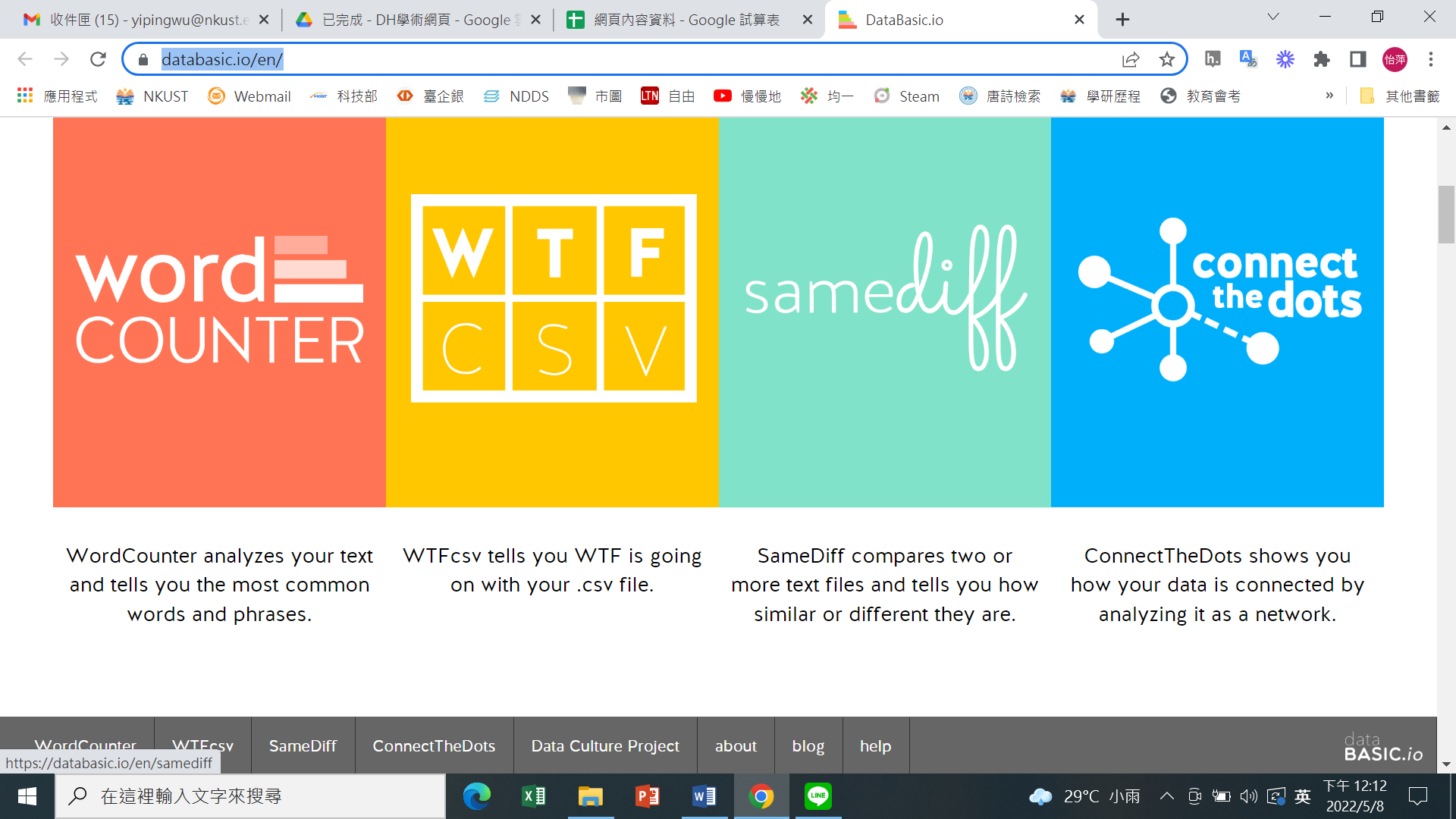
本文摘自[Rahul Bhargava](http://www.datatherapy.org/) and [Catherine D'Ignazio](http://www.kanarinka.com/)的〈[Data Basic](https://databasic.io/en/)〉，由侯湘珆統整後撰文。

Data Basic是一個適用於初學者的網路工具。撇除那些雖然非常好用但對於初學者來說很難懂的數據分析工具，Data Basic能夠提供初學者去執行基本的數據分析，幫助初學者們學習理解CSV(Comma-Separated Values，字元分隔值)中的內容，網站提供四個小工具WordCounter、WTFcsv、SameDiff和ConnectTheDots。



第一種工具是WordCounter，主要是用來分析文本中的單字和短句，即一個文本中使用頻率最高的單字及短句(包含二字短句及三字短句)是什麼，讓讀者們在一開始就能先對於所要讀的文本有基礎概念，或是利用這個方式去推算出故事主要內容圍繞在甚麼主題，而這一步驟通常也是量化文本分析的第一步。此工具操作起來非常的簡單易懂，有範例可以先讓使用者試做看看，不然也可以自已貼上文本內容、上傳檔案或是直接貼上連結進行操作，其中可以自己選擇是否要忽略文本中的大小寫及停用詞，選擇好之後直接按下「COUNT」就可以進行分析。另外除了英文的文本之外，中文文本也可以使用此工具，只是中文文本在分析字元上會有所偏差，會出現較長的句子。

第二種工具是WTFcsv，這個工具是以表格及圖表的方式來幫助讀者們分析內文，但這個工具的限制較多，只接受csv、xls和xlsx類型的文件，並且在使用上只能從網站所提供的範本去進行試用，或是以上傳檔案的方式進行分析。雖然中英文文本都是可以使用的，但也並非是所有excel均能正常操作，像是報表類的就無法使用。舉例來說，裡面有一個範本是鐵達尼號乘客名單，這項工具可以從檔案分析出乘客的編號、性別、年齡、票價等，以圖表方式呈現，讓讀者可以快速地去對應到想要知道的資訊，名單式的檔案是相對容易執行的。

第三種工具是SameDiff，這個工具可以把兩個文本(包還歌詞、文章、演講)去做比較，可以看出兩個文本之間同時都有出現的字以及頻率有多少，另外還可以單獨看到只有各自文本中才出現的字，可以把這個工具所呈現的結果下載下來後去做更深入的探討及分析，例如﹕為什麼兩個文本會同時都出現相同的文字?是否代表了某種意義?但這個工具主要用於英文文本(最好連檔案的名稱都使用英文)並且為txt、docx、rtf格式的檔案，不然呈現出來的結果並不理想，此頁面也有範例可供使用者練習。

第四種工具是ConnectTheDots，它是透過節點及邊去呈現出一個網路圖，然後顯示出人事物跟其他人事物之間的關係，中英文文本都可以進行操作，但這個工具僅限用在csv、xls、xlsx類型的文件或是直接複製資料上去，並且這些文件需要是事物和事物之間互相有關係，或是需呈現兩列排列的數據才能進行分析，此頁面同樣也有範本可供使用者練習，也可以自行複製自己的數據資料或是直接上傳文件。分析結果出來的時候每個分類會有不同的顏色可以讓使用者更清楚地去做區分，當然如果不需要去做區分的話也可以從community轉換成n/a，這樣點的地方顏色就不會做區分了。

這個網站提供簡單的工具與實作範例，相當適合初學者使用，可作為數位人文課程的基礎入門工具。