觀察數據世界的二種視角：探索性vs視覺化

江信昱

世新大學資訊傳播學系助理教授

壹、前言

自2008年9月《Nature》出版大數據特刊[1]、2011年2月《Science》出版大數據處理專題[2]後已超過十年，從政府組織、科學研究、市場應用等皆已發生變化，例如：2012年美國白宮公布《大數據研發先導計畫》（Big Data Research and Development Initiative）[3]屬於國家層次的重要政策、科學研究出現第四典範轉移，即轉向「資料密集型的科學發現」（Data-Intensive Scientific Discovery）（Hey, Tansley & Tolle, 2009），或社會科學也衍生出「計算社會科學」（Computational Social Science）（Lazer, Pentland, Adamic, et al., 2009）等新方向；在產業端，企業紛紛提出大數據平臺或解決方案（如BigQuery、InfoSphere BigInsights、Azure HDInsight等），串聯合作間也逐漸形塑出獨特的大數據產業結構。在在顯示，大數據被多個領域視為下一代資訊技術與資料分析管理的熱點，影響著人們思維與生活、企業經營與銷售、國家決策與治理、全球經濟與產業發展等層面，儼然已進入「大數據時代」（Manyika, Chui, Brown, et al., 2011）。

1 網址：https://www.nature.com/nature/volumes/455/issues/7209

2 網址：https://science.sciencemag.org/content/331/6018

3 網址：https://obamawhitehouse.archives.gov/sites/default/files/microsites/ostp/big_data_press_release_final_2.pdf

在新的資料環境中，龐大的資料量潛藏著無法用人腦立即探知的未知關係與結果（李廣建與江信昱，2018）。若能善用分析與視覺化工具，深度提取海量資料中的有用資訊並形成結論，加速其決策效率，將有助於組織單位在經營管理或行銷推廣。對典藏各類資料資源的圖書館而言，善用資料視覺化工具挖掘與展示結果，應可協助圖書館單位或從業人員思索資料加值的新形式或新服務。本文引述英國資料記者David McCandless的說法：在資料充斥的現在環境，透過資料視覺化能夠幫助人們去探索資料，尤其當人們在資料大海裡迷失方向時，視覺化的引導將起強大作用。此一說法，恰好替本講題內容引出作者看待此事的觀點。

"By visualizing information, we turn it into a landscape that you can explore with your eyes. When you're lost in information, an information map is kind of useful."

為此，本文先以圖書館館員與讀者眼中的數據作為引言，闡述二者之間關聯，再梳理資料視覺化的發展脈絡與相關工具，輔以國內外圖資領域之因應做法與實例，解釋二種觀察數據世界的視角所帶來的差異、價值與困難點，期望為圖書館服務提供一些新的啟發。

貳、圖書館館員與讀者眼中的數據

圖書館的數據在哪裡？在館員與讀者的眼中，並非全然一致。首先，從圖書館館員眼中所看到的各式數據，包括借還書數據、資料庫使用統計、入館人數、網頁點擊率及館藏統計等，都與日常館務有極大關聯。但是當轉換到讀者角度時，讀者主要關心圖書館有多少空間可供借用，以及現在熱門圖書為何？簡言之，館員及讀者是基於不同目的看待資料這件事，前者基於館藏服務與圖書館營運目的，後者則基於資源取用目的。姑且不論目的各異之情況，在二者眼中皆是顯而易見的館務統計結果。本文進一步思考是否還存在一些鮮少關注但對圖書館經營有利的其他數據，例如：成大Dcard社群評論中的文本數據，或成大圖書館官方Twitter貼文被名人或圖書館界先進分享連結的關係數據；另外，讀者透過圖書館留言板反映想法或意見，可視為讀者反饋的文本數據，機構典藏中也隱藏著學者社群的關係數據。倘若館務數據能再進一步結合上述社群、文本及關係等數據，應有利於圖書館長久經營。

延續此一數據資源擴充想法，本文認為體現出二點變化：

一、數據異質（Variety）：是大數據4V的一項特性（McAfee & Brynjolfsson, 2012），其本質是在探討利用多源數據反映事物發展的全部面貌。由於同一事實或規律可以同時隱藏在不同的數據形式中，但取用與分析館藏服務、讀者行為等館務數據僅能反映圖書館經營的一個面向，若結合社群、文本、關係等數據則可再揭示圖書館經營的其他面向，用以描繪圖書館經營的全貌。

二、量化一切：「讓數據發聲」是大數據時代的重要特性（Debackere, 2016），當圖書館蒐整的數據資源越多，所構成的事物景象及其關係會越詳細，利於決策者或讀者從不同維度觀察與分析數據世界。正因數據讓一切加速發展，連帶影響人們的分析思維：第一，強調全體數據，大數據容許忽略微觀層面的精準度，以換取更多數據來反映整體服務及讀者樣態；第二，追求速度優勢，效率優先於精準，藉由機器處理並快速發現洞見成為未來管理趨勢；第三，重視相關分析，過去追求的分析是建立在有限的抽樣範圍且利於人腦推論其因果關係，但當數據量大到一定程度，其關係就容易被複雜結構所掩蓋，再加上現實發展多屬於複雜情況，人為直觀推論的因果關係也可能會忽略一些潛在因素，因此須先找出潛藏在各類數據間之相關性，再判斷其因果成為大數據時代的思維轉變之一（Mayer-Schönberger & Cukier, 2013；Hubbard, 2014）。

因應大數據趨勢，圖書館既持續累積各類館務相關數據，也紛紛進行資料視覺化展現，其目的除了希望提升整體服務外，更為了爭取經費，因此需要客觀數據來管理圖書資源（羅思嘉, 2010；Finch & Flenner, 2017；Yadav & Ahmed, 2020）；就推廣服務而言，資料視覺化亦可針對校內科學社群提供學術研究決策及資料上的互動服務（Jiang & Carter, 2018；Ogier & Stamper, 2018），例如：清華大學圖書館透過挖掘資料，連結教師著作與聯合國17個永續發展目標[4]。本文認為資料服務將是未來圖書館非常重要的業務，站在服務讀者的角度，資料視覺化除能促進資源與資訊的推廣，更能達成視覺溝通。

參、資料視覺化進展

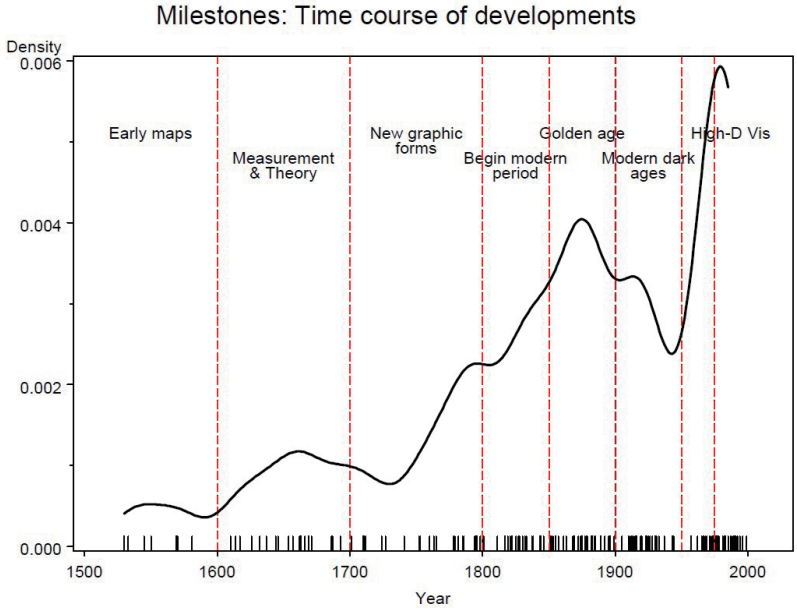
什麼是資料視覺化（Data Visualization）？各個領域都在談資料視覺化，包括工程領

4 網址：https://sdgs.nthu.edu.tw/

域、電腦科學，甚至圖書資訊科學都有論述。圖書資訊領域強調透過資料分析可輔助事物的理解以幫助決策（Börner, Chen, & Boyack, 2003；Phetteplace, 2012；Chen, 2017），電腦科學領域中的資料視覺化是從物理層次探討資料物件之間的關係（Chen, Unwin, & Härdle, 2008；Bikakis, 2019）。不論圖書資訊或電腦科學領域皆有一共通核心思想，即資料能幫助理解事物發展，更有助於以客觀角度進行後續決策。

一、資料視覺化發展史

資料視覺化最早可追溯至十五、十六世紀，該時期的圖像或地圖多採取符號形式的表徵，主要是以量測方式描繪空間概念，到了十七世紀出現新的圖像形式（如等值線和等高線）並嘗試應用在地質、經濟等主題，開始將資料視覺化帶往高峰發展，一路至十八世紀逐漸達到第一波高峰（圖一），其中1850年到1900年期間是整個資料視覺化發展史的黃金時代，此時期著重美觀，而忽略資料能反映真實的本質。1900～1950年逐漸下降至峰底，係因二次世界戰爭爆發，對資料視覺化的應用需求較少，直到1950年之後相關工具開始發展，整體資料環境不斷建設， 2000年之後再次攀向另一波發展高峰（Friendly, 2005）。



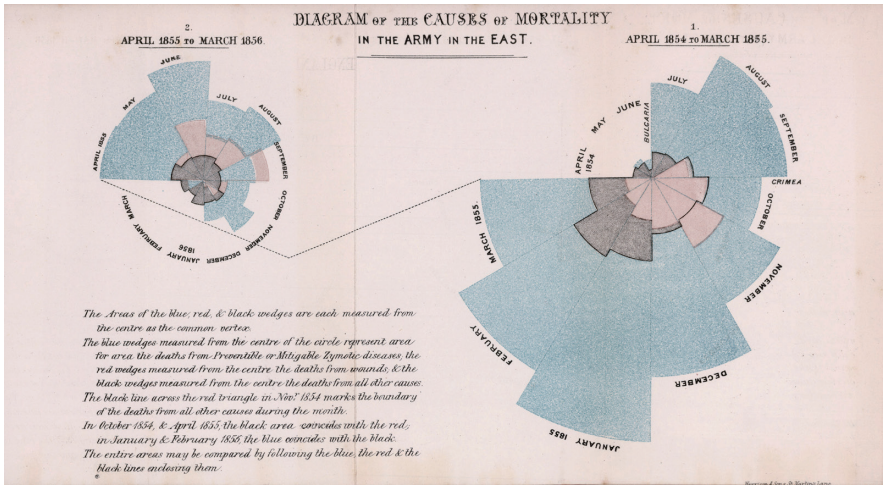
圖一　資料視覺化發展史

資料來源：Friendly, M. (2005). Milestones in the History of Data Visualization: A Case Study in Statistical Historiography. In: Weihs C., Gaul W. (eds) Classification – the Ubiquitous Challenge. Studies in Classification, Data Analysis, and Knowledge Organization. Springer, Berlin.

二、1850～1900年：統計圖的黃金年代

㈠ 南丁格爾玫瑰圖（1856年製圖）

1850年到1900年這段黃金年代有兩張代表圖示，也是資料視覺化研究的經典案例。第一張是南丁格爾的玫瑰圖（圖二），南丁格爾是位護士，也是數學家，更是統計學會的第一位女性會員。此圖右側（1854～1855年）及左側（1855～1856年）分別以視覺化呈現二個時間區間在戰場上的死亡人數差異，扇形面積大小反映每月死亡人數多寡，另輔以三種顏色分別表達感染、槍傷與其他死亡原因。圖二可見左側之藍色面積大幅縮小，說明因感染而死亡的士兵人數有明顯下降，其原因可歸功於醫院的衛生條件改善。進一步補充，當時南丁格爾製圖的用意是想表達戰爭期間的傷員死亡與衛生環境的關係，透過視覺化統計結果，避免繁雜表格閱讀，更能讓決策者快速掌握傷兵現況。

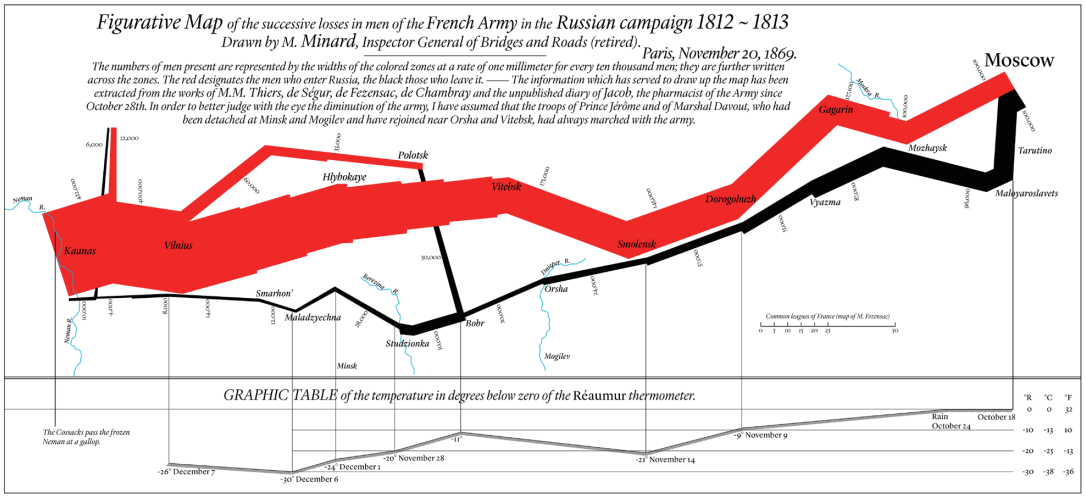


圖二　南丁格爾玫瑰圖
資料來源：https://cwfn.uoguelph.ca/

㈡ 拿破崙遠征莫斯科圖（1869年製圖）

拿破崙遠征圖（圖三）的訊息量更多於南丁格爾玫瑰圖，此圖記錄1812年法國軍隊東征莫斯科往返的實際路徑。其中，紅色象徵軍隊往莫斯科時的「出征」，黑色象徵莫斯科折返法國的「回程」，而紅色到黑色之寬度，則代表軍隊人數的多寡。從色塊的寬細變化隱喻軍隊人數往返的銳減程度，同時搭配圖中的城市地名與圖下方的地點溫度，既呈現法

國到莫斯科的距離感，也能看出遠征過程中氣溫變化對軍隊的影響。在法軍撤退的10～12月，溫度從列式溫度0°一路下降到-30°，法軍遭受著氣溫寒冽、長途跋涉及傷兵負累，致使軍隊回程人數一路下滑。整體而言，此圖包含了時間、空間、方向、人數等符號，運用紅色的激昂感與黑色的挫敗感隱喻軍隊往返狀態，直觀地呈現軍力由盛轉衰的動態變化。



圖三　拿破崙遠征莫斯科圖
資料來源：https://edspace.american.edu/visualwar/minard/

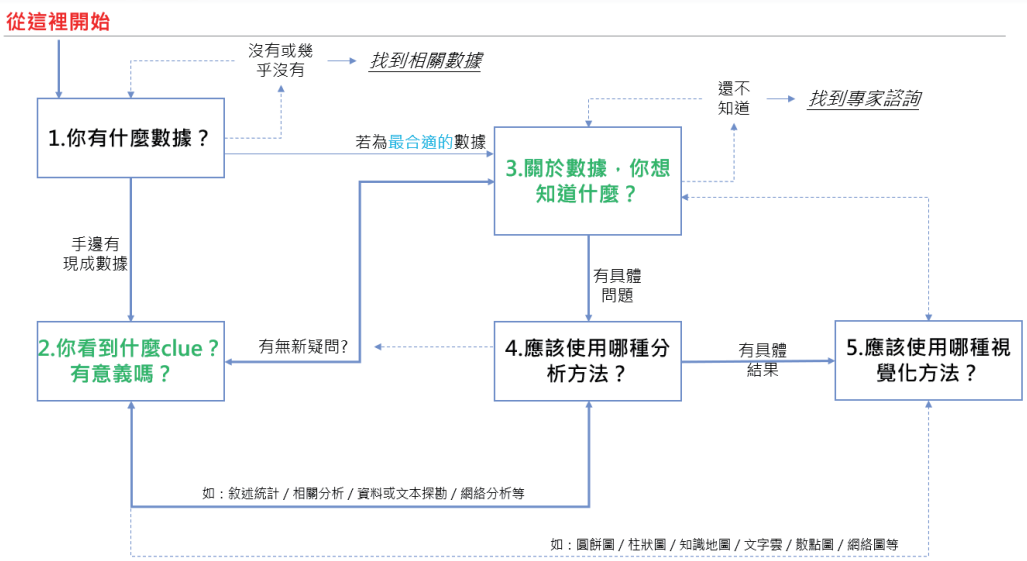
以上兩張圖示未將資料導入相關工具進行分析，皆以人工方式處理資料，再以視覺化方式繪製當中的符號涵意或洞見結果。然而，伴隨相關工具發展至今日，會看到近期資料視覺化相關作品饒富趣味，例如：報導者在2018年進行縣市長選舉報導[5]時，蒐集社群媒體上對各縣市候選人的討論文本數據及聲量進行視覺化展現；CNN也於2020年結合開票數據，即時呈現美國總統大選之各州選情進度[6]；另外，現因政府開放資料盛行，許多資料愛好者會應用這些開放資料，結合地理圖資替整體視覺性、指向性與互動性提供查詢服務，如用數據看臺南的「臺南交通事故都在哪」[7]，或是新冠疫情期間的口罩供需地圖或本土案例確診足跡地圖等，這些都是資料視覺化的另一應用場景。

5　網址：https://www.twreporter.org/i/2018-election-report-sharevoice-gcs
6　網址：https://edition.cnn.com/election/2020/results/president
7　網址：https://www.taiwanstat.com/vp-tainan/traffic_report.html

三、資料視覺化工作流（Data Visualization Pipeline）

要完成一個資料視覺化作品，通常會採行一工作流程來逐步完成資料分析與視覺化。例如：微軟研究院劉世霞等人（Liu et al., 2014）調查資訊視覺化研究進展時歸納出一套視覺化處理流程，首先資料要經過轉換與清洗，再透過篩選找到符合目的之資料結果，經由資料映射（Mapping）連接資料屬性後，進行資料表格或統計結果的描繪或渲染（Rendering），最後透過使用者介面進行各種互動；美國伍斯特理工學院 Mattew D. Ward等人（2010）則從資料分析觀點來看視覺化流程，首先將原始資料轉換成機器可讀取的表格，再把當中的特徵屬性進行視覺化映射（Visual Mapping）與呈現，從獲取資料到最後產製結果都是清楚的資料處理過程。不難看出，各學者對資料視覺化流程雖有細微差異，但整體仍是可看作數據經一系列處理階段並得到轉化的過程。

於是，將上述流程結合至資料新聞學之實務教學時，觀察到傳播學院的同學很難理解此流程。原因在於他們很少接觸資料，不像資工或資管領域同學每天與資料打交道，因此須配合傳播學院的學習情境重新修改資料視覺化流程，且是能實際做操作的內容（圖四）。在執行過程中，先問自己有什麼樣的資料，再問資料中的線索是有意義的嗎？接著才會詢問：「關於這些數據，想知道什麼？」。當有了對問題的初步想像後，再決定要用何種視覺化方法與工具。



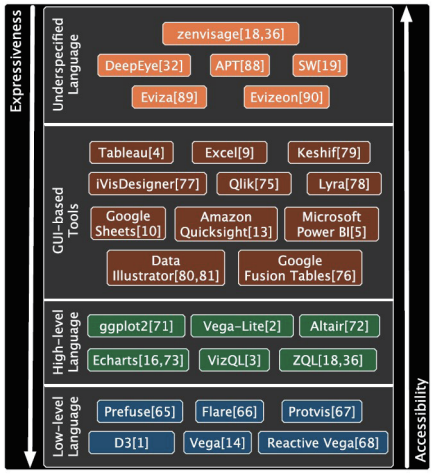
圖四　課堂實務操作之資料視覺化流程

另一方面，在課程教學現場，考量傳播學院的學生情境及須向閱聽眾傳遞正確的圖表與詮釋內容，會搭配「圖表選擇矩陣」[8]，協助新手學生在繪製各類圖表時，可依據想表達的意涵（如揭示資料間之關係、比較不同資料、呈現資料分佈等）選擇合適的呈現圖示，以避免用錯圖表或錯誤理解圖表意義。

四、資料視覺化常用工具及呈現原則

㈠ 常用工具

一般而言，資訊管理領域人員會透過程式撰寫展現資料視覺化結果，如D3.js[9]，但對於沒有程式背景的學生或館員，建議可善用現成工具或軟體，包括Tableau Public[10]、Google Data Studio[11]、RAWGraphs[12]、國內自行研發的PlotDB[13]以及歐洲新聞中心開發的Flourish[14]，多數工具隸屬於圖五的圖像使用者介面（Graphical User Interface, 簡稱GUI）工具區塊。



圖五　資料視覺化常見工具或軟體

資料來源：Qin, X., Luo, Y., Tang, N. et al. (2020). Making data visualization more efficient and effective: a survey. The VLDB Journal, 29, 93-117.

8　網址：https://thelead.io/chart-selection-diagram-data-visualization
9　網址：https://d3js.org/
10　網址：https://public.tableau.com/zh-tw/s/
11　網址：https://datastudio.google.com/
12　網址：https://rawgraphs.io/
13　網址：https://plotdb.com/
14　網址：https://flourish.studio/

㈡ 呈現原則

即便我們已掌握資料與工具，也知道使用什麼圖表，但在呈現結果時，往往會把非常複雜的訊息同時擺在一張圖中，對讀者來說，既難以理解，也較難有效傳遞訊息。為解決此問題，本文認為新手須先掌握三項基本原則：

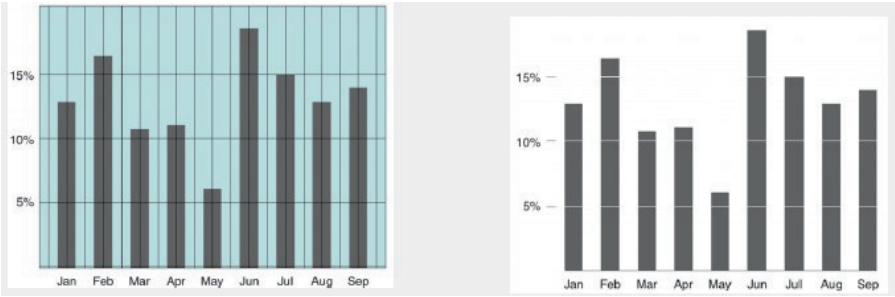
1. 正確性：一個好的資料視覺化，須優先保證資料結果的正確性，並注意資料混用情況，讀者才能正確解讀結果。
2. 單一性：人的記憶有限且存在遺忘曲線（Forgetting Curve），為使資料結果能深刻記憶於人腦，一張圖表只講一件事，以區隔主要訊息與次要訊息之層次，避免過多雜訊干擾，讀者才能清楚接收圖表的主要訊息訴求。
3. 舒適性：人是視覺動物，當訊息以圖表的方式呈現時，與大段文字相比，更可以快速了解其含義。因此透過符號隱喻（如箭頭表示方向、柱狀表示長度、圈叉表示對錯等）、編排技巧（如字體大小、行距高度、留白等）或色彩策略（如選用低彩度色系、避免大面積使用原色、善用對比色或相似色等）等方式視覺化分析結果，可以使人的眼睛能舒適地閱讀圖表。

五、由訊息派與視覺派的爭論，引出本文探索數據世界的二種視角

資料視覺化從1500年至今已有五百多年歷史，發展過程中衍生出兩個派系，一個是視覺派，圖表會做得非常花俏但人性化，讓讀者一眼就能記住；另一派是訊息派，主張刪去冗餘的裝飾用元素或符號，僅強調圖表裡的關鍵訊息。以下簡介兩個派系的代表人物與主張：

㈠ 訊息派訴求

代表人物是耶魯大學統計學與電腦科學的Edward R. Tufte，Tufte（1983）認為呈現資料圖表時，應優先呈現訊息，其他如配色、裝飾，甚至文字，對圖表都是干擾，所以Tufte教授提出「資料墨水比」（Data-Ink Ratio）計算方式，來計算圖表中資訊量與使用油墨量的比值，比值越高表示作品能以最少量的圖像傳達最多資訊。圖六之左側的藍色區塊會干擾讀者，屬於低Data-Ink Ratio；圖六之右側圖示拿掉藍色區塊後，當中所隱藏的重點訊息就會浮現出來，屬於高Data-Ink Ratio。

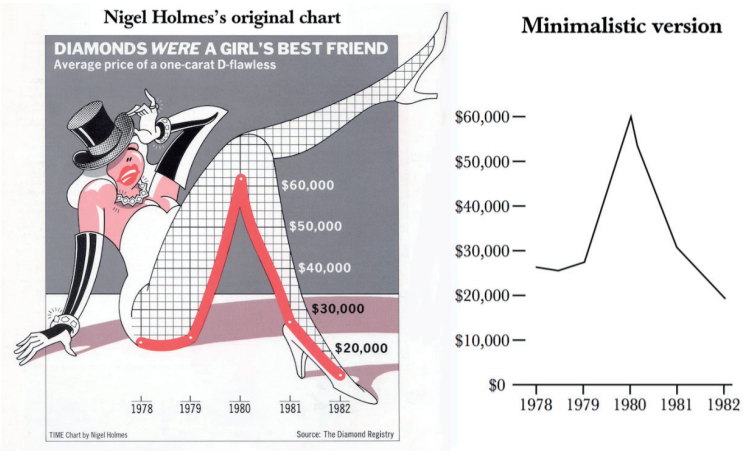


圖六　資料墨水比之高低差異示意

資料來源：Tufte, E. R. (1983). The visual display of quantitative information. Graphics Press, Cheshire, CN.

㈡ 視覺派訴求

代表人物是美國的圖形設計師Nigel Holmes，Holmes（2006）認為圖表應人性化，枯燥乏味的圖表是一種威脅，而一張有趣的圖表有助於讓讀者記住所要傳遞的訊息，也因此美觀的、有記憶點的圖表設計是此一派系的核心訴求。在1980年代，Holmes擔任《時代雜誌》的視覺總監期間，曾列舉一個視覺化案例（圖七），此圖的資料結果呈現了 1978 年至 1982 年的鑽石價格走勢。然而，比起看過圖七右側折線圖的人來說，圖七左側經裝飾過的圖表更讓人難忘，大家都會記得這位女性的腳部曲線剛好對應了 80 年代初期的鑽石價格，此圖當時在《時代雜誌》造成極大迴響。



圖七　鑽石價格趨勢圖

資料來源：Holmes, N., & Heller, S. (2006). Nigel Holes on Information Design. Jorge Pinto Books, New York.

整體而言，二個派系訴求對想要進行資料視覺化的人而言，不只是嚴謹實用與美觀有趣、乾淨與裝飾之爭，更多的是如何把握訊息展示的內容量，二者各有論點，至今雖仍無共識，卻也開始往彼此主張靠攏。反觀圖書館，即便有大量的讀者資料與豐富館藏資料，也尚未完全發揮這些資料的潛在價值。對圖書館館員而言，運用統計方法進行資料結果之視覺展現並不是難事，較難的是為什麼要做這件事？是基於何種目的進行資料視覺化。因此由上述二派爭論延伸本講題內容的思考，引出圖書館館員若要觀察數據世界並進行資料視覺化，館員應先釐清重點是完整呈現訊息訴求與分析洞見，或是為了透過分析成果向讀者溝通，前者傾向訊息派角度，後者傾向視覺派角度。若一開始沒有釐清資料視覺化的用意是「分析」（探索性）或是「溝通」（視覺化），在資料結果呈現與圖表設計上便容易產生資訊是否有效傳播之問題。以下，本文介紹數個案例，引導圖書館館員逐步思考不同性質的館務數據應以何種取向呈現資料結果。

肆、圖書館界之資料視覺化案例

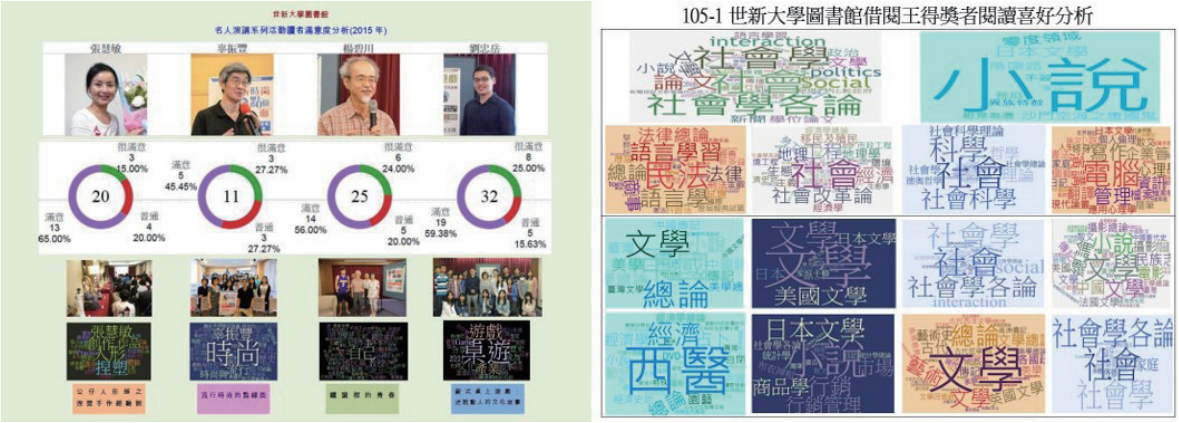
國內圖書館其實已存有許多資料視覺化成果，但較多出現於文獻報告或內部調查報告，例如：圖書館年鑑、委外調查報告或單位論文發表等。隨著資料庫技術與網際網路的普及，圖書館開始將各類館務數據或統計報表轉化為線上系統，既向讀者提供檢索服務，也快速展示資料基本統計圖表，例如：國家圖書館的「全國圖書館調查統計系統[15]」或國立公共資訊圖書館的「公共圖書館統計系統[16]」，把原存在於文獻報告裡的資料或報表以系統化方式呈現，透過系統的互動讓讀者了解圖書館服務的量能。以下為國內外圖書館界的各種形式資料視覺化案例：

一、世新大學圖書館的資料加值視覺化

世新大學圖書館投入諸多心力於活動成果視覺化，屬於資料加值服務，另外也整理借閱紀錄主題並以文字雲呈現（圖八），最大目的是希望能讓學生了解圖書館的借閱狀況，並藉由視覺傳播的方式慢慢貼近世新大學學生讀者群的喜好。

15 網址：http://libstat.ncl.edu.tw/statFront/index.jsp

16 網址：https://publibstat.nlpi.edu.tw/#tw



圖八　世新大學圖書館之各項資料視覺圖表
資料來源：http://lib.shu.edu.tw/datastory/

二、以電子看板即時呈現入館與借閱紀錄結果

上海圖書館於入口處放置一智慧牆（圖九），透過讀者借閱證的入館登記與借閱紀錄，即時呈現館內人數與館藏使用人次，另外這些分析結果也會定期刊登在圖書館報告，以一併考量讀者閱讀數位報告時的視覺感受（趙斌，2015），本文認為其主要目的不是為了分析也並非探索，而是為了與讀者溝通。

圖九　上海圖書館之智慧牆

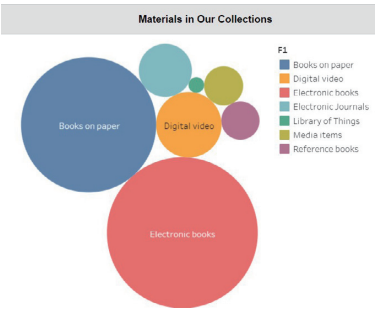
與上海圖書館類似的作法，還有美國西雅圖公共圖書館，2005～2014年該館執行「Making Visible the Invisible」計畫時也曾採取類似作法，在入口處放置6塊電子看板（圖

十），分別呈現館內人數、借閱數據，甚至運用杜威分類法串聯每本書的年代、標題、作者及關鍵字等。前述兩館皆擅長運用大螢幕在館內入口處「即時告知」館內現況。

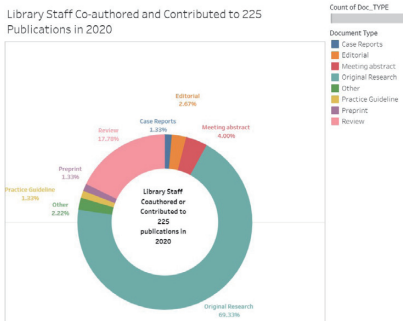
圖十　西雅圖公共圖書館之電子看板
資料來源：https://www.mat.ucsb.edu/~g.legrady/glWeb/Projects/spl/spl.html

三、以資料儀表板（Dashboard）展現資料結果

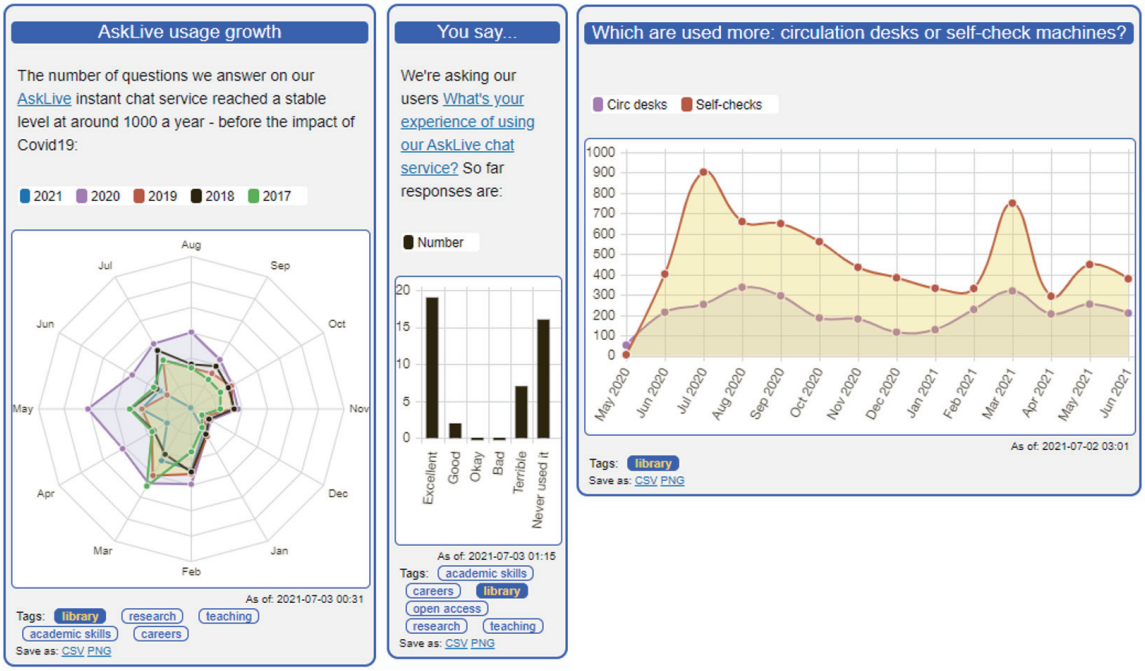
有別於前二類的視覺化之溝通目的，以下列舉案例則是採取探索性之分析目的。美國休士頓社區大學圖書館（圖十一）與德州大學醫學圖書館（圖十二）運用Tableau呈現館藏資料，紐西蘭圖書館（圖十三）雖未運用Tableau工具，也能以視覺化形式呈現年鑑或年報統計結果，且這些圖書館的資料視覺化介面皆可提供互動服務。由此可看出從過往大螢幕顯示已逐漸出現資料儀表板形式服務的雛型。國內圖書館較少看到運用資料儀表板的功能，若能運用此免費工具，可讓讀者清楚掌握館藏與流通情況，且能依讀者需求，彈性調整欲了解之資料維度，用以觀察更細緻的資料結果。



圖十一　休士頓社區大學
資料來源：https://library.hccs.edu/c.php?g=811635&p=7879644



圖十二　德州大學醫學圖書館
資料來源：https://mdanderson.libguides.com/librarydashboard

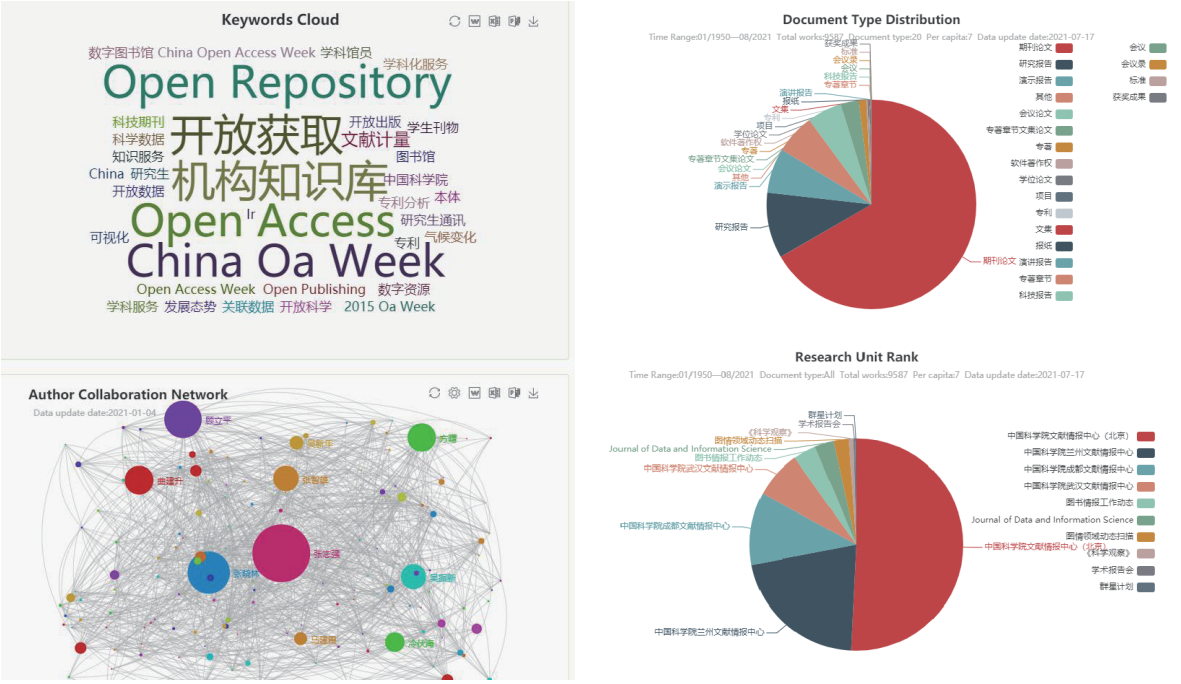


圖十三　紐西蘭圖書館
資料來源：https://library2.lincoln.ac.nz/dashboard/#library

四、機構典藏系統支持的資料視覺化服務

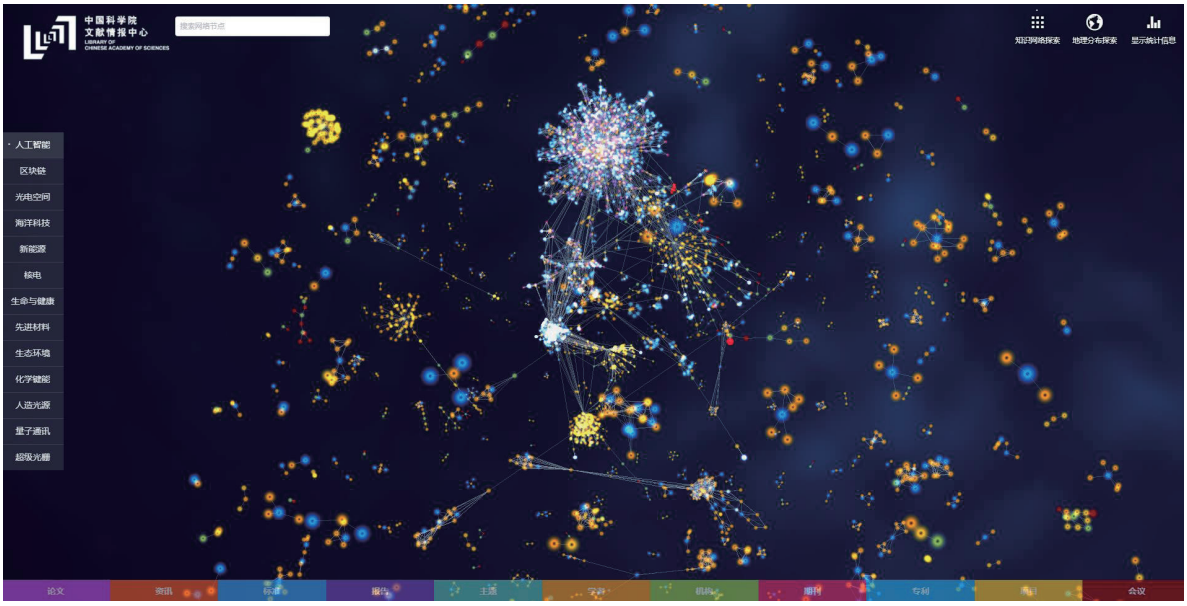
中國科學院文獻情報中心基於探索性分析角度，運用機構典藏中各學科表現、教師學術產出及相關合作網絡等多方數據，向中國科學院所有相關研究所提供資料視覺化查詢服務，包括學術研究的產出統計、關鍵字之文字雲或機構所屬作者之合作網絡（圖十四）。其中，該系統更將教師著作進行分類標記，所標記內容可再做衍生應用，例如圖十五之學科主題多元資料視覺化，既可以檢視主題、機構和作者之間的關聯，畫面下方列出論文、學者、機構及專利等分項資料，可再就各分項資料個別展示重要成果，其學科主題、單位、研究者等面向之研發能量因此清晰可見。

須進一步說明，這些圖表產出的前提是機構典藏系統中具備完善的後設資料（Metadata），且進行的標註與分類結果也須經相關專家檢核，才有辦法產製出具學科發展脈絡的探索性視覺化圖像。因此，可以看到資料視覺化已進階到資料探索階段，具情報決策的服務型態也順應而生。



圖十四　文字雲、作者合作網絡圖譜及學術產出統計

資料來源：http://ir.las.ac.cn/knowledge-analysis?type=topicsmap



圖十五　學科主題多元資料視覺化結果

資料來源：http://ir.las.ac.cn/knowledge-analysis?type=advancedmap

五、COVID-19期間，圖書館的資料服務支援

在新冠疫情發生期間，世界衛生組織（World Health Organization, WHO）與國家科技圖書文獻中心分別整理疫情相關文獻，以提供學術社群進行文獻研究或加值應用，WHO雖未進行視覺化，但已匯聚大量文獻資料並建立完善的Metadata，可供快速檢索之用（圖十六）。國家科技圖書文獻中心除了匯集文獻，也建構館內的互動式資料視覺化服務，例如點選地圖上的特定國家，相關文獻會同步進行重新篩選（圖十七）。可以看出，目前中國在資料視覺化的基礎上，往探索性的資料服務逐步邁進已是明確方向，且在此目的下所分析出的視覺化結果也與傳統視覺化結果有所差異。

圖十六　WHO之COVID-19全球學術文獻資料查詢服務
資料來源：https://search.bvsalud.org/global-literature-on-novel-coronavirus-2019-ncov/

圖十七　國家科技圖書文獻中心之疫情文獻視覺化展示
資料來源：https://www.nstl.gov.cn/service/coronaviridae/index.html

伍、資料視覺化之思考

一、探索性與視覺化的差異

回應本講題內容欲探究之核心，在爬梳資料視覺化相關研究與觀察課程教學現場過程中，已注意到資料視覺化開始出現不同取徑，其差異歸納如表一所示，研究人員若從探索性（分析）角度用於決策與分析上，本質上會採取深度挖掘資料結構，從資料間的內在關係連結產生新的洞見，並強調交互式的服務功能，意即可透過線上系統的介面操作及不同資料維度展現，幫讀者獲取有價值的訊息。本文認為此系統的程式功能會隨著讀者目的不同，其展現的分析層次會從全局結果，逐層拆解資料結構至局部訊息呈現，即從宏觀層次逐步限縮至微觀層次之資料結果。

反觀視覺化（溝通）角度，是一種感性的訴求，目的是簡化複雜的資料，著重視覺溝通與展示，並讓讀者記得資料結果或訊息訴求。所以本文認為視覺化取向，須考量美觀舒適的閱讀（瀏覽）功能，讓讀者了解資料內容或理解複雜議題，展現的層次是以局部角度切入，一張圖只講解一件事，因此在設計圖表時，會特別思考此圖表結果是否能與讀者直覺相輔，例如看到加油的符號就能聯想到加油站。總之，此兩種取徑因面向不同，在結果設計與呈現上也截然不同。

表一　探索性與視覺化之差異

面向	探索性	視覺化
取向	分析腦（理性）	美觀腦（感性）
著重	決策／分析	溝通／展示
本質	深度挖掘數據	簡化複雜數據
功能	交互式	閱讀式
目的	幫助讀者得到有價值的訊息	幫助讀者了解複雜議題
層次	全局結果→局部特點（逐層拆解）	局部特點（一張圖只講一件事）
形式	將數據的各個屬性值以多維形式表示，讓讀者可從不同維度觀察數據	易於理解、與人直覺相符的圖形／符號傳遞訊息的一種形式

二、資料視覺化的價值與困難點

㈠ 價值

本文認為資料視覺化最終仍要回歸到「資料」本身，因為資料是客觀記錄現實事件發生的結果，所以從資料裡挖掘的訊息，某種程度都能反映真實，而此結果是具公信力且有科學基礎，不致產生偏差、失真或歧義疑慮。再者，好的資料視覺化成果能夠清楚展現產製者的訴求，或利於更快發現資料間潛藏的關係連接，可減少不同讀者在感知過程的耗時。最後，有效的圖表訴求若傳遞明確，讀者在感知與解讀資料結果意義上不僅更為便利，其閱讀過程也能夠使人賞心悅目，進而留下深刻印象。

㈡ 困難點

在實際教學現場中，學生即便設計非常美觀的視覺圖像，但在解讀資料意義上，仍有一定的困難。尤其面對同一份數據結果，不同的領域專家解讀各有差異。因此，開始反思

資料視覺化並非單純只用全局訊息角度解讀，例如全臺癌症歷年發生人數，除了流於表面意義解釋外，若要深度挖掘，還需透過醫護、公衛、疾病預防等不同領域專家進行結果解讀的交叉驗證。此外，資料的準確度與資料的維度劃分，例如全臺癌症地圖是以縣市、城市、鄉鎮區或是村里等不同層級區分，所反映的癌症情況各有不同，至今尚無可參照的標準，這些都會加深資料視覺化在結果呈現與傳播之難度。另一方面，在圖表中錯用色彩或符號進而誤導讀者是常見的問題狀況，例如政治議題會因不同的政黨而使用對應的色彩或符號，產製者若無一定的敏感度，可能會為了作品美觀而重新配色，就容易誤導讀者，甚至影響訊息的有效傳播。

陸、結語

不論圖書館或各領域進行資料視覺化的最終目的，皆希望能讓讀者理解產製者所傳達的訴求，甚至幫助讀者從資料中挖掘隱藏的暗示或洞見。圖書館發展至今，已累積豐厚的資料資源，並逐步達成館務資料的「可分析」與「可視化」，最後能從龐大的資料挖掘到「可變現」的洞見，能夠反饋在圖書館永續經營，本文認為此三層運用資料的核心價值，對圖書館在管理決策與整體服務上，期待產生新的啟發與思考。

柒、問題與答覆

一、**問題：**請問以江教授的觀點，訊息派與視覺派的平衡點該如何取捨？是否能舉例？

答覆：訊息派與視覺派發展至今，能發現兩派開始能彼此互相體諒，前提是因為資料訊息量變多，訊息派開始採用裝飾符號，視覺派則逐漸突顯圖片中訊息重點。在資料視覺化的製作過程中，發現傳播學院的同學還無法從資料裡看到重要訊息。舉例來說，零售業在疫情發生前後的營業額與人力資源管理的極大落差是可突顯的重點，但同學很難從此角度進行比較，僅能呈現歷年下滑的趨勢，因而較偏重在以視覺派角度處理新聞報導。但當目標服務對象及決策管理的目的非常明確時，就可由訊息派導向挖掘資料當中可能的意涵或重要訊息。

二、**問題：**請問若圖書館館員想深入進修，有關剛才提到的視覺化工具，您最優先建議哪幾個工具？

答覆：個人很推薦館員優先學習Tableau Public與Google Flourish這兩套工具，尤其是

Tableau Public免費且容易上手，國外非常多圖書館已將Tableau內嵌於網頁上，世新大學與臺灣大學的校務管理系統也是運用此軟體。另外Google Flourish是一個通用平台，在資料表格的操作上，會比Excel或Google表單來得更容易上手。

三、**問題：**請問老師有什麼方法可以培養自己處理資料視覺化的敏銳度？

答覆：一開始可先訓練自己看到一個全局情況，但往往會漏掉20%的細節內容，而此部份可能就是潛在的暗示。例如內政部的性別統計中心曾公開過去新北市家暴案件的年齡層分布，當年齡層分布轉化成表格形式閱讀時，整體年齡層明顯偏高，但若細看能發現12歲以下被通報為家暴者，而這就有違以全局看待資料表格的常理，因此我們要不斷練習從一個80%的全局資料結果裡看到20%的暗示。

四、**問題：**請教老師過去是否會請一些不了解資料的圈外人閱讀，看是否能成功傳達原意？或如何協助閱讀者來獲取資料的重點呢？

答覆：一則新聞報導的目的是讓一般人看懂圖表的用意，所以我們會請相關領域專家協助檢查圖表資料的正確性，再請一般民眾、大學老師或政府官員閱讀這份資料，從他們的回答中交叉比對是否與原意一致。若出現不同認知時，會再回頭詢問領域專家在資料上是否有所遺漏，或是分析維度和顆粒度的不同導致認知差異。這段過程會耗時非常久，原因在於每個人反應皆不同，如何取捨也是一道難題。另外也發現在呈現資料視覺化時，一張圖沒有加註任何說明很難讓讀者看懂，所以除了告知圖表的重點之外，也會讓讀者點選提醒方塊看到圖示的完整發展脈絡，同時維持視覺上的舒適度。

五、**問題：**資料視覺化運用在媒體上時如何避免或善用媒體偏見（Media Bias）？

答覆：在現今媒體快速處理即時新聞的環境中，都是為了視覺化而視覺化，對整體議題擴散或討論並無太大幫助，這也是為什麼我們在處理即時新聞報導時並不會特別去強調視覺化議題，而將重點擺在深度或專題報導上，為的就是要避免媒體濫用資料視覺化。

六、**問題：**中國圖書館界在視覺化上的發展如何？

答覆：因早期在北京大學求學的緣故，所以相關舉例皆以北京的學校為主。北京各圖書館入口處透過大螢幕即時呈現館內情況已是常態。另外會透過館內電子資源或機構

庫提供科研決策服務。文獻情報中心是等同國家圖書館層級，於2010年已開始進行資料視覺化互動服務，只是當時僅以基礎建設為主，還未發展眾多子系統或資料標註業務，這一兩年因疫情關係，沒有辦法到圖書館取用文獻，上述功能發展反而突顯出來。

七、**問題：**老師提到機構典藏有很多資料可呈現，但現在圖書館較少涉獵，請問是否有哪一間圖書館已在這方面有實際成效，值得其他圖書館學習？

答覆：臺灣大學圖書館所做的機構典藏累積了非常多的Metadata，但目前還未看到臺大圖書館有把這些資料重新加值應用。目前所蒐集到的個案只有大陸文獻情報中心在做。

八、**問題：**視覺化前的資料整理（Data Cleaning）是一個關鍵。在中國的視覺化資訊是否會有國家政策的導引？是否有學術獨立性的議題？

答覆：在做資料視覺化或任何加值應用時，資料的乾淨與否會影響後續分析與呈現，當資料沒有標註好或Metadata沒有整理好，後面的分析呈現就沒有參考性，這也是為什麼中國花了非常多人力進行資料清洗，尤其在中國，資料標註已是一個新的職業。資料標註是整個國家政策所支持的研究發展方向，也是這兩三年社科基金、自然科學基金所列入的年度重點選題之一。應該會有國家政策引導，只是層級不會在國務院等級，可能更多是由社科基金與自然基金主導。至於有沒有學術獨立性的問題，是有可能的，因為對中國來說，資料是很重要的戰略資產。學術獨立性比較多的問題是當要運用時，會不會受限於資料整理出來的結果，因此獨立研究中所看到的結果可能需要非常多領域專家協助交叉驗證。至於有沒有相關的議題，以這兩三年的社科基金與自然科學基金審核通過後的題目來看，目前還未看到相關領域專家針對資料清洗或資料標註做專門性研究。

參考文獻

李廣建、江信昱（2018）。論計算型情报分析 。*中國圖書館學報，44*(2)，4-16。

趙斌（2015）。數據可視化在上海圖書館數據展示服務中的應用。*圖書館雜誌，34*(2)，23-29。

羅思嘉（2010）。圖書館營運成效分析。*國立成功大學圖書館館刊，19*，8-16。

Bikakis, N. (2019). Big Data Visualization Tools. In S. Sakr & A. Y. Zomaya (Eds.), *Encyclopedia of Big Data Technologies* (336-340). Cham: Springer International Publishing.

Börner, K., Chen, C., & Boyack, K. W. (2003). Visualizing knowledge domains. *Annual Review of Information Science and Technology, 37*(1), 179-255.

Chen , H. W. (2017). An Overview of Information Visualization. *Library Technology Reports, 53*(3), 5-7.

Chen, C. H., Härdle, W. K., & Unwin, A. (2008). *Handbook of Data Visualization*. Springer-Verlag Berlin Heidelberg.

Debackere, K. (2016). Let the Data Speak for Themselves: Opportunities and Caveats. *Journal of Data and Information Science, 1*(1), 3-5. doi.org/10.20309/jdis.201602

Finch, J., & Flenner, A. (2017). Using Data Visualization to Examine an Academic Library Collection. *College & Research Libraries, 77*(6), 765.

Friendly, M. (2005). Milestones in the History of Data Visualization: A Case Study in Statistical Historiography. In: Weihs C., Gaul W. (eds) *Classification–the Ubiquitous Challenge*. Berlin, Heidelberg: Springer Berlin Heidelberg.

Hey, T., Tansley, S., & Tolle, K. (2009). *The Fourth Paradigm: Data-Intensive Scientific Discovery*. Redmond, USA: Microsoft Research.

Holmes, N., & Heller, S. (2006). *Nigel Holmes on Information Design*. Jorge Pinto Books, New York.

Hubbard, D. W. (2014). *How to measure anything: Finding the value of intangibles in business*. New Jersey: Wiley.

Jiang, Z. H. & Carter, R. (2018). Visualizing library data interactively: two demonstrations using R language. *Library Hi Tech News, 35*(5), 14-17.

Lazer, D., Pentland, A., Adamic, L., Aral, S., Barabasi, A. L., Brewer, D., ... Van Alstyne, M. (2009). Computational social science. *Science, 323*(5915), 721-723. doi.org/10.1126/science.1167742

Liu, S., Cui, W., Wu, Y., & Liu, M. (2014). A survey on information visualization: recent

advances and challenges. *The Visual Computer, 30*, 1373-1393.

Manyika, J., Chui, M., Brown, B., Bughin, J., Dobbs, R., Roxburgh, C., & Byers, A. H. (2011). *Big Data: The next frontier for innovation, competition, and productivity*. McKinsey Global Institute.

Mayer-Schönberger, V., & Cukier, K. (2013). *Big Data: A Revolution That Will Transform How We Live, Work and Think*. Hachette UK: John Murray Publishers.

McAfee, A., & Brynjolfsson, E. (2012). Big data: the management revolution. *Harvard Business Review, 90*(10), 60-66, 68, 128.

Ogier, A. L., & Stamper, M. J. (2018). Data Visualization as a Library Service: Embedding Visualization Services in the Library Research Lifecycle. *Journal of eScience Librarianship, 7*(1), e1126.

Phetteplace, E. (2012). Effectively Visualizing Library Data. *Reference & User Services Quarterly, 52*(2), 93-97.

Qin, X., Luo, Y., Tang, N., & Li, G. (2020). Making data visualization more efficient and effective: a survey. *The VLDB Journal, 29*, 93-117.

Tufte, E. (1983). *The visual display of quantitative information*. Cheshire, Conn: Graphics Press.

Ward, M. O., Grinstein, G., & Keim, D. A. (2010). *Interactive Data Visualization: Foundations, Techniques, and Application*. Massachusetts, New Yourk, NY: A. K. Peters, Ltd.

Yadav, A., & Ahmed, S. (2020). Visualizing Academic Library Data to Understand Use Patterns and Trends. *SRELS Journal of Information Management, 57*(2), 55-62.

本文為知識經濟時代之圖書館服務系列二十八「社會責任及永續發展目標／數據視覺化」研討會（110.04.22）之專題演講紀錄，由張榮輝、周志文協助整理，並經主講者潤稿並同意刊登。