

生生不行 and happy の孕將



第八組 好想停修

外文四 陳玠青 外文四 蔡宜軒
外文四 吳尚勳 經濟四 盛子芸

2021-06-23

目錄

簡介	3
方法	4
結果	6
討論與貢獻	10
組員分工表	10

簡介

(一) 動機：

在今年四月美國中情局發布的報告顯示，臺灣每位育齡婦女平均只生1.07個小孩，臺灣的生育率在世界敬陪末座。另外，在2020年時，臺灣的生育率也首次低於死亡率，國家進入了史無前例的人口負成長階段。這個消息引起了軒然大波，許多人開始探討大家為何越生越少，甚至是不生小孩。而我們也閱讀了數篇相關報導與研究，對此我們發現，多數人將低生育率的現象歸因於高房價，然而有研究指出，高房價的都會區未必是生育率最低的區域；以過去的資料來看，房價上漲的時期也未伴隨著生育率的下降。除此之外，也有報導推測，少子化與已婚夫妻生或不生小孩的關聯性不大，關鍵點在於女性逐漸晚婚或選擇不婚相關。由於大家對於生小孩的意願有所不同，各自也持有不同的考量因素，因此我們想藉著這個報告，了解更多人的想法，並歸納在生育與否的背後，多數人權衡的因素為何。

(二) 想回答的問題：

我們想知道社會大眾對於生小孩的看法，對於生或不生大家各自有什麼考量。

1. 想生/不想生小孩的討論熱度？

由於生育率低已是事實，對此我們假設不想生的人多於想生的人。

2. 想生/不想生小孩的理由？

由於不想生的人占多數，我們推測在網路上的討論中，會有一大部分的人在文章中提出「想生的人在想什麼」、「為什麼有人想生小孩」這類的問題，所以我們假設「想生」的關鍵詞會多於「不想生」的關鍵詞。此外，因為事前閱讀的資料將生育率低迷的現象歸因於女性逐漸晚婚，同時我們也發現臺灣非婚生子女所占比例極低，所以對於我們假設婚姻與生子的議題息息相關。

方法

這份原始碼是在"R version 4.1.0"底下運行

(一) 資料取得(PTT網頁爬蟲)

1. 使用套件: `httr`, `rvest`, `tidyverse`, `readr`, `stringr`, `tm`

2. 使用函數

`get_content_link`: 取得文章超連結

`get_data`: 獲得文章的作者、標題、日期、內文

3. 變數

`rep1`: 輸入標題關鍵字「生小孩」

`rep2`: 輸入標題關鍵字「養小孩」

`rep3`: 輸入標題關鍵字「生子」

`rep4`: 輸入標題關鍵字「生孩子」

`content_ques`: 將收集到的資料(共392篇文章), 資料年份為2017~2021年, 進行資料清理, 去除日期、網頁、引述、全形標點符號、半形標點符號、空白及斷行符號。

4. 將整理好的資料儲存至"Womentalk_content.csv"

(二) 原始碼運作說明

a. 人工標記

(1) 將整理好的資料進行人工標記

(2) 將文章分為「無關」、「想生」、「不想生」、「要生」、「不要生」、「隨便」、「隨便」七大類, 並將「無關」類文章(共232篇)刪除, 將剩下的160篇文章存入

"content.csv"

(3) 分類依據:

I. 想生、不想生指得是是自己對於生小孩的意願

II. 要生、不要生則可能是告誡別人, 或是講述自己未來的計畫(將要...)

III. 猶豫是發文者在文中分析生與不生的利弊, 但是無法下決定, 想詢問大家意見

IV. 隨便是發文者不在乎生或不生, 或是抱持都可以的看法

b. Ngram處理

(1) 使用函數`get_ngrams`把清理後的內文做bigram~heptagram

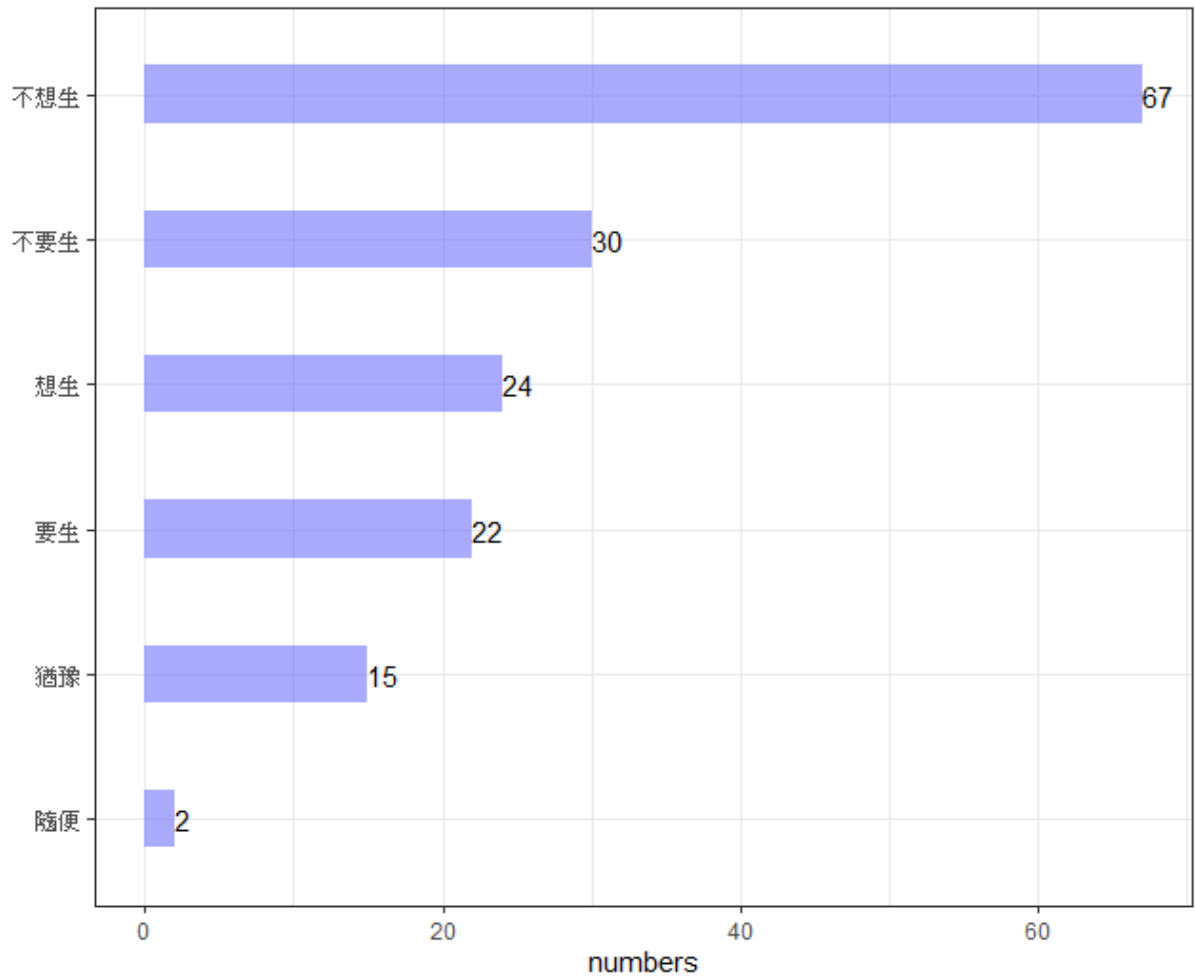
- (2) 篩選有「不」的關鍵詞，存成"nwords.csv"
- (3) 人工篩選"nwords.csv"中詞頻為2以上，和「想不想生」相關的詞彙。
- (4) 因為用ngram會有正向關鍵字重複計算的情況，所以在人工篩選完反向關鍵字("nwords.csv")後，將這些關鍵詞存成"keywords.csv"。
- (5) 將"keywords.csv"中的「不」去除，存成positive_words，再利用filter函數篩選包含這些正面詞彙的詞頻。
- (6) 因無法使正反面詞彙對稱，用人工的方式相減得出最後確切的詞頻，存成"birth_keywords.csv"。

c. 視覺化

- (1) 使用套件: ggplot2
- (2) 繪製分類標記圖: 將人工標記的類別繪製成長條圖，輸出成"f1.png"
 - a. 因為「隨便」類別只有兩篇，數量過少，所以並沒有進行後續詞頻比較
- (3) 利用前面相減出的"birth_keywords.csv"畫詞頻前十名長條圖:
 - a. 不生相關負面關鍵詞: "f2.png"
 - b. 想生相關正面關鍵詞: "f3.png"
- (4) 原因部分利用自建的"user_dict.txt"，篩選我們想要看到的可能原因，因為各類文章數能找出來的關鍵字不多，且類別過多較難比較，因此最後我們決定把想生和要生合併，不想生和不要生合併後，進行繪圖，最後呈現三張以類別分類的詞頻前十名長條圖
 - a. 想生相關原因關鍵詞: "f4.png"
 - b. 不想生相關原因關鍵詞: "f5.png"
 - c. 猶豫相關原因關鍵詞: "f6.png"

結果

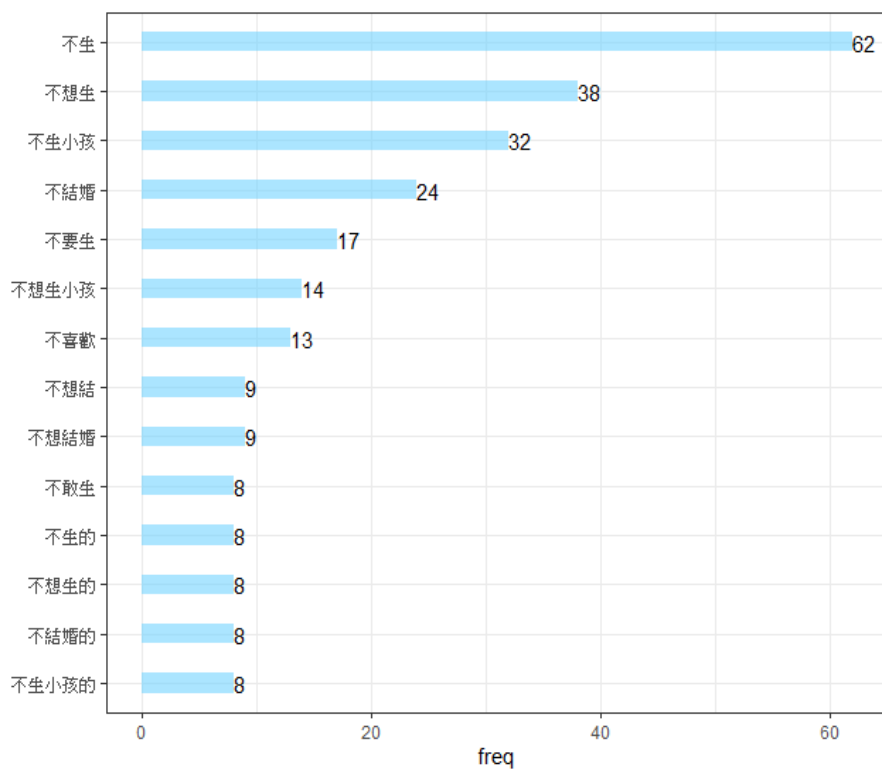
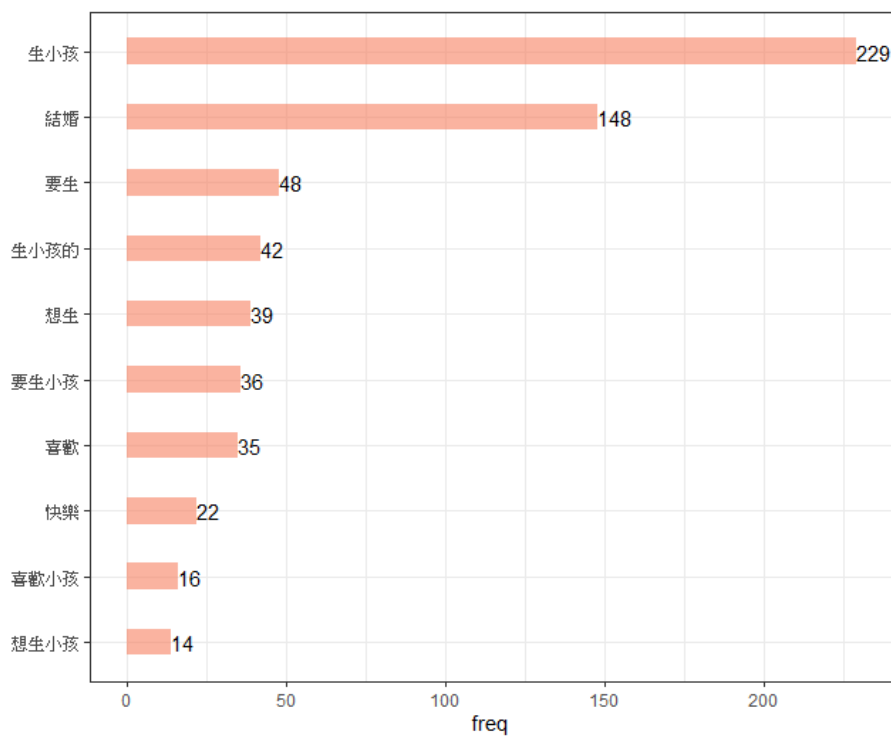
(一)想生 vs. 不想生 文章比例



1. 在版上討論不想生的文章居多

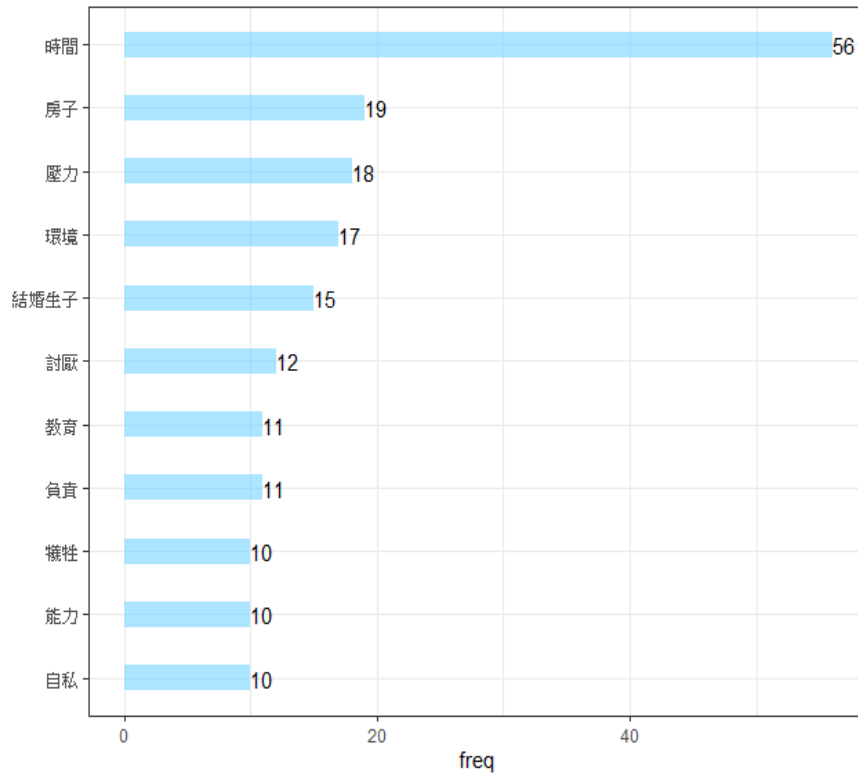
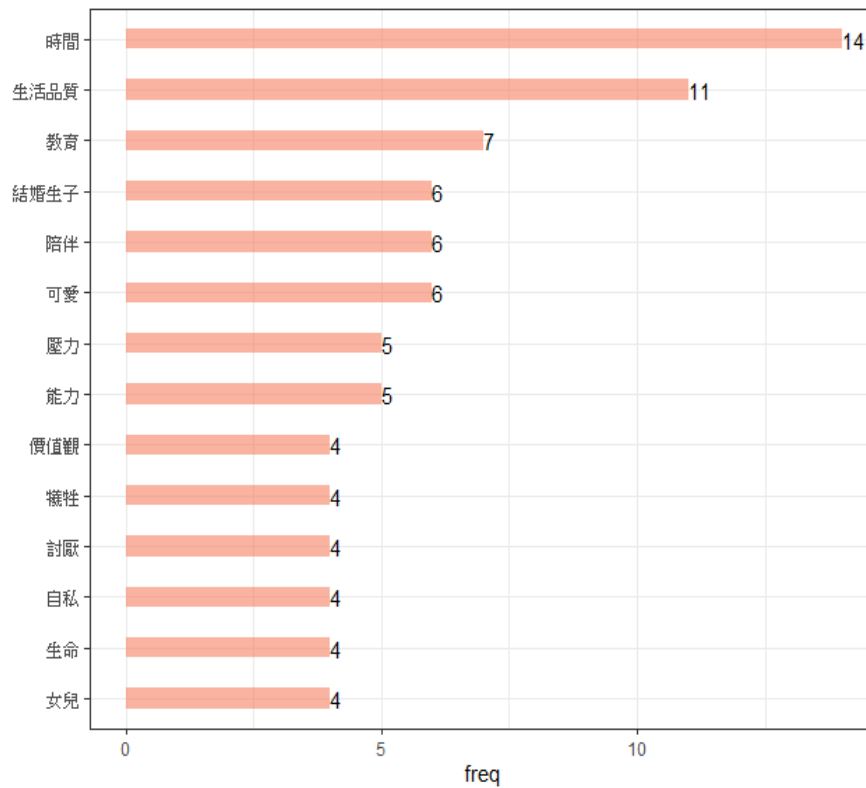
a. 不想生:想生 = 97:46

(二)想生 vs. 不想生 文章類型比較



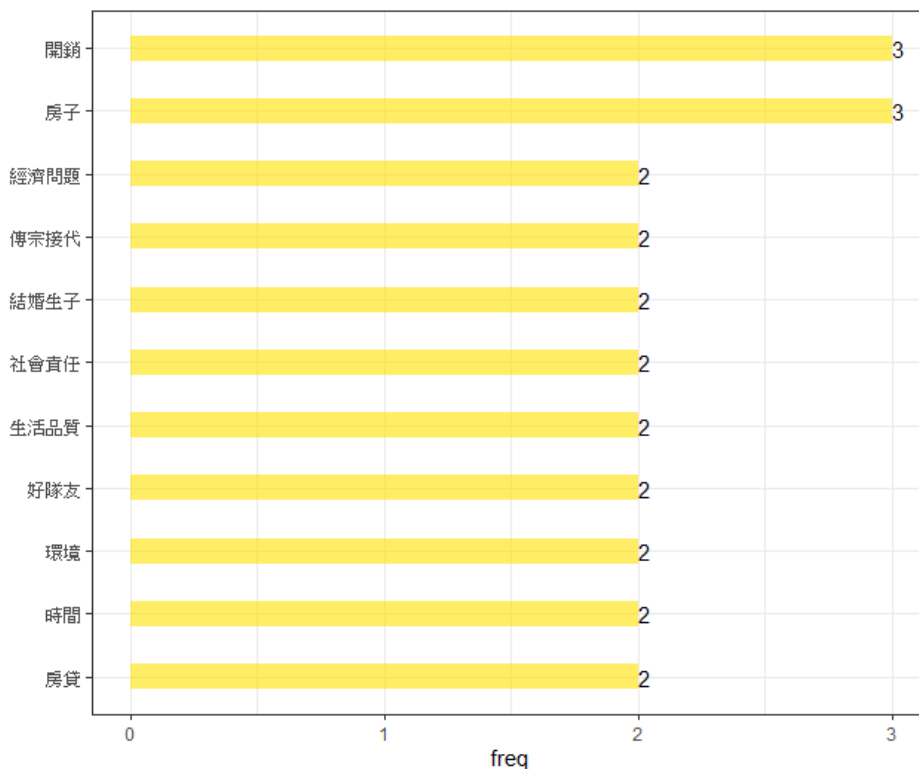
1. 討論「生」或「不生」的文章中，都有被「結婚」和本身喜不喜歡小孩影響。
2. 其中想生的人則更常討論「喜歡小孩」。

(三)想生 vs. 不想生 關鍵詞比較



1. 在兩類文章中有明顯最高討論度是「時間」因素，而和「時間」這個關鍵詞一同出現的句子大致上有自己的時間(4)、時間到了(3)、時間成本、大量的時間(2) (括弧內的數字是出現頻率)
 - a. 想生和不想生對時間討論各有不同
 - i. 不想生:生小孩需要花時間、犧牲自己的時間
 - ii. 想生:時間到了就會想生
2. 再來共同有討論的因素是:壓力、教育、結婚、能力, 和較低頻的討厭、犧牲、自私。
3. 而各類分別單獨提到的詞如下
 - a. 只有不想生提到:房子、環境、負責。
 - b. 只有想生提到:生活品質、價值觀、陪伴、可愛、生命、女兒
 - i. 同時, 想生文章的高頻詞中完全沒有「房子或房貸」的討論。

(四)猶豫的關鍵詞



1. 猶豫的人主要考量在開銷(房子、經濟等)因素。
2. 猶豫的文章抓下來的數量只有15篇, 因為這個類別的樣本比較少, 所以和前面兩個類別相比, 猶豫的文章討論關鍵詞出現的頻率也較低, 在2到3次區間。

討論與貢獻

1. 在版上討論不想生的文章居多，不想生的人可能承擔「理所當然應該生孩子」的社會期待，為了支持自己的想法而發文尋求認同，甚至對於和自己抱持不同意見的人感到好奇，因而發文詢問大家為什麼想生小孩。而想生本身較能順社會期待，則不需要特別討論。
2. 在討論生小孩的文章中，結婚這個因素也息息相關，除了「結婚生子」一詞時常併用之外，重點討論的面向還有對於不想生的人，會擔心結婚後有生小孩的壓力，對於想生的人，則會計劃結婚多久要生。
3. 在兩類文章中有明顯最高討論度是「時間」因素，這和先前新聞報導的主要原因「晚婚與房貸」有所差異，或許在政策上除了鼓勵結婚與減輕經濟壓力之外，這個世代給予父母親有喘息機會、有自己的時間追求理想也很重要。
4. 若想要針對不想生的人的生育意願，可以針對在不想生文章中獨有的「房子」、「環境」、「負責」做出對策，而在想生的文章中，則發現沒有討論到房子的關鍵字，推測是因為排除了這個阻礙，可能就增加想生的意願。
5. 只有想生提到「女兒」這個詞，隱約可以看出當大家在討論生子問題時，已經不如過去重男輕女，甚至有部分的人更想生女兒。

組員分工表

	內容	投入程度
陳玠青	爬資料、分類文章、製作 user.dict、n-gram、畫圖	6
蔡宜軒	爬資料、分類文章、製作 user.dict、n-gram、簡報	6
吳尚勳	爬資料、分類文章、製作 user.dict、n-gram、R markdown	6
盛子芸	爬資料、分類文章、製作 user.dict、n-gram、簡報	6