劇本裡的千層套路

從情感分析談談電影



Outline

• 主題介紹

• 方法概述

• 結果分析

• 限制與展望

主題介紹

簡介

看電影的時候,有沒有一種感覺好像這個橋段跟某部電影好像:

男女主角雖然翻臉了,但最後還是復合了。 這個時候部隊是不是會被某處的敵軍偷襲? 怪物才不會那麼容易就死掉……牠果然還活著!

好奇同類型劇本是否存在某種發展模式嗎?透過情感分析來一探究竟吧。



分析內容

本專案挑選三個電影種類-懸疑片、動作片和愛情片,每種種類大約有6~7篇劇本。

以下是我想討論的主題:

驚悚片:我們如何預防不被鬼嚇到一驚悚畫面的分布點。

戰爭片:戰爭中的煙硝味—衝突情緒被安排在電影中哪個環節?

愛情片:你的愛情計畫通不只是因為你帥(美),而是你吵得起架。



影片挑選

HOLIVWOOD

驚悚片	戰爭片	愛情片
黑天鵝	美國狙擊手	回到17歲
逃出絕命鎮	梅爾吉勃遜之英雄本色	美女與野獸
戰慄遊戲	末代武士	約會喔麥尬
孤兒怨	珍珠港	愛情昏迷中
驚魂記	搶救雷恩大兵	鐵達尼號
火線追緝令	決戰時刻	暮光之城
沉默的羔羊		暮光之城2:新月

方法概述

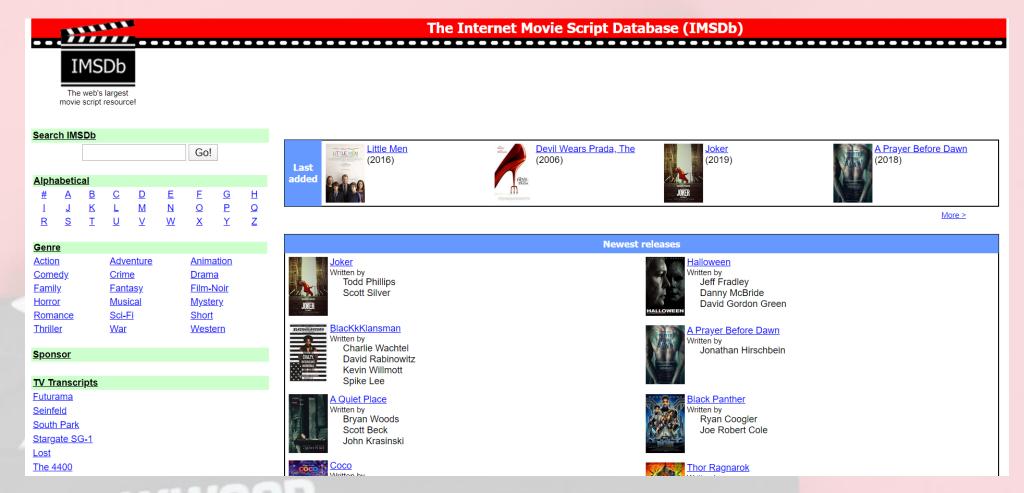
流程

- 1. 劇本蒐集
- 2. 文本預處理
- 3. 分析處理
- 4. 視覺化、動態化

本專案研究方法參考 Shuyi Wang:如何用Python和R對故事情節做情緒分析?

劇本蒐集

來源: The Internet Movie Script Database (IMSDb)



文本預處理

將文本分為列表並輸出成CSV檔,經過觀察,大多數的劇本會用數字標示篇章或是使用 "INT., EXT., INSERT."標示場景轉移,我將利用這些標記來分割文本。

主要使用函數: strsplit()

兩種正規式:

1. 以數字標示作為分割

"
$$[0-9]$$
\\. | $[0-9][0-9]$ \\. | $[0-9][0-9][0-9]$ \\." 或 " $[0-9][0-9][0-9]$ \\" 或 " $[0-9][0-9][0-9]$ "

- 2. 以場景轉移標示作為分割
- "INT. | EXT. | INSERT. "

處理後文本展示

	line	content
1	mic	1 I appeal to you therefore, brothers, by the mercies of God, to present your bodies as a living sacrifice, holy and acceptable to God, which is your spiritual worship. Do not be conform
1		
2		2 THE SHAW'S HOUSE - FRONT LAWN - NIGHT A perfect suburban house with bay windows and a front lawn. The SHAW family. Caucasian and warm - RICHARD, 34; NAN
3		3 SUBURBAN STREET - NIGHT ANDRE, 29, an African-American man runs down the sidewalk in sweats. He listens to jazz on his phone. The music stops. He stops running and
4		4 THE SHAW'S HOUSE - DINING ROOM - CONTINUOUS Richard looks up from his table to see Andre standing there. The rest of the family doesn't notice. Nancy tends to May
5		5 SUBURBAN STREET - CONTINUOUS Andre walks on. It's getting dark. He is suddenly very alone. A vintage creme-colored Porsche Boxter with tinted windows and a roof CR
6		6 SPORTS CAR - CONTINUOUS Driver's POV watching Andre. His BREATH ECHOES deep and tinny as if it were into a coffee can. Through the car's system we hear an English
7		7 SUBURBAN STREET - CONTINUOUS Andre, feeling followed, stops to tie his shoe. The car also stops. Andre waves at the unseen DRIVER. There is no response. The ENGIN
8		8 SPORTS CAR - CONTINUOUS Driver's POV. Andre continues to talk, but isn't heard. TUTORIAL Can you show me the nearest bathroom? Pouvez-vous me montrer la salle de
9		9 SUBURBAN STREET - CONTINUOUS Andre hears nothing from inside. ANDRE (under his breath) This is some shit right here. Andre begins walking. The car creeps along
10	1	10 SPORTS CAR - CONTINUOUS Driver's POV. Andre bangs on the window.
11	1	31 SUBURBAN STREET - CONTINUOUS Andre bangs harder on the window three times. The car's passenger-side window rolls down. TUTORIAL Can you help me find a hotel?
12	1	12 THE SHAW'S HOUSE - FRONT LAWN - CONTINUOUS Andre falls in front of the Shaw's lawn. A dart sticks out of his back. He reaches towards the Shaw family. The masked
13	1	13 THE SHAW'S HOUSE - DINING ROOM - CONTINUOUS The Shaw family eats while Andre lays on their front lawn. This time Richard, buried in his tablet, doesn't notice. the
14	1	14 SUBURBAN STREET - CONTINUOUS The driver carries Andre to the car. JOSHUA (O.S.) Why? RICHARD (O.S.) Because he hasn't aged in 100 years. The dri
15	1	15 BROOKLYN PARK SLOPE - DAWN The sun rises over the city. Autumn. Beautiful.
16	1	16 BROOKLYN LOFT - LIVING ROOM - MORNING We move slowly through the small but clean apartment. The walls are decorated with fascinating urban photography.
17	1	17 BROOKLYN LOFT - BATHROOM - MORNING CHRIS WASHINGTON, 26, a handsome African-American man shuts the medicine cabinet. He's shirtless and naturally athletic
18	1	18 STARBUCKS COUNTER - MORNING ROSE ARMITAGE. 28 - Caucasian. brunette with freckles - cool and beautiful like an old Summer Camp crush. Rose looks at pastries th

分析處理

- 1. 讀取已處理文本
- 2. 使用tidytext套件中的unnest_tokens函數將句子分解成單詞
- 3. 使用tidytext內建的停用詞並與自訂停用詞結合成新的停用詞集
- 4. 進行情感分析
- 5. 将分析結果匯出成CSV檔
- 6. 視覺化、動態化

將分析結果匯出成CSV檔

註:

本專案進行多篇劇本集合觀察,為了避免視覺化後難以判讀,因此這些情感分析數據需要平滑化。本專案調整每部劇的分割段落侷限在25~30的長度。因此使用:

mutate(paragraph = (line %/% 調校值) + 1)



視覺化、動態化

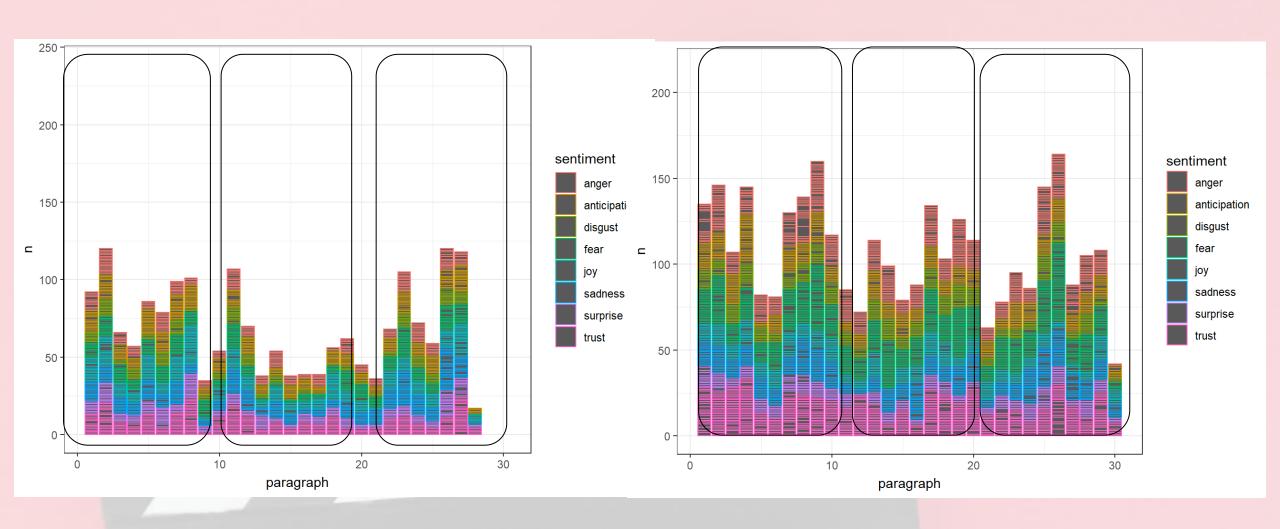
- 1. 在所有情感分析檔案中新增欄位作為電影名稱標記
- 2. 對於過高的情感值進行調校。
- 3. 將所有情感分析檔案合成總表。
- 4. 使用gganimate套件的transition_states生成動態圖。

註:對過高的情感值進行調校是由於本專案進行多篇的動態化,不同的頻率規模會導致圖表y軸數值根據頻率最高的劇本最高值而定,使頻率規模較低的劇本情感值起伏十分不明顯,影響可看性。



結果分析

好萊塢劇本風格

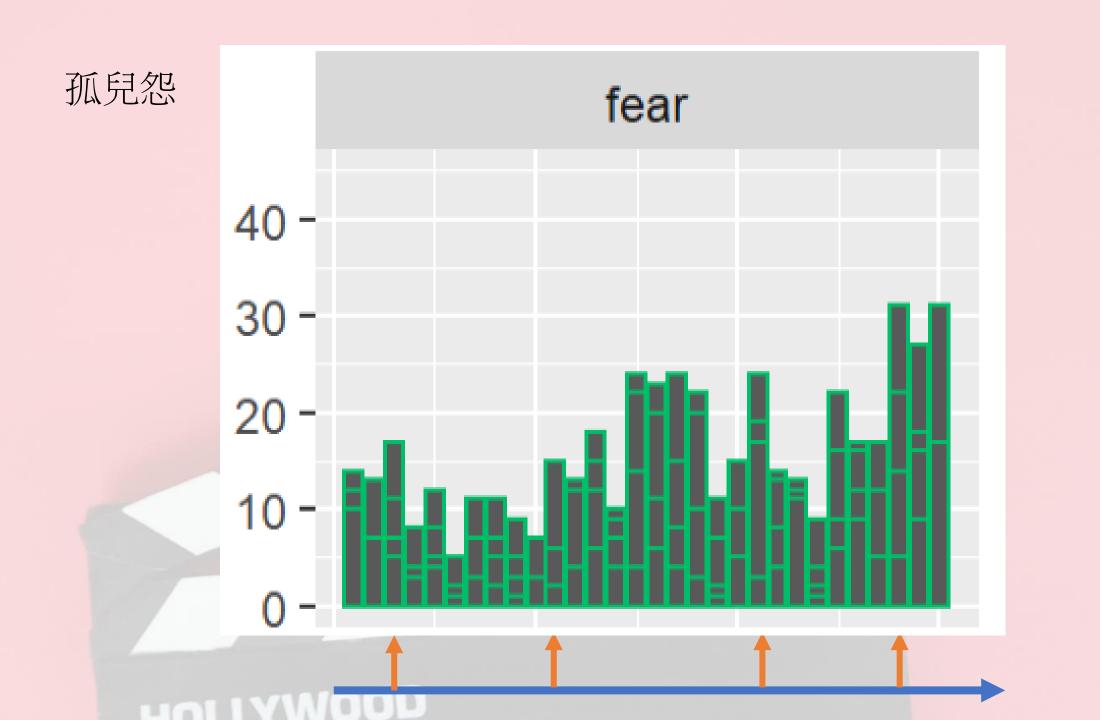


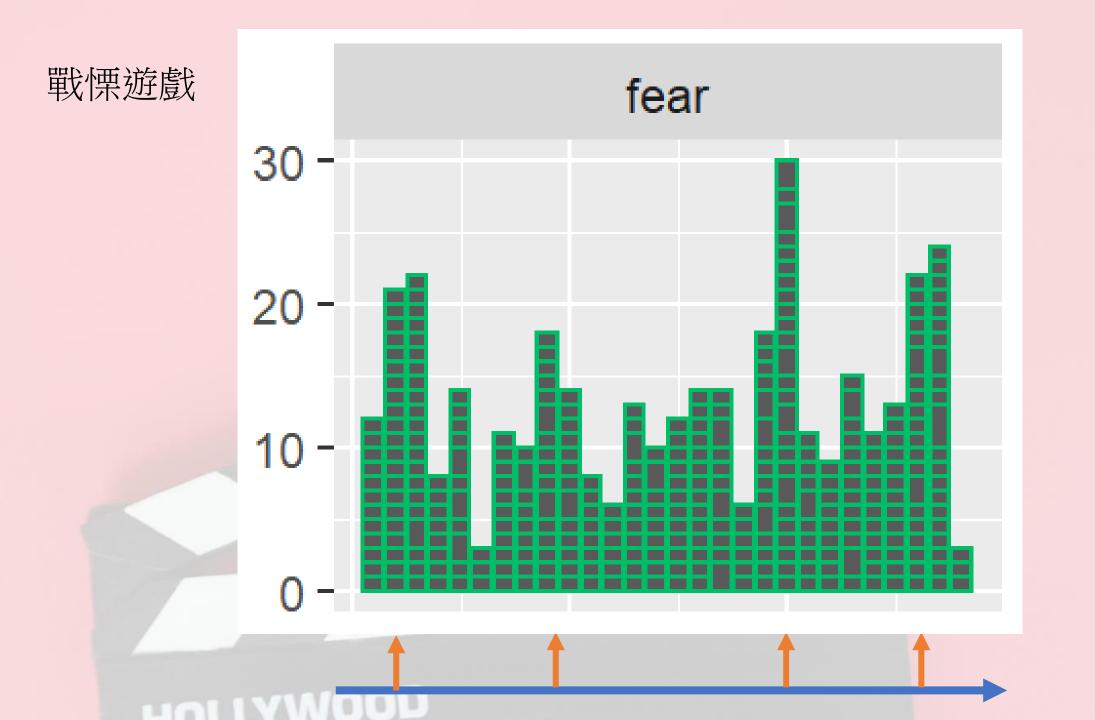
驚悚片的千層套路

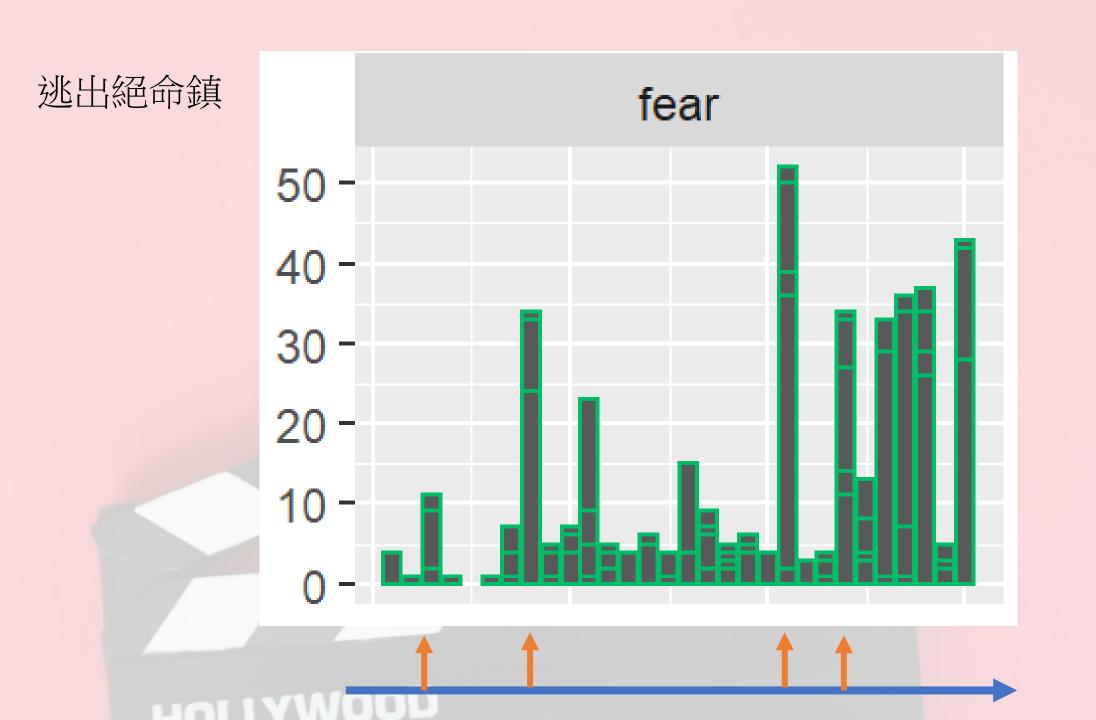
• 我們如何預防不被鬼嚇到一驚悚畫面的分布點。

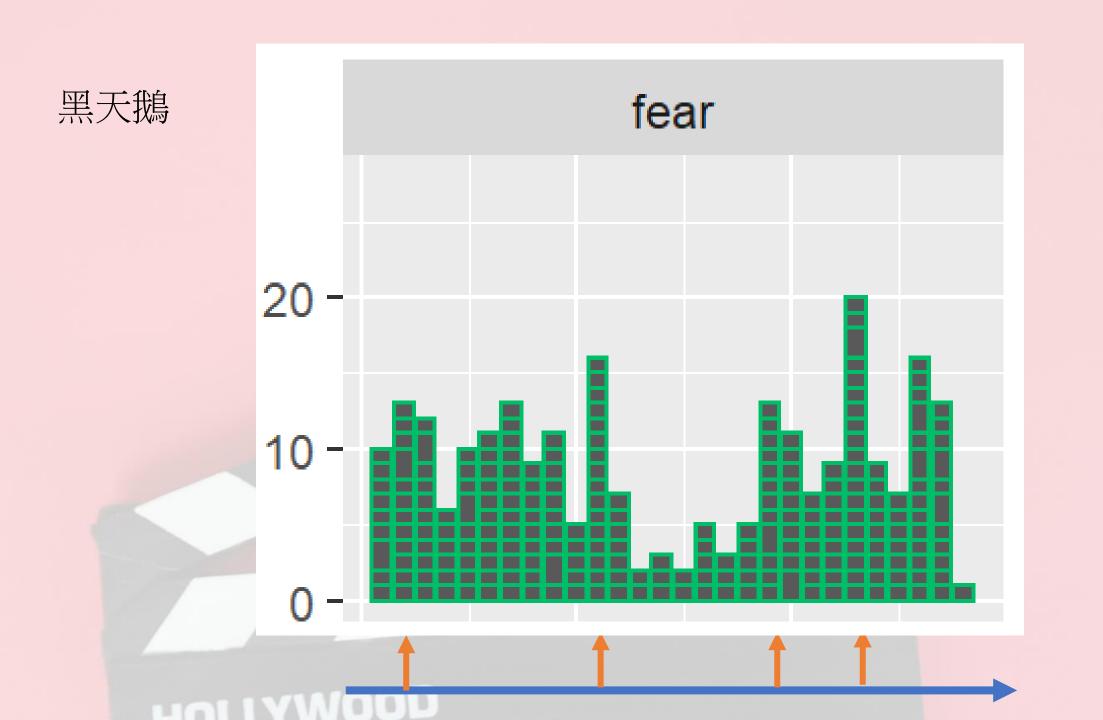
•何時前方高能? 影片進度的10%、30%、70%、90%處最有可能

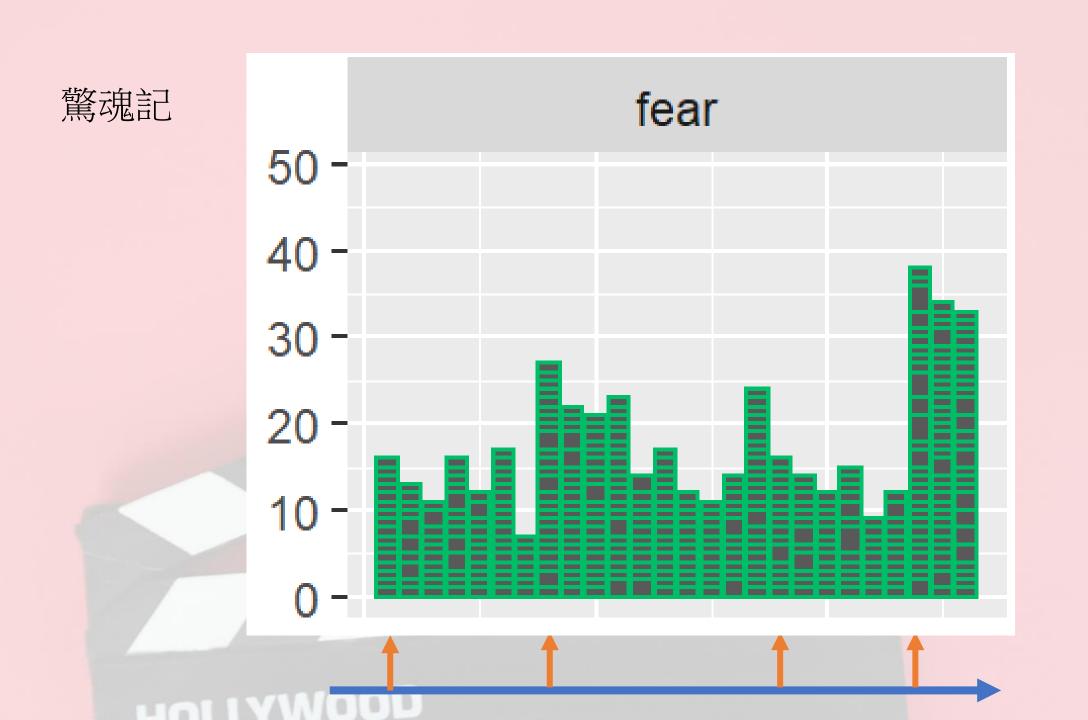
沉默的羔羊 fear 50 -

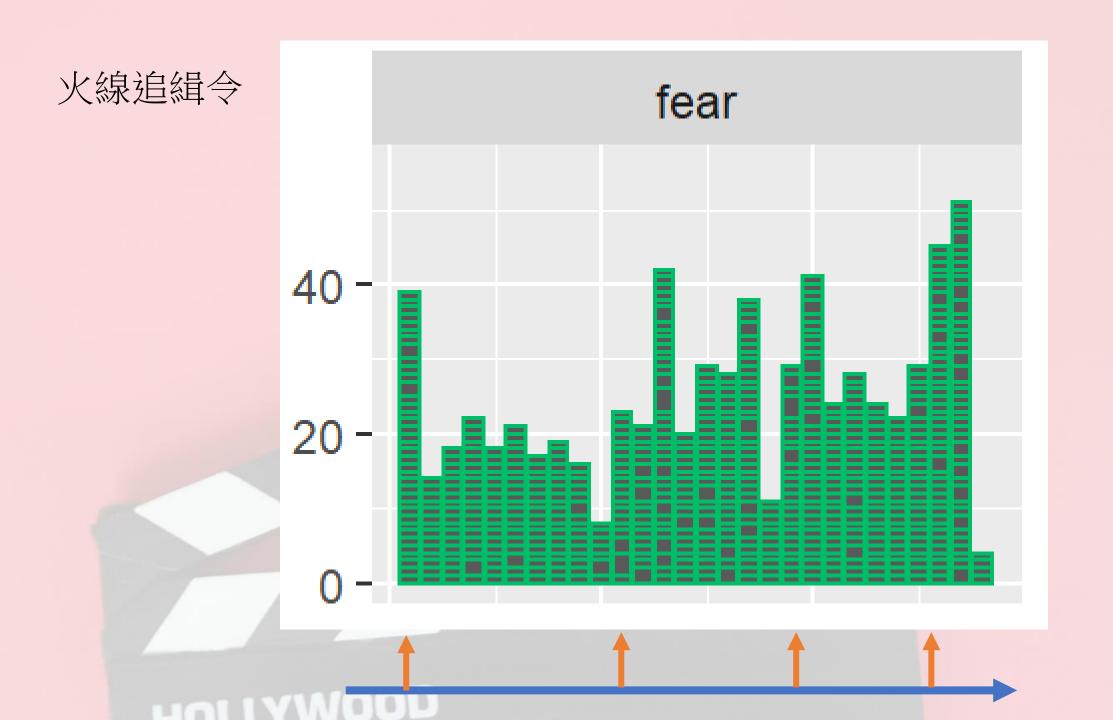












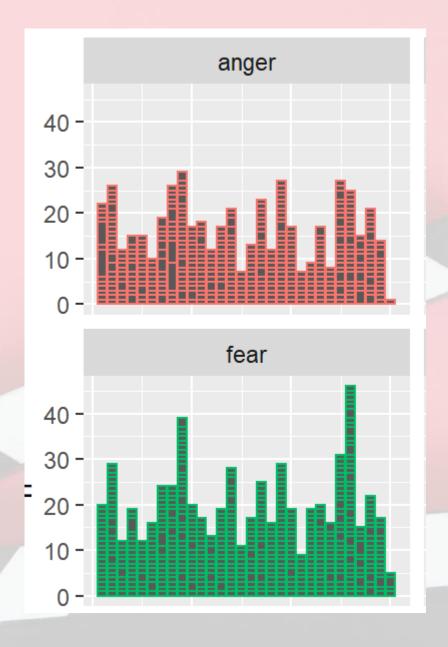
戰爭片的千層套路

• 戰爭中的煙硝味—衝突情緒被安排在電影中哪個環節?

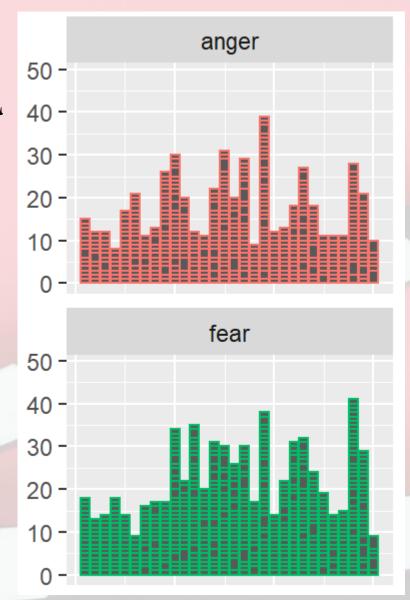
影片的開頭充斥著憤怒與恐懼,想必那是充滿著煙硝味的世界。而這些情緒大抵跟著三幕劇的形式,在圖表中能觀察出至少三個明顯的高點。



美國狙擊手

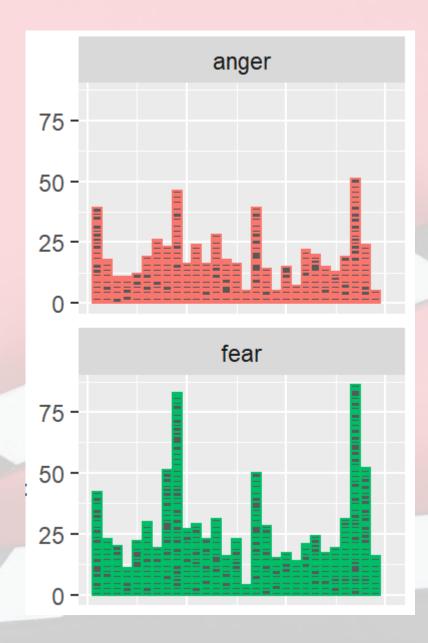


梅爾吉勃遜之英雄本色

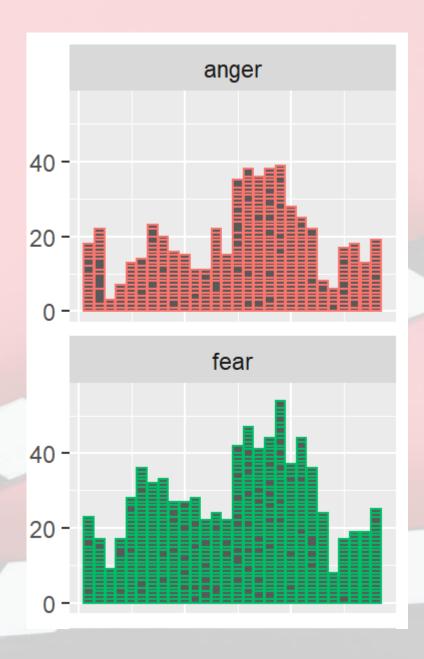




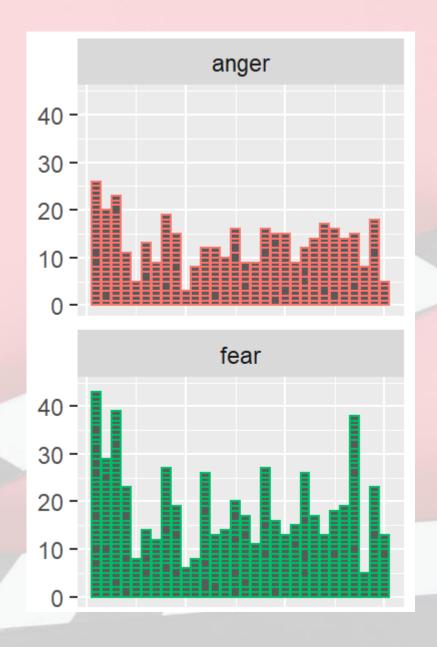
末代武士



珍珠港



搶救雷恩大兵



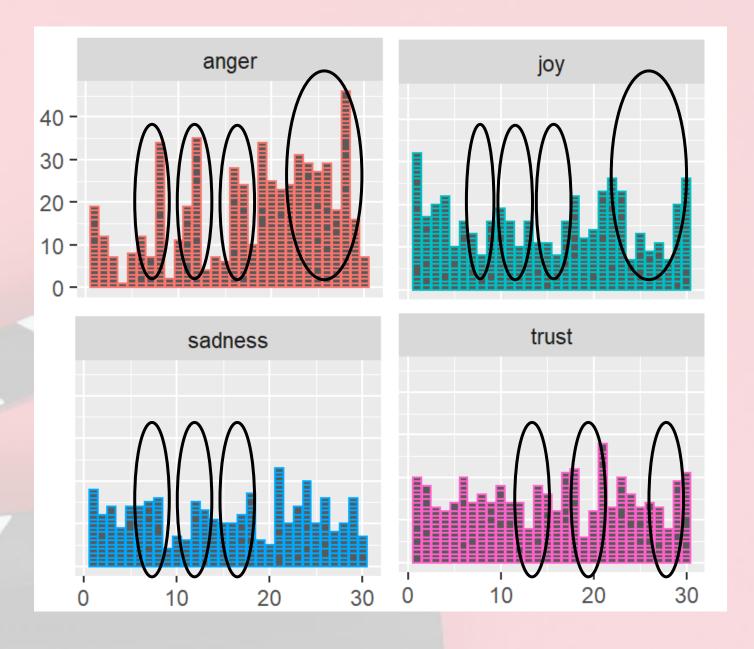
愛情片的千層套路

- 横刀奪愛才是真愛?
- 真金不怕火煉,真愛不怕唇槍舌劍?
- 與世不爭使你憑實力單身?

愛情片中設計至少三個以上的衝突,讓觀眾看得直揪心,但Happy ending讓觀眾帶著暖暖的心情回家。



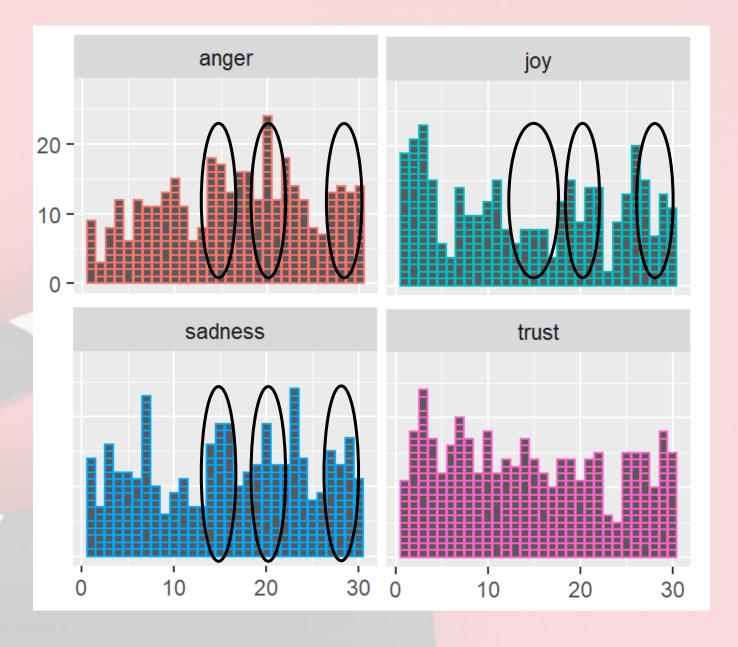
美女與野獸



衝突次數:4

HOLLYMOOD

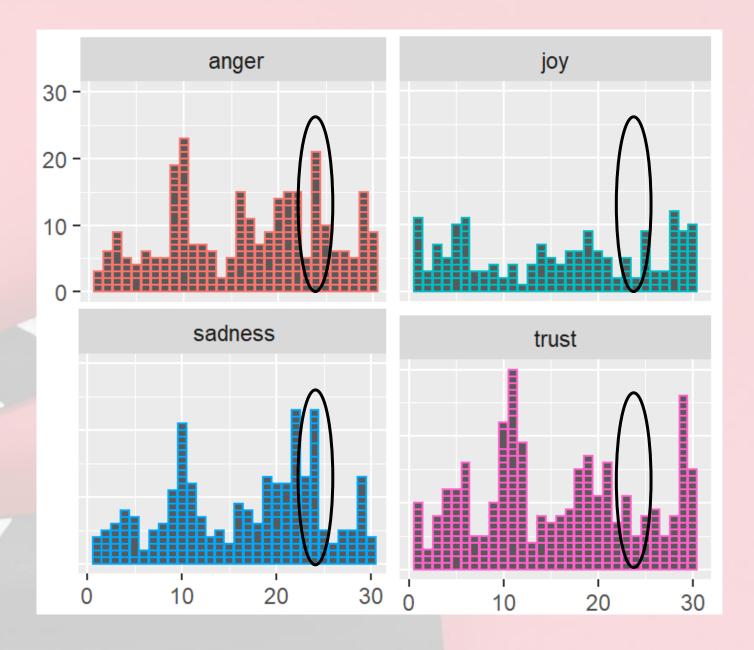
約會喔麥尬



衝突次數:3

HOLLYWOOD

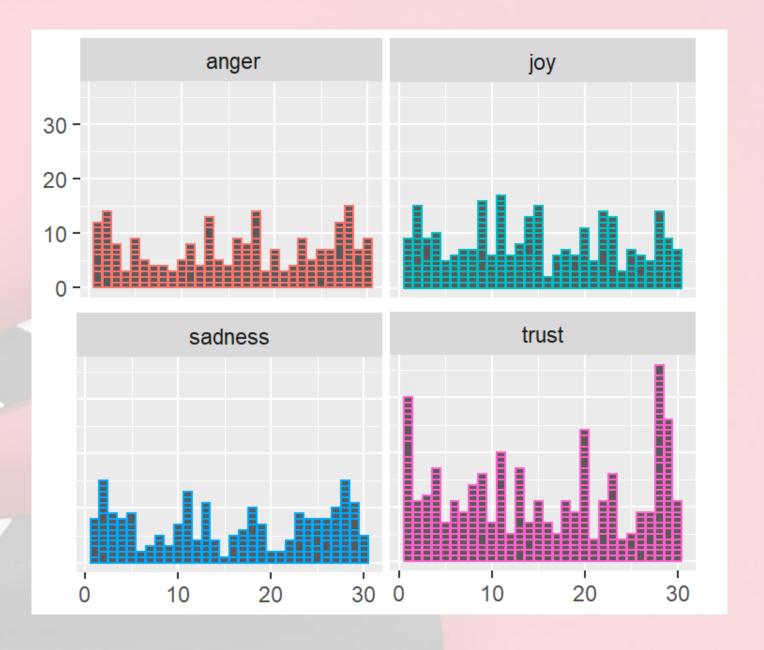
回到17歲



衝突次數 ≌ 1

HOLLYWOOD

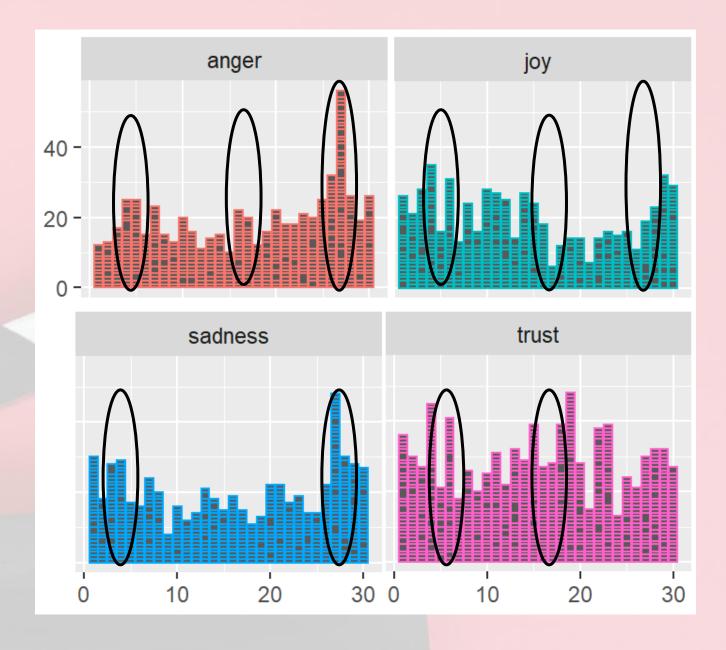
愛情昏迷中



衝突次數:難以判斷

LINI I VWOO

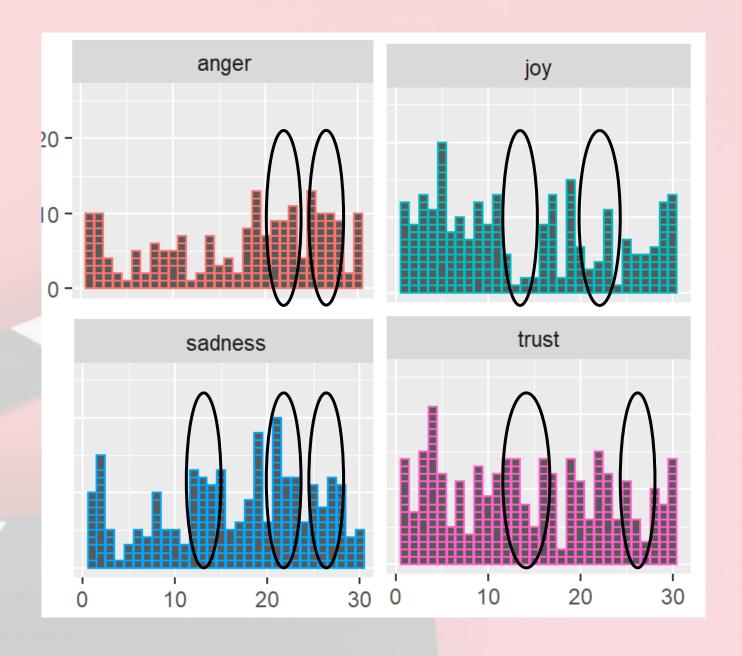
鐵達尼號



衝突次數:3

uni IV

暮光之城

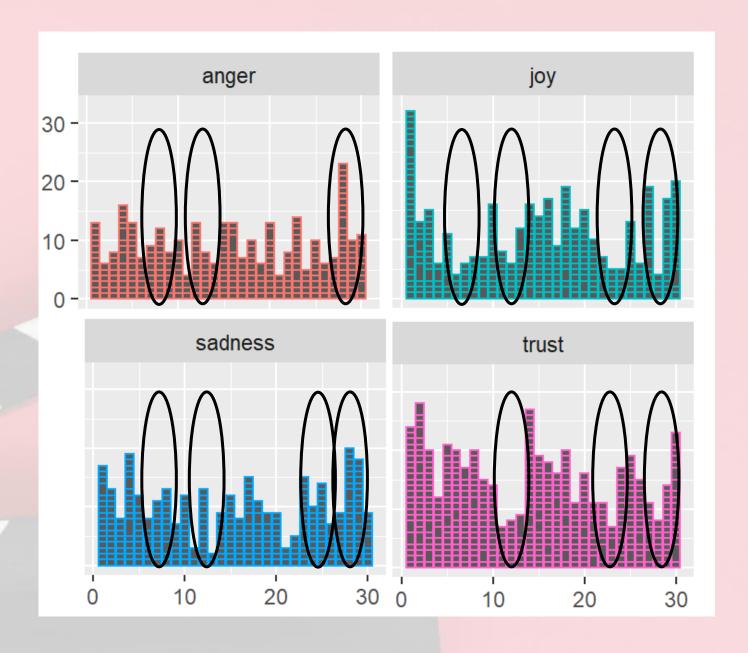


衝突次數:3

LIOLIV

暮光之城2: 新月

LIOLIVW



衝突次數:4

限制與展望

限制:

• 劇本只呈現人物對話,但情感更多從肢體動作、眼神所傳達出來,因此在本專案中,我們尚有被忽略的因素沒有兼顧到。

按照議題所針對的情感變動趨勢,我使用的是肉眼辨識,而非利用統計方法,相較之下會稍嫌主觀。

展望:

如果再對劇本的情感分析進一步探討,可發現情感分析的變動趨勢或許存在一些規律,若日後能力精進,可在此專案的基礎上,帶入機器學習等技術,以達到藉由情感分析模型來預測電影分類的目的。