

課程編號：8N2033

全國大學校院夏季學院通識教育課程

課程計畫書

學校名稱	國立臺灣大學		
課程中文名稱	資料科學程式設計		
課程英文名稱	Data Science Programming		
授課教師姓名	蔡芸琇	單位/系所	共同教育中心 / 共同教育組

第一部份、課程規劃	
開課學校	國立臺灣大學
中文課程名稱	資料科學程式設計
全英語授課	<input checked="" type="checkbox"/> 否
課程屬性	<input checked="" type="checkbox"/> 基本能力課程
課程類別	<input checked="" type="checkbox"/> A 類討論課 <input checked="" type="checkbox"/> 翻轉教室
課程領域	CS+X 基本能力 (請依貴校通識課程領域分類方式,填寫本課程歸屬之領域別)
修課人數上限	150 人
特殊限制	★是否 開放 準大學生 (高三升大一之新生) 修課 <input checked="" type="checkbox"/> 是 ★是否 開放 研究生修課 <input checked="" type="checkbox"/> 是 ★是否 開放 旁聽 <input checked="" type="checkbox"/> 是
學分數	2 學分 (每學分上課時數 (含考試) 至少應滿 18 小時)
上課起迄日	107 年 07 月 04 日至 08 月 08 日
上課總週數	上課共 6 週, 是否連續每週排課? <input checked="" type="checkbox"/> 是, 上課時間連續數週不中斷
上課時間及節數	日期 7/4, 11, 18, 25, 8/1, 8 週三 9：10~16：20 (午休 1 節) 每次 6 節、共 6 次 上課時數共計 36 節
上課地點	<input checked="" type="checkbox"/> 國立臺灣大學校總區
是否提供住宿	<input checked="" type="checkbox"/> 否
課程目標	1. 拜寬頻網路普及化之賜，大數據熱潮來襲，許多資料科學相關的需求因應而出，《哈佛商業評論》將會寫程式、又有專業領域知識的資料科學家，封為「21 世紀最性感行業」。本課程結合 R 語言與其他電腦資訊技術，運用在各專業領域上。例如：以虛擬社群做社會行為的量化分析、以量化研究方法做人口統計學上的社會變遷、以質性研究方法針對虛擬社群做網誌調查、資料視覺化與結合地圖資訊系統等。  2. 想透過這門課，培養學生們成為台灣的資料英雄。進行資料分析不是為了要批評或鬥爭，而是想要讓我們所處在的社會可以透過資訊技能，看到問題點在哪裡，進而得到更好的決策參考。批評很容易，放炮也很簡單。最難的是要如何在舊有的體制內做軟性的創新，讓大家都願意一同進步，期待熱血的

	你們加入。		
教學內容 及進度	次別	上課日期/時間	課程內容
	1	107/07/04(三) 9:10-12:00	1. 上課規則、計分規則、作業說明、Github 教學。 2. 介紹 R 語言、上手 R 語言、如何進行翻轉教學。
	2	107/07/04(三) 13:20-16:10	1. R 語言的資料型態。 2. 生活中的資料科學。 3. 作業一：完成第一個 RMarkdown 的基本報告。
	3	107/07/11(三) 9:10-12:00	1. 活用 dplyr 與 data frame。 2. 上手 ggplot2 繪圖工具包。
	4	107/07/11(三) 13:20-16:10	1. 透過 R apply family 執行文本蒐集。 2. 詞頻與文字雲。 3. 主題討論一：使用 R 進行自動化分析，政治人物的臉書都在討論甚麼？詞頻與文字雲。 4. 作業二：使用 R 進行自動化分析，你所關注的臉書都在說些甚麼？詞頻與文字雲。
	5	107/07/18(三) 9:10-12:00	1. Term to Document Matrix 與文本分析。 2. TF-IDF 與文本分析。 3. 初探 PCA 主成分分析。 4. 透過 K-Means 分群演算法應用於文字分析中。
	6	107/07/18(三) 13:20-16:10	1. 主題討論二：使用 R 進行自動化分析，政治人物的臉書都在討論甚麼？時間變化與討論主題關聯。 2. 作業三：使用 R 進行自動化分析，你所關注的臉書都在說些甚麼？時間變化與討論主題關聯。
	7	107/07/25(三) 9:10-12:00	1. 如何使用統計工具於資料科學問題中。 2. 回歸分析、雙變數分析、ANOVA 基本概念介紹。
	8	107/07/25(三) 13:20-16:10	1. 如何使用機器學習工具於資料科學問題中。 2. 決策樹、SVM、NN 基本概念介紹。 3. 作業四：使用 R 進行自動化分析，你所關注的臉書都在說些甚麼？合併其他資料源進行文字與數值複合分析。
	9	107/08/01(三) 9:10-12:00	利用 R 和 Shiny 製作網頁作品。
	10	107/08/01(三) 13:20-16:10	1. 主題討論三：使用 R 進行自動化分析，政治人物的臉書都在討論甚麼？合併其他資料源進行文字與數值複合分析。透過 R Shiny 呈現網頁成品。 2. 作業五：使用 R 進行自動化分析，你所關注的臉書都在說些甚麼？合併其他資料源進行文字與數值複合分析。透過 R Shiny 呈現網頁成品。
	11	107/08/08(三) 9:10-12:00	個人作業分享
	12	107/08/08(三) 13:20-16:10	小組專題成果發表
教學助理規劃	請勾選教學助理類型，並預估需求人數： <input checked="" type="checkbox"/> 申請 A 類討論課教學助理，預估 TA <u>  2  </u> 人		

	<p>請說明運用教學助理之規劃：</p> <p>★設有 A 類教學助理之課程，請說明分組討論相關規劃。</p> <p>三次主題討論時段，由 TA 協助學生進行現場實作：</p> <ol style="list-style-type: none"> <li>1. 主題討論一：使用 R 進行自動化分析，政治人物的臉書都在討論甚麼？詞頻與文字雲。</li> <li>2. 主題討論二：使用 R 進行自動化分析，政治人物的臉書都在討論甚麼？時間變化與討論主題關聯。</li> <li>3. 主題討論三：使用 R 進行自動化分析，政治人物的臉書都在討論甚麼？合併其他資料源進行文字與數值複合分析。透過 R Shiny 呈現網頁成品。</li> </ol> <p>★翻轉教室請詳述討論題綱與進行方式等課程安排。</p> <ol style="list-style-type: none"> <li>1. 每週作業進度皆搭配指定用書所使用的網路課程內容，學生須完成指定用書所提供的線上課程內容閱讀，若製作作業過程中遇到問題，可向 TA 尋求協助。</li> <li>2. 當學生每週完成基礎語法自學後，並完成每週指定作業。於隔週課堂上才可順利進行主題討論，並較能吸收授課老師所安排的教學內容。</li> </ol>
指定用書	<ol style="list-style-type: none"> <li>1. <a href="https://csx.aca.ntu.edu.tw/1061CSX0014_01">https://csx.aca.ntu.edu.tw/1061CSX0014_01</a></li> <li>2. <a href="https://www.udacity.com/course/data-analysis-with-r--ud651">https://www.udacity.com/course/data-analysis-with-r--ud651</a></li> <li>3. <a href="https://in.udacity.com/course/data-analysis-with-r--ud651">https://in.udacity.com/course/data-analysis-with-r--ud651</a></li> <li>4. <a href="https://feeatherwind.shinyapps.io/final/">https://feeatherwind.shinyapps.io/final/</a></li> </ol>
參考書籍	<a href="http://datascienceandr.org/">http://datascienceandr.org/</a>
作業設計	<ol style="list-style-type: none"> <li>1. 作業一：完成第一個 RMarkdown 的基本報告。</li> <li>2. 作業二：使用 R 進行自動化分析，你所關注的臉書都在說些甚麼？詞頻與文字雲。</li> <li>3. 作業三：使用 R 進行自動化分析，你所關注的臉書都在說些甚麼？時間變化與討論主題關聯。</li> <li>4. 作業四：使用 R 進行自動化分析，你所關注的臉書都在說些甚麼？合併其他資料源進行文字與數值複合分析。</li> <li>5. 作業五：使用 R 進行自動化分析，你所關注的臉書都在說些甚麼？合併其他資料源進行文字與數值複合分析。透過 R Shiny 呈現網頁成品。</li> <li>6. 小組專題。</li> </ol>
成績評定方式	<ol style="list-style-type: none"> <li>1. 個人作業 60%，共 5 次，每次佔個人總分的 12%，每週上課作業公布，於隔週上課前繳交，每次作業分數依照指示給分，於公布作業當天公告，遲交斟酌扣分。</li> <li>2. 期末專題 40%，共 1 次，於第三週上課進行分組，至多五人一組，挑選出完成度前 10%的組別，於期末進行發表。</li> </ol>
特殊創意規劃	<ol style="list-style-type: none"> <li>1. 於課堂結束後，輔導完成度較高的學生，繼續參與以資料科學相關的競賽。例如：105-2 學期修畢台大資料科學程式設計的學生，其中台大校園用電專題作品完成度較高，授課教師於課程結束後繼續輔導該組學生參與</li> </ol>

	<p>2017 台電黑客松，榮獲第二名。</p> <p>2. 於課堂結束後，提供學生參與 Taiwan R User Group 閃電秀的機會。<a href="https://www.meetup.com/Taiwan-R/events/245806606/">https://www.meetup.com/Taiwan-R/events/245806606/</a>、<a href="https://www.meetup.com/Taiwan-R/">https://www.meetup.com/Taiwan-R/</a>。</p>
課程網址	(網址： <a href="http://homepage.ntu.edu.tw/~pecutsai/index.html">http://homepage.ntu.edu.tw/~pecutsai/index.html</a> )
其他補充資料	<p><a href="http://csx.ntu.edu.tw/">http://csx.ntu.edu.tw/</a></p> <p><a href="https://ntu-csx-datascience.github.io/NTUCSX-DS-FinalProject/">https://ntu-csx-datascience.github.io/NTUCSX-DS-FinalProject/</a></p> <p><a href="https://www.facebook.com/ntucsx/">https://www.facebook.com/ntucsx/</a></p> <p><a href="http://datantu.azurewebsites.net/">http://datantu.azurewebsites.net/</a></p>