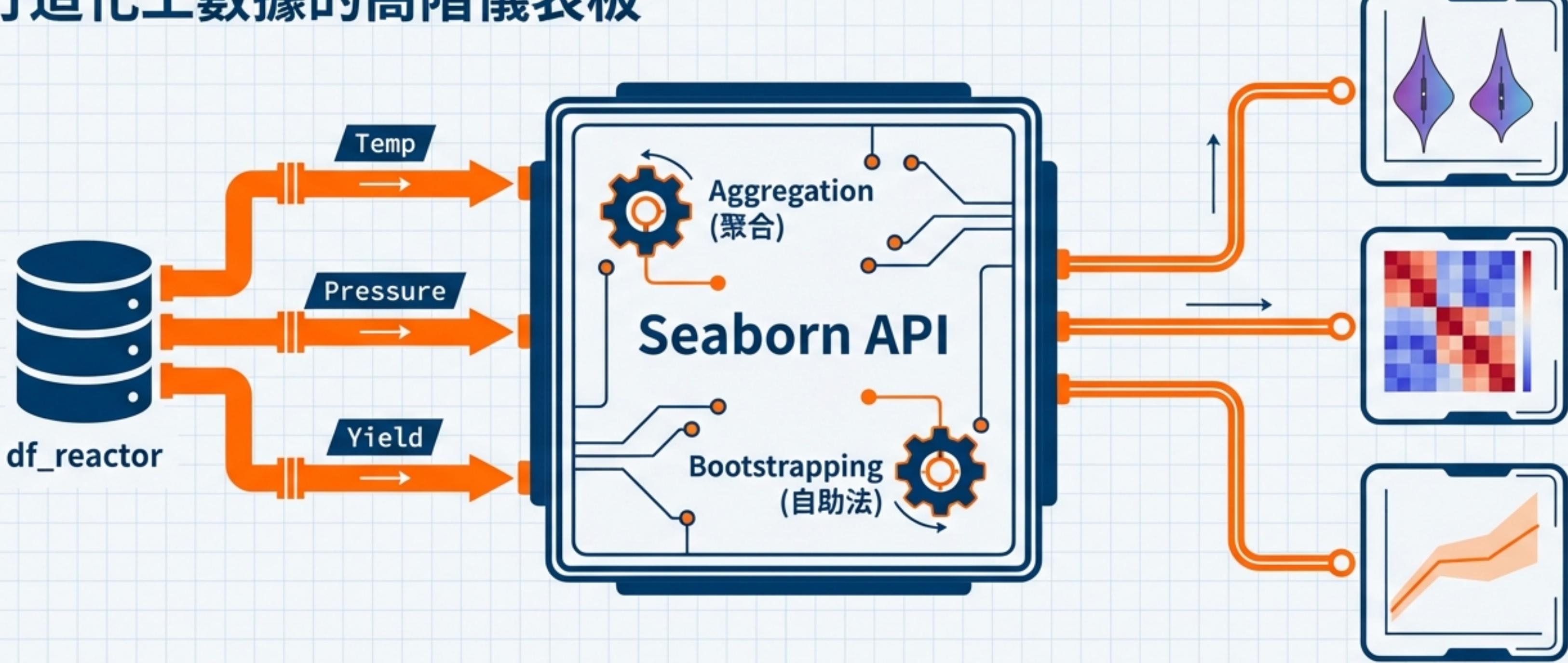


Unit 04 Seaborn 統計資料視覺化

打造化工數據的高階儀表板



典範轉移：從低階繪圖到高階統計

Matplotlib (低階控制)

```
1 import matplotlib.pyplot as plt  
2 import matplotlib as plt  
3  
4  
5 x = [...]  
6 y = [...]  
7  
8 fig, ax = plt.subplots()  
9 ax.scatter(x, y)  
10 ax.set_title('Data')  
11 ax.set_xlabel('X')  
12 ax.set_ylabel('Y')  
13 plt.show()  
14  
15 # some constants  
16 # parameters = Matplotlib API  
17 # much easier, prettier  
18 plt.close()
```

- 客製化精細控制
- 需手動計算統計量
- 適合：底層開發與特殊圖表

Abstraction
(抽象化)

Paradigm Shift

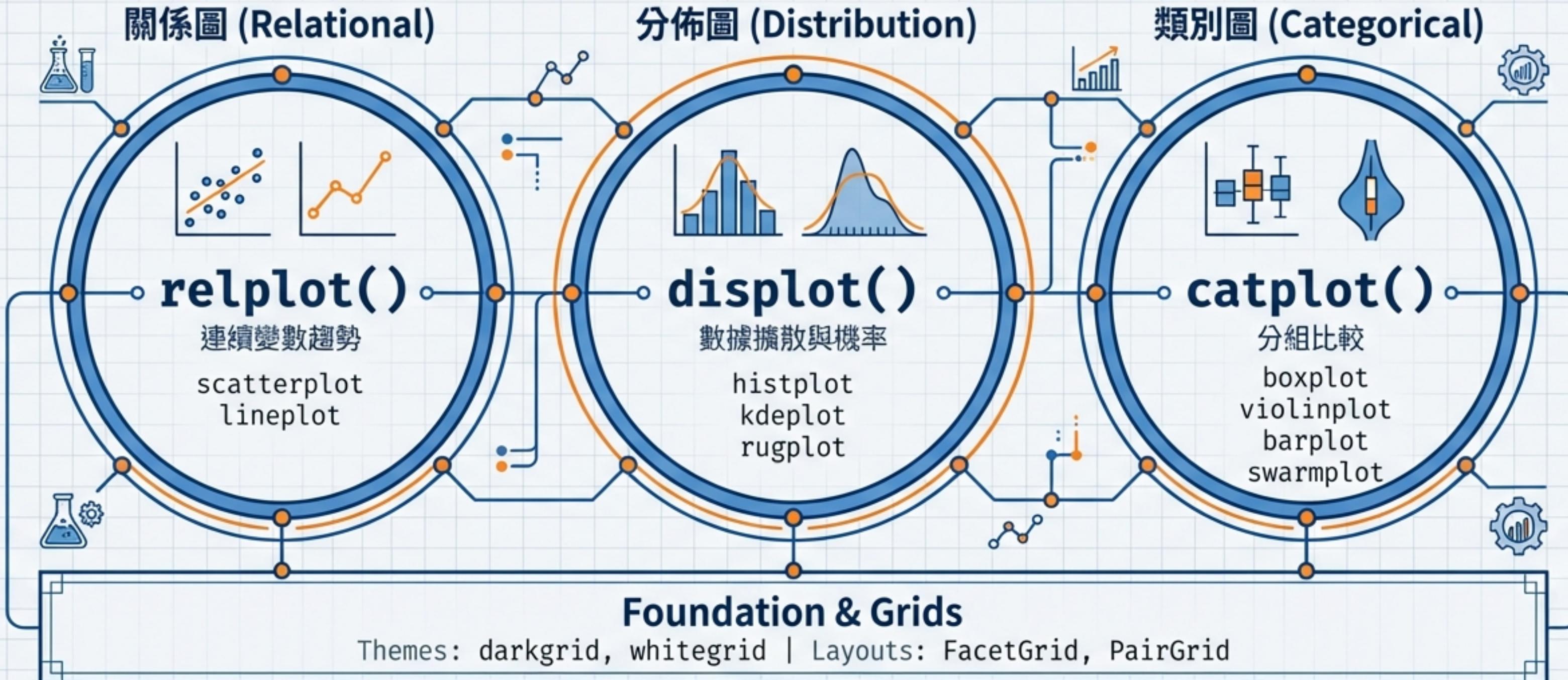
Seaborn (高階洞察)

```
sns.regplot(x="T", y="Yield")
```

- 基於 Matplotlib 構建
- 內建統計模型 (信賴區間/回歸)
- 原生支援 Pandas DataFrame

關鍵洞察：Seaborn 讓工程師專注於化學意義，而非像素座標。

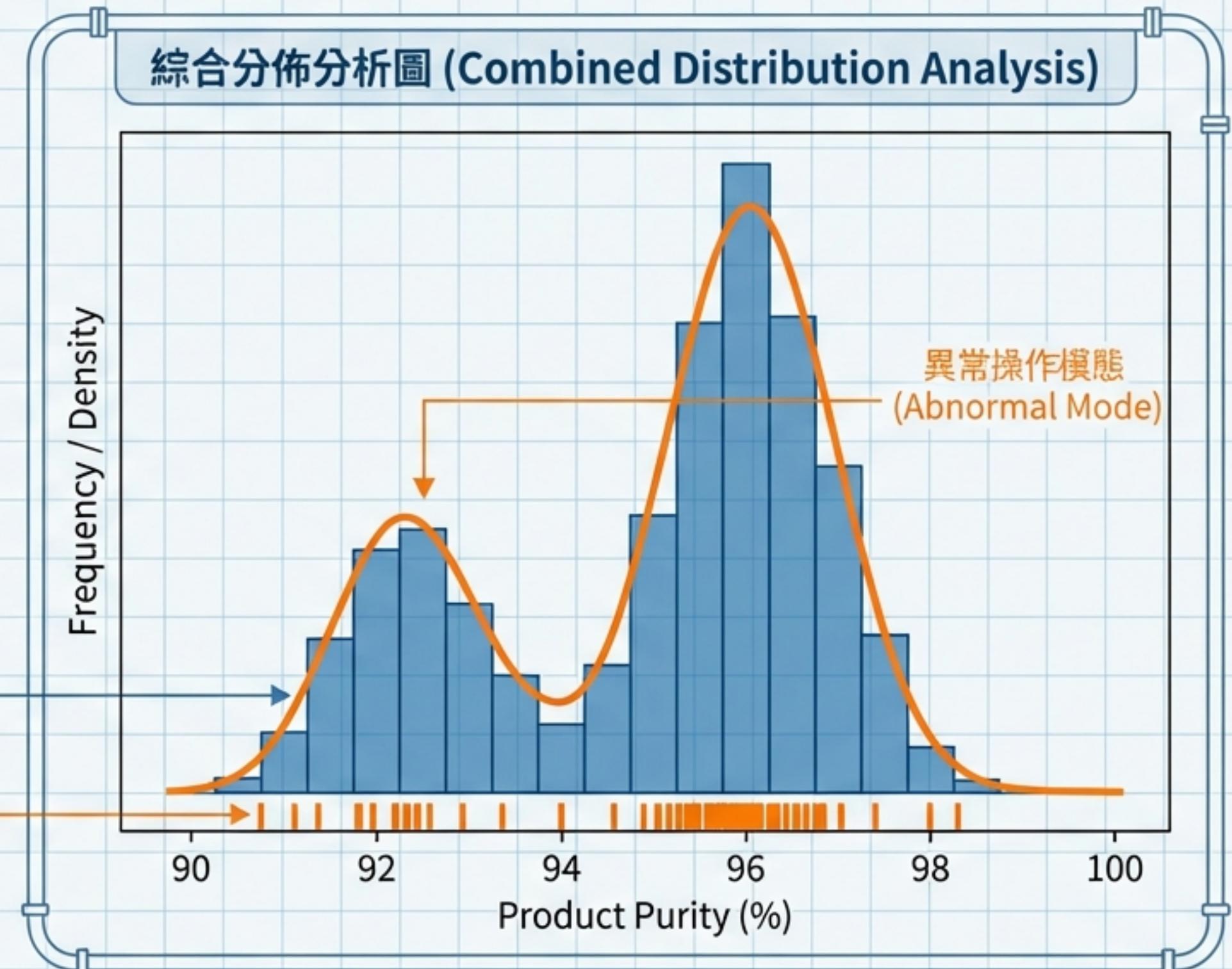
Seaborn 函數地圖：理解 API 層次結構



關鍵洞察：Seaborn 讓工程師專注於化學意義，而非像素座標。

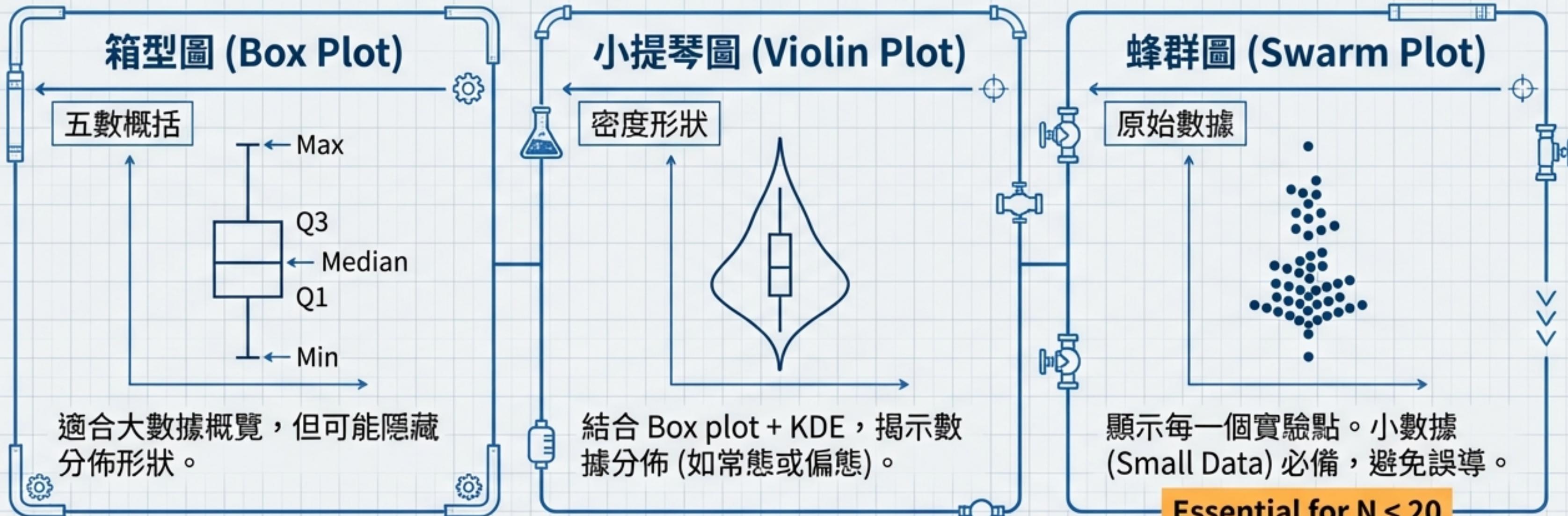
分佈圖：分析製程穩定性

化工應用：監控反應器出口的產品純度 (Product Purity)

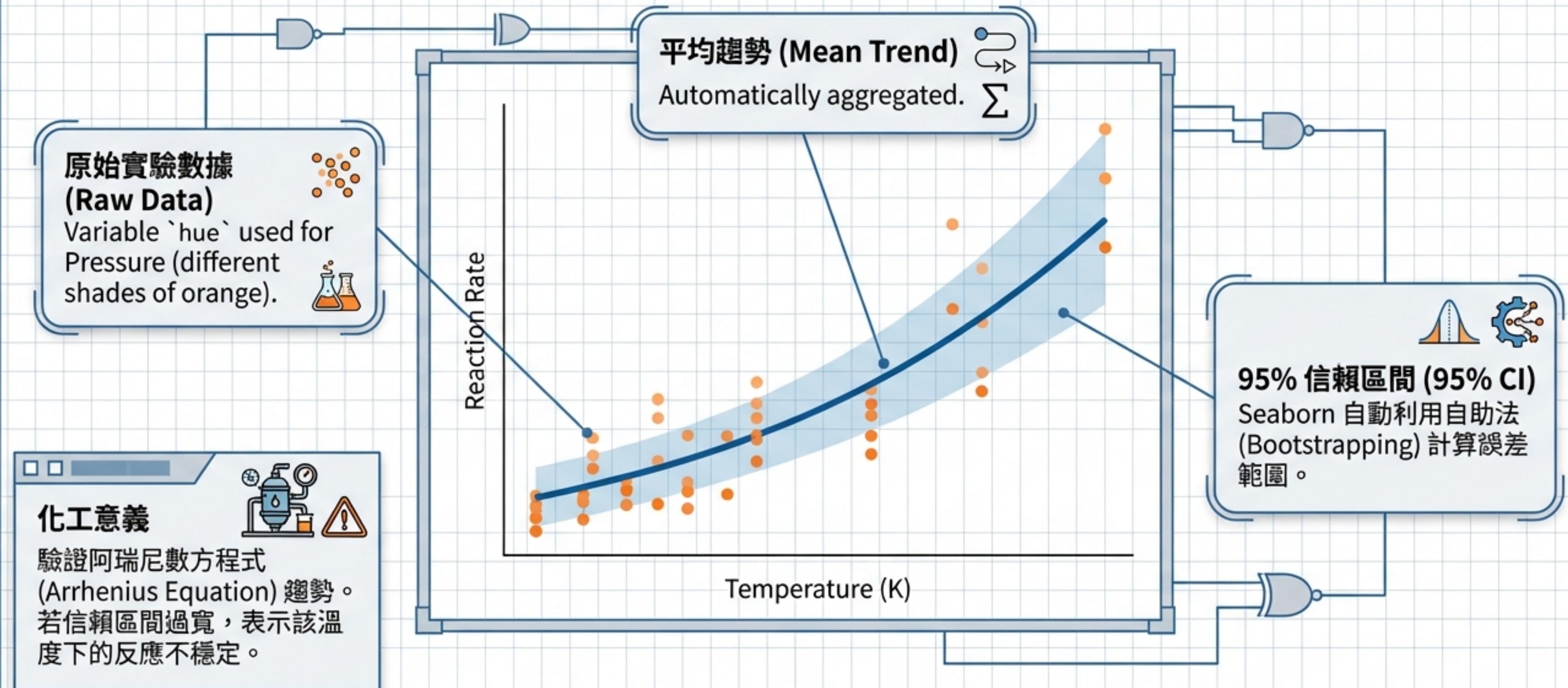


類別圖：實驗條件比較與篩選

化工應用：比較不同催化劑 (Catalyst A vs. B) 的產率



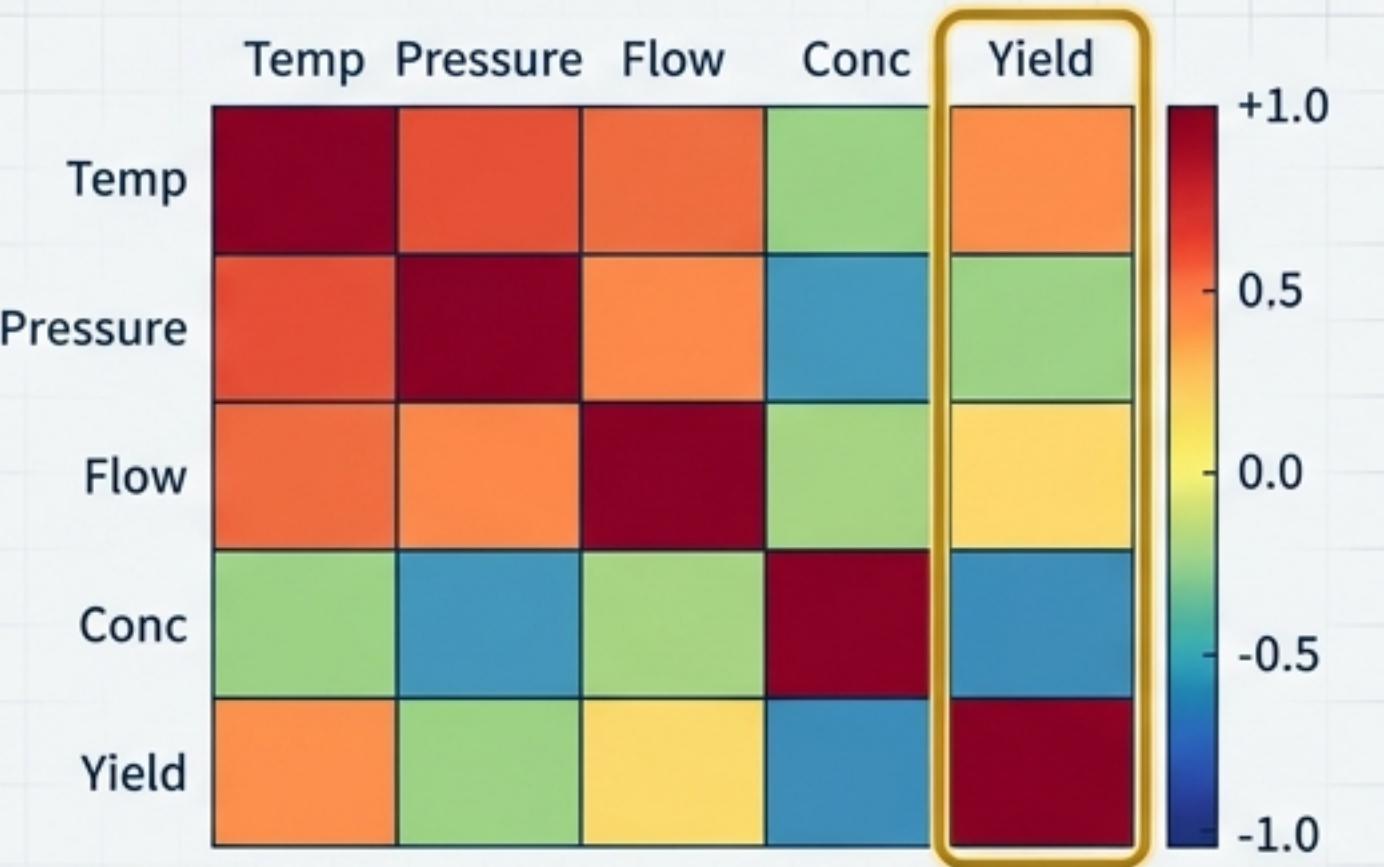
關係圖：動力學趨勢與信賴區間



矩陣圖：相關性分析與特徵篩選

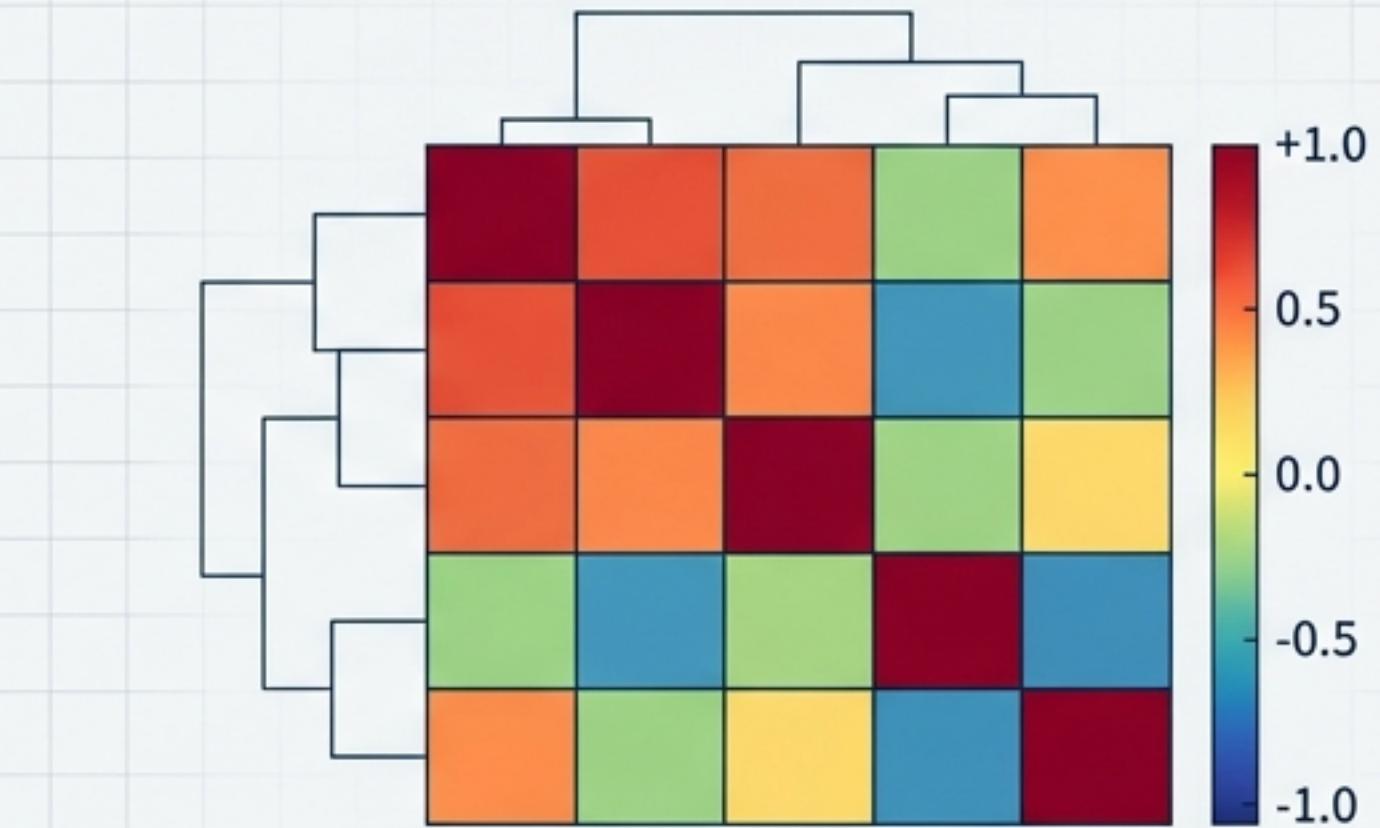
Feature Selection for Process Modeling

矩陣圖 (Heatmap)



尋找與 Yield 高度相關的變數
(Finding Impact Factors)。

階層式聚類 (Clustermap)



階層式聚類 (Hierarchical Clustering)。
自動將相似的操作批次 (Batches) 分組，
識別不同製程狀態。

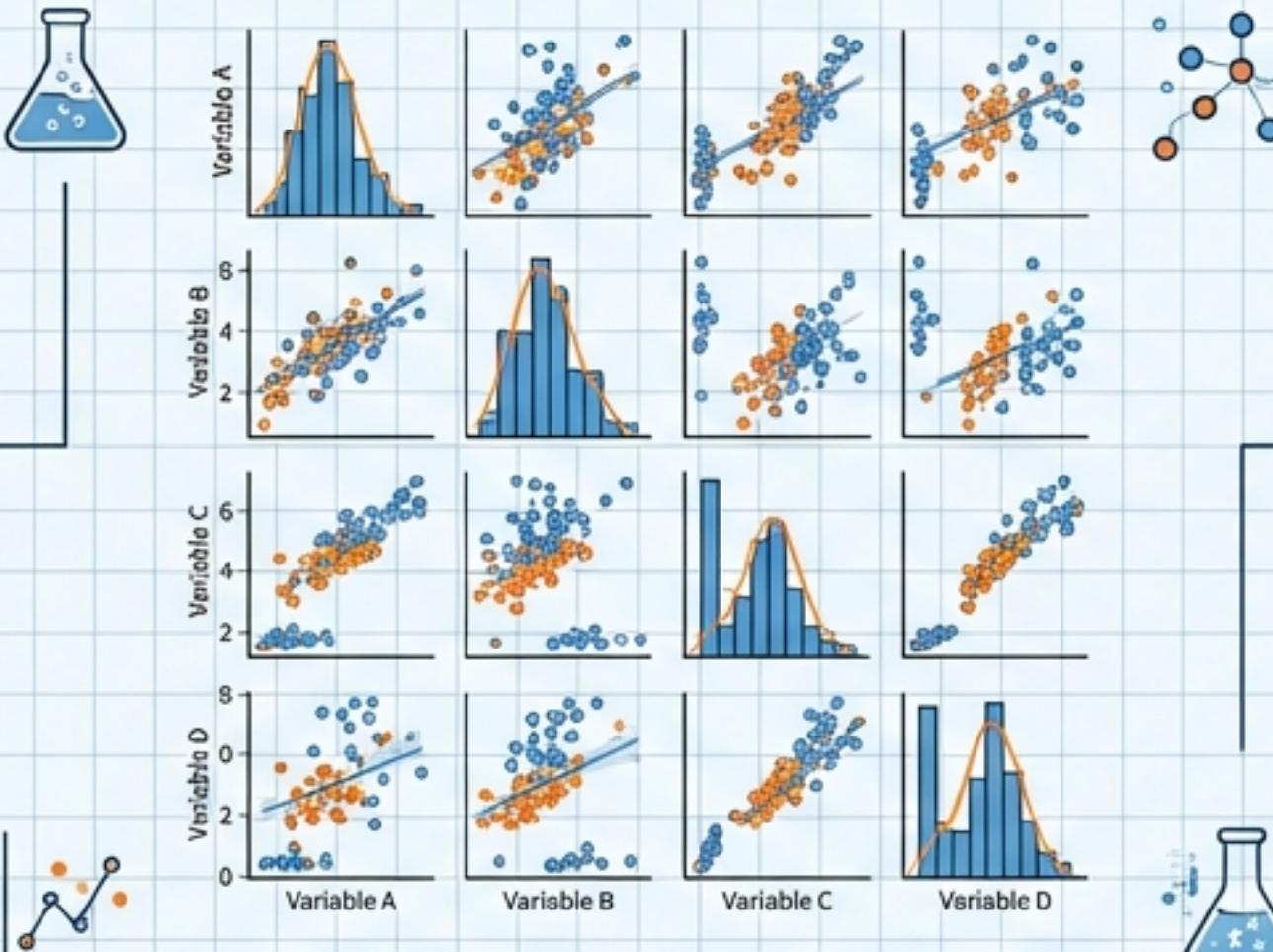


前置作業：在進入機器學習 (Unit 10) 前，必須先移除相關性過高 (Collinear) 的冗餘變數。

多變數探索：數據的全知視角

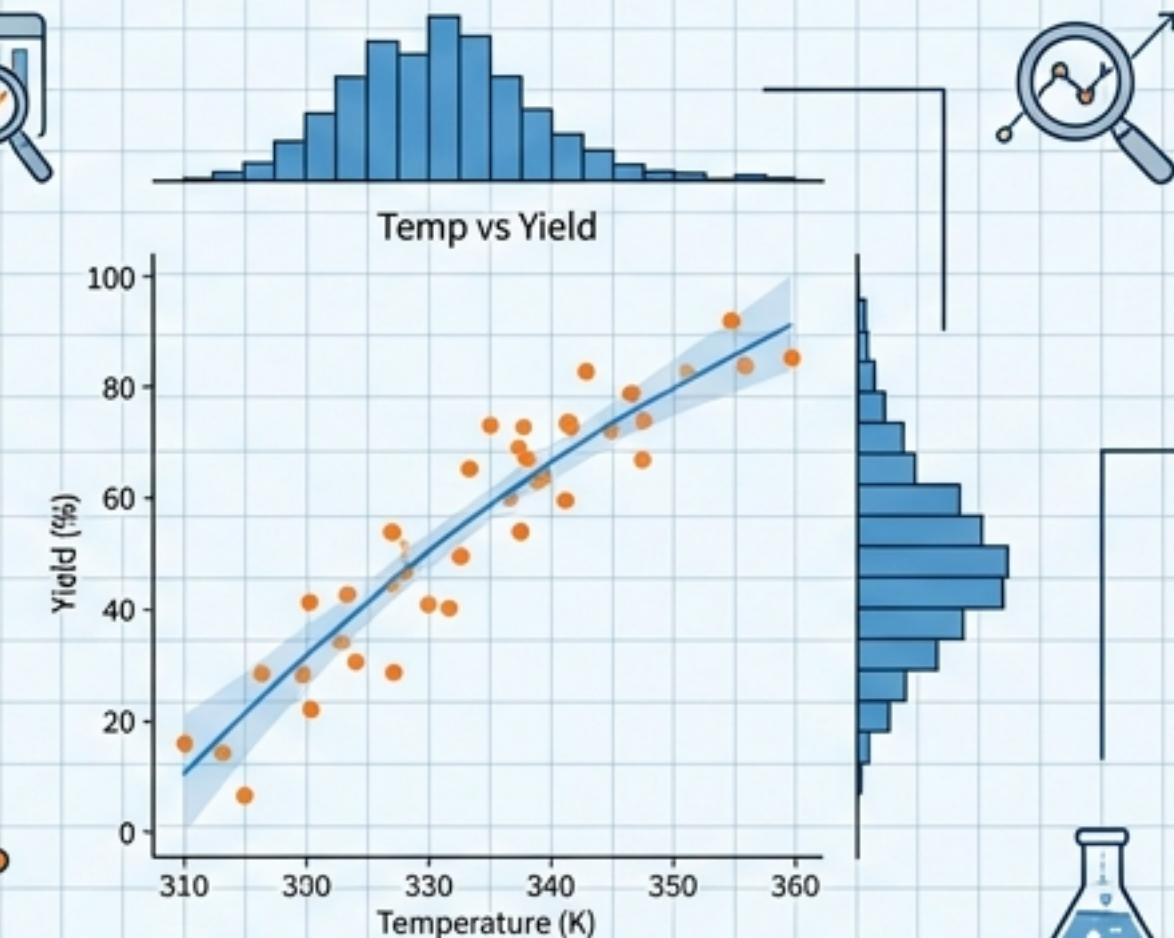
The God's Eye View for Pilot Plant Data

配對圖 (Pair Plot - 'pairplot')



快速掃描所有變數間的兩兩關係
探索反應條件多參數空間。

聯合分佈圖 (Joint Plot - 'jointplot')



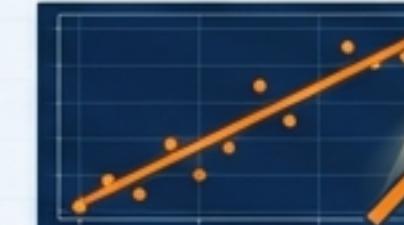
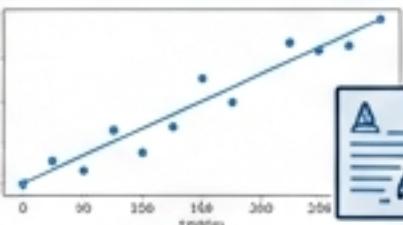
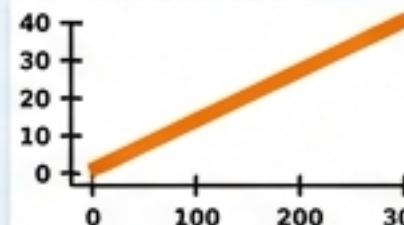
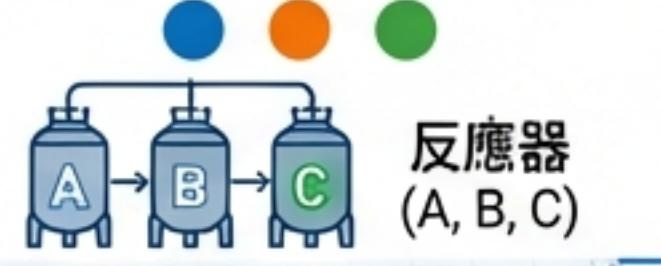
同時檢視相關性與邊際分佈

Key Takeaway : 使用這些高階網格圖表，
在一張圖中完成 80% 的初步數據掃描。

視覺美學：從 Notebook 到學術發表

優化圖表風格，提升學術溝通效率與影響力。

樣式與情境控制 (Style & Context Control)

1	Style Control (`set_style`)	'whitegrid' (數據讀取) 數據讀取 (Data Reading)		↔ 風格切換 (Style Switching)	'dark' (強調趨勢) 強調趨勢 (Trend Emphasis)		
2	Context Control (`set_context`)	'paper' (論文發表) 論文發表 (Paper Publication)		↔ 情境調整 (Context Adjustment)	'talk' (簡報演講) 簡報演講 (Presentation Talk)		
3	調色盤 (Color Palettes)	Sequential (連續) 溫度 (0-100°C)		Diverging (發散) 偏差量 (Error +/-)		Qualitative (類別) 反應器 (A, B, C)	

Key Takeaway：選擇正確的視覺樣式與情境，是學術發表中清晰傳達專業知識的關鍵步驟。

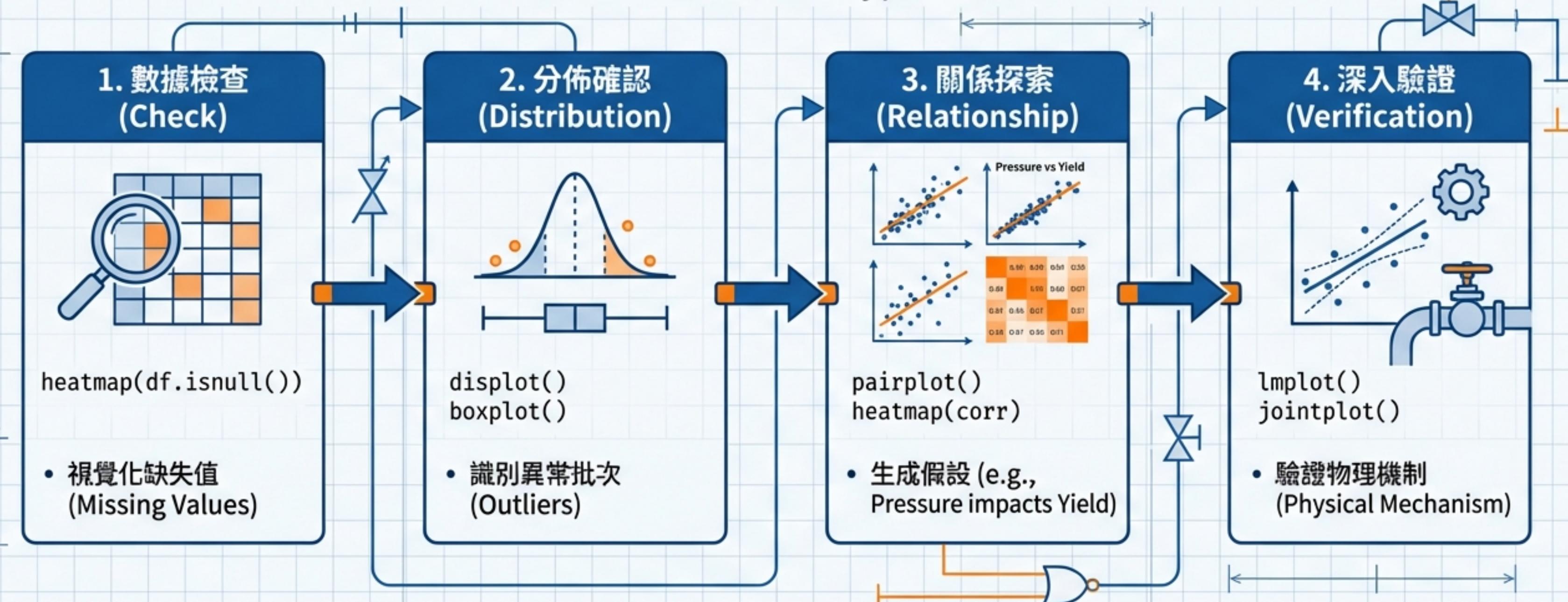




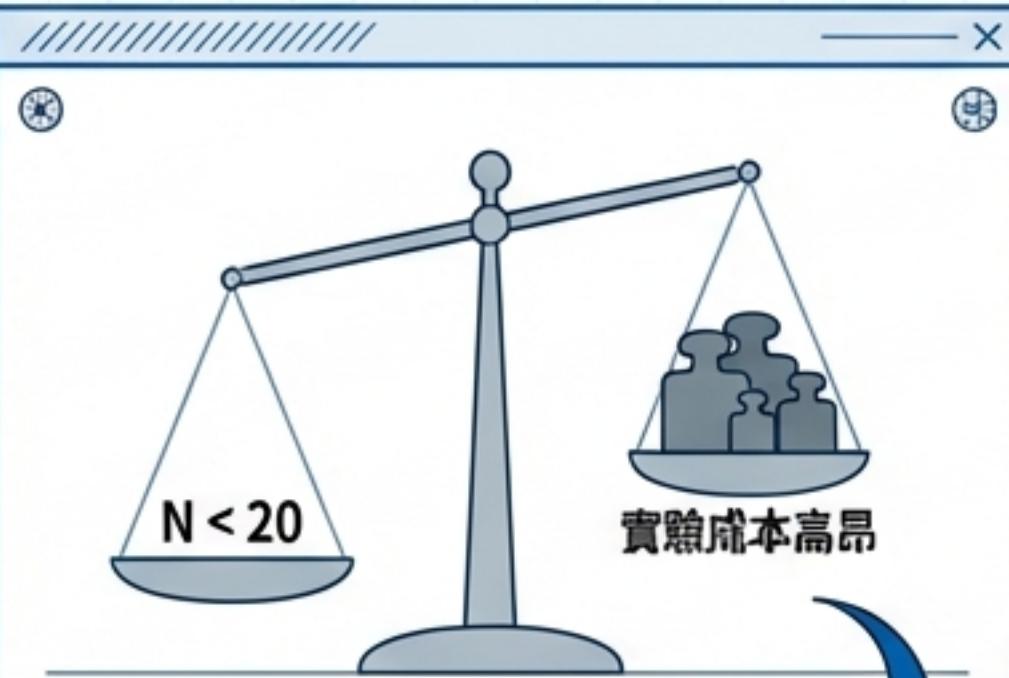
原則：Garbage In, Garbage Out (垃圾進，垃圾出)

化工 EDA 工作流程 (ChemE EDA Workflow)

From Raw Data to Hypothesis



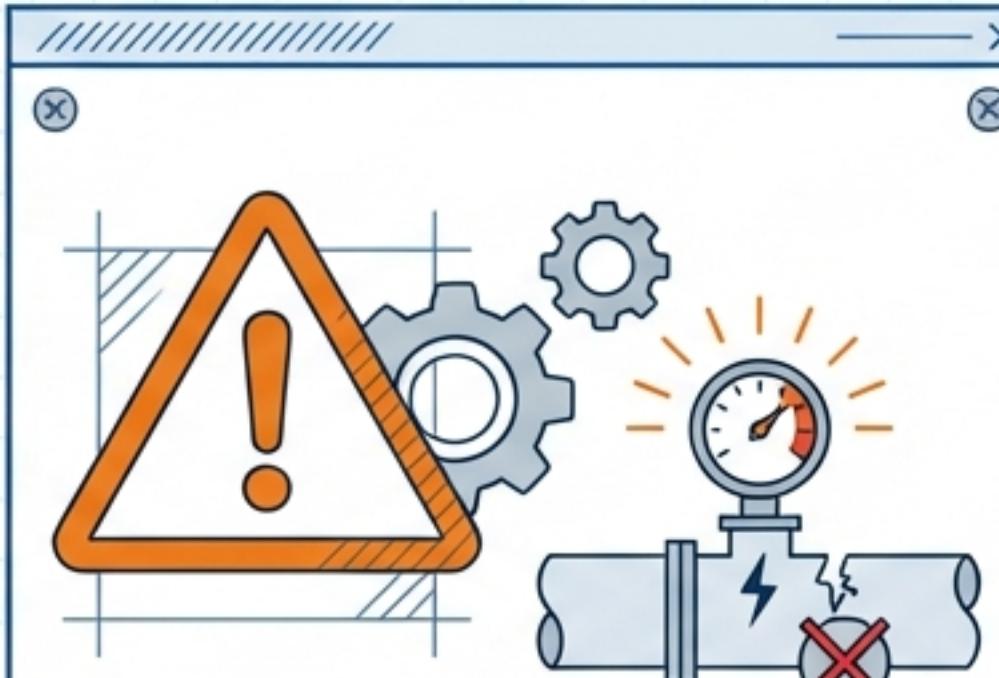
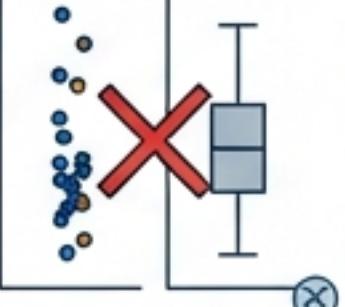
現實世界的挑戰：小數據與離群值



小數據 (Small Data)
實驗成本高昂，樣本數 $N < 20$ 。

→ 解決方案 (Solution):

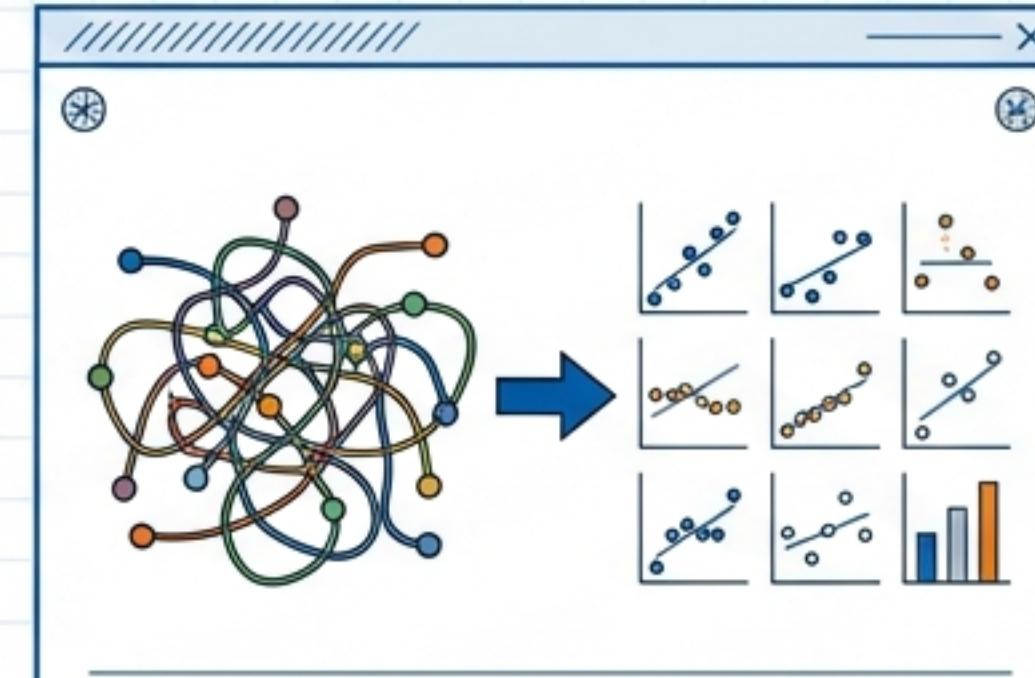
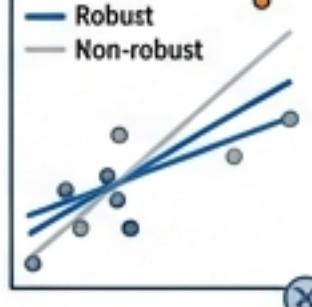
- 使用 stripplot。
- 不要使用 Boxplot 隱藏數據。



離群值 (Outliers)
感測器故障或操作失誤導致極端值。

→ 解決方案 (Solution):

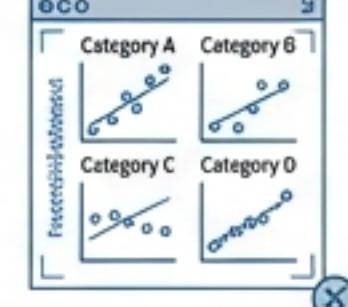
- 使用 regplot (robust=True)。
- 降低異常值權重。



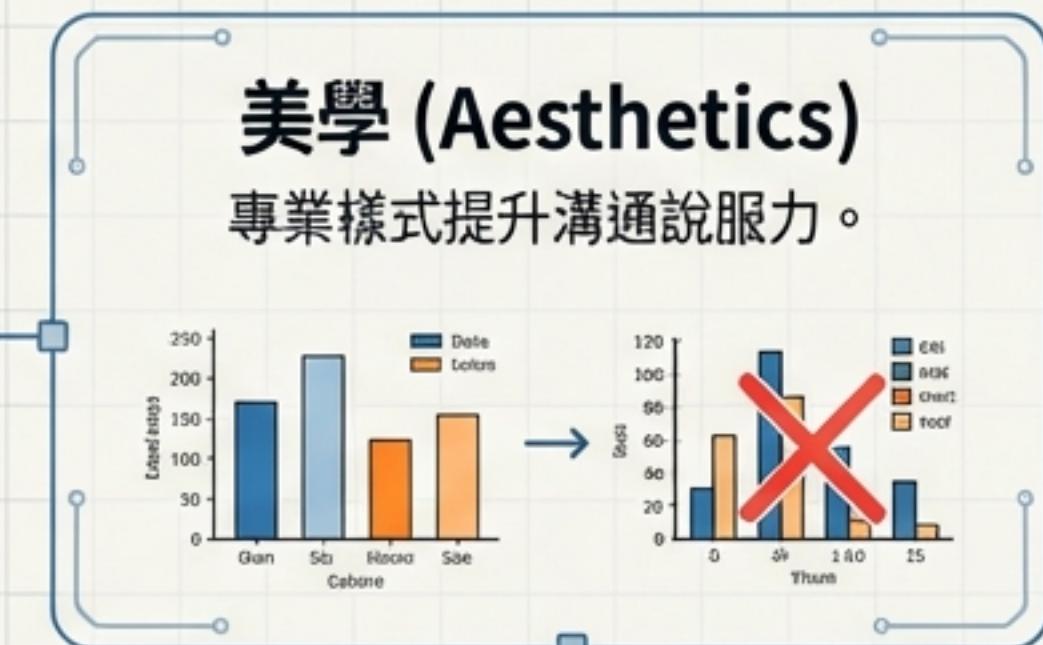
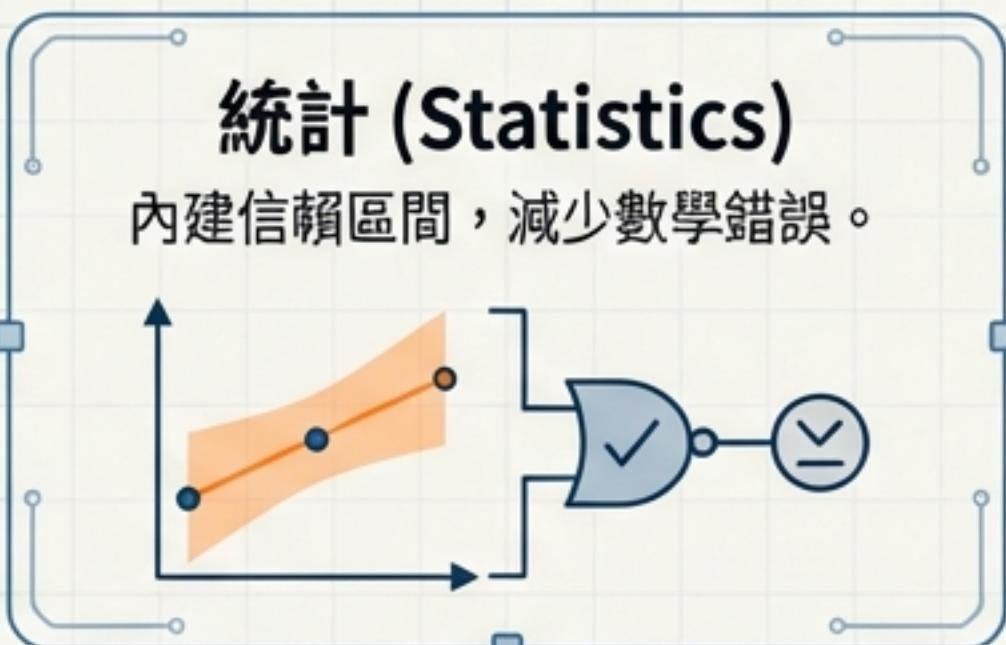
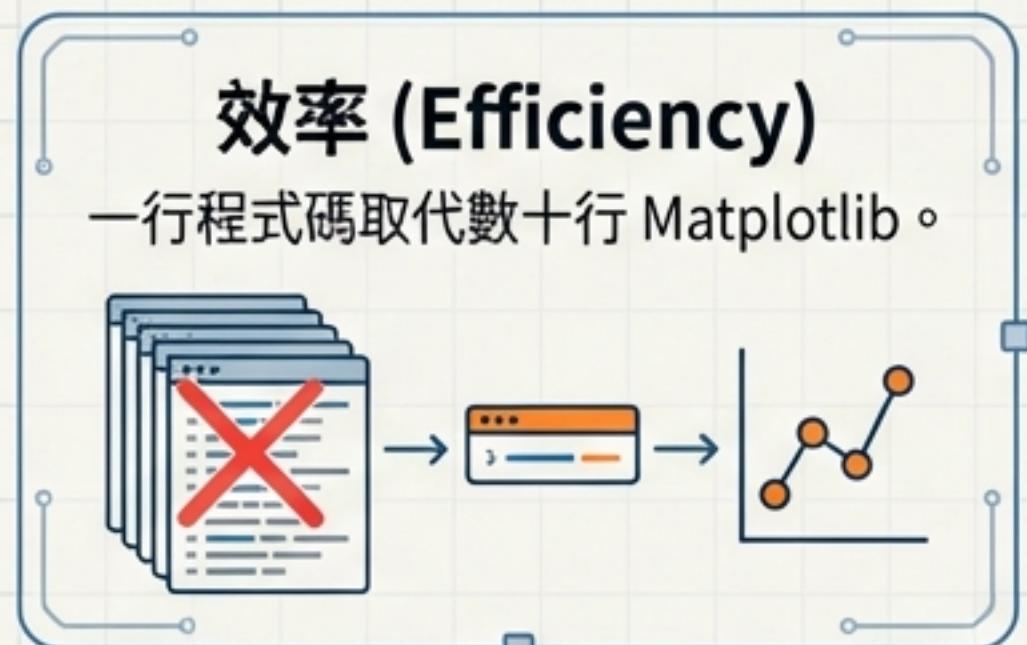
多維度混亂 (High Dim)
變數太多無法同時觀察。

→ 解決方案 (Solution):

- 使用 FacetGrid (Small Multiples) 依類別拆分視圖。



總結：從數據繪圖到科學決策



Unit 04 (Completed)

數據視覺化
(Visualization)

從原始數據到視覺洞察。

下一步

Unit 05 - 非監督式學習
(Unsupervised Learning)

利用機器學習演算法自動發現分群與異常。

