
簡介

為什麼要進行資料視覺化？

在這本書中，你會透過結合了設計原則和步驟教學的章節，來學習如何製作真實而有意義的資料視覺化，使這些以資訊為基礎的分析和論述，更具洞察力和吸引力。正如句子若有證據和來源注釋會更有說服力一樣，當資料驅動的文章搭配上適當的表格、圖表或地圖時，也變得更加強大。文字告訴我們故事，但視覺化將定量、關係或空間上的模式轉換為影像，向我們展示資料故事。經過精心設計的視覺化，會將我們的注意力吸引到資料中最重要的部分，而這是很難單靠文字來溝通的。

我們的書中使用了許多免費、易學的數位工具來製作資料視覺化。我們廣義地將它定義為「圖表」，它將資料編碼為影像，並加上空間維度成為地圖。雖然表格無法以相同的方式說明資料，但我們也將它們收錄在書中，因為在指導新學習者的決策過程當中，通常結論都是製作這三種形式的其中之一。此外在這個數位時代，我們將資料視覺化定義為：只要「修改通常儲存在資料文件中的基礎資訊，就能輕鬆地重新利用」的影像。這與通常設計成單次使用的資訊圖（*infographics*）是不同的¹。

身為教育者，我們藉由本書來介紹關鍵概念，並為新學習者提供步驟教學。你可以用它來自學，或使用這本書來教別人。此外，相較於許多只將焦點放在一種工具上的技術書籍，本書推薦了 20 多種免費且易於使用的視覺化工具中供你選擇。最後，雖然有些書籍將焦點放在只能在紙本或 PDF 文件上發佈的靜態視覺化內容，但本書示範了如何設計互動式

1 請注意，其他資料視覺化書籍使用這些術語的方法可能不同。例如，在 Alberto Cairo 的《*How Charts Lie: Getting Smarter About Visual Information*》（W.W. Norton & Company, 2019）一書中（<https://oreil.ly/wXcBX>, p. 23），所有的視覺化都被定義為「圖表」（charts）。

表格、圖表和地圖，並將其嵌入到網路上。互動式的視覺化能引發受眾與資料進行互動、探索他們感興趣的模式、下載需要的檔案，並在社群媒體上輕鬆分享你的成果。

在過去的十年中，資料視覺化已在網路上廣泛散播。在現今的網頁瀏覽器中，我們遇到的數位圖表和地圖比以往的印刷時代多。但是快速成長也帶來了嚴重的問題。現在「資訊時代」與「虛假時代」交疊了。幾乎每個人都可以在網上發佈資訊，那麼要如何才能明智地判斷要信任誰？當你看到關於分裂性的政策議題（例如社會不平等或氣候變遷）等相互矛盾的資料故事時，你該相信哪一個？在下一單元中，我們要深入討論這個棘手的話題，探討哪種證據具有說服力及其原因。我們將分享一個關於資料視覺化的暗黑小秘密：它照亮了我們追求真理的道路，但同時也讓我們掌握了欺騙和說謊的能力。

你能相信什麼？

首先，你如何知道是否能相信本書的作者？我們有可能正在欺騙你嗎？你如何確定哪些資訊是真實的？讓我們從一個簡單的單句陳述開始。

範例 I-1

自 1970 年代以降，美國的經濟不平等現象急劇上升。

你是否相信此一聲明？也許你過去從未用這樣的方式思考過這個主題（若是如此，有很多資訊都可以為你提供幫助）。你的回應可能取決於此陳述與你先前的信念一致或者相反。或者，也許你被教導要對缺乏支援證據的主張抱持懷疑態度（如果是的話，要感謝你的老師）。好，讓我們前往下一個更複雜的雙句陳述，其中一句話引用了一個來源。

範例 I-2

在 1970 年，美國收入最高的 10% 成年人平均收入為 \$135,000（現今幣值），而收入最低的 50% 成年人平均收入約為 \$ 16,500。依據「世界不平等資料庫」（World Inequality Database）的資料，這種不平等差距在隨後的五十年急劇擴大，最高收入攀升至約 \$350,000 美元，而最低 50% 的收入僅增至 19,000 美元²。

2 World Inequality Database, "Income Inequality, USA, 1913–2019," accessed 2020, <https://oreil.ly/eUYZn>.

有些圖片更具說服力

現在，讓我們用資料視覺化（尤其是圖 I-1 中的折線圖）來代替表格，比較一下哪個更有說服力。

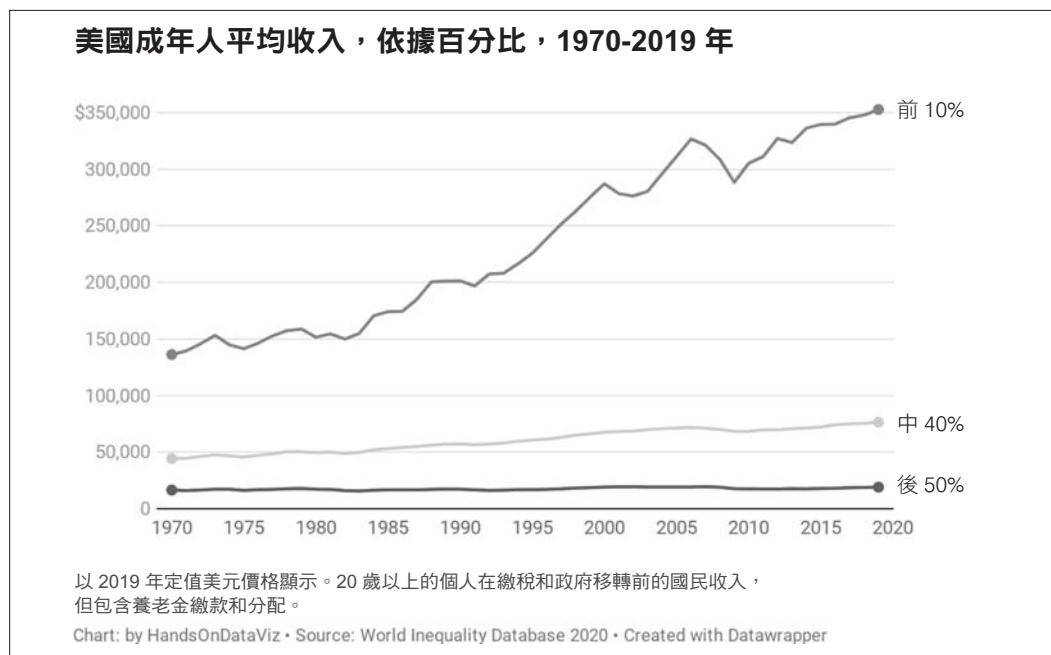


圖 I-1 探索一下美國成年人收入不平等之互動式折線圖 (<https://oreil.ly/x0Phg>)。

圖 I-1 比表 I-1 更具說服力嗎？由於折線圖內含的歷史起點和終點與表格相同，因此應該不會有任何區別。但它同時傳達了一個強大、視覺化的收入差距的資料故事，比表格更有效地吸引了你的注意力。

當你的視線沿著頁面上的水平彩色線條移動時，頂層、中間層和底層之間的不平等現象逐漸擴大。此圖表也將大量的詳細資訊塞進一張影像中。仔細觀察，你還會發現頂層收入等級在 1970 年代相對穩定，接著從 1980 年代飆升到現在，並且與其他收入線之間的距離越來越遠。同期間，中等收入階層隨著時間略有上升，而最低階層的命運仍然相對平穩，在 2007 年達到頂峰，然後在過去十年的大部分時間回落。常言道：富者更富，窮者更窮。但此圖表顯示了這些財富的成長有多快，而貧困仍然原地踏步。

現在讓我們來看看圖 I-2。它內含了與圖 I-1 相同的資料，但是以不同的格式顯示。你應該相信哪個圖表呢？別忘了我們警告過，要留心那些使用資料視覺化來說謊的人。

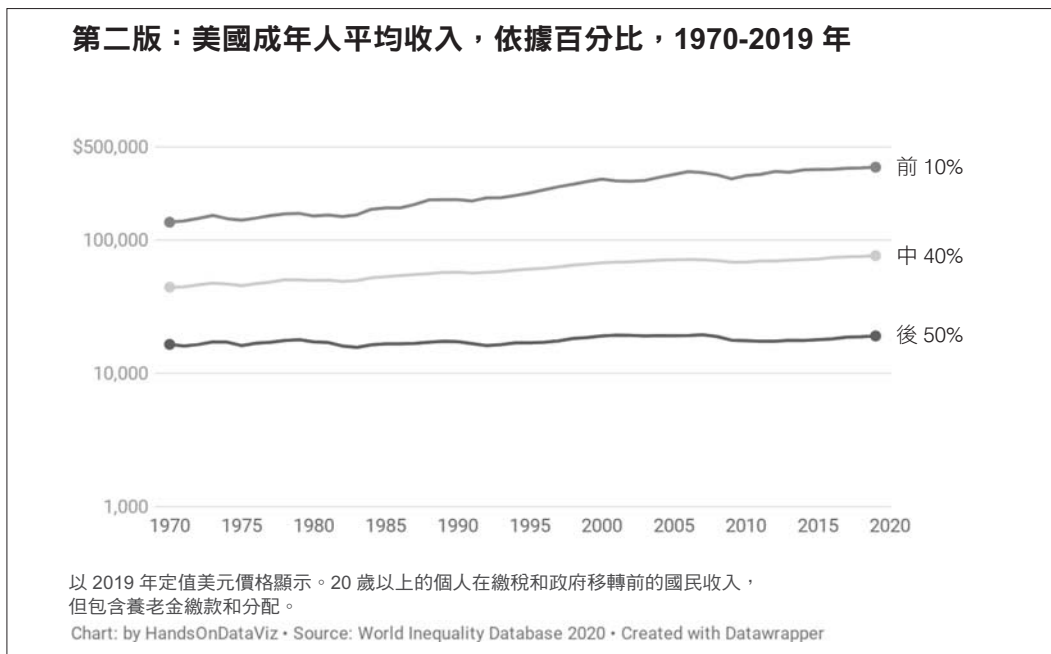


圖 I-2 探索美國成年人之收入不平等隨時間變化的互動式折線圖的替代版本 (<https://oreil.ly/vECje>)，使用與第一版相同的資料。

這是怎麼回事？如果圖 I-2 的資料與圖 I-1 相同，為什麼看起來截然不同？原本不平等差距的巨幅成長似乎已經消除了？發生了什麼事？危機突然消失了嗎？只是個騙局嗎？

雖然圖 I-2 中的圖表在技術上是準確的，但我們故意將它設計來誤導讀者。仔細查看縱軸上的標籤。第一個數字和第二個數字之間（\$1,000 到 \$10,000）的距離，與第二個和第三個數字（\$ 10,000 到 \$ 100,000）之間的距離相同，但呈現了差距極大的金額（\$9,000 對 \$90,000）。這是因為此圖表是用對數尺度 (<https://oreil.ly/Hr4dL>) 製作的，它最適合顯示指數成長。

你可能還記得在 COVID-19 疫情期間看到的對數尺度，將它們用來呈現傳統線性尺度很難顯示的極高成長率是很適合的。第二張圖表的資料點和尺度標籤是吻合的，因此在技術上是準確的，但它具有誤導性，因為沒有理由使用對數尺度來呈現此收入資料，除非是想要隱瞞這場危機。圖表可以用來闡明事實，也可以用來掩蓋事實。

真相的不同陰影

讓我們將收入不平等的分析，擴充到單一國家之外。範例 I-3 加入了比較性的證據及其來源。與先前美國範例顯示了三個收入層級之歷史資料不同，這個全球範例將重點放在每個國家中收入最高的 1% 的最新資料。此外，這項全球性的比較，衡量的並非美元收入，而是收入最高的 1% 人口占了國民收入的百分比。換句話說，它呈現每個國家中最富有的 1% 人口占了圓餅圖的多少範圍。

範例 I-3

在美國，收入不平等現象更為嚴重，美國最富有的 1% 人口目前占全國收入的 20%。相比之下，在大多數歐洲國家中，最富有的 1% 人口所占比例較小，介於全國收入的 6% 至 15% 之間⁴。

讓我們延續相同的思路，對範例 I-3 進行視覺化補充以評估其說服力。雖然我們可以製作表格或圖表，但這並不是快速展示 120 多個國家資訊的最有效方法。因為這是空間資料，所以讓我們將它轉換為互動式地圖，協助我們辨識出地域性模式，並鼓勵讀者去探究全球收入等級，如圖 I-3 所示。

與範例 I-3 相比，圖 I-3 是否更具說服力？地圖和文字提供了關於美國和歐洲的收入不平等的相同資料，所以應該沒有差別，但是地圖將你帶入一個有力的故事中，生動地描繪了貧富之間的鴻溝，類似上面的圖表範例。地圖中的顏色預示著危機，因為紅色在許多文化中表現出了急迫性。美國（以及包括俄羅斯和巴西在內的其他幾個國家）的收入不平等情況，以圖例中的最高的深紅色凸顯出來，收入最高的 1% 占國民收入的 19% 或以上。相較之下，當你的視線跨到大西洋另一端時，幾乎所有的歐洲國家都呈現淺米色和橘色，代表沒有緊急危機，因為他們的最高所得者在國民收入中所占的比例較小。

現在讓我們來介紹圖 I-4 的替代地圖，它包含的資料與圖 I-3 相同，但以不同的格式呈現。你應該相信哪張地圖？

4 World Inequality Database, “Top 1% National Income Share,” 2020, accessed 2020, <https://oreil.ly/fwQQV>.

清理凌亂的資料

通常資料集是凌亂而且難以立即視覺化的。裡面會有缺少值、不同格式的日期、純數值欄出現文字、同一欄出現多筆資料、同一名稱有各種拼法，以及其他意料之外的東西。範例見圖 4-1。如果你發現自己花在清理資料的時間比進行分析和視覺化還多，不用感到驚訝。

Year	City	Amount
1990	New York City	\$1,123,456.00
1995-96		2.2 mil
2000's	NYC	No data
2020	New_York	5000000+

圖 4-1 原始資料通常看起來很亂。

在本章中你會學到不同的工具，可幫助你決定使用哪種工具來有效清理資料。我們將從第 68 頁的「使用 Google 試算表進行智慧清理」、第 69 頁的「尋找並取代為空白」、第 71 頁的「轉置列和欄」、第 72 頁的「將資料拆分為個別的欄」，和第 75 頁的「將資料合併為一欄」，從 Google 試算表的基本清理開始著手。雖然我們在範例中使用的是 Google 試算表，但當中的許多原理（在某些情況下，它們的公式相同）適用於 Microsoft Excel、LibreOffice Calc、Mac 的 Numbers 或其他試算表套裝軟體。接下來，你會學到如何使用 Tabula 從文字類型 PDF 檔案中擷取表格資料。免費工具 Tabula 是全球資料新聞工作者和研究人員用來分析支出資料、健康報告，以及許多封在 PDF 中的各種資料集（請參閱第 76 頁的「使用 Tabula 從 PDF 擷取表格」。最後，我們將在第 79 頁的「使用 OpenRefine 來

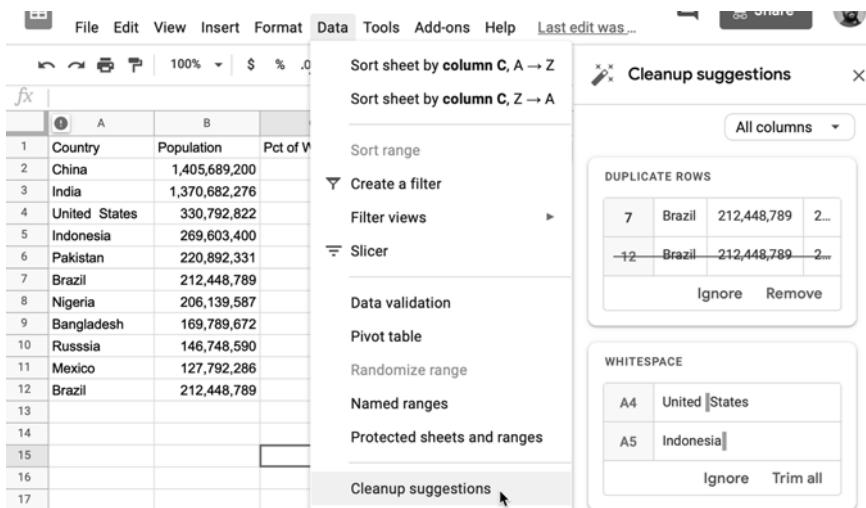
清理資料」中介紹 OpenRefine，這是一種功能強大且用途廣泛的工具，可清理最混亂的試算表，例如包含了同一名稱卻有數十種不同拼法的試算表。

使用 Google 試算表進行智慧清理

使用 Google 試算表來處理資料的最新原因之一，是使用它的智慧清理（Smart Cleanup）功能協助你識別和建議更正不正確的資料。此工具會開啟一個側欄選單，抓出潛在問題，然後讓你決定是否接受建議。

讓我們使用全球 10 個人口最多的國家的樣本資料（其中包括我們刻意加上的一些問題資料），來了解 Smart Cleanup 會發現和漏掉哪些類型的問題：

1. 在 Google 試算表中打開 Smart Cleanup 範例資料檔案（<https://oreil.ly/NxGPN>），使用你的帳戶登入，然後到「File」>「Make a copy」，製作一個可以在 Google 雲端硬碟中編輯的版本。
2. 到「Data」>「Cleanup suggestions」，檢視側欄中出現的項目。



Smart Cleanup 功能成功抓出了重複的條目（第 12 行），以及儲存格 A4 和 A5 中的空白。點按綠色的「Remove」和「Trim All」按鈕，確認 Google 試算表應該執行清理。

將資料轉成圖表

圖表吸引讀者深入你的故事。折線圖的斜線，或散佈圖上的密集點點等影像，會比文字或表格更有效地將證據傳遞到讀者眼中。但是，要製作有意義的圖表來吸引讀者關注資料中的關鍵洞見，需要對設計決策進行清晰的思考。

在本章裡，我們將在第 103 頁的「圖表設計原則」中，學到如何區分好的圖表和不好的圖表。你會讀到適用於所有圖表的重要規則，以及一些客製化設計時應遵循的美學準則。雖然許多工具都可以將圖表下載為靜態影像，但本書也示範了如何製作**互動式**圖表，以邀請讀者在自己的網頁瀏覽器中瀏覽資料。稍後，你會在第 9 章中學習如何嵌入互動式圖表到網站上。

表 6-1 介紹了你在本書中將製作的不同類型的圖表。圖表類型的決定，基於兩個主要因素：資料的格式，和你希望講述的故事的類型。例如，折線圖最適合顯示一系列連續的資料點（例如隨時間變化），而範圍圖更適合強調資料類別之間的距離（例如不平等之差距）。選擇好你要的圖表類型後，請跟隨我們的工具建議和步驟教學。本章介紹的都是使用拖放式選單的簡單工具，如第 114 頁的「Google 試算表圖表」、第 131 頁的「Datawrapper 圖表」，和第 146 頁的「Tableau Public 圖表」。此表格也將介紹提供了更多操控來客製化和託管視覺化的強大工具，例如第 11 章的 Chart.js 和 Highcharts 程式碼模版。這些進階工具需要具備在 GitHub 編輯和託管程式碼模版的知識，這在第 10 章會有介紹。



我們將條形圖和柱形圖放在一起介紹，因為它們本質上是相同的，差別只在於條形圖是水平的，而柱形圖是垂直的。它們最主要的區別是資料標籤的長度。請使用條形圖來顯示更長的標籤（例如「摩卡星冰樂 24 盎司」和「雙層牛肉吉事堡」），因為它們需要更多的水平閱讀空間。不需要太多空間的短標籤（例如「星巴克」和「麥當勞」）則條形圖或柱形圖都適合。你還會注意到，本章中的所有範例都是食物（因為我們在寫書的時候很餓）和健康飲食（因為我們也需要減肥）。

表 6-1 基本圖表類型、最佳用途和教學^a

圖形	最佳用途與本書中的教學
分組的條形圖或柱形圖	<p>最適合並列進行比較。如果標籤很長，請使用水平條代替垂直柱。</p> <ul style="list-style-type: none">簡單工具：第 114 頁上的「條形圖和柱形圖」或第 131 頁上的「Datawrapper 圖表」強大工具：第 11 章中的 Chart.js 和 Highcharts 模版
分組條形圖或柱形圖	<p>最適合在單獨的群集中進行比較。如果標籤很長，請使用水平條代替垂直柱。</p> <ul style="list-style-type: none">簡單工具：第 114 頁上的「條形圖和柱形圖」或第 131 頁上的「Datawrapper 圖表」強大工具：第 11 章中的 Chart.js 和 Highcharts 模版
堆疊條形圖或柱形圖	<p>最適合進行子類別或整體各部分之間的比較。如果標籤很長，請使用水平條代替垂直柱。</p> <ul style="list-style-type: none">簡單工具：第 114 頁上的「條形圖和柱形圖」或第 131 頁上的「Datawrapper 圖表」強大工具：第 11 章中的 Chart.js 和 Highcharts 模版

時間序列資料中的行與欄

表格和圖表是以相反的方向來處理時間序列資料的。在設計表格時，正確的方法是將日期當作欄標題來水平放置，以便從左讀到右，如下所示：⁴

年	2000	2010	2020
序列 1
序列 2

在 Google 試算表和其他一些工具中設計折線圖時，我們得將日期垂直排列在第一欄的下方，以便工具可以讀取為資料組的標籤，如下所示：

年	序列 1	序列 2
2000
2010
2020

到第 71 頁的「轉置列和欄」中，了解如何將資料從表格轉換為圖表。

堆疊面積圖

面積圖類似折線圖，但線下有填充的空間。最有用的類型是堆疊面積圖，它最適合結合前面提到的兩個概念：顯示整體的一部分（如圓餅圖）和隨時間推移的連續資料（如折線圖）。舉例來說，折線圖顯示了三種不同肉類的供應量隨時間變化的情況。但是，如果你也希望顯示這些肉類的總供應量隨著時間如何上升或下降的話，則很難在折線圖中看到這一點。你可以改變做法，使用堆疊的折線圖來直覺顯示每種肉的供應量，以及一段期間內的每人平均總供應量。堆疊的折線圖可同時顯示資料的兩個方面。

要製作堆疊的面積圖，請依照與圖 6-22 中的折線圖相同的方式來整理資料。現在你可以輕鬆地製作一張堆疊的折線圖，顯示每種肉類及其組合總量的供應量隨時間的變化，如圖 6-24 所示。整體而言，我們可以看到在 1930 年代的大蕭條時期後的總肉類供應量提高，以及 1970 年以後，雞肉成為總數的較大部分。

⁴ Few, Show Me the Numbers, p. 166

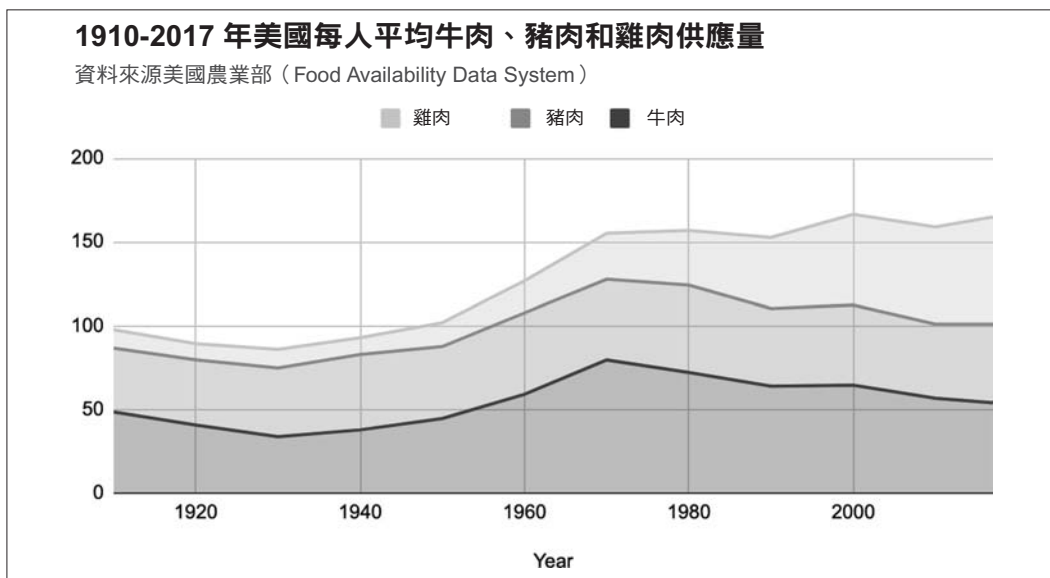


圖 6-24 堆疊面積圖；瀏覽互動式版本 (<https://oreil.ly/mdZlY>)。資料來自美國農業部 (<https://oreil.ly/PXFbI>)。

使用 Google 試算表中的堆疊面積圖樣版 (<https://oreil.ly/PXFbI>) 製作你自己的版本。這些步驟與本章之前的 Google 試算表圖表教學中的步驟相似。到「File」>「Make a copy」製作一個副本，以便在 Google 雲端硬碟中進行編輯。將資料整理成和折線圖相同的方式，第一欄為 X 軸的時間單位，並將資料組放到個別的欄中。選擇資料，然後到「Insert」>「Chart」。在「Chart editor」視窗的「Setup」標籤中，在「Chart type」下拉選單中選擇「Stacked area chart」。

現在你已經在 Google 試算表中製作了一些基本圖表，在下一單元中，我們將在另一種工具「Datawrapper」中製作一些更進階的圖表。

Datawrapper 圖表

另一個製作互動式圖表的免費協作工具是 Datawrapper (<https://www.datawrapper.de>)，它比 Google 試算表更具優勢。首先，即使沒有註冊帳號，你也可以立即在瀏覽器中使用 Datawrapper，而且它的四步驟流程對於許多新使用者來說都十分直覺。其次，你可以加上製作者、資料來源連結，甚至讓觀眾從你發佈的線上 Datawrapper 視覺化中的按鈕下載

資料，這一點會使你的工作更有可信度，而且更方便存取。第三，Datawrapper 支援的互動式圖表類型比 Google 試算表更廣，還有地圖（將在第 7 章中討論）和表格（第 8 章）。使用 Datawrapper，你可以製作本章到目前為止討論的所有基本圖表，以及我們將在第 132 頁的「帶注釋的圖表」，第 137 頁的「範圍圖」和第 140 頁的「散佈圖和泡泡圖」中介紹的三種新類型。稍後，你會在第 9 章中了解如何在網站上嵌入互動式 Datawrapper 圖表。

雖然沒有一種工具可以完成所有任務，但我們建議你考慮同時並用 Google 試算表和 Datawrapper，將這一組容易使用的工具轉變為視覺化的強大工具。首先如第 2 章所述，使用 Google 試算表當成試算表來整理和分析資料，記錄詳細的來源注釋，依照第 3 章所述的方法儲存原始資料檔案，並依照第 4 章所述清理資料。雖然 Datawrapper 可以轉置資料（交換行和欄），但它無法製作資料透視表，也無法像試算表一樣能夠尋找和合併資料。接著，將資料從 Google 試算表匯入到 Datawrapper 中以製作視覺化，因為正如我們稍後會介紹的，後者能讓你控制外觀、注釋和其他功能。Datawrapper 可以直接連結到資料存放的地方，因此與 Google 試算表能配合得很好。將 Google 試算表和 Datawrapper 搭配起來是強大的組合。

此外，我們強烈建議你使用高品質的 Datawrapper 學院支援頁面、大量範例（<https://oreil.ly/mldeT>），以及精心設計的培訓資料（https://oreil.ly/LbCo_）。閱讀這些內容不僅可以學到該按哪些按鈕，而且更重要的是學到如何設計更好的視覺化效果來講述關於資料的、真實且有意義的故事。在編寫本書時，我們從 Datawrapper 學院學到了很多東西，並且在下面的單元中提供了資源和明確的連結。最後還有一個優點是，Datawrapper Core 是開源程式碼（<https://oreil.ly/xQjHJ>），不過它不適用於大多數平台外掛來製作圖表和地圖。

現在，你已經準備好使用 Datawrapper 來製作超越基礎知識的新型圖表了。但是如果本單元中的 Datawrapper 或圖表類型不能滿足你的需求，請參考表 6-1 了解其他工具和教學，或者前面幾章關於試算表、資料來源和清理資料的內容。

帶注釋的圖表

帶注釋的圖表最適合凸顯特定資料，或在視覺化內加上相關說明。設計精良的說明能夠簡潔指出圖表中資料的重要性，以及透過後續更詳細說明的句子或段落，幫助回答「那又如何？」的提問。做注釋時要留意，因為避免加上不必要的「圖表垃圾」非常重要，如第 103 頁的「圖表設計原則」所述。