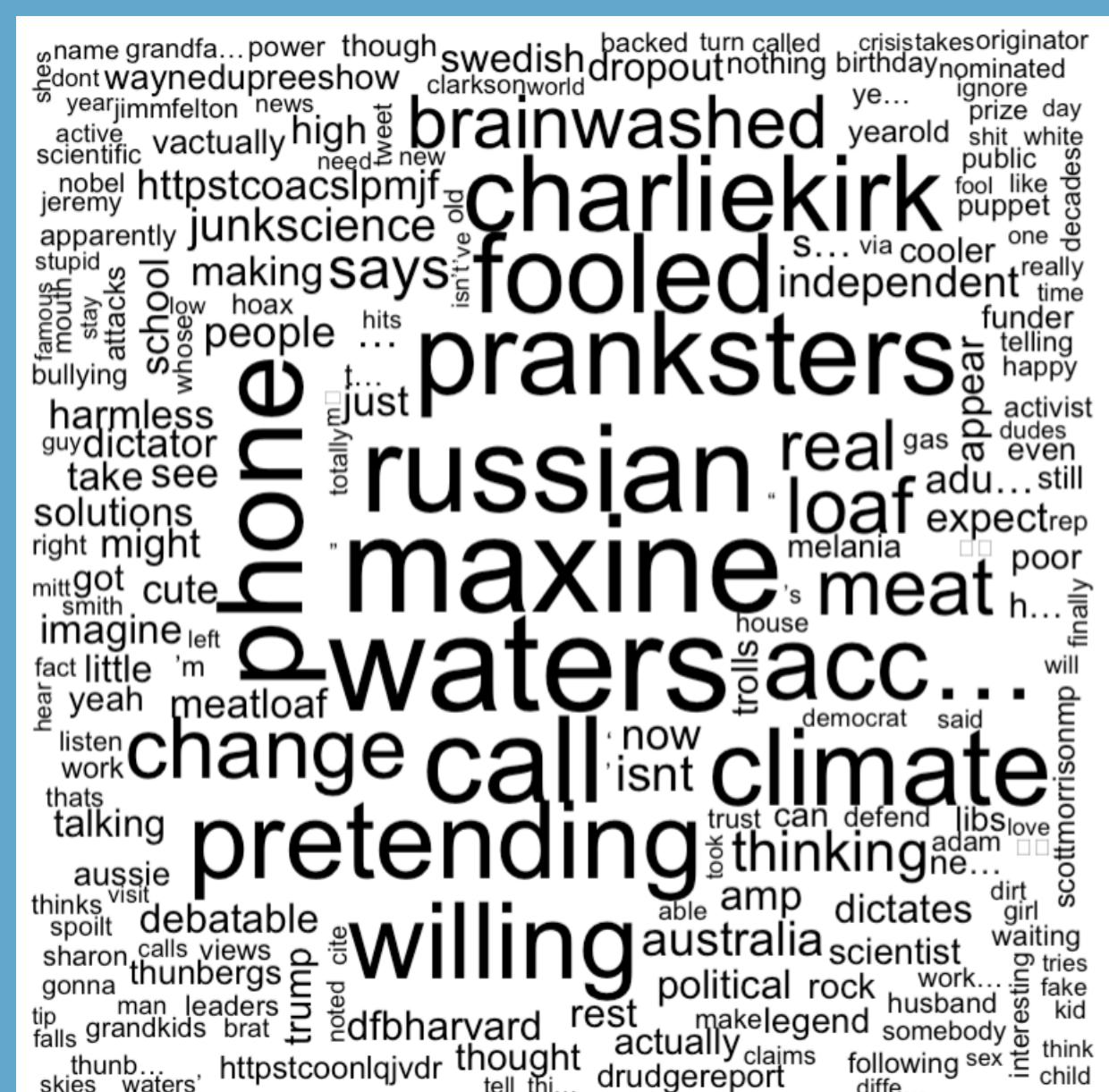




Twitter資料探勘： 環保少女與香港示威者討論熱度比較

第十六組【一七】 經濟一 黃顥 | 經濟一 陳宇澤 | 經濟一 葉秀軒 | 會計一 蕭羽騏



1000筆分別出現「greta thunberg」及「hong kong」貼文中的高頻字



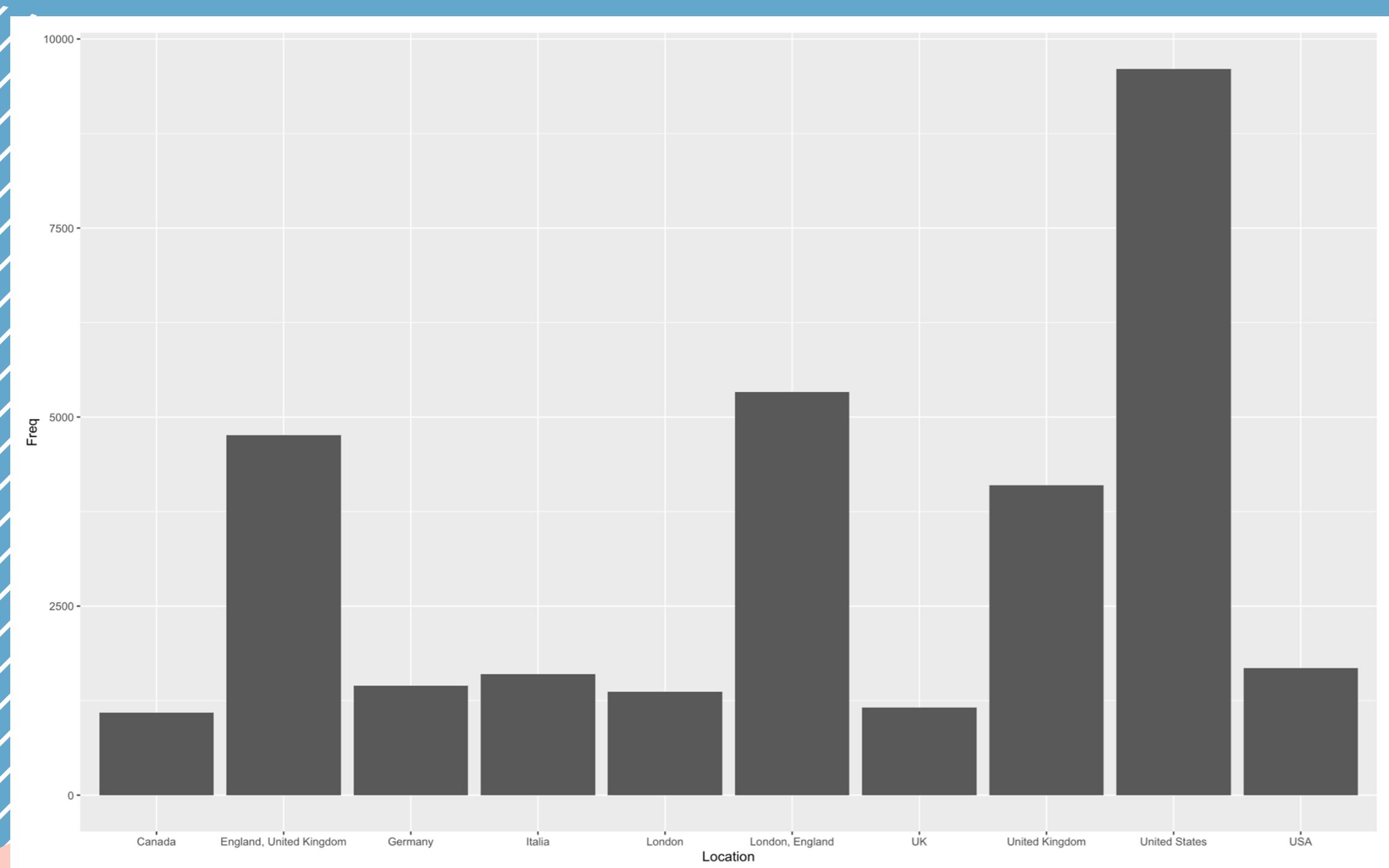
Goals

備受爭議的瑞典環保少女——格蕾塔·桑伯格（Greta Thunberg），擊敗網路投票票數最高的「香港示威者」，獲頒時代雜誌2019年度風雲人物。我們希望藉由社群平台Twitter龐大的資料庫來一窺桑伯格在社群上獲得的評價以及討論熱度，並以香港示威者做為對照。

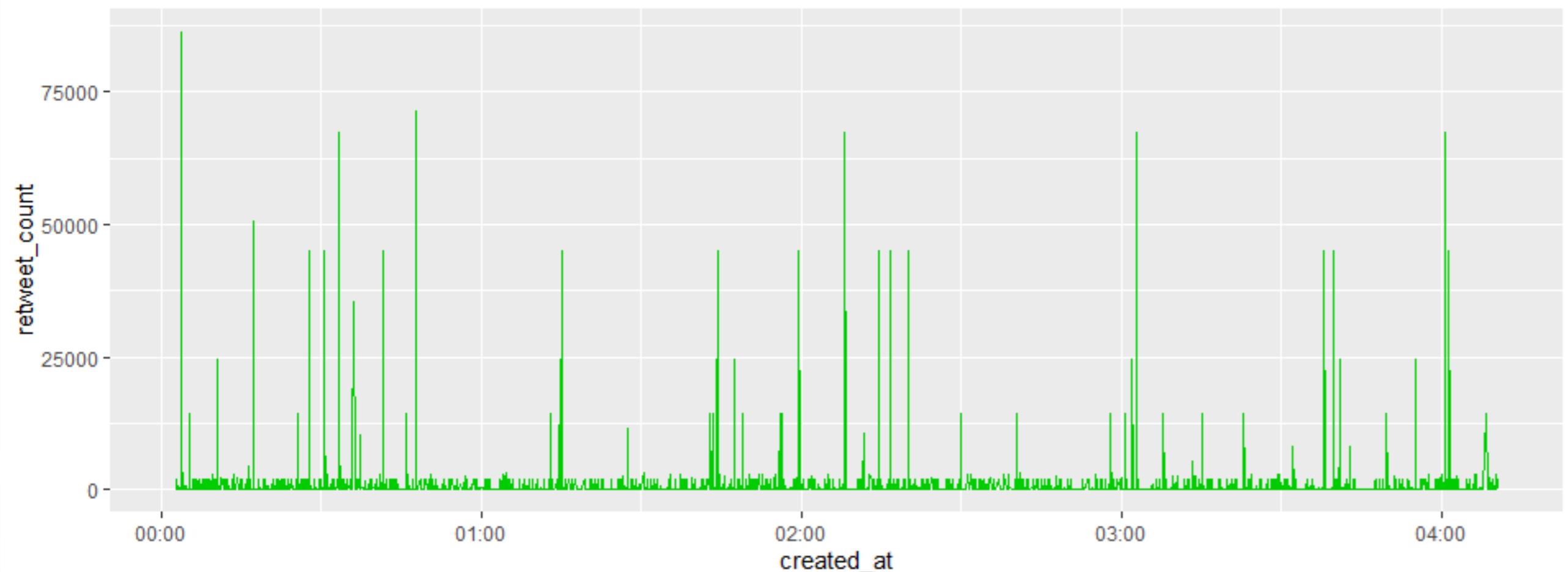
Methods

文字雲的部分主要運用 tm 以及 twitteR 套件，其餘統計圖表則利用 rtweet 與 tidyverse 進行資料的爬取與整理，加以ggplot2製圖。另外，也需要 Twitter API 以及 Google API。

* 研究限制：Twitter API 資料爬取僅限7-9天內之資料



10000則貼文中「#gretathunberg」的發訊區域（ location ）統計



近3200則提及「Greta Thunberg」的貼文中，被轉發（retweet）的次數

Results

根據長條圖，可以發現多數「#hongkong」來自香港本地，相對於「#gretathunberg」，呈現較為區域化的現象，亞洲國家感受度相對強烈。另外，透過文字雲可以發現與桑伯格相關的高頻字出現「brainwashed」、「fooled」等等較為負面之詞彙；香港部分則出現意料之中的「police」、「arrested」等字詞。最後，再由折線圖可得知，隨著時間變化，桑伯格的關注度是相對來說較穩定的，香港會因為特定的文突然增加關注度（特定的時間點或發生特定事件），但是就每篇文被retweet的總數來看，香港是大於桑伯格（以爬取資料當下的結果而言）。