

# 中國文化大學教師教學創新暨教材研發獎勵成果報告書

壹、計畫名稱：資料視覺化教材之研發

貳、實施課程、授課教師姓名

課程名稱：I276 巨量資料分析

開課系級：MBMIM 資管碩 1

教師姓名：資管系 游源治

參、前言

設計合宜的教材可以協助教師對於學生學習成效分析，並藉此了解學生學習路徑及學習狀態、進而得到教學回饋以調整教學實務，精進日後教學活動。

然而對資料科學相關課程而言，教材不只是科學理論及實務應用的講解而已，還包括資料本身的配合呈現，而當今網路世代的資料是巨量的，來源多樣性的、資料變異多的、需要更快的處理速度；特別是這些資料量往往龐大到資料庫系統無法在合理時間內進行儲存、運算、處理，分析；因此若在短短的課程時間中，要去導入這些巨大資料集來進行資料處理案例教學，勢必是一項艱難的工作，因為學生不清楚這些資料集外在及內在本質是什麼樣子，若以傳統方式來解釋這些資料集的輪廓(profile)，則必需要先介紹資料的組織綱目，並取樣數筆資料搭配不同的分佈尺度來解釋，儘量想辦法讓學生了解資料集的內外輪廓，以教學實務來看，這會造成寶貴教學時間的浪費，教學節奏受到打擾，內容也變的複雜化，因此我們需要一種教學工具或教材，讓這些巨大複雜的資料集可以被快速而簡潔的呈現及被解釋。

這些資料集必須透過某些層次的抽象化或實體化做法，學生才有辦法去理解那些數據。而資料視覺化技術則提供了一個很好的抽象化或實體化解決方式，也就是我們在準備教材中的資料集時，事先用視覺化工具把它們形成可操作的視覺化介面，然後在案例教學中，以檢視操作方式，快速視覺化的導入資料輪廓，讓學生可以清楚而直覺的切入資料內涵，並專心在核心理論的驗證學習，也就是把教材中的資料集圖式化及可操作化，讓學生直接看的到，也看得懂。

以資料科學領域而言，教學常用的經典資料集頗多，然而在此計劃案中，我們將配合資料科學教學進度規劃，選擇適當的資料集，製作出這些資料集的視覺化操作介面，並把它包裹封裝成為教材的一部份，以協助課程教學實務之進行。

## 肆、計畫特色及具體內容

### 〔特色〕

資料視覺化基本概念就是一當資料很多，內容很複雜時，透過各種圖像或是互動的方式，幫助人們理解資料。將一股腦的複雜數據轉化成為圖像，不但能夠看得生動，而且也能讓人有視解式理解資料的機會。

然而本計劃所產出的視覺化教材與一般的資料視覺化有何主要不同呢？其主要特色在於可操作性，也就是資料集可以被視覺化操作，呈現它的不同面向，而這種操作所呈現的互動是動態的，可調適的，讓教學者每次可以用不同的故事說同樣的資料集，這是傳統預設腳本的互動式教材所無法呈現的，其被後原理在於操作介面是即時串接到資料集本身，當操作指令變化時，可以取得資料的不同維度，並反應在視覺化介面之中。

### 〔具體內容〕

資料視覺化就是利用人類的各種天生的特性，將資料轉化為更方便我們理解的形式。人類天生擁有許多天然的特性，有擅長與不擅長的事物。看到百元紙鈔沒什麼感覺，但若你神經地把它全換成一元硬幣，就能真切地感受到它的份量。雖然百元紙鈔和一百個一元硬幣有一樣的意義，但換成一百個一元硬幣後，我們就能更能直覺的理解，原來那張紙真的代表了「一百個」一元

資料視覺化對資料科學家而言，首先可以節省他們判斷資料的時間，再者，因為透過不同面向的閱讀，也可能藉此找出自己的思考盲點，發現問題，找到隱藏的關聯性等；而對資料科學課程的學生而言，圖像化可以增加閱讀的意願，有利於資訊的傳播，而且也能減少閱讀的障礙，因為那些教材已經將資料轉化為圖像，便已融入了豐富的資訊，可以讓人用同樣的視覺角度看問題。

然而資料視覺化不會只是美工裝飾，資料視覺化的目的並不僅僅是為了好看，更關鍵是為了能讓人看懂；前者相對容易，但後者則必須花上不少心思。要使別人易懂，自己也要先了解才行，如果連自己都弄不明白，那你做的圖就稱不上是「視覺化」了。好的設計可以幫助人理解資料、增加閱讀意願、方便傳播等。這有一個共通的特性，即是通常如果做得很好，讀者其實是沒感覺的。好比說交通狀態良好時，你不會記得什麼，但是一旦常發生塞車，你便會特別不耐。所以我們在做資料視覺化時，是盡可能的在不簡化資料的情況下，利用各種手段把閱讀門檻降低到目標讀者的程度。

#### ● 資料視覺化的設計要點

在設計之初，我們要盡可能的理解資料，進行資料探索（data explore），然後選用適合的圖表表達資料的關聯性。每一種圖表都有其適合與不適合，無論是長條圖、圓餅圖、折線圖還是直方圖，這需要經驗和對各種圖表的熟悉掌握程度。這比較像是基本功，但用得好可以讓使用者更能輕易地理解資料。

1.顏色：顏色最能抓住讀者的注意，速度甚至快到「讀者都沒開始理解其中的意義時，就已經進腦袋裡了」。也因為顏色非常「強烈」，所以常用來表示資料最重要的維度。

好的設計不隨便使用顏色，因為顏色太過醒目，讀者很容易分心。圖表上如果有太多顏色，讀者會無法分辨何者為重要，何者為不重要。而且我們的腦袋時常會偷偷越過我們的理性，自行判讀顏色的意義，假若其中的意義和資料本身不相符合，就會使人困惑。例如彩虹色便是錯用顏色的經典例子，許多人會用彩虹色來製圖表，認為這是物理上波長連續的一系列顏色，所以代表一系列的數值非常符合自然。聽起來好像有道理，但人類的視覺系統並不認為彩虹是連續的一整體，而會自動為顏色分群，我們會潛意識為這些顏色賦予意義，而這個意義不見得與資料相符合。

2.形狀：形狀非常籠統，形狀、角度、距離、方向等都包含在內，這些要素都能抓住讀者的注意，不過人類對不同的要素敏感度不一。比如說人類對「長度」的判斷敏感，因此當有資料要比較時，長度是不錯的選擇。相對而言，人對角度雖然也算敏感，不過沒有像對「長度」的感受那麼細緻。所以要比較差異時，長條圖會比圓餅圖還要好，但圓餅圖還是有自己的優勢，在看整體比例時，圓餅圖效果還是很不錯的。相比起來，人對面積就不是這麼敏感，不容易判斷資料差異的倍數。如果有一大一小的圖形，我們會傾向低估大圖形的面積，以為大的沒這麼大。

3.動作：人對會動的事物也會有反應，這也是初學者特別喜歡使用的方式，希望藉此讓圖表看起來豐富多采以用來掩飾貧乏的內容。但事實上，這種方式可能不但沒有加分，還會反過來扣分。因為我們的腦已經演化出一種精密的方式可以自動地篩選動態的資訊，以幫助我們在很短的時間內做出正確的判斷，所以如果錯誤的使用動畫，反而可能會干擾資訊解讀。

4.空間的運用：所有的圖表都是在一個空間表達資料，在一個大的空間呈現和在一個小的空間呈現是不同的概念。在電腦呈現的方式不應該與手機相同，因為我們閱讀的方式並不一樣。所以在設計之時，也要考慮在資料顯示的情境為何。

適當的扭曲資料裡的空間有時也是一種表達資料的手段，例如「捷運地圖」，台北捷運的路線顯然不都是直線，轉彎也不都是直角和四十五度角，但透過這種扭曲空間的方式，反而能方便我們理解。當然你也可以解釋說，因為捷運的實際的路線不是我們在乎的，所以被簡化了。

5.互動的運用：互動的運用，設計師可以提供一些工具來讓讀者探索資料，讓使用者更能了解資料的本質。互動也可以提升趣味性，增加使用者的閱讀意願。

6.可操作化的運用：可操作化為本次教材研發的重點，藉由可操作化介面，視覺化教材變成活的，這好比一個線上資料分析系統一樣，決策者可以看到總成的管理數據，然後藉由操作去對多維形式的資料採取切片（Slice）、切塊（Dice）、向上擷取（Roll-up）、向下擷取（Drill-down）、旋轉（Pivot）等各種分析動作，以求剖析資料，使學生能從多個角度觀察資料庫中的數據，從而深入理解包含在數據中的內在

(insight)。

●資料視覺化介面處理的範例資料集

而本計劃所研發的教材中，將處理以下資料集，把它們包裹成可操作化的視覺介面，

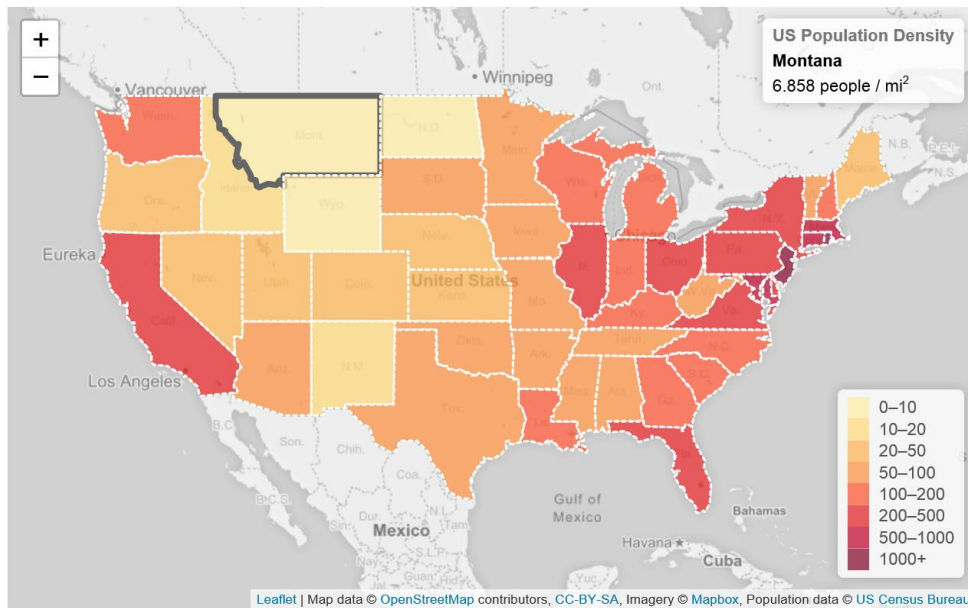
Name	Description
Auto	Gas mileage, horsepower, and other information for cars.
Boston	Housing values and other information about Boston suburbs.
Caravan	Information about individuals offered caravan insurance.
Carseats	Information about car seat sales in 400 stores.
College	Demographic characteristics, tuition, and more for USA colleges.
Default	Customer default records for a credit card company.
Hitters	Records and salaries for baseball players.
Khan	Gene expression measurements for four cancer types.
NCI60	Gene expression measurements for 64 cancer cell lines.
OJ	Sales information for Citrus Hill and Minute Maid orange juice.
Portfolio	Past values of financial assets, for use in portfolio allocation.
Smarket	Daily percentage returns for S&P 500 over a 5-year period.
USArrests	Crime statistics per 100,000 residents in 50 states of USA.
Wage	Income survey data for males in central Atlantic region of USA.
Weekly	1,089 weekly stock market returns for 21 years.

TABLE 1. A list of data sets needed to perform the labs and exercises

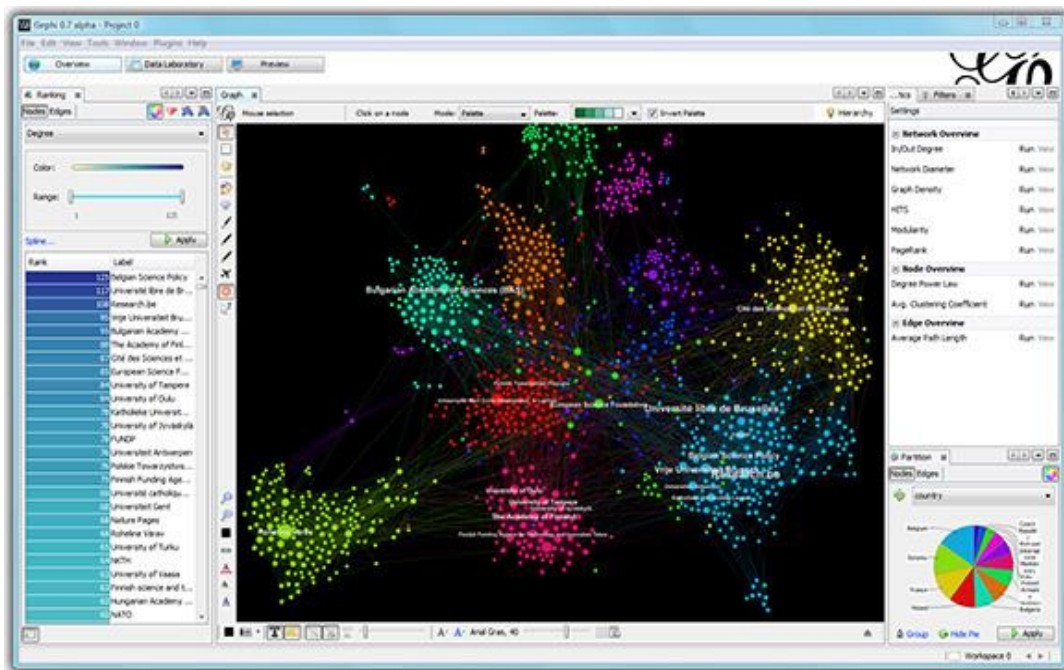
●資料視覺化介面處理的工具

而本計劃所研發的教材中，將利用以下工具處理資料集，把它們包裹成可操作化的視覺介面，成為教材的一部份

1. R 語言:作為一種開源電腦語言和工具，R 日漸成為資料專業人的心頭好。如：資料科學網站 EnhanceDataScience.com 就分兩期介紹了 R 語言的一些常用互動式資訊圖工具包。
2. Leaflet: Leaflet 是現代化開源的 JavaScript library，是為了手機優化的良好互動地圖。容量只占 JavaScript 的 31KB，擁有所有大部分開發商所需要線上地圖的特色，Leaflet 融合了簡單、效能及實用性三者的設計，有效率的跨桌上型及行動手機平台高效能地運作，雖以 HTML5 及 CSS3 為優勢，但仍能夠讓版本較早的瀏覽器進行瀏覽動作。



3. Google Charts: Google Charts 提供了非常棒的方式能在網站上視覺化資料，也就是教材可以成為網頁放在雲端；從一條簡單的線圖到很複雜的階層型樹狀圖，圖庫藝廊提供了非常多可使用的圖表類型，對於專業視覺畫的圖表如地理圖、動態壓力圖等，是十分好用的工具；另外包內建含動畫和使用者互動控制的功能。
4. Gephi： Gephi 對於所有類型的網絡及複雜的系統、動態性和階層圖等，都兼具互動性視覺化以及搜尋平台。以圖表為基礎的視覺化者以及資料搜尋者，不只能夠應付大量的資料集，還能創出美麗的圖表，同樣也能讓你篩選、清除不必要的資料



視覺化學習也有些挑戰要面對克服、當教材中的某些部份被資料視圖化後，這些資料有了不同的展現形態，至於能不能從這些資料中看出端倪，使得這些資料變成資訊，進而轉化為知識，就要看教學者與學習者的實務表現了。

## 伍、實施成效及影響（量化及質化）

### 〔實施成就〕

(1). 完成之視覺化教材：分別以 R 語言，python，excel 為視覺化分析呈現語言，完成了以自我學習為主的五個實驗單元式資料視覺化教材，如光碟檔案的教材目錄所示：

Lab00	以 R 直接匯入 Csv 資料後，做視覺化資料互動探索
Lab01	線性迴歸視覺化分析
Lab02	最小平方的多元線性回歸視覺化分析
Lab03	推薦系統及回歸視覺化分析
Lab04	Power View 的 OLAP 視覺化分析

(2). 教材之驗證實驗：

- 以單元報告來評估學生利用視覺化教材課後學習狀態及成效
- 分析學習歷程紀錄，分析教學活動的反應和學習途徑之變化
- 執行課程中評量活動，包括：滿意度調查。
- 依主要的學習指標，針對不同表現程度做量化及質化的描述，用於執行結果的評估。
- 記錄並規劃後續研究。

(3). 學習成效的期待：

- i、 互動性及多媒體化，可協助老師導入視覺化多媒體教學

ii、 減少教學複雜度，讓學員更專心在核心理論的學習。

iii、 重覆性深化學習：由於教材是動態的，學生可以於課後，操作性複習，深化學習成果。

#### (4). 課程品質的可能改善

i、 強化研究所課程之教學深度及創新。

ii、 促進 e 化學習：透過互動式視覺化教材，不凝聚學員對課程的關注力，特別是整合 e-Learning 教材及系統，擴展 e 化學習的範疇

iii、 課程的智慧化應用。利用視覺化教材，可以搭配教學回饋分析，去分析學習者狀態，以協助教學者在教學實務的設計與決策。

#### [ 影響評估 ]

(1). 視覺化學習也有些挑戰要面對克服、當教材中的某些部份被資料視圖化後，這些資料有了不同的展現形態，至於能不能從這些資料中看出端倪，使得這些資料變成資訊，進而轉化為知識，就要看教學者與學習者的實務表現了。

(2). 此外學員的參與及注意力很重要，資料視覺化後，還是有賴學員去用心探索。

(3). 主要技術瓶頸在於視覺化介面的使用技巧，除了要符合數位學習的教材標準之後，如何整合數位學習系統，並讓學員熟習數位學習系統的操作模式，將會是關鍵技能。

(4). 本次計劃目前以單一課程為雛形，待可行性被驗證後，結合教學進度驗證機制是未來發展方向。



## 伍、 結論

資料視覺化 (Data Visualization)，指的是利用圖形化工具 (如：各式統計圖表、立體模型等) 從龐大繁雜的數據庫中萃取有用的資料，使其成為易於閱讀、理解的資訊。日常生活中，我們時常運用到資料視覺化工具，而在教學的運用上，其實可以有很多發想及實務應用，如：將全班各科成績以柱狀圖顯示、捷運路線圖的規劃等。

資料視覺化有三個要點必須考量：

- 資料的正確性：在簡化資料時保有數據的正確性
- 讀者的閱讀動機：讀者能否突破心理障礙，去理解不熟悉的領域資訊
- 傳遞資訊的效率：讀者理解圖表所耗費的時間、能否留下印象等

因此，在處理視覺化資料教材時，必須斟酌這三個要點的比重。思考如何運用在教學實務之上，而本專案教材主要在於：引導學生正確定義問題，從龐大繁雜的數據庫中萃取有用的資料，建立運算規則整批處理數據資料集，進而運用資料之視覺表現形式解說結果與發現，藉助圖形化方式清晰有效地傳達與溝通訊息。特別是學位學時代到來，教學實務的數位化，行動化，經由環境感測，不斷增加的資料量，包括課程進度因持續進行所積累的活動數據等。龐大的資料帶來了更多值得研究的隱藏資訊，有助於教學實務進行。我們認為資料視覺化可以快速為使用者提供更好的方式去理解資料，而剛好這門課是大數據分析，於是我們進行了這個專案。

## 陸、 執行計畫活動照片

由於本學期課程，安排在一般教室，學生無法在課堂時操作視覺化教材，故以課後學習方式來散佈相關教材，以下是借用電腦教室，做教材相關的驗證實驗。

借用電腦教室，做資料視覺化教材驗證實驗



借用電腦教室，做資料視覺化教材驗證實驗



捌、附件

備註：

1. 本報告書大綱得視需要自行增列項目。
2. 成果報告書須另以光碟儲存，並附加執行計畫活動照片電子檔。

柒、

捌、