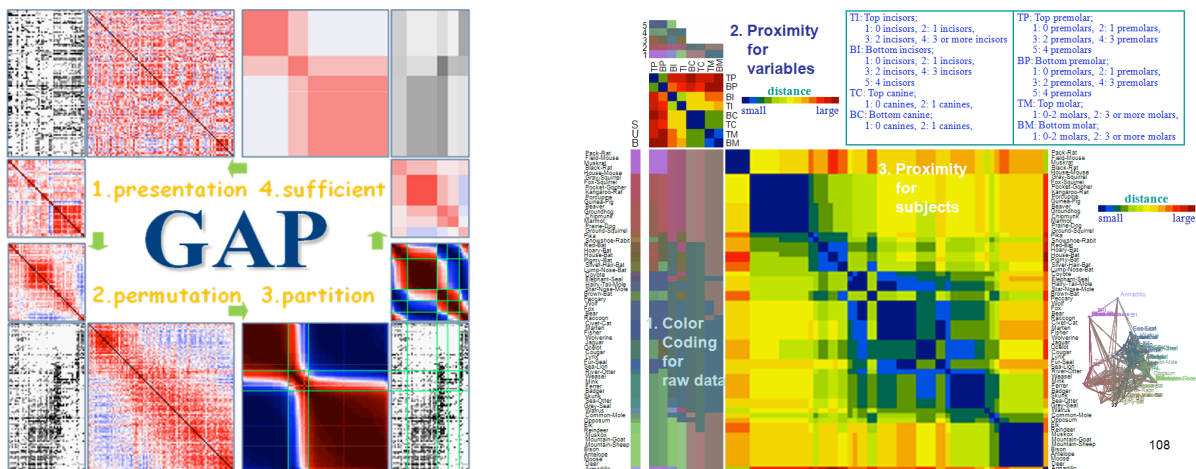


John W. Tukey 在探索式資料分析(Exploratory Data Analysis, EDA)書中開宗明義提到“*It is important to understand what you CAN DO before you learn to measure how WELL you seem to have DONE it.*” 學習可以做什麼，有助於在資料分析的過程中達到事半功倍的效果。EDA 的作用在於從「看」資料獲得資料所傳達的訊息，所注重的是簡單的算術與容易建構的圖、表。透過 EDA 對於圖表中所顯露的型樣(pattern)做一初步的認知與描述，再進一步以人類的心智(mind)對所接受的訊息做全面的分析與判斷，以探索潛藏於資料中的訊息。強調的是探索式的分析而非嚴謹的模式確認。

現代資料分析與統計應用面臨高維度與高複雜度之巨量資料 (BIG Data) 艱難挑戰，而現有之統計繪圖法與視覺化方法不足以讓使用者一窺資料之全貌，以選擇合適之分析方法與統計模型。EDA 已經問世四十年了，而箱型圖(box-and-whisker plot)與散佈圖(scatterplot) 仍然是二十一世紀最主要之繪圖與視覺化工具。發展有效率之繪圖法與視覺化環境以幫助使用者在處理由先進科技與複雜實驗所產生之資料時瞭解 what one CAN DO 乃統計學家重要職責。

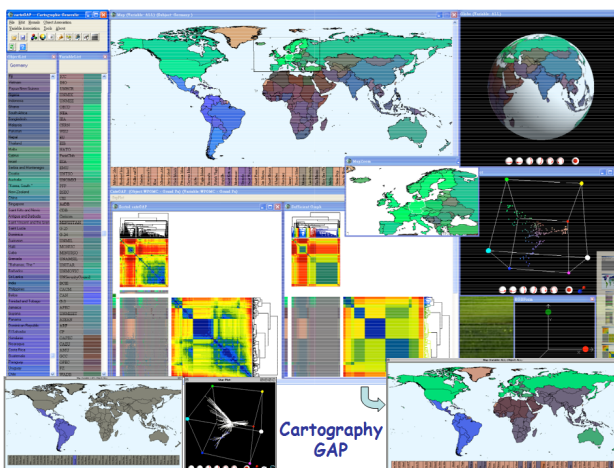
本演講將介紹相關研究團隊多年來發展之矩陣視覺化環境: Generalized Association Plots (GAP 廣義相關圖)。我們以矩陣視覺化來進行統計分析，包含敘述統計、建模、推論，以至於診斷；我們也將統計概念引進矩陣視覺化以處理更複雜之資料形態。GAP 可以處理上萬個樣本在上萬個維度(變數)的各式資料矩陣，包含連續、二元、有序、類別、象徵性 (symbolic data)、及與地圖學連結之各類資料。本演講將以深入淺出方式經由各種實例介紹各種主要 GAP 工作環境。有興趣者請瀏覽 GAP 網頁:
<http://gap.stat.sinica.edu.tw/>

Generalized Association Plots (GAP 廣義相關圖) for Matrix Visualization

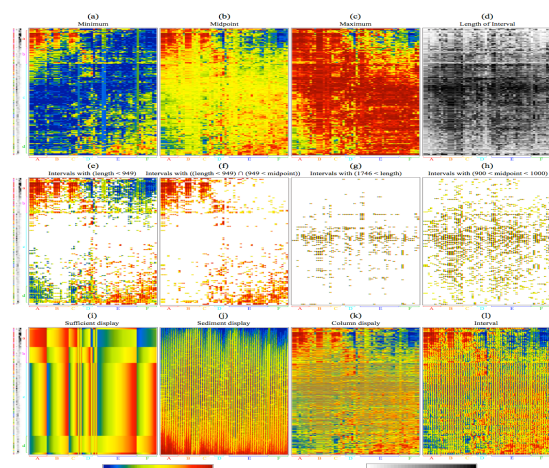


1. GAP 基本概念

2. GAP 類別性資料



3. GAP 地圖學資料



4. GAP 象徵性資料