

「商業資料分析與管理決策」 ——分析與決策

Spring 2025

授課教師：統計系余清祥

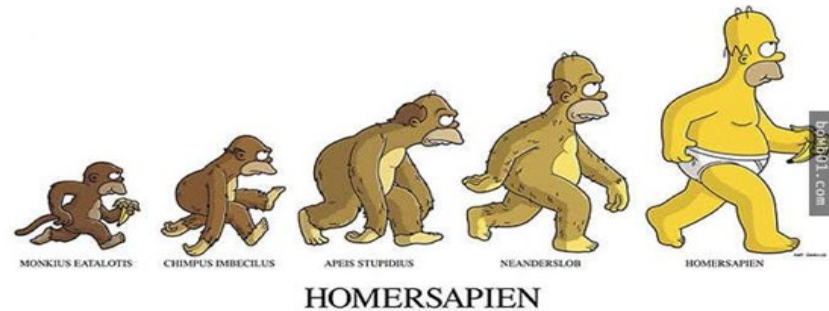
2025年4月13日

授課內容：統計分析_Part 1

課程下載：csyue.nccu.edu.tw



資訊與決策

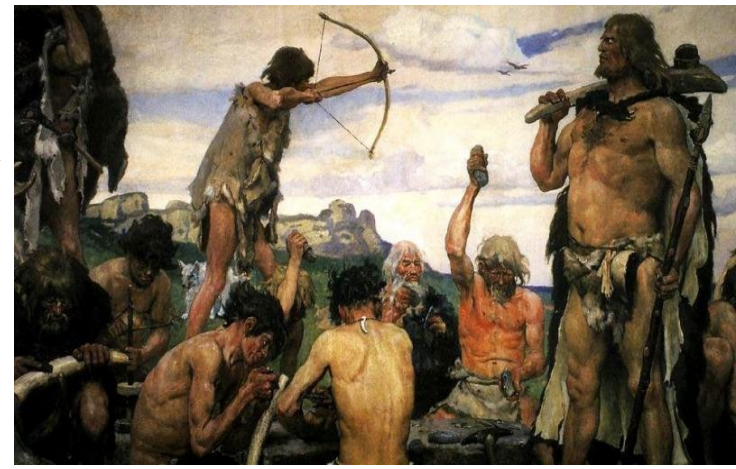


- 過去擁有土地、資金等的資本家，因為勞力及資本密集具有優勢；21世紀是知識經濟的時代，擁有及解讀資訊者掌握絕對優勢。

(Amazon、Google！)

- 問題：人類歷史（包括石器時代）各年代的生存關鍵是什麼？

註：天下文化《人類大命運》
《人類大歷史》。



天下文化 遠見

人類大命運

從智人到神人

Homo Deus

A Brief History of Tomorrow

by Yuval Noah Harari

哈拉瑞 著 林俊宏 譯

天下文化

人類大歷史

從野獸到扮演上帝

Sapiens

[From Animals Into Gods]

A Brief History of Humankind

by Yuval Noah Harari

哈拉瑞 著 林俊宏 譯

資料分析策略

■ 「觀察」、「推論」、「驗證」三步驟

- 先檢查資料品質，避免Garbage in, garbage out，通常會佔用一半以上的時間。
- 接著是探索性資料分析(EDA)，逐步找出資料重要特性，作為進一步推論(如迴歸分析)的依據。
- 驗證性資料分析(CDA)則是最後步驟，分析結果應與EDA接近，否則需重頭檢查。

資料偵錯

資料輸入錯誤、尋找可能的離群值。

初步探索資料特性

資料的集中、散佈趨勢

驗證已知的結果

是否與已知的結果相同？

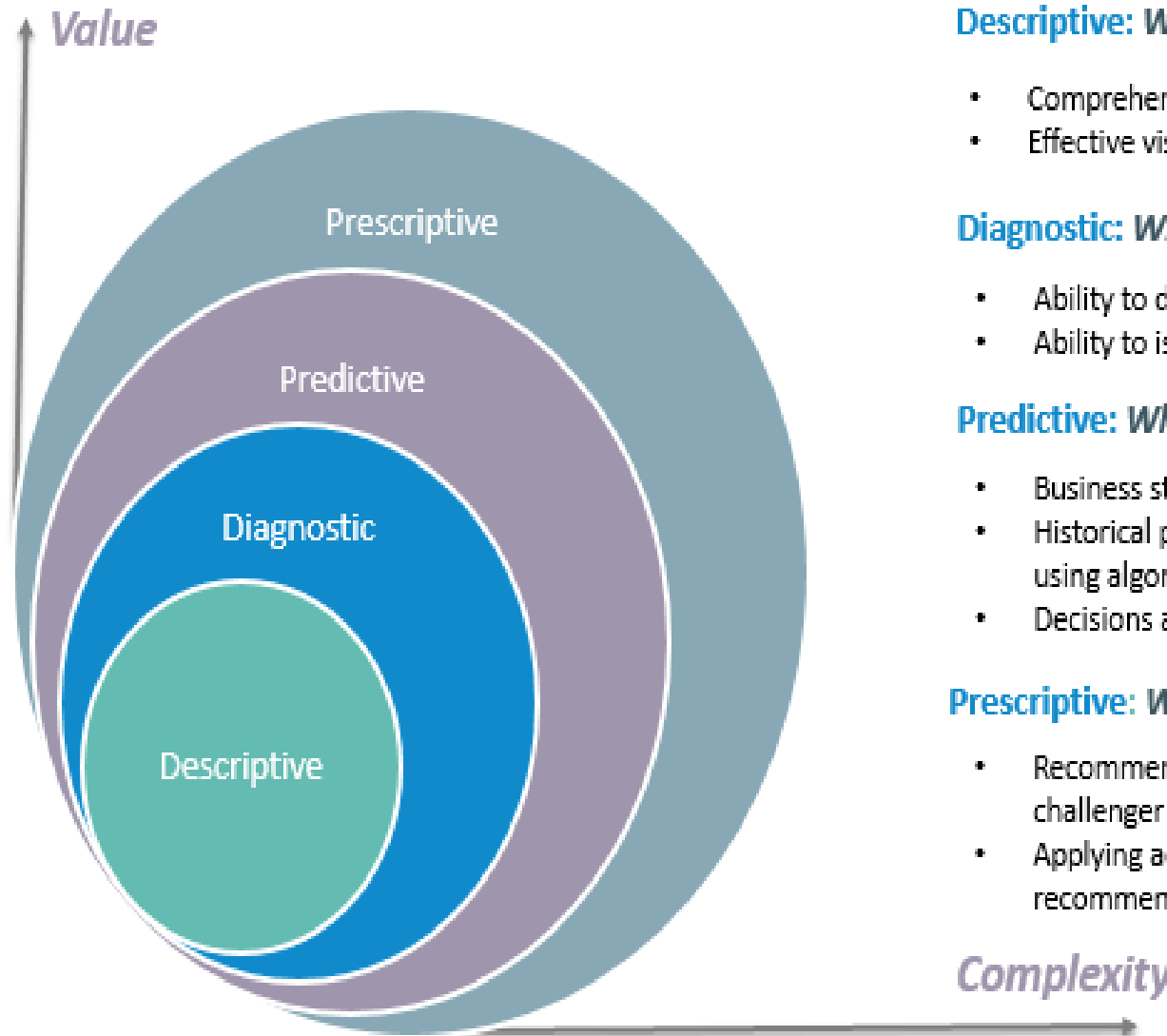


EDA and CDA

- EDA (or **Descriptive analytics**) tells us what happened up from the data.
- CDA can be separated into two parts:
 - **Predictive analytics** give us clues about the future, given data and domain knowledge.
 - **Prescriptive analytics** provide suggestions for optimizing the future.

1. **Descriptive.** Traditional HR metrics are largely efficiency metrics (turnover rate, time to fill, cost of hire, number hired and trained, etc.). The primary focus here is on cost reduction and process improvement. Descriptive HR analytics reveal and describe *relationships* and *current and historical data patterns*. This is the foundation of your analytics effort. It includes, for example, dashboards and scorecards; workforce segmentation; data mining for basic patterns; and periodic reports.
2. **Predictive.** Predictive analysis covers a variety of techniques (statistics, modeling, data mining) that use current and historical facts to make predictions about the future. It's about probabilities and potential impact. It involves, for example, models used for increasing the probability of selecting the right people to hire, train, and promote.
3. **Prescriptive.** Prescriptive analytics goes beyond predictions and outlines decision options and workforce optimization. It is used to analyze complex data to predict outcomes, provide decision options, and show alternative business impacts. It involves, for example, models used for understanding how alternative learning investments impact the bottom line (rare in HR).

4 types of Data Analytics



What is the data telling you?

Descriptive: *What's happening in my business?*

- Comprehensive, accurate and live data
- Effective visualisation

Diagnostic: *Why is it happening?*

- Ability to drill down to the root-cause
- Ability to isolate all confounding information

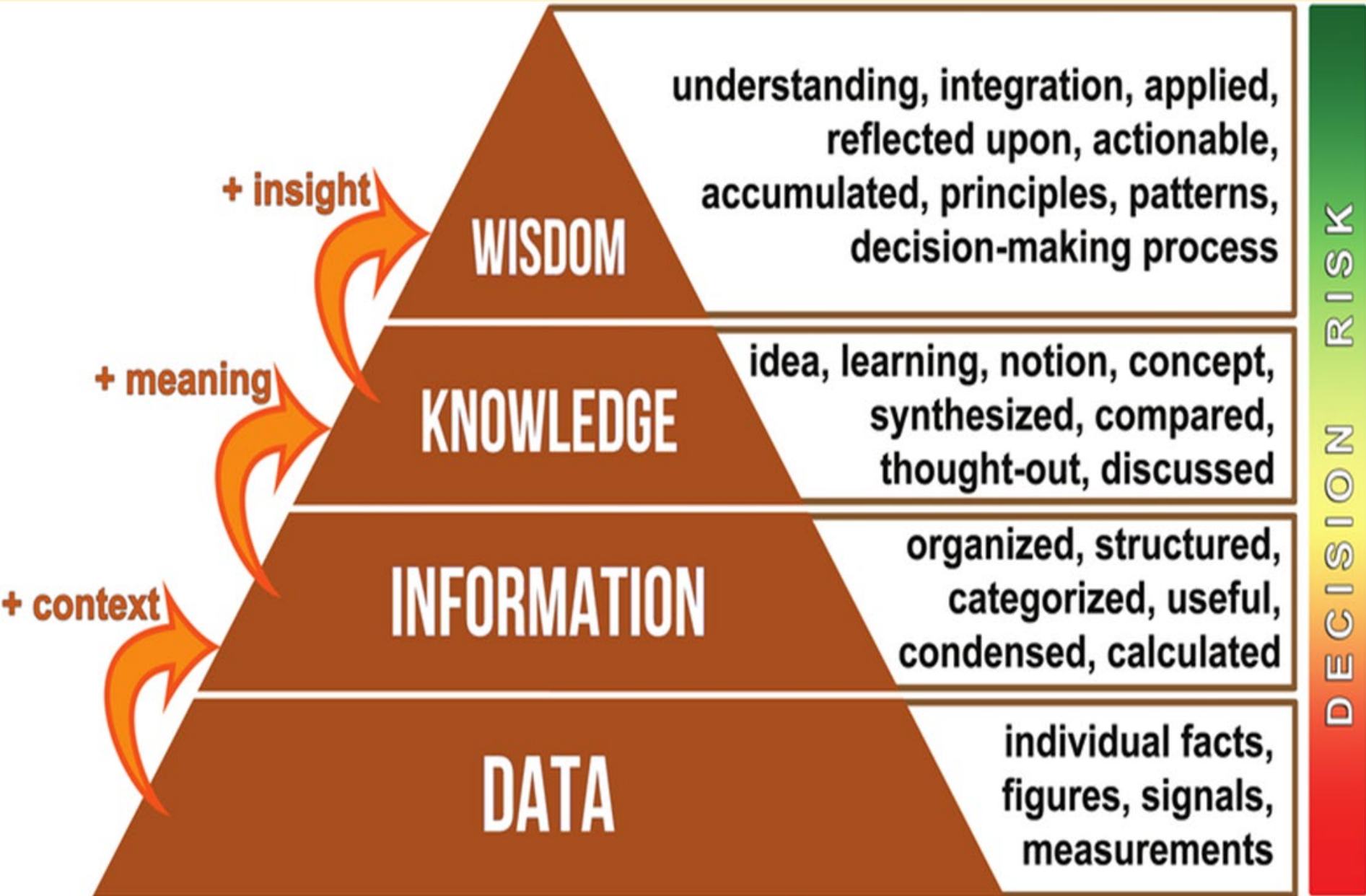
Predictive: *What's likely to happen?*

- Business strategies have remained fairly consistent over time
- Historical patterns being used to predict specific outcomes using algorithms
- Decisions are automated using algorithms and technology

Prescriptive: *What do I need to do?*

- Recommended actions and strategies based on champion / challenger testing strategy outcomes
- Applying advanced analytical techniques to make specific recommendations

DIKW 金字塔 (Data-Information-Knowledge-Wisdom)





<https://corticare.com/wp-content/uploads/revslider/home-new/clinical-trials-bg.jpg>

如何建立模型？



<https://www.bigskyassociates.com/hs-fs/hub/283820/file-2615489650-jpg/statistical-modeling-methods-for-security-data-analysis.jpg.jpg?width=217&name=statistical-modeling-methods-for-security-data-analysis.jpg.jpg>

驗證性資料分析

- 透過量化模型描述觀察結果：

$$\text{觀察現象} = \text{模型} + \text{誤差}$$

或是

$$y = f(x) + \text{error} ; \text{觀察值} = \text{訊號} + \text{雜訊}$$

- 數量化模型的關鍵：

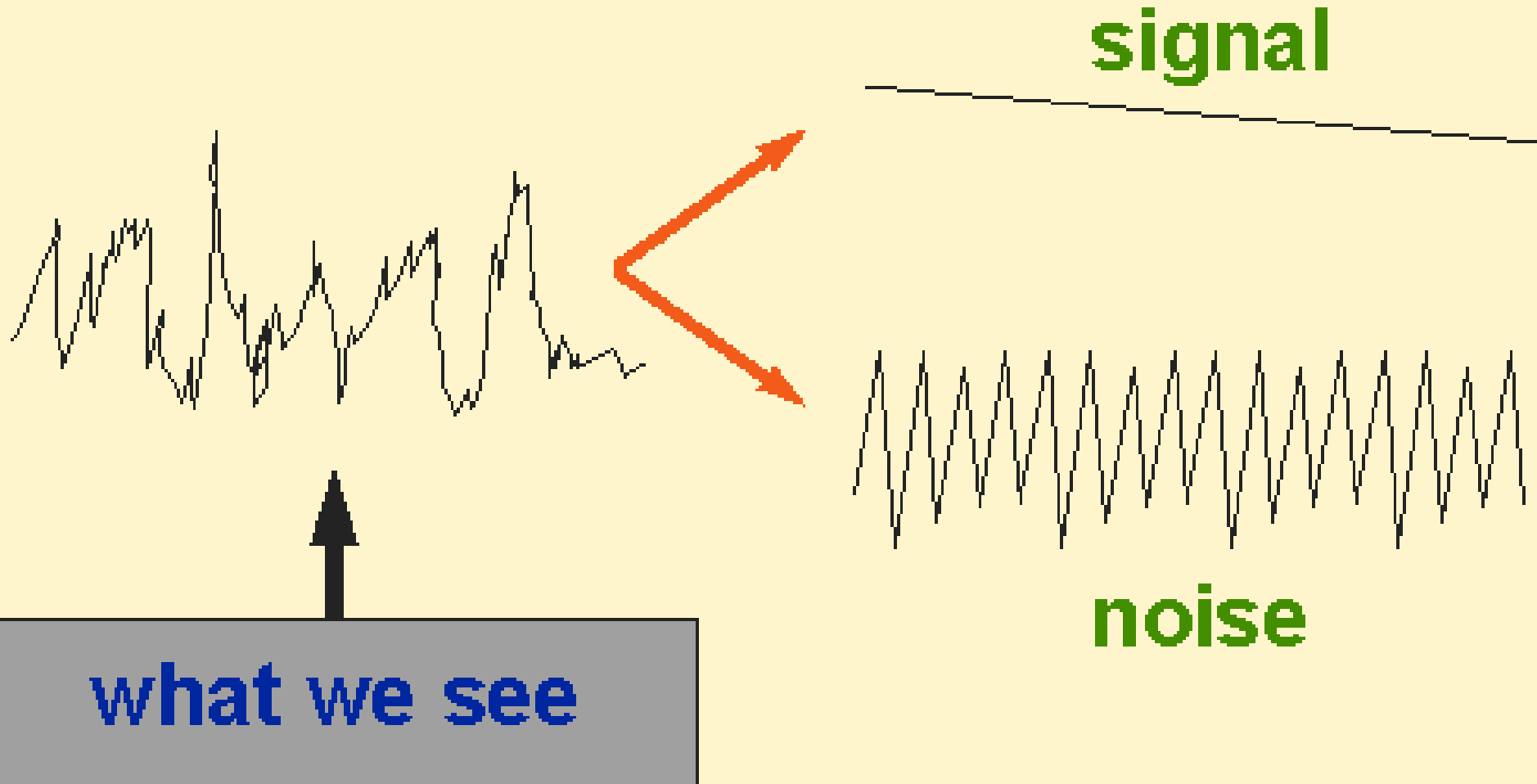
→ 量化目標值 y ：定義問題！

→ 選取關鍵變數： $x_1, x_2, \dots, x_p$

→ 建立量化模型：統計學習、機器學習。

觀察現象 = 模型 + 誤差

What we observe can be divided into:



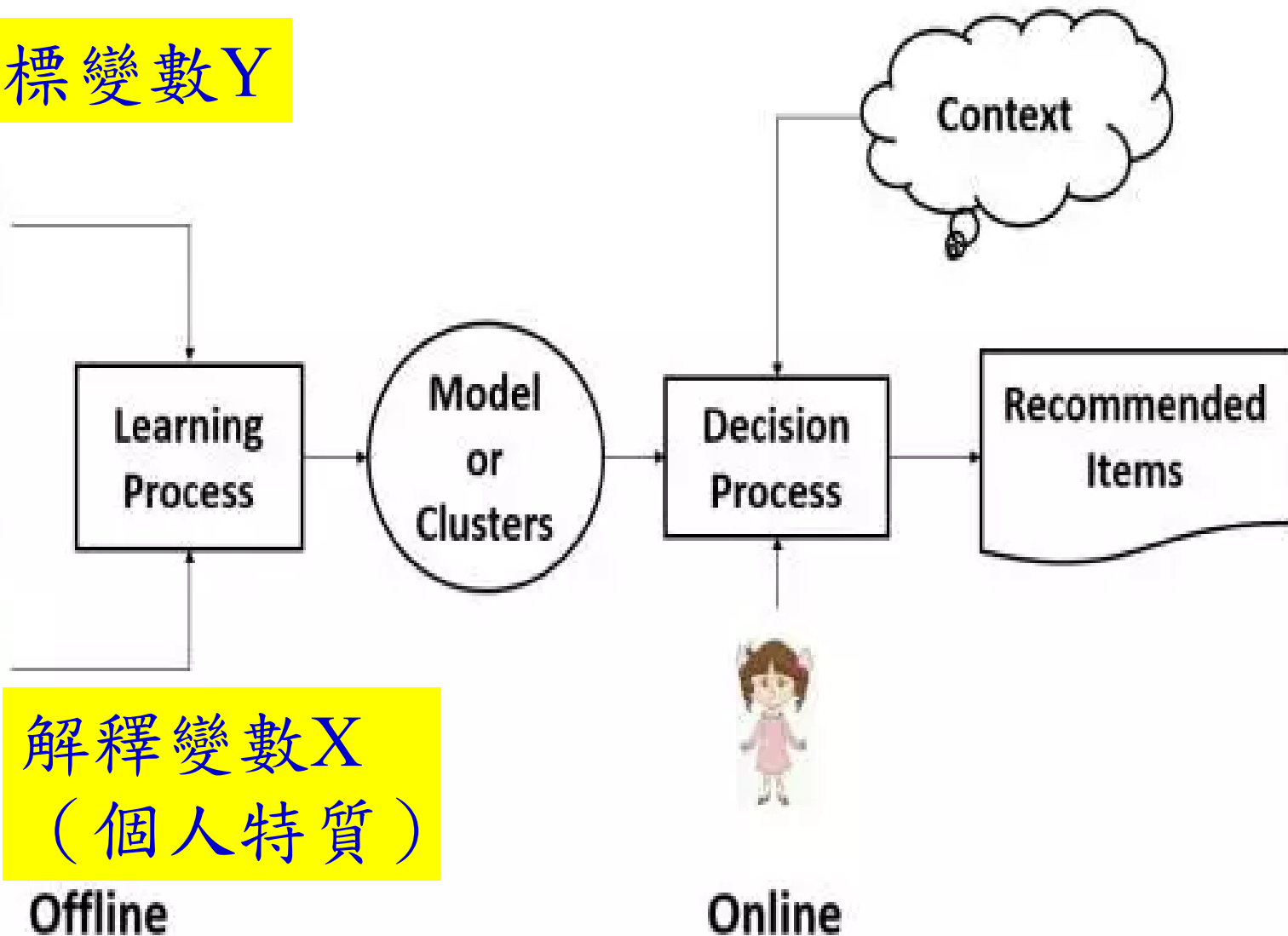
推薦系統(Recommendation System)分析流程

目標變數Y



解釋變數X
(個人特質)

Offline

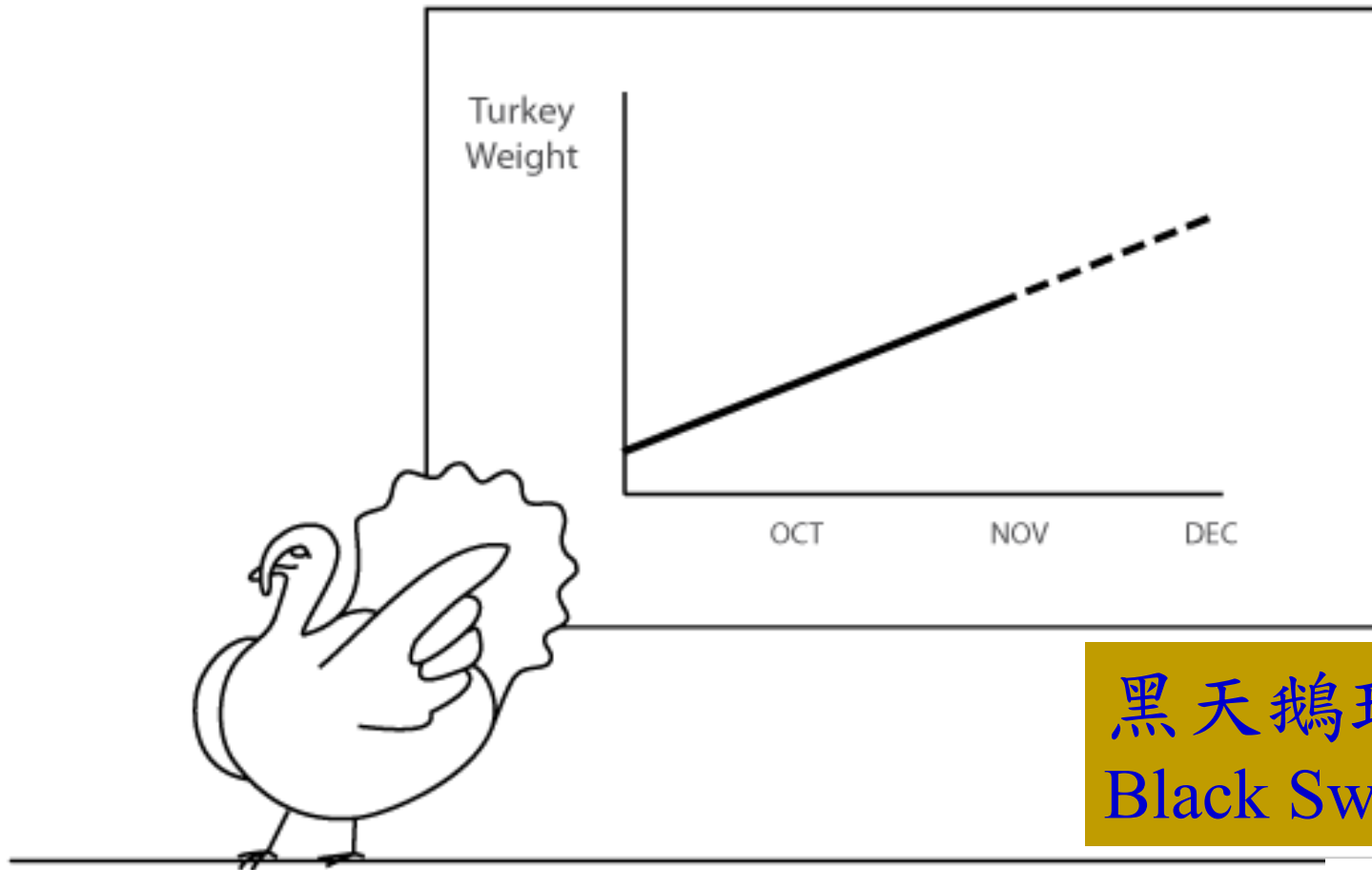


Online

解決問題的考量因素？

- 生活上遇到問題時，解決過程大致包括：
 - 確定問題目標及需求 Y ；
 - 蒐集相關資訊 X （資料來源、資料可信度、資料量及其品質）；
 - 資料整理、分析比對；
 - 決策評估標準 $f(X)$ （不同決策衍生出的優勢及劣勢、短期及中長期的影響）。

THANKSGIVING PREDICTIVE ANALYTICS



黑天鵝理論
Black Swan Theory

"I see no reason why excellent growth shouldn't continue.."

<https://i0.wp.com/timoelliott.com/blog/wp-content/uploads/2013/07/turkey-predictive-analytics.png?ssl=1>

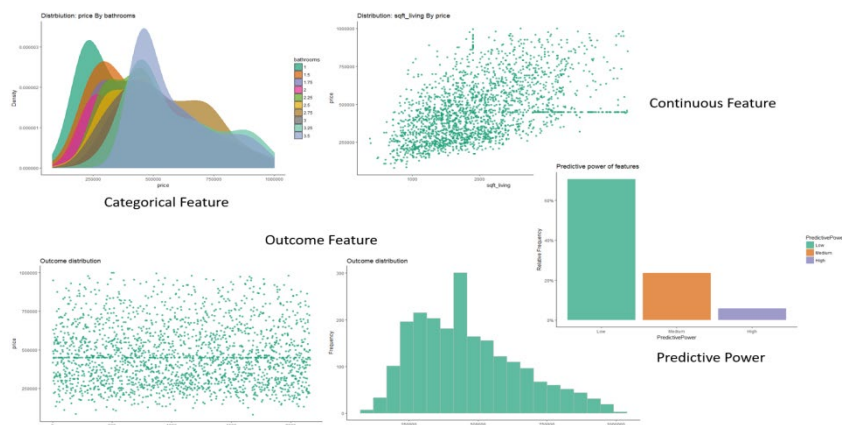
Q：灰犀牛理論？



<https://cdn.techinasia.com/wp-content/uploads/2016/10/bs1.jpg>

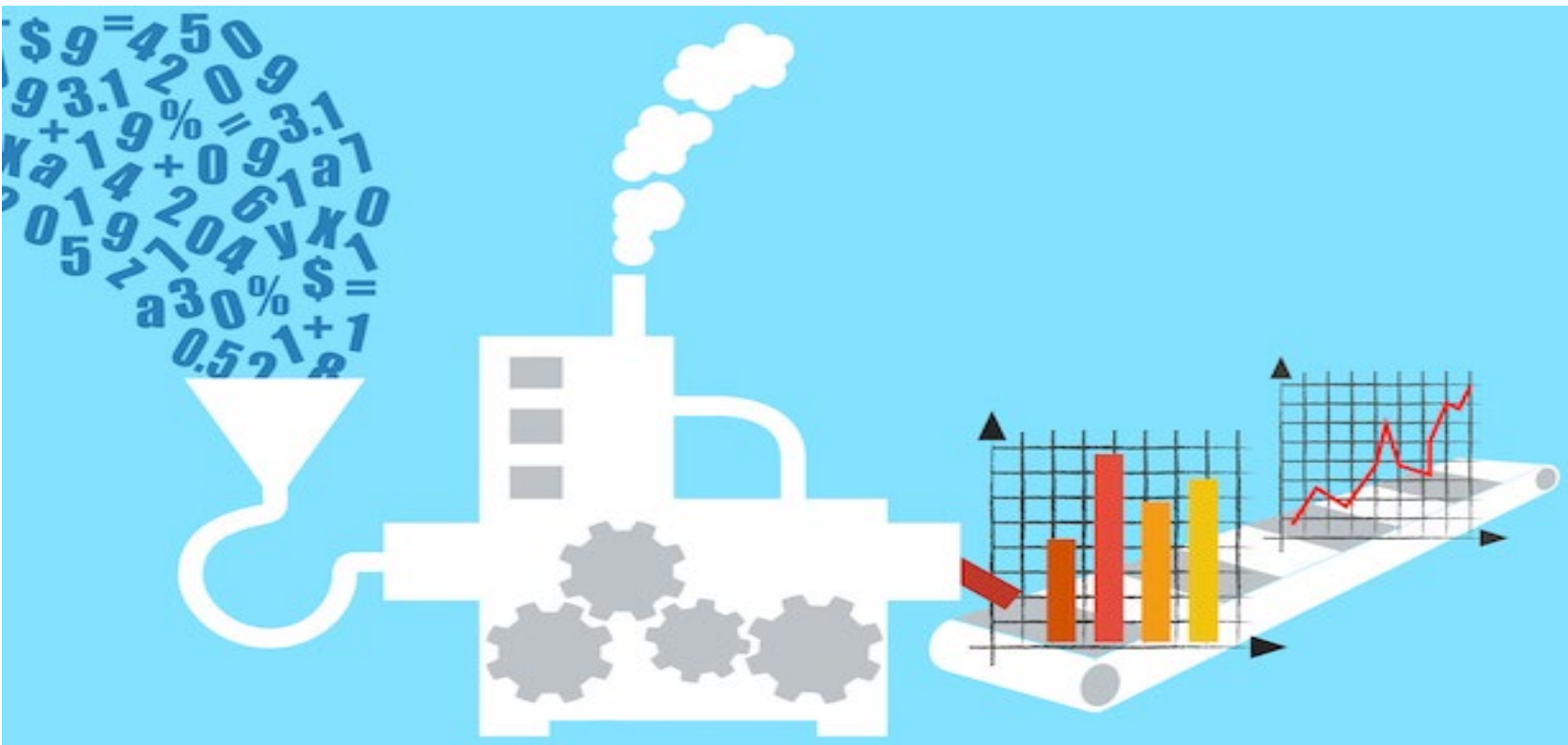
探索性資料分析(EDA)

- EDA首要目的在於資料偵錯、獲得資料的大略資訊、驗證已知結果。
- 圖形、表格在EDA中扮演重要的角色；並由分析結果中尋找合適的下一步分析方法。
- 使用任何的統計方法前，先確定該方法需要的假設條件是否滿足。



探索性資料分析(資料驅動)

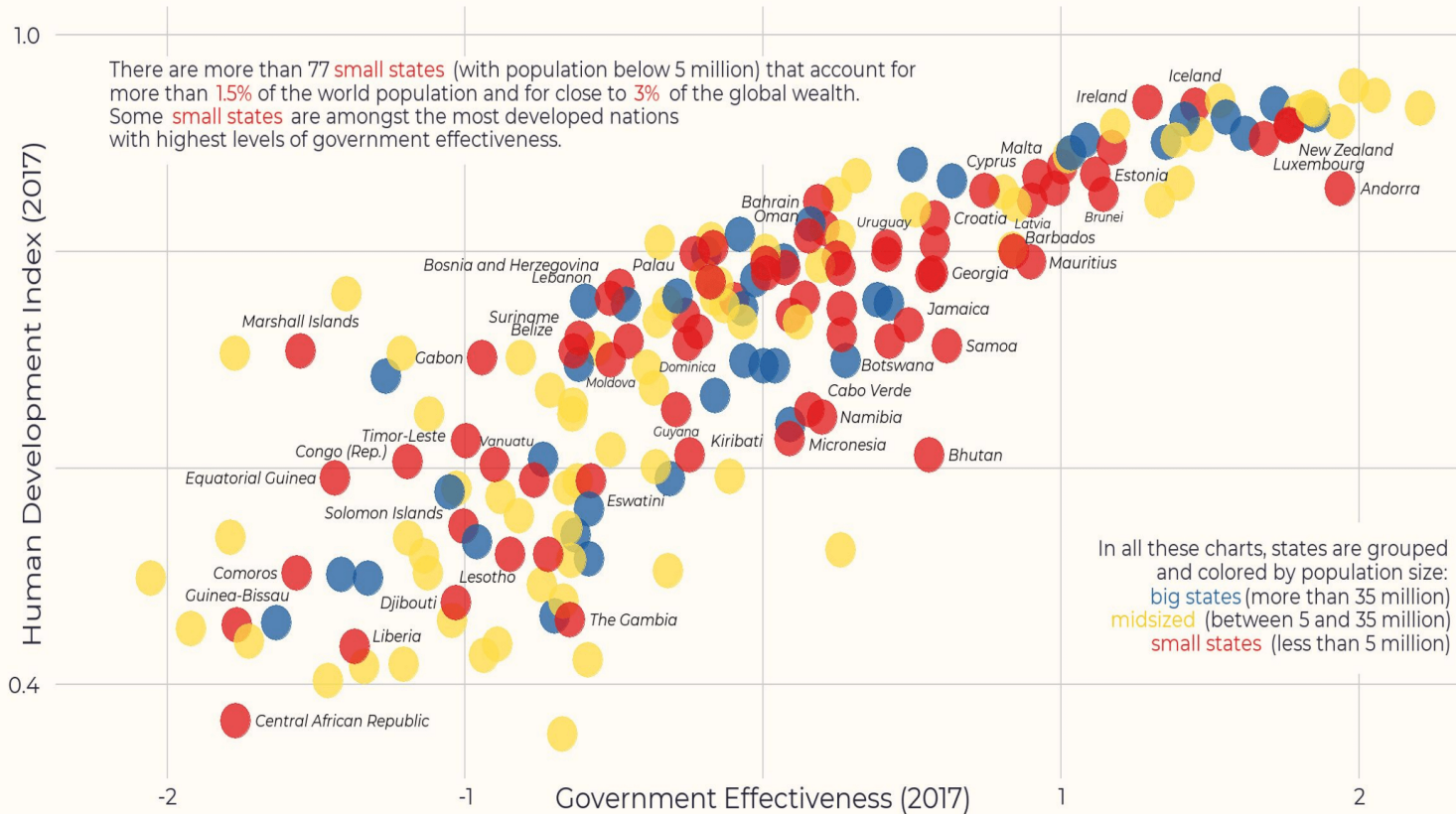
Exploratory data analysis (EDA) is an approach to analyzing data sets to summarize their main characteristics ... EDA is for seeing what the data can tell us beyond the formal modeling. ---Wikipedia



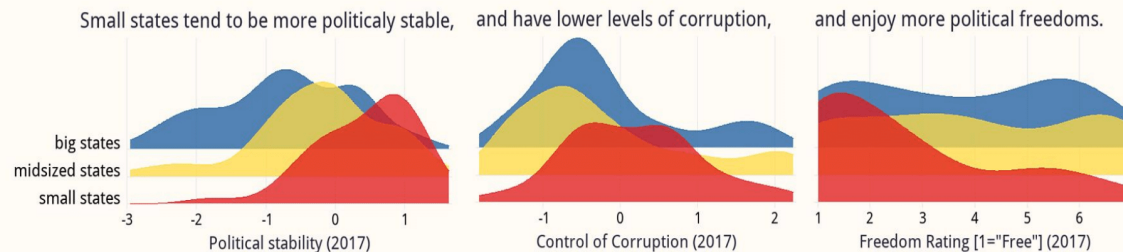
https://www.google.com/url?sa=i&url=https%3A%2F%2Fwww.aiche.org%2Facademy%2Fwebinars%2Fapplied-statistics-exploratory-data-analysis&psig=AOvVaw36ZuxAJqz27dLqU5IFzBMO&ust=1570108849384000&source=images&cd=vfe&ved=0CAIQjRxqFwoTCJC1qLXV_eQCFQAAAAAdAAAAAAAJ

第五個V：資料視覺化(Visualization)

Small States Can Be Big Players in Development and Good Governance



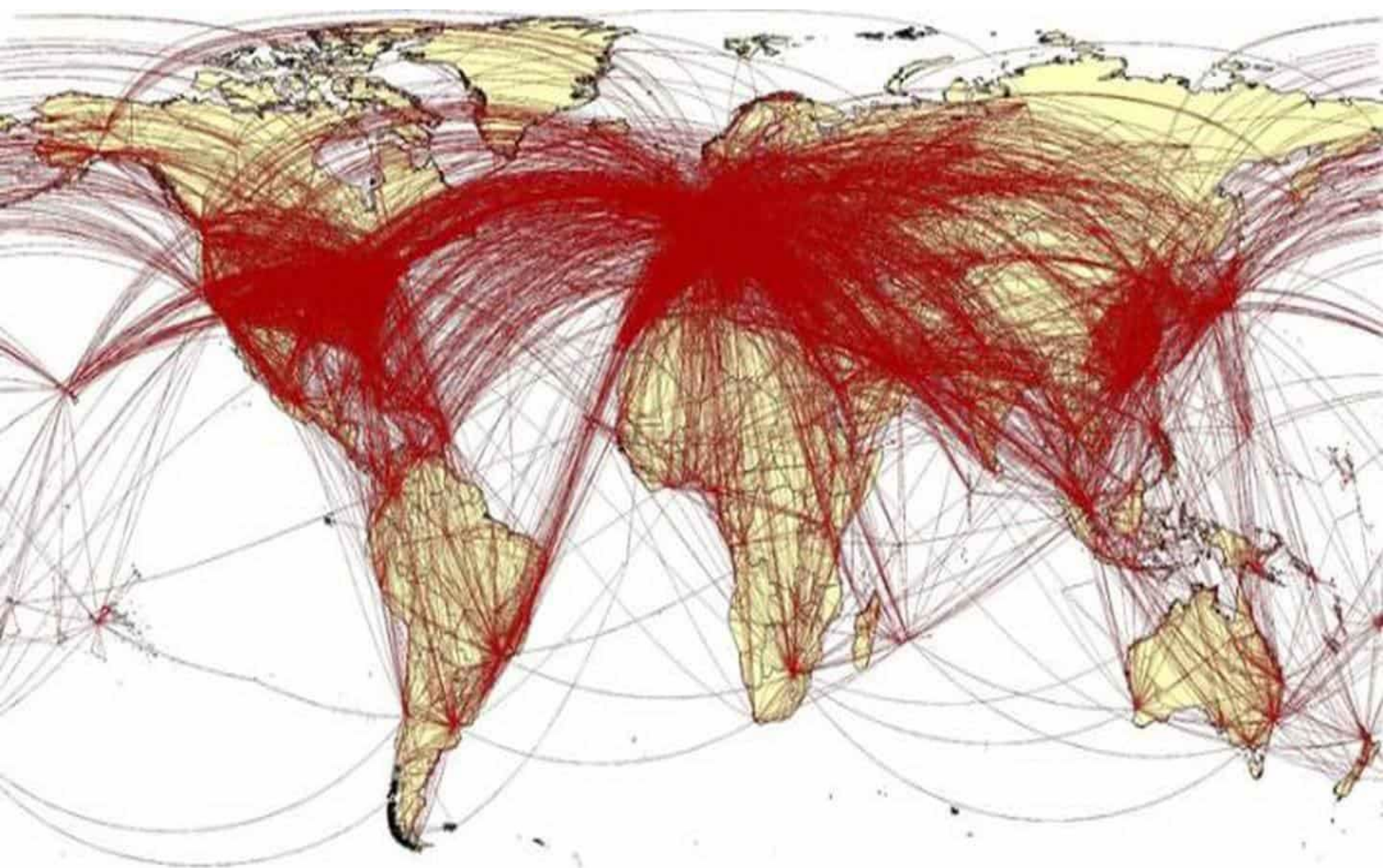
These charts show the distribution densities of different variables by three groups based on state size:



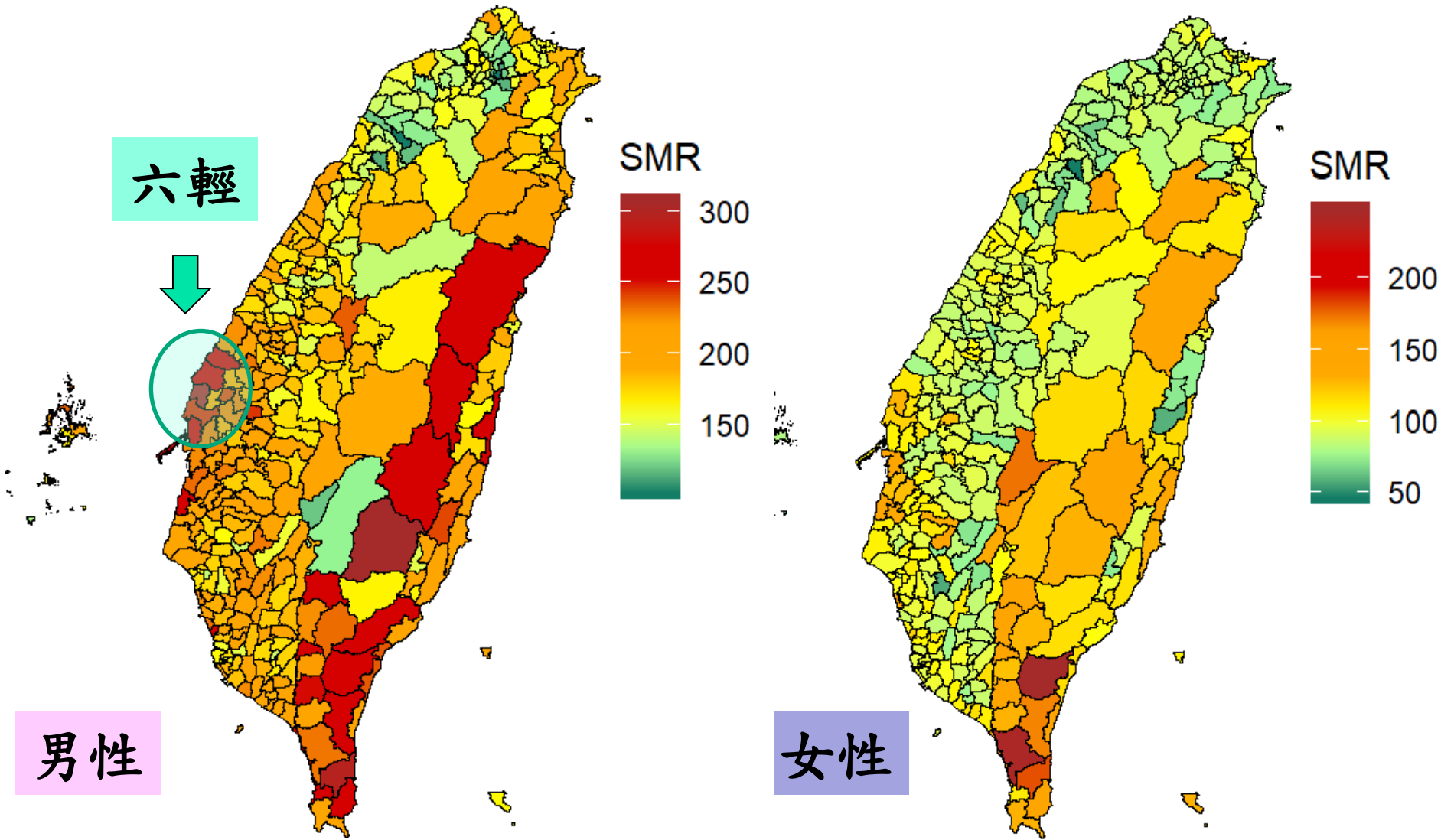
الجائزة العالمية
لفن عرض للبيانات
WORLD DATA
VISUALIZATION PRIZE

Data: Human Development Index from Human Development Reports. Government Effectiveness, Control of Corruption, and Political Stability from Worldwide Governance Indicators. Freedom Rating from Freedom House.

武漢大逃亡 Worldpop 分析路線圖

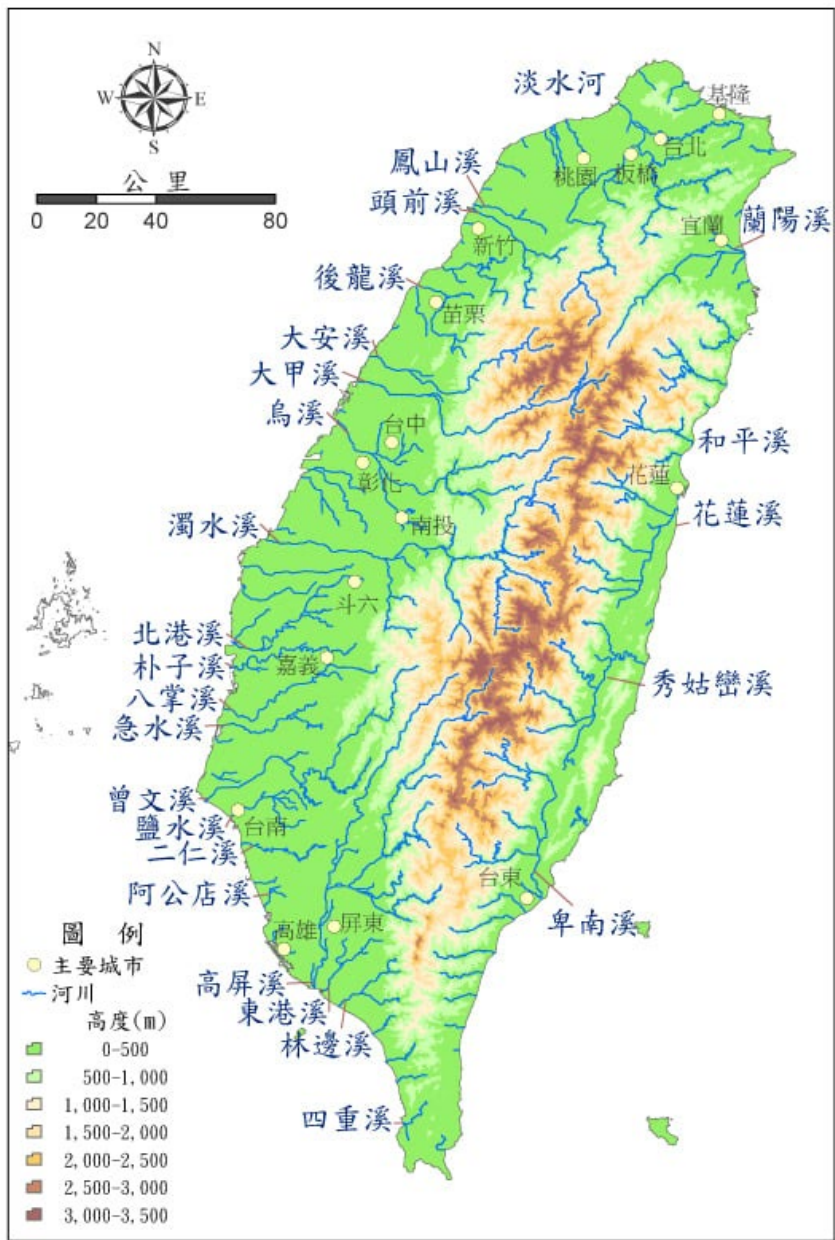


台灣2014~18年鄉鎮市區癌症死亡SMR



註：標準死亡比(SMR, Standard Mortality Ratio)可視為死亡率比值。

台灣主要河川分布圖



資料來源：經濟部水利署
地圖製作：余紀忠文教基金會 林書楷
中華民國100年9月

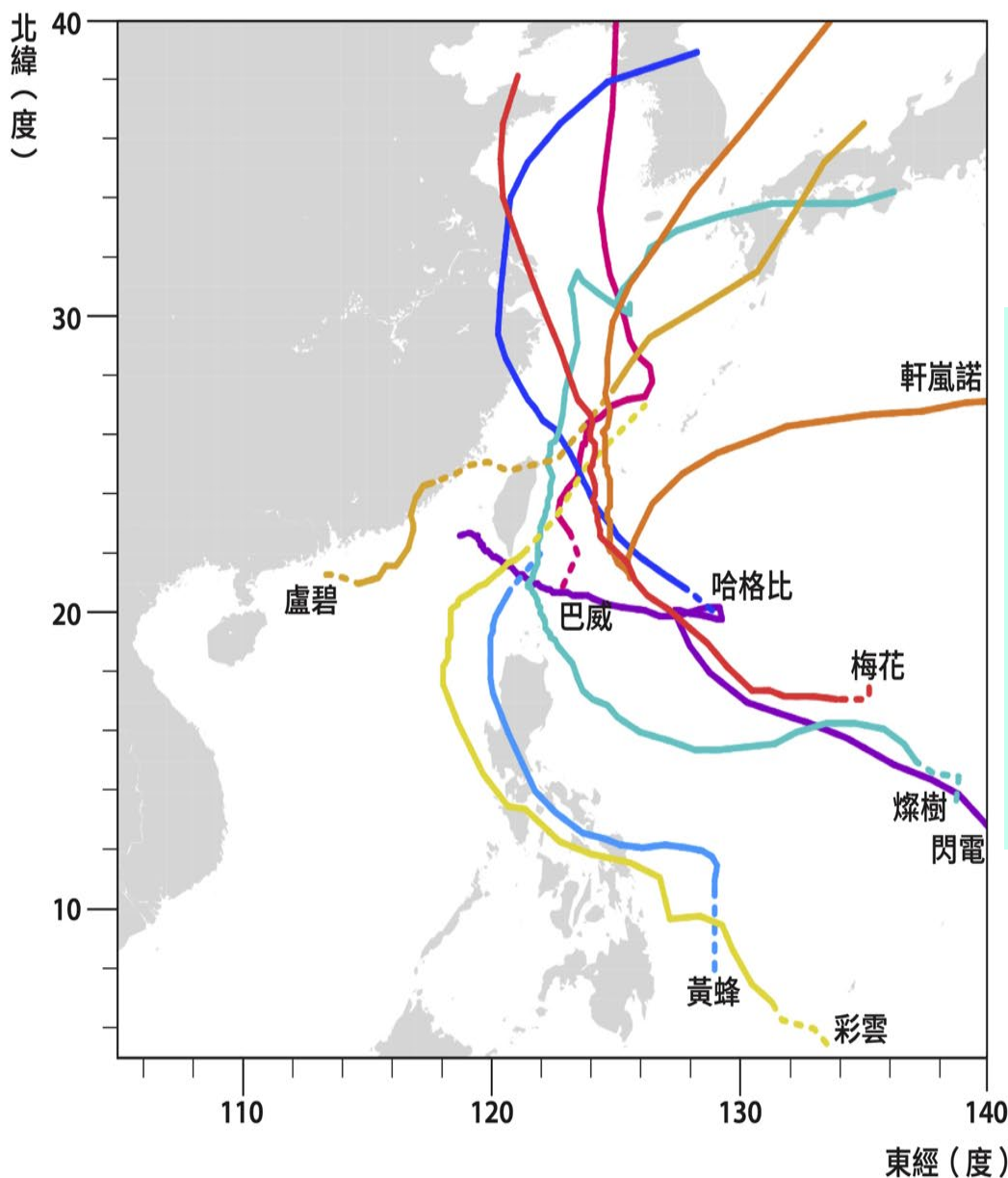
癌症地圖

癌症集中在河流下游



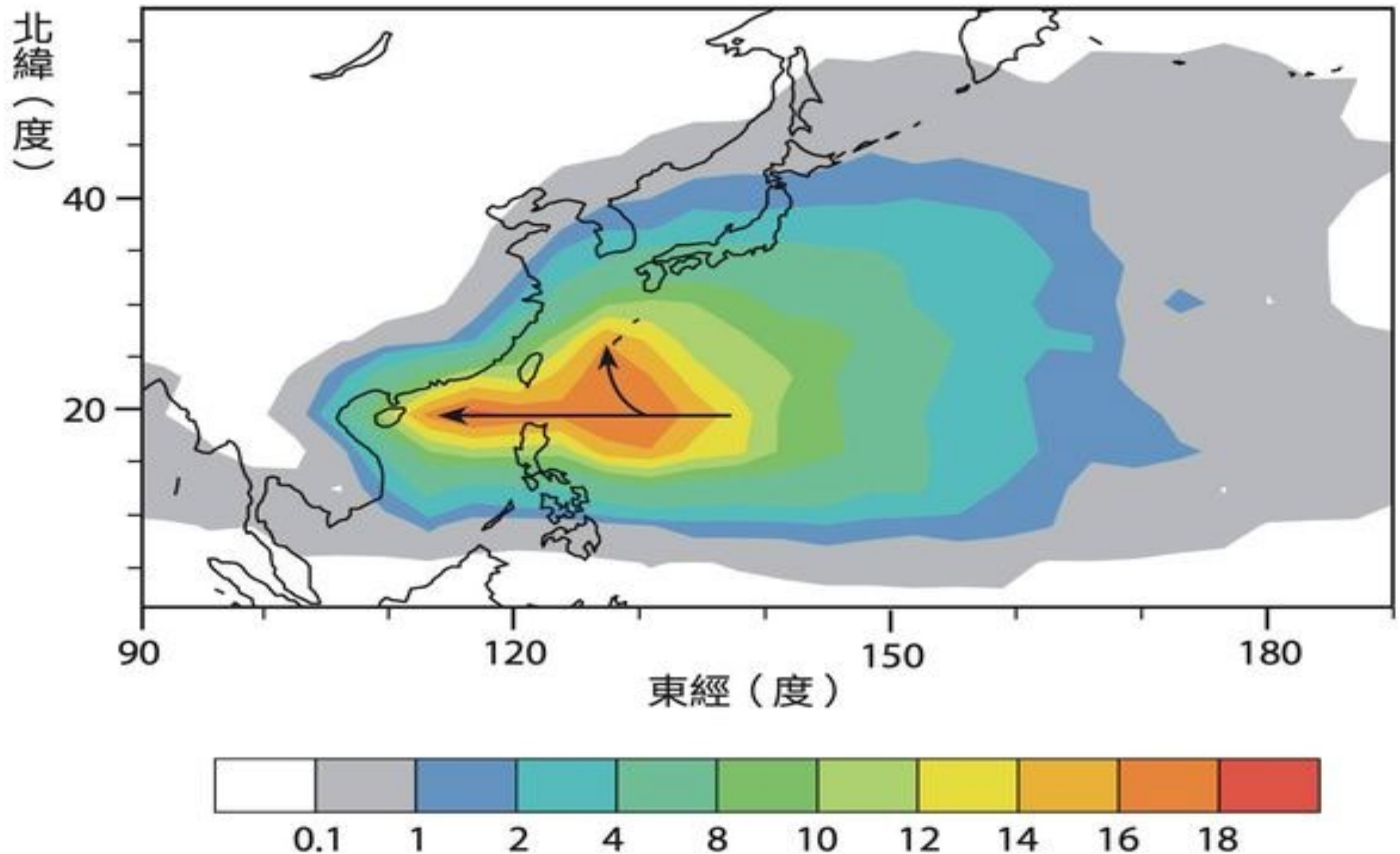
<https://www.masters.tw/270/%E7%99%8C%E7%97%87%E6%82%A3%E8%80%85%E7%82%BA%E4%BD%95%E9%9B%86%E4%B8%AD%E5%9C%A8%E6%B2%B3%E6%B5%81%E4%B8%8B%E6%B8%B8>

為什麼颱風又沒登陸！？(2020~2022)



東折西繞：
2020~2022
年共有九個
颱風行經台
灣周邊，皆
未登陸。

歷年颱風的路徑統計



1945~2022 年 6~11 月西北太平洋颱風路徑之頻率

圖形有時描述地更傳神！

A Picture



is worth
=

Creamy, delicious, yummy,
fudge ice cream, smooth,
chocolate-chip mint ice
cream, strawberry ice
cream with real chunks of
strawberry, colored sugar
sprinkles, waffle sugar
cone, sweet, wonderful,
tastes great, cold, nice to
eat, dessert, good yummy
toppings, chocolate
sprinkles, comforting,
good fun, dipping, terrific,

A thousand words.

©2003 E. Aoyama

圖形與表格

■除了基本的敘述統計量外，圖形與表格可以輔助判斷資料的特性。

→常見的圖形：Boxplot、Histogram

■這些圖表看似簡單，但仔細判讀仍可發現重要訊息，甚至不需進階統計分析，即能約略猜出分析的結論。

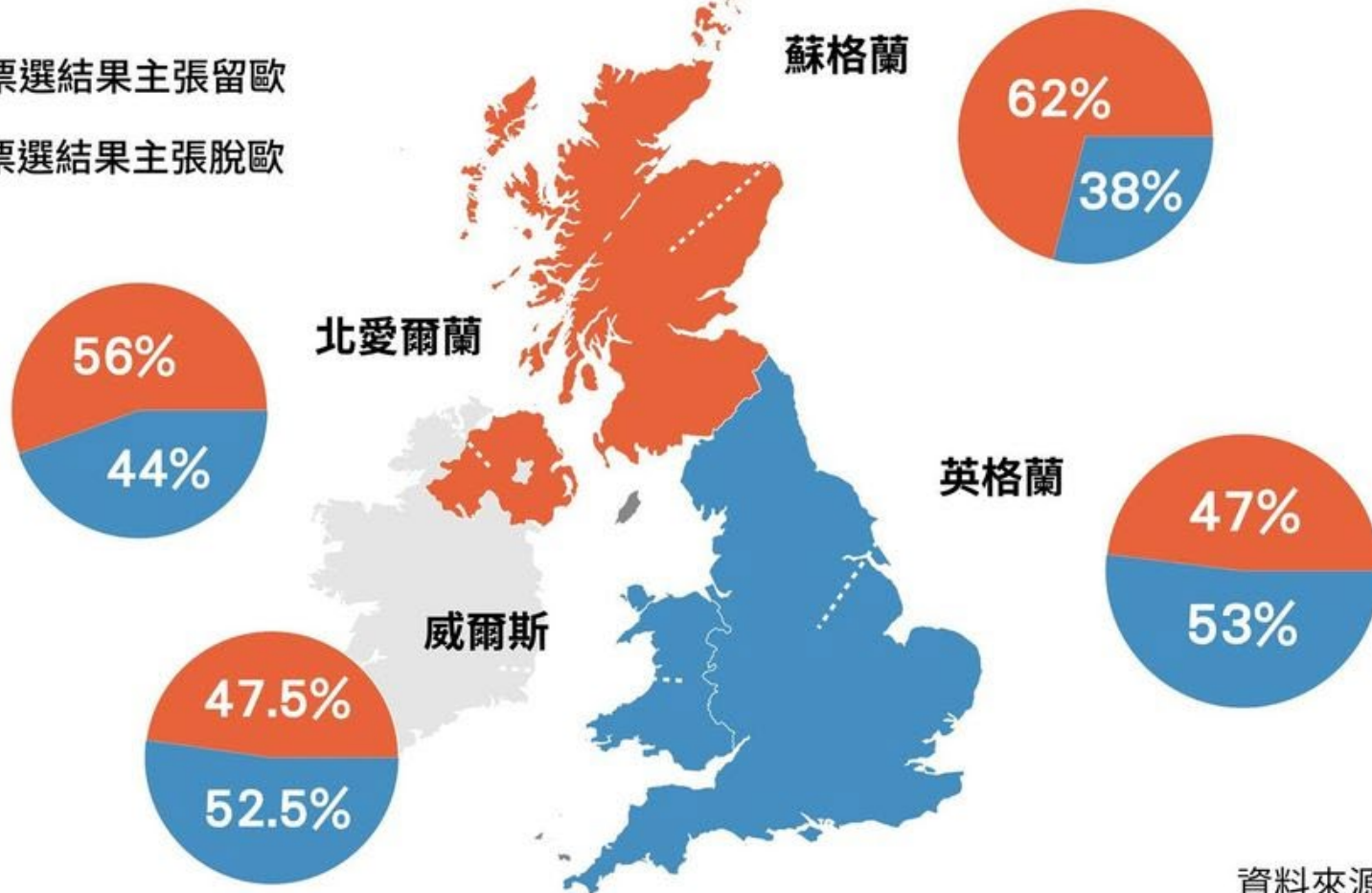
註：A picture is worth a thousand words!



英國四大構成 脫歐留歐各有所好

各構成國投票結果

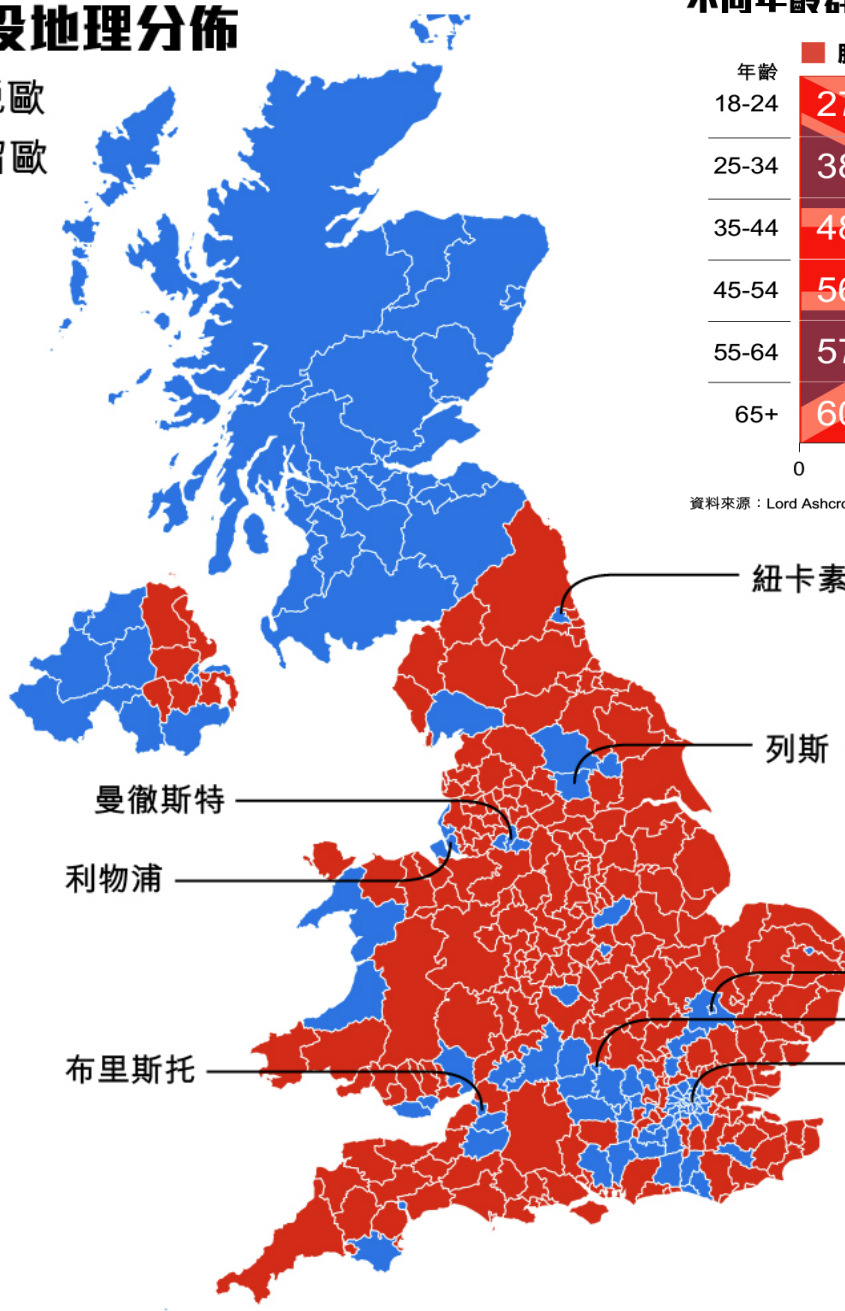
票選結果主張留歐
票選結果主張脫歐



資料來源:BBC

脫歐選舉公投地理分佈

■ 大多數支持脫歐
■ 大多數支持留歐



紐卡素 (台譯新堡)

列斯 (台譯里茲)

曼徹斯特

利物浦

劍橋

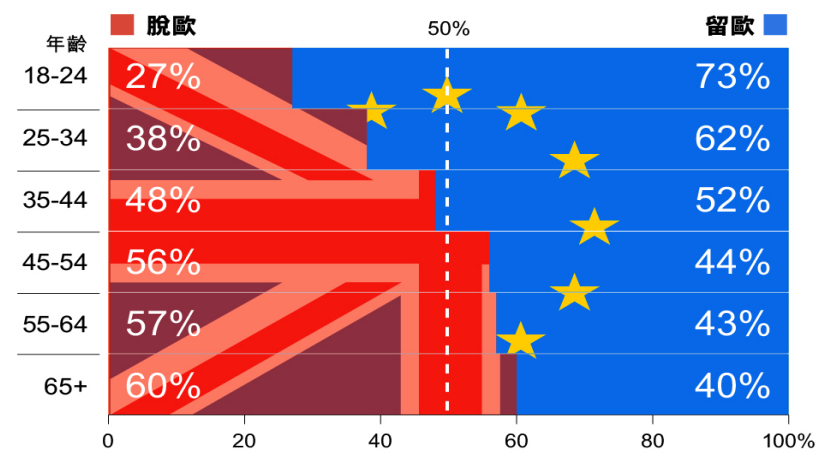
牛津

倫敦

布里斯托



不同年齡群投票結果



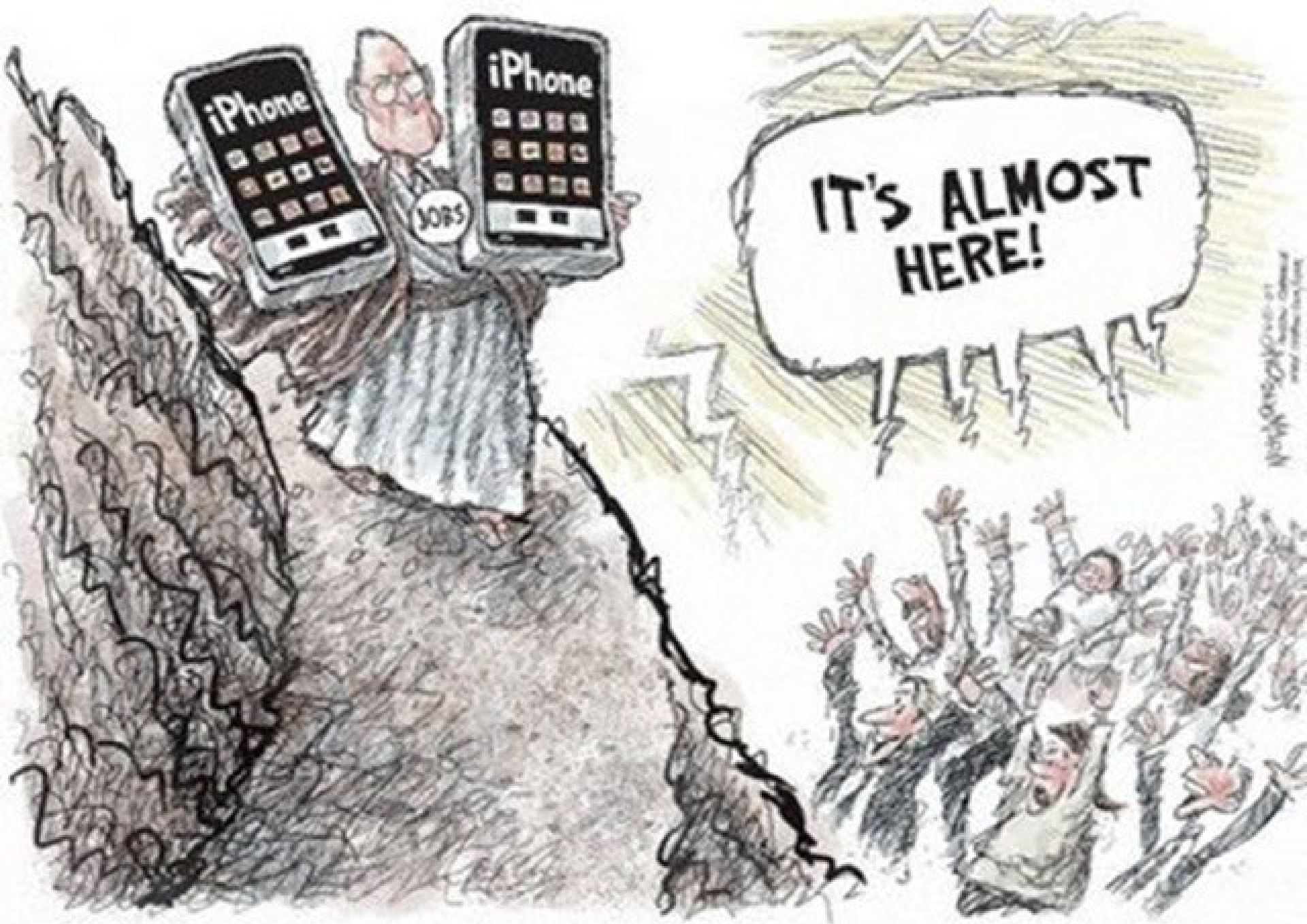
資料來源：Lord Ashcroft Polls、BBC

<https://theinitium.com/article/20160627-opinion-britishandeu/>



<https://theinitium.com/article/20160627-opinion-britishandeu/>

資料來源：BBC

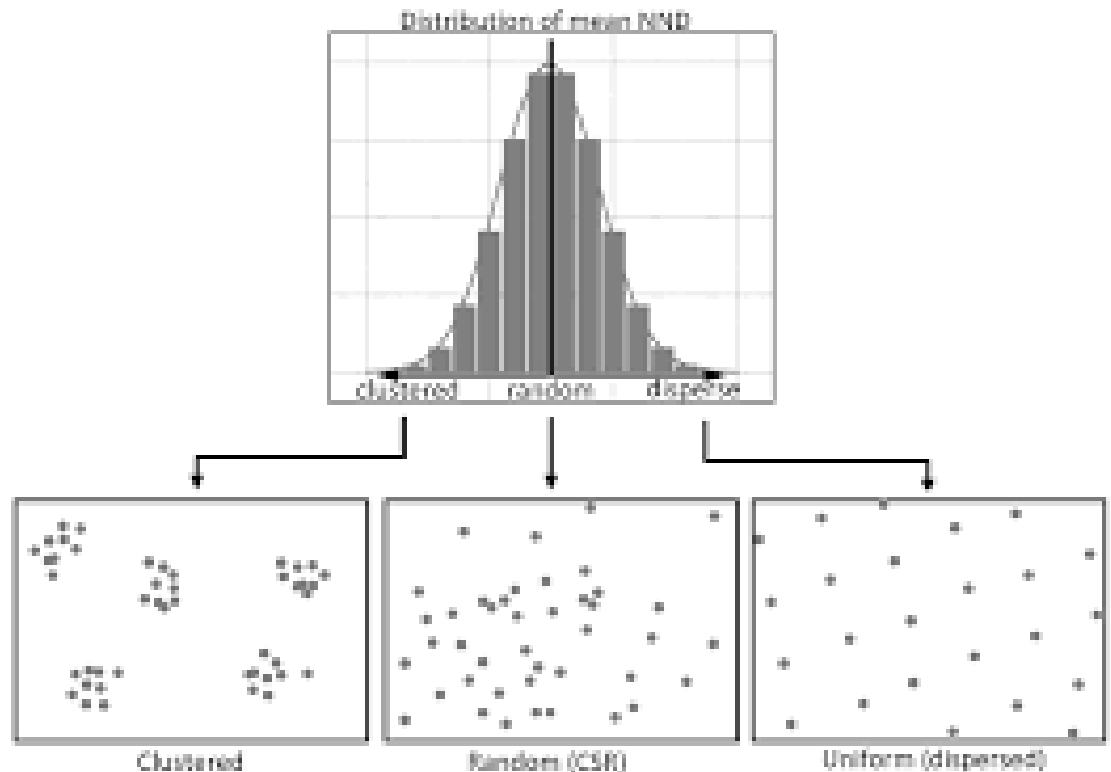


統計與規律

■ 透過不同角度發掘資料的關鍵特性：

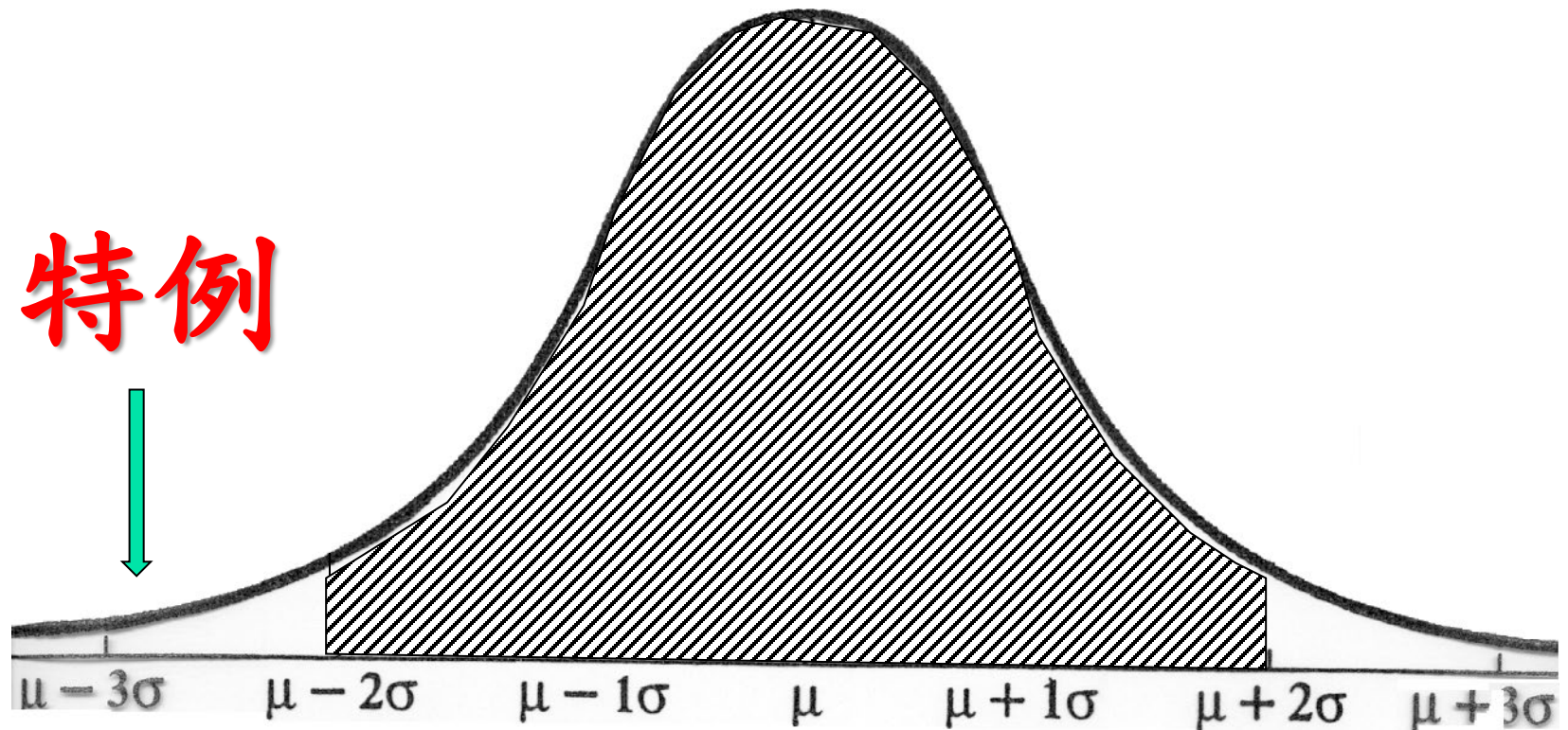
→ 共同特徵(Regularity)；

→ 特殊性質。

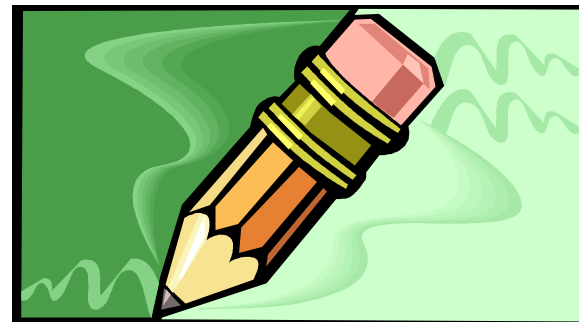


原則!!

特例



敘述性統計量



- 探索性或基本資料分析(Explanatory Data Analysis或Initial Data Analysis)是資料分析中最基本、也是非常重要的一個步驟，資料分析的成敗往往在這個步驟中決定。
- 敘述性統計量包括資料的基本特性，例如：平均數、標準差、所佔比例(圖表)等。
- 一般的分類方式為：
集中趨勢量數、差異量數



基本資料分析

- 資料偵錯

- 資料輸入錯誤、尋找可能的離群值。

- 初步探索資料的特性

- 資料的集中、散佈趨勢。

- 驗證已知的結果

- 是否與已知的結果相同？



資料類型

資料通常可分成四種類型：

Nominal (名目)

Interval (區間)

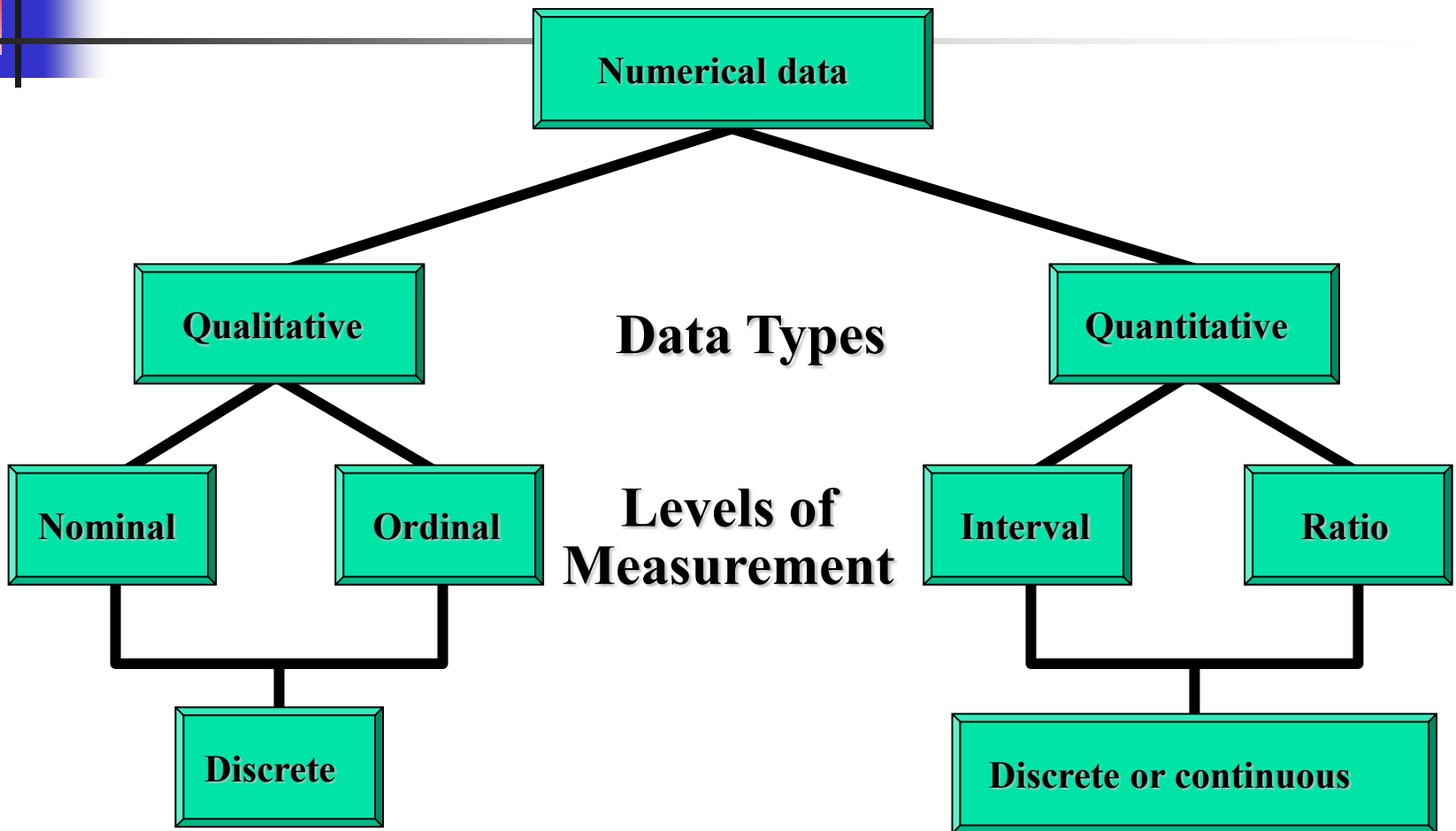
Ordinal (順序)

Ratio (比值)

蘊含的資訊與資料類型有關！

資料分析方式也與其類型有關！

Types of Data (資料類型)



■ 資料類型將直接影響分析方法的選取，並非所有資料都適合常見的統計方法，任意使用分析方法可能會得出令人啼笑皆非的結果。

→ $A > B, B > C$ 是否代表 $A > C$? (遞移律!)

	甲城市	乙城市	丙城市
A候選人	1	2	3
B候選人	2	3	1
C候選人	3	1	2

註：1代表最喜歡，3代表最不喜歡。



集中趨勢量數

- Mean (平均數)
- Median (中位數)
- Mode (眾數)
- Percentiles (百分位數)
- Quartiles (四分位數)

敘述統計量(範例)

例題一、試以文字詮釋以下隨機抽出某公司業務部門20位員工的年齡：

41	25	25	33	27	31	42
35	36	32	36	41	34	29
34	31	34	35	32	35	

→ 平均數 = 33.4，中位數 = 34.0，
標準差 = 4.75，全距 = 17。



敘述統計量(續)

例題二、試以文字詮釋以下隨機抽出某公司20位員工去年請假的天數：

0 0 0 0 0 0 0 0 1 1

1 2 2 3 4 5 5 6 7 42



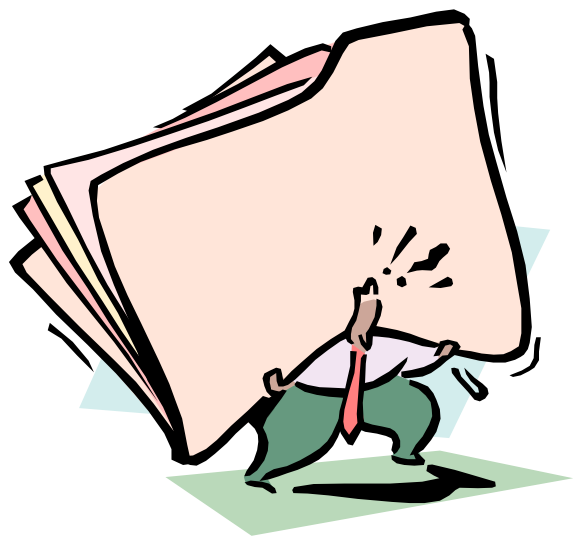
→ 你/妳 看到了甚麼現象？

敘述統計量(續)

例題三、街頭隨機訪問20位成年受訪者去年閱讀某月刊的期數：

0	1	11	0	0	0	2	12	0	0
12	1	0	0	0	0	12	0	11	0

→ 請問這是甚麼樣的月刊？



敘述統計量(續)

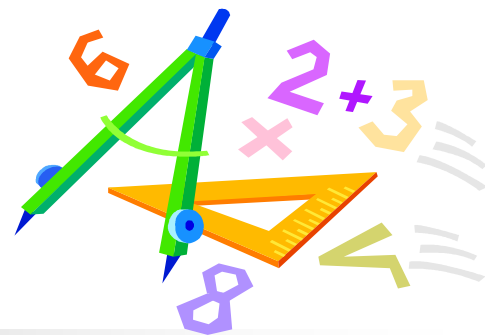
例題四、以下為隨機抽出某地區16位孕婦的身高(單位：公尺)：

1.57 1.55 1.60 1.52 1.68 1.57 1.62 1.55
1.65 1.52 2.55 1.60 1.55 1.60 1.62 1.57

→ 請問你/妳看到資料有何特性？

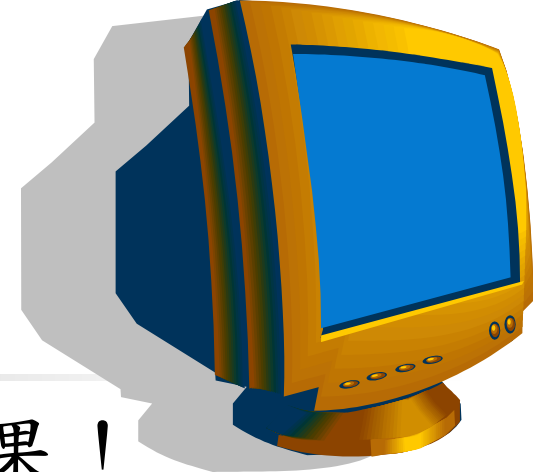


基本資料分析



- 基本資料分析的首要目的在於資料偵錯、獲得資料的大略資訊、驗證已知結果。(例如：正常 vs. 異常！)
- 因此，圖形、表格在基本資料分析中扮演重要的角色；並由基本資料分析的結果中尋找合適的下一步分析方法。
- 使用任何的統計方法前，先確定該方法需要的假設條件是否滿足。

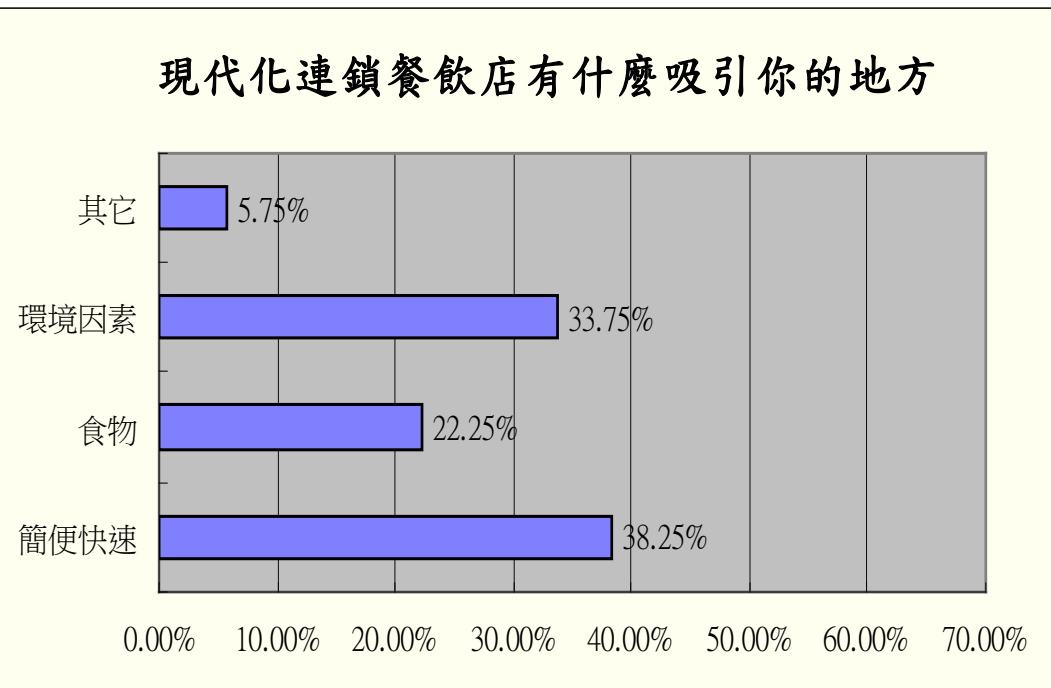
資料分析範例



■ 以表格、圖形展示資料更有效果！

長條圖(Bar Chart)

	選項	人數	百分比
1	簡便快速	153	38.25
2	食物	89	22.25
3	環境因素	135	33.75
4	其它	23	5.75
	N =	400	





Example: Singer Heights Story

Each singer in the NY Choral Society in 1979 self-reported his or her height to the nearest inch. Their voice parts in order from highest pitch to lowest pitch are Soprano, Alto, Tenor, Bass. The first two are typically sung by female voices and the last two by male voices.

→ What is the best way(s) to describe the heights of these singers?

Note: You may use find the data from web site

<http://lib.stat.cmu.edu/DASL/>

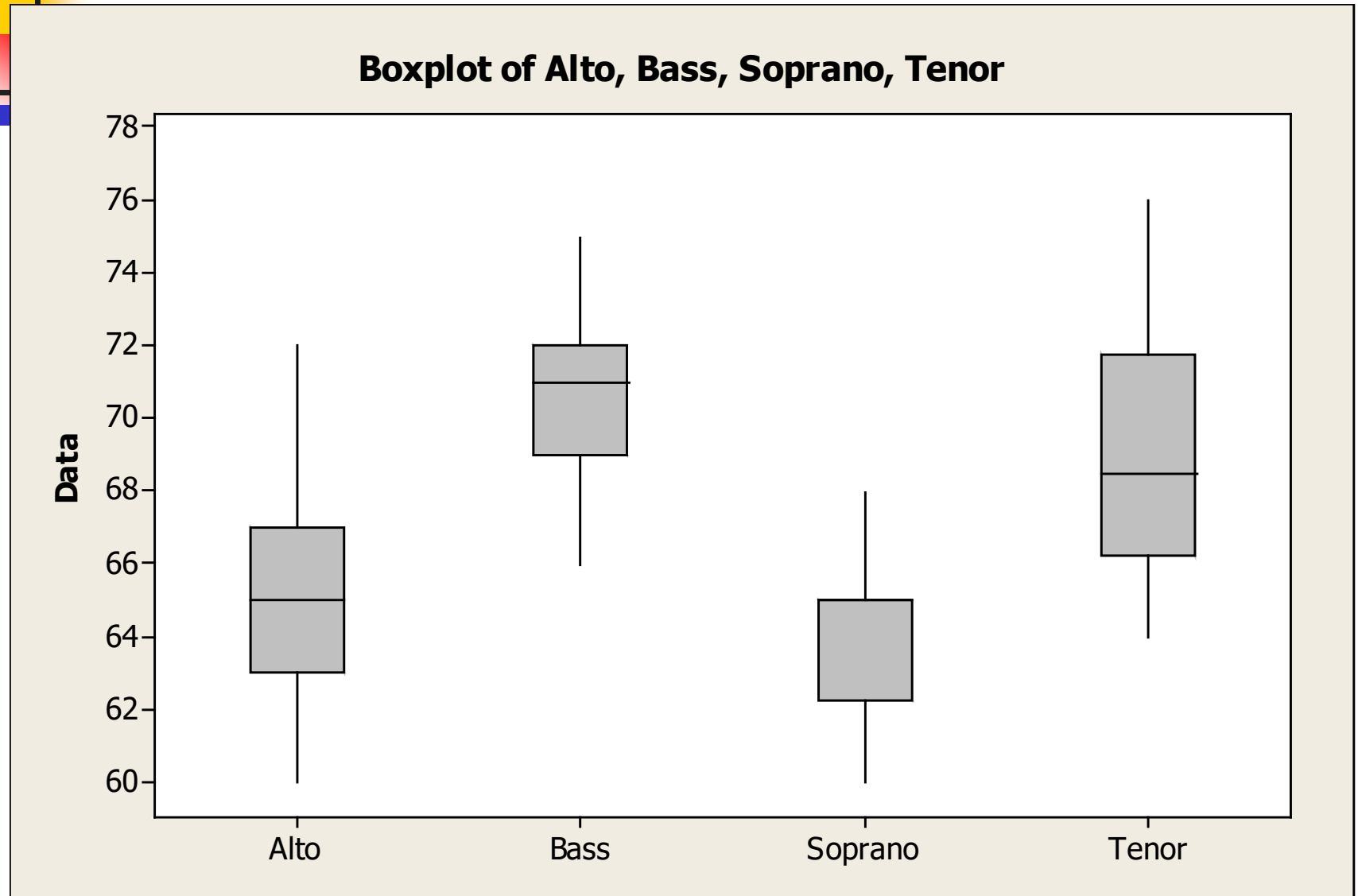


Some descriptive statistics

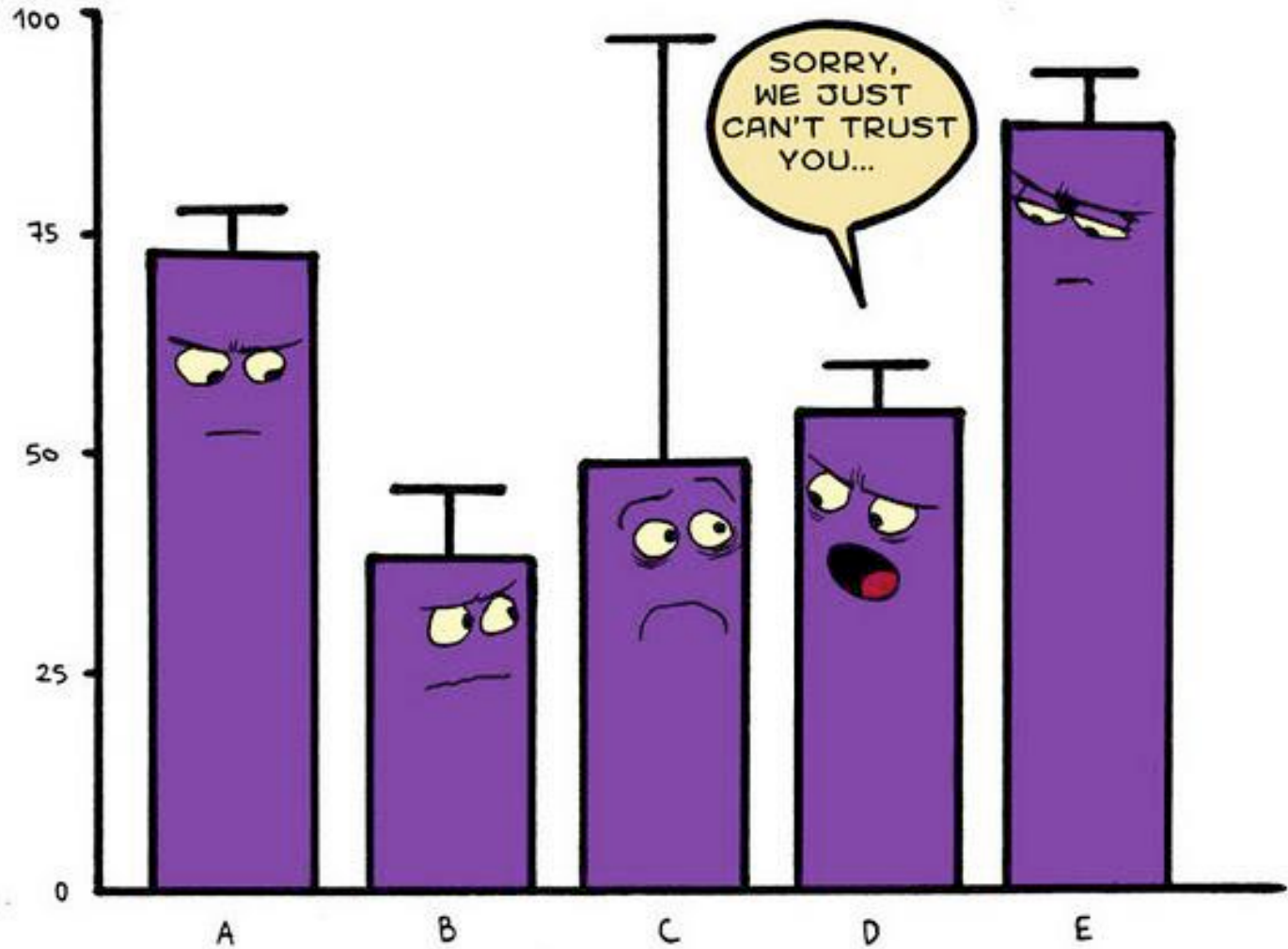
Variable	N	N*	Mean	SE Mean	StDev
Soprano	36	0	64.250	0.312	1.873
Alto	35	0	64.886	0.472	2.795
Tenor	20	0	69.150	0.719	3.216
Bass	39	0	70.718	0.378	2.361

Variable	Minimum	Q1	Median	Q3	Maximum
Soprano	60.000	62.250	65.000	65.000	68.000
Alto	60.000	63.000	65.000	67.000	72.000
Tenor	64.000	66.250	68.500	71.750	76.000
Bass	66.000	69.000	71.000	72.000	75.000

Compare the differences!



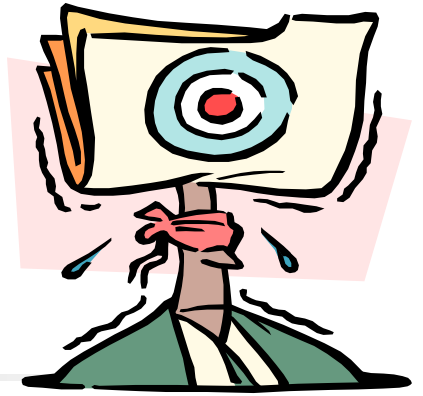
非我族類其心必異！



結果詮釋與推論限制

- 分析結果若能合理闡述，可達到「畫龍點睛」之效，但最忌諱忽略關鍵點，純粹就數字面來詮釋，反而變成「畫虎不成反類犬」。
- 研究結果的推論也需注意，不能僅從統計結果來看，也需瞭解推論結果代表的意義。





驟下結論(範例)

- 多數車禍發生在車速40~60公里/時，僅有少數在車速超過100公里。
- 開快車比較安全？
- 美國亞歷桑那州死於肺結核的比例最高。
- 亞歷桑那州的天氣易於感染肺結核？
- 調查小學生的拼字能力，發現腳愈大的拼字能力也較強。
- 腳的大小影響拼字能力？



驟下結論(續)

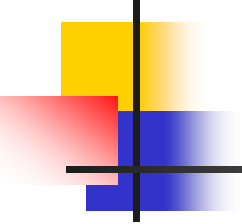
- 2001年大陸調查發現長壽者中，排行老大者最多。

→ 排行老大較長壽？

→ 抑或是排行老大者佔了多數？

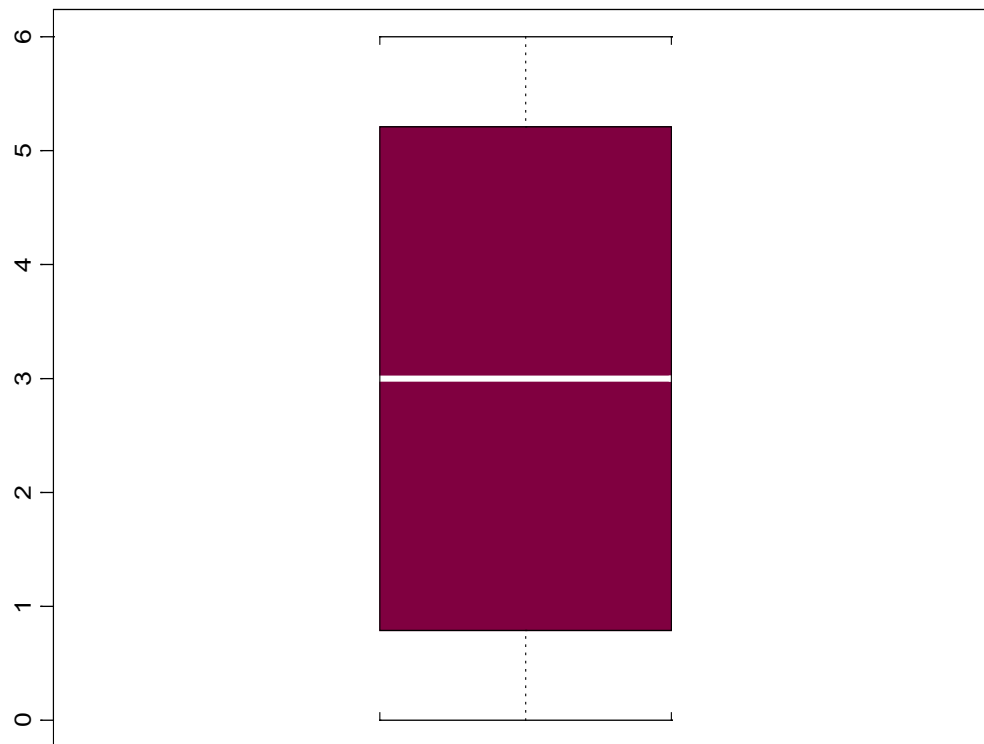
- 英國公務統計顯示在家裡生產者，發生意外的比例較在醫院生產者高，因此孕婦都應該在醫院生產。

→ 為什麼有些孕婦會在醫院以外的地方生產？

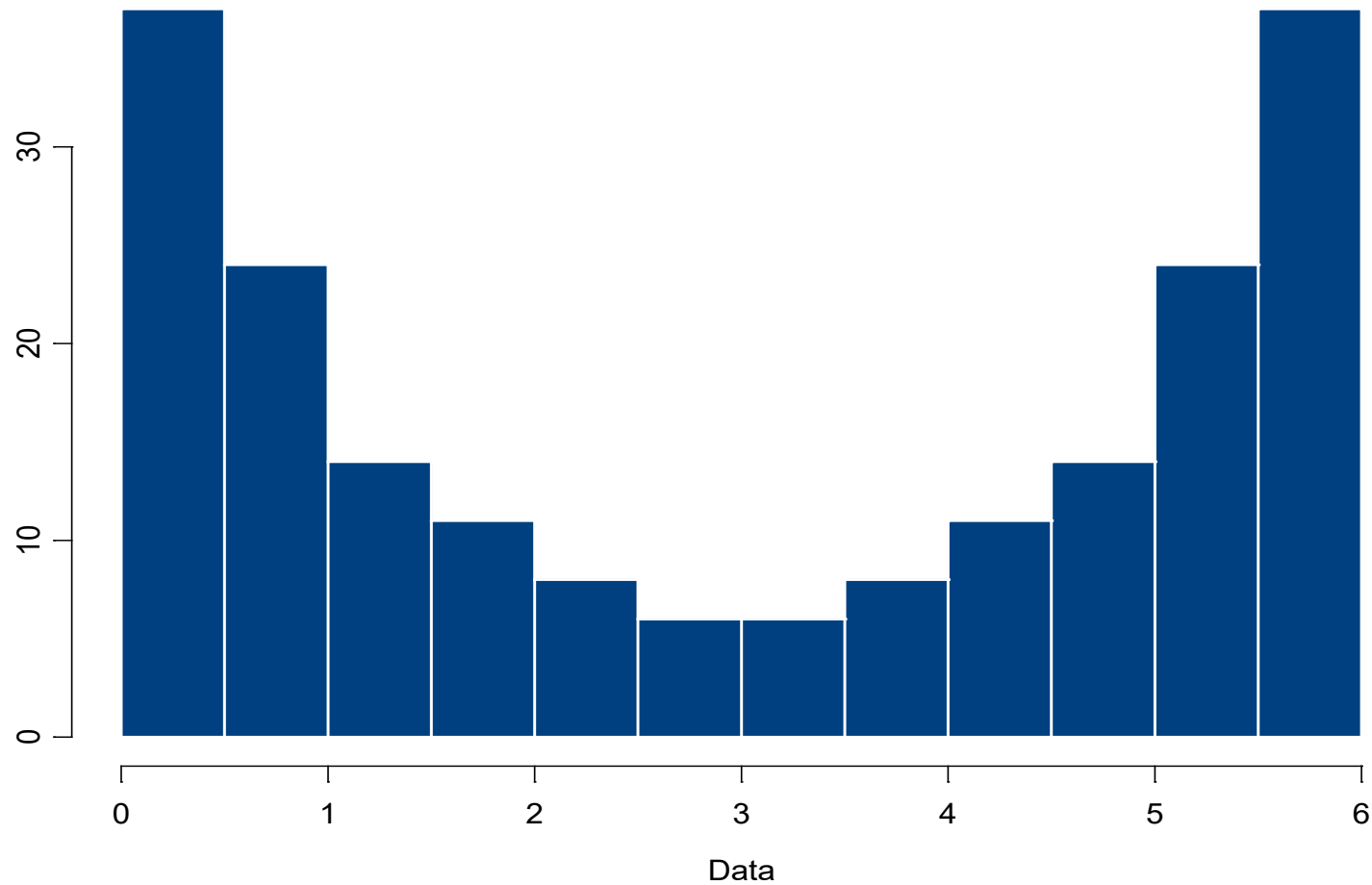


表格與圖形(範例)

例題五、大略敘述下圖資料的特性。



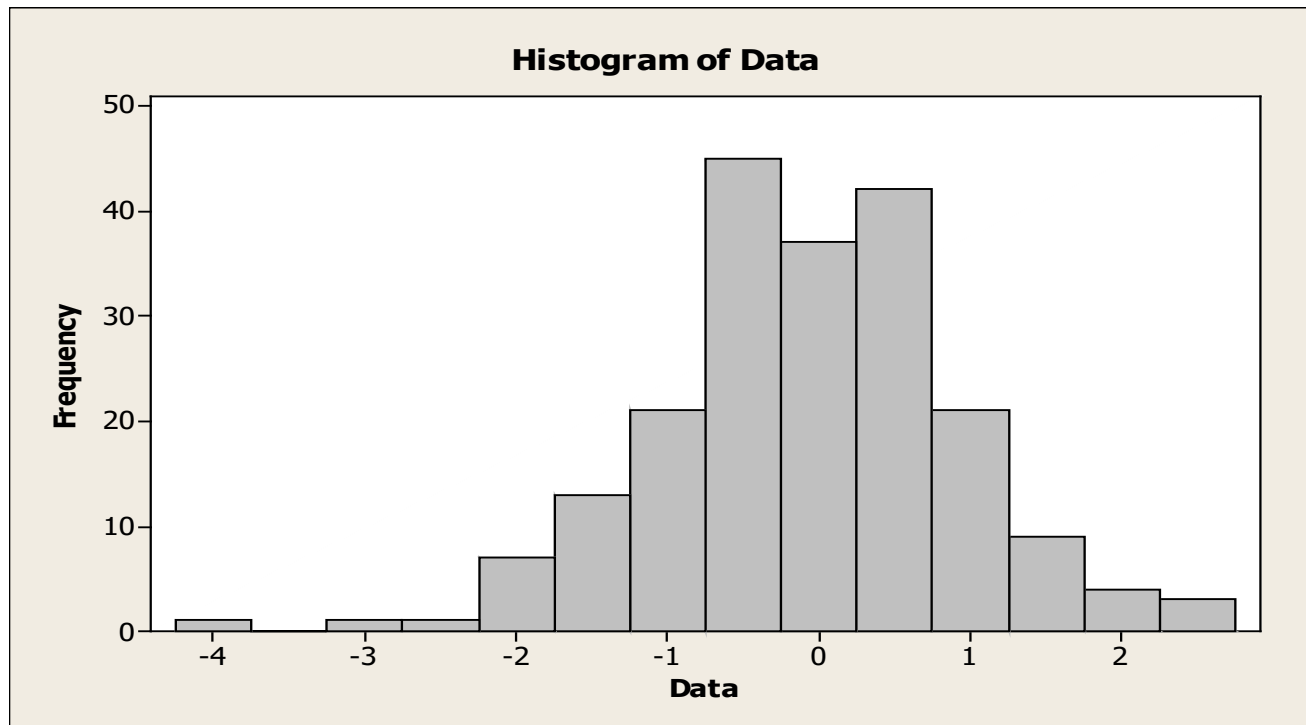
原始資料(你/妳猜對了嗎?)

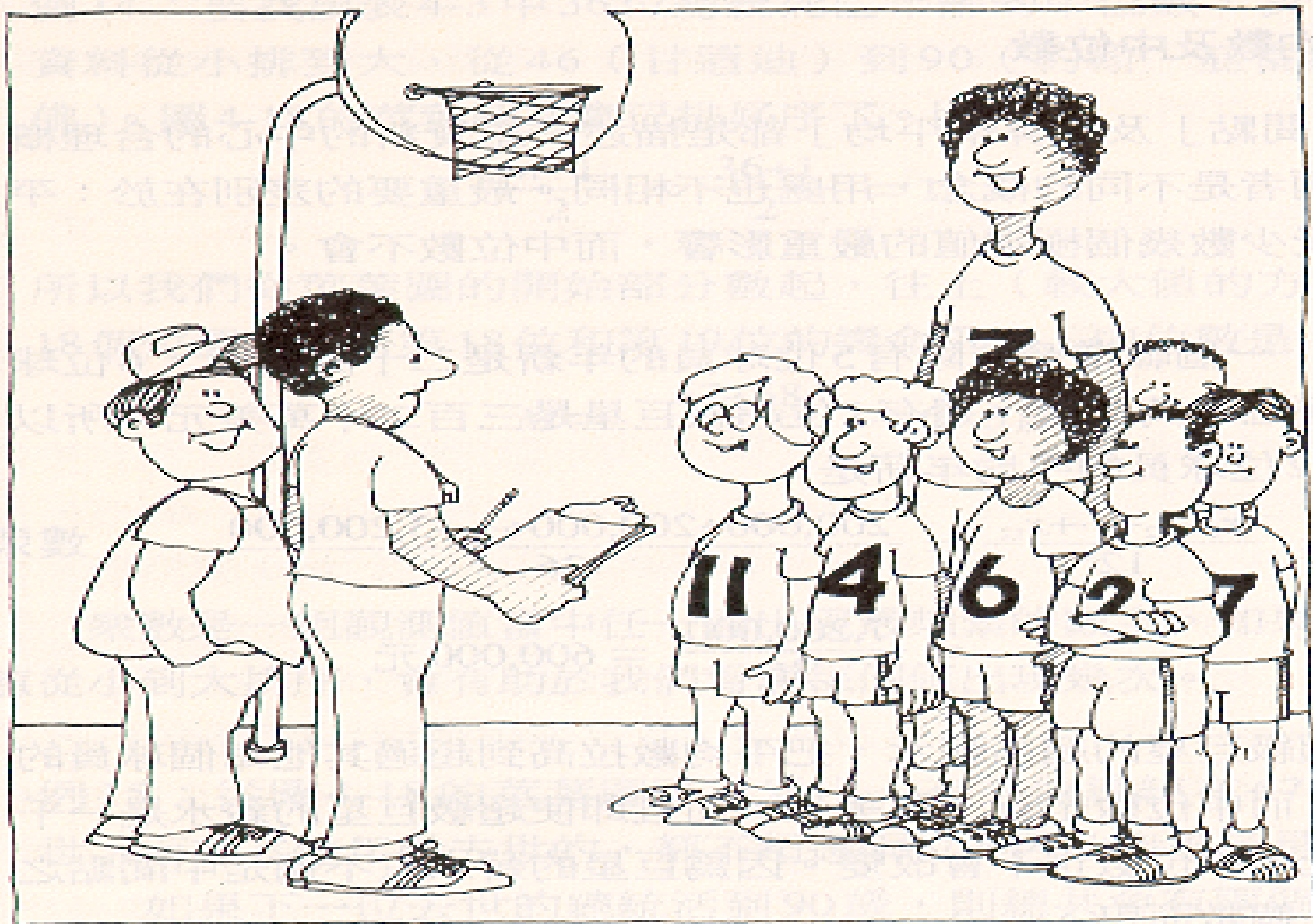


問題：常態分配的箱型圖具有哪些特性？

平均數與中位數

- 左偏(Left-skewed)：當少數觀察明顯小於一般觀察值(如下圖)，平均數將被這些觀察值拉下，但中位數較不受影響，此時中位數大於平均數。
- 少數觀察值較大時稱為右偏(Right-skewed)。





「我們是應該宣布我們的平均高度來嚇死對手，還是宣布我們的中位數高度來消除他們的戒心呢？」



老謝曝「很多優秀球員是姜豐年去求來的」：無所求的無私奉獻最感動人！

