



性別統計資料庫的探索、運用與視覺化¹

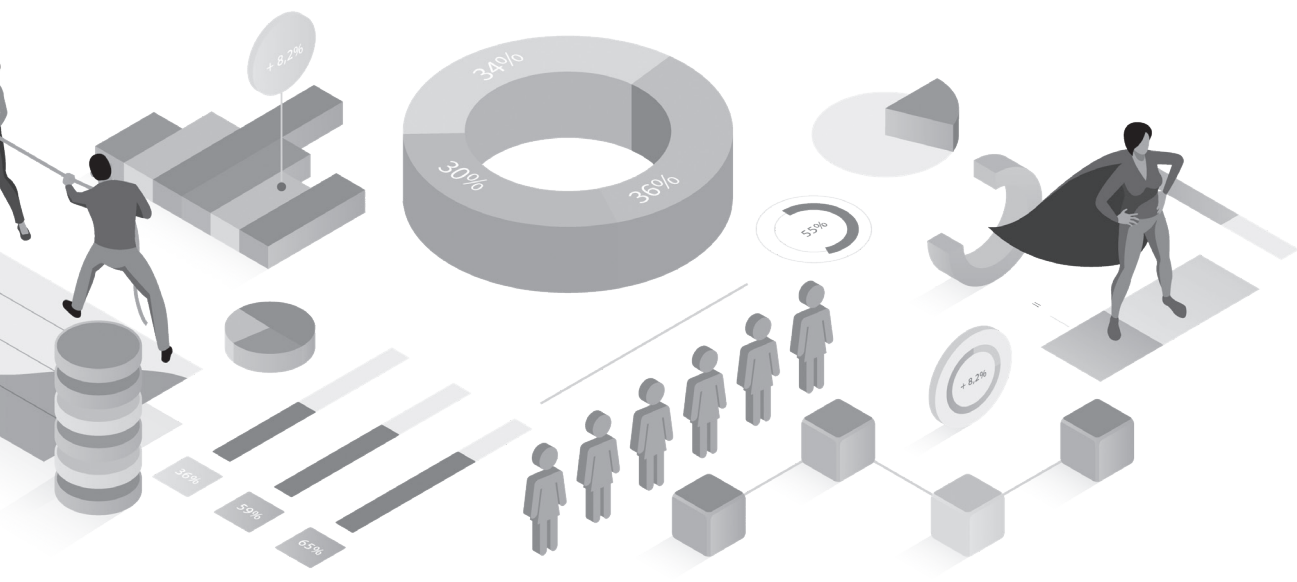
文 | 編輯室

不論是在數十萬年前的篝火旁，或是在今日時髦的咖啡店裡，我們都對「說故事」充滿熱情。伴隨科技的長足進展，人們如今不僅擁有文字、影音與圖表等豐富多元的敘事媒介，亦因身處數位時代而能輕鬆獲得可供敘事的大量資料。問題在於，我們該如何善用資源，從而道出令人著迷的故事？為回答這一重要問題，迄今已有許多人文及社

會科學研究者付出心力。除了嘗試運用數位科技外，他們也不斷思索，學人的研究視野在數據與新工具的刺激之下能如何擴展與轉化。

在不勝枚舉的數位研究工具中，Tableau 是有助於使用者視覺化資料的新興平臺，共有多種產品。其中，Tableau Desktop 的功能在於可直接透

1 本文經講者檢閱修訂。

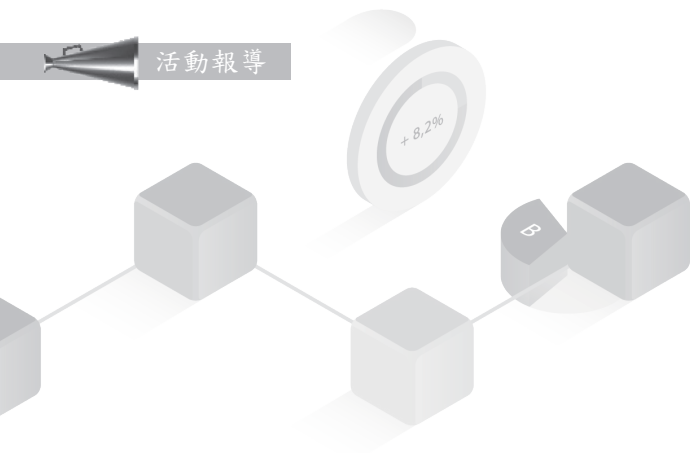


過拖拉進行資料清理，而不必寫程式，包含：讀取資料（幾乎所有格式皆可讀取，甚至可連結資料庫，讀取 PDF 表格及地理資訊等）、清理資料（提供介面預視每個變數的概況，包含數值型態與變項型態）、操作資料（包含轉置、合併、聯集等操作多個表格的功能）。不止如此，Tableau 亦可透過拖拉來進行視覺化，便於使用者切換各種視覺化方式，進行探索性分析、敘事和發掘新視角。

Tableau Prep Builder 則有助於更快地處理複雜且多來源的資料、按照所需欄位彙總數值、按照所需欄列位置轉置資料、將同類型資料併集在一起（例如將分散在不同檔案的多年資料整合起來）、將不同類型資料以共同項目聯集起來。

2023 年 1 月 6 日，臺灣大學人口與性別研究中心特別邀請高雄醫學大學性別研究所余貞誼助理教授主講「性別統計資料庫的探索、運用與視覺化」工作坊。本場次工作坊分為兩部分，余貞誼在 Part I 講論數據的後設思考與視覺化運用，並在 Part II 以行政院建置啟用的「重要性別統計資料庫」為例，帶領與會者實地操作和應用 Tableau 進行數據的視覺化。由於 Part II 內容所涉及的技術問題較為繁複而難敘明，因此，本篇紀錄以 Part I 內容為主，並輔以余貞誼歷來的相關著述，重在呈現對數據工作的批判性理解上。

首先，余貞誼將統計資料的視覺化過程概分為五階段：第一，獲取原始資料（Raw Data），即未經處理和分析的數據。第二，清整資料、處理缺漏值、



摘要彙算為視覺化所需的變數型態。第三，用各種視覺化方法探索資料中有價值的切入點。第四，在選定的圖表上加入標籤、標記來凸顯敘事的重點。第五，以文字和多媒體進行敘事。

不論是處於資料視覺化歷程的哪一階段，皆須保持對權力的批判性警覺。相比於文本修辭，數字和視覺化較

具權威性，較容易被接受為一種證據，理由在於文本很容易形成反面論述，但圖像和數字卻難以輕易推翻（須先取得特定工具方能進行反證），因此，科學溝通中使用數字和圖像較容易被人接受，也較不易遭質疑其解釋的效力（Rieder & Röhle, 2012, p. 74，引自余貞誼，2020，頁 169）。儘管數據經常被視為客觀中性的存在，但余貞誼提醒我們在獲取數據的同時，也需要看見資料中的權力，亦即對數據進行後設思考，例如，我們拿得到什麼資料？這些資料生產的邏輯是什麼？是否已經蘊含特權視角的偏見？

常見的問題在於，原始資料蘊含了某些性別刻板印象，因而限縮了研究

15-64歲女性平均每日無酬照顧工作時間

中華民國108年6月底

單位：分鐘

	總計	照顧未滿12歲兒童	照顧65歲以上家人	照顧12-64歲家人	做家務	志工服務
總計	190.9	62.2	14.1	6.0	103.9	4.6
按年齡分						
15-24歲	58.7	9.0	2.7	1.3	42.3	3.4
25-34歲	158.6	83.5	3.2	3.1	67.7	1.1
35-44歲	252.1	130.3	8.5	6.2	104.7	2.3
45-54歲	198.5	26.1	23.8	9.7	132.8	6.1
55-64歲	243.7	44.7	28.9	8.2	152.2	9.7

註：1. 資料來源：108年15-64歲婦女生活狀況調查。

2. 本表資料為每位婦女之平均每日無酬照顧時間，包含完全未從事者。

3. 「未滿12歲兒童」包含子女以外之其他兒童，如孫子女或其他親屬。

4. 「65歲以上家人」為日常生活起居活動需要協助之65歲以上家人。

5. 「12-64歲家人」為日常生活起居活動需要協助之12-64歲家人。

表 1：15-64 歲女性平均每日無酬照顧工作時間（衛生福利部，2023a）。

本會農村婦女教育訓練成果(農會家政班)				
縣市別	109年度			
	總計 (人)	男性 (人)	女性 (人)	新住民人數 (人)
新北市	2,180	20	2,160	15
台北市	1,105	8	1,097	0
台中市	12,121	267	11,854	94
台南市	6,858	286	6,572	13
高雄市	4,056	73	3,983	35
宜蘭縣	3,002	47	2,955	5
桃園市	5,763	69	5,694	11
新竹縣	3,362	156	3,206	5
苗栗縣	6,016	258	5,758	19
彰化縣	7,530	88	7,443	41
南投縣	3,243	93	3,149	48
雲林縣	3,977	150	3,827	19
嘉義縣	7,200	116	7,084	10
屏東縣	3,722	178	3,544	11
台東縣	1,173	119	1,054	18
花蓮縣	2,316	6	2,310	30
澎湖縣	194	11	183	1
基隆市	236	0	236	2
新竹市	323	15	308	0
嘉義市	180	3	177	0
合計	74,557	1,963	72,594	377

表 2：109 年度農委會農村婦女教育訓練成果（農會家政班）（行政院農業委員會，2023）。

本會農村婦女教育訓練成果(農會家政班)					
縣市別	110年度				單位：人
	合計	男性	新住民	女性	新住民
宜蘭縣	3,149	55	-	3,094	5
新竹縣	3,217	163	-	3,054	7
苗栗縣	5,953	240	2	5,713	25
彰化縣	8,387	116	-	8,271	45
南投縣	4,133	132	-	4,001	52
雲林縣	4,913	194	-	4,719	20
嘉義縣	7,806	144	-	7,662	12
屏東縣	4,036	141	-	3,895	18
臺東縣	1,902	167	-	1,735	18
花蓮縣	2,698	18	-	2,680	34
澎湖縣	206	21	-	185	-
基隆市	236	0	-	236	2
新竹市	323	15	-	308	-
嘉義市	180	3	-	177	-
合計	79,662	2,133	2	77,529	406

表 3：110 年度農委會農村婦女教育訓練成果（農會家政班）（行政院農業委員會，2023）。

者伸展研究觸角的可能性。例如，在無酬照顧工作時間的統計上，資料庫有「15-64 歲女性平均每日無酬照顧工作時間」表（見表 1），但卻缺少同年齡層的男性資料，使得我們難以探見男性擔負無酬照顧工作的概況。此外，該表也缺少按教育程度分群的資料，以致使用者無法藉此了解照顧工作者的年齡與教育程度間有何關聯，以及不同類型的無酬照顧工作與照顧工作者的教育程度間有何關聯。

又如，資料收集經常涉及分類的向度，而這個分類本身也經常透露出我

們對於何種類型是重要的、或值得統計的想像。如「農委會農村婦女教育訓練成果」表中，在中華民國 101 至 109 年內的分類中只有「男性」、「女性」、「新住民」這三個平行的類別，以致使用者無法看出新住民中有多少男性和女性接受農會家政班的教育訓練（見表 2）。直至民國 110 年，該表方把新住民的性別放入統計欄位，在「男性」與「女性」項下分別設置「新住民」的附屬類別（見表 3）。

再以衛福部「歷年新生兒、嬰兒及孕產婦死亡概況」表為例，該表未按





表32 歷年新生兒、嬰兒及孕產婦死亡概況

年 別	新 生 兒				嬰 兒				孕 產 婦	
	計 (人)	男 (人)	女 (人)	死亡率 (每千活產)	計 (人)	男 (人)	女 (人)	死亡率 (每千活產)	死亡數 (人)	死亡率 (每千活產)
民國85年	1,129	621	508	3.5	2,169	1,183	986	6.7	25	7.7
民國86年	1,064	578	486	3.3	2,071	1,123	948	6.4	30	9.2
民國87年	918	546	372	3.4	1,784	1,015	769	6.6	24	8.9
民國88年	980	554	426	3.4	1,721	977	744	6.1	24	8.4
民國89年	1,038	575	463	3.4	1,789	992	797	5.8	24	7.8
民國90年	865	485	380	3.4	1,559	863	696	6.0	18	7.0
民國91年	745	415	330	3.0	1,325	721	604	5.4	19	7.7
民國92年	624	324	300	2.7	1,105	574	531	4.9	15	6.6
民國93年	623	358	265	2.9	1,146	645	501	5.3	12	5.5
民國94年	605	343	262	2.9	1,026	563	463	5.0	15	7.3
民國95年	554	307	247	2.7	943	524	419	4.6	15	7.3
民國96年	588	342	246	2.9	959	547	412	4.7	14	6.9
民國97年	538	291	247	2.7	897	496	401	4.6	13	6.6
民國98年	452	233	219	2.4	778	417	361	4.0	16	8.3
民國99年	429	246	183	2.6	705	403	302	4.2	7	4.2
民國100年	530	293	237	2.7	832	452	380	4.2	10	5.0

表 4：歷年新生兒、嬰兒及孕產婦死亡概況（局部）（衛生福利部，2023b）。

縣市及族群分群，也缺少詳細的死因分類（見表 4）。如果表格能夠呈現出上述資訊，那麼，便有機會了解孕產婦的死因是否與其族群或縣市所擁有的資源相關，因而思索哪些孕產婦的死亡狀況是可避免的。

不只原始資料的產製過程與權力密切相關，在獲得原始資料後必須進行的數據清理工作亦不例外。危險的是，一旦數據在清理過程中被移除於其所在的脈絡之外，便容易讓統治群體以自己的認識觀點來凌駕於他人之上而造成危害（D'Ignazio & Klein, 2020，引自余貞誼，2021，頁 236）。

基於上述可知，當數據發布者決定哪些數據要納入收集範疇時，便已

形成可供偏見與價值關聯滲入的空間，而這些皆可能影響資料代表性、測量信效度和資料品質等問題（Parks, 2014，引自余貞誼，2020，頁 168）。「數據會自己說話」的主張，忽略了數據在收集時都已經過處理，是社會、政治和歷史環境下的產物。因此，已有學者提倡將數據與脈絡相連，考慮數據產製的歷程，從而討論數據在功能上何種限制、負有何種倫理義務，或是生產它所涉及的權力和特權會不會蒙蔽某些事實等等（D'Ignazio & Klein, 2020，引自余貞誼，2021，頁 237）。

值得注意的是，視覺化本身亦可能為權力和特權所利用。視覺化圖表多透過選擇性的壓縮或化約，以凸顯一個現象的某些層面。有時這是出於實用

理性的考量（例如出於螢幕尺寸或解析度需求而裁減資料面向），有時則是基於特定的操縱意圖（例如更改級距和顏色，以傳達對特定對象有利的結果）（Rieder & Röhle, 2012, pp. 73-74; Wainer, 2009，引自余貞誼，2020，頁 169）。因此，有學者指出視覺化圖表具有認知、情感、社會等三個層面的風險。認知是指不適切地選擇圖表要素（例如要將哪些要素放進圖表？是否過度簡化或複雜化？），情感涉及圖表設計如何連結到反感的情緒，社會面則是指使用者的跨文化差異會影響對圖表的解讀（例如東亞國家對圖表方向的解讀、顏色的意義認知，就不同於西方國家）（Bresciani & Eppler, 2008，引自余貞誼，2020，頁 169-170）。

在談及視覺化圖表的三種風險後，余貞誼以數個實例講述如何以視覺化形式呈現數據並進行探索。廣為人知的例子是 Charles Joseph Minard (1781-1870) 的「1812-1813 年對俄戰爭中，法軍人力持續損失示意圖」（見圖 1），該圖以兩個維度同時呈現六種數據，包含拿破崙（Napoleon Bonaparte, 1769-1821）軍隊的人數、經緯度、時—地關係、移動方向、溫度。圖中褐色長塊表示法國向莫斯科推進的大軍人數，黑色則表示從莫斯科撤退的法軍人數。

值得一提的是，圖 1 以簡要的方式呈現法俄戰爭中著名而慘烈的別列津納河戰役（Battle of Berezina）：撤退的法軍在渡河前尚有五萬人，但渡河

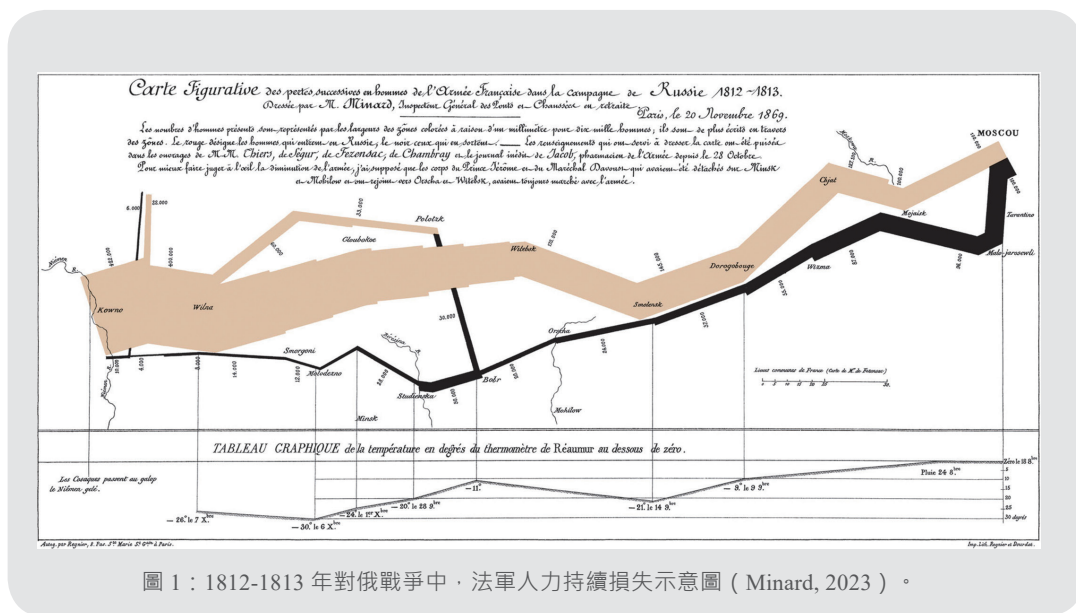


圖 1：1812-1813 年對俄戰爭中，法軍人力持續損失示意圖（Minard, 2023）。

後卻僅剩 28,000 人。儘管圖 1 的讀者可能不諳法俄戰爭史，但只要透過這幅同時將六種數據包含在內的圖表，不難想見法軍搶渡過程中的大量死亡或與當時的溫度有關。圖 1 雖因其成功地以簡馭繁而被譽為最優秀的統計圖表之一，但它可能也過度簡化戰爭的血腥等面向，這是讀者需要加以留意之處。

數據視覺化的其他傑出範例，可見於《紐約時報》（*The New York Times*）的〈這張圖裡頭正在發生什麼事？〉（What's Going On in This Graph?）系列文章（*The New York Times*, n.d.）。余貞誼從中挑選一例加以說明：〈婦女生孩子的年齡：差距是如何分化美國的〉（*The Age That Women Have Babies: How a Gap Divides America*）一文，分析近 40 年來的美國孕產情況，指出婦女成為母親的年齡因其地理區位與教育程度而有顯著差異。文中 2016 年有關新手媽媽（first-time mothers）年齡分群的圖表呈現雙峰分布，可看出「無大學學位」與「有大學學位」兩個次群體（Bui & Miller, 2018）。

除了上述的直條圖外，原始數據還可供製作出包含新手媽媽所在地理區位與教育程度等資訊的圖表（Bui & Miller, 2018），有助於使用者進行延伸思考，例如探討美國婦女的首次孕產及其所在地區與教育程度有何關聯。



余貞誼（高雄醫學大學性別研究所）

在結束對圖表範例的解說後，余貞誼也根據《紐約時報》籲請讀者在進行數據視覺化時應思考四個問題，邀請大家在面對一張圖表時可以往此探詢：你注意到什麼？你想知道什麼？這跟你和你的社群有什麼關係？這張圖裡頭正在發生什麼事？且應設法創造一個吸引人的標題來捉住圖的核心想法。

余貞誼總結道，數據分析具有兩種重要的功能：第一，探勘資料，理解事件樣態，並找出故事中值得挖掘的現象；第二，視覺化呈現，輔助敘事，並凸顯所欲構框的重點。而 Tableau 因其免寫程式即可處理複雜資料、進行資料聯集，且能以視覺化形式呈現統計數據的特點，有益於學人更快速、不費力地完成前述的數據分析工作。然而，在利用數位工具多方嘗試的同時，亦應時時保持警覺：若不去意識大數據研究中權力與特權的存在和運作，那麼，數據工作將無法看顧不同立場的需求與觀點，無助於讓遭受邊緣化及消音的主體重回舞臺（余貞誼，2021）。

參考文獻

行政院農業委員會（2023 年 3 月 21 日）：〈農委會農村婦女教育訓練成果〉。重要性別統計資料庫。https://www.gender ey.gov.tw/gecdb/Stat_Statistics_DetailData.aspx?sn=SG70KcMSWlugiJvSVZD7Xw%40%40&d=m9ww9odNZAz2Rc5Ooj%24wIQ%40%40

余貞誼（2020）：〈從經驗中對話：初探計算社會科學的立論基礎與應用挑戰〉。《調查研究 - 方法與應用》，45，155-196。

余貞誼（2021）：〈數據始終來自於權力——評 Catherine D'Ignazio and Lauren F. Klein (2020). Data Feminism. Cambridge, MA: The MIT Press.〉。《科技、醫療與社會》，32，233-239。https://doi.org/10.6464/TJSSTM.202104_(32).0005

衛生福利部（2023 年 5 月 31 日 a）：〈15-64 歲婦女平均每日無酬照顧時間（分鐘）.pdf〉。重要性別統計資料庫。https://www.gender ey.gov.tw/gecdb/Stat_Statistics_DetailData.aspx?sn=cd25txZGJ35JhMi30z7uEw%40%40&d=194q2o4!otzoYO!8OAMYew%40%40

衛生福利部（2023 年 5 月 31 日 b）：〈3.1-02 嬰兒與孕產婦死亡數及死亡率歷年死亡概況 1110826.pdf〉。重要性別統計資料庫。https://www.gender ey.gov.tw/gecdb/Stat_Statistics_DetailData.aspx?sn=%244h7q0oY5fyYspcAyybs1g%40%40

Bui, Quoc Trung & Miller, Claire Cain (2018, August 4). *The Age That Women Have Babies: How a Gap Divides America*. The New York Times. https://www.nytimes.com/interactive/2018/08/04/upshot/up-birth-age-gap.html

Minard, Charles Joseph (2023, June 22). *Charles Minard's 1869 chart showing the number of men in Napoleon's 1812 Russian campaign army, their movements, as well as the temperature they encountered on the return path. Lithograph, 62 x 30 cm*. Wikipedia. https://en.wikipedia.org/wiki/File:Minard.png

The New York Times. (n.d.). *What's Going On in This Graph? Graphs, maps and charts from The Times — and an invitation to students to discuss them live*. The New York Times. https://www.nytimes.com/column/whats-going-on-in-this-graph

