

frilly / October 23, 2015 09:57PM

[你應該知道，資料視覺化的七個秘密](#)

資料視覺化

[url=http://www.finereport.com/tw/]資料視覺化[url]的道路上充滿了不可見的陷阱和迷宮，最近ClearStory Data的兩位資料視覺化開發人員分享了他們總結出來的資料視覺化開發的7個不宣之秘，普通開發者了解這些方法能提升視野，少走彎路。

資料視覺化，特別是基於Web的資料視覺化的時代已經到來了。類似JavaScript的可視化庫如D3.js, Raphaël, 以及Paper.js, 以及最新瀏覽器所支持的如Canvas和SVG, 以及使得那些過去只能由計算機專家和專業設計人員開發的複雜的可視化變得越來越簡單了。

資料視覺化如今成為了很多網站項目的必備功能。而類似於Platfora, DatameerClearStory

Data以及Chartio等初創公司則可以利用基於瀏覽器的分析平台融到數百萬美元的投資。

資料視覺化是資料探索以及資料表現的重要方式，然而，對於資料視覺化的開發者來說，

依然有很多挑戰要去面對。這些迎接這些挑戰的方法，

則是很多專業的資料視覺化開發者不願意讓別人知道的秘密。ClearStory Data的兩位資料視覺化開發人員Nate Argrin和 Nick Rabinowitz 在

netmagzine.com上分享了他們總結出來的資料視覺化開發的7個秘密以及在實踐中如何應對的方式。

IT經理網編譯如下：

秘密一：現實中的資料往往很醜

大部分的資料視覺化的教程，都會讓你輕鬆地從一個原始資料集開始。

無論你是學習基本的柱狀圖還是力導向的網路圖，你的資料都是乾淨的，經過整理的資料。

這些完美的JSON或者CSV文件就像電視里的廚藝節目中的灶台那樣乾淨整潔。而實際上，

當你在處理現實中的真正的資料是，你80%的時間得用來搜尋，獲取，載入，清洗以及轉換你的資料。

這樣的過程，有時候可以用自動化的工具來完成。不過，差不多任何需要針對兩個以上的資料集進行清洗的工作總

會需要或多或少的人工的工作。有很多工具能夠把XLS文件轉化為XML的格式或者把時間戳轉換為其他日期格式。但是，

要想把一個公司的內部使用的銷售類型與競爭對手進行比對，或者對輸入錯誤進行檢查，

或者對不同的Encoding或者OCR產生出來的文字進行檢查時，就只能靠手工來處理了。

工具及處理方式：

1) 在資料視覺化項目中給資料清洗留出足夠的時間，特別是在需要處理多個資料源，需要手工錄入或者OCR資料，進行不同類別的配比，或者需要處理一些非標準格式時，需要留出更多的時間。

2) Google Refine (編者：需要翻牆)是一個很好的資料清洗工具，儘管在有些地方，特別是處理非表格化資料時有些不足。此外，還有一些資料清洗專用的工具如Data Wranger 和 Mr. Data Converter。不過，

很多的資料清洗工作仍然需要你熟悉腳本語言如Python或者需要你在Excel里進行一些手工工作。

記得把你的腳本存檔，你以後肯定用得上。

3) 用簡單的一些散點圖或者直方圖來發現一些超正常範圍的錯誤資料。

秘密二：柱狀圖往往更好

[img]http://www.finereport.com/tw/wp-content/themes/BusinessNews/images/15102301.jpg[/img]

柱狀圖

和柱狀圖比起來，氣泡圖可以在同樣的空間表現更多地資料，餅圖可以更清晰地表現整體和局部的關係，

樹狀圖能夠更好地表現分層的結構。然而，這些圖在簡單明了方面都無法與柱狀圖相比。

在考慮資料視覺化設計方案時，我們要問自己的第一個問題就是：「這個方案比柱狀圖好嗎？」

如果你需要在一個單一維度上可視化一個量化的資料集，那麼很少有別的方式能比得上柱狀圖。類似的，

時間序列最好表現為線狀圖，而散點圖一般用來表現兩個線性度量的相關性。在資料視覺化設計中，

使用這些從18世紀以來就一直使用的圖風險最低。而柱狀圖對於進行資料比較的可視化來說是最佳方式。

因為我們人眼最習慣的比較方式就是將兩個東西並排比較。

關於柱狀圖優先，其實揭示了資料視覺化中一個最大的秘密，那就是，那些最酷的可視化往往用處反而最小。

最求新奇以及美觀的可視化往往帶來一個問題，那就是資料的可理解問題。

很多柱狀圖的替代圖迫使人們用他們並不擅長的方式進行比較，如比較面積，角度，色彩，或者透明度等。

這些比較，說好聽的，是增加了比較的難度，說的嚴重一些，可能會對資料進行扭曲，

導致使用者得出錯誤的結論。

工具及處理方式：

1) 不要輕易拋棄那些傳統的可視化方式，如果這些方式能夠表現你的資料。先試試柱狀圖或者線狀圖，如果你的資料真的需要其他的再考慮其他圖。

2) 理解其他形式的圖的表現優勢，比如，氣泡圖支持更多地資料範圍，餅圖支持局部全局的對比，

樹狀圖能夠支持分層結構等等。

3) 柱狀圖是可視化最容易的圖形之一，你可以手工編寫一段HTML代碼，僅僅使用CSS或者很少量的JavaScript，或者從Excel裡面的一個公式，就可以生成一個有效的柱狀圖來。
秘密三：真實資料不可替代

對一個資料集進行清洗和格式化已經很繁瑣了，如果你需要設計一個基於多個資料集的可視化呢？比如你需要把公司不同部門的資料進行可視化，而這些部門各自有各自的資料庫，而且你也沒有時間手工把每個資料集進行清洗。這時候，人們的第一想法可能是抓一些Demo的資料來進行可視化。而且你的可視化庫里可能就有一些標準的樣本資料。很不幸，真實資料不可替代。Demo資料一般遵循正態分布而且資料量有限。是為了展示可視化用的。而一個看上去完美的柱狀圖，並不能幫助你解決那些資料缺失，異常資料或者現實中的真實問題。如果你過度依賴Demo資料，當你用真實資料時，你就會發現你的資料視覺化設計並不能真正滿足你的資料分析或者資料表現的需求。

工具及處理方式：

- 1) 如果你無法訪問整個資料集，不妨先試試從真實資料集中隨機取些樣本資料。
- 2) 保留無效或者缺失資料，如果你的資料集在可視化前不準備進行資料清洗，那麼也不要清洗樣本資料。
- 3) 真實資料集也許過大。在你使用樣本資料時，在生成最終的可視化圖前，等比例調整樣本資料規模。

秘密四：細節的地方才最頭痛

[img]http://www.finereport.com/tw/wp-content/themes/BusinessNews/images/15102302.jpg[/img]
信息圖

如上圖，當你水平排列資料標識時，資料標識會看不清，如果旋轉90度，資料標識是看清楚了，不過又浪費了很大一塊空間。選擇一個合適的資料標識格式對有些可視化來說是個解決方案，不過也不是對所有方案都適用。

設計資料標識，注釋或者橫軸縱軸通常都是在初始可視化後才考慮的。不過這些元素對可視化來說非常重要，而且可能會很困難或者需要大量時間才能把它們做好。特別是在你無法事先預知你的資料的情況下。

在設計你的可視化的時候，你需要留出相當部分的空間以便你可能需要添加標識只用，通常要在你的圖周圍留出相對較大的空間。橫縱軸上的標識要保證它們不相互覆蓋而且可讀。如果必要的話，可以將標識進行旋轉來增加可讀性。如果有一塊空間標識過於集中，而你又需要這些標識可讀，你可以讓考慮把標識離它們所指的元素遠一些，然後用連接線把標識和元素連起來。

另外一種方式就是把標識整合成一個組，用標識工具提示的方式來進行可視化。如果標識的文字過長，可以考慮進行縮寫或者把超出的文字剪掉等方式。

類似的，對圖的注釋也需要事先計劃好。最簡單的方式就是在可視化中保留一部分區域來方便添加注釋。不過，這樣意味著你的圖所佔的部分就會減小。為了保留空間，把注釋放在圖上的空白部分。或者把注釋做成可拖拽，這樣用戶可以把注釋移開來看注釋遮蓋的部分。

工具及處理方式：

- 1) 在設計時把資料標識，資料軸及注釋的空間在圖上留好。
- 2) 對資料標識，定義最大字元數，超出部分需要裁掉。把相近的標識組合在一起，在用戶點到時再顯示。
- 3) 對長注釋，可以考慮用滾動或者展開的方式
- 4) 無論如何，不要忽視這些元素。資料標識在你專注圖形設計的時候，可能不是你的最主要考慮，不過它們對可視化的用戶來說非常重要。

秘密五：需要的時候才用動畫

可視化的設計者經常希望能夠在最終設計上加上動畫。動畫是一種連接資料和變化趨勢的非常有用的工具。

不過動畫也常常會導致對你的資料的錯誤理解。你需要對它會如何影響你的最終效果進行評估，而不是簡單地在最後加上動畫效果。動畫最適合表現的，是揭示資料如何在不同狀態下組合在一起，如何隨時間變化或者是如何相互影響的等場合。

一般的設計原則是，動畫要簡單，可預測並且可以重新播放。讓用戶能夠多次播放動畫，可以讓他們看到動畫元素從哪裡開始到哪裡停止。要避免不同元素在移動中互相覆蓋，不要讓元素的運動不可預測。對於複雜的動畫，研究表明，可以把動畫分解為幾個不同的階段，在每個階段暫停一會給用戶一些時間來體會。這樣有助於提高用戶的理解。

工具及處理方式：

- 1) 儘可能讓動畫簡單
- 2) 如果動畫複雜或者有很多動畫元素，可以考慮分階段動畫
- 3) 一開始動畫往往能夠給人新鮮感，不過會很快讓用戶感到厭倦。

不要僅僅因為你會加動畫就在你的可視化你加上動畫。

秘密六：資料視覺化不是分析

資料視覺化可以產生一些分析結果，不過需要指出的是，可視化是一個輔助分析的工具，而不是資料分析的替代，它也不是統計的替代：你的圖形可能揭示了一些資料差異或者資料的相關性。不過，

要得出存在這些差異和相關性的可靠結論，還需要運用統計的方法。要對你的資料真正了解，需要分析的技能，以及專業的知識。不要指望可視化能夠給你這些。因此，在進行可視化項目的時候，要調整客戶或者你的CEO的期望值。

工具及處理方式：

1) 除非你就是資料分析師，你對資料視覺化得出的結論不要輕易下判斷。如果需要進行結論，最好找一個統計師或者專業人士一起驗證後再給結論。

2) 一些細小的設計改變，比如調色板的變化，對某個變數的可視化方式等，都可能改變可視化得出的結論。如果你用可視化進行分析，一定要試試多種可視化方式，而不要依賴於一種方式。

3) Stephen Few的書「Now You See It」裡面介紹了利用可視化進行商業分析的方式，包括對開發者如何設計可供分析使用的可視化工具的一些建議，讀者可以參考

秘密七：資料視覺化不僅僅是編程

現在大量的可視化編程庫和教程使得普通的人員在進行基於Web的可視化中，也可以設計出高質量的可視化產品。

然而，要想真正設計一個能夠提供深入見解，或者能夠清楚表達的可視化產品，除了編程之外，還需要很多其他的技能。比如圖像設計，資料分析，交互設計，以及對人們認知的了解等。這些技能，是那些可視化編程庫提供不了的。（參考本站文章：資料視覺化是科學，也是藝術）

不過，好消息是，如果你堅持採用一些資料視覺化的基本原則的話。

你也不需要對這些技能了解太多。對於初學者來說，需要堅持一些最基本的原則，比如，盡量使用柱狀圖，不要把圓半徑設置按線性比例設定（編者：在面積比較時會給用戶錯誤理解），設計要簡單（不要用3D，少用動畫，不要用陰影）等。按照一些好的可視化樣本，初學者也可以創造出好的可視化作品來。

4500+企業選擇FineReport報表與 [\[url=http://www.finereport.com/tw/\]](http://www.finereport.com/tw/)BI[\[url\]](#)

[\[url=http://www.finereport.com/tw/\]](http://www.finereport.com/tw/)商業智慧[\[url\]](#)工具【免費下載】

opensource開發，類excel設計，全方位異質資料庫整合，資料填報、Flash列印、權限控制、行動應用、客制化、交互分析、報表協同作業管理系統。

分享自：巨量資料
