

資料視覺化的奇幻之旅

彭其捷

視覺化數據分析技巧

作者介紹

彭其捷鑽研網路服務多年，出版過 3 本 UX 使用者經驗專書，曾大量接觸設計、工程，使用者體驗等相關工作，近年因大數據 & 物聯網概念蓬勃發展，觀察到越來越多數據導向的服務興起。然而，艱澀的數據需要良好的設計輔助，才能創造良好的閱讀體驗，其中特別依賴資料視覺化的相關能力。因此，本專欄特別針對各類網路服務的資料呈現提出美學觀點，分享一些國內外資料視覺化的概念、工具與案例。

系列文章介紹

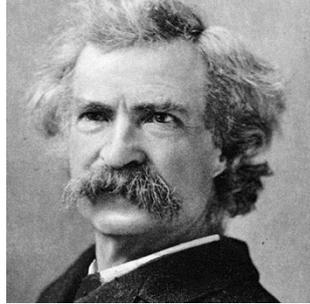
FINDIT 的目標是『發現趨勢，看見未來』，事實上，眾多的趨勢就隱藏在眾多的數據當中，等待著人們去發現、去解讀。透過『資料視覺化』的輔助，能夠把冷冰冰的數據圖像化，協助人們掌握趨勢，更能夠協助新創企業在創新創業過程中找出正確方向，或是幫助投資人找到潛在投資標的，是一項強大的武器。

本系列文章將從資料視覺化的概念開始著手，之後會陸續分享視覺化的經典案例，相關工具介紹等等，希望讀者們能夠在大數據時代，透過視覺化的輔助，說出好的故事。

數據分析是大數據的核心技巧之一，也是數據應用的關鍵所在，大數據當中蘊含許多內藏價值，但需要透過數據分析的過程，才能找出背後潛藏的數據價值，並推理驅動決策改變的證據。而數據分析的方法五花八門，其中視覺分析便是一項重要的技巧，能夠在許多數據分析流程當中帶給我們幫助。本期內容具聚焦在視覺化輔助數據分析三大面向，讓讀者學習如何利用視覺化輔助數據分析技巧。

數字不會說謊，但說謊者會操弄數字

- 馬克吐溫

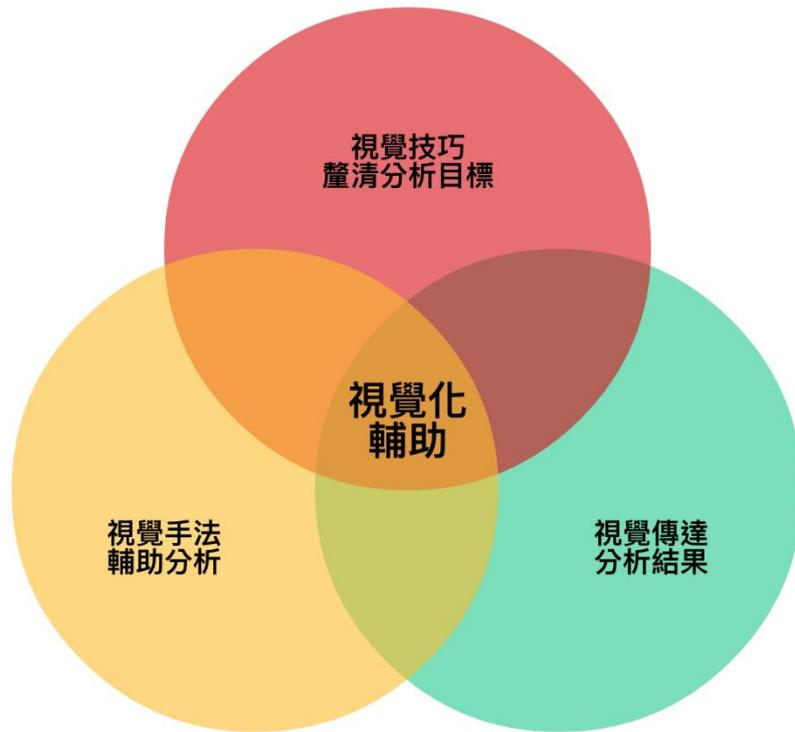


數據分析，指的是透過數據找出有趣的訊息，過程當中會有許多技術性流程，像是數據清理、轉換、建模、統計等等，最終獲得對數據洞察的一連串過程，可協助相關人士進行決策。數據分析是大數據的核心技巧之一，也是數據應用的關鍵所在，大數據當中蘊含許多內藏價值，但需要透過數據分析的過程，找出背後潛藏的數據價值，並推理驅動決策改變的證據。

然而，從茫茫數據海中萃取有價資訊是深具挑戰的，我們需要反覆在分析過程自我答辯，包括像是：是否已經確認分析目標？使用者真正想要看到的訊息是什麼？我們想要透過怎樣的手法進行論證？而最終又想要傳達怎樣的訊息給閱讀者？隨著呈現手法的不同，閱讀者吸收到的資訊也會有所差異。

數據分析的方法五花八門，其中視覺分析便是一項重要的技巧，能夠在許多數據分析流程當中帶給我們幫助，本篇主要介紹其中的三個環節：

- 透過視覺技巧，幫助我們釐清分析目標
- 透過視覺手法，輔助我們進行分析
- 透過視覺傳達，講述正確的數據故事



資料來源：本文繪製。

圖 2-1、本篇分享的視覺化輔助數據分析三大面向

一、透過視覺化技巧釐清分析目標

數據分析的第一大問題，是釐清分析目標，如果一開始的問題定義錯誤，後面的方向很容易偏移軌道，真正有感的分析任務應該要跟業務問題有重要連結，而分析結果則需要實質對業務帶來幫助。



資料來源：本文繪製。

圖 2-2、在釐清分析目標階段加入視覺化技巧的輔助

數據分析的目標包括許多細節，可能每個人心中所想也不完全相同，舉例來

說：

- 數據分析的最後的目標族群為何？
- 分析結果是想要給誰看呢？
- 這些人想要看到的是什麼呢？

一連串分析問題包括了目標統一與溝通的過程，以下介紹幾種視覺化輔助技巧，能夠幫助我們更快釐清分析目標：

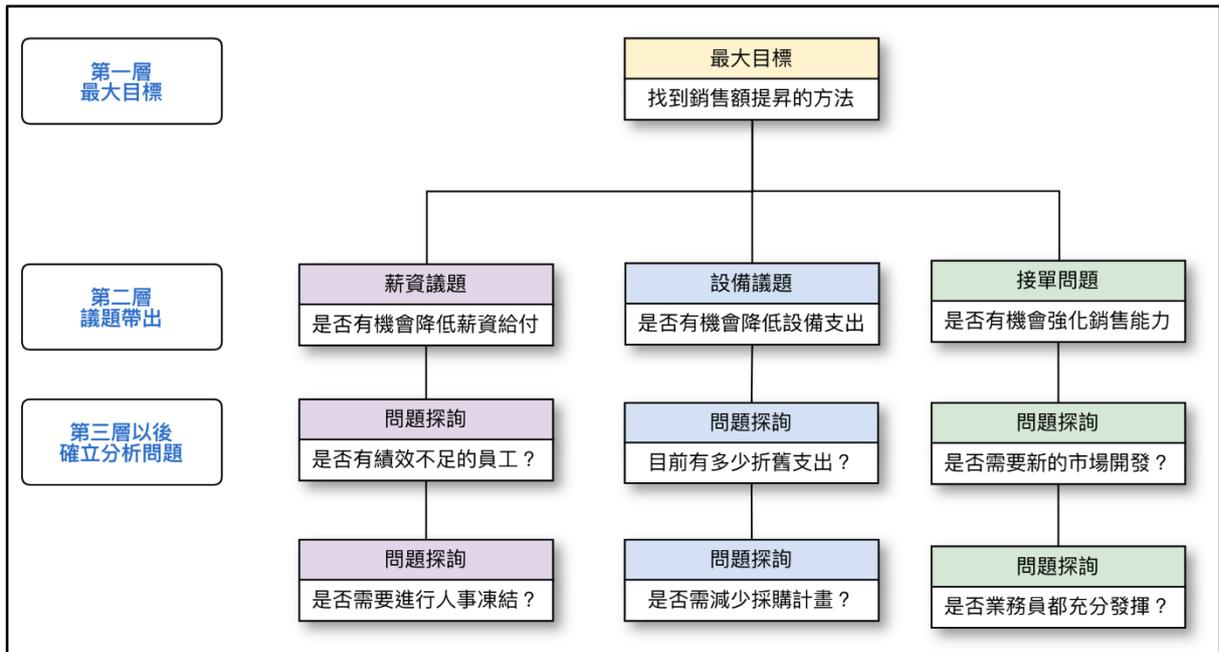
- 金字塔法：確立分析目標階層
- 視覺化數據側寫：視覺化業務相關統計數據
- 便利貼紀錄法：留住每一個精彩的點子
- 視覺雛形設計法：強化溝通素材
- 走出辦公室法：帶著視覺圖與關鍵人物溝通

金字塔法：確立分析目標階層

拿到一批數據後，因為可以分析的問題與方向有太多排列組合，很容易陷入數據當中無法自拔，有沒有好用的技巧可以幫助整理問題呢？本段介紹的：『金字塔法』就很適合幫助我們進行目標與問題的溝通與釐清任務。

金字塔法指的是透過：『最大目標』-->『次要目標』-->『分析任務』的任務拆解技巧，透過金字塔形狀的視覺呈現，在一個畫面中彙整我們感興趣的分析問題。金字塔方法可以區分多層，而最常見的是分成三個階層，分別是：

- 第一層：確認研究目標
- 第二層：列出重點議題與假設
- 第三層：列出每項議題需驗證的分析問題



資料來源：本文繪製。

圖 2-3、透過金字塔視覺呈現方法，確立分析目標與問題

金字塔呈現的是一個普羅大眾都能讀懂的重要性階層，所以是一個非常好用的溝通工具，在數據分析的初期可先放進簡報當中，拿來與相關人員進行溝通與目標確立，也能快速引導出有趣的分析問題。

視覺化數據側寫：視覺化業務相關統計數據

如果我們一開始不知道從哪裡著手分析，可先進行視覺化數據側寫任務，指的是針對部分數據直接進行統計圖表的製作，像是長條圖、線圖、地圖等等基本圖表即可，這裡的重點在於『利用短時間製作出基本圖表』，避免跳進太深的分析問題當中，這些繪製任務必須在很短的時間內完成（可透過 Excel, Tableau 等等視覺化工具點擊達到），卻很有機會引導我們找到許多有趣的分析議題。



圖 2-4、透過簡單的視覺統計圖表側寫，能夠幫我們快速找出有趣的研究議題

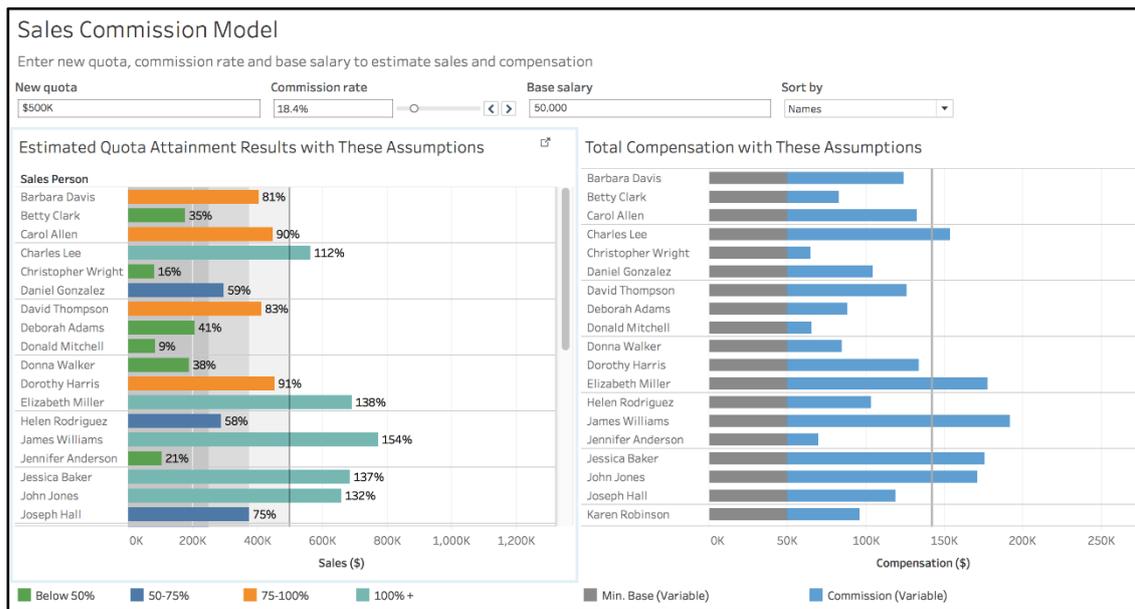


圖 2-5、透過基本的統計視覺化，就能幫助我們觀察到許多資料的特性

便利貼紀錄法：留住每一個精彩的點子

數據分析過程中，常常會有非常多的點子與問題產出，當我們透過溝通、討論得到一些感興趣議題的同時，務必當下把這所有好點子完整記錄下來，這些素材都可能成為未來分析的養分。

然而，傳統的『會議紀錄』或是『大腦記憶法』很難有效率的收錄相關細節，它比較像是一個討論的黑盒子，由於會議紀錄通常是私人所有，只會有某些特定的人才能看到，所以與會者可能較無法深入參與討論與確認點子的過程，所以這裡推薦可用『便利貼』作為點子溝通的紀錄與分享工具。

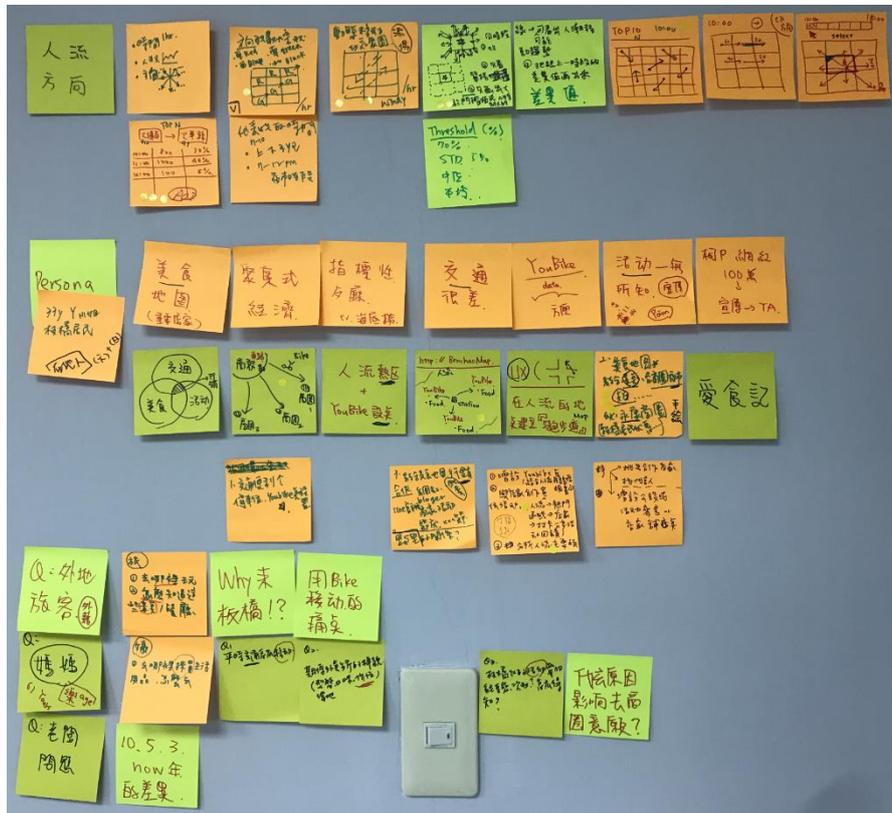


圖 2-6、數據分析討論時的超好用的視覺輔助工具 - 便利貼

便利貼的強項在於能夠快速產出，也能快速改變資訊的排列組合，或是快速修正點子（揉掉就好了），最強大的部分則是快速可視化大家的點子細節，引導所有人充分思考，相較於傳統的私人『會議紀錄』，透過便利貼來進行點子的溝通交流，是數據分析溝通與點子紀錄的好用技巧。

超好用的便利貼工具



圖 2-7、小小的便利貼擁有許多難以取代的優點

視覺雛形設計法：強化溝通素材

透過前述的金字塔法、數據側寫法、便利貼法等等的腦力激盪之後，我們應該已經可以訂出一些分析目標，例如：

- 希望能夠看到近半年的訂單數據變化
- 希望能夠同時觀察訪客性別與下單轉換率之間的關聯性
- 希望能夠同時比較忠實用戶與訪客之間的比較
- 希望能夠看到各個業務銷售額比較表

然而，看起來合情合理的分析目標，細節還是有許多可琢磨之處，舉例來說：『希望能夠看到近半年的訂單數據變化』這句話，指的是哪一個數據呢？或是：『希望能夠看到各個業務銷售額比較表』，這個分析目標指的是希望看到最近的結果嗎？還是想要看數年長期的績效報表呢？

也就是說，就算是大家都點頭同意，或是白紙黑字紀錄下來的分析目標，對於實際進行數據分析任務的人來說，依然存在許多模糊地帶，我們如果直接就開工，有可能在一開始就制定了錯誤的方向，為了解決這樣的問題，我們可以透過『視覺雛形』的技巧來釐清，主要是透過紙筆輔助繪製手繪草圖，視覺化來溝通並確認分析規格。

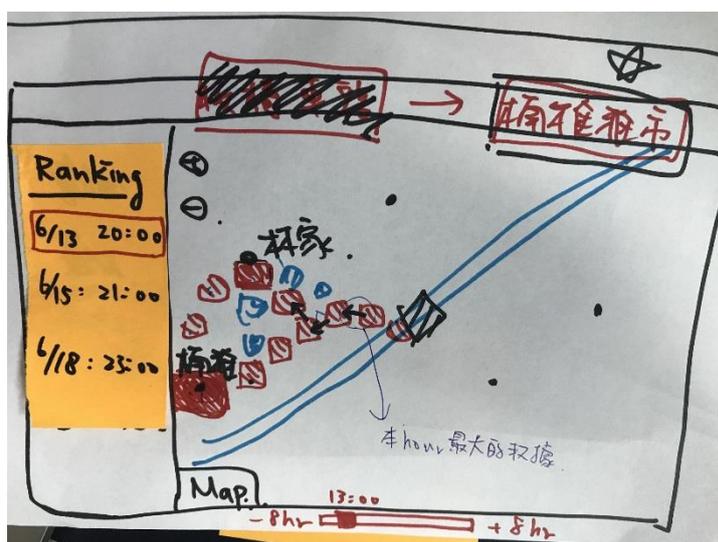


圖 2-8、確認視覺化目標的時候，可先用手繪的方式簡單勾勒視覺樣式
可以帶來更多的討論，釐清更完整的分析細節

走出辦公室法：帶著視覺圖與關鍵人物溝通

數據分析並非全然是一項邏輯技能，許多時候『情報』更是帶來重要觀點的分析技巧，尤其當我們自己並非該議題的專門人士的時候，務必找到該議題的關鍵意見領袖，透過優秀的溝通與訪談技巧取得許多有趣的數據故事與情報。

然而，許多人基於各種原因，並沒有把與關鍵人物溝通的任務放進分析流程中：

- 受限於實驗室或是辦公室的框架，出門需要上級同意
- 還要特別安排出差行程，流程好麻煩
- 要跟不認識的人接觸，但是我天性害羞

- 我自己看過數據，我知道我在分析的是什麼
- 會不會對方覺得我的分析問題大有問題，如果是的話怎麼辦



圖 2-9、避免讓辦公室的框架減少了我們了解問題的好奇心，大膽走出舒適圈吧

以上種種，都讓許多分析人員選擇較輕鬆的執行方式，也就是『待在辦公室裡面，想像使用者的反應，並做出改善』，由自己定義各種分析問題，缺少重要人物的觀點。在此強烈建議在數據分析的過程中，務必使用『走出辦公室法』找人聊聊天，不論是銷售人員、一般民眾、路人、可疑對象等等，可以約在一個舒服的咖啡廳環境，聽聽對方的想法，透過溝通的過程釐清分析問題，很可能成為數據分析的重要觀點。

然而，由於直接與目標對象進行訪談有時容易太發散，建議可以整合上面的『視覺雛形設計法』，將設計出來的分析雛形給對方看，作為溝通工具，對方就能直接針對現有資訊說出『好』與『不好』的地方，幫助我們取得關鍵分析情報。



圖 2-10、透過視覺統計雛形圖表，能夠引導對方快速進入重點分析問題的討論

二、視覺輔助分析六大心法

數據分析並非單一方向的邏輯推論，更像是複雜與交錯的驗證過程，在實際分析階段，視覺技巧能夠幫助我們釐清許多思路，看出資料之間有趣的故事，以下就介紹若干手法，讓我們利用強大的視覺技巧進行分析。

心法一：宏觀與微觀視覺階層切換法

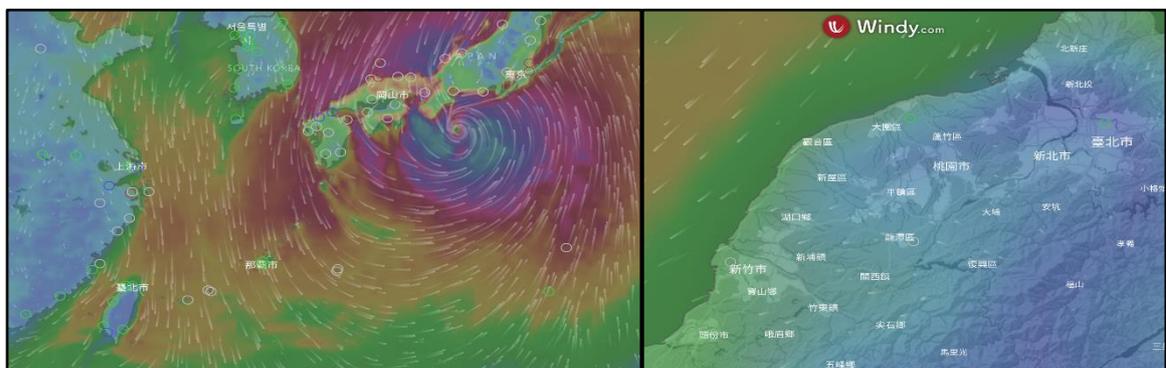
進行數據分析任務時，很容易跟真正的答案擦肩而過，例如我們很容易分析到錯誤的時間區段，可能觀看的是分鐘的變化，但真正數據特性卻出現在小時變化或是秒值變化時；又或是地理資料的分析，特性故事只會出現在特定的地理階層中，但我們卻查看了過大或是過於細節的地理位置數據。也就是說，執行數據分析任務時，我們需要不斷切換宏觀和微觀層面，用交互式的方式探索並發現數據出現價值的地方，下表整理了兩種觀點的分析特性：

表 2-1、宏觀與微觀視覺

觀點	分析特性	說明
宏觀觀點	觀察	透過更廣域的時間、地理特性，觀察數據的整體特性
微觀觀點	解析	鎖定細部的分析議題，像是特定區域、特定時間、特定資料的細微觀察，找出細節中的資料特性

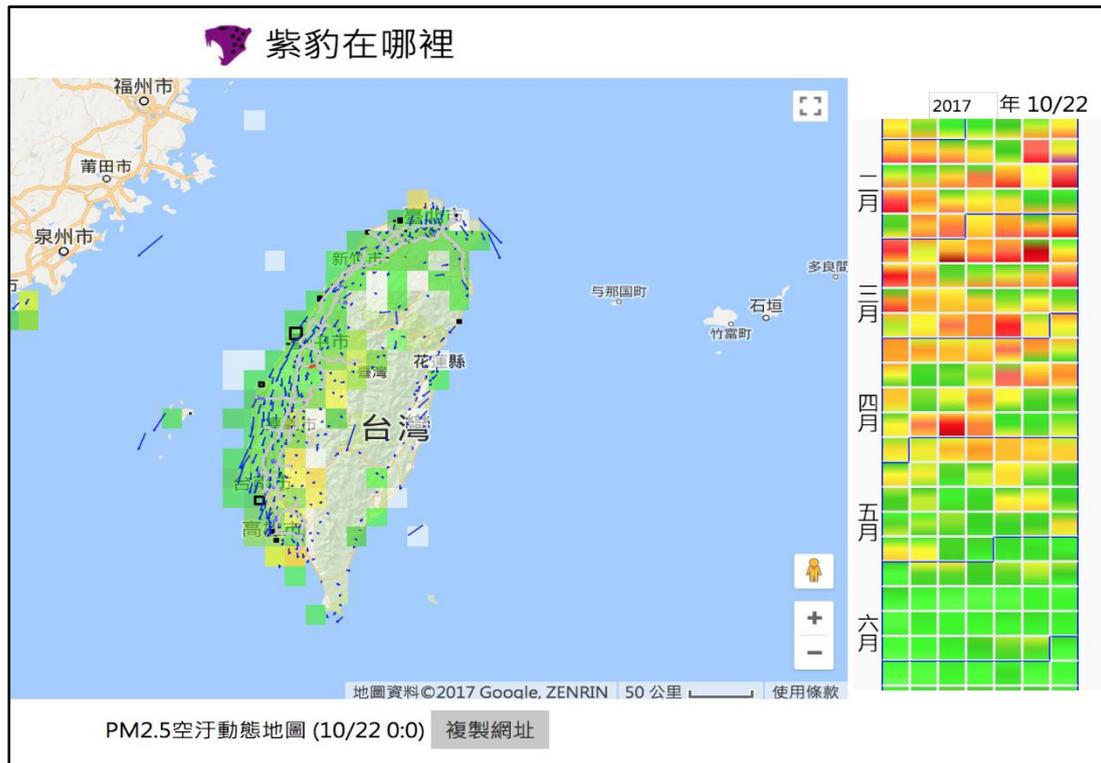
資料來源：本文整理。

數據探索時，我們可不斷在宏觀層面與微觀層面之間進行切換，不需要侷限自己的分析框架，如果我們解析資料的細節卡關時，不妨跳脫出去，拉到更大的地理區域，或是更長的時間區段，有時資料的故事便會自己浮現。



資料來源：<https://www.windy.com>

圖 2-11、不同觀點的地理視覺階層，資料能夠解釋的故事也不同



資料來源：<https://purbao.lass-net.org>

圖 2-11、紫豹在哪裡，顯示的是台灣的空汙數據，右邊區域並列全年的空汙指數資料視覺化，用時間的宏觀觀點進行大範圍比較

心法二：嘗試多種呈現形式

取得資料後，同一份資料可以嘗試多種呈現形式，例如從線圖改成地圖的方式呈現，或是將表格改成用熱度圖的方式呈現等等，有時即使是一些細小的設計改變，比如色彩配置的變化，對某個變量的視覺呈現方式等等，都可能可以看出截然不同的資料特性。

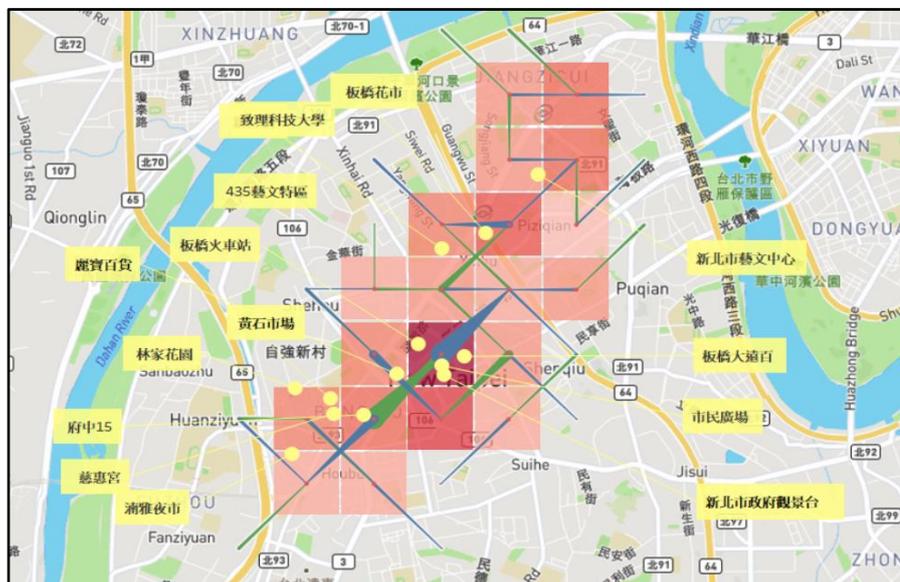
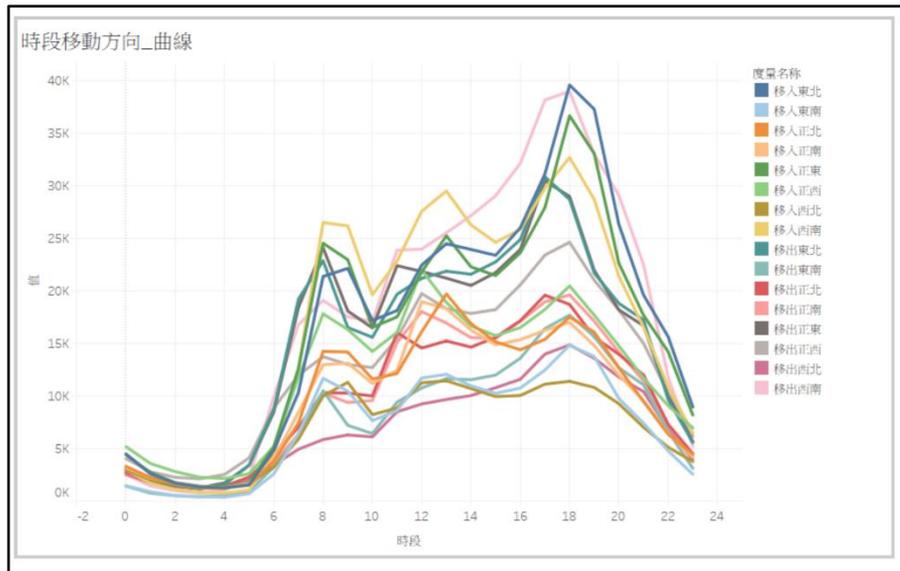


圖 2-12、同樣的資料採用不同的呈現方法，分析結果可能會有很大不同

心法三：視覺聚焦與過濾法

當一張圖同時藏有太多資訊時，我們可以使用視覺聚焦的技巧幫助大腦過濾重要資訊，這個技巧非常簡單好用，也就是忽略整張圖出現的大部分資訊，刻意只觀察其中一個區塊，或是乾脆把跟目前分析問題無關的資訊都隱藏起來，這樣可以讓我們的大腦更專心解析眼前資訊，降低分析所需的心智負擔。

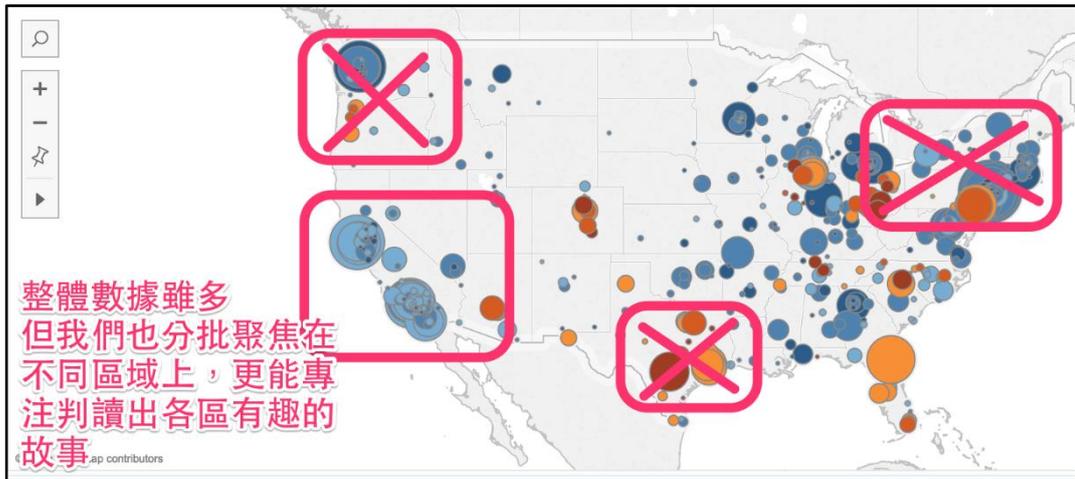


圖 2-13、專注觀察局部資訊，有時更能捕捉到數據細節

視覺聚焦法的目標在於讓資訊降維，其中可以使用的技巧像是：

- 隱藏部分資訊圖層
- 透過過濾器關閉部分資訊
- 重點顯示局部資訊
- 只觀察局部資訊

我們也可直接進行重點數據標記，特別把感興趣的資訊做重點標記出來，或是畫面中乾脆只留下想要觀察的特定資訊。



圖 2-14、重點顯示 (Highlight) 的呈現形式

心法四：留心反常與極端值

一般來說，如果跟機器或是人有關的數據，大多隱藏著某種慣性，將資料視覺化之後，我們可以特別留心那些跟平常長得不太一樣的圖像，因為之中可能隱藏著某個精彩的數據故事。

- 這裏舉幾個反常的分析情境，像是：
- 整天就只有這個小時的數據特別高
- 這三個月都呈現穩定的數字變化，但是其中一天出現一個極大的變異值
- 全區都是男性參與者較多，但只有這一區是女性參與者較多

我們總是會對這些反常的數據特別感興趣。

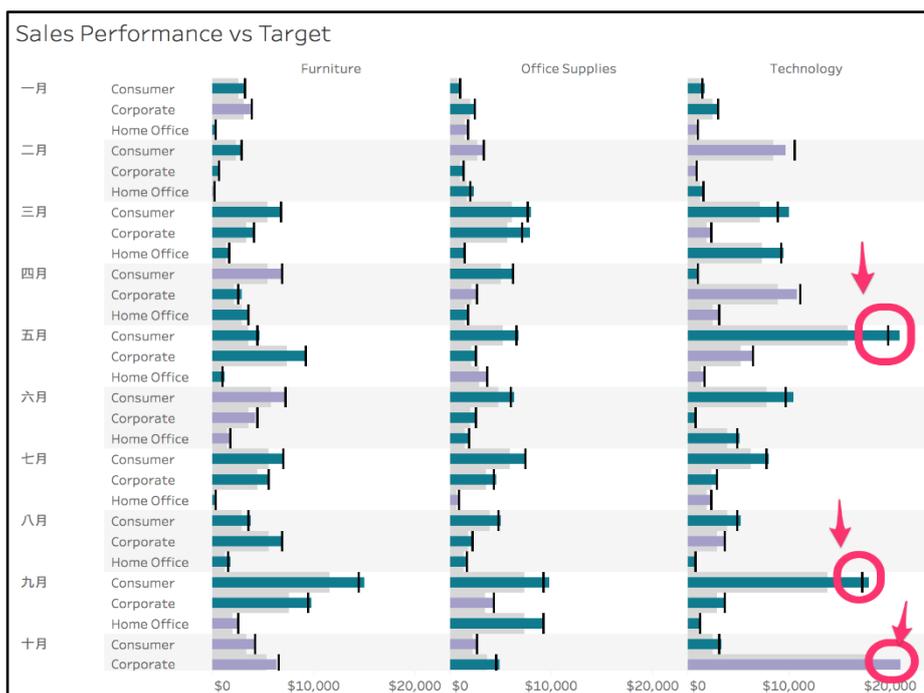
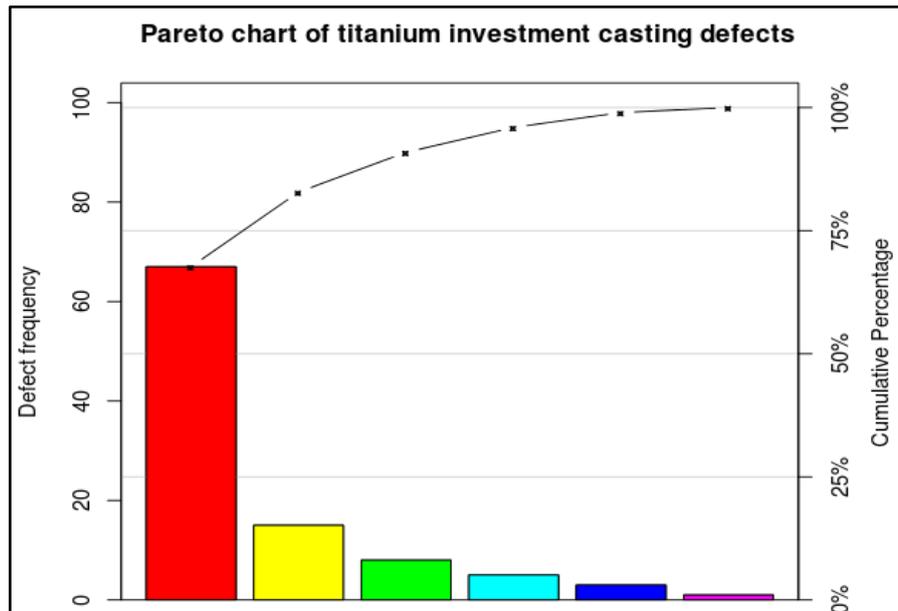


圖 2-15、我們可以特別觀察一些特別高或是特別低的反常數據，觀察其資料特性

此外，反常也可延伸為極端值的議題，像是觀察績效表現特好或特不好的、數據特多或特少的、地理位置最偏遠或是最近的等等，可作為其中的分析重點，

我們可以搭配許多軟體都內建有的排序功能來撈出這些極端資料。

這裡分享『柏拉圖 (Pareto Chart)』方法，又稱排列圖法，是資料品質檢驗上常用的方法，透過長條圖顯示問題原因發生頻率的排序，並以遞減的方式呈現，同時透過線圖也可以觀察累進的百分比例，簡單而易用，我們可以應用此觀念來分析許多問題。



資料來源：[維基百科](#)

圖 2-16、柏拉圖 (Pareto Chart) 方法圖，就是應用排序找極端值法找出重大品質問題

如果是地理資料的視覺化分析，我們也可純顯示較重要、較多的數據，來避免資訊的過度呈現，提升決策與判讀的能力。

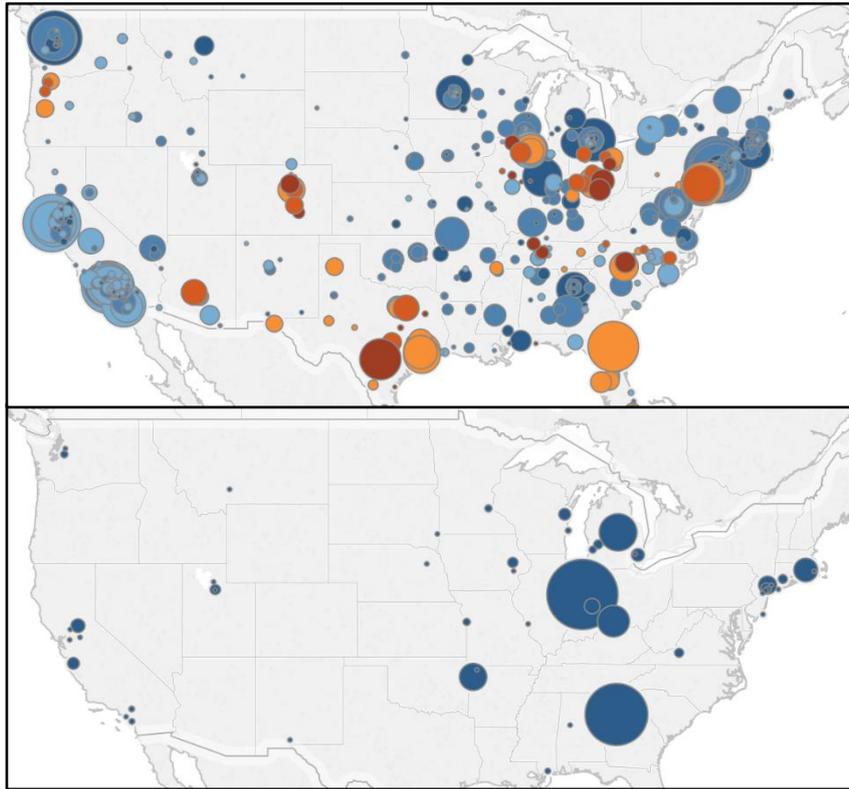


圖 2-17、上圖是完整數據，下圖則是排序後只顯示高利潤比例的資料點

心法五：輔助使用參考線或參考區間

如果我們想要快速比較一些整體資料差異性，或是區域資料的差異性時，可以善用參考線或參考區間的方法，用視覺元素標記參考資訊，強化分析者的判讀速度。

參考線是在畫布當中放置一條讓分析者參考的數據常數線，例如我們可以設定一個『分數 = 60』的水平線，如此一來只要位在該線以下的都是不及格的學生，或是在畫布當中標記一些統計資訊的參考線，例如：平均值、中位數、最大值、最小值等等，分析者能更快看出許多有趣的資料故事。

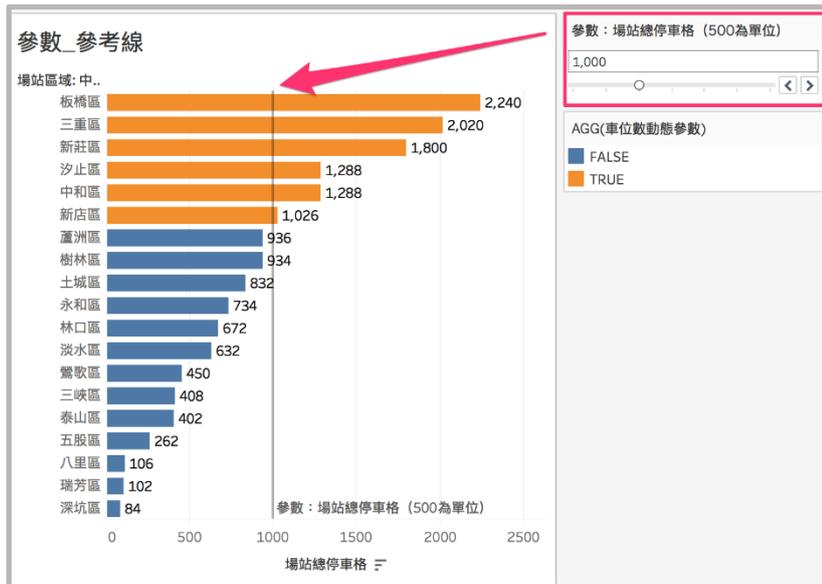


圖 2-18、拉一個 1000 的常數參考線，幫助分析者快速識別大於或小於

參考區間的概念跟參考線很像，不過標記的是一塊區域而非一條線，常用於幫助離群值或是資料的分佈狀況，這些資訊在沒有參考區間輔助時要判斷是比較辛苦的。

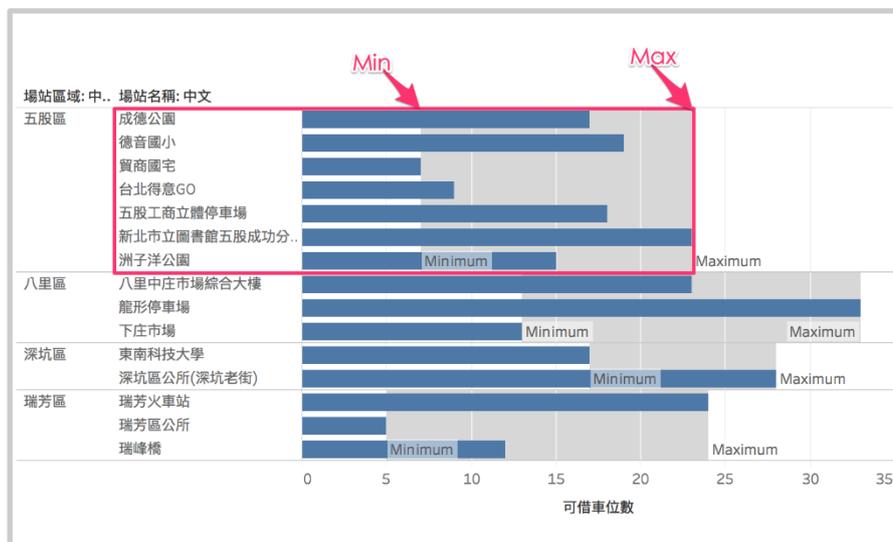


圖 2-19、參考區間標記的是一整塊區域，能夠方便進行資訊的區間識別

三、視覺化傳達分析結果

視覺化的技巧，對於資訊的揭露上影響重大，就算是極佳的分析成果，如果使用了錯誤的表達形式，也會對閱讀者的吸收程度大打折扣，本段分享一些視覺成果分享的心得法則，供讀者參考。

有效傳達資訊重點：長條圖

商業資料視覺化的目標，並非只是做出漂亮的圖表，而是能夠有效傳達正確資訊的圖表，因此，在考慮數據視覺化設計方案時，我們要先問問自己：『目前的呈現方式，會比長條圖清楚嗎？』

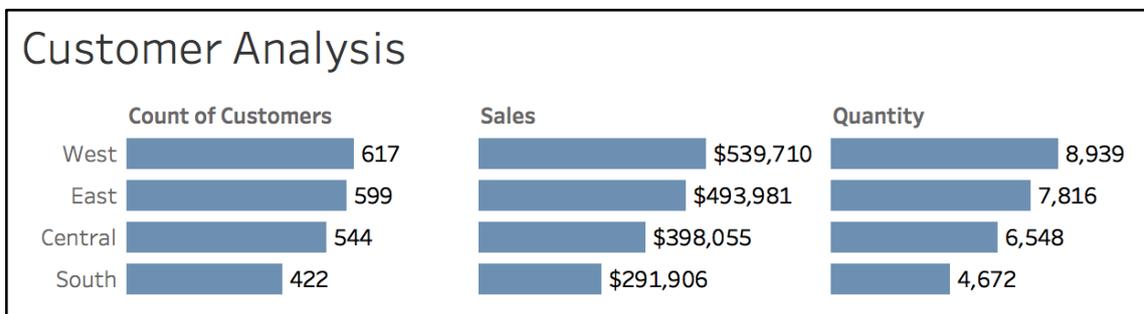


圖 2-20、長條圖雖然簡單，卻是最好理解的視覺化形式

在進行多項數據大小比較時，很少有其他呈現方式比長條圖更好，原因包括：

- 不需要學習，多數人都看得懂
- 製作容易，所有統計工具都內建相關功能
- 人類非常擅長使用長度進行大小的比較
- 呈現方式簡單、簡潔、俐落，不會傳達額外的、錯誤的資訊

我們在視覺圖像上有許多選擇，如果是希望能夠快速吸引目光，可以考慮使用許多新奇酷炫的表達形式，但如果真正的目的是希望讓閱讀者能夠吸收到正確的資訊，長條圖很可能是最佳的策略。

展示更多的證據：表格

許多讀者可能會好奇，『表格』算是資料視覺化嗎？這個答案當然是 YES，表格最擅長的就是忠實將結果陳列出來，把資料整整齊齊的排放在一起呈現，如果能夠適當加入一些顏色或是視覺上的輔助，清楚又有說服力，能夠說服許多對於分析成果有意見的人！

		一月	二月	三月	四月
Furniture	2014	\$6,243	\$1,840	\$14,574	\$7,945
	2015	\$11,740	\$3,134	\$12,500	\$10,476
	2016	\$7,623	\$3,926	\$12,801	\$13,212
	2017	\$5,964	\$6,866	\$10,893	\$9,066
Office Supplies	2014	\$4,851	\$1,072	\$8,606	\$11,155
	2015	\$1,809	\$5,368	\$15,883	\$12,559
	2016	\$5,300	\$6,794	\$17,347	\$10,647
	2017	\$21,274	\$7,408	\$14,550	\$15,072
Technology	2014	\$3,143	\$1,609	\$32,511	\$9,195
	2015	\$4,625	\$3,449	\$10,344	\$11,161
	2016	\$5,620	\$12,259	\$21,568	\$14,891
	2017	\$16,733	\$6,027	\$33,429	\$12,383

圖 2-21、表格本身即是一個超好的視覺化呈現手法，也可根據數字改變標記顏色的深淺

展現趨勢與未來預測：線圖

如果我們希望呈現資料隨著時間序列的改變，最好的表達形式是線圖，線圖很適合呈現像是市場的變化、成本的變化等等，能清楚呈現指標數據變化。

此外，線圖也很適合呈現未來的數字預測，前方用實線繪製已經發生的資訊，後方則搭配虛線或是較淺的顏色，暗示尚未發生的預測數值，視覺上可在中間加入中斷，或是在未來預測區域加上數字的上下誤差區間，幫助閱讀者判斷未來可能的數字預測變化。

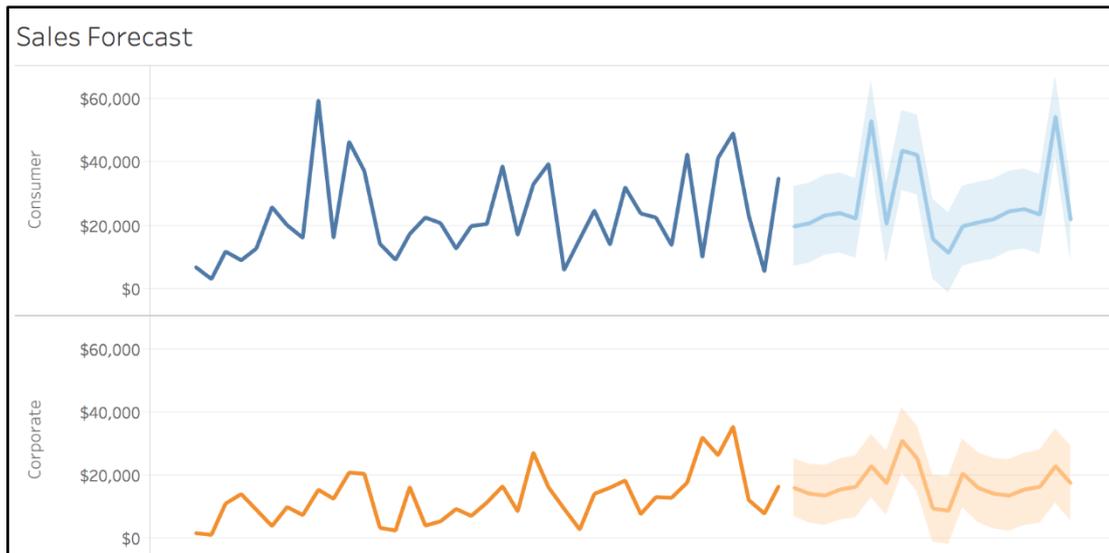


圖 2-22、線圖很適合呈現目前數據變化與未來的預測結果

表達相關性與離群值：散佈圖

散佈圖主要用來表現兩個度量值的相關性，透過資料點位置的標記，帶領我們判讀資料的特性，像是資料的正相關、負相關、弱相關、強相關、不相關等等。另外，散佈圖也能夠幫我們快速判讀離群值，或是展現迴歸與最佳化關聯關係。

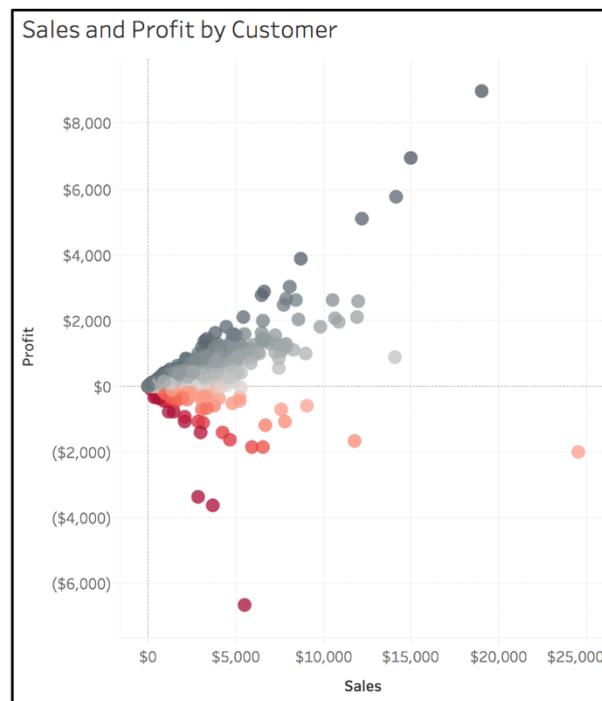


圖 2-23、觀察散佈圖呈現的結果，就可大略看出其 X,Y 維度的相關性以及離群值

如果要繪製散佈圖，我們需要成對的資料點（例如 X 與 Y），通常會用垂直 Y 軸表現數據度量，並用水平 X 軸表達相關因素數值，搭配呈現維度之間的兩兩對應關係，像是『下雨機率』與『氣壓』之間的關係，或是『銷售能力』與『年資』之間的關係等等。

展現多樣化分析結果：儀表板

資料儀表板，指的是在同一個畫面中同時呈現多種數據內容，放在同一個畫面的好處在於讓認知降維，讓人們減少圖表切換的動作，也同時減少相關切換所導致的智力損耗，運用超強大腦看出跨數據的整合特性。

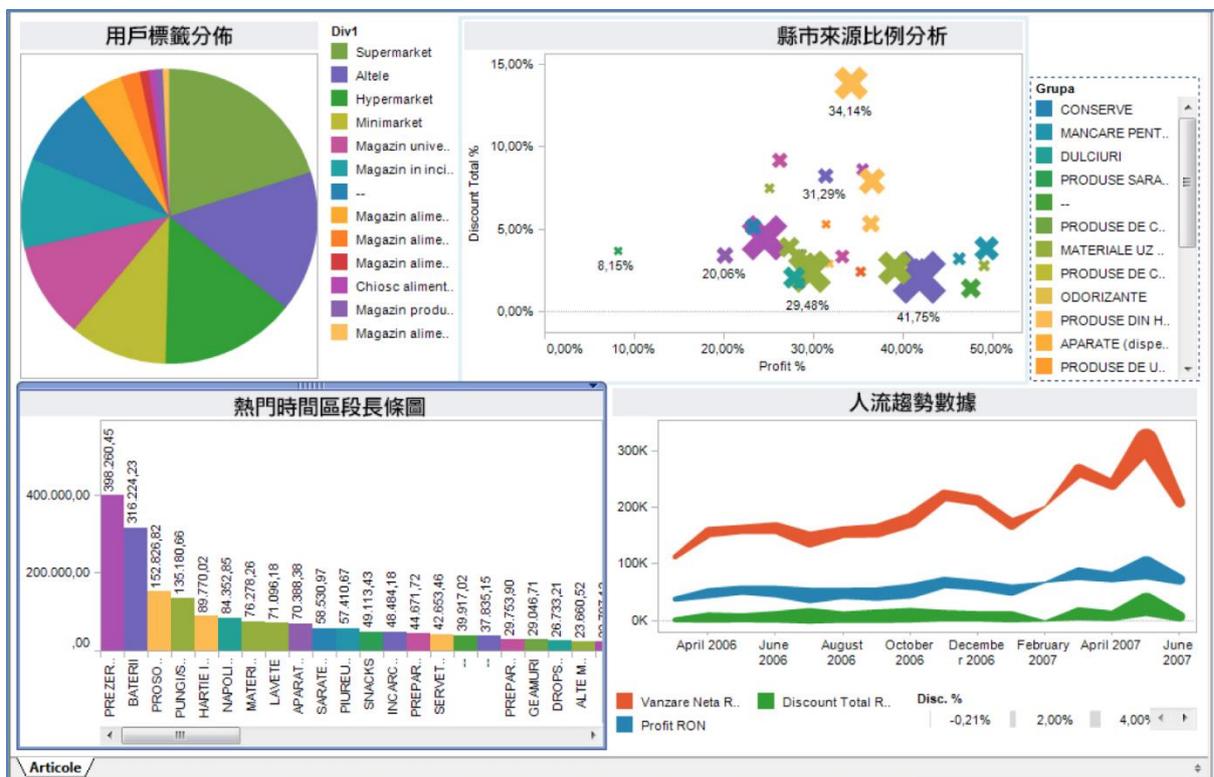
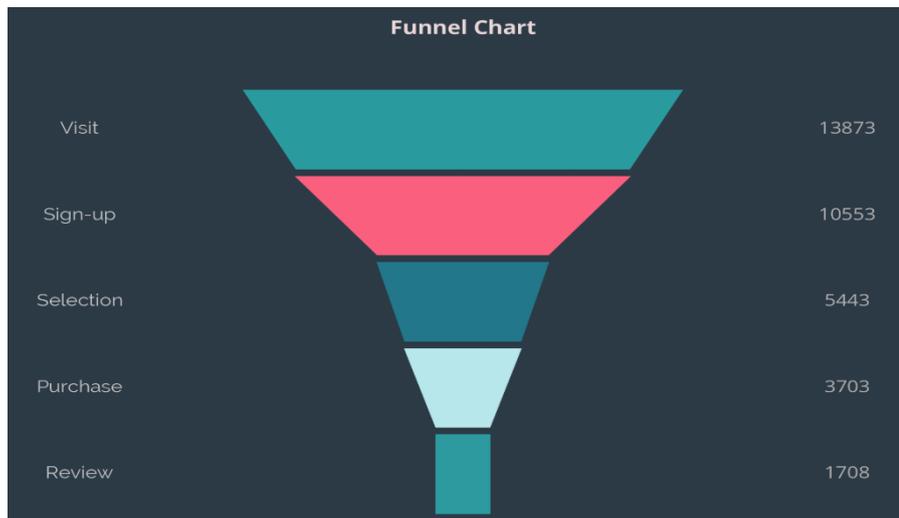


圖 2-24、視覺儀表板，能夠善用大腦的多維分析能力，讓使用者同時觀察多重數據關係

展現資訊轉換與過濾：數據漏斗

我們在分析商業行為時，很常使用數據漏斗進行呈現，像是用戶註冊轉化漏斗、電商網站下單漏斗、頁面操作漏斗等等，漏斗圖擅長判斷不同階層流失的數據（訂單、註冊人數等等），讓我們可以有效鎖定設計上的重大瑕疵，並挑選重

大問題做進一步的成因分析，最佳化每一個業務流程的轉換效益。



資料來源：<https://moderndata.plot.ly/funnel-charts-in-python-using-plotly/>

圖 2-25、數據漏斗能夠幫助我們判斷每一層的數據轉換

強化結論：使用提示框

對於分析者來說，每張圖表都經過精心設計並反覆觀看過，因此對於畫面所有資訊都了然於胸。然而，對於初次閱讀者來說，他期待能夠在很短時間就看懂你的圖表，不然很可能產出挫折感，甚至放棄閱讀！這時，建議可以在圖上添加重點視覺提示框，記得盡量放在背景較空白的位置，讓資訊清楚可讀，其有關鍵的引導性作用，提升閱讀者對於圖表的理解力，也提升理解信心。



圖 2-26、小小的提示框，能夠帶給閱讀者莫大的信心

結論

數據分析的成果有時會讓人看起來好像理所當然，但其實所有的論述要確立下來，需要反覆的檢視，分析者沒有經過系統的培訓，要勝任這項工作並不容易。

數據分析任務包括了許多環節，本篇主要介紹的是視覺技巧的輔助，除了一開始能幫助我們釐清分析目標之外，也能在分析階段提供許多輔助，或是最終幫助我們產出讓閱讀者好讀的分析成果。