

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：
男女主角是不是要親吻了
部隊這個時候是不是會被偷襲
怪物才不會那麼容易就死掉……你看牠果然還活著！

你們好奇同類型劇本是不是存在某種淺規則嗎？一起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：怎麼預防我們不被嚇——驚悚點大約分布在影片的哪部分？

戰爭片：爆炸效果很刺激？看看導演們是怎麼燒錢的。

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

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)



The web's largest
movie script resource!

Alphabetical

#	A	B	C	D	E	F	G	H
I	J	K	L	M	N	O	P	Q
R	S	T	U	V	W	X	Y	Z

Genre


Action	Adventure	Animation
Comedy	Crime	Drama
Family	Fantasy	Film-Noir
Horror	Musical	Mystery
Romance	Sci-Fi	Short
Thriller	War	Western


Sponsor


TV Transcripts


[Futurama](#)
[Seinfeld](#)
[South Park](#)
[Stargate SG-1](#)
[Lost](#)
[The 4400](#)

Last added

[Little Men](#)
(2016)


[Devil Wears Prada, The](#)
(2006)


[Joker](#)
(2019)


[A Prayer Before Dawn](#)
(2018)


[More >](#)


Newest releases


[Joker](#)
Written by
Todd Phillips
Scott Silver


[BlacKkKlansman](#)
Written by
Charlie Wachtel
David Rabinowitz
Kevin Willmott
Spike Lee


[A Quiet Place](#)
Written by
Bryan Woods
Scott Beck
John Krasinski

[Coco](#)
Written by

[Halloween](#)
Written by
Jeff Fradley
Danny McBride
David Gordon Green

[A Prayer Before Dawn](#)
Written by
Jonathan Hirschbein

[Black Panther](#)
Written by
Ryan Coogler
Joe Robert Cole

[Thor Ragnarok](#)
Written by

文本預處理

這個環節主要把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來分篇章或是使用 “INT., EXT., INSERT.” 標示場景轉移，我將利用這些標記作為斷點來分割文本。

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] | [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

[illegible]

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. 將分析結果匯出成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，因此這些情感分析數據需要整齊化。為了避免視覺化後難以判讀，本專案調整每部劇的分析段落應落在25~30的長度。因此使用：

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

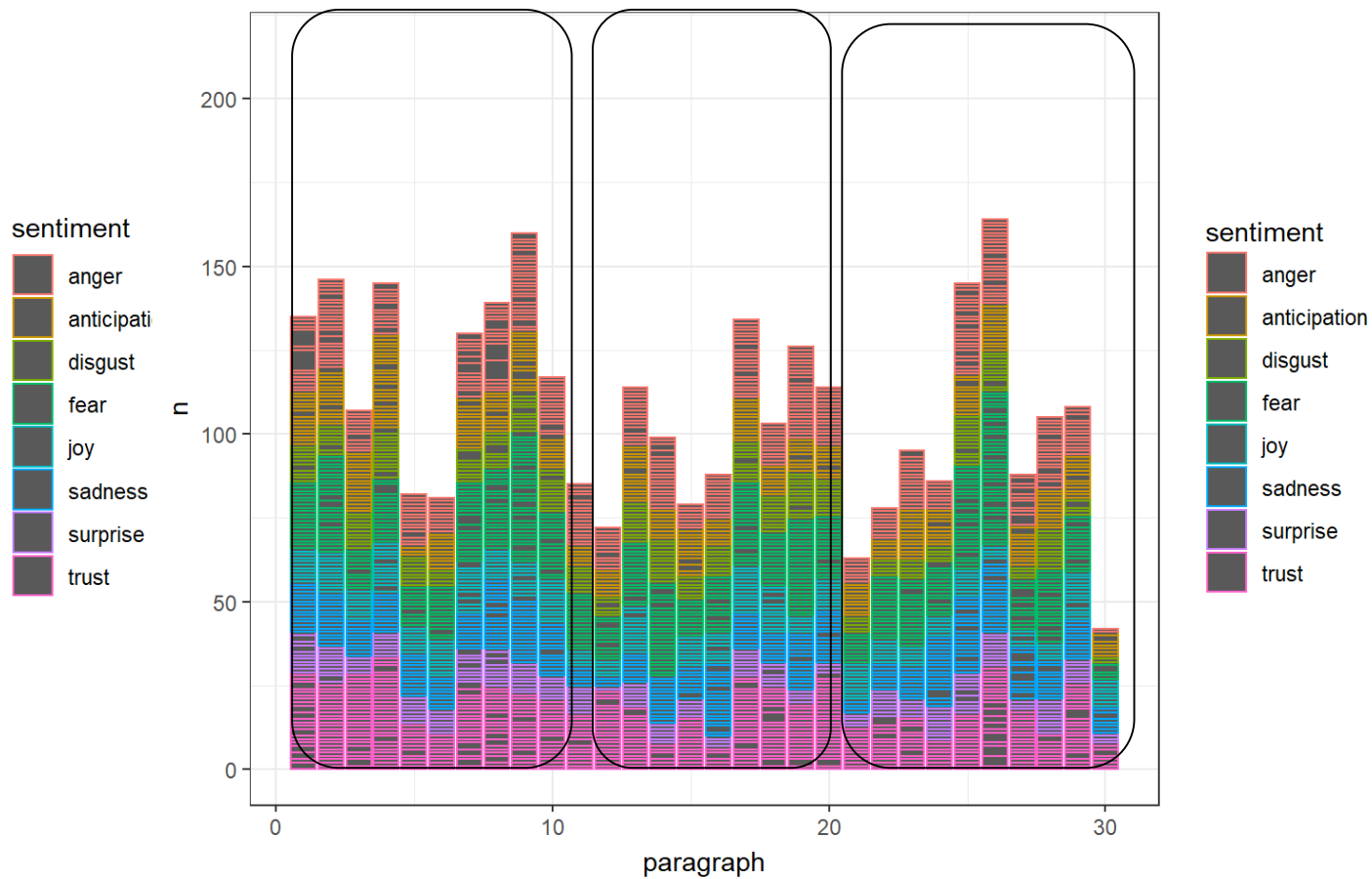
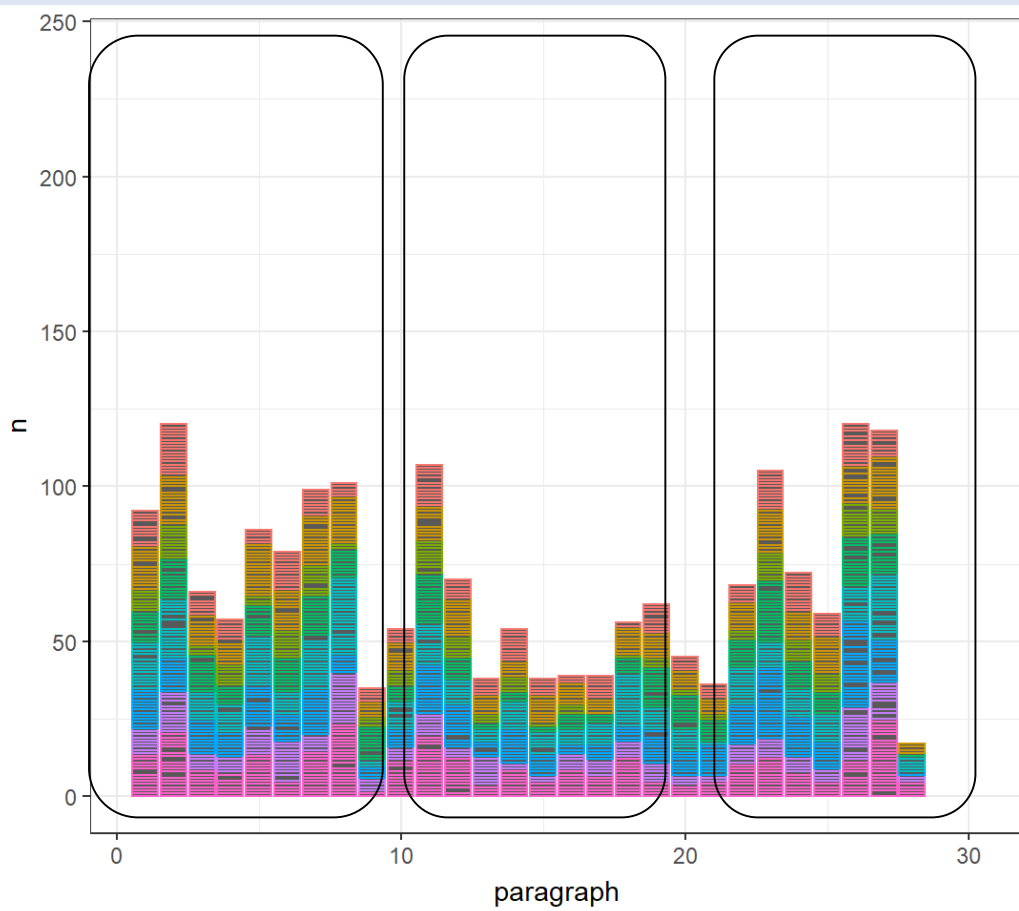
視覺化、動態化

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

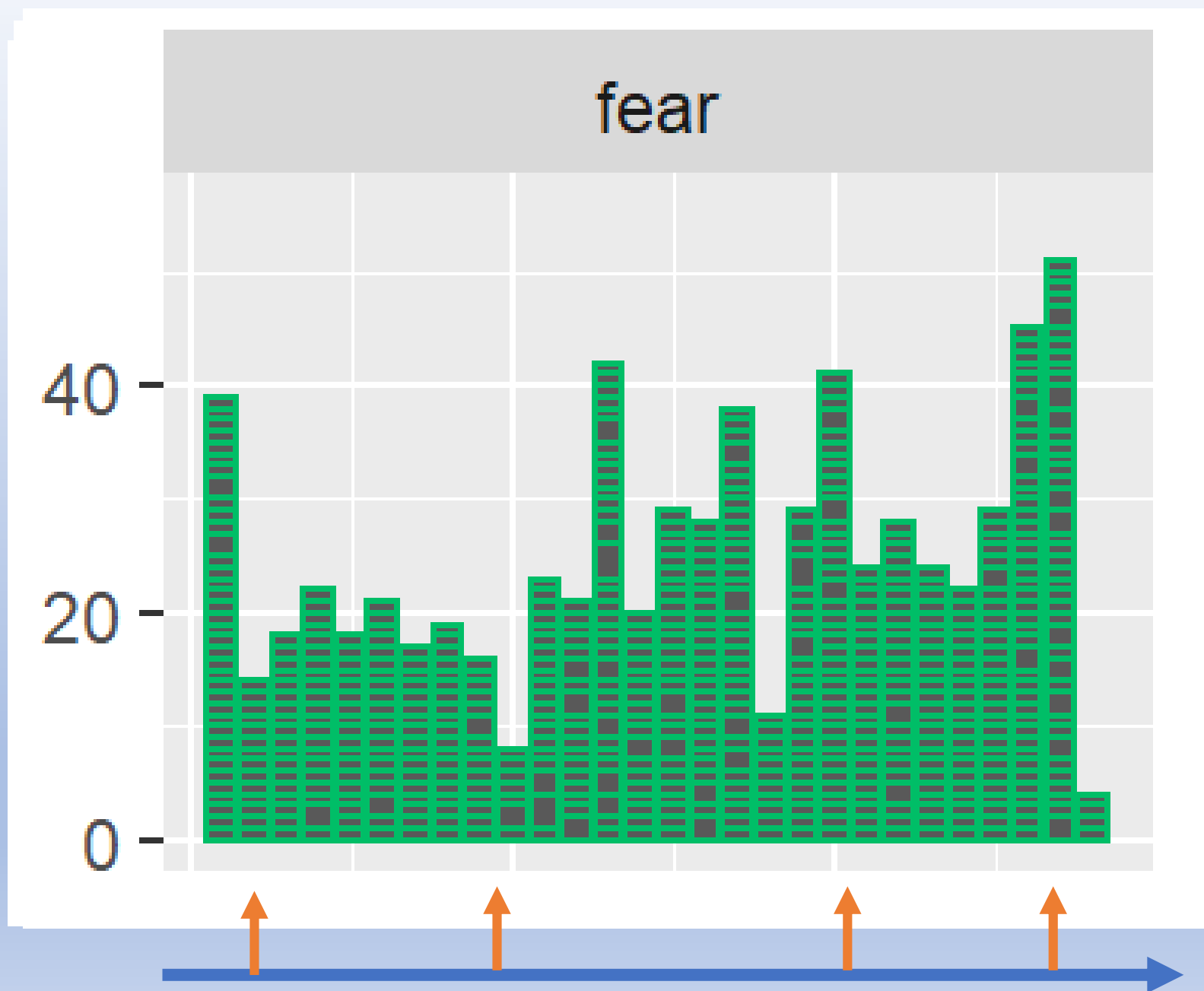


驚悚片的千層套路

- 何時前方高能？

影片進度的10%、30%、70%、90%處最有可能

雞籠通鎮

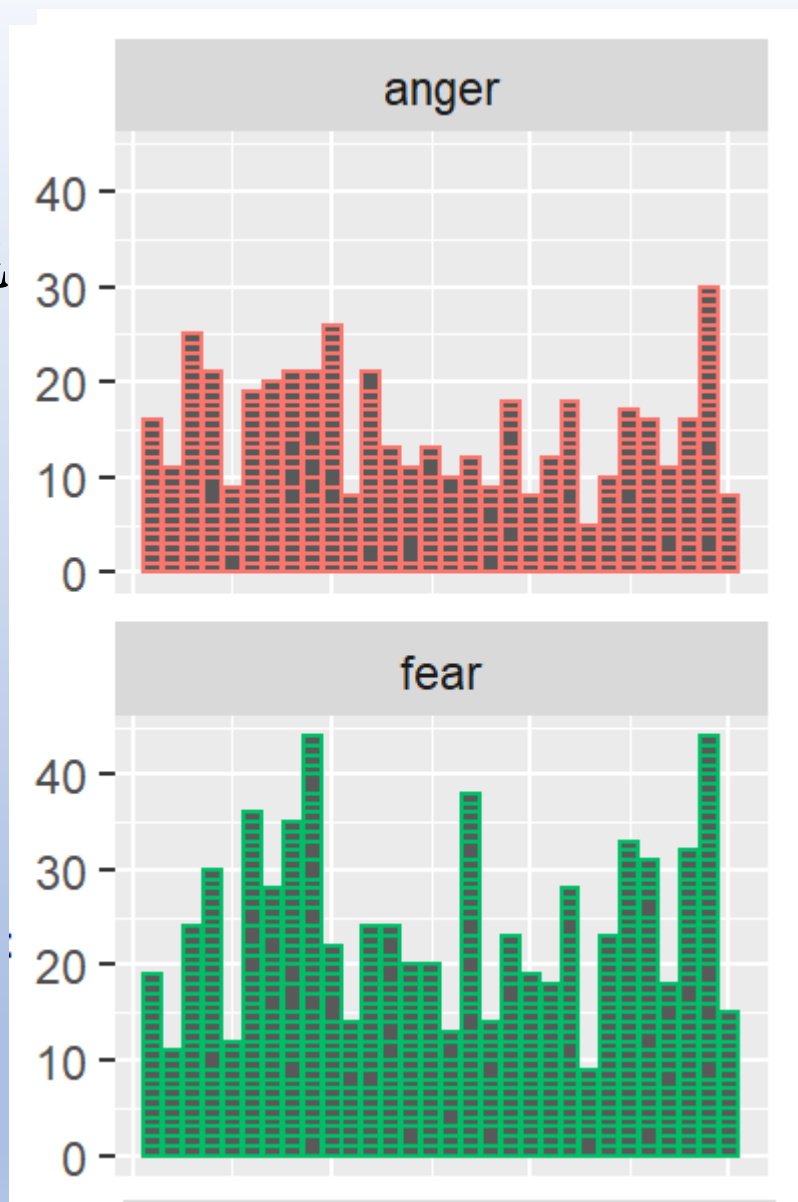


戰爭片的千層套路

- 爆炸效果很刺激？看看導演們是怎麼燒錢的。

開頭就是大把大把鈔票在燃燒的味道

梅爾歐羅亞英雄本色



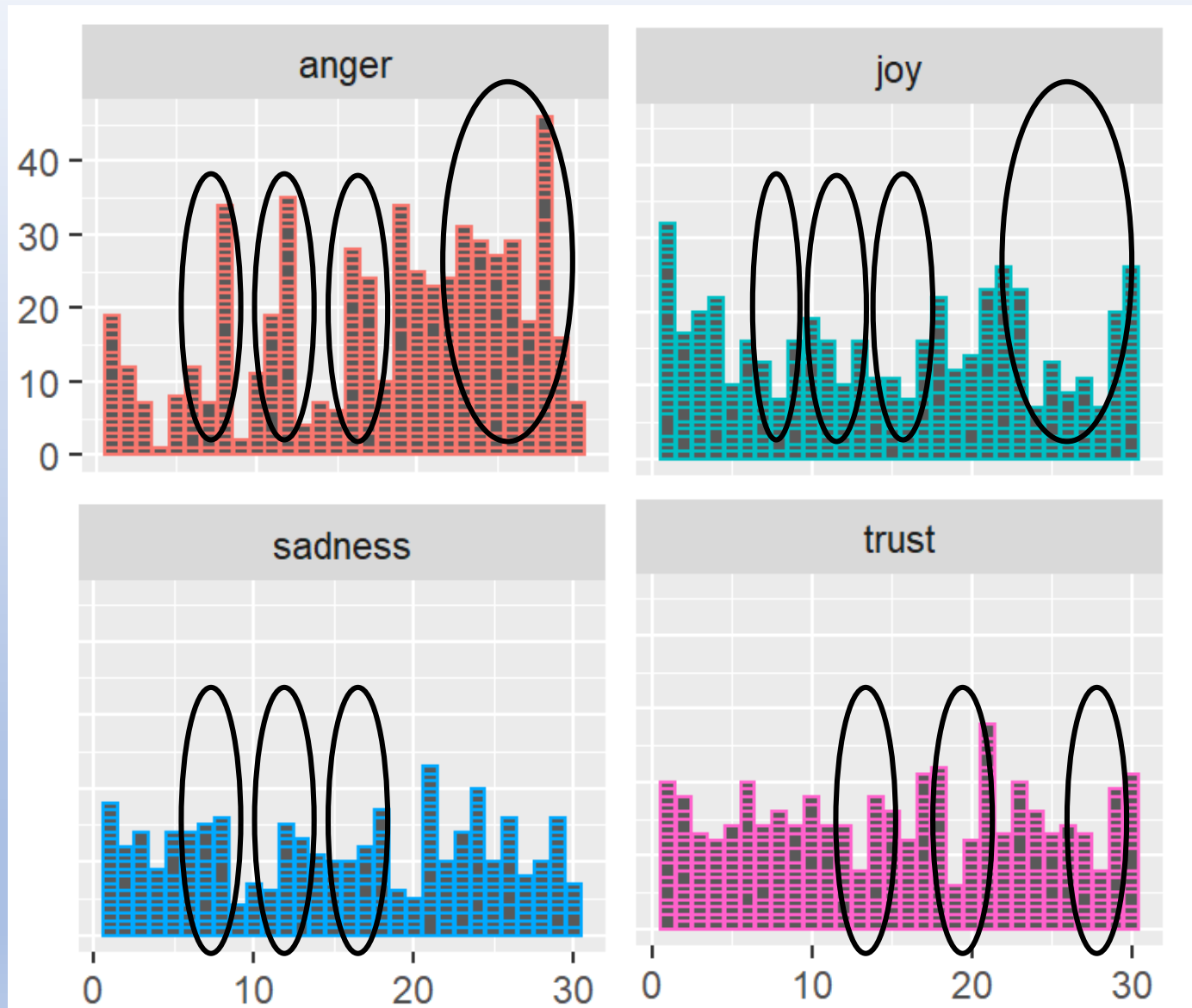
愛情片的千層套路

- 橫刀奪愛才是真愛？
- 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？

愛情片設計多少衝突，讓人心裡小鹿亂撞？

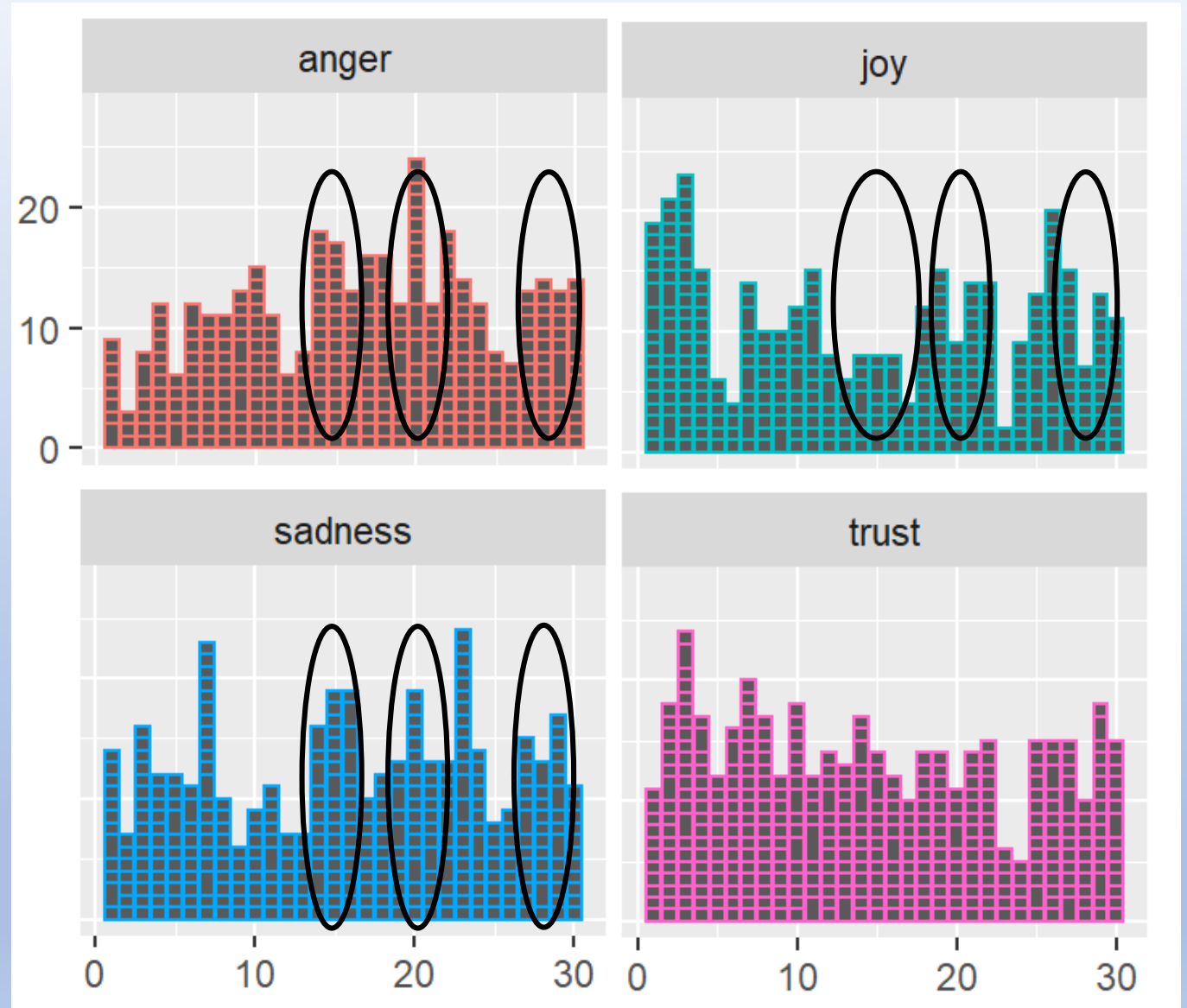
美女與野獸

衝突次數：4



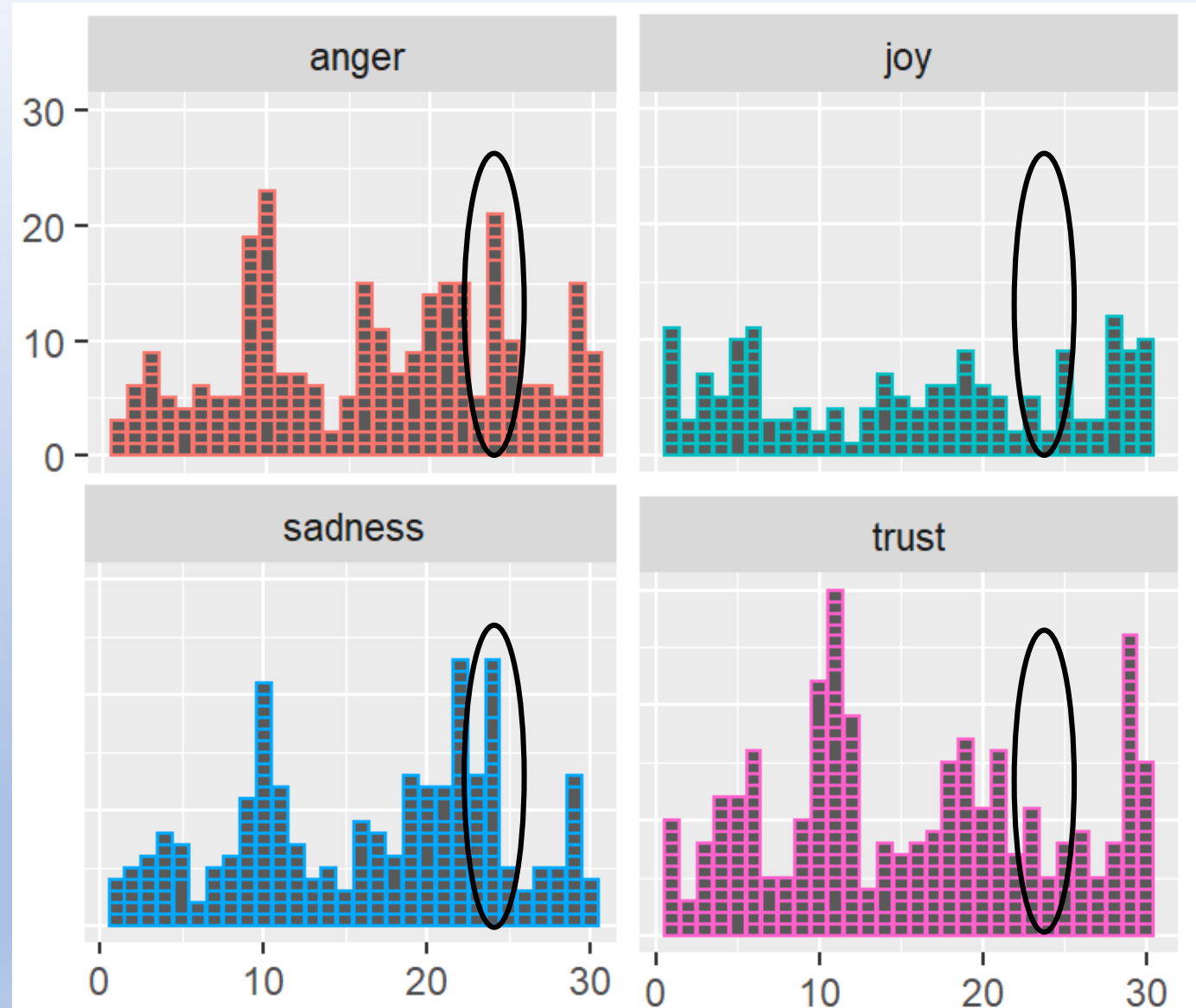
約會喔麥尬

衝突次數：3



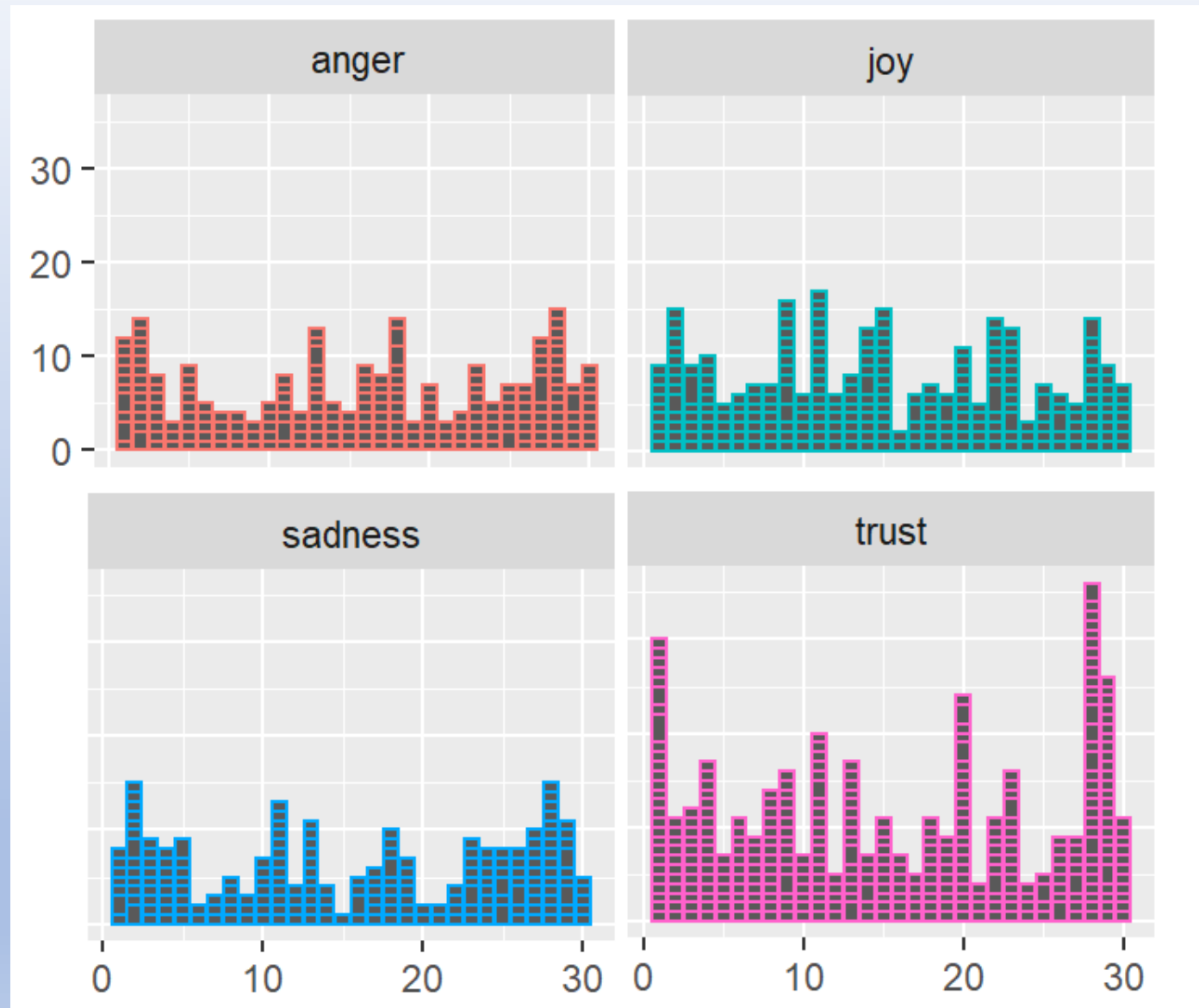
回到17歲

衝突次數 $\cong 1$



愛情昏迷中

衝突次數：難以判斷



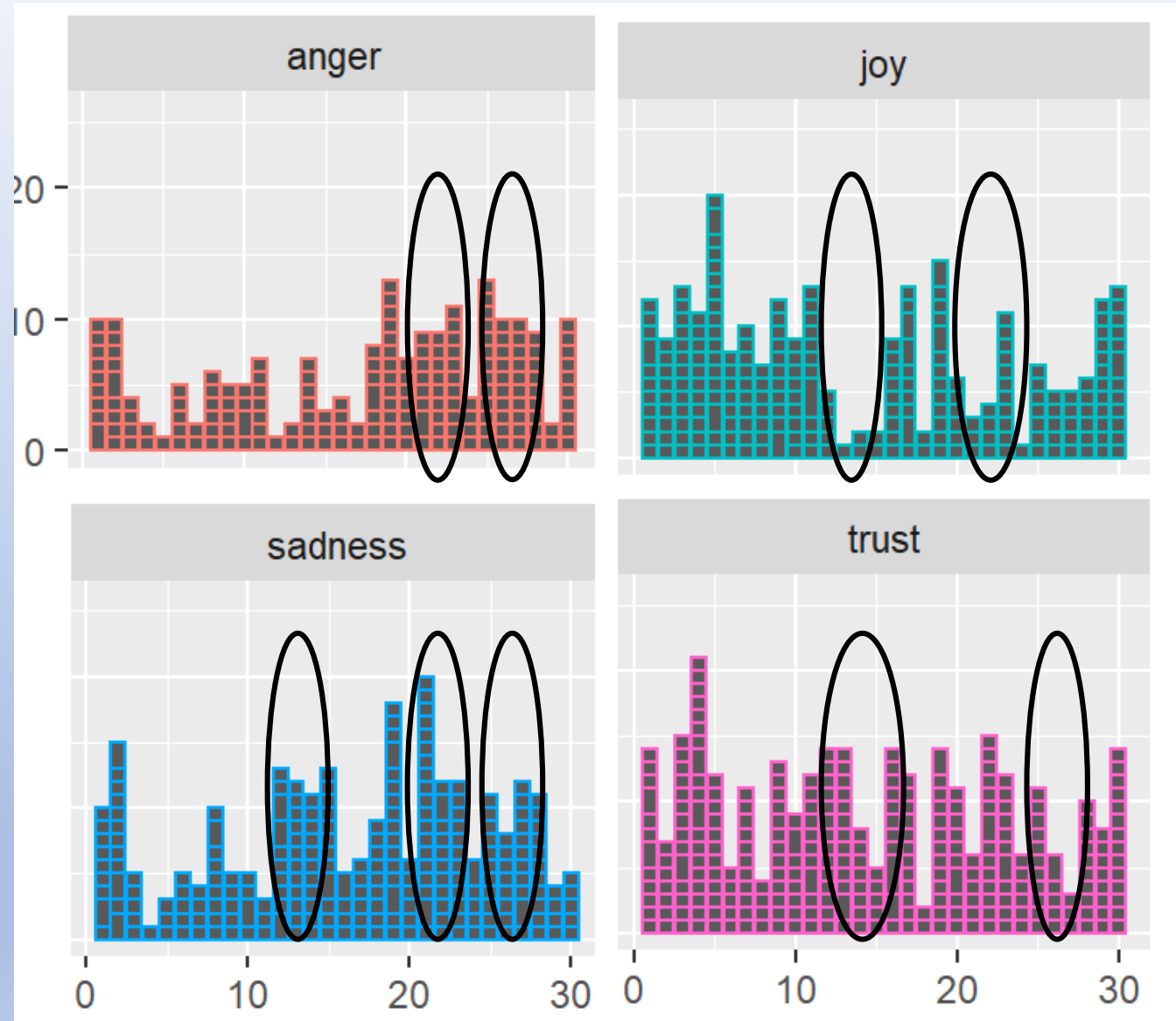
鐵達尼號

衝突次數：3



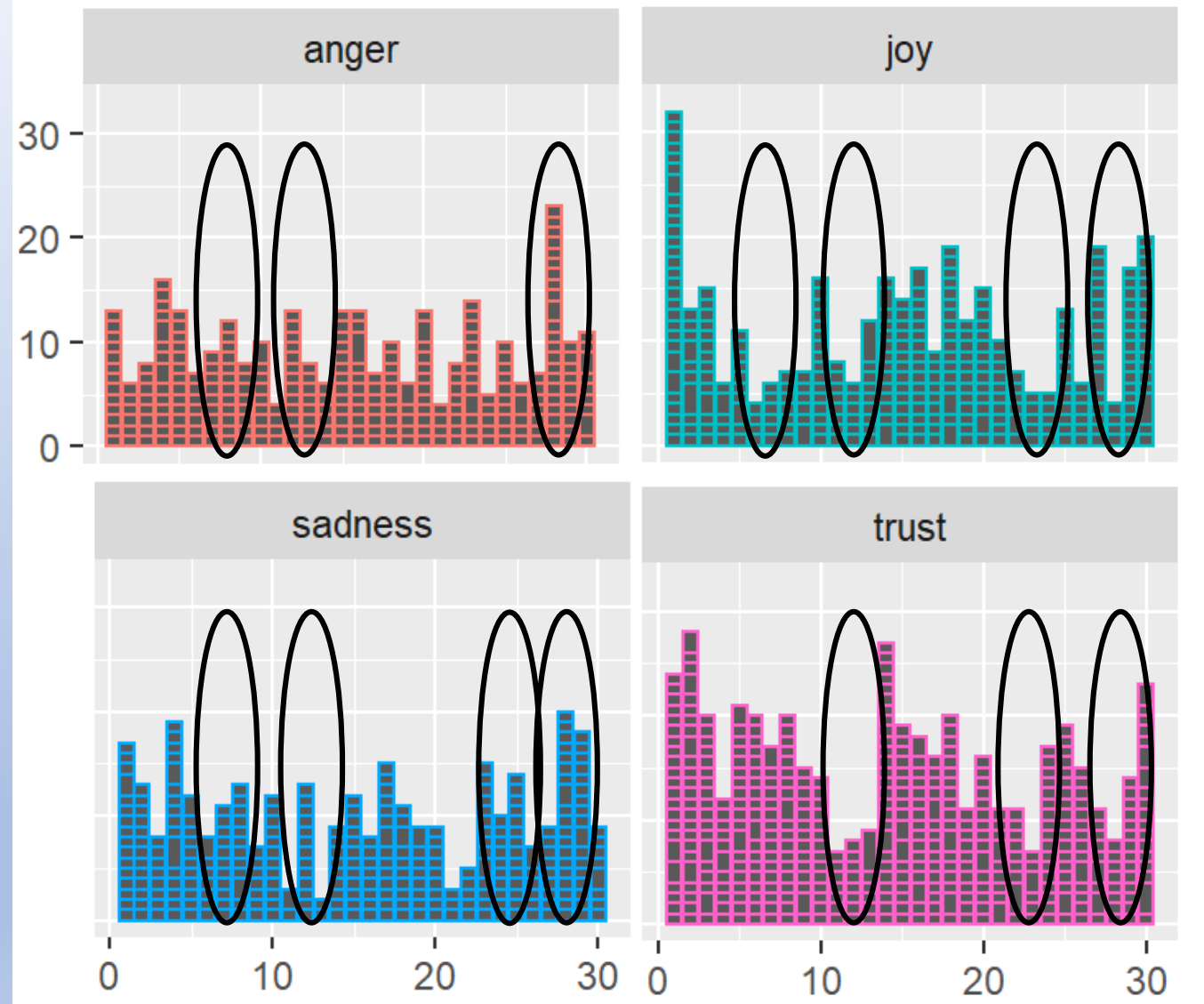
暮光之城

衝突次數：3



暮光之城2： 新月

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。