

## 見聞科技有限公司 函

地 址：台北市北投區建民路116巷11號2樓  
聯絡人：吳泰輝 電話：(02)2823-9270  
e-mail：tkirby@plotdb.com

受文者：文化內容策進院

發文日期：中華民國111年 1月 12 日  
發文字號：見聞函字第11101120001號  
速別：普通  
密等及解密條件或保密期限：無  
附件：如文

主旨：檢送「文化內容產業數據儀表板設計研究勞務採購」案結案文件，請查收。

說明：依據契約第五條辦理。

正本：文化內容策進院  
副本：

## 文件清單

依服務建議書所列項目之對應文件表列於下。

### 「潛在擴充資料分析」與「依資料分析結果需求之介面更新、設計調整」

彙整於「文化內容產業數據儀表板設計研究」報告中

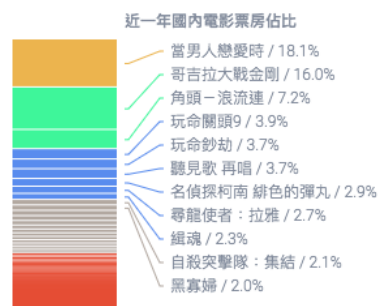
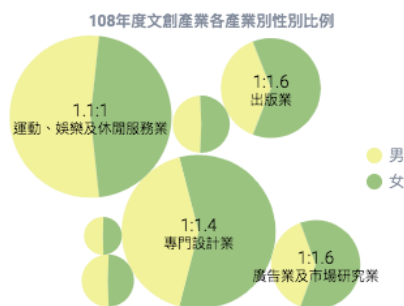
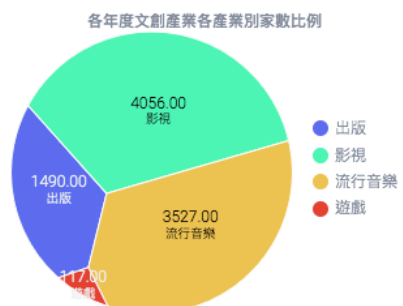
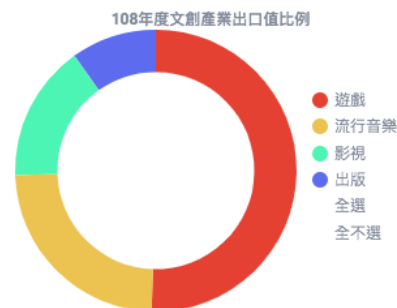
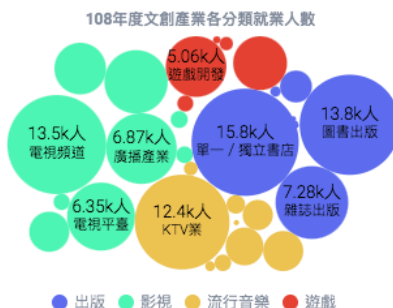
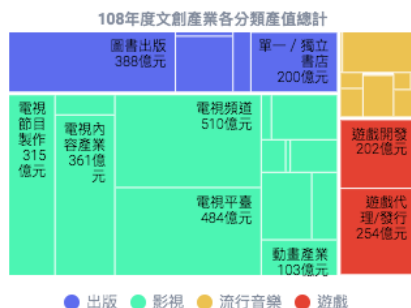
### 「資料格式說明與技術支援」

資料格式說明記載於「資料檔格式簡介」報告、技術支援之相關使用文件則為「文創產調視覺化使用手冊」

### 「供外部網站展示之嵌入元件」

六組元件之網址與元件之截圖分別如下

- 108 年度文創產業各分類產值總計：<https://vis.taicca.tw/banner/vis/?favor=1>
- 108年度文創產業各分類就業人數：<https://vis.taicca.tw/banner/vis/?favor=2>
- 108年度文創產業出口值比例：<https://vis.taicca.tw/banner/vis/?favor=3>
- 各年度文創產業各產業別家數比例：<https://vis.taicca.tw/banner/vis/?favor=4>
- 108年度文創產業各產業別性別比例：<https://vis.taicca.tw/banner/vis/?favor=5>
- 近一年國內電影票房佔比：<https://vis.taicca.tw/banner/vis/?favor=6>



### 「分析調查期間供雙方檢視之資料展示站台」

以於專案執行期間設立於 <https://vis.taicca.tw>

# 文化內容產業 數據儀表板設計研究

本專案針對下列兩組產業類型之常見資料集進行結構分析，並提出建議的視覺化與儀表板規劃。資料集包括：

- 假想之線上影音數據 (以流行音樂產業為主)
- 假想之獨立書店數據 (以出版產業為主)

此外，我們亦針對現已有之 友善書業供給合作社提供之獨立書店相關數據 併 假想之出版產業獨立書店銷售數據 統整進行出版產業獨立書店部份之設計規劃。

## 研究方法

此設計研究分別針對兩組資料 (線上影音、獨立書店) 進行三步驟的分析：

- 資料分析
- 儀表板構成
- 參考設計

在 資料分析 部份，將分為 資料標記 → 資料轉換 → 視覺化 三項子步驟做處理，並在視覺化步驟匯整目標讀者可能感興趣的資料觀點。

在 儀表板構成 部份，我們將以資料觀點出發，構思儀表板布局以及對應適合的視覺化手法，並簡單討論可能的儀表板設計改良方向。

在兩組資料討論完後，我們再統一匯整所有提及的儀表板 設計改良 部份做總結，並依使用情境考量額外的功能設計，提出額外的設計規畫，以提升儀表板整體使用體驗及有效性。

資料分析的 資料標記 → 資料轉換 → 視覺化 步驟，我們將以系統性的方式進行，分項描述於後。

# 1. 資料標記

除了友善書業已提供之資料集外，我們並不確定實際上最終將取得的資料類型，因此我們將分析各領域現有之服務或已釋出之資料集，歸納資料之並做初步分析，做為研究的基礎。

為了簡化資料分析的過程並更容易（甚至自動地）生成視覺化圖表，我們將使用 **視覺化資料分類法** 標記歸納後的各產業資料。視覺化資料分類法將資料以 欄（Column）為單位，依據其對應的型態標記其形式。

## 基本分類標記

我們預計使用的 **基本分類標記** 包括：

- **N** 名目式 / Nominal  
單純的文字、名稱資料。以 **N** 標註。任何資料均可視為名目式資料。
- **C** 分類式 / Categorical  
亦為文字資料，但各筆資料間共用少數的幾組文字。以 **C** 標註。
- **O** 有序式 / Ordinal  
有序、可以建立順序的，如年份、金額。  
任何資料均可轉化為有序式。以 **O** 標註。
- **T** 時序式 / Temporal  
資料的本質是時間。亦為區間式（I）或有序式資料（O）。以 **T** 標註。
- **I** 區間式 / Interval  
數值資料，但各筆資料間相加不具實質意義，如年齡。以 **I** 標註。
- **R** 比例式 / Ratio  
可以相加的數值資料 以 **R** 標註。

例如，以假想的 **逐年各類圖書銷售量** 來說，我們可能有如下表的資料：

年份	類型	銷售量(本)
2018	漫畫	120
2018	科普	421
2018	商業	132
2018	小說	220
2019	漫畫	123
2019	科普	521
2019	商業	332
2019	小說	122

由於視覺化考慮的是各欄位的分類標記，我們需要依視覺化資料分類法將各欄位做標記。例如，

- 年份 為時序資料，標記為 **T**。
- 類型 為分類式資料，標記為 **C**。
- 銷售量 則為比例式資料，標記為 **R**。

標記過後之資料如下：

年份	類型	銷售量(本)
T	C	R

為了方便，我們使用 **T x C x R** 來表示此資料的資料分類標記。

### 進階性質

除了基本分類標記以外，資料有時會有其它亦值得標注的、適合用於視覺化的概念。舉例來說，地理資料中常有 **縣市 → 鄉鎮** 這類型的資料，若使用資料標記來描述為 **C x C**，但這個描述方式並沒有表現到兩者之間上對下的關係；亦或者若我們使用多個欄位表達六都的人口數，資料標記為 **R x R x R x R x R x R** 但標記並沒有說明這六個欄位表現的為同質性的人口數資料。

為了更完整的表現此類資料的特性，我們額外定義 **進階分類標記**：

- **同質資料**  
在不同欄位表現同質性資料時，使用方括號表現。  
例如，用 **[ R x R ]** 表示 台北與新北的人口數。
- **對立資料**  
同質性但包含對立概念的資料，用角括號表現。  
例如，用 **< R x R >** 表示 男人口數 vs 女人口數。
- **階層資料**  
欄位間有階層性時，使用箭頭做標記。  
例如，用 **C → C** 表是 縣市與鄉鎮 欄位。
- **階層資料**  
若我們有不定數量的額外欄位可追加時，使用星號標記。  
例如，用 **R\*** 表示一組以上的數值資料。

以人口數資料為例，**台灣各縣市、各鄉鎮 之 未成年、壯年、老年男女人口數** 之資料分類如下：

縣市	鄉鎮	男人口(未成年)	女人口(未成年)	男人口(壯年)	女人口(壯年)	男人口(老年)	女人口(老年)
C	C	R	R	R	R	R	R

其資料標記可以寫成：

**C → C x [ < R x R > x < R x R > x < R x R > ]**

## 2. 資料轉換

在資料標記完成後，我們應該就資料的內涵考慮可能的資料轉換。

資料轉換可以想成是透過重新整理資料讓我們可以從另一個面向來觀察資料。以上例 逐年各類圖書銷售量來說，梳理完的資料類型為  $T \times C \times R$ ，表示的是在 各個年度、各別分類 的圖書、其 各別銷售數量 為何。然而，我們亦可以銷售數量依年度總計，得到 各個年度、在該年度的書籍總銷售量 為何，分類標記為  $T \times R$ 。以這個例子來說，我們以 年度相同 為條件，將 不同分類的銷售量 做 總和， 分類 欄位因此消失，留下的只有做為條件的 年度 以及統計過後的 總和。

在此研究中，我們可以使用以下幾種轉換方式：

### 叢聚轉換 (Clustering)

將資料依範圍做分組轉換。例如，將年齡欄位 (R) 按 未成年 (0 ~ 18)、壯年 (18 ~ 65)、老年 (65+) 分為三組。資料經轉換後，會變成分類式 (C) 資料。若資料集中有其它欄位，則其它欄位需要進一步進行聚合轉換。

叢聚轉換會消滅目標欄位，將其轉換為標記類型為 C 的新欄位。

### 聚合轉換 (Aggregation)

依資料分類重新統計各欄位。如上例中，依年齡欄位重新分組後，我們可以將各分組中的身高以 平均值 (Average) 表示，亦可將各組中的財產依 總和 (Sum) 表示，其中 平均 與 總和 即為 聚合轉換 的一種。

### 資料表合併 (Join)

若有多個資料表指涉到同一個實體 (例如，兩個資料表都有身份證字號)，這兩個資料表則可以透過身份證字號合併 (Join)。

例如，下列兩表分別代表某學生的考試成績：

學號	國文成績	學號	數學成績
A001	50	A001	80
A002	60	A002	72

兩者可透過 學號 一欄做合併成：

學號	國文成績	數學成績
A001	50	80
A002	60	72

樞紐 / 反樞紐轉換 ( Pivot / Unpivot )

我們可以將分組欄位按分組重新編排成獨立的欄位，例如上例中 年齡分類 欄位中的 未成年、壯年 與 老年，可以搭配 身高 欄位重新分出三個獨立的欄位：

轉換前		轉換後		
年齡分類	身高	身高(未成年)	身高(壯年)	身高(老年)
C	R	R	R	R

如上表所示，原本 C x R 的資料分類經轉換後變成 [ R x R x R ]。

3. 視覺化

在資料標記完成後，再進一步使用各資料集的視覺化分類歸納出適合的視覺化呈現方式。

不同的視覺化方式有各種不同適合的對應分類，而資料亦可以透過各種轉換後再套用到原先不適用的視覺化類型上。

舉例來說，以最常見的 折線圖 ( Line Chart ) 來說，可以對應到 時序式 (T) x 一組以上的 比例式 (R) 的資料組，因此我們只要有 T x [ R x R ... ] 類型的資料，即可使用折線圖來呈現。

就上例「逐年各類圖書銷售量」來說，雖然資料標記為 T x C x R，但我們可以透過樞紐轉換 ( Pivoting ) 將其中的 C ( 類型 ) 轉成多組的 R ( 銷量 )，變成 T x [ R x R ... ]：

年份	漫畫銷量	科普銷量	商業銷量	小說銷量
2018	120	421	132	220
2019	123	521	332	122

轉換之後，我們得到了一個年份與四個銷量的表格，資料標記變為 T x [ R x R x R x R ]，恰好可適用於多重折線圖。

除了資料轉換外，我們亦可透過額外的補助工具加強圖表所能呈現的資料類型。比方說，我們可以透過選單元件，將 圖書分類 ( C ) 單獨呈現供選擇；此時資料集則為依選擇過濾後的結果，例如：

過濾元件

漫畫

科普

商業

小說

過濾後的資料集 ( 省略完全相同的 類型 欄後即為 T x R )

年份	類型	銷售量(本)
2018	漫畫	120
2019	漫畫	123



考慮資料標記對應到視覺化與強化元件有多種可能，再加上本案為延伸前案之研究案，目的主要為擴充前案之設計，因此談及視覺化時，我們會優先考慮尚未使用過的視覺化類型為主進行討論。其它亦合適但已實用於現有儀表板中之視覺化，基本上我們不另外討論。

## 線上影音數據

在線上影音數據部份，我們將以 KKBOX 之線上影音數據為例。

## 資料分析

KKBOX 線上影音服務透過其網站與應用程式對其 註冊會員 以 訂閱 / 單時段消費模式 提供 隨選即播線上目錄中之音樂。從這個基本設定來看，此類線上影音服務一般有以下數據：

- 會員資訊 ( 註冊人數、月活躍人數、訂閱會員數、會員組成分析 等 )
- 營收資訊 ( 月營收、人均月營收、LVT 、營收組成 等 )
- 樂曲資訊 ( 樂曲 / 影片資訊、歌手、出版商資訊 等 )

除此三種資訊外，亦有資訊之間衍生的額外資訊 ( 如「樂曲好評數」)。接下來我們將對此三種分類與其衍生資訊做逐一分析。

### 會員資訊

會員以人組成，因此帶有人為基礎的基本資訊；同時會員在平台上的活動亦可納入考量，比方說「收聽音樂的時段」、「日均播放長度」等等資訊。

會員本身可能有的資訊以及其對應的分類如下：

- N - 會員代碼
- C - 性別
- C - 職業
- C - 地區
- C - 使用裝置
- C - 會員等級
- I - 年齡
- I - 會員年資
- T - 註冊日期

其中「年齡」「年資」這類資料常見的使用情境亦為分組分析 ( 如，18 ~ 30 歲、30 ~ 45 歲做分組 )，我們亦可將這兩類資料視做分組式資料 ( C ) 來處理。因此可以說會員本身的資料大多為分類式 ( C ) 為主。

總地來說，我們可以用  $N \times C \times C \times C \times C \times C \times I \times I \times T$  描述會員資料的分類標記。



## 樂曲資訊

單一樂曲所具有的常見資料欄位有：

- N - 樂曲代碼
  - N - 樂曲名稱
  - N - 所屬專輯
  - N - 歌手 / 演出
  - N - 發行人
  - N - 製作人
- C - 來源地區
  - C - 語系
  - C - 類型 ( Genre )
  - C - 子類 ( Subgenre )
  - C - 年代 ( Decade )
  - R - 長度

其中歌手、製作人、發行人等資料均可延伸出額外的資料內容，例如歌手的性別、國籍、發行人的資本額等等。

然而欲使用此類額外資料，往往會需要一個完整的音樂人資料庫，此類服務通常會是商業用途並需要付費使用；考量到此研究案的性質並不一定需要深入到如此細節的資料中，因此我們先不進一步分析此類資料的額外資訊。

樂取資訊的分類標記為：**N x N x N x N x N x N x C x C x [ C > C ] x C x R**

## 營收資訊

營收模式會隨企業經營而調整，並且一般不一定會願意釋出相關資料供使用，因此我們僅依現有模式粗略做整理。單就服務的營收來看，至少有以下兩個面向可以考慮：

- **統計時段 ( T )**  
營收一般均為統計一定時段之內的收益，因此我們可以考慮「天」、「週」、「月」、「季」、「年」等不同週期的營收統計數據，但以儀表板來說我們若能取得越小區間的資料則越能呈現更豐富的數據與讀者。
  - **營收類型 ( C )**  
服務收入往往不會是單一類型，以 KKBOX 來說營收至少還可區分為：

類型	說明
<b>訂閱</b> 月租	用戶選擇持續訂閱 KKBOX 的付費服務。
<b>單次</b> 計時制	用戶選擇一次性的付款，以換取在特定時段內的付費功能。
<b>廣告</b> 歌單、插播等	向其它企業販售版位、插播或任何可供置入內容的廣告。

從上述面向分析，我們可以想像類似下者的營收資訊：

T - 時段	C - 類型	R - 金額
以下為範例資料，非真實數據		
2021/11/01 ~ 2021/11/02	訂閱	210000
2021/11/01 ~ 2021/11/02	單次	150000
2021/11/01 ~ 2021/11/02	廣告	350000

由於消費牽涉到會員，我們亦可以將各筆營收統計展開至最細節的數據，加入會員對應的內部代碼；此外，雖然 KKBOX 已不提供單曲購買服務，我們亦可想像類似的服務提供此類購買形式，因此我們亦可額外加入樂曲對應的內部代碼用以代表該筆消費與特定樂曲直接相關。

有了會員與樂曲內部代碼，我們即可以連動營收與會員、樂曲資料，進行進一步的資料分析：

T - 時段	C - 類型	N - 會員代碼	N - 樂曲代碼	R - 金額
以下為範例資料，非真實數據				
2021/11/01 ~ 2021/12/01	訂閱	8388608	-	278
2021/11/01 ~ 2021/11/04	單次	1048576	12345	180
2021/11/02 ~ 2021/11/05	單次	3355443	24671	180
2021/11/03 ~ 2021/11/06	單次	3939889	62612	180

總地來說，我們可以使用 **T x C x N x N x R** 描述營收資訊的分類標記。

## 衍生資訊

當考慮樂曲在平台使用的狀況，便可以從 **樂曲** 與 **用戶** 互動的內容探索更多資料，進而做下列統計：

- N - 會員代碼
- N - 樂曲代碼
- R - 播放次數
- R - 播放分鐘數
- R - 收藏數

小結

就 KKBOX 可能有的資料集來說，我們已列舉了下列四種資料：

會員資訊

N	會員代碼
C	性別
C	職業
C	地區
C	使用裝置
C	會員等級
I	年齡
I	會員年資
T	註冊日期

樂曲資訊

N	樂曲代碼
N	樂曲名稱
N	所屬專輯
N	歌手 / 演出
N	發行人
N	製作人
C	來源地區
C	類型 ( Genre )
C	子類 ( Subgenre )
C	年代 ( Decade )
R	長度

衍生資訊

N	會員代碼
N	樂曲代碼
R	播放次數
R	播放分鐘數
R	收藏數

營收資訊

T	時段
C	類型
N	會員代碼
N	樂曲代碼
R	金額

## I 視覺化

在完成資料分析後，我們還可以進行資料轉換來取得調整過的分類標記。例如：

- 將營收與會員表透過 會員代碼 合併 ( Join )
- 再將相同 地區 ( C ) 的各筆資料做聚合 ( Aggregation )
- 便取得特定地區 ( C ) 會員對 營收 ( R ) 的貢獻，標記為  $C \times R$

或者

- 將營收與會員表透過 會員代碼 合併 ( Join )
- 依會員年齡叢聚 ( Clustering )、分成數個年齡層 ( I 轉 C )
- 依年齡層、消費類型對 單次購買量 ( R ) 與 訂閱量 ( R ) 做 聚合轉換 ( Aggregation )
- 最後取得各年齡層 ( C ) 單次購買量 ( R ) 與 訂閱量 ( R ) 的比例性資料，標記為  $C \times \langle R \times R \rangle$

取得各別的資料標記後，我們就可以來嘗試調性相合的視覺化方式，例如：

- 標記  $R \times C$  ( 特定地區營收 ) 恰好與 圓餅圖 相符
- 標記  $C \times \langle R \times R \rangle$  ( 各年齡層消費類型比例 ) 則與 圓餅泡泡圖 相符

然而，要選出什麼樣的資料組合，以及運用哪一個相符的視覺圖表，這是個相當主觀的問題。由於我們可以呈現的資料相當的多樣，在視覺化時我們可以考慮優先提供讀者可能感興趣的資訊。

為了了解這個問題，我們先試著大略列出一些讀者可能感興趣的題目：

- 最近哪首歌最受歡迎？
- 年輕人現在都在聽什麼？
- 音樂產業發展狀況？

這些問題大致上可以透過下面幾種資訊來說明：

- 排行榜
- 營收狀況
- 用戶數

## 排行榜

排行榜可以告訴我們在 **指定的時間範圍** 內，哪些樂曲 / 歌手最受歡迎。除此之外，**特定分類** 的排名也是必要的，這有助於幫助讀者了解特定類型的音樂現況。最後，排名勢必要依照樂曲 / 歌手的表現 - 我們可以使用衍生資料中的 **播放次數**、**播放時數** 做為排序依據，亦或者使用營收資訊中的 **金額** 做為依據。

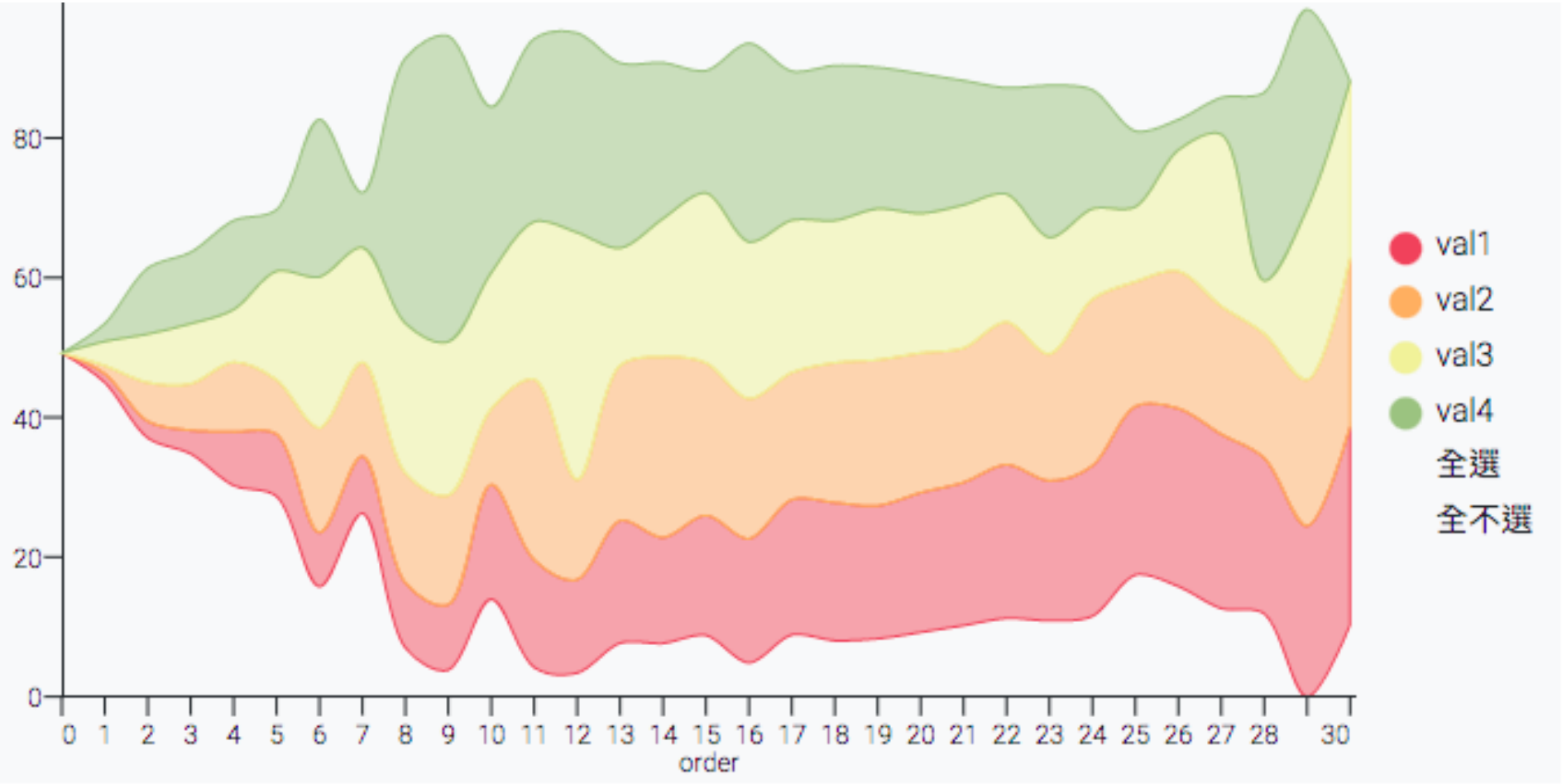
上述四點分別包含了時段、樂曲 / 歌手、分類 以及 金額 / 次數 等四組資訊：

資料分類	欄位	說明
T	時段	可按週、月、年等
N	名稱	排名對象，樂曲、歌手、專輯或製作人等
C	分類	排名領域，如男性歌手、90 年代等。可以同時使用多組分類
R	數值	排名依據，如播放次數、播放時數等。亦可同時使用多組數值

大致上來說，我們的分類標記可以寫成 **T x N x C\* x R\***。

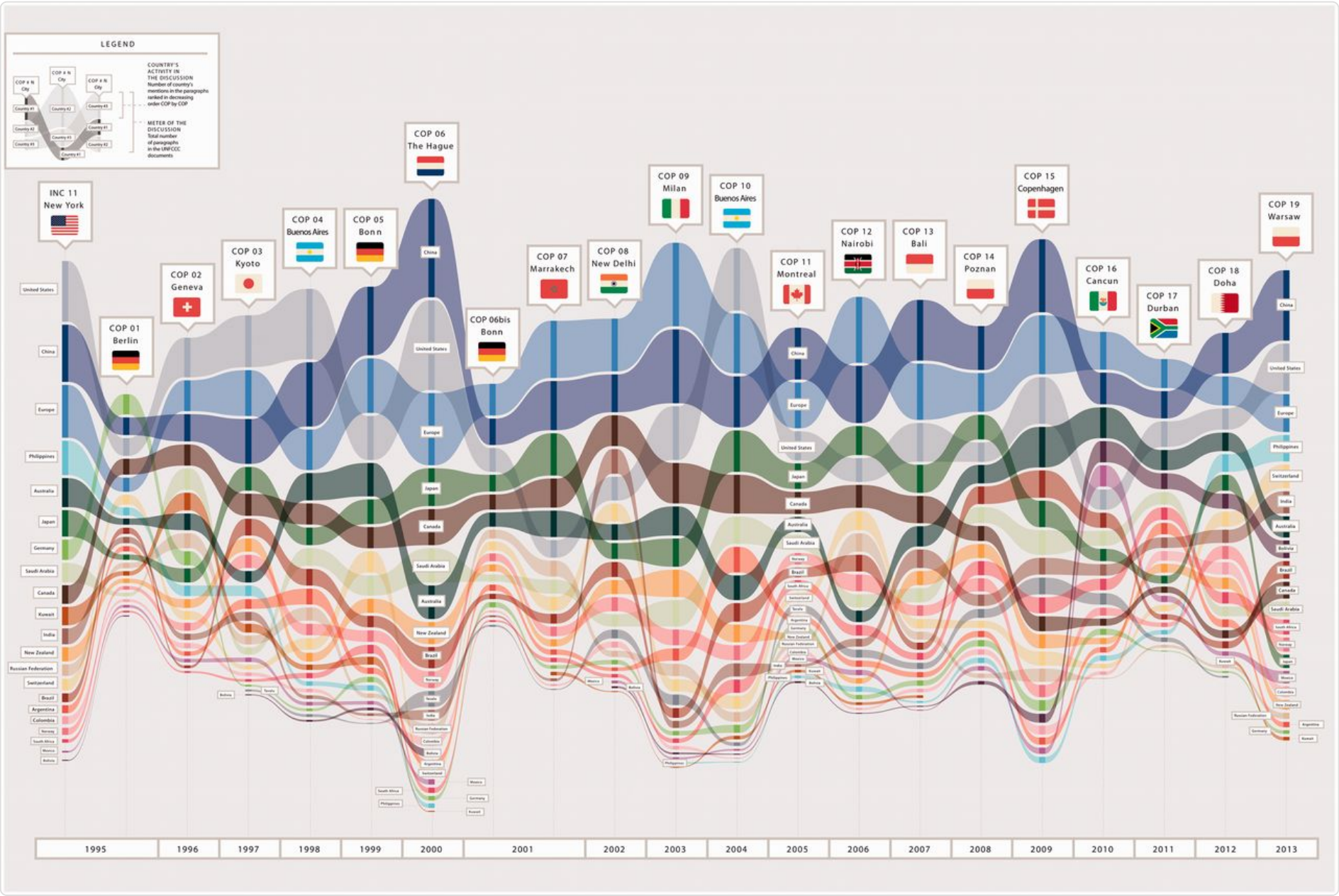
## 主題河流

若只考慮 **T x N x R**，我們可以將其中的 **N (名稱)** 經樞紐轉換 (Pivoting) 變換成 **T x [R\*]** 式的資料，適合用於折線圖示的圖表，其中常用於聲量統計的 **主題河流 (Themeriver / Streamgraph)** 圖表便適合呈現此類排行資料；若我們需要使用被忽略的 **分類 (C)** 欄位，可透過額外的選單元件供用戶切換：





因為主題河流在 Y 軸僅以高度為指標，不考慮相對座標軸原點的值，我們亦可將各數值在 Y 軸上排序，得到排序主題河流 ( Sorted Streamgraph )，一種類似於凹凸圖 ( Bump Chart ) 的面積圖版變形：



The Safest Cars of the Last Eleven Years, from *titleproloans.com*

### 散布圖與其變種

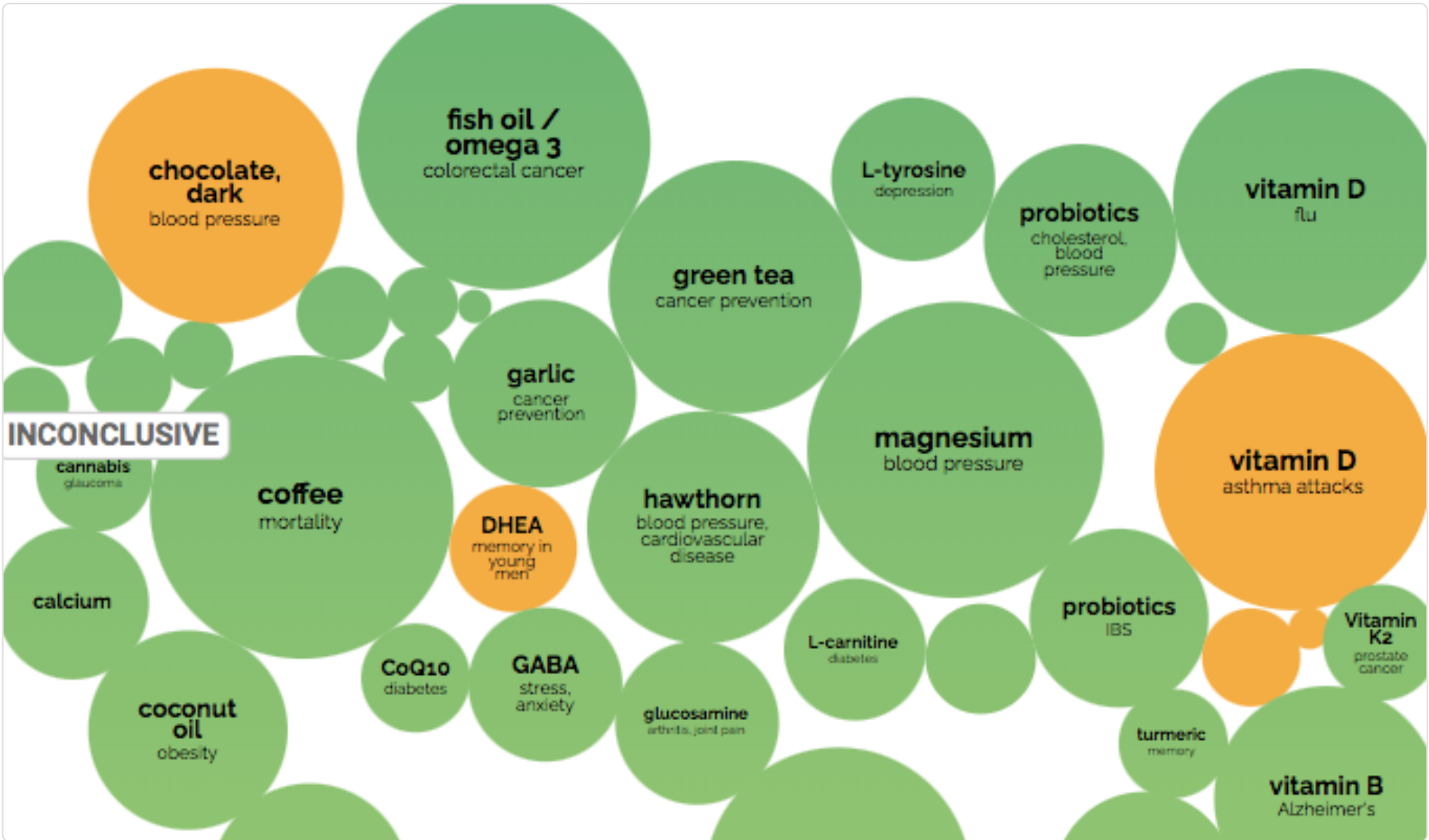
若希望直接透過圖表呈現分類資料，我們可以使用散布圖 ( Scatter Plot ) 與其變種；散布圖的變種有許多類型，依設計不同有很大的差異，但一般無法呈現時序資料 - 在這個情況下，我們仍可以統計特定時段的資料做排行，分類標記則為  $N \times C^* \times R^*$  。

舉例來說，"[Snake Oil Supplements?](#)" ( from Information is Beautiful ) 使用依 Y 軸排序的泡泡圖，再透過顏色 ( C ) 與選單元件 ( C ) 提供額外的變量，支援  $N \times C \times C \times R \times R$  形式的資料，亦適合用來呈現排行榜式的概念。

若以此圖例類推，我們可以做下列維度對應：

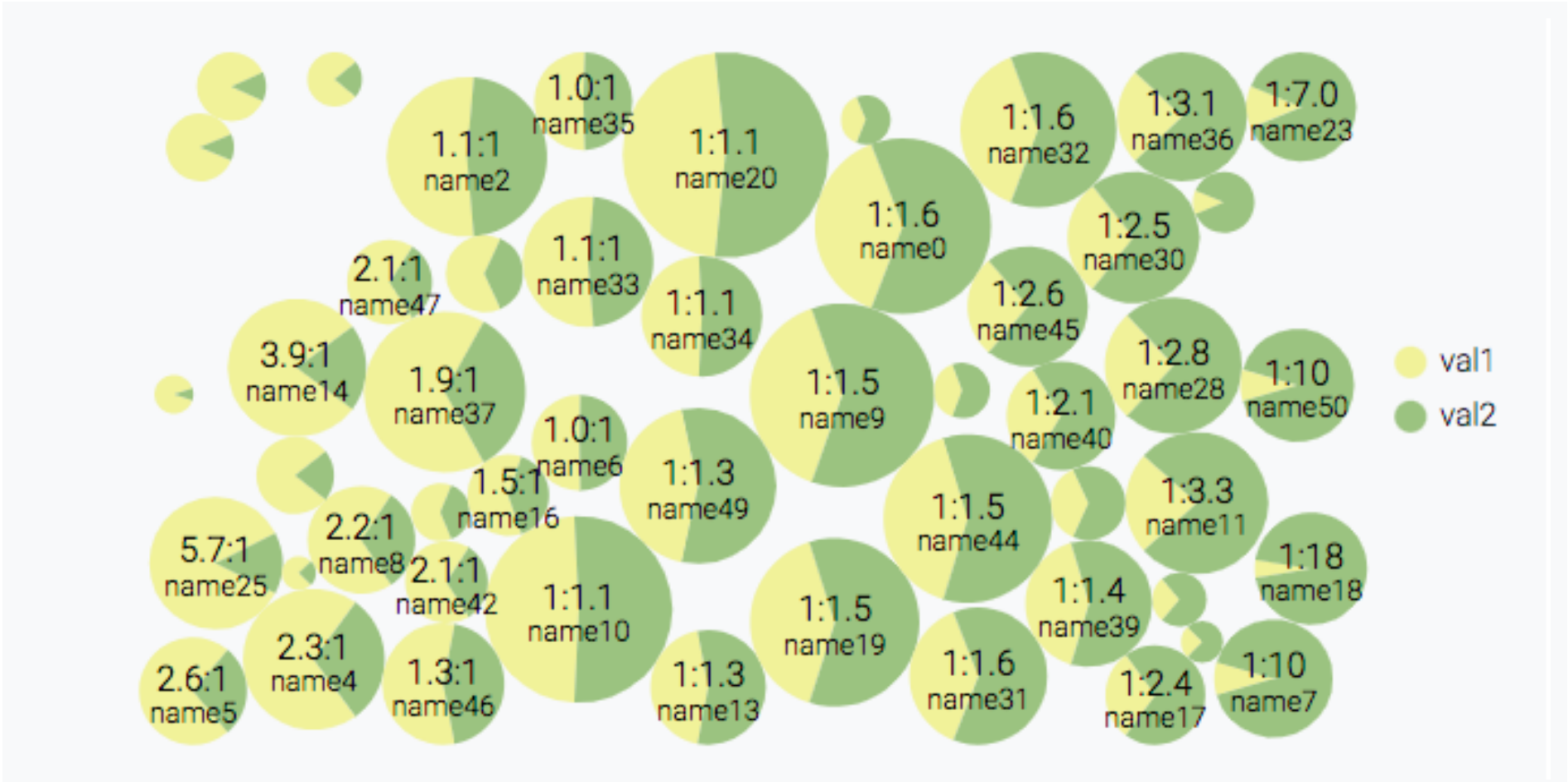
- N - 歌手名稱，顯示為泡泡之名稱
- C - 來源地區，以洲別對應到泡泡顏色表現
- C - 曲風，透過選單元件選取
- R - 播放時數，做為 Y 軸排序依據，越上方越長
- R - 貢獻營收，做為泡泡大小依據，越大營收越高

下圖為此類型泡泡圖之示意圖，摘自 *Information is Beautiful*：



Snake Oil Supplements? from *Information is Beautiful*

亦或者，若恰好僅有兩組 比例式 ( R )、對立性質的資料，則泡泡圓餅圖  $C \times R \times R$  恰好適合此類型的資料：



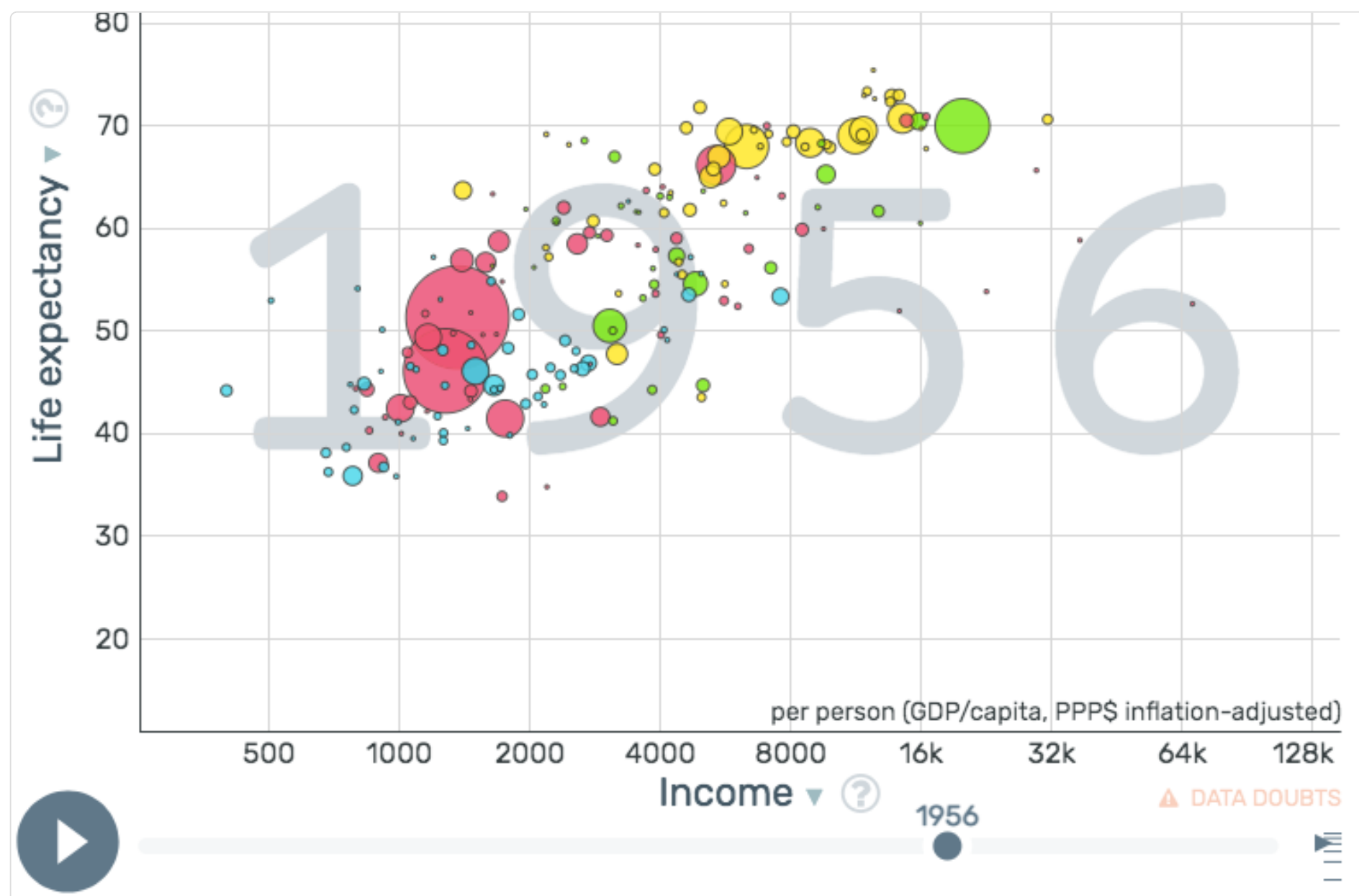
舉例來說，性別 ( 男 / 女 ) 為一種對立性質，我們可以依曲風 ( C ) 將個曲風男、女歌手之營收金額做聚合，便可實現以下對應：

- C - 曲風，每個泡泡代表一種曲風
- R - 特定曲風之男歌手貢獻營收
- R - 特定曲風之女歌手貢獻營收

同時間泡泡大小則對應到各曲風之總營收。



若必須要在泡泡圖中表現**時序資訊 (T)**，我們可以使用額外的時間軸元件，此種視覺化方式即為相當有名的 "[Gapminder](#)" 動畫泡泡圖視覺化：

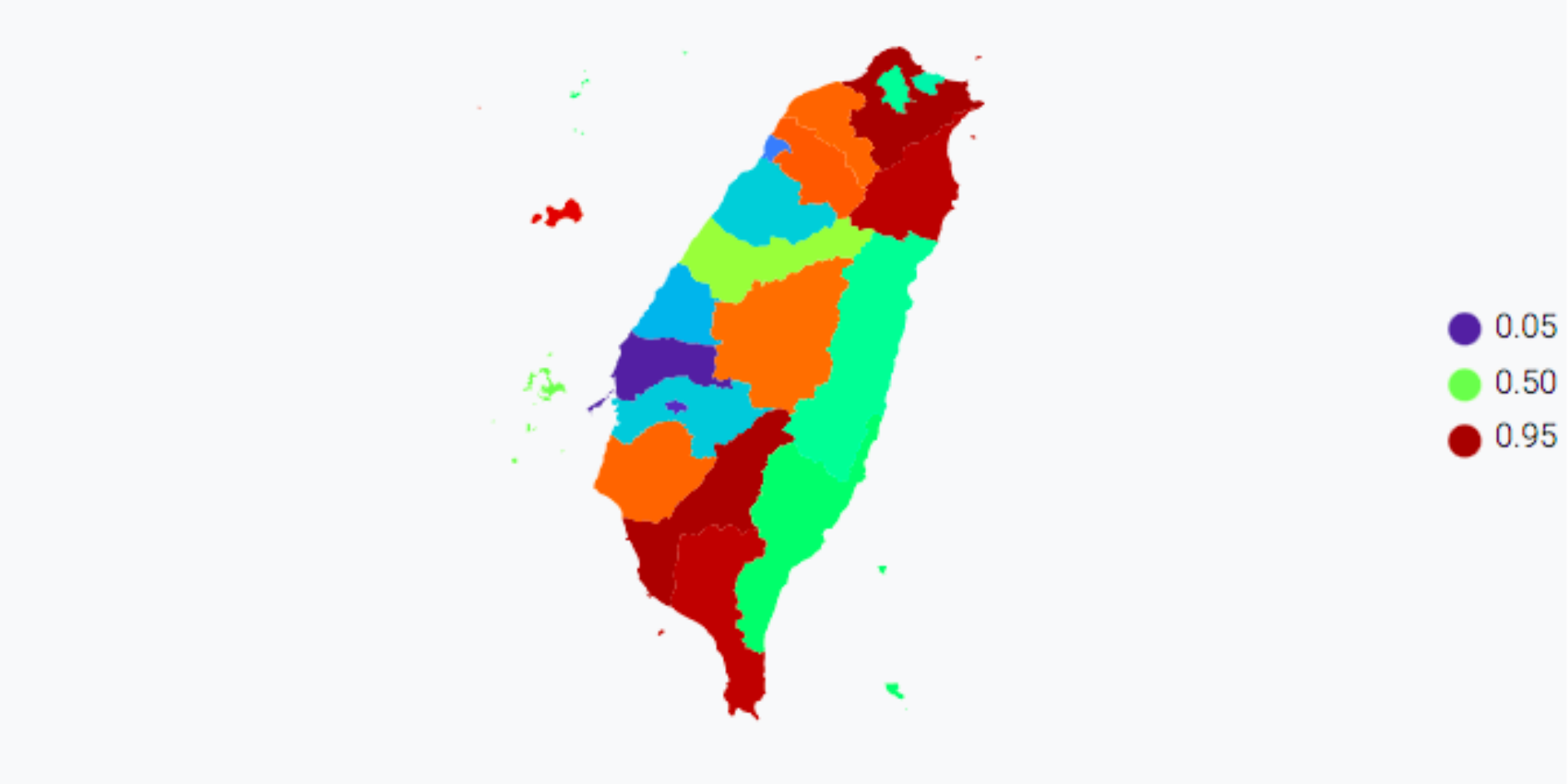


Gapminder Animation Tool, from *Gapminder*

然而，由於儀表板本身已帶有特定時段資訊，我們不太容易在圖表中再次置入時間軸動畫，因為這樣可能會導致混亂。較有可能的是在用戶所選定的時段中做動畫輪播，這可視實際資料的需求決定是否我們有必要用這種方式呈現資料。

熱度地圖

當我們只考慮當期數據時，不會用到 **T (時段)** 資料，資料標記為 **N x C x R**；考慮使用多組分類的情況，亦可能是 **N x C x C x R** 或 **N x C x C x C x R** 等。此類資料一般適合使用樹狀圖 (Treemap) 來呈現，亦可考慮其中分類的進階特性來選擇圖表，例如當分類為地理資料 (縣市、鄉鎮) 時，我們可以使用台灣熱度地圖 (C x R) 視覺化：



由於熱度地圖使用顏色表示各分類 (C) 的數值 (R)，因此我們必須將 **N x C x R** 透過聚合轉換 (Aggregation) 變成 **C x R** 才能適用，並透過使用者介面提供 **名稱 (N)** 的切換機制。這個情況下，無法有效的表現出排名的概念，但可以做為觀察個別名稱 (專輯、歌手) 的分類分布時使用。

營收狀況

在營收部份，我們重視的通常是 **金額 (R)** 與 **交易量 (R)** 的 **走勢或變化 (T)** 狀況，也會考慮不同分類 (C) 的營收來做分組對比。由於消費模式是有週期性的，我們也常會在前後區間做對比 (比方說 今年與去年的營業額對比)。

在這個前提下，營收狀況的分類標記可寫為

時段

分類

金 額

交 易 量

T

x

C\*

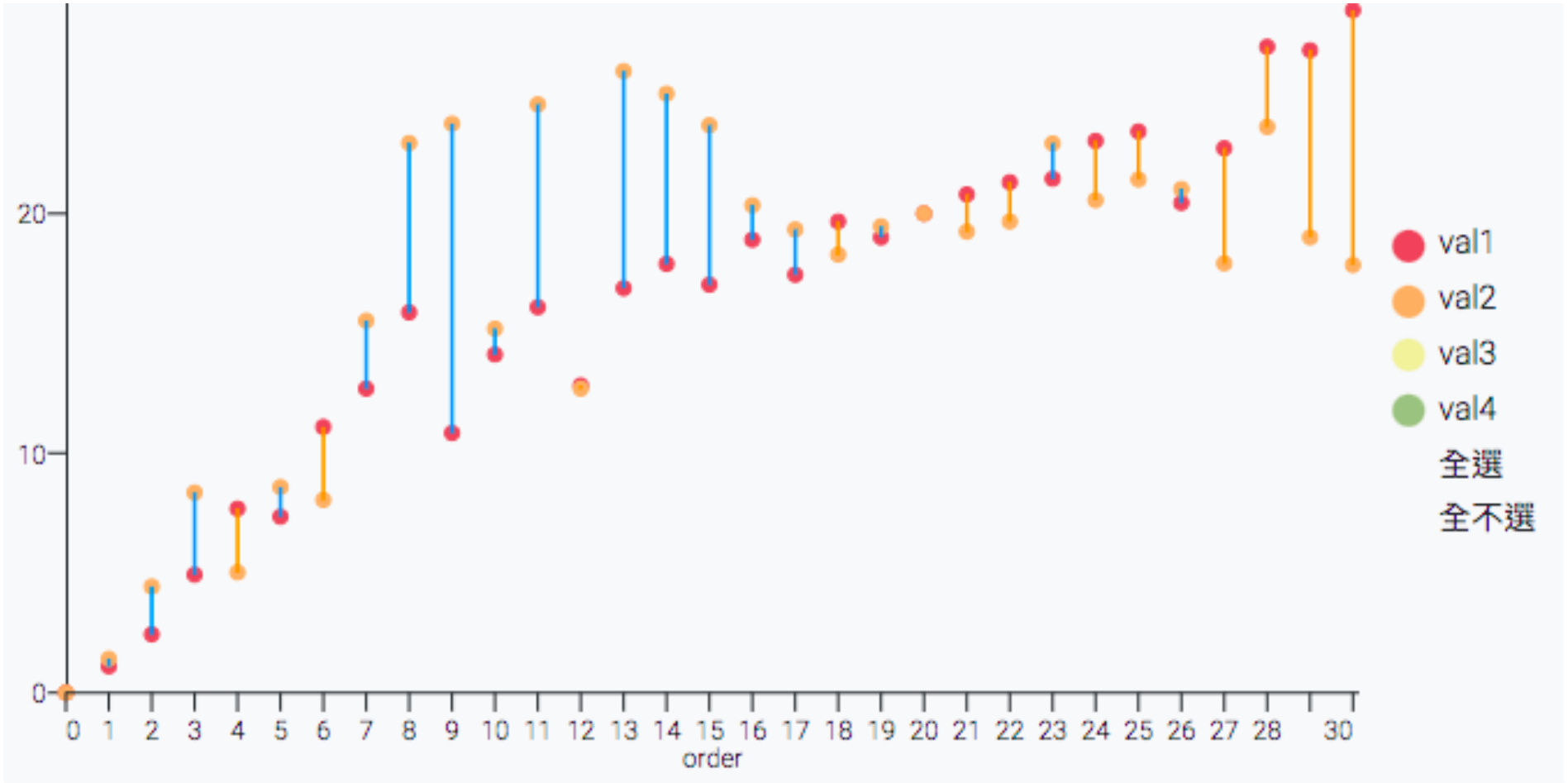
x

[R x R]

x

[R x R]

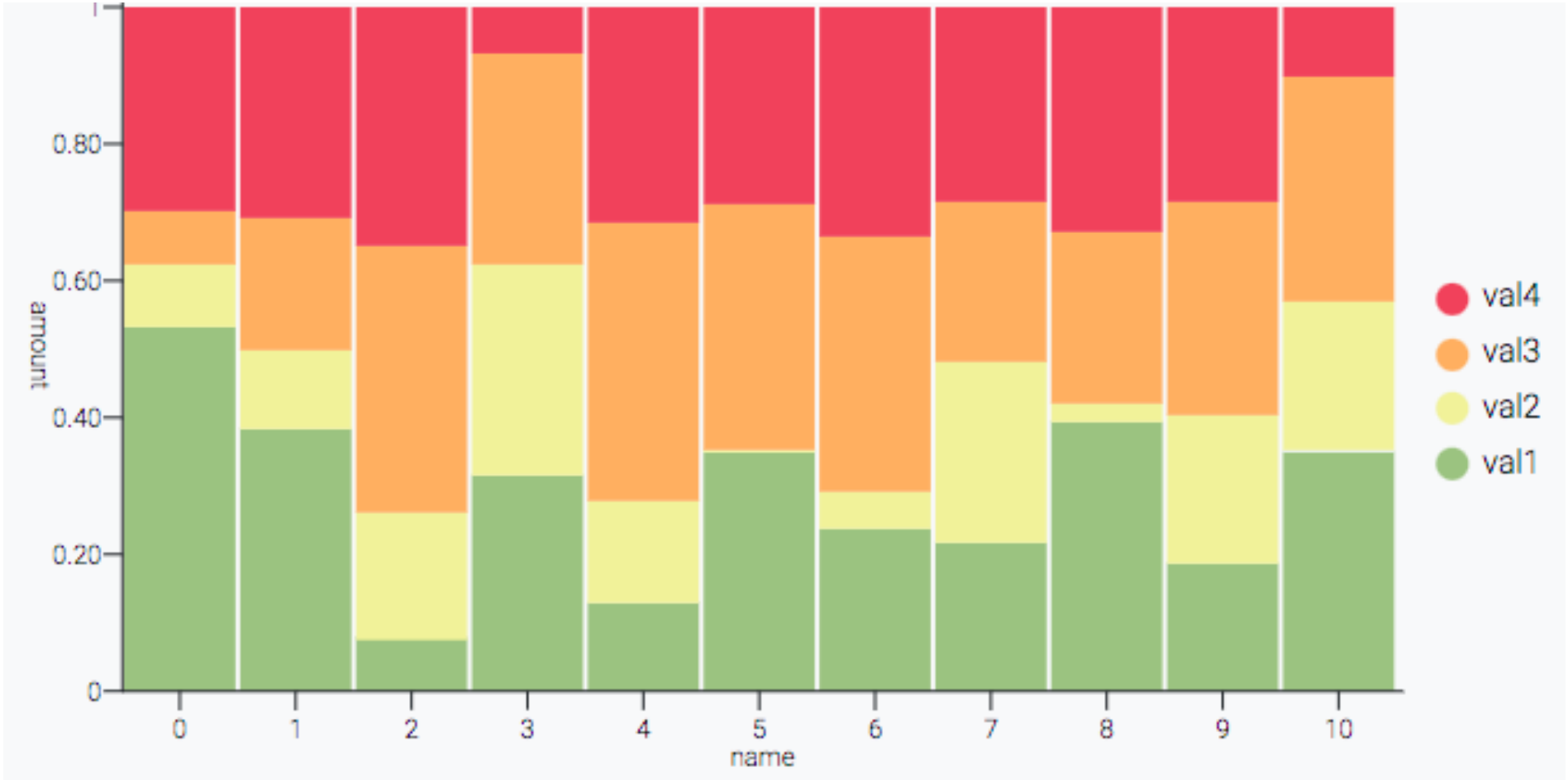
這類型的典型圖表即為 折線圖 或 組合圖 (Combo Chart, 合併折線與長條圖以呈現金額與數量)。在考慮呈現 **變化** 的部份，我們可以使用啞鈴圖 (Dumbbell Chart, < **N x R x R** >) 強調變化的數值：



由於啞鈴圖主要僅使用一對比例式欄位 ( $< R \times R >$ )，在使用前我們必須針對 **時序欄位 (T)** 做 **叢聚轉換**，並試情況提供選單元件處理 **分類 (C)** 資料。

## 用戶數

如同用戶數，我們通常關心的是人數的成長、同時段跨年度的變化比較、使用時間等等，這類型需求透過經典圖表即可以處理。若需進一步分析，用戶還可以透過性別、地理位置、留存率等不同的分類方式來觀察，這些需求通常都需要專用的圖表來處理，但在不考慮資料語意的情況下，我們一般使用 **堆積長條圖** 或 **100% 堆積長條圖** 即可初步呈現資料：



## I 儀表板設計

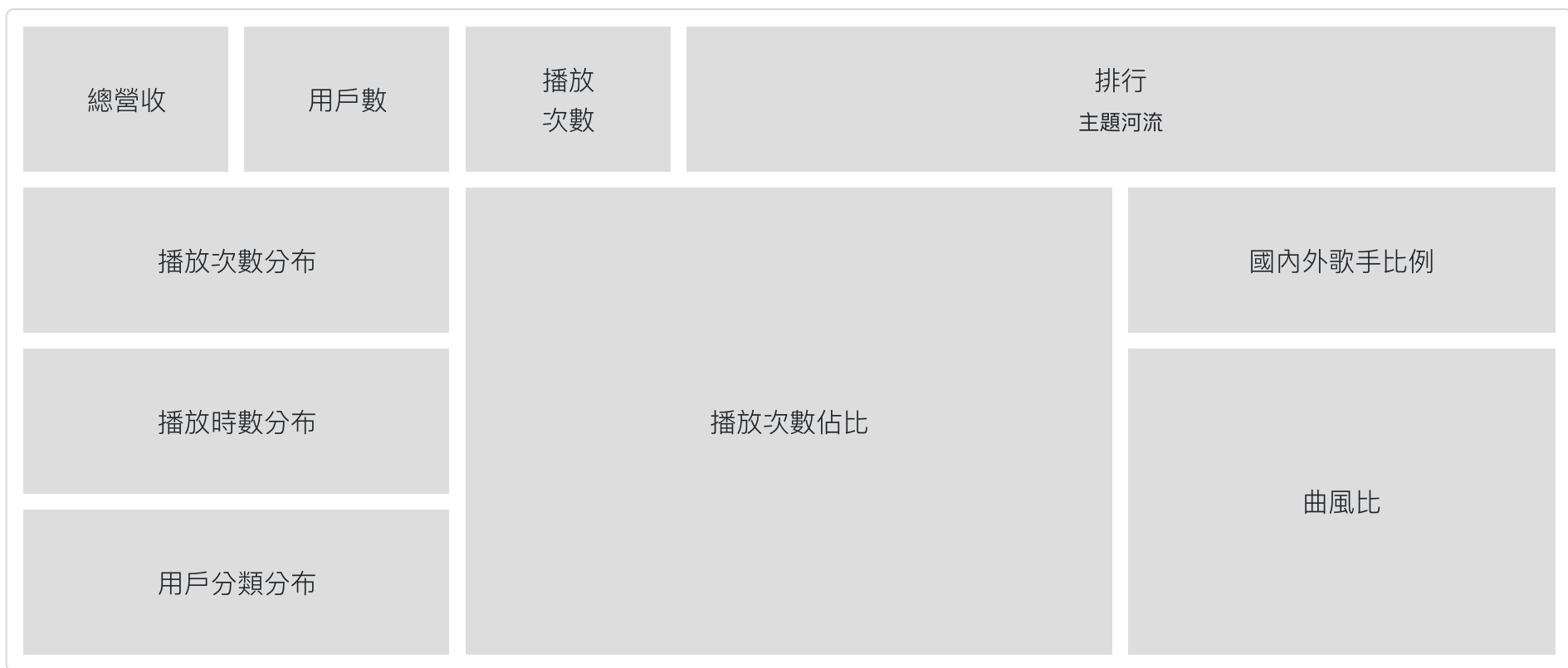
以 排行榜、營收狀況 與 用戶數 三者為主要指標、同時參考先前的討論，我們可以歸納出以下幾個儀表板設計方向：

- 提供主要指標 ( 營收、用戶數 ) 統計數據
- 提供歌手、樂曲的排行資訊
- 依歌手 / 樂曲、用戶組成提供統計資訊
- 基於組成資訊提供進一步的與過濾功能

以下為建議使用的儀表板欄位：

- 總營收 ( 純數字 )
- 活躍用戶人數 ( 純數字 )
- 播放總次數 ( 純數字 )
- 歌手播放次數排行 ( 主題河流 )
- 所有歌手播放次數佔比 ( 互動式比例長條圖 )
- 播放次數分布 ( 直方圖 )
- 播放時數分布 ( 直方圖 )
- 用戶年齡層 ( 或其它分類 ) 分布 ( 直方圖 )
- 播放曲風比例 ( 圓餅圖 )
- 國內外歌手播放次數比 ( 100% 堆積長條圖 )

為降低使用者操作儀表板時的複雜度，我們以電影產業儀表板的版面做為藍本，在排版上將相似機能的面版設置於同樣的位置：



## 獨立書店數據

在獨立書店數據部份，我們將以友善書業提供之獨立書店資料為例。

## 資料分析

雖然圖書出版、銷售之數據變化性可能遠多於友善書業提供之資料，但由於我們仍需考慮到實際所能取得之資料，因此在目前已有資料的前提下，我們將優先考量現有資料的分析，並以此為依據提供儀表版設計建議。

友善書業提供獨立書店數據包含以下內容：

### 僅最新資料

各縣市書店量 (C x R)

- (C) 縣市別
- (R) 書店量

### 逐年更新 (T x C x R)

社員資訊: 逐年 / 縣市別入社統計

- (T) 入社年度
- (C) 縣市
- (R) 入社數量

### 逐季更新

出貨書目 (亦有年度統計表格)

( [T x T] x N x C x < R x R > x R x R x < O x O > x O x O )

- (T) 年
- (T) 季
- (N) 書名
- (C) 出版社
- (R) 出貨量
- (R) 退貨量
- (R) 客訂量
- (R) 活動量
- (O) 出貨排名
- (O) 退貨排名
- (O) 客訂排名
- (O) 活動排名

### 逐月更新

書訊筆數 (T x R)

- (T) 年 / 月
- (R) 書訊建立筆數

#### 交易供應商數量 ( T x R )

- ( T ) 年 / 月
- ( R ) 交易供應商數

#### 交易出版社數量 ( T x R )

- ( T ) 年 / 月
- ( R ) 交易出版社數

#### 書目出貨種類 ( 亦有年度統計表 ) ( T x R )

- ( T ) 年 / 月
- ( R ) 書目出貨種類

#### 銷貨數據 ( 亦有年度統計表 ) ( T x R x R x R x R x R x R x R x I )

- ( T ) 年 / 月
- ( R ) 銷貨收入
- ( R ) 銷貨成本
- ( R ) 銷貨退回收入
- ( R ) ( 銷貨收入 - 銷貨退回收入 )
- ( R ) 銷貨退回成本
- ( R ) 實際銷貨毛利 ( 銷貨收入 - 銷貨成本 - (銷貨退回收入 - 銷貨退回成本) )
- ( R ) 銷貨數量
- ( R ) 銷退數量
- ( R ) 實際銷貨數量
- ( I ) 實際銷貨毛利率(%) ( 毛利 / ( 銷貨收入 - 銷貨退回收入) )

### 資料解析度問題

互動儀表板為了實現過濾機制，需要較細節且詳盡的資料；例如，銷貨數據尚須書目、書店、出版社等欄位，我們才能進一步統計各別書目、書店的銷貨數據。友善書業目前所提供的數據中，銷貨數據與出貨書目是獨立的兩組資料集，因此我們無法同時分析營收與銷貨量，也沒辦法依照書店來統計營業額。

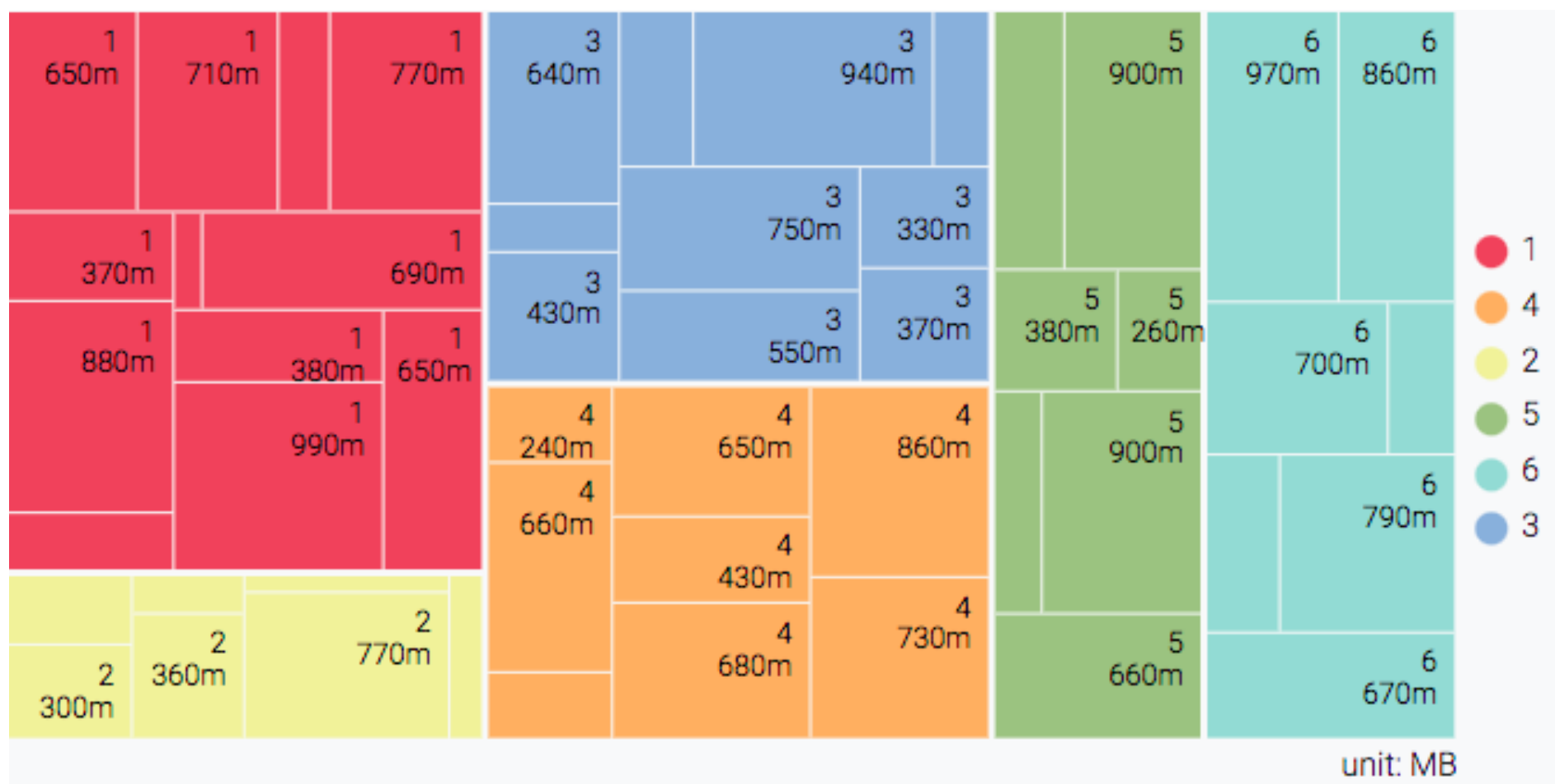
我們仍能夠假想有一組這樣的資料集進而做為設計儀表板的依據，然而在取得更詳細實際數據之前，我們僅能進行概略的設計，在獲得更完整數據之後再依實際狀況做設計調整。以下為我們假想的詳細數據，資料至少以月為單位收集：

- ( T ) 年/月
- ( N ) 書店
- ( C ) 書店所在縣市
- ( C ) 出版社
- ( N ) 書名
- ( R ) 出貨量
- ( R ) 退貨量
- ( R ) 客訂量
- ( R ) 活動量
- ( R ) 銷貨收入
- ( R ) 銷貨成本
- ( R ) 銷貨退回收入
- ( R ) 銷貨退回成本

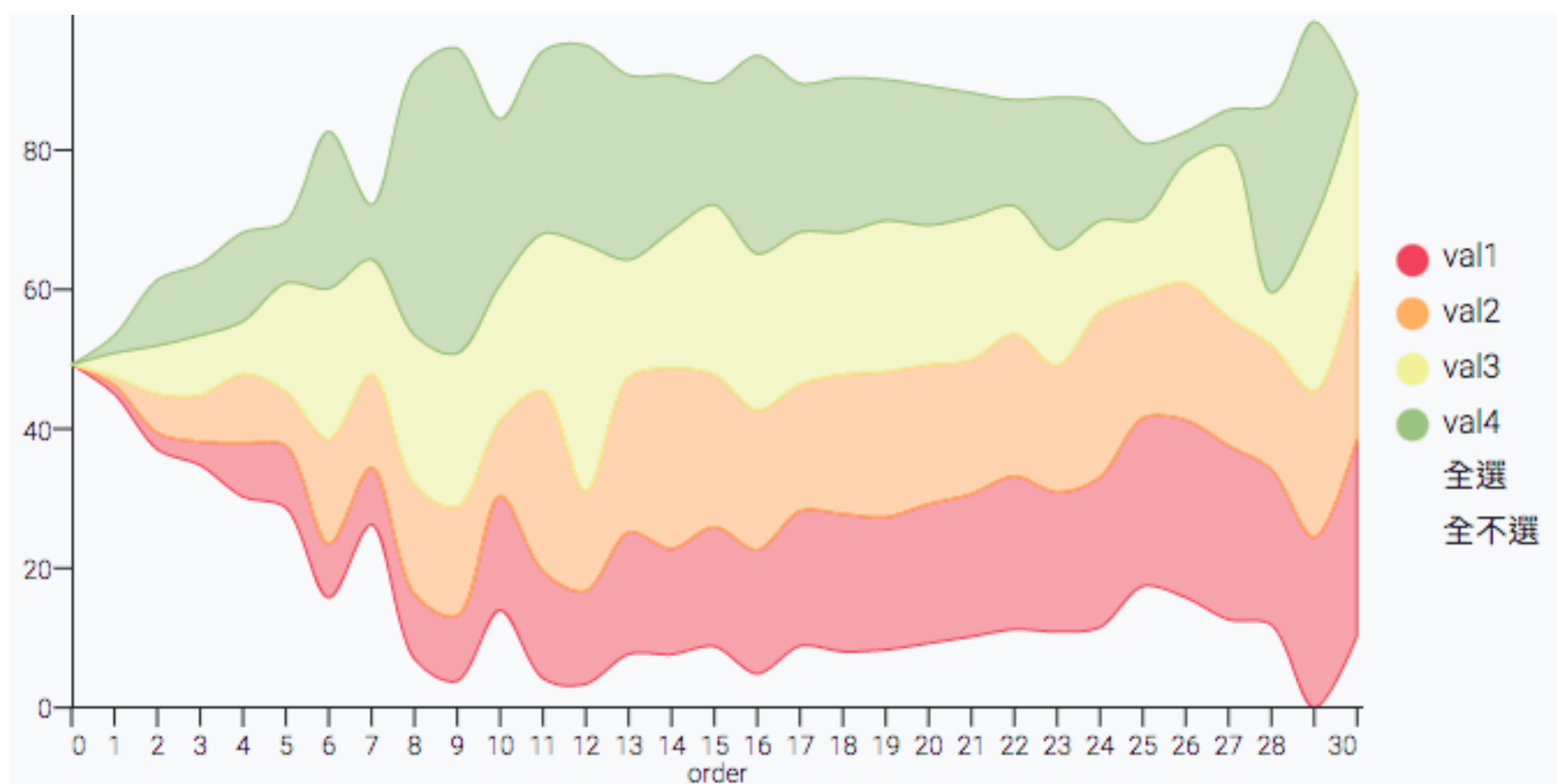
## 視覺化

獨立書店數據以時序資料 (T x R) 為主，輔以縣市分區社員統計 (T x C x R)，以及較為細節的出貨數據 (出貨書目 / 銷貨數據) 等。縣市別入社統計、書訊筆數、交易供應商數量、交易出版社數量、書目出貨種類等數據互相獨立，無法對應儀表板交互過濾機制 (Cross Filter) 實現資料探索，僅適合單獨使用折線圖或長條圖方式呈現。(縣市別入社亦可使用熱度地圖表現)

出貨書目、銷貨數據有較細節之資訊，其中出貨書目較具有變化性，可單獨對出版社統計、對比出退貨量、亦可對書名做個別統計及排名。出版社統計可以使用樹狀圖呈現 (樹狀圖分類標示為 C\* x R，不包含時序資料，因此需捨棄時間資訊)：



亦可使用主題河流，用以表現 獨立書店青睞的出版社排名：





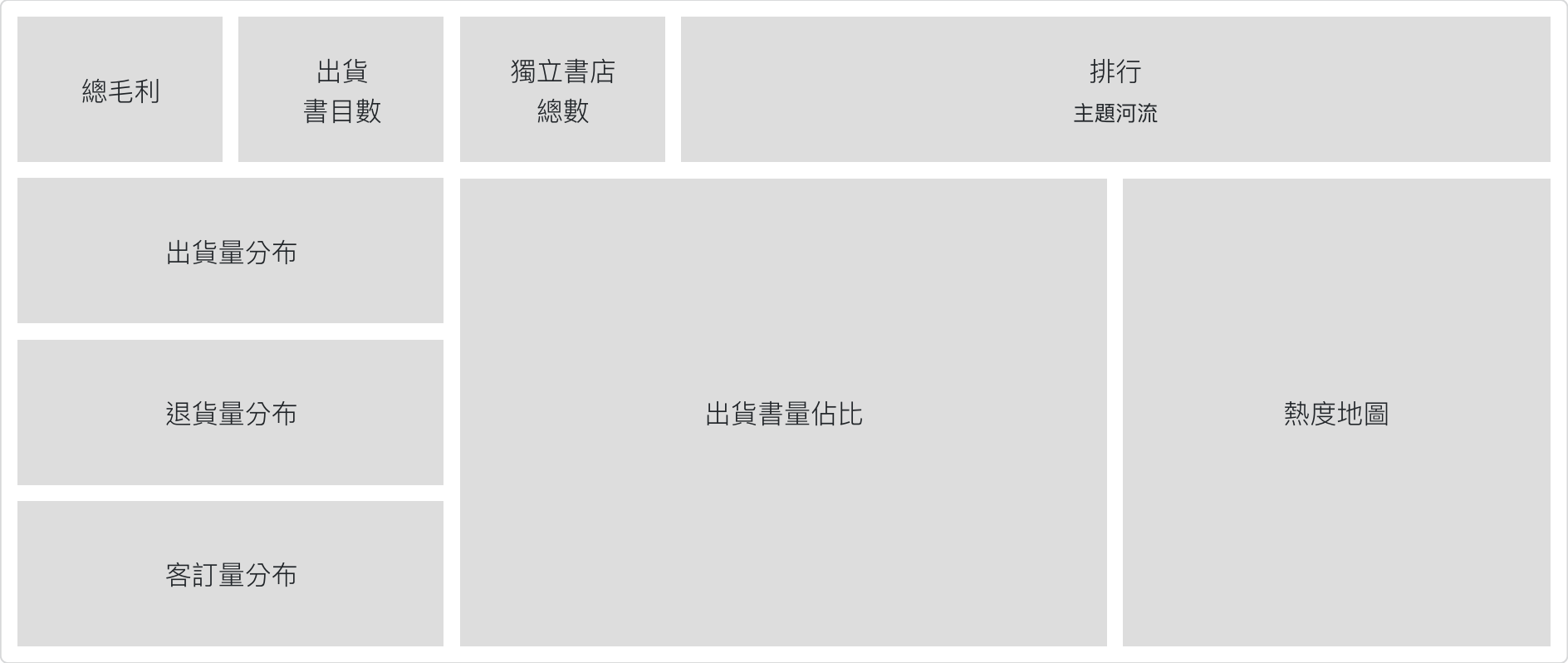
考慮到獨立書店眾數據間獨立性較高，我們亦可考慮多觀察維度較多的出貨書目資料來構成儀表板，以實現交互過濾機制。

## 儀表板設計

以下為建議使用的儀表板欄位

- 總毛利 ( 純數字 )
- 出貨書目數 ( 純數字 )
- 獨立書店總數 ( 純數字 )
- 前十大出版商排名 ( 主題河流 )
- 各書目出貨書量佔比 ( 互動式比例長條圖 )
- 出貨量分布 ( 直方圖 )
- 出貨量分布 ( 直方圖 )
- 退貨量分布 ( 直方圖 )
- 各縣市書店量 ( 熱度地圖 )

類似線上影音數據的設計考量，為降低使用者操作儀表板時的複雜度，我們以電影產業儀表板的版面做為藍本，在排版上將相似機能的面版設置於同樣的位置：



考慮現有資料並非統一資料集，以此資料集做設計時無法實作有效的過濾機制，因此我們亦考慮提供非連動式的儀表板，單純做為資料瀏覽之用。為避免用戶在使用時有混淆的情況，我們得依情況將「出貨量分布」、「退貨量分布」、「客訂量分布」圖表改為折線圖呈現近期走勢。

## 參考設計

基於原先儀表板之設計以及後續討論之調整建議，我們提出以下幾個設計擴充，可供未來增強儀表板功能使用：

- 時段對比  
選取特定時段時，同時在儀表板中將該時段與過往同期時段做數值對比
- 進階檢視  
透過介面轉換檢視更複雜的細節資料
- 切換資料欄  
利用選單、按鈕機制供用戶切換同類型的不同資料欄位

## 時段對比

對某些資訊來說，我們可能更感興趣的是數值相對於過往同一時段的變化。要表現此類資訊，我們可以透過：

- 調整圖表的呈現方式，以提供「變化」「對比」的機制
- 在各項圖表的標題上，額外提供變化比例的數值

以下表為例，我們除了擴充數字方塊以呈現前期資訊外，亦在標題列額外表現數值的漲跌：

總毛利	+1.5%	出貨書目數	-2.3%
1.387B		68.23M	
前期: 1.366B		前期: 69.83M	

除了圖表外，我們亦需要對時段選擇器做額外的調整，至少提供以下機制：

- 時段對比功能是否開啟
- 想要對比的對應時段

我們可以在時段選擇器中追加額外的控制項來實現這個機制（如右圖）；因為一般都以相同長度的時段做對比，因此我們僅需要提供對比時段起點做選擇以做為對比的基準點，或者亦可以提供快選功能（如「前一年」、「前一季」等選項）。

全部時段

最近一個月

時段起點

2021/02/21

▼

時段終點

2021/08/22

▼

啟用時段對比

☒

對比時段起點

2020/02/21

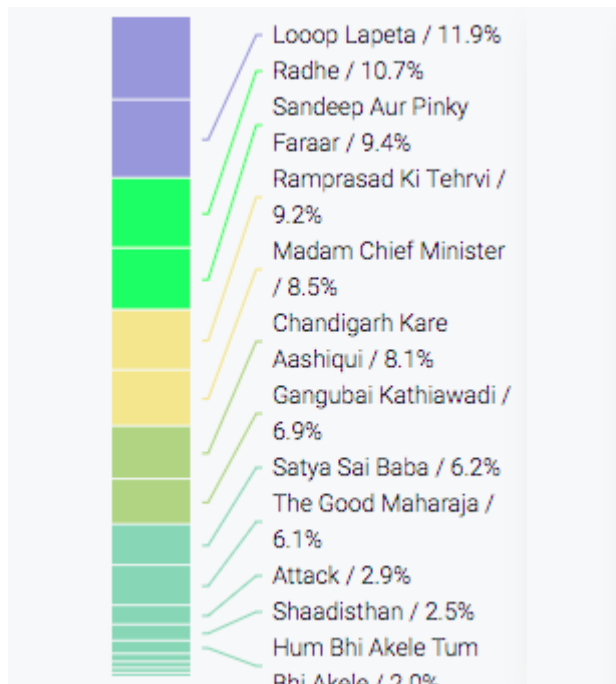
▼

## 進階檢視

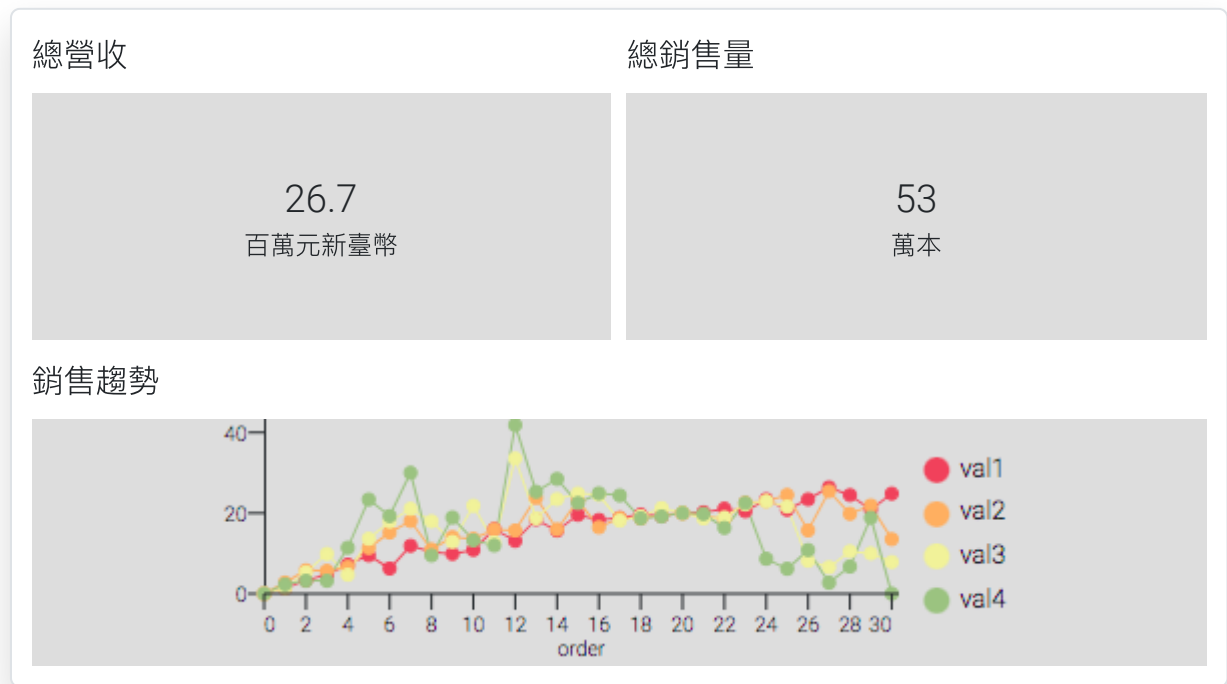
儀表板受限於空間限制無法呈現細節資料或複雜圖表，這時我們可以採用互動形式透過點選、切換的方式進入「進階檢視」以瀏覽進一步的資料細節。

進階檢視所呈現的依圖表而異，需要個別按設計跟規畫，此處我們以「排行榜」做為範例，點選方塊後即彈出進階檢視：

銷售排行



點擊左側方塊後彈出以下結果



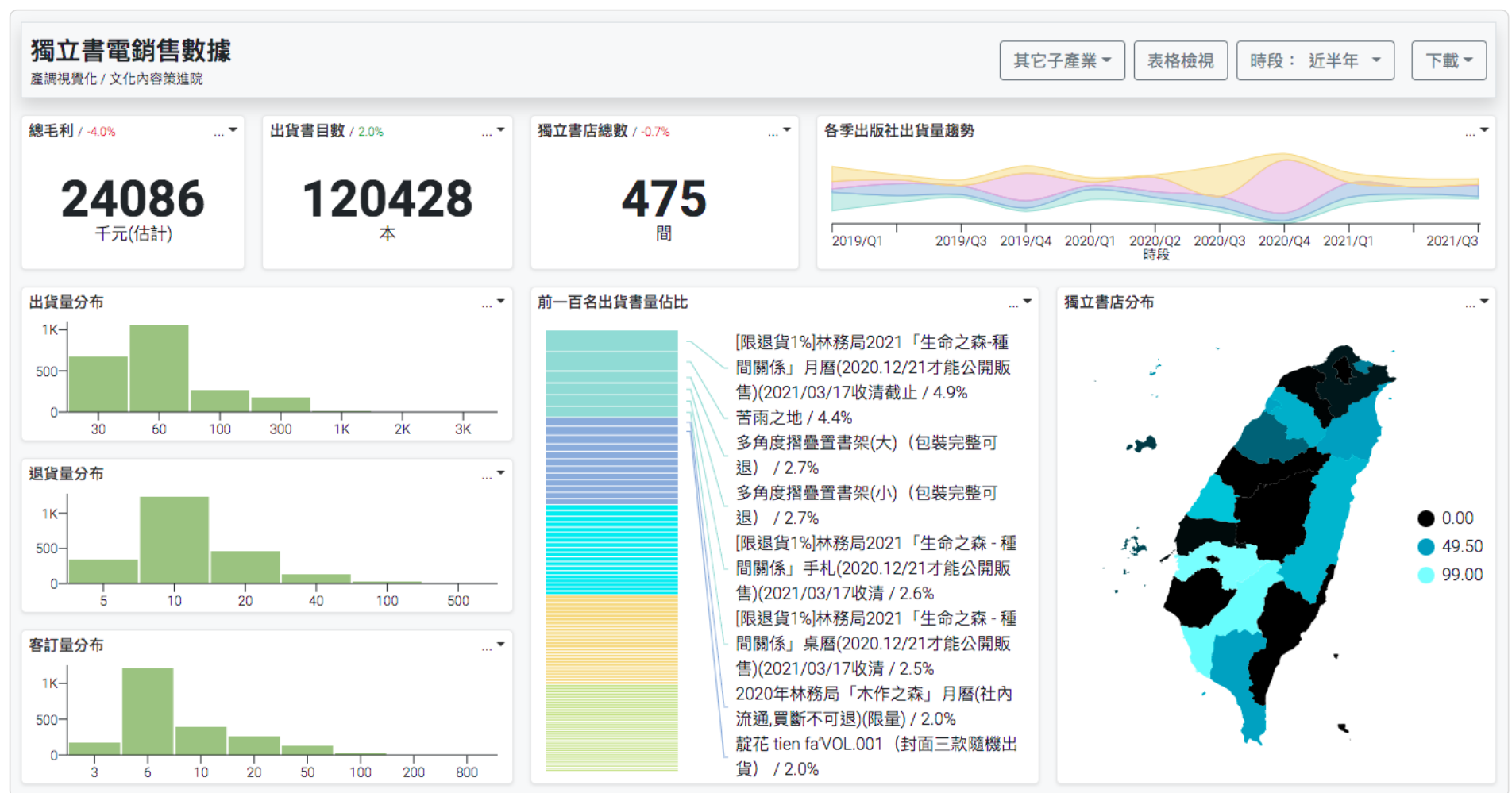
## 切換資料欄

當多組相似數值因空間問題僅能表現一組時，我們可透過額外的控制面板來實現切換，例如使用前文提及之選單元件或時間軸元件。下圖即為使用同一塊面板表現多組數據之方式：



## 範例

由於我們僅有獨立書店的數據，此處我們提供一組依前文設計規劃之獨立書店數據儀表板範本；線上影音可以此範本做為概念引申。



# 文創產調視覺化使用手冊

產調視覺化與主題產業數據儀表版使用說明與簡介

## 簡介

文創產調視覺化 ( 後稱本視覺化) 旨在提供文創相關產業各種數據之視覺化檢視介面，提供更直觀、便利使用的資訊管道。本視覺化主要分為三個面向：

### 產業概況

以掌握文創產業之概況為主，依產業別大方向式的呈現主要產業指標，目標在快速了解整體產業之樣貌。

### 詳細數據

以深入探索產業特有的資料為主，從產業結構、內外銷、從業人員組成等，切分成多組數據各自以單一圖表呈現。

### 即時資訊

從各項即時的領域數據 ( 目前以電影票房為主 ) 了解產業最新脈動，並透過互動提供分項探索的空間。

## 產業概況

文化內容產業概況即為本視覺化的主要入口頁面，提供產業別大方向式的呈現主要產業指標。

**數據方面**，本視覺化提供四大產業中四種跨年度的主要產業數據供參考；**介面方面**，我們使用樹狀圖等四種圖表搭配可客製的介面操做供您做簡單的資料探索。

接下來我們則分別就「**數據**」與「**介面**」角度為您介紹此「**產業概況**」項目。

## 數據

產業概況項目以概況為主，呈現包含 **影視**、**出版**、**流行音樂**、**遊戲** 等四大產業中的 **產值**、**出口值**、**就業人數**、**家數** 等四組數據。各產業數據以年度為單位統計，但由於僅 **影視** 部份有較長期的資料，因此我們仍以單年度數據為呈現重點。

為有效率的呈現數據，我們設計了包含樹狀圖、泡泡圖等五種不同的資料檢視模式，並提供產業別分組調整的功能，供您快速框列感興趣的產業子類別。



## 介面

就介面來說，產業概況分成三個面板：



各種功能選項，如圖表切換、表格檢視、詳細檢視等。



顯示細節並供切換數據種類、年度；亦為客製設定的面板。



主要用於顯示當前數據對應的視覺化互動圖表。

下面將針對三個不同面板分別做簡單介紹。

## 資訊面板

此面板主要為呈現數據細節與調整設定之用途。

如右圖所示，資訊面板預設顯示產業總值與子產業之統計，並提供切換「產值」、「出口值」、「就業人數」與「家數」等四大指標之功能。圖表面板將依照您的選擇即時呈現對應的數據。

預設數據顯示為最新年度 (圖中為 108 年度)，但可點擊年度觸發**年度選單**，切換至其它年度。請注意部份分類沒有歷年數據，沒有數據的分類將直接忽略。

除重點數據外，本面板亦會依您的操作顯示其它不同的內容，例如**圖表客製**或者**各子分類細節**，相關細節我們將於接下來的內容進一步做介紹。



## 導覽列



如上圖之導覽列截圖所示，導覽列提供多種不同的圖表、表格檢視與設定項目供您使用，以利您從不同面向檢視資料。一般來說，您會想先在資訊面板選定想瀏覽的數據類型後，再於此選擇感興趣的檢視模式。

導覽列依功能區分，提供以下選項：

- **切換檢視** - 供您在五種檢視中做切換，以不同角度觀察圖表。
- **設定** - 客製化圖表繪製 ( 包含顏色、字體尺寸等 )
- **下載** - 將當前檢視中的圖表或資料另存成本地端檔案
- **詳細統計數據** - 進入「詳細數據」瀏覽頁

接下來我們分別就這四個選項做介紹。



※ 切換檢視

透過切換檢視，您可以變換圖表的檢視方式。我們目前提供四種不同的圖表與一個表格檢視，分別包括：

樹狀圖	泡泡圖
樹狀圖利用面積表現數據，並透過顏色與方塊呈現兩級式的分層資料，方便我們同時比較各分類與子類的數值大小。	泡泡圖類似樹狀圖，但改以泡泡呈現，相較於樹狀圖來說更趨向各別子分類間的比較。
折線圖	排名線圖
折線圖同時呈現各子類別數值的歷年走勢，主要用來比較跟觀察各數據間的消長與變化。	排名線圖以折線圖為基礎，但以強調排名為主，在資料包含極端數值時能避免數據混雜的問題。

此外我們亦提供表格檢視附帶基本基本搜尋功能，呈現完整的原始數據檔案，供您直接在線上查看原始資料之用。

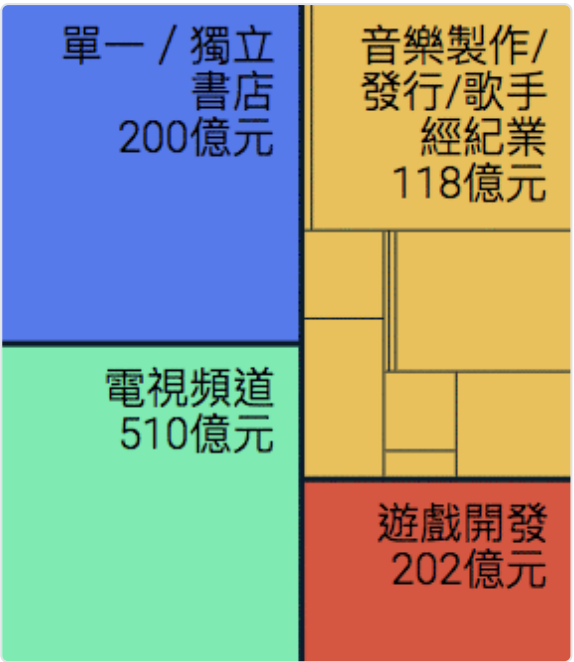
※ 設定

設定項目供我們做簡單的圖表客製化。右圖即為設定面板的預覽圖。

點擊「設定」後，「設定面板」會覆蓋並顯示於「資訊面板」之上。如右圖所示，目前設定面板提供三項功能：

字體大小

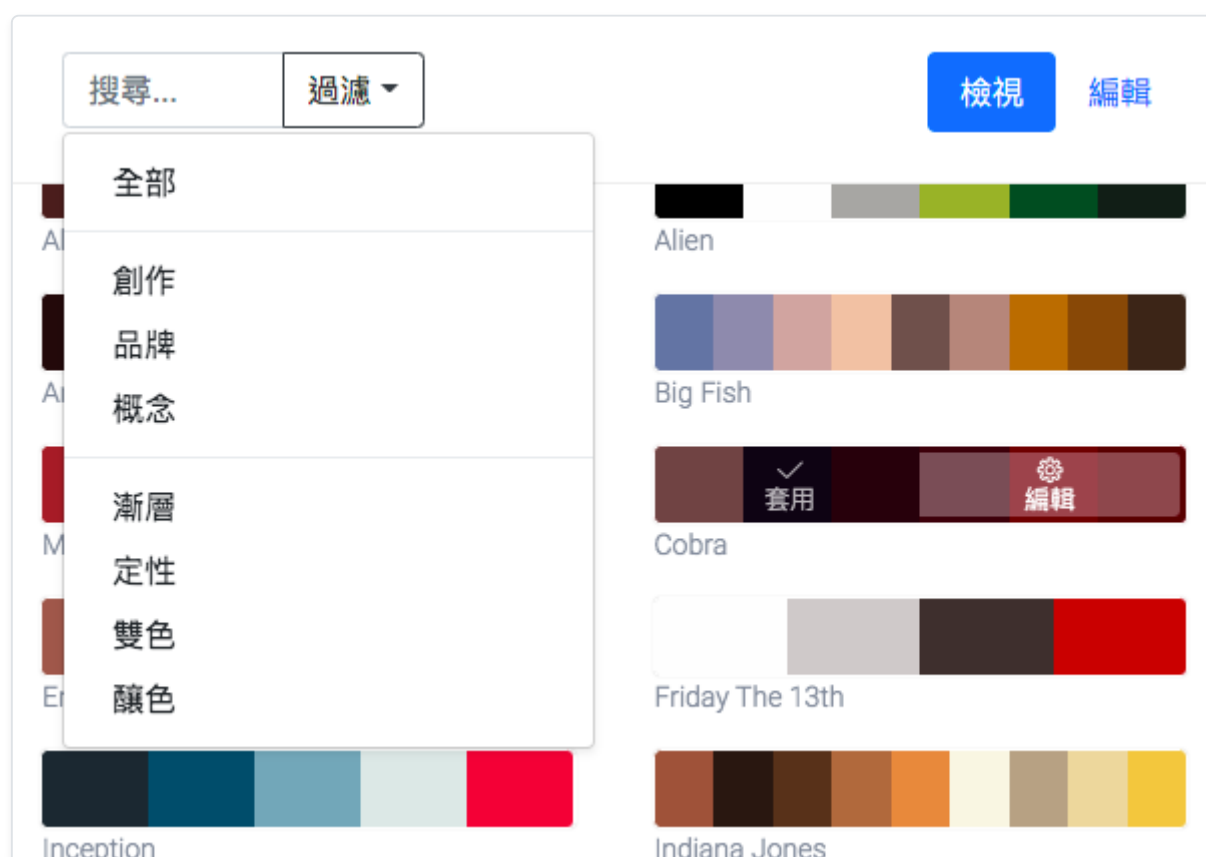
視覺化圖表有時會有因為空間限制無法顯示文字的情形，這時您可以透過調整字體大小來改變繪製的結果。右圖展示了樹狀圖在兩種不同字體尺寸時其中的文字效果。



配色

調整圖表的配色。我們提供 1085 種包含文策院色盤等不同的預建色盤，並依色盤屬性分類供您挑選。點選「配色」色盤後，即彈出色盤選擇器，此選擇器並預建色盤編輯器，您可隨意客製任何色盤，只要在選取色盤時點選「編輯」選項即可。

下圖即為色盤選擇器，除了可看到畫面中的十來組色盤外，下拉式選單亦提供分類篩選功能。其中「Cobra」色盤處於選擇狀態，點擊其上的「編輯」鈕即可進入該色盤的客製模式。



## 下載

若您想利用生成的圖表於它處，您可以依照下載選單中的不同檔案格式利用下載功能取得視覺化圖檔。我們目前提供以下幾種格式：

- SVG - 向量圖，可用於做圖表後製編輯
- PNG - 點陣圖，方便直接嵌入使用
- CSV - 原始資料檔，分「完整資料」與「客製檢視資料」。關於客製檢視，請參考後文說明。
- ZIP - 圖檔 + 資料檔之壓縮檔

## 客製分組

客製分組讓您重組各項子類別，合成為單一項目。例如，若您想將出版與影視的行銷相關產業自成一組，即可使用客製分組功能。

下圖為客製分組面板的長相：

The screenshot shows the '客製分組' (Custom Grouping) panel. At the top, there are three toggle switches: '啟用' (Enabled) which is turned on, '顯示未選取的分類' (Show unselected categories) which is turned off, and '顯示客製分組的細項' (Show details of custom groups) which is turned on. The main area is divided into two sections. On the left is a sidebar with a blue header '分組一' (Group 1) and a red 'X' icon. Below the header is a text input field labeled '新增...' (Add...). At the bottom of the sidebar are '確定' (Confirm) and '取消' (Cancel) buttons. The right section contains a grid of category buttons. The top row has three buttons: '圖書出版 / 圖書出版', '漫畫出版 / 漫畫出版', and '雜誌出版 / 雜誌出版'. The second row has three buttons: '數位出版 / 數位平臺', '原創圖像 / 插畫經紀', and '出版行銷通路 / 單一 / 獨立書店'. The third row has two buttons: '出版行銷通路 / 連鎖書店 (總公司)' and '出版行銷通路 / 網路書店 (狹義)'. The fourth row has three buttons: '出版行銷通路 / 二手書店', '電視產業 / 電視節目製作', and '電視產業 / 電視節目後製'. The fifth row has three buttons: '電視產業 / 電視節目發行', '電視產業 / 電視內容產業', and '電視產業 / 電視頻道'. The bottom row has three buttons: '電視產業 / 電視平臺', '電視產業 / 網路影音節目製作', and '電視產業 / 數位發行與播送'.

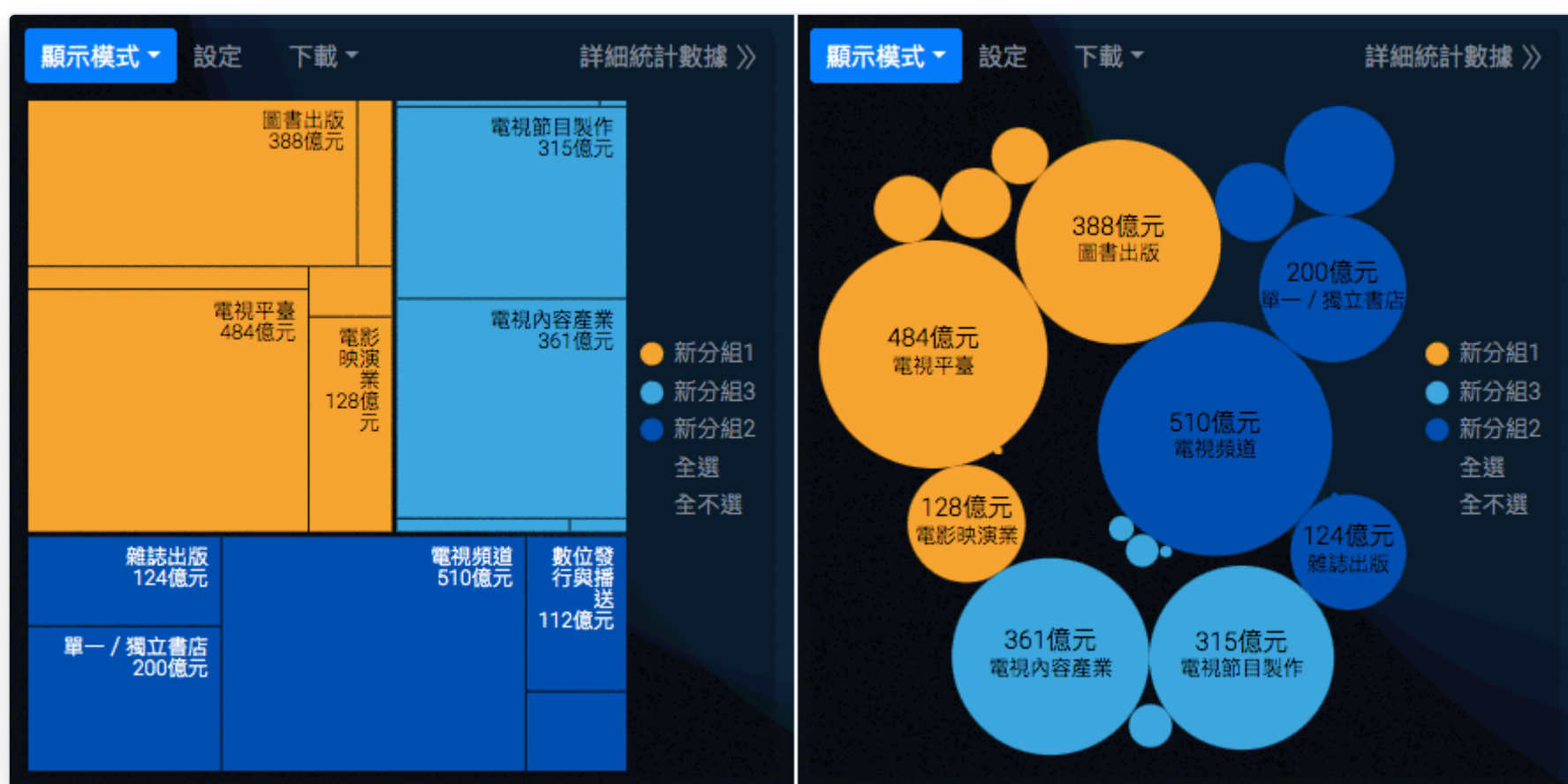
使用方式的簡單描述如下：

- 點擊左方「新增」字樣即可建立新分組。點擊新建立的分組文字，可客製分組名稱。
- 在左方選取欲編輯的新建分組後，可在右方點選想框列的子分類。您亦可拖曳滑鼠做選取。
- 點擊完成，新繪製的圖表即會依照您設定的分組來呈現。

客製分組面板亦於右上角提供三個選項供設定，分別是：

- **啟用** - 可切換是否套用客製分組。
- **顯示未選取的分類** - 若啟用，未框列進客製分組的分類亦會被繪製出來。反之則僅呈現框列的分類。
- **顯示客製分組的細項** - 若啟用，分組細項亦會被繪製出來。反之則僅呈現分組總數據。

下圖為啟用客製分組、將分類切成三組後的視覺化結果。請注意由於折線圖並非分多層分組呈現用的圖表，客製分組在折線圖中可能無法顯示預期的效果。



## ※ 詳細統計數據

點擊「詳細統計數據」即可進入詳細資料頁面。關於詳細資料，請參考後文說明。

## 圖表面板

佔據畫面最大的圖表面板主要功能為呈現視覺化後的數據內容，但您亦可透過與其簡單的互動從而取得更深入的資訊。您所能進行的互動包括：

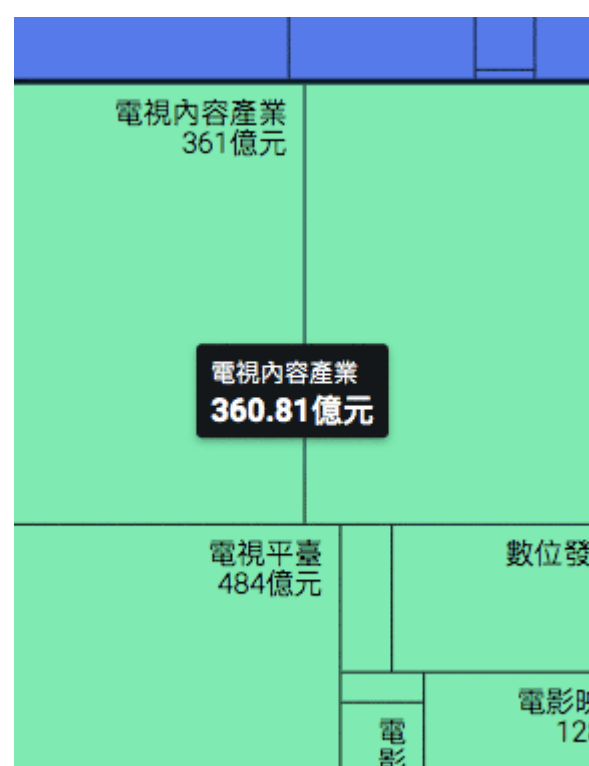
- 滑鼠移入顯示 **資料細節**
- 點擊區塊顯示 **次類別資料**
- 透過 **圖例選取** 功能開關特定資料組

下面我們以實例簡單介紹各種互動的效果。

## ※ 資料細節

視覺化為清楚呈現資料，必須取捨所顯示的數據完整性。就本視覺化來說，有些數據會因為顯示區域過小直接隱藏，而不會顯示在圖表中。您可以透過滑鼠移入各個資料節點來觸發資料泡泡，其中便含有該組數據的完整資料。

如右圖中，滑鼠移至「電視內容產業」即可顯示當前數據的數值 (360.81億元)



## ※ 次類別資料

在「比較」形式的圖表中 ( 包括 **分組比較/樹狀圖** 與 **整體比較/泡泡圖** )，您可透過點擊個別子類別色塊來切換子類別資訊面板，此時子類別資訊將會呈現在左方的資訊面板中。

子類別資訊並沒有任何額外的新數據，但他將單一子類別的資訊匯整顯示，包含當**前年度的四組屬性**，並以折線圖呈現子類別當前屬性之**歷年走勢**。如果特定子類別的當前屬性沒有歷年資料，則以空白表現。

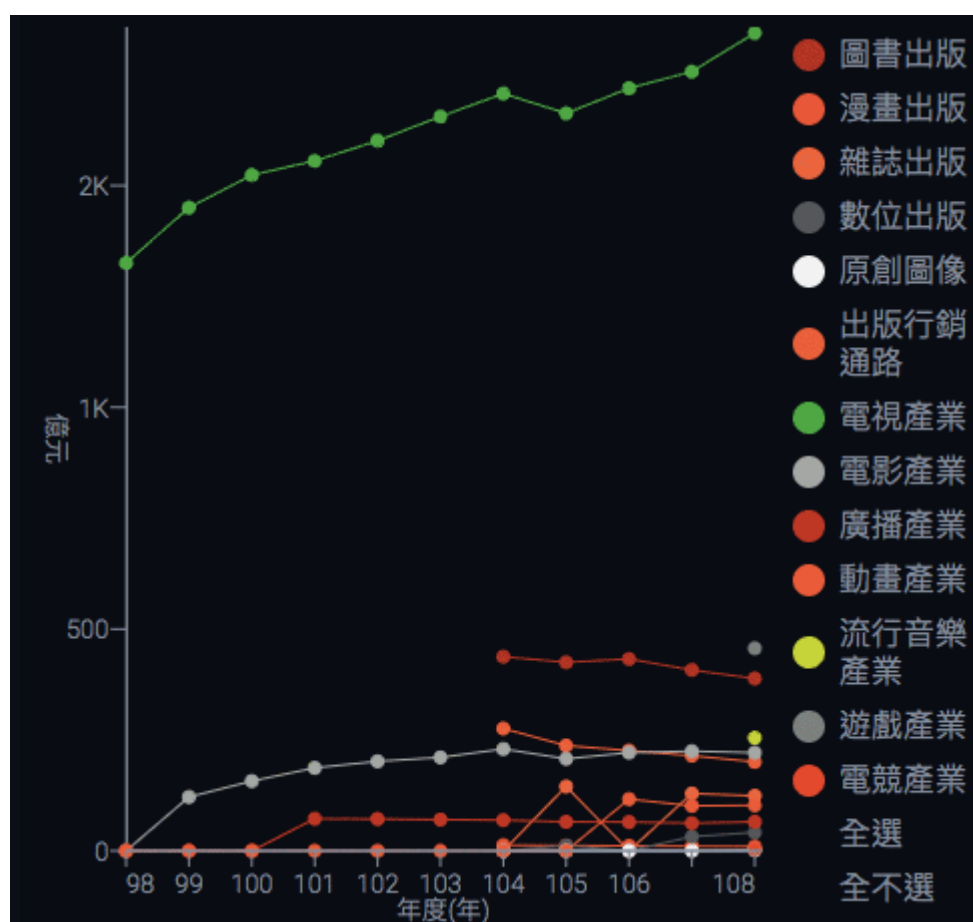
以右圖中的資訊面板為例，所呈現的次類別為「**影視 / 電視產業**」中的「**電視內容產業**」，下方除了四大數據外，亦呈現當前屬性「**產值**」的歷年走勢。



## ※ 圖例選取

文創產業除「影視」、「出版」、「流行音樂」、「遊戲」等四大類外，各類別中亦有許多子類別；我們可以透過點擊圖表中各項圖例來開關各類別的顯示，達到聚焦於特定類別的效果。

以右圖來說，顯示完整資料時，因為電視產業在產值上的領先，其它各產業數值相對受到壓縮而較不易閱讀。我們透過點擊圖例中「電視產業」與「遊戲產業」兩者將最高的兩條數據隱藏，即可將焦點聚集在下方較為密集的資料區域。





# 產業詳細數據

您可經由產業概況頁面右上角的連結進入「產業詳細數據」頁。這裡提供依產業分類、各產業細節資料的數據視覺化，包括如「內容產製數」、「各類型營收比重」、「內外銷結構」甚至如「ISBN 申請量」、「電競戰隊經紀約比例」特定產業之數據等等。

## 數據

與產業概況不同，產業詳細數據頁面並非提供單一資料集數據，而提供了四大產業共 200 餘組的資料集；資料集性質各異，無法統一呈現，因此我們使用清單形式供您選取感興趣的資料集單獨呈現。

## 介面

類似產業概況頁，詳細數據頁面亦分成三個面板：



各種功能選項，如表格檢視、設定、概況檢視等。



提供資料集搜尋與當前資料集的基本資訊。



主要用於顯示當前數據對應的視覺化互動圖表。

下面將針對三個不同面板分別做簡單介紹。

## 資訊面板

產業詳細數據中有相當多的資料可供選擇，而且彼此大多不相關，因此詳細數據頁的資訊面板以呈現資訊清單供選擇為主，並提示用戶目前已經選擇的資料集名稱與分類。

如右圖所示：

- **當前選擇的資料集**以獨立方塊形式顯示於最上方，並標示其所屬的產業別；
- **可選資料集清單**則顯示於下方，以色塊區分產業別
- 提供 **簡易搜尋功能**，方便快速尋找特定資料集。

除資料集清單外，類似產業概況頁面，開啟設定功能時，此處亦會用來顯示設定項目。



## 導覽列



不同於產業概況頁面，產業詳細數據頁面由於資料種類多差異大，因此圖表主要透過程式自動選擇為主。也因此，導覽列中並沒有多組供選擇的圖表，僅有一組預設啟用的「圖表檢視」。

其它部份則類似概況頁面，仍有表格檢視、設定、下載等功能供選用，亦提供返回「產業概況」的快速連結。

## 圖表面板

類似產業概況頁面，在產業詳細數據頁面中的圖表面板亦提供圖例選取與資料細節等功能，但由於各組資料內容不同，我們無法依子類別提供資訊，因此沒有「次類別資料」選單可供瀏覽。

取而代之的則是，當選取的資料本身帶有多重不相容屬性、無法呈現於同一張圖之中時，程式將自動整合同類型的資料，於面板上方表列可供切換的屬性給您做選擇，並依您的選擇再自動生成對應的圖表。

此外，各資料對應的註解則陳列在圖表底端供參。



下圖為「我國108年各類型新書之電子書零售價格及出版家數」之視覺化預覽，可看見上方的三個按鈕分別是「種數」、「價格(平均),價格(眾數)..等兩項」、「廠商家數」，下圖並以 GIF 動畫展示三張圖表切換後各別的結果。



## § 主題產業數據儀表板

---

有別於產業概況與產業詳細數據，「[主題產業數據儀表板](#)」(後稱儀表板) 以不同網址進入，為一獨立頁面。

儀表板以主題方式呈現不同子產業之即時數據，例如電影產業之「每週電影票房資訊」；以下並以電影票房資訊為範本簡單介紹儀表板介面的使用方式。

### 數據

儀表板資料以即時資料為主，更新頻率則不一，以電影票房資訊來說，資料為逐週統計，因此**更新頻率亦為逐週更新**。資料主要來自[國影中心](#)，包含以下數據：

- 國別/地區
- 中文片名
- 上映日期
- 上映院數
- 累計銷售票數
- 累計銷售金額
- 申請人
- 出品

雖然資料為逐週更新數據，累積下來的數據亦為長期統計資料，因此我們不僅提供最新資訊，亦提供特定時段範圍的查詢功能，供您檢索過去的電影票房統計資料。

### 介面

儀表板介面切分成上下兩塊，包含上方的導覽列與下方的圖表面板。

- **導覽列**：提供表格檢視、時段選取、下載與整體排名等功能選項
- **圖表面板**：以區塊方式提供不同統計資訊與圖表。特定圖表提供**資料篩選功能**、並連動到其它圖表。可切換**國內**、**全球** 兩種數據檢視。

# 導覽列

主題產業數據儀表板  
文化內容策進院·策略研究處

每週電影票房資訊

出版產業

流行音樂產業

OTT

整體排名

表格檢視

時段：由 2021/02/21 至 2021/08/22 ▾

下載 ▾

導覽列如上圖所示，提供瀏覽指引以及與整體數據相關的功能，包括下列：

- 整體排名
- 表格檢視
- 時段選擇
- 下載

各別項目簡介如下。

## ※ 整體排名

整體排名提供選定時段範圍內，國內外所有電影的票房總排名，供您快速了解近期票房銷售狀況。國內電影以獨立顏色顯示，方便您比較國內外的票房成績。

右圖為整體排名資訊的部份截圖，可看到票房前五名的電影國別、票房與片名。其中國片「孤味」名列第四，總票房為 161756545 元。請注意總票房為「累計金額」，即從上映日起計之總票房。

排名	國別	總票房 (新臺幣)	片名
1	中華民國	405380495 元	當男人戀愛時
2	美國	357855616 元	哥吉拉大戰金剛
3	美國	184653736 元	靈魂急轉彎
4	中華民國	161756545 元	角頭－浪流連
5	美國	86851461 元	玩命關頭9
6	中華民國	82183108 元	聽見歌再唱
7	美國	81868136 元	玩命鈔劫
8	日本	65229041 元	名偵探柯南緋色的彈丸
9	美國	60504885 元	尋龍使者：拉雅

## ※ 表格檢視

您可於表格檢視中瀏覽完整細節資料。表格檢視提供選定時段中所有電影的詳細資訊 (即前述「國別/地區」、「中文片名」、「上映日期」、「上映院數」等資訊)，並提供全文搜尋功能以及欄位排序功能，您可使用搜尋功能過濾特定文字後，再使用排名功能比較各片表現。

例如下圖所示，我們搜尋「2021/」年列出所有於 2021 年上映之電影後，再讓各部電影依「上映院數」排序，得出 2021 年迄九月止上映院數最多的影片分別為「自殺突擊隊：集結」、「真人快打」、「那些要我死的人」等。請注意時段功能為影片銷售的日期，實際上映日期可能會早於指定範圍，此為正常現象。

2021/					
	A	B	C	D	
1	國別地區	中文片名	上映日期	上映院數 ^	累計
12	美國	自殺突擊隊：集結	2021/08/05	96	13
13	美國	真人快打	2021/04/08	96	10
14	美國	那些要我死的人	2021/05/06	95	32
15	美國	尋龍使者：拉雅	2021/03/05	95	13
16	美國	叢林奇航	2021/07/30	94	45
17	美國	湯姆貓與傑利鼠	2021/02/10	94	11
18	美國	厲陰宅3：是惡魔逼我	2021/07/22	93	24
19	南韓	永生戰	2021/04/12	92	10
20	美國	無名弒	2021/04/01	92	85
21	美國	花漾女子	2021/03/12	90	24
22	美國	詭老	2021/07/30	89	11
23	中華民國	緝魂	2021/01/29	89	20
24	美國	路卡的夏天	2021/08/06	87	22

## ※ 時段選擇

時段選擇供您指定想要呈現的數據時段範圍，區間則以週為單位。指定範圍後，統計資訊即以「該範圍內有銷售的影片資訊」為主，因此有可能會在設定指定範圍後，仍看到上映日期早於範圍內的影片。

例如，羅賓威廉斯主演的「春風化雨」上映日期為 1990 年，但在 2021 年重新上映了數位修復版，因此即便將時段設定在 2021 年間，仍能看到該片的統計數字。



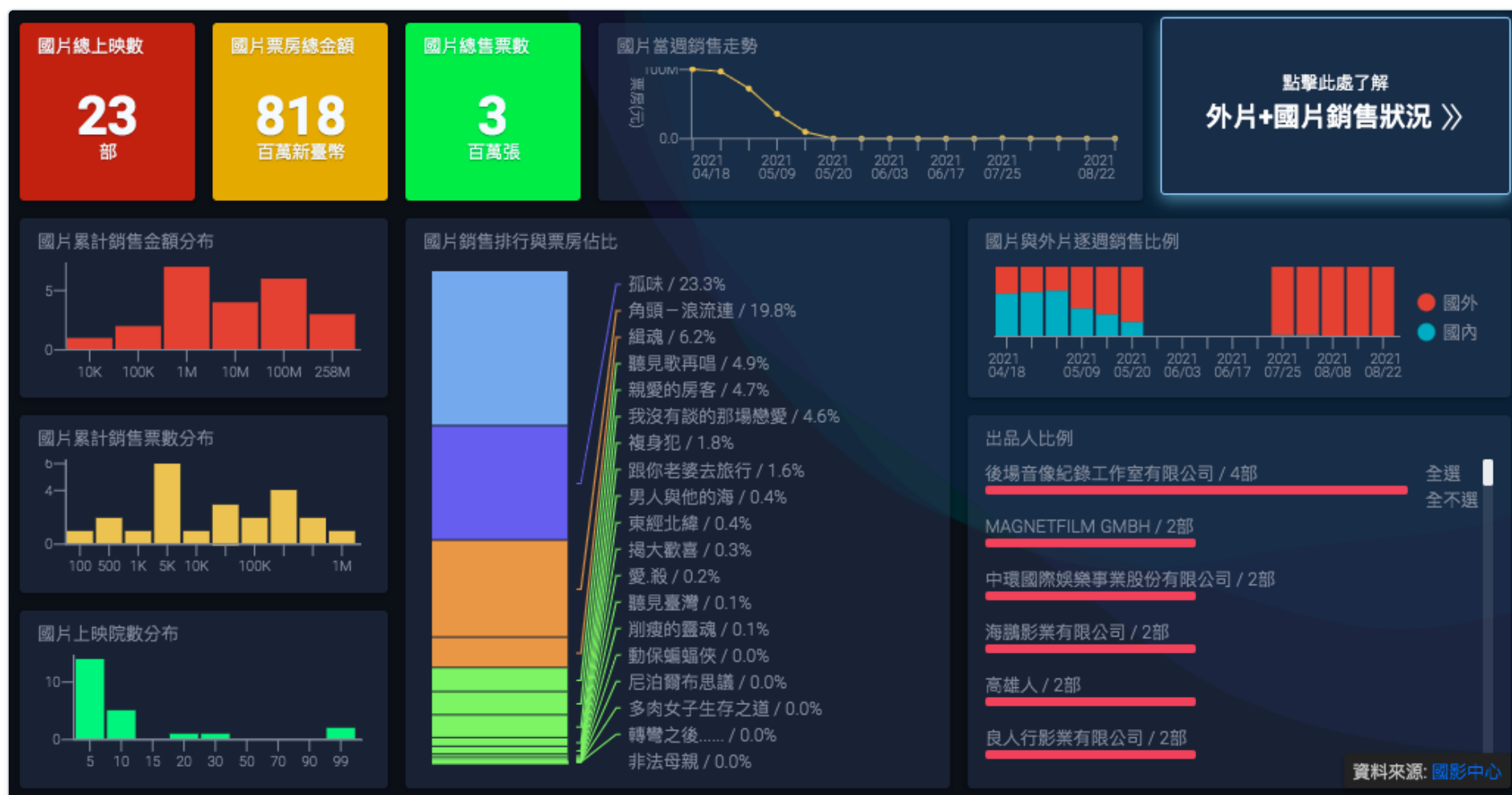
## ※ 下載

下載功能提供原始數據下載功能，其中包含「全時段完整數據」或「已選時段之數據」。此外，若您已透過儀表板圖表篩選過部份資料，亦可選取「篩選後之資料」中「僅國片」或「國片及外片」的數據。



## 圖表面板

圖表面板即為儀表板的主體資訊面板，分為「國片」與「全球」兩個面向，透過右方的「點此處了解」做兩者間的切換。下圖為圖表面板的樣貌：



面板各區塊呈現不同統計數據，包括「總上映院數」、「總上映影片數」、「總票房」等重點數據，以及銷售走勢、金額分布、各票房排名等。

部份區塊包含互動過濾功能，可透過滑鼠設定範圍限制呈現的數據量。其中包含：

### ※ 分布區間

在左方三項「**累計銷售金額**」、「**累計銷售票數**」、「**累計上映院數**」等分布圖，您可使用滑鼠拖選指定的分布範圍，儀表板將會即時依選取範圍重新計算各項統計數字。

例如，在右圖中我們同時針對「金額」與「院數」篩選，顯示其票房大於一千萬（10M ~ 258M）、上映院數少於 10 間的影片。

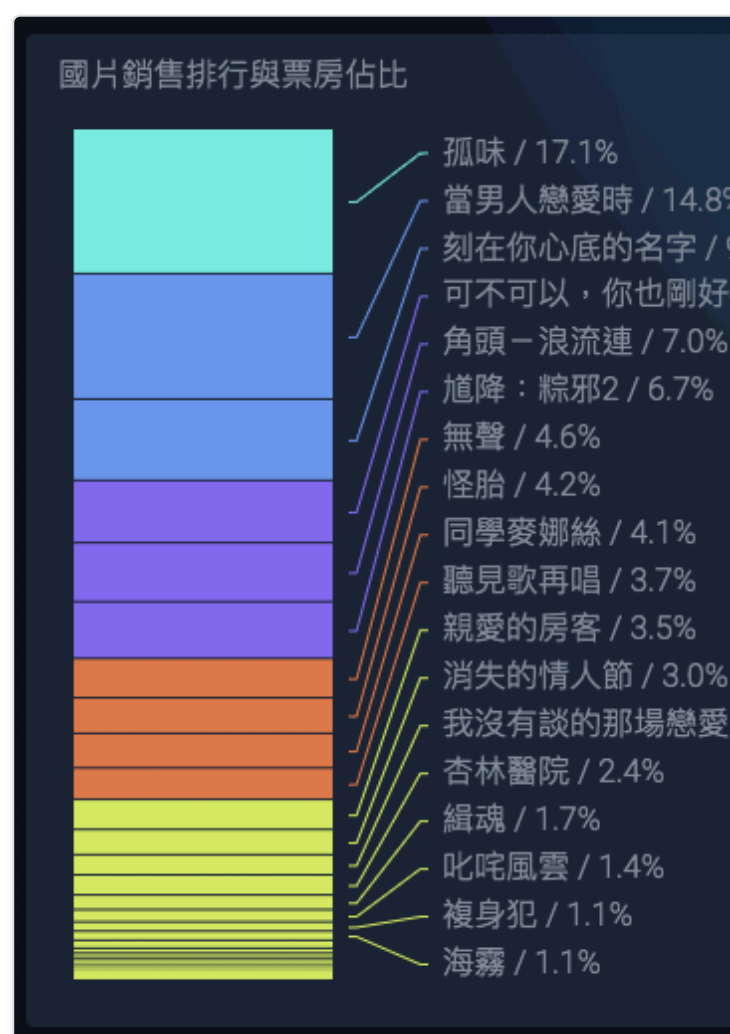




## ※ 排行與佔比

在「**國片銷售排行與票房佔比**」中，我們以單一比例長條圖呈現所有影片佔整體的票房比例、並以票房做排名；但是由於影片數眾多，無法同時顯示所有片名。

為此，我們實作了區段片名顯示功能，您只需要將滑鼠移至長條上，視覺化即會顯示對應區塊一定範圍內的影片名。右圖動畫即展示不同位置顯示不同片名的效果。



## ※ 出品人比例

在「**出品人比例**」圖表部份，我們想呈現選定範圍內出品人的比例，但由於家數眾多，我們使用排序的長條圖列表搭配捲軸的形式較完整的呈現資訊，同時您亦可點擊各個出品人來開關特定出品人的電影。

若欲關注特定出品人，您可點選「全不選」選項後，再選取特定出品人，即可讓儀表板僅呈現該出品人、時段範圍內的所有電影票房等統計資料。

由於此圖表亦同時顯示出品人家數，為方便選取，未被選取的出品人仍會顯示於圖表中，但所出品的電影部數將顯示為「0 部」。





# 資料檔格式簡介

用於文化內容策進院之「文創產調視覺化」之資料檔說明文件

---

此文件說明如何建立、編輯使用於文創產調視覺化之資料檔案。首先，您可能會最關心的應該就是「檔案格式」跟「編輯軟體」了：

## ※ 檔案格式



文化內容策進院之「文創產調視覺化」（後稱產調視覺化）為一線上網站，使用 JavaScript 即時載入並繪製資料檔成互動圖表。

為了方便直接透過線上程式讀取，資料檔格式以 **CSV (逗號分格數值)** 格式為主，可以透過試算表軟體（如 Google Spreadsheet 與 Microsoft Excel）中編輯後匯出。此份文件將**著重在試算表中資料的編排方式**，不會涉及到檔案格式等較低階的細節。

## ※ 編輯軟體

要編輯此類檔案，您僅需要準備 Microsoft Excel 或 Google Spreadsheet 等軟體；若牽涉到軟體使用的部份，此文件將以 **Google Spreadsheet 做為示範工具**，但兩者的概念大多通用，若您希望使用 Microsoft Excel 或其它相似的軟體，您可將文件說明的部份類推至您的軟體上使用。



此文件分為以下章節：

- 簡介
- 試算表結構規範
- 多重資料表
- 屬性規劃方針
- NOIR 資料分類法

# 簡介

為了讓程式能夠順利讀取您的資料，您的資料必須要「結構化」。結構化不容易簡單說明，但大致上，資料必須要有一致性，避免混雜排版、自然語言等資訊，並且盡量將複合性的資訊分開記載。

## ※ 基本架構

為方便接下來的說明，我們簡單描繪一下表格的長相。如下圖：

- 資料以表格呈現
- 由上而下為逐筆資料 ( data row )
- 由左至右則為各筆資料的不同欄位 ( data column / field )

各筆資料			各個欄位		
第一筆資料			第一欄	第二欄	第三欄
第二筆資料					
第三筆資料					

## ※ 資料一致性

以下例來說，左表在數字中混雜了文字、縮寫跟單位，缺乏一致性；中表在數字中使用文字加註說明；右表在同一個欄位中記載兩組資料：

缺乏一致性	缺乏一致性	混合資料程式不能用
5k	102年	203 / 五折
256	103年	198 / 七折
五千元	104年 (僅 A 類)	200 / 六折

改進的方式並不難，首先統一格式，接著抽離單位，最後將混合欄位分開：

元	年	分類	元	折數
5000	102	全部	203	5
256	103	全部	198	7
5000	104	A	200	6

由於大多數的資料都有單位的概念，因此上表此類「標頭 ( header ) + 資料 ( body )」的基本格式可說是我們表格的最基本格式。

又或者，下表某真實案例中，記載各類型廣播電臺的涵蓋電臺，等於將多組資料寫在單一儲存格中，並用頓號排板，前後文結構又不一致，程式無法正確分析：

類型	涵蓋電臺
公營電臺 (10)	內政部警政署警察廣播電臺（大功率）、國立教育廣播電臺（大功率）、漢聲廣播電臺（國防部心理作戰大隊）（大功率）、復興廣播電臺（大功率）、行政院農業委員會漁業署漁業廣播電臺（大功率）、臺北廣播電臺（大功率）、高雄廣播電臺（大功率）、財團法人中央廣播電臺（海外短波）、講客廣播電臺（客家委員會）（大功率）、財團法人原住民族文化事業基金會（大功率）
調幅電臺 (16)	中華、天南、天聲、民本、民立、先聲、成功、建國、電聲、益世、國聲、基隆、勝利之聲、華聲、鳳鳴、燕聲

以上例來說，我們可以做以下改進：

- 數字並不需要，程式即可做統計
- 各電臺分別陳列在不同筆資料之中

調整後的資料表如下：

電臺	類型
內政部警政署警察廣播電臺（大功率）	公營電臺
國立教育廣播電臺（大功率）	公營電臺
漢聲廣播電臺（國防部心理作戰大隊）（大功率）	公營電臺
中華	調幅電臺
天南	調幅電臺
天聲	調幅電臺
民本	調幅電臺
... 下略 ...	

※ 嚴禁合併儲存格

為了將各筆資料明確分割，嚴禁使用合併儲存格。的確，由於統計上的困難，有時無法準確的將資料切分開來，然而這樣的資料就算您做成了資料檔，程式也是無法呈現。

為了盡量呈現您手中的資料，您可以選擇：

- 標註為「無資料」（留空白, 或使用橫線「-」）
- 將表格依欄位分割後，再合併的資料組重新編排

請不要這樣做

年度	金額	數量
2018	500	25
2019		16
2020	250	

寧可省略

年度	金額	數量
2018	-	25
2019	-	-
2020	250	-

切割表格後留下金額

年度	金額
2018 ~ 2019	500
2000	250

切割表格後留下數量

年度	數量
2018	25
2019 ~ 2000	16

進一步結構化

起始年度	結束年度	金額
2018	2019	500
2000	2000	250

起始年度	結束年度	數量
2018	2018	25
2019	2000	16

※ 摘要

依上述的結構化要求，一個最基本的資料表格如下：

欄位1	欄位2	欄位3	欄位4
第一筆資料 欄位1的值	第一筆資料 欄位2的值	第一筆資料 欄位3的值	第一筆資料 欄位4的值
第二筆資料 欄位1的值	第二筆資料 欄位2的值	第二筆資料 欄位3的值	第二筆資料 欄位4的值
... ( 後續其它筆資料 )			
... ( 後續其它筆資料 )			

最後，請確保單一檔案僅包含一個表格。如果您有多組表格，只要無法合理併成一個表格，就應該另外存成獨立的檔案，避免程式無法正確解析。

# 試算表結構規範

除了一般結構化的概念，我們還必須要考慮到程式的特定用途，在此處即為「產調視覺化」線上圖表繪製的用途。為了生成圖表，我們需要以下資訊：

- 說明性資料－資料集標題、註解等
- 欄位資訊－各欄位名稱、型態、單位等
- 資料內容－實際數字、文字等本身
- 保留欄位－格式版本等程式未來可能會需要的擴充機制

一般來說，這些資訊通常會使用程式常見的專用格式來儲存 ( 如 JSON 或 XML )，但由於此類文件通常較不容易讓一般人編輯，但一般較容易編輯的 Microsoft Excel 檔案卻又過於一般性，沒有這些資訊的規範，因此我們必須明確規定如何在試算表檔案中加入這些資訊，並請編輯人員遵照規定編輯試算表檔案。

首先，在一個普通的試算表中，我們在上面保留 7 行用以記載額外的資訊；第 8 行以後則為單純的資料區域。前 7 行資料又分為「表格」與「欄位」的中介資訊：

- 表格中介資訊：說明這個表格，包括標題、註解、程式用保留資料等，共 3 行。
- 欄位中介資料：說明表格中各個欄位的單位，意含跟資料類型等，共 4 行。

以圖像化的方式來呈現：

1	標題
2	保留列 ( 請留白 )
3	註解
4	資料類型 ( 文字、數值等，以 NOIR 分類法表示，於後續章節說明 )
5	單位 ( 例如：元、人、頁 )
6	保留列 ( 請留白 )
7	欄位名 ( 如年份、出口值、人次、頁數等 )
8	資料1
9	資料2
10	...

以下以一個假想案例來做範例：（注意前兩行我們僅用到第一欄的儲存格。第二欄以後的儲存格因為未定義用途，應保持空白）

1	文創產業樣貌		
2			
3	自 104 年至 108 年，樣本 320 間公司，其中 105 年度樣本數較少，採正規化後推估。		
4	T	C	R
5	年		元新臺幣
6			
7	年度	產業別	產值
8	104	出版	2881
9	104	影視	3078
10	105	出版	2763
10	105	影視	3159



# 多重資料表

如同在前文中所提到的，我們禁止在單一檔案放多個表格。

每個資料檔僅能含有一個表格，再強調一次，不得將多組資料記載在同一份資料表中。

下圖為一錯誤的示範，將多個表格拼貼在一個試算表中，其中包括「電影放映業」、「電影後製特效業」等兩產業，又分別將兩產業各年度之資料分不同表格記載：

	A	B	C	D	E	F	G	H
1	電影映演業	2019年	電影後製特效業	2018年	電影後製特效業	2017年	電影後製特效業	2016年
2	電影放映收入	90.02%	電影後製收入	23.32%	電影後製收入	20.90%	剪輯	22.78%
3	廣告收入	0.64%	電影特效製作收入	25.51%	電影特效製作收入	20.88%	音樂製作	3.45%
4	食物飲料販售	8.41%	其他內容製作收入	2.67%	其他內容製作收入	7.81%	後製錄音	9.43%
5	衍生商品販售	0.34%	其他內容後製收入	13.03%	其他內容後製收入	11.87%	視覺特效	35.07%
6	其他電影商務收入	0.09%	其他內容特效收入	6.48%	其他內容特效收入	10.75%	字幕、拷貝	2.99%
7	其他收入	0.51%	影視相關設備器材或場地租賃收入	15.09%	影視相關設備器材或場地租賃收入	16.10%	SR Dolby	2.44%
8			藝人經紀	0.39%	置入性行銷、贊助收入	1.47%	器材及設備提供	19.93%
9			其他	4.58%	其他	5.35%	其他	3.91%
10			電影版權銷售收入	4.31%	電影版權銷售收入	4.66%		
11			電影於網路數位發行之相關收入	0.22%	電影於網路數位發行之相關收入	0.15%		
12			電影衍生商品授權收入	3.36%	電影票房分成	0.05%		
13			電影代理發行收入	1.03%				
14								
15			電影映演業	2018年	2017年	2016年		
16			電影放映收入	94.66%	93.47%	89.47%		

當多重表格混合在一起時，程式完全無法知道該如何判讀資料，也因此這樣的資料只能透過人工使用，無法輸入程式使用，這也是為什麼我們不應該這樣整理資料。

然而，分成多檔案有時會讓整理資料變得相當不容易。為方便管理，我們可以將多個表格以 **不同工作簿** 的形式儲存在同一個試算表檔案中；只要將相關的資料放在同一個試算表中，資料整理或檔案管理都會變得相對容易。下圖展示了使用 Google Spreadsheet 做多重表格管理的範例，各個工作簿以表格標題命名，內容則依上述資料表結構編排：

年度	參與國際市場展	輔導參與國際市	入圍國際影展國	榮獲國際影展獎	輔導參與國際影展國片人次		
97	208	-	91	16	69		
98	211	-	125	12	110		
99	239	82	124	32	77		
100	206	80	122	23	51		
101	244	122	96	35	60		
102	294	138	86	31	53		
103	289	122	86	16	56		
104	222	153	88	17	49		
105	214	14	60	12	44		
106	167	66	10	78			
107	155	70	20	83			
108	133	76	8	113			
109	41	12	15				

- 國片製作兼發行情形
- 跨國合作影片數量
- 國片導演兼任編劇情形
- 國際影展活動參與度
- 收益結構
- 全台電影票房
- 台北市電影票房
- 人均觀影次數

+

三

情形

跨國合作影片數量

國片導演兼任編劇情形

國際影展活動參與度

收益結構

全台電影

由於程式使用資料時是以單一 CSV 檔為單位，而 CSV 檔並不支援工作簿的概念，我們仍需要將工作簿輸出成多個 CSV 檔，才能供給程式使用。資料表數量眾多，手工逐一輸出資料較費時，因此這時我們可以使用輔助程式如 VBA 或 Script Tool 來協助我們自動生成 CSV 檔。

輔助程式屬編輯軟體操作的範疇，並非此文件所涵蓋的範圍，但相關教學可以在網路上找到，這裡附上一個用於 Google Spreadsheet 的工具程式，可協助我們生成所有工作簿的 CSV 壓縮檔：

程式碼出處：[Michael DeRazon](#)

此程式需透過 Google Spreadsheet 的 Script Editor ( 選單 → 工具/Tool → 指令碼編輯器/Script Editor )，關於使用 Script Editor 的教學，請另參考 Google Spreadsheet 的使用文件。

# 欄位規劃方針

有時候，即便我們已經依照上述方式來整理資料，仍不免將資訊藏進欄位名稱中，減損了資料可再利用性。

比方說，下表利用多個欄位分「年度」來呈現「產值」這個數據，由於「年度」之值儲存在欄位中，並沒有對應的描述性資料 ( 年度的資料分類、單位等資訊都缺失了 )，我們在利用資訊時就變得更為侷限；此外，「產值」的概念也被混進年度之中，更難正確的解析：

產業別	105年產值	106年產值	107年產值
出版	2517	1923	1498
影視	3011	3262	3119

這裡我們應考慮將欄位重新匯整成單一欄位，並將資料做對應的轉換。此動作一般稱之為「反向樞紐 ( Unpivot )」。

以上表為例，我們先將「年度」獨立成一欄，產業別則會因年份而需要重覆輸入，如下表左。接著，再增加一欄「產值」，將「產業別 x 年份」原本對應到的數值填回此欄，如下表右。

先將年份獨立成一欄

產業別	年份
出版	105
影視	105
出版	106
影視	106
出版	107
影視	107

再依對應位置合併原本多欄的產值

產業別	年份	產值
出版	105	2517
影視	105	3011
出版	106	1923
影視	106	3262
出版	107	1498
影視	107	3119

反向樞紐後的資料表仍能透過正向樞紐轉換變成原本的形式，卻多了額外欄位的資訊，同時程式也更容易依範圍過濾、使用額外的欄位做更多樣化的資料呈現，因此較建議在編撰資料時使用此種形式的欄位規劃。

# NOIR 資料分類法

掌握正確良好的資料結構後，讓我們來看看資料數據本身。

數據有不同的型態，最常見的就是「文字」與「數字」的分野：

我是文字	5447
------	------

數字跟文字可以做的事情不太一樣。數字可以加總、平均，文字可以排序、做分類，因此若能妥善的判斷數據的類型，我們才能正確的實現資料的視覺化。

程式基本上可以自動判斷數據的類型；然而問題是當我們有多組數據時，往往會因為一些例外狀況而使得判斷數據類型變得困難。比方說，以下兩個表格中的數字完全不應該做為數字使用，但程式很難自行判讀：

公司名稱	公車路線
1111	508
104	260
591	紅15

也因此，我們在整理資料時，最好同時將各欄資料的類型標註在表格中，這也就是我們在「試算表結構規範」一節中，表格第四行「資料類型」的用途。然而除了「文字」、「數字」外，資料還可細分成相當多不同的種類；為了避免讓資料梳理的過程變得過度複雜，我們僅依照資料視覺化的需求，使用「NOIR 資料分類法」將資料大致上分成以下幾種：

## N Nominal / 記名式

代表標示、名稱，如代號型資料，如產品編號，國家、產業別、姓名等。

## O Ordinal / 有序式

可以排序的資料都可視為有序式，但一般以直觀認定其是否有序來標註即可。如月份、日期、流水號。

## I Interval / 區間式

以數字為主，若加總無意義（如年齡相加），或數值實為相對值（如攝式溫度），則視為區間式資料。

## R Ratio / 比例式

非區間式的數字即為比例式。如金額、人數、面積、降雨量等等。

此外, 為更精準進行視覺化, 我們進一步將 N 與 O 類資料分為下列兩種子類：

C

Categorical / 分類式

當一組資料中此欄位中的數值僅有少數幾種數值, 可視為分類用途, 即可進一步標註為 C.

T

Temporal / 時序式

時序式資料. 當有序式資料為時間的概念, 可另標註為 T

了解資料分類方式後，我們在編撰資料表時，僅需將各種資料的類型以其對應代號 ( **N O I R C T** ) 填入第四行之「資料類型」欄位即可。下表即為一範例，依各欄位類型填入了對應的各代碼：

文創產業樣貌					
無註解					
T	C	N	O	I	R
年				%	元新臺幣
年度	產業別	公司名	資本額排名	員工男女比	產值
104	出版	雜誌A	23	25	1237
104	出版	雜誌B	15	28	672
104	影視	電視臺A	2	48	7826
104	影視	電視臺B	9	46	6271