

以視覺化呈現體育運動資料分析之可行性

陳秋丰 / 國立臺灣師範大學

葉劭緯 / 湖南商學院

摘要

在資通訊快速轉換的社會趨勢下，資料量不斷加倍成長，龐大的資料帶來了更多值得研究的隱藏資訊。次級資料分析便是從與主題相關的文獻或過往的資料中整理蒐集完成後進行的分析。次級資料分析是透過既存的大型研究資料庫（例如：中央研究院社會變遷調查）或政府統計資料（例如：工商普查）來對該研究進行深化實證探究的方法。而資料視覺化是透過圖像化工具（例如：各種統計圖表、立體模型等）從複雜浩大的資料庫中篩選出合適且可用的數據資料，經由轉化或介接分析後，進而成為簡易閱讀、容易理解的可靠訊息，資料視覺化可以快速提供即時的方式理解資料。此外，政府開放資料為近年研究的熱門議題，我國政府也將其資料列為政策一大重點，並要求其所屬機關開放資料給予社會公眾使用。以教育部體育署而言，目前有將近 20 個資料庫，因為委辦或承辦單位不同，目前較缺少整合或介接分析。因此，本研究目的是從非反應式研究角度，探究不同資料庫整合後，視覺化分析呈現的樣態。本研究以體育運動資訊相關之運動現況調查、運動消費支出、體適能等資料庫為例，探究不同資料庫進行視覺化分析的樣態。本研究發現任何的資料分析及視覺化之前，首先應將資料再整理與計算，有乾淨、完整和有意義的資料，才能呈現真實的現況。其次，根據研究目的與設計，將資料庫進行介接或合併，例如：歷年規律運動人口與縣市對應之趨勢；各縣市運動現況與消費支出之關聯。最後，從資料可呈現性的角度，以多元圖像方式呈現其資料視覺化分析，亦可展現出更有效果的資料訊息傳遞之能力，但是，從學術研究角度仍須強化資料的可檢驗性與價值性，並進一步進行量化的技術分析。

關鍵詞：非反應式研究、次級資料、資料探勘

通訊作者：葉劭緯 單位：湖南商學院體育教研部 Email：shaowei_yeh@163.com

壹、前言

隨著數據科學逐漸成為學術研究的新方法，近年來有越來越多關於資料庫整合的研究（倪瑛蓮、陳龍弘、張仁和，2017；許懷中、黃致豪，2017；O' Halloran, Tan, Pham, Bateman, & Moere, 2016）。一般而言，資料庫的數據即為大量、快速、複雜且多變的資料集合而成，大量的數據能夠更清楚的呈現一個現象的現況，進而透過分析來解讀這些資料背後的意涵。因此，目前之前瞻性議題和做法則是數據之分析、資料庫之整合以及趨勢之決策，若是研究的數據量大，資料的品質高，就能增加了資料庫研究的應用價值，瞭解社會現象變項的關聯性，或是能夠驗證一些研究假設或理論，尤其是體育運動資料庫的應用近年來更是廣泛地被討論及研究（謝邦昌，2017）。

同時，科技部人文司為瞭解數據科學在人文與社會科學的應用，探究量化研究的新實證典範發展，特別進行各學門在資料庫應用與數據科學分析的調查，包括常用資料庫來源、發表現況等（梁定澎，2017）。本文從非反應式研究角度，探究不同資料庫整合後，視覺化分析呈現的樣態。以體育運動資訊相關之運動現況調查、運動消費支出、體適能等資料庫為例，探究不同資料庫進行視覺化分析的樣態。

貳、非反應式之次級資料研究

實驗法和調查法多是反應式 (reactive) 研究，被調查和被實驗的對象知道他們被研究，且對象必須反應意見和反應刺激情境（關秉寅，2008）。非反應式研究 (non-reactive research) 則是指各類非反應式測量及非介入式觀察、內容分析，也包括既有統計 / 文獻資料庫等分析各類次級資料分析（關秉寅，2008）。次級資料分析是透過既存的大型研究資料庫（例如：中央研究院社會變遷調查）或政府統計資料（例如：工商普查）來對該研究進行深化實證探究的方法。在研究統計之資料庫中通常包含多期之調查資料（包含橫斷、縱貫性），因此能對研

究主題的跨時變化進行分析與探測。隨著統計系統技術之提升與網絡科技的進步，因此想取得公開次級資料庫進行統計數據之分析更是日益方便。

關於次級資料庫在數據科學分析上的應用，國立臺灣師範大學利用國中生體適能與基測資料庫分析探討臺灣7到9年級學生之身體適應能力以及學業方面成就之間的關係。研究透過40萬筆國中生身體適應能力與基本學力測驗之成績進行分析，發現國中三年級身體適應能力有持續上升的學生，基本學力測驗成績則越好，此研究結果則顯示了身體適應之能力表現越好，學業方面成就相對的也會變高的趨勢（許易旭，2016）。另外，倪瑛蓮、陳龍弘與張仁和（2017）為瞭解臺灣12至17歲青少年身體質量指數（BMI）分布情形及其肥胖盛行率，並探討社經地位與臺灣青少年肥胖關係。運用教育部體適能資料庫與中華民國統計資訊網之縣市家庭收支重要統計指標進行資料庫介接與分析，發現各年齡階段之青少年其縣市平均家戶收入與縣市體重過重及肥胖率大致呈現負向關聯。其他體育運動資料庫之應用分析也包括體育班與專任運動教練資料數據整合分析研究；應用地理資訊系統探討學校游泳池設置與學生溺水發生之關聯性（簡全亮、康正男，2016）；學校運動空間與學生體適能之相關研究（施登堯，2017）；國小體育專長教師比例與學生體適能變化情形研究（江詠宸、魏正、陳秀惠、徐振德，2017）。

更進一步而言，在資料庫分析中，資料視覺化是透過圖像化工具（例如：各種統計圖表、立體模型等）從複雜浩大的資料庫中萃取合適且可用之數據，經由轉化或介接分析後，進而成為簡易閱讀、容易理解的可靠訊息（O' Donoghue 等, 2010），資料視覺化可以快速提供即時的方式理解資料。

參、視覺化智能分析

資料收集、清理、儲存、分析與視覺化呈現是目前資料庫分析結果呈現的趨勢 (Markowetz, Blaszkiwicz, Montag, Switala, & Schlaepfer, 2014)。其中，視覺化呈現是資料庫智能分析結果呈現的最高層次。

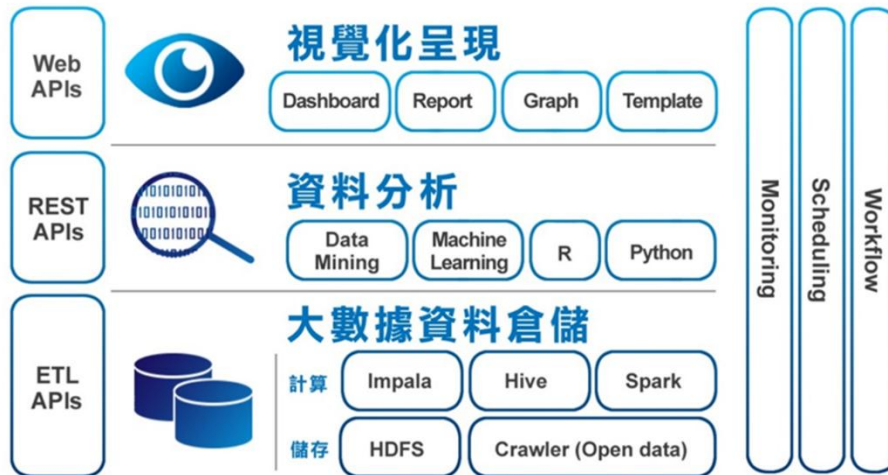


圖 1 數據分析的層次

一、資料視覺化 (Data Visualization)

資料視覺化指將數據資料圖形化 (例如：各統計圖表、立體模型等) 從浩大複雜的數據庫中篩選可用的資料，使其成為簡易閱讀、容易理解的訊息 (謝邦昌，2016)。人腦對視覺資訊的處理比書面資訊容易得多，透過圖表的方式來統整複雜的訊息，以便確保對關係的理解要比那些文字的報告或表格更迅速一些，資料視覺化工具可以提供即時資訊 (O' Halloran et al., 2016)。因此，以視覺方式 (例如：圖表、圖形、彩色地圖) 呈現資料以及分析複雜的統計數據會比用口頭或冗長的文字報告來的有效。

資料是資料視覺化的基礎，也是一開始最重要的工作。一般而言，資料的來源大多為開放資料 (Open data) 庫，這些資料是指篩選過與被允許的資料，這些資料不受其他管理機制所限制 (包括著作權與專利權)，可開放給任何人使用，

任何社會公眾都可以自由出版使用。經過大數據資料倉儲、資料探勘或文字探勘後，進行資料分析，最後再以視覺化方式呈現。視覺化呈現在智能化分析上有許多優點，首先是人腦對視覺的訊息理解分析要比書面資訊容易得多，使用圖表將繁雜的資料分析簡易化，便提供了一種清楚簡單的溝通方式。其次，視覺化呈現能與資料互動，與靜態圖表不同，互動式資料視覺化可以讓使用者依據不同變項和期間（例如：年份、週期等），呈現不同的趨勢分析，讓資料處理與分析更具智能。

二、體育運動資料庫視覺化分析樣態

在體育運動領域中，以 IBM 的視覺化分析為例，IBM 為美國網球公開賽 (US Open Tennis Championships) 開發官方 App，透過大數據即時分析各球員過去參賽之特質，能讓球迷更瞭解球員的基本資料、近況介紹與分析，亦可使球迷隨時隨地的掌握最新賽程與戰況，此 App 功能可利於使用者在賽前預測出每位運動員的致勝關鍵指標，亦可協助運動員瞭解自己的參賽效率值。此外，在 MLB Advanced Media 平台上，每天平均可產生 1,000 萬個串流，將其線上轉播的品質穩定化，支援各相關平台（例：手機、平板、電腦）讓閱聽眾不受任何地域的限制能隨時隨地收看 MLB 職棒賽事。之前獲得世界大賽小熊隊總裁營運總裁艾普斯坦 (Theo Epstein) 與運動神經認知科學家合作，開發一套像電腦遊戲般的系統，去測驗球員的手眼協調等能力，並進行數據分析，瞭解球員的潛在能力。

在國內的體育運動資料庫多達 20 個以上，其中有屬於競技型運動或全民運動型，也有屬於場館型或產業型的（見圖 2）。若能透過資料倉儲與介接方式，進行資料分析，進而視覺化呈現結果，便能充分應用體育運動資料庫的功能，瞭解更多的現況與趨勢，做為體育運動政策制訂的依據。

- 全國各級學校運動人才資料庫 (由中正大學)
- 運動產業產值調查
- 運動消費支出調查
- 裁判教練資料庫 (由體育總會、單項協會)
- 運動場館資訊網
- 運動員運動科學訓練 (由國訓中心)
- 運動場地器材與設備調查
- 奧亞運單項運動團體評鑑 (由台評會)
- 體育統計年報(由中山大學)
- 運動城市與現況調查
- 體適能檢測 (學生、國民)
- 學校體育績優團體及個人資訊網
- (由體育學會、輔仁大學)
- 臺灣體育運動詞庫
- 學生游泳能力121網站(由臺灣體育運動管理學會)
- 國光體育獎章資料庫 (由體育總會)
- 全國自行車道分布
- 學校運動志工服務平臺
- 運科與藥檢資料庫
- 運動學協會資料庫
- 運動健康資訊雲
- 企業贊助資料庫

圖 2 體育運動資料庫彙整

下圖3和4便是以教育部體育署運動現況調查結果和場館資訊網整合為例，可以呈現歷年參與運動人口趨勢分析和各縣市的分配（面積越大的縣市其參與人口或場館空間越大）。

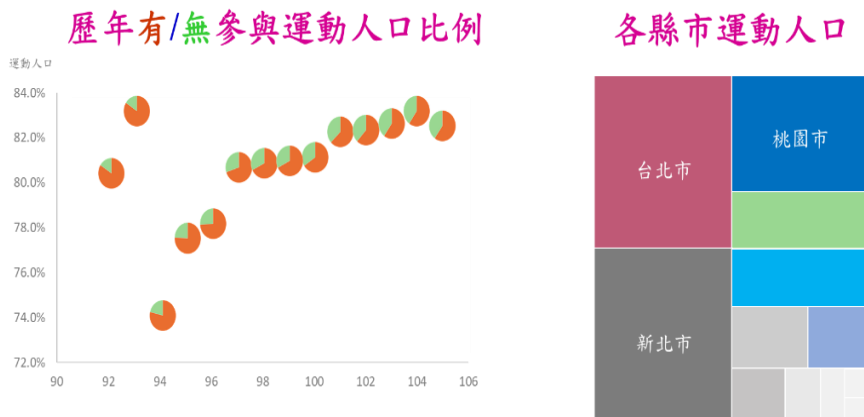


圖 3 臺灣歷年參與運動人口趨勢分析

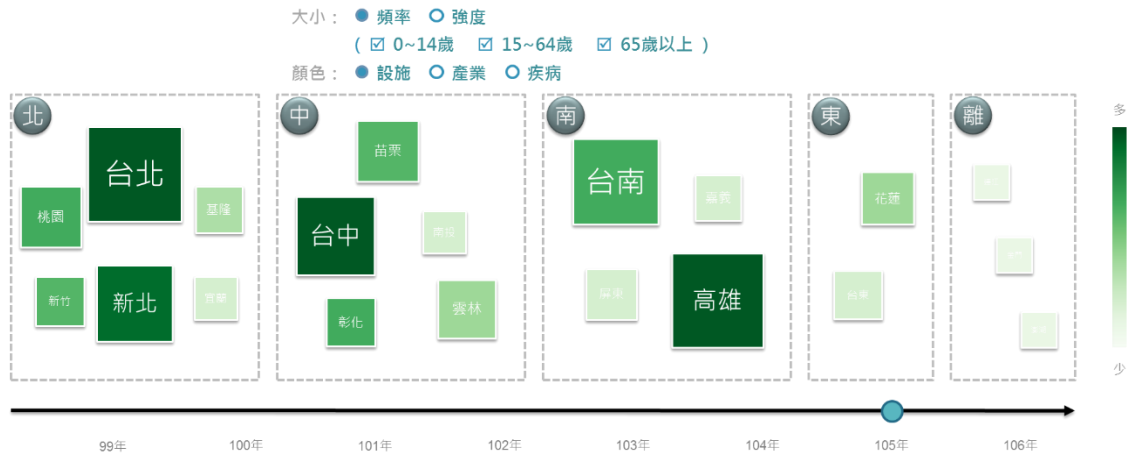


圖 4 臺灣參與運動人口與場館整合分析

肆、結語

視覺化的數據分析廣泛應用在各領域，例如：大型職業賽事（美國職棒、職業網球）在組織內部幾乎都安排有資料收集分析家也可稱為情蒐員，而情蒐員會根據球員與球的位置，彈跳的高度及其角度等，擬定出有效的技、戰術，或是分析球員身體能力之強度等級；將其用在電子商務領域中，可以得出各種特質的消費者會偏愛的產品種類；或者透過各個城市搜集來的空氣數據來得出污染源或空污指數等。但是，任何的資料分析及視覺化之前，應將資料再整理與計算，有乾淨、完整和有意義的資料，才能呈現真實的現況。根據研究目的與設計，將資料庫進行介接或合併，例如：歷年規律運動人口與縣市對應之趨勢；各縣市運動現況與消費支出之關聯。最後，從資料可呈現性的角度，以多元圖像方式呈現其資料視覺化分析，亦可展現出更有效果的資料訊息傳遞之能力，但是，從學術研究角度仍須強化資料的可檢驗性與價值性，並進一步進行量化的技術分析。

參考文獻

- 江詠宸、魏正、陳秀惠、徐振德 (2017) 。我國運動健身教練證照現況分析。 *運動管理*，(38)，47-64。
- 施登堯 (2017) 。建構優質學校體育學習環境>談提升體育教師專業養成與持續發展的政策介入。 *學校體育*，(160)，4-6。
- 倪瑛蓮、陳龍弘、張仁和 (2017) 。臺灣地區青少年身體質量指數、肥胖盛行率與社經地位的關聯：2007-2012 體適能資料庫追蹤。 *體育學報*，50(S)，47-60。
- 梁定澎 (2017) 。臺灣管理學術研究的一些回憶。 *人文與社會科學簡訊*，18(3)，175-179。
- 許易旭 (2016 年 5 月 18 日) 。 *運動補腦！研究：體適能與基測成績正相關*。風向新聞 | Kairos.news，取自 <https://kairos.news/36459>
- 許懷中、黃致豪 (2017) 。以大數據分析球員技術面表現、對戰組合與中華職棒歷年票房之相關性。 *體育學報*，50(S)，79-90。
- 謝邦昌 (2016) 。大數據分析中軌跡數據挖掘的現狀與挑戰。 *中國統計*，(8)，13-15。
- 謝邦昌 (2017) 。大數據和雲計算成為橋梁從此人腦接計算機的時代來臨。 *中國統計*，(9)，16-17。
- 簡全亮、康正男 (2016) 。開放水域戲水和游泳活動風險因子與其因應方式。 *學校體育*，(157)，34-43。
- 關秉寅(2008)。非反應式研究和次級資料分析。取自：<http://www3.nccu.edu.tw/~soci1005/nonreactive%20research.ppt>
- Church, R. M. (2001). The Effective Use of Secondary Data. *Learning and Motivation*, 33, 32-45.

Markowetz, A., Blaszkiewicz, K., Montag, C., Switala, & Schlaepfer, T. E. (2014).

Psycho-Informatics: Big Data shaping modern psychometrics. *Medical Hypotheses*, 82(4), 405-411.

O' Donoghue, S. I., Gavin, A., Gehlenborg, N., Goodsell, D. S., Hériché, J., Nielsen, C. B., North, C., Olson, A. J., Procter, J. B., Shattuck, D. W., Walter, T., & Wong, B. (2010). Visualizing biological data—now and in the future. *Nature Methods*, 7, S2–S4.

O'Halloran, K. L., Tan, S., Pham, D., Bateman, J., & Moere, A. V. (2016). A Digital Mixed Methods Research Design: Integrating Multimodal Analysis With Data Mining and Information Visualization for Big Data Analytics. *Journal of Mixed Methods Research*, 12(1), 11-30.