巨量資料與統計分析 Fall 2017

第二次作業（2017/11/2繳交）

1. 除了上課提到阿基米德跟費曼面對問題的範例外，還有哪些名人應用創意思考？請提供2位名人與其事蹟，並詳列參考文獻及出處。
2. 不當使用大數據建立的模型，可能威脅個人生計、甚至造成重大損害，2008年金融大海嘯即是值得我們警惕的教訓，這類型大數據模型大多具有三項特質：不透明、大規模應用、造成傷害。請以各組在第一次作業中的應用為例，說明不當地套用大數據模型，會衍生出那些不公平現象及社會問題。（註：參考《大數據的傲慢與偏見》Weapons of Math Destruction。）
3. 資料偵錯、敘述性統計量（集中趨勢、散佈趨勢）、變數的關聯性分析，大致可視為資料探索性分析（EDA）的三個重要單元，適當地輔以圖形及表格的視覺化呈現，可以提供我們重要資訊。請以美國交通違規(Traffic Violation)的資料為目標，說明上述三項EDA的執行方式及重要成果。（註：同學可在我的網站下載資料，或是逕行至美國的開放資料網站data.gov查詢相關訊息<https://catalog.data.gov/dataset/traffic-violations-56dda>。另外，本題沒有指定分析方向，同學自行決定交通違規的探討議題。）
4. 大數據與我們生活的關係愈來愈密切，根據日常生活蒐集的資料，可以獲得尿布、啤酒之類的關聯性連結，再從這些分析結果推敲出高附加價值的應用。請各組同學以上一題交通違規的分析為藍本，說明哪些關聯性連結值得注意，並且討論如何從中獲致重要應用。