

## 推薦序

在強調以數據驅動治理的當代，有效的數據溝通技能儼然是就業市場中必備的重要能力之一。身處教育前沿的工作者，深知資料視覺化不僅是技術，更是一門兼具美感溝通與傳播的藝術。基於這樣的認識，個人高度肯定本校劉育津教授、申元洪教授以及溫志皓教授三位共同撰作本書，提供從基礎到進階的全面視覺化觀念、工具和應用等知識。

本著作首先廣泛介紹視覺化觀念，如視覺化的科學原則，以及教導讀者如何減少視覺雜訊並增強資訊的清晰度和吸引力；接續更深入探討視覺化工具在實際應用中的科學基礎和設計原則。此外，妥善加入實用的工具解說，特別是 Power BI 的運用，書中深入探討視覺化多樣圖形之應用情境，期使讀者選擇最適合呈現數據的方式，並靈活運用這些工具解決實際問題以達溝通與傳播之效。從基礎建立到企業實戰，三位教授以結構清晰的章節，引導讀者理解和應用資料視覺化的多樣化技術，並透過案例學習如何將數據轉化為洞察力。

作為大學校長，由衷為世新大學能出版如此高品質和實用的教材感到自豪。《商業大數據的視覺化設計與 Power BI 實作應用》是一本必備的參考書，無論學生、教師還是專業人士都能從中獲益匪淺。此一專業著作的出版不僅展示我們學術團隊的專業知識和創新精神，也反映了幾位教授對學術與實踐能力的承諾。

陳清河

世新大學校長  
2024 年於世新大學

## 推薦序

隨著科技風雲變幻，數位江湖波濤洶湧，資料視覺化、機器學習與人工智慧仿如令人目眩的武功心法，不斷更迭，引領風騷；更強度關山，創立諸多門派，伏妖降魔，扶持正道，成為現代人江湖行走不可或缺的知識。我輩身處風波難測的險惡江湖，又非練武奇才，如何才能不被淹沒，甚至乘此長風破此強浪？相信各位道友都已有答案——一本能夠一窺絕世心法的武功秘笈。

今有三位集理論、實作與教學經驗於一身的武林宗師，溫志皓主任、申元洪教授、劉育津主任，為提攜武林後進，不惜耗費大量功力，嘔心瀝血、閉關兩年，著成此寶典。期間用心之苦、用力之艱實難與外人道。今大著初成，武林絕學得以承續，特著此序感謝三位大師無怨無悔的付出。

本秘笈揭示資料視覺化的理論心法、兵器招式與實務應用，深入淺出、層層疊進。資料視覺化內功心法強調將數據複雜內涵呈現於眼前，使其無所遁形；而 Power BI 則如神兵利器，將心法威力發揮到極致，使視覺化達到前所未有的便利與高度。本秘笈因此也強調使用此利器搭配心法。

資料視覺化與大數據分析與人工智慧，這兩門武功相輔相成。看透資料與數據隱藏的特性，方能用最恰當的武功心法與利器、克敵制勝。

誠摯希望各位武林同道皆能從本秘笈獲得寶貴的經驗和啟示，以期武功精進，技藝大成。

江湖路遠，正氣長存，願與諸君共勉。

許秉瑜 敬筆

國立中央大學副校長  
中華企業資源規劃學會秘書長  
民國一百一十三年八月

# 自序

Power BI 被公認為一款易學且功能齊全的資料視覺化工具，同時也是在學術及職場環境中最為普遍且易於獲取的軟體之一。我們衷心感謝中華企業資源規劃學會秘書長暨中央大學副校長—許秉瑜教授的盛情邀請，提供我們這個難得的機會，親身參與並貢獻於本書的創作。

本書的寫作角度，是從資料面出發，與坊間大多是從功能面的介紹有著明顯的差異。從資料面的出發，著重在我們於工作時會接觸到的商業大數據。面對資料，我們應該如何理解並選擇適當的圖表去呈現其中的資訊，進而用資料說故事。使用九種的大類別，以及其中各自不同的圖表去讓資料說一個好故事。同時，我們也會在其中，說明每一種不同的類別與圖表，在設計與應用時的優劣分析，減少使用者在資料視覺化的試誤成本。

感謝一路上協助我們的每一位夥伴及指導我們的每一位師長。最後，我們對作者團隊的密切合作表示感謝，正是這份團隊精神使得本書得以問世。

溫志皓、申元洪、劉育津

2024 年於世新大學大數據暨智慧企業研究中心

# 資料視覺化的科學基礎 與設計原則

資料視覺化可協助我們在進行後續的資料分析與探勘前，對手上的資料集先具備初步的認識，這也是為業界目前廣泛使用探索式資料分析（Exploratory Data Analysis，EDA）的一環，透過資料視覺化、基本統計與相關圖表等工具，我們可以瞭解資料的特徵、結構與數值差異，也可理解資料內各特徵之間的關聯性並找出重要的特徵，最後也可以檢查資料內是否有離群值或異常值，檢視資料是否有誤；在這個過程中，資料視覺化扮演重要角色，協助資料分析人員正確掌握資料現況。

進行資料視覺化之首要任務為視覺化的圖形必須正確且清楚地傳達用來建構圖形之資料，不能誤導閱讀者或是讓閱讀者產生混淆；在此同時，資料視覺化也應具備協調之顏色，並與文字搭配形成良好的視覺化構圖，以吸引閱讀者閱讀並協助閱聽者進行正確的解讀。倘若一個視覺化圖形包含著令人困惑的顏色、不協調的視覺元素以及混亂的圖像排列，閱讀者將會難以了解圖形呈現之意涵，影響閱讀者的閱讀動機以及對於資料的正確解釋。因此，資料視覺化可說是包含了資料科學、色彩學、版面設計…等專業之結合，透過選擇合適且精準顯示之圖形，輔以容易辨識以及令人感受協調之色彩，往往可以讓閱讀者更準確地理解資料的結構、數值與特徵，並可進行在不同群體間、不同時間或不同類別的比較。簡而言之，資料視覺化需要使用正確的資料、選擇合適的圖形，同時在圖表設計的過程，需要增加吸引閱讀者且讓閱讀者感興趣繼續閱讀的視覺效果，而這部分更是需要設計與藝術的成分才能完成。

本章第一節將說明色彩與心理之間的關係，第二節與第三節則採用業界廣泛利用的格式塔理論（Gestalt theory）與前注意處理（Pre-Attentive Processing）理論說明圖形與視覺之關係，透過格式塔理論與前注意處理理論描繪出資料視覺化的相關概念，並透過

上述概念提出幾項資料視覺化之通用規則於第四章節，進行常見問題的討論與提醒，也請讀者注意，本書專注於資料視覺化的過程，並不強調於資訊圖表（Infographic）的製作。資料視覺化與資訊圖表兩者之間差異在於，資料視覺化專注於一組特定資料集所產生的統計圖表或是圖像，以閱讀者容易理解且較客觀的方式表現事實，為閱讀者後續進行解釋和分析留下空間，目前已有許多的商業智慧（Business Intelligence）軟體、統計軟體與資料分析工具可以協助自動生成資料視覺化圖形。上述工具除了內建多樣式的圖表與地圖功能外，並會協助判斷現有資料類型與格式適合呈現之圖形；而資訊圖表則是嘗試將多個資料集整合，呈現出一個具備某一特定主題且比上述的資料視覺化圖形更全面的故事故事，資訊圖表通常帶有圖表製作者個人明顯的主觀觀點，經常出現在報章雜誌中的主題故事（Cover story）或深度報導內容以及小冊子、傳單…等傳統印刷品中，而近幾年盛行之社群媒體懶人包、網路文章、梗圖…也屬於資訊圖表的一類，資訊圖表主要希望達到對閱讀者的教育與資訊流通的意義，並期望引導閱讀者產生一個贊成或近似於圖表製作者論述之結論，帶有傳播理論中的「說服」目的，因此製作資訊圖表更需要具備色彩與版面設計…等設計層面的專業。

## 2.1 色彩與心理之關係

色彩是人類視覺的一部分，色彩與視覺的關係是相互的，色彩是光特定頻率的反射，我們的眼睛感知這些頻率、接收和處理後，向大腦傳遞色彩的感知訊息，我們的大腦透過對這些訊息的解釋和理解來認知色彩，因此，色彩和視覺是緊密相關的，它們共同構成了我們認知世界的方式；而色彩與心理也有密切的關係，色彩可以影響人的情緒和心理狀態。不同的色彩具有不同的心理效應，例如：紅色被認為能激發人的熱情，藍色則常常被認為是使人放鬆和冷靜的，因此色彩被廣泛用於商業和美學領域，例如廣告與商業設計。

西元 1666 年艾薩克·牛頓爵士（Sir Isaac Newton）以三稜鏡實驗發現彩色光譜，可謂是色彩學（Colour Theory）研究的先驅，也建立了後續研究的基礎。在本章節中，我們透過色環（Color Wheel）對色彩進行詳細的說明，色環是色彩學的一個工具，將可見光區域的顏色以圓形圖來表示，基礎是 12 種顏色，色輪上的顏色是屬於純色，也稱色相（Hue），包含基礎三原色（無法混合其他顏色來調配出來，是指紅色、黃色和藍色）、二次色（透過兩種不同原色等比例調配而成的顏色，如綠色、橙色、紫色）、三次色（混合相鄰的原色和二次色調配而成，如黃綠色、黃橘色、紅橘色、紅紫色、藍紫色、藍綠色），基礎的 12 色色環由約翰尼斯·伊登（Johannes Itten）所提出，因此又稱為伊登 12 色環。

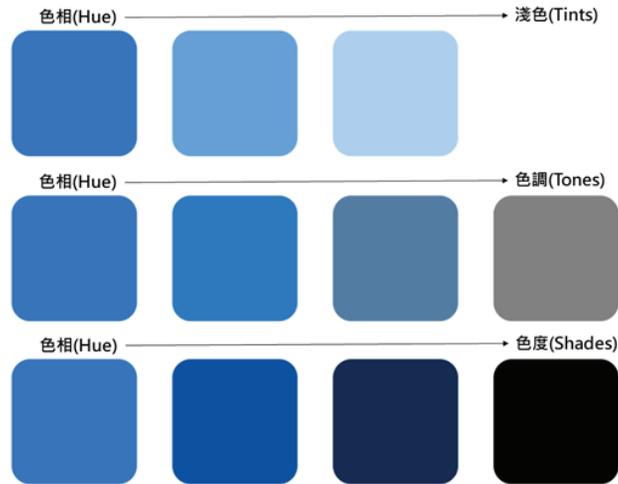


△ 圖 2.1 伊登 12 色環與色彩學名詞

資料來源：Canva 色環範例

在色環上還有一些常會聽到的名詞，除了三原色（Primary colors）外，還包含互補色（Complementary colors）是指彼此對立的顏色，又可稱為對比色，當兩者混合一起時會產生相互抵銷的狀態，也就是灰階色彩，例如白色或黑色，常見的互補色為紅色和綠色、黃色和紫色、藍色和橙色等。類比色（Analogous colors）是在色環上相鄰的顏色，如紅色和橙色、綠色和黃色、藍色和紫色等，類比色不像互補色如此強烈對比，顏色相較柔和，使用類比色可以讓一件作品在視覺上賞心悅目，提供較為純淨顏色的視覺感受。

上述的顏色並沒有混入白色或黑色這些中性色的純色，雖然黑色和白色不是色環上的顏色，但也因其混和顏色所產生的變化，所以有下列三個名詞來代表，一為淺色（Tints）是指加入白色的色相，會讓顏色變淡及變亮，如粉彩色系；二為色調（Tones）是指加入灰色的色相，會使顏色強度更加暗淡；三是色度（Shades）是加入了黑色色相，將使顏色變暗。



△ 圖 2.2 色相、淺色、色調與色度圖例

使用色環可以製作任何配色或組合，但有些配色會比其他配色更美觀。就像混合顏色來創造新的顏色一樣，顏色搭配得好就可以創造出令人賞心悅目的組合。顏色也會影響受眾的感覺，如紅色、橙色、黃色屬於暖色系，給人以溫暖、活力等感覺；藍色、綠色、紫色屬於冷色系，給人以冷靜、沉穩等感覺。不同的顏色在不同的文化和背景下有不同的意義，以下是一些常見的顏色代表的意義：

表 2.1 顏色代表的意義

顏色	正向意涵	負向意涵
紅色	熱情、愛、力量、勇氣	危險、警告、血腥、攻擊性、困難
橙色	活力、陽光、快樂、創造力、年輕、開放	挫折、不成熟、虛偽
黃色	希望、喜悅、活力、聰明、財富	警告、挫折、憤怒、飢餓、不穩定
綠色	平靜、安全、健康、成長、自然	忌妒、唯物主義
藍色	穩定、專業、信任、智慧、沉著	冷漠、不友善
紫色	靈性、尊貴、高貴、神秘	自省、傲慢、不切實際
粉紅色	浪漫、愛情、幸福、柔和、甜美、善良	軟弱、不成熟
灰色	中立、平衡、穩重、實際	乏味、壓抑
黑色	權威、力量、老練、魅力、威嚴	哀悼、壓迫、邪惡
白色	純潔、無邪、平和、寬敞、希望	空的、無聊

色彩是圖表設計中非常重要的一個元素，可以幫助圖表更加生動、鮮明，透過基礎的色彩學知識，可以協助選擇適當的色彩來表達數據和資訊，設計出更具吸引力、表達力和有效性的圖表。

## 2.2 圖表色彩的設計原則

選擇圖表主題相關的色彩時，我們除了考量使用者對色彩的認知和情感反應外，透過色彩的明亮度、飽和度、色相所形成的色彩對比，是首要的參考原則，色彩對比講求的是兩種或多種顏色之間的差異，在圖表一起呈現時會產生明顯的區別，可以幫助強調數據的重點、提高數據的可讀性，色彩對比的呈現可以從下列方式選擇：

1. **互補色相對比**：如紅色與綠色，藍色與橙色等，這些色相對比通常最為明顯，因為它們在色輪上相距較遠。
2. **明亮度對比**：色彩明亮度，也就是色彩的色值（Value），代表了色彩的明暗程度，當色彩明亮度降到最低時，顏色就變成了黑色，但調得太亮色彩又會偏向於白色，不同明亮度之間的對比。例如，黑色與白色之間的對比非常強烈，而灰色之間的對比則相對較弱。
3. **飽和度對比**：色彩飽和度（Saturation）指的是色彩強度，色彩飽和度愈高表示顏色越鮮艷，色彩飽和度越低顏色則會變得黯淡，例如，鮮紅色與灰紅色之間的對比。在圖表設計中，如果色彩強度過高或過低都不利於數據傳達，需根據資料的特點和設計風格來選擇適當的色彩飽和度。
4. **暖色系與冷色系對比**：冷色與暖色之間的對比。冷色如藍色、綠色，暖色如紅色、橙色。這種對比會影響到整體的色調感覺。

其次是配色的方式，通常使用幾種類型，一為單色配色法：通常採取使用一個色相的淺色、色調和色度作搭配變化，單一色調配色初步聽起很單調無聊，但比起使用多種顏色使得閱讀者眼花撩亂，失去閱讀重心而言，單一色相的細微顏色變化有助於簡化設計，這種配色方式使用單一顏色的變化來代表不同的資料數列或類別，是一個簡單而乾淨的配色方法，很適合用於比較資料列之間的差異。

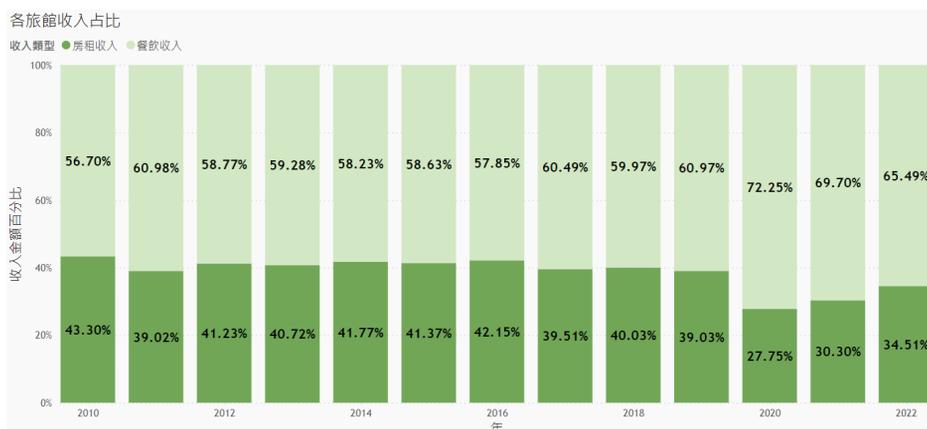


圖 2.3 單色配色法範例

二為類比色（相似色）配色法：使用色環上相鄰的顏色，創造出和諧感，相鄰顏色的搭配有著類似於單色配色的和諧美感，但類別的對比顯示上較不如互補色的強烈。

消費者最喜愛的冰淇淋口味調查？

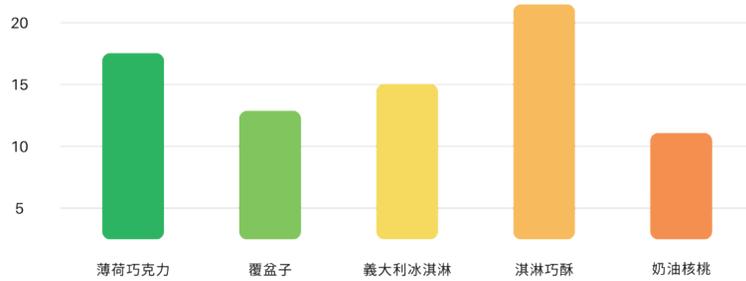


圖 2.4 類比色配色法範例（使用 Canva 繪製）

三為互補色配色法，使用色輪上彼此相對的顏色，讓受眾一眼就看到強烈視覺效果的圖表，更容易吸引注意力，建議用於報表須特別強調的地方，但如果使用過多的顏色時，反而會沒有重點讓人感到混亂，本書建議可採用色彩 60-30-10 原則，使用三種主要顏色來做為圖表搭配，色彩 60-30-10 原則是指一種顏色配色方案，其中使用三種顏色來創建平衡和視覺吸引力的設計，60-30-10 指的是每種顏色在設計中所占的百分比。具體來說，60-30-10 原則建議使用三種顏色：主色（60%）：該顏色應該在圖表設計中占主導地位，通常用於背景或大面積的元素；輔助色（30%）：這種顏色用於設計的輔助，例如用於標題、標籤或是需要注意的文字；強調色（10%）：這種顏色用以吸引閱讀者目光注意，例如用於小圖示、圖標、連結或行動呼籲按鈕。使用色彩 60-30-10 的配色方案可以讓圖表看起來更加協調和諧，且不會讓人感到過於單調。

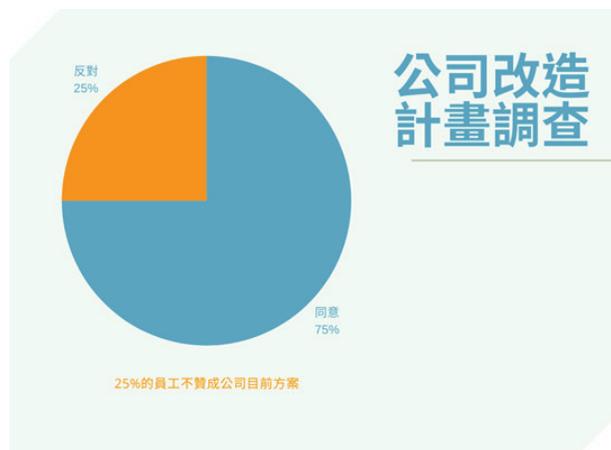


圖 2.5 色彩 60-30-10 原則範例，60% 主色為淺藍白底色；30% 輔助色為藍色；10% 強調色為橘色（使用 Canva 繪製）

# 關聯性之視覺化 ( Relationship )

## 6.1 關聯性視覺化圖表特色及使用之資料格式

在進行數據分析時，探索資料間的關聯是很重要的前置步驟之一，當對資料的分佈或特性有一定了解時，有助於進一步預測資料的趨勢走向，進一步對相關主題做出有利的判斷或決策參考。

相關性主要是用於展現兩（多）項變數間的連動關係，一般而言平面視覺圖表主要是用來展現兩（多）個變數間的數值變化關係，常被用來展示兩變數間是否進一步存在因果關係（例如一個變數的變化如何影響另一個變數變化）。

常見的關聯性視覺化圖表包括散佈圖（Scatter plot）、折線 + 柱狀圖（Line + Column）、連接散佈圖（Connected scatterplot）、泡泡圖（Bubble chart）、XY 熱圖等分述於本章。而常見的圖表應用例如：年齡與薪資的關係、所得收入與居住房價的關係、分析市場需求和產品利潤率之間的關係…等，可協助讓閱圖者掌握現況以及預測趨勢走向。

## 6.2 圖形介紹

### 6.2.1 散佈圖 (Scatter plot)

- 圖表名稱：散佈圖
- 資料格式：二維（以上）－ 數值 vs 數值
- 元件展示方式：點、水平 / 垂直、顏色
- 用途：關聯性（relationship）（兩數值變數間）
- 特點：兩變數（以上）間其數值的變化與分佈關係。
- 範例意涵：某動物頭部大小（數值）和體重（數值）之間的關係，用以觀測是否頭部愈大體重愈重。（圖 6.1）

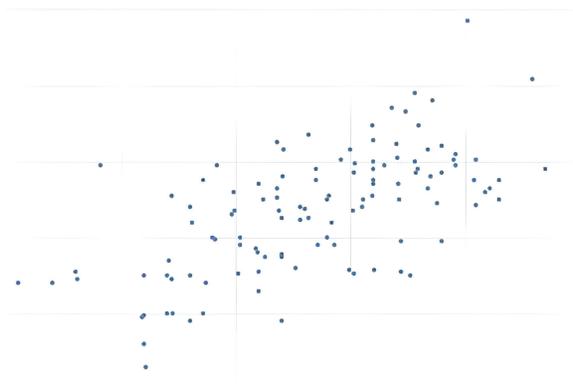


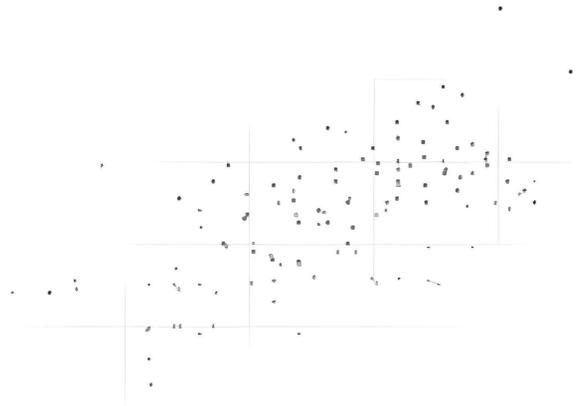
圖 6.1 散佈圖 (Scatter plot) 範例 1

散佈圖 (Scatter plot) 是最廣泛使用的圖表之一，需要兩個數值變數的配對資料，例如 x 軸值和 y 軸值。其通常是被用來觀察兩數值資料彼此間的連動關係，以利圖表閱讀者理解兩數值變項之關聯 (relationship) 性。一般來說它很常被用於探索變數之間的線性關係，而散佈圖可以顯示出包括正向關係、負向關係和沒有關係。當有明顯之關聯存在，可再進一步加上輔助線以展示其可能存在之關係。當兩數值呈現正向關係，散佈圖上的點會形成一條往右上的直線；如果呈現負向關係，則散佈圖上的點會形成一條傾向右下的直線；如果呈現無關係，則散佈圖上的點會呈現出隨機分散的形狀。

散佈圖中的每個點代表一組個別之資料，所有資料點組合起來亦可觀測出數據的趨勢；同時倘若數據中含有離群值或極端值時，製圖者可用不同的標記或顏色將其標識出來，以增進對數據的理解。

此外亦可再引入其它類別變項，綜合展示可能存在的關係；如圖 6.2 展示出繪圖者使用顏色將數據區分成“公”和“母”兩組，閱圖者可進一步觀察是否不同性別（類別變項），此二數值變項間各有何的關聯或影響。

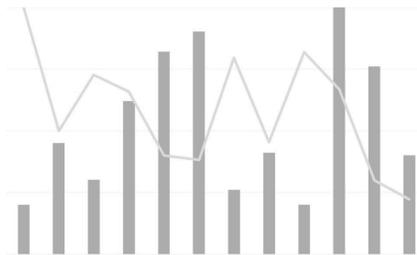
簡言之，散佈圖是一種簡單而有效的圖形工具，可協助人們探索數值變數間的關係、視覺化個別資料點、描繪整體趨勢以及識別離群值。



△ 圖 6.2 散佈圖 (Scatter plot) 範例 2

## 📊 6.2.2 折線 + 柱狀圖 (Line + Column)

- 圖表名稱：折線 + 柱狀圖
- 資料格式：三維 - 類別 (順序) vs 數值 vs 數值
- 用途：比較 (compare) (兩數值變數間)、關聯 (relationship) (數值的變化關係)。
- 元件展示方式：點、線條、長度、水平 / 垂直、顏色
- 特點：展現兩種不同數值在有序類別上的變化關係。
- 範例意涵：如不同大城市 (有序類別)，總人口數 (數值) 與疫苗施打率 (數值) 間的關係。(圖 6.3)

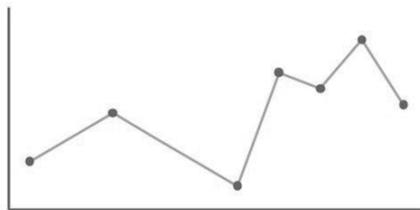


△ 圖 6.3 折線 + 柱狀圖 (Line + Column chart) 範例

折線 + 柱狀圖顧名思義是由折線圖和柱狀圖所組成，兩個圖形以一個有序的類別變項做連結。藉由此有序變項可同步觀察兩變項之數值的變化情形，方便閱圖者理解與推想變項間的關係。例如我們可從圖 6.3 中觀察出，最左側的城市人口數最多，但疫苗施打率並不是最高的。而針對有序類別，折線圖可以輕易展示出數值的趨勢走向；倘若該類別變數沒有特定排列方式，則建議選定折線或長條圖擇一數值做升 / 降冪排列，以避免資訊太過發散，以致不利觀察出數值間的關聯。

### 6.2.3 連接散佈圖 (Connected scatterplot)

- 圖表名稱：連接散佈圖
- 資料格式：二維 - 類別（有序）vs 數值
- 元件展示方式：點、線條、長度、水平 / 垂直、顏色
- 用途：比較（compare）（類別與數值變數間）、關聯（relationship）（類別與數值變數的變化關係）
- 特點：從圖中可直接觀察數值數列隨時序（順序）類別上升或下跌的趨勢走向。
- 範例意涵：不同年度（順序類別）的就業比率（數量）。（圖 6.4）



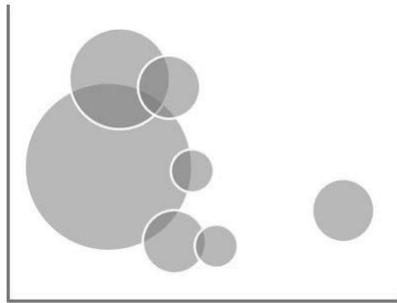
^ 圖 6.4 連接散佈圖 (Connected scatterplot) 範例

連接散佈圖 (Connected scatterplot) 顧名思義是一種複合圖表，將散佈圖和連接線結合在一起，與折線圖 (Line chart) 也有相同意涵；折線圖的概念就是將資料點以線條型式連接起來，如此易於觀察出數值資料點隨有序類別變數的散佈和整體趨勢走向、易於觀察出離群異常值，以及比較多個數值變項數列等特色。折線圖表清晰易閱讀，數值資料透過線條連接，特別適用在時間序列的資料中，顯示連續期間的值趨勢走向和分佈的模式。

使用連接散佈圖 (Connected scatterplot) 和使用折線圖來展示資料時，都需留意時間顆粒度的問題，同樣的資料如果採用日、月、季或年等不同顆粒度視為橫軸刻度時，可能會有某些週期起伏被忽略了而引發資料解讀的不夠精準。（請參考圖 5-7、圖 5-8）

## 6.2.4 泡泡圖 (Bubble chart)

- 圖表名稱：泡泡圖
- 資料格式：三維 - 數值 vs 數值 vs 數值 (類別 vs 類別 vs 數值)
- 元件展示方式：點、面積、水平 / 垂直、顏色
- 用途：關聯 (relationship) (兩數值變數的分佈關係)、比較 (compare) (加上第三數值的大小)
- 特點：從平面圖中直接觀察第三軸數值的大小 (泡泡大小)。
- 範例意涵：某地區人口的身高 (數量) 和體重 (數量) 的分佈與人數的關係 (數量, 泡泡)。(圖 6.5)



△ 圖 6.5 泡泡圖 (Bubble chart) 範例

泡泡圖主要用來展示兩個及以上的連續變數之間的關係，相較於散佈圖，泡泡圖在視覺上更增添趣味感；泡泡之大小為原圖中加強展現第三變量的數值，增加原二維圖形所能呈現出來的資訊量，故常見於各種視覺化圖表中。

在實務上，除了泡泡大小可增加展現出數值變項的訊息外，亦可將泡泡塗上色彩以再加突顯出第四個類別型維度的資訊。

在準備泡泡圖數據時，還應考慮去除任何不必要的資料點，例如缺失數據或極端值，如此有助於更有效地運用泡泡圖來探索數據潛在的關係和洞察。另，常見將不同時間點依序繪製平面泡泡圖，最後以動畫的形式依時間軸做播放，可協助讀者對各泡泡依不同時點的大小變化有更直覺的觀察。此外，泡泡的大小可能會導致數據重疊，故泡泡圖可同步採用互動式顆粒大小的縮放選擇，讓閱圖者自行截切所欲分析數據的範圍。

## 6.2.5 XY 熱圖 (XY heatmap)

- 圖表名稱：XY 熱圖
- 資料格式：數值 vs 數值
- 用途：關聯 (relationship)、比較 (compare)
- 元件展示方式：水平 / 垂直、面積、密度、顏色
- 特點：展示兩個連續變數間強度、方向、密度關係，不適合展示數據的細微差異。
- 範例意涵：氣象數據中，溫度 (數量) 和降雨量 (數量) 之間的關係。(圖 6.6)



圖 6.6 XY 熱圖 (XY heatmap) 範例

XY 熱圖用於顯示兩個連續變數之間關係的二維散布之延伸，相較於散佈圖，XY 熱圖在視覺上更增添豐富感；透過顏色或大小的使用來表示變數之間的強度、方向或密度，例如強度強 / 密度高的數值會採用較深的顏色，而強度弱 / 密度低則用淺色表示。

在延伸應用下，XY 熱圖可以額外增加類別變項，以區分不同類型的資料點，例如可以引進性別變項，使用不同的色系區隔之。另外，依據不同的資料和分析目的，製圖者可能在圖中加入特定權重變項，以控制加權不同資料點在圖表中的強度（以色系的深淺強度表示）。

## 6.3 關聯性視覺化圖表之優缺點比較

在 6.2 節中我們介紹了用來表達關聯性的五個圖形，匯整如表 6.1 所列。

表 6.1 關聯性圖形彙整

圖表名稱	資料格式	元件展示方式	用途	特點
散佈圖	二維 數值 vs 數值	點、水平 / 垂直、顏色	關聯性 (relationship) (兩 數值變數間)	變數間其數值的 變化與分佈關係
折線 + 柱狀圖	三維 類別 (順序類別) vs 數值 vs 數值	點、線條、長 度、水平 / 垂 直、顏色	比較 (compare) (兩 數值變數間)、關聯 (數值的變化關係)	展現兩種不同數 值在順序類別上 的變化關係
連接散佈圖	二維 類別 (有序) vs 數值	點、線條、長 度、水平 / 垂 直、顏色	比較 (類別與數值變 數間)、關聯 (類別 與數值變數的變化 關係)	觀察數值隨時序 (順序) 類別上 升或下跌的趨勢 走向
泡泡圖	三維 數值 vs 數值 vs 數值 或 類別 vs 類別 vs 數值	點、面積、 水平 / 垂直、 顏色	關聯 (兩數值變數 的分佈關係)、比較 (加上第三數值的大 小)	從平面圖中可直 接觀察第三軸數 值的大小 (泡泡 大小)
XY 熱圖	二維 數值 vs 數值	水平 / 垂直、 面積、密度、 顏色	關聯、比較	展示兩個連續 變數間強度、方 向、密度關係

這五張圖基本上最大的共通點就是依二維軸去展現出資料的數值分佈資訊，從單純的資料描點如散佈圖以了解二維資料是否呈現出因果變化等相關情形；到將描點數值做連線以了解其可能的隨順序資料的變化趨勢（如：折線 + 柱狀圖和連接散佈圖）。除此之外，亦可擴展呈現出第三維以上的資訊；如折線 + 柱狀圖多加圖形來揭示另一個數值變量、而連接散佈圖再加不同顏色的線條來展示不同類型的數值趨勢。最後泡泡圖和熱點圖，進一步可使用顏色、大小、透明度、深淺等各種不同變化來擴增到三維四維資訊的揭露。惟以上這些表達的形式，人類可能無法非常精準的做絕對大小的判斷，故原則上用來揭示方向性的指引使用。一般來說，當二維的圖形被擴增用來表達三維以上的訊息時，多少會讓圖形的解讀更加複雜化，讀者在解讀時難免有可能會出錯或不易觀察出資料的樣貌，是故製圖者首要掌握的是繪製圖形時最主要想傳達的訊息是什麼，最終的成品是否有達到該效果；行有餘力時才考量能否加上其它變項訊息，並需要再次確

認新變項訊息在加上去之後，是否會妨礙原本要傳達的主要訊息，以便挑選出最適合的圖形。

不同的圖表可用於不同的數據探索和分析任務，以展現出不同的訊息；在選擇圖表時，需綜合考量數據的特性、製圖目標和閱圖者的理解情形等始能達到預期之成效。

本章所介紹的五張圖的優缺點比較表整理於表 6.2。

表 6.2 關聯性圖形之優缺點比較

圖表名稱	優點	缺點
散佈圖	簡潔清晰	僅涵蓋兩個維度的資料，所呈現的資料範圍有限
折線 + 柱狀圖	1. 簡潔清晰 2. 可展現三維以上的資料	1. 需留意折線和柱狀圖軸線刻度是否為不同的值域範圍 2. 所呈現的資料範圍有限
連接散佈圖	1. 簡潔清晰 2. 可展現三維以上的資料	相同時序資料點，如日、月、季或年等採不同橫軸顆粒度時，可能有週期起伏被忽略
泡泡圖	1. 圖形親和有趣 2. 可展現出三維以上的資料	1. 視覺上無法精準理解泡泡所代表的絕對大小 2. 泡泡的重疊造成判斷不夠精準 3. 不適合展示數據的細微差異
XY 熱圖	1. 圖形親和有趣 2. 可同時表現出三維以上的資料	1. 視覺上熱圖的密度無法被精準理解其值的絕對大小 2. 不適合展示數據的細微差異

## 6.4 實作與解釋

打開 Ch06\_start.pbix，選擇【另存新檔】將檔案另存在指定目錄下，可將檔案命名成“Ch06\_prac.pbix”。

本節主要是示範相關圖表之製作，實際商務應用需依讀者自有之資料，針對所欲探索之數值搭配到相關資料軸線或其它設定，以達查看資料關聯性之效。

另，各子節中的圖形都可依循下列步驟做視覺效果的調整：點按【圖】，依序在【視覺效果】中，將所欲調整的項目嘗試做不同設定，以完成更精美的製圖。

## 6.4.1 散佈圖 (Scatter plot) 的製作

- 製圖目的：展示銷售大分類銷售金額和銷售次數間的散佈關係

- STEP01 請在左上方點選【報表檢視】，接著於下方頁籤按【第 1 頁】重新命名為【散佈圖】。
- STEP02 在【視覺效果】視窗中的【組建視覺效果】先按下【散佈圖】。
- STEP03 在製圖區將圖表的區域拉大成適中的大小。
- STEP04 在【視覺效果】視窗中的【組建視覺效果】，將【Transaction】表中的【銷售金額】拖曳到【X 軸】，再將【Transaction】表中的【銷售金額】拖曳到【Y 軸】並同時下拉改成【計數】。
- STEP05 將【Product】表中的【大分類】拖曳到【值】，即完成製圖。



圖 6.7

## 6.4.2 折線 + 柱狀圖 (Line + Column chart) 的製作

- 製圖目的：展示大分類銷售金額 (柱狀) 和銷售次數 (折線) 的分佈

STEP01 請在左上方點【報表檢視】，接著於下方頁籤按【+】新增頁籤並重新命名為【折線 + 柱狀圖】。

STEP02 在【視覺效果】視窗中的【組建視覺效果】先按下【折線 + 柱狀圖】。

STEP03 在製圖區將圖表的區域拉大成適中的大小。

STEP04 在【視覺效果】視窗中的【組建視覺效果】，將【Product】表中的【大分類】拖曳到【X 軸】，再將【Transaction】表中的【銷售金額】拖曳到【資料行 y 軸】，再將【Transaction】表中的【銷售金額】拖曳到【線條 y 軸】並同時下拉改成【計數】，即完成製圖。

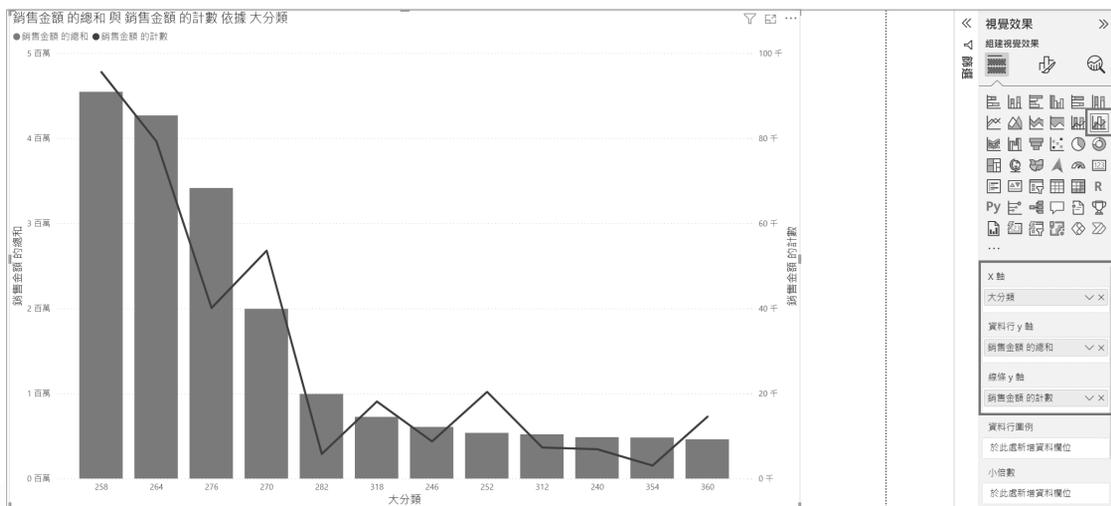


圖 6.8

## 6.4.3 連接散佈圖 (Connected scatterplot) 之製作

- 製圖目的：依銷售日期顯示每天的總銷售金額

STEP01 請在左上方點【報表檢視】，接著於下方頁籤按【+】新增頁籤並重新命名為【連接散佈圖】。

STEP02 在【視覺效果】視窗中的【組建視覺效果】先按下【折線圖】。

STEP03 在製圖區將圖表的區域拉大成適中的大小。

# 部分和整體關係之視覺化

## ( Part-to-Whole )

### 11.1

#### 部分和整體關係視覺化之圖表特色及使用之資料格式

部分和整體關係之視覺化 (Part-to-Whole) 的目的是要能顯示出一個整體如何被拆解成不同組成，展示部分與整體的關係，或是解釋整體的某些部分。如果只是想瞭解個別成分的大小，建議改用比較量大小 (Magnitude) 的圖表。部分和整體關係之視覺化較常使用的圖形有堆疊圖 (Stacked chart)、比例堆疊條形圖 (Marimekko)、圓餅圖 (Pie chart)、甜甜圈圖 (Donut chart)、樹狀圖 (Tree map)、網格 (Grid plot)、…等 (Smith, 2022)。

任何時候，你嘗試說明正在談論、報告的事物，其累加起來的總值是 100%，這就是在講述一個關於部分和整體關係的故事。當報告需要整體的某些部分時，有可能是你要描述某個群體的比例組成結構時，也有可能是想要說明研究對象的敘述性特徵。以上這些都是在談論關於整體的其中一部分，也就是關於部分和整體關係的故事。部分和整體關係一類的描述可以是如下的範例：

- 手機中剩餘電量的百分比。
- 休旅車的銷售市場主要是 A、B、C 三個品牌占了大宗。
- 我們的顧客群主要是女性，大約是男性的兩倍。

- 我們這間分店的本月營業額已經達到了 15% 的年度目標。
- 學生構成為 75% 的免費或減價午餐和 25% 的付費午餐。
- 以下是本研究受試者的收入分佈情形。
- 此圖顯示了來自台灣各縣市的使用者百分比。
- 我們的大多數訪客年齡在 10 歲以下和 56 歲以上。
- 市場數據顯示我們擁有三個主要的客戶概況。

當我們要製作的圖表主題中包括以下如：份額、占比、總數百分比、占百分比多少等詞彙時，需要製作部分和整體關係的對比關係圖表來呈現較為合宜。

## 11.2 圖形介紹

### 11.2.1 堆疊圖 (Stacked chart/graph)

堆疊圖是最常被使用來呈現部分與整體關係的圖形。堆疊長條 / 橫條圖 (Stacked column/bar chart) 與一般的長條圖或是橫條圖不同，它是多個資料集相互疊加，以顯示較大類別如何劃分為較小類別及其與總量的關係。以分成幾個部分的條狀圖形，分別表示一些變數在整體中所占之數值或是比例，常用來顯示簡單的總數明細。這種圖形主要的優點是可以替代圓餅圖。因為堆疊圖能夠明確顯示主次類別的占比，較圓餅圖可處理更多類別，可橫或直堆疊的方式來呈現。

基本上，堆疊圖可以分為兩種類型：

1. **簡單堆疊直條 / 橫條圖 (Simple stacked column/bar chart)**：簡單堆疊條形圖將類別的每個值放在前一個值之後。顯示條形的總值是所有類別值加在一起。非常適合比較每個類別條的總量。
2. **100% 堆疊長條 / 橫條圖 (100% Stack column/bar chart)**：100% 堆疊條形圖藉由繪製每個值占每個組中總量的百分比來顯示整體百分比。這樣可以更輕鬆地查看每個類別中數量之間的相對差異。

依照堆疊的方向（垂直或水準）的不同，以及呈現內容（數值或比例）的差異，整理成如表 11.1 所示的四種形式。

表 11.1 堆疊圖的分類

堆疊方向	呈現內容	
	數值	比例
垂直堆疊	簡單堆疊直條圖	100% 堆疊直條圖
水準堆疊	簡單堆疊橫條圖	100% 堆疊橫條圖

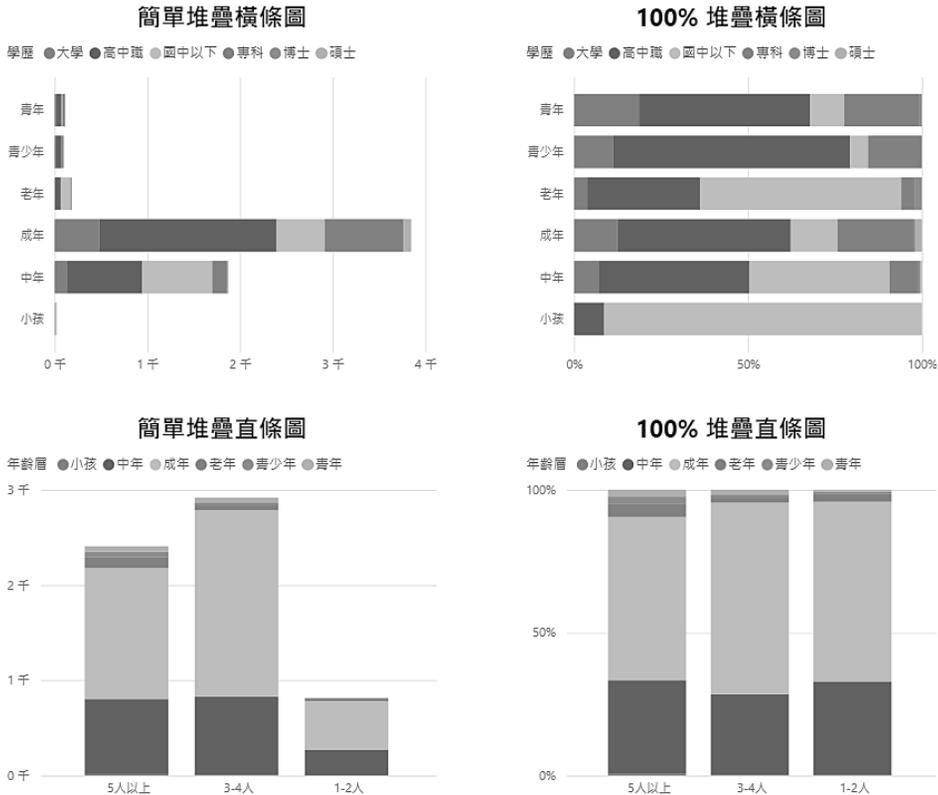


圖 11.1 四種不同的堆疊圖

圖 11.1 是使用本書範例資料繪製的四種不同的堆疊圖。簡單堆疊橫條圖與 100% 堆疊橫條圖使用了顧客學歷與顧客年齡層的欄位來繪製。簡單堆疊直條圖與 100% 堆疊直條圖則是使用了顧客家庭人口與顧客年齡層的欄位來繪製。在圖中可以看到包含太多類別，或是一次顯示太多堆疊長條時，較難看出差異與變化。堆疊圖的主要缺點是，當每個條形圖的分類越多，它們就越難以閱讀。此外，將每個部分與另一個部分進行比較很困難，因為它們沒有在共同的基線上對齊。這是一種能突顯出部分和整體關係的簡單方式，但如果組成部分過多會造成理解困難。

## 11.2.2 比例堆疊條形圖 (Marimekko)

Marimekko 圖，也稱為馬賽克圖 (Mosaic)、鑲嵌圖和比例堆疊條形圖，或者簡稱為 Mekko。適合用來同時呈現資料的大小與占比，只要資料不是太複雜。Marimekko 是一家 1951 年成立於芬蘭的時尚家居公司的名字，它著名的是充滿條紋、格子以及色彩強烈的抽象花卉圖案的產品。據說，Marimekko 圖的靈感來自於前美國第一夫人賈桂琳·甘迺迪 (Jacqueline Kennedy)。在 1960 年美國總統競選期間的造型，身穿 Marimekko 的產品，加上背景的鮮豔色彩方塊 (如圖 11.2 所示)，就讓諮詢顧問們將此圖取名為 Marimekko (Smith, 2017)。



圖 11.2 賈桂琳·甘迺迪與馬賽克圖

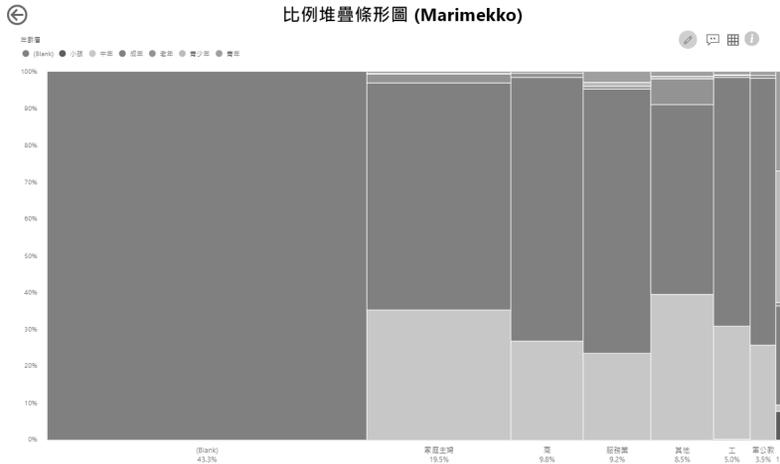
資料來源：Smith, A. (2017). *How to apply Marimekko to data*. *Financial Times*.  
<https://www.ft.com/content/3ee98782-9149-11e7-a9e6-11d2f0ebb7f0>

Marimekko 圖是一種使用不同寬度的堆疊橫 (或長) 條圖來顯示分類資料的二維堆疊圖形的表示方式。此類圖是表示分類樣本資料的理想選擇。在 Marimekko 圖中，X 軸和 Y 軸都是帶有百分比刻度的變數，它決定了每個分類的寬度和高度。因此，Marimekko 圖呈現了一種近似於雙變數 100% 堆疊條形圖的樣貌。因此，我們可以藉由 Marimekko 圖同時檢視 X 軸和 Y 軸兩個類別變數之間的關係。

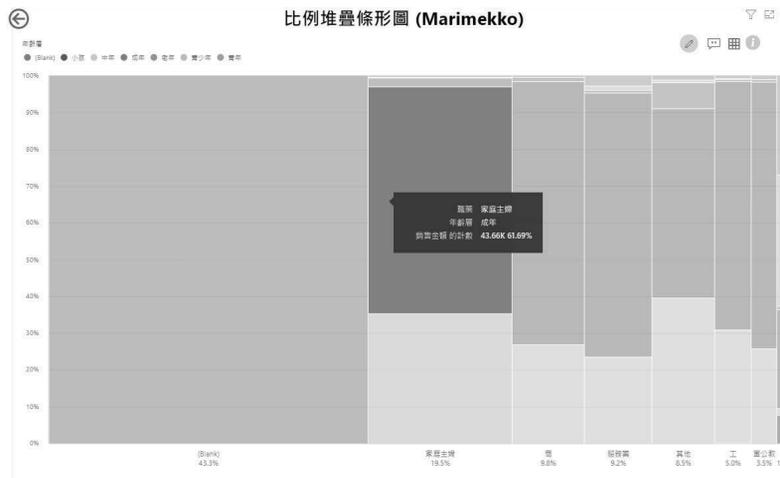
Marimekko 圖的主要缺點是當 X 軸和 Y 軸的類別數過多時，展現的圖形將令人難以閱讀與理解。此外，Marimekko 圖難以準確地比較每個變數中的每一個類別，因為每一個類別的 X 軸和 Y 軸的基準線不同。因此，Marimekko 圖比較適合提供更全面的資料總體概述 (general overview)。

圖 11.3 及圖 11.4 是使用本書的範例資料繪製的 Marimekko 圖。使用的欄位是職業與年齡層。由於本書所附的範例資料是真實的零售業交易資料，可以在資料中看到一個有趣的現象。在交易的對象中，有大約 43.3% 的顧客，我們不知道他 / 她的個人屬性。所以在進行資料分析時，就會有一定的資訊落差。圖 11.3 可以看到整體顧客資料所呈現的職業與年齡層的分佈情形。職業比例最高的是家庭主婦 (19.5%)，依序則是商 (9.8%)、服務業 (9.2%)、...等。由於 X 軸和 Y 軸的基準線不同，所以我們難以直

接比較【職業：家庭主婦】和【職業：商】之中的【年齡層：成年】，何者比較多。現在，透過 Power BI 的圖形，我們可以直接將滑鼠移到指定的區塊上面，就可以獲得絕對的資料，直接理解並進行比較（如圖 11.4）。



^ 圖 11.3 使用範例資料集繪製的【顧客職業】與【年齡層】比例堆疊條形圖



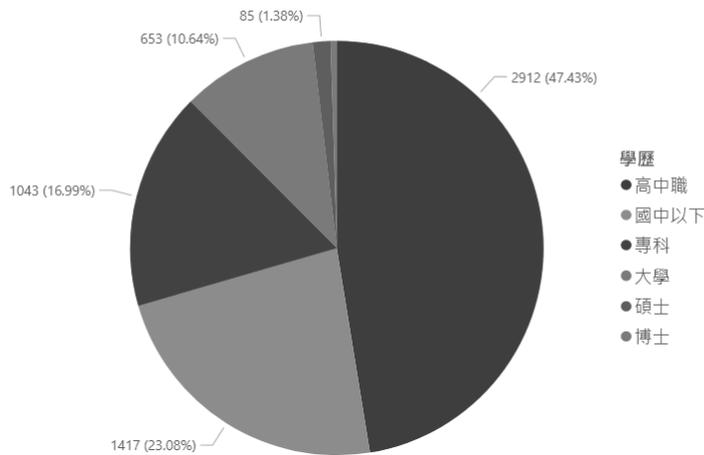
^ 圖 11.4 使用範例資料集繪製的【顧客職業】與【年齡層】比例堆疊條形圖

### 11.2.3 圓餅圖 (Pie chart)

威廉·普萊費爾 (William Playfair) 是一位活躍於 18.19 世紀時的蘇格蘭工程師和政治經濟學家，是圖形統計方法的創始人，也被廣泛認為是圓餅圖的發明者。圓餅圖將一個圓切成幾個部分，每部分都表示整體總值不同的比例，經常用來呈現部分和整體之間的關係。圓餅圖使用了數個扇形而組合成的一個圓形圖表，主要是用於描述數量、頻率或百分比之間的相對關係。這個圖表非常適合在我們不是直接對每一個扇形面積進行比較，而是想要特別突顯某一個類別在整體之中所占的比例時使用。完整的一個圓就代

表了所有資料的總和，應該要等於 100%。圓餅圖的呈現方式十分直觀，相當適合讓讀者快速瞭解資料的比例分佈（如圖 11.5）。圖 11.5 的內容在於呈現範例資料集之中顧客學歷的分佈。由於學歷的類別較為簡單，僅有六種類別，因此使用圓餅圖來呈現這樣的資料相當適合。

顧客學歷的圓餅圖

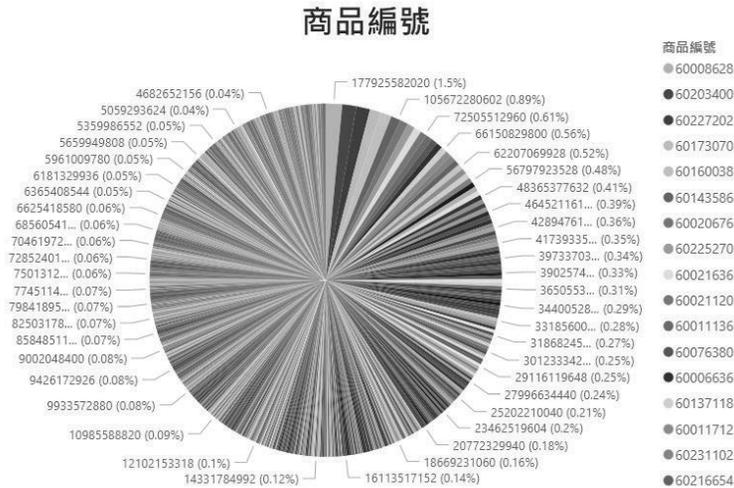


△ 圖 11.5 使用範例資料集繪製的【顧客學歷】圓餅圖

Chen（2015）對於圓餅圖的使用，有以下幾個建議：

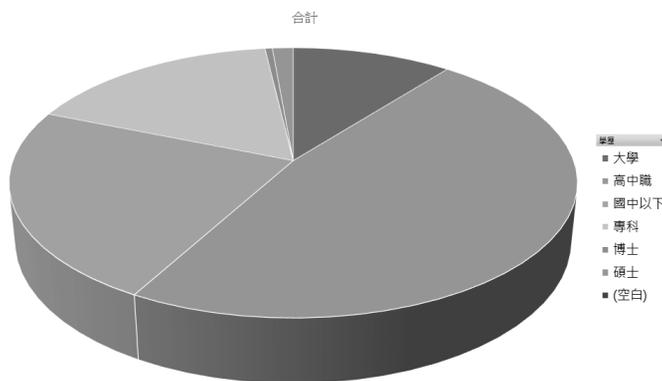
1. 圖形必須包括所有組成部份；組合也必須是 100%。
2. 從 12 點鐘方向開始畫圖。
3. 組成項目請勿過多。
4. 避免使用分裂式圓形圖。
5. 請勿使用 3D 圓形圖。

因此，我們需要注意的是，圓餅圖並不適合去精確比較出不同類別組成的大小。直觀上無法直接評估圓餅中扇形的各部分面積（如圖 11.6）。尤其是當分切的數量較多時，更是難以辨識與估量，不適合在圓餅圖組之間進行準確的比較。因為隨著顯示值的數量增加，每個片段 / 切片的大小會變小。因此，圓餅圖不適用於具有許多類別的大型資料集。圖 11.6 是使用本書範例資料集的顧客交易資料進行購買商品的視覺化。將購買商品的種類（意即商品編號）以圓餅圖呈現後，因為類別過多，反而導致此種圖表無法呈現資料中的訊息。若是我們需要呈現多個類別之間的比例或是數量的關係時，可能採用如長條圖或圓點圖等圖表會顯得較為合適。



△ 圖 11.6 使用範例資料集繪製的錯誤【商品編號】圓餅圖示意

此外，圓餅圖尤其不適合以 3D 的方式來呈現或使用。因為這樣的視覺效果，對於不同類別項目的扇形面積或是比例呈現，會有被扭曲或是誤解的可能（如圖 11.7）。在圖 11.7 中的 3D 呈現效果，我們可以看到，國中以下（灰色）和專科（黃色）的扇形面積好像相差不大，而專科（黃色）和大學（藍色）的扇形面積好像也是非常相近。但是它們真實的比率卻是 23.08%：16.99%：10.54%。



△ 圖 11.7 使用範例資料集繪製的【顧客學歷】3D 圓餅圖

許多圖表都不適合使用 3D 形式來呈現，尤其是圓形圖。3D 圓形圖較靠近受眾的部份會看起來面積較大，較遠的部份則會顯小，容易扭曲對資料的理解。賈伯斯（Steve Jobs）在 2008 年 MacWorld 的演說中介紹各廠牌的占有率時，畫面上以 3D 圓形圖來呈現，而無巧不巧，Apple 的 19.5% 就被安排在最靠近聽眾的位置，而 21.2% 的【其他】則在最遠的位置。雖然這完全符合前面介紹的圓形圖排列建議，但 3D 卻造成了 19.5% 比 21.2% 大的神奇效果（Chen, 2015）（如圖 11.8 所示）。所以圓餅圖的 3D 使用要很小心，以免被誤解。