

我沒做錯，警察別碰我

分析加州交通事件警察攔查以及警察暴力致死事件數據資料

一、目錄

- 一、目錄
- 二、簡介
- 三、資料取得
 - (一) 資料前處理
 - (二) 資料觀察
- 四、資料探勘與分析方法
 - (一) 搜查逮捕年曆
 - (二) 各種族被警察攔查的處遇
 - (三) 加州警察執法地圖
- 五、結果
 - (一) 搜查逮捕年曆
 - (二) 各種族被警察攔查的處遇
 - (三) 加州警察執法地圖
- 六、回饋回覆與小研究
- 七、分工表

二、簡介

在台灣長大生活多年的我們常稱呼警察為人民的保姆，社會上對警察作為社區好鄰居的形象，觀感普遍甚佳。然而，在橫跨一個太平洋的美國，警察恐怕不是什麼人民想見到的角色。我們平時在社會新聞裡往往觀察得到美國警察執法過當或是偏頗的消息，例如在追逐其實並不具有攻擊能力的歹徒時，還是選擇以開槍的手法處理。在法律學中有所謂的「比例原則」，即是對於立法者訂立的法律、政府機關訂定的規則進行檢驗，必須符合適當性、必要性、衡量性，適當性要求手段有助於目的的達成，必要性則著重為達成目的時必須採取對人民侵害最小的手段。美國警察對於許多案件的處理，恐怕無法通過必要性的檢驗。

為了正視這個問題，我們決定找尋相關資料，看看警察在什麼時間、被攔查對象是什麼種族、在哪個地區會有不同的執法處理，並關心執法過當致死的問題。

三、資料取得

- (一) 資料來源

本專案使用的資料包括：Stanford Open Policing Project - California: Data on Traffic and Pedestrian stops by Police in California，Campaign Zero: Mapping Police Violence，Californian Geographical Data 以及 NBC News California Presidential Election Results 2020。第一組資料及著重加州警察對交通事件攔查個案的處理數據，第二組強調加州所有警察暴力執法過當致死的數據，第三組則是加州地理空間資料，第四組是 2020 年美國總統大選各個郡川普的得票率。

至於資料的選擇方面，會以加州為對象是因為該州資料較完整，且數據量較多。加州的原始資料就有約 1453 萬筆，本文以為，以較豐富的數據資料進行分析可以獲得較全面性的結果。

（二）資料前處理

除了一般的清理以外，我們有做在交通事件中，被警察攔查後遭到搜查的比率，以及被攔查後遭到逮捕的比率，以下將簡稱為搜查率與逮捕率。此外，我們也有計算每個郡警察執法過當致死的比率，以下將簡稱為警暴率。

上述三種比率，我們定義如下：

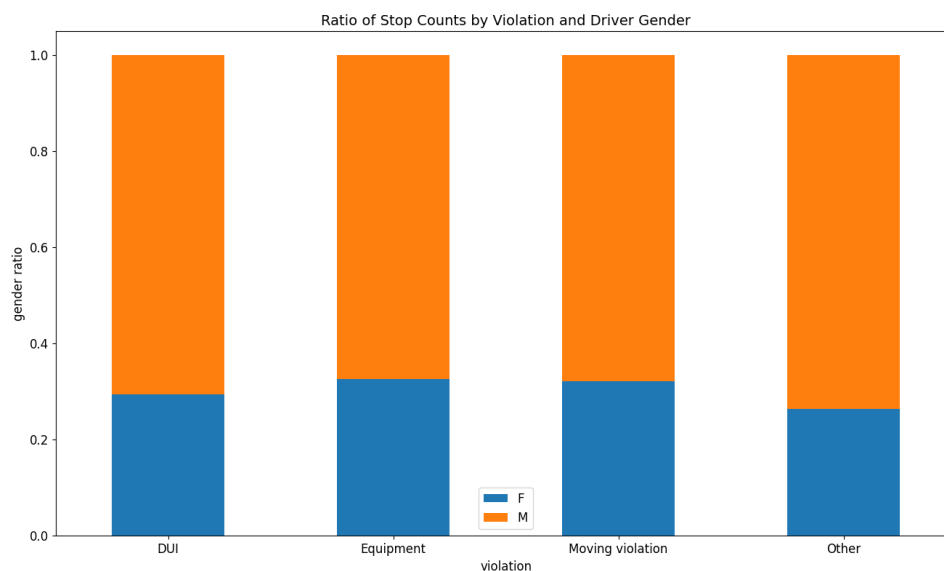
搜查率：被警察攔查後執行搜查人數/被警察攔查人數

逮捕率：被警察攔查後執行逮捕人數/被警察攔查人數

警暴率：因警察暴力執法過當致死人數/該郡 2013-2016 年平均人口數

（三）資料觀察

1. 性別：依照各種不同常見的交通違規事件分類，看觸犯規定的男女比（例如酒駕可能男性比例較高）？



圖一、因為各種違規被攔停的男女比

由上長條圖表可知，無論是在酒駕、設備（有無違法改裝、違規未開頭燈等）、駕駛違規（未按規定讓行、違規超車等），皆是男性遠高於女性，組間差距不顯著。

2. 年齡：各年齡段的搜查率、逮捕率、警暴致死率分別是多少，以及何年齡層居冠？

表一、各年齡段搜查逮捕率

| driver_age | search_conducted | is_arrested | stop_count | search_rate | arrest_rate |
|------------|------------------|-------------|------------|-------------|-------------|
| 0-14 | 894 | 2653 | 24743 | 3.61% | 10.72% |
| 15-25 | 117761 | 88295 | 2423934 | 4.86% | 3.64% |
| 25-32 | 153598 | 116989 | 3437799 | 4.47% | 3.40% |
| 33-39 | 81727 | 61353 | 2721297 | 3.00% | 2.25% |
| 40-48 | 62276 | 47594 | 2698382 | 2.31% | 1.76% |
| 49+ | 62471 | 49884 | 3230183 | 1.93% | 1.54% |

由上表可知，0-14 歲居冠，因為 0-14 歲兒童與少年本不應該駕駛，故一旦被警察攔查發現，自然有較高機率被搜查和逮捕。

3. 人數：加州警察在 2013-2016 年間因為各種原因在執法中殺了多少人？又因為交通事件違規逮捕多少人？

2013-2016 年間，加州警察攔停後搜查了 448738 人，逮捕了 344562 人，且因為暴力執法過當致死 12339 人。由此可知，加州警察恐怕具有執法過當的問題，為了讓此說法更有公信力，我們也有上網另外查詢資料，發現加州每 100 萬人裡面就會有 3.36 人因警察暴力執法過當而死，但是麻州卻只有 0.71 人。

4. 地區：哪個郡的搜查率最高？哪個郡的警察暴力致死事件最多？

表二、搜查率（郡）

| county | search_rate | arrest_rate | killed_rate* 100000 |
|--------------------|-------------|-------------|------------------------|
| Lake County | 0.064686 | 0.062399 | 3.124951 |
| Tulare County | 0.06056 | 0.039003 | 2.192987 |
| Humboldt County | 0.055455 | 0.043206 | 3.701729 |
| Fresno County | 0.054865 | 0.031257 | 2.591748 |
| San Joaquin County | 0.054625 | 0.040876 | 0 |

表三、逮捕率（郡）

| county | search_rate | arrest_rate | killed_rate* 100000 |
|--------------------|-------------|-------------|------------------------|
| Lake County | 0.064686 | 0.062399 | 3.124951 |
| Humboldt County | 0.055455 | 0.043206 | 3.701729 |
| San Joaquin County | 0.054625 | 0.040876 | 0 |
| Stanislaus County | 0.05299 | 0.040687 | 1.50664 |
| Tulare County | 0.06056 | 0.039003 | 2.192987 |

表四、警暴率（郡）

| county | search_rate | arrest_rate | killed_rate* 100000 |
|---------------------------|-------------|-------------|------------------------|
| Trinity County | 0.043279 | 0.032447 | 7.622532 |
| Plumas County | 0.028893 | 0.021067 | 5.361355 |
| Sutter County | 0.035511 | 0.026654 | 4.207116 |
| Kern County | 0.029583 | 0.019723 | 4.013688 |
| Shasta County | 0.029697 | 0.022282 | 3.920866 |

湖郡在搜查率、逮捕率皆居冠，三一郡則在警暴率居冠。

四、資料探勘與分析方法

（一）搜查與逮捕年曆

先使用 Pandas 中的 groupby 將資料按照 year、month、day 的組合分組，用以計算各個組合日期下的搜查率與逮捕率。接著，利用 Pandas 中的 groupby 及 agg 把每一年中相對應日期的數據取平均，取得年曆中每一格的數值。例如：1 月 1 日的搜查率 = (2013 年 1 月 1 日搜查率 + 2014 年 1 月 1 日搜查率 + 2015 年 1 月 1 日搜查率 + 2016 年 1 月 1 日搜查率) / 4。再來，使用 Pandas 中的 pivot 進行樞紐分析，將數值按照月份及日期排列備用。最後使用 Seaborn 套件製成 heatmap 熱力圖，期望找出加州警察在交通執法是否有時間上的趨勢或規律。

（二）各種族被警察攔查的處遇

先使用 Pandas 中的 pivot 將資料按照違規事項及種族分組，進行樞紐分析。接著再將資料製成 bar plot 長條圖，期望看到加州警察在交通執法是否會因為被攔查人的種族區別而有不同趨勢。

（三）加州警察執法地圖與相關性分析

先使用 Pandas 中的 groupby 及 agg 將資料按照郡別分組並取得各郡的搜查率與逮捕率。接著，利用 Seaborn 中的 regplot 將各郡的資料繪製在散布圖上，並透過 Pandas 中的 data.corr 得到逮捕率與搜查率的相關係數。警暴率分別搜查率、逮捕率的相關性分析作法同上。關於警察執法地圖的製作方面，我們原先打算運用 Q-GIS 軟體（高中時期處理地理空間資料最常用的第三方軟體），但為了更大程度地結合本課程，我們決定跳出舒適圈直接運用 python 套件製作。取得空間資料後，先使

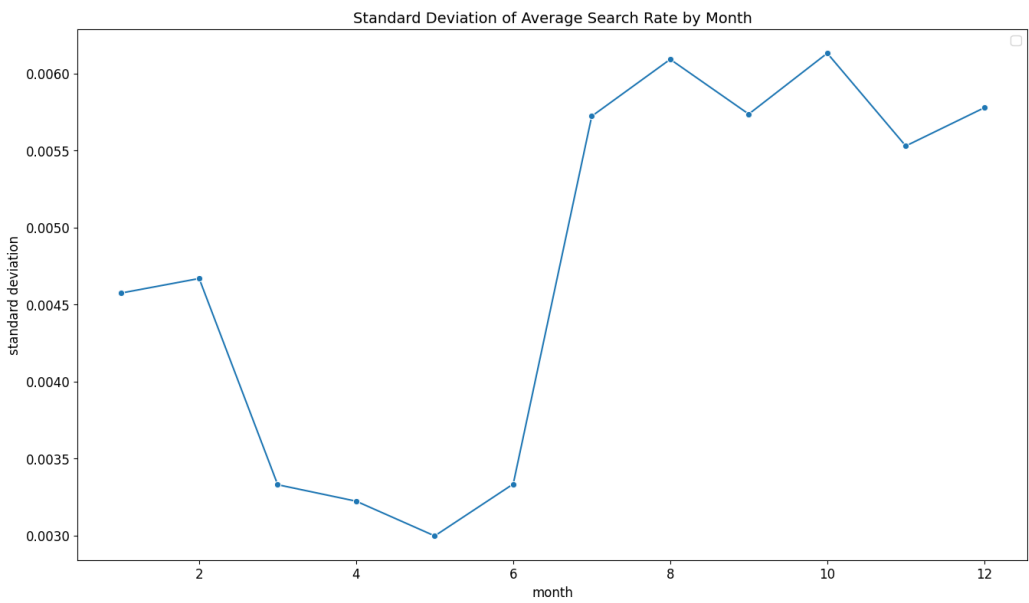
用 explode 拆分 MultiPolygon，接著再將拆分過後的空間資料與原先的統計數據（搜查率、逮捕率、警暴率）merge 起來，並運用 Geoplot 繪製 Choropleth（面量圖），以看出加州警察在不同地區的執法趨勢。

五、結果

(一) 搜查與逮捕年曆

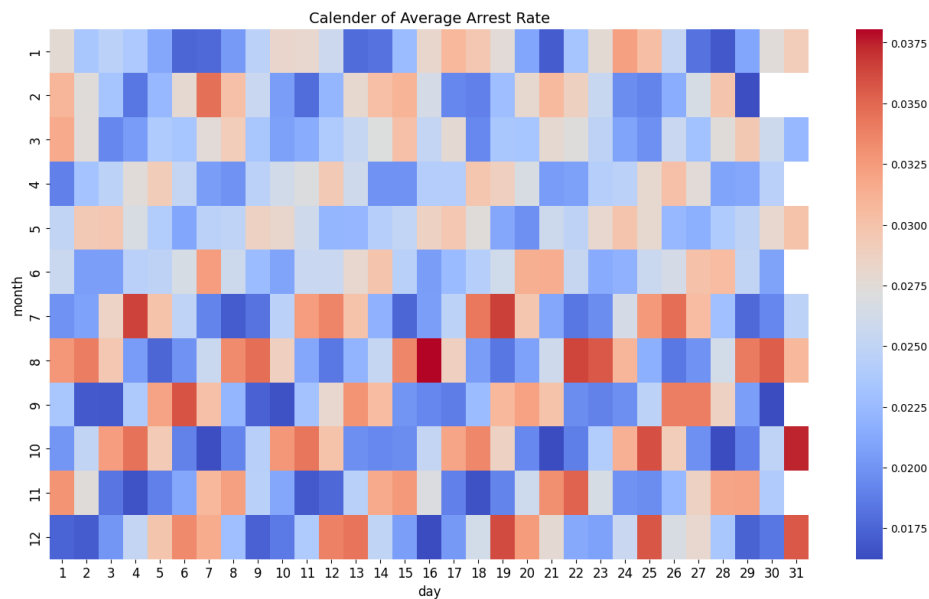


圖二、搜查年曆

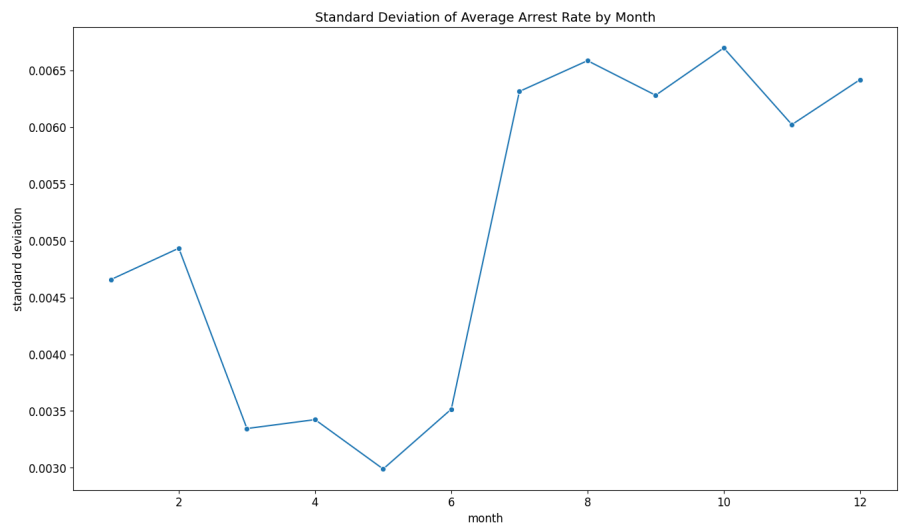


圖三、各月份搜查率標準差

綜合搜查年曆與各月搜查率標準差，我們觀察到，1-6 月搜查率相對中庸，尤以 5 月最為平均；7-12 月搜查率大起大落，尤以 8、10 月變異最大；其中又以 8 月平均搜查率最高；關於紅藍相間的趨勢，則有三天紅、四天藍的規律可循，我們推測可能是和警察執勤時數相關，以 LAPD 洛杉磯警察局為例，警局規定警察每週工作三天 12 小時，四天 10 小時，而這恰好與搜查年曆的趨勢相符。



圖四、逮捕年曆



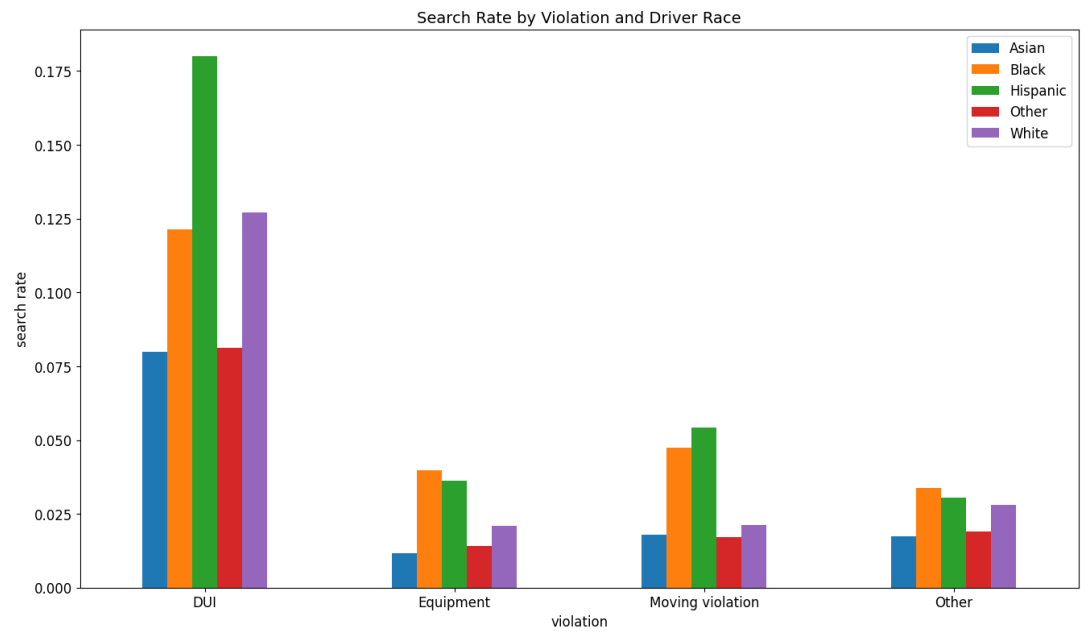
圖五、各月份逮捕率標準差

綜合逮捕年曆與各月逮捕率標準差，我們觀察到，1-6 月搜查率相對中庸，尤以 5 月最為平均；7-12 月搜查率大起大落，尤以 8、10 月變異最大；其中又以 8 月平均逮捕率最高；關於紅藍相間的趨勢，則有三天紅、四天藍的規律可循，我們推測可

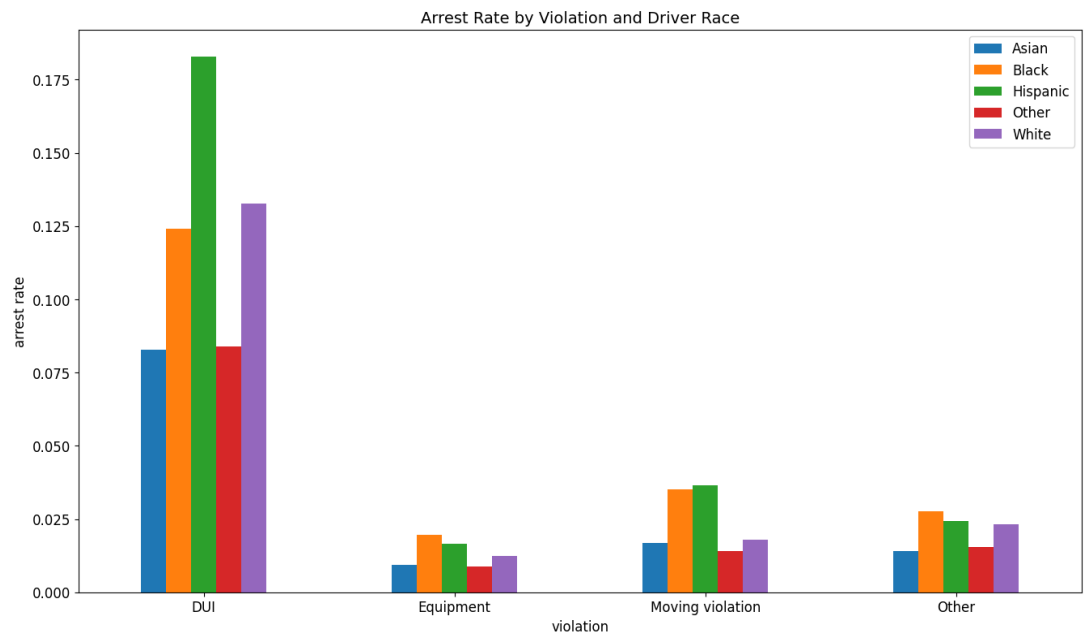
能是和警察執勤時數相關，以 LAPD 洛杉磯警察局為例，警局規定警察每週工作三天 12 小時，四天 10 小時，而這恰好與逮捕年曆的趨勢相符。

我們原先假設加州交通警察會因為業績壓力在月底或年底比平常更積極地進行搜查與逮捕。經過數據統計、視覺化等探勘後，我們發現無類似趨勢，反而是上下半年數據的變異程度差異以及紅藍相間的規律更加明顯。

(二) 各種族被警察攔查的處遇

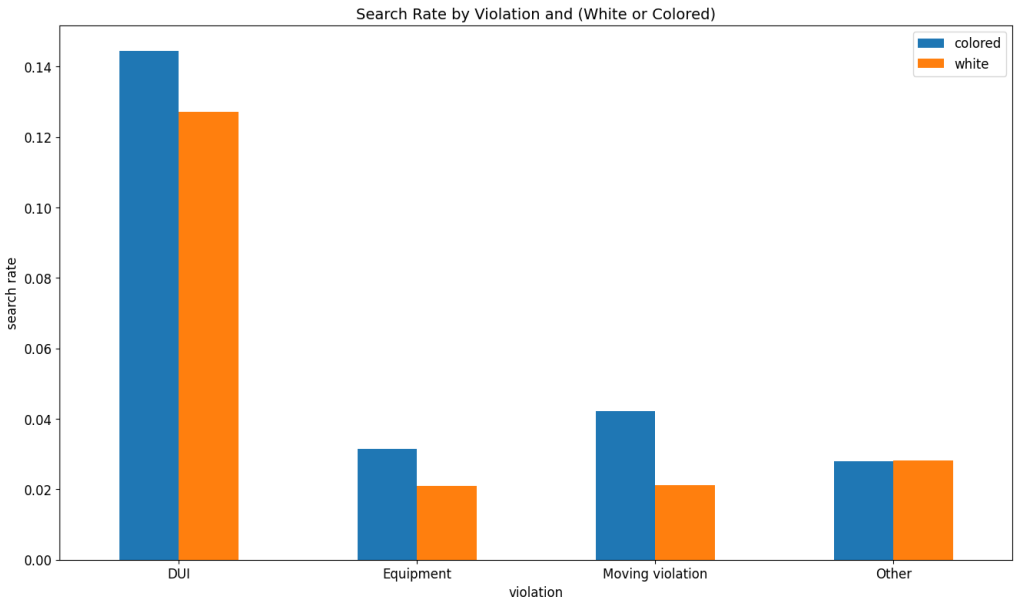


圖六、依罪名分類各種族的搜查率

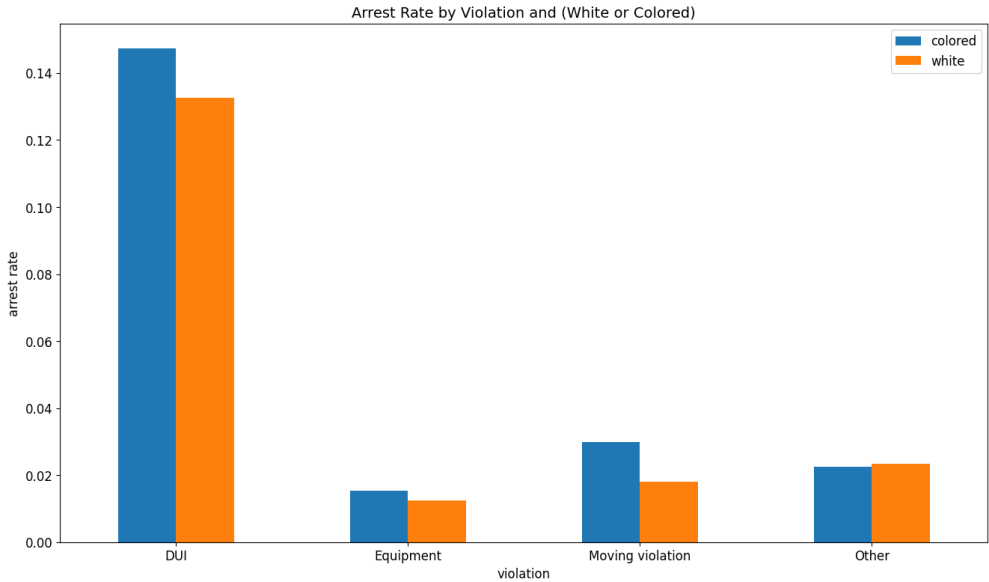


圖七、依罪名分類各種族的逮捕率

排除 other（可能為其他種類的違規罪名，如環境類的行政罰，車子超過一定時間未熄火等，種類過於雜亂不利分析統計），在各個罪行中亞裔是搜查率最低的族群，推測可能是因為亞裔犯罪率本來就比較低的緣故。在酒駕、設備、駕駛違規中，拉丁裔美國人對比白人都有較高的搜查率；而在設備、駕駛違規中，非裔美國人對比白人都有較高的搜查率；最後根據數據，加州交通警察最容易對拉丁裔進行搜查。逮捕率部分與搜查率也具有相同趨勢。

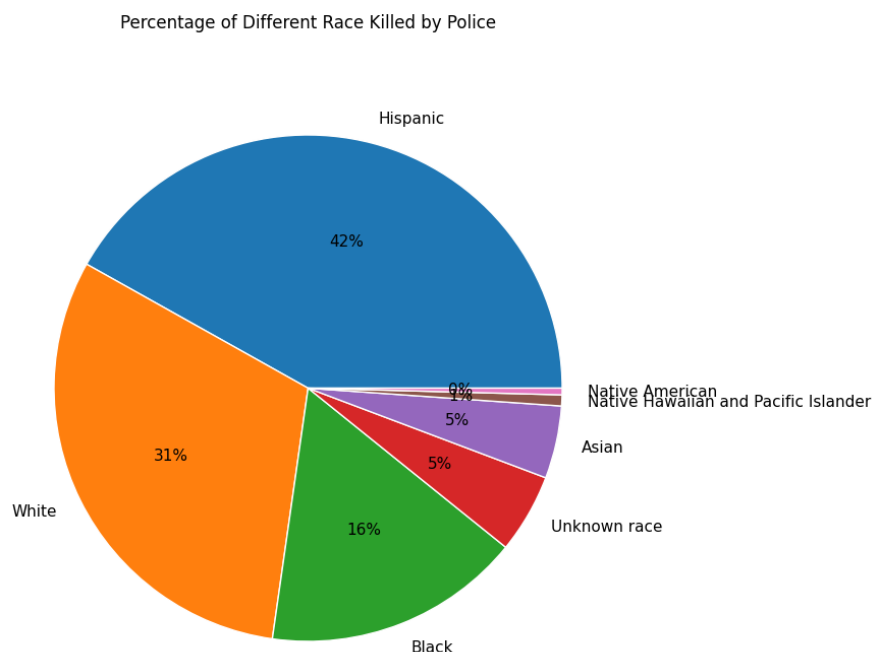


圖九、有色人種與白人的搜查率比較



圖十、有色人種與白人的逮捕率比較

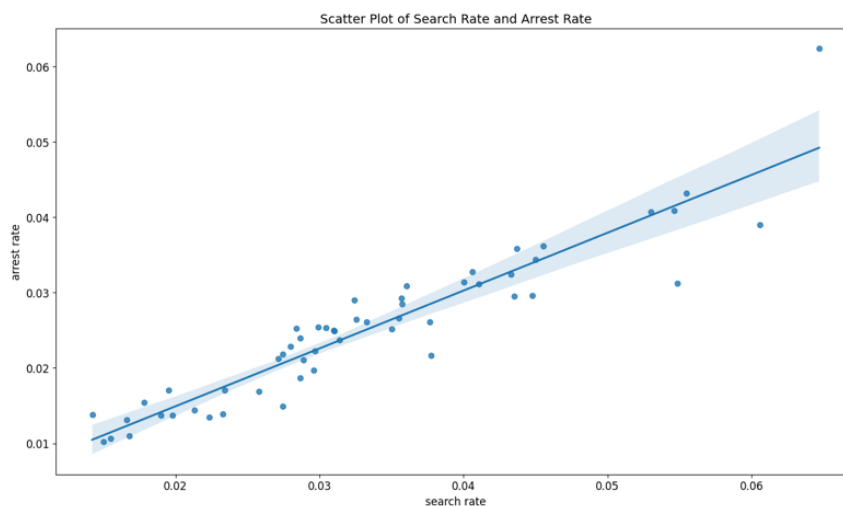
排除 other 類別（雜含各類違規），有色人種在各類罪行的搜查率都比白人高，逮捕率亦同。



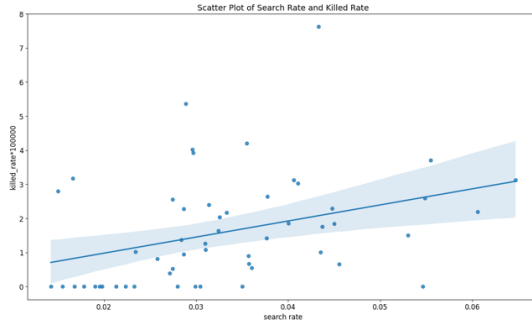
圖八、警暴致死的種族比例

由上圖可知，因為警察暴力執法過當致死的人數以拉丁裔佔最多數，其次是白人，非裔佔第三名。其實前兩名與人口比例分佈是相符合的，加州有 39%的居民自認是拉丁裔，35%居民屬於白人。然而，而僅僅 5%的居民屬於非裔，**卻佔了警暴致死的 16%**，這就與人口分布趨勢不符，雖然統計上無法直接證明一定就與種族歧視有關，但是可以合理推測，加州警察在執法時受到種族影響此一陳述，尚非無疑。

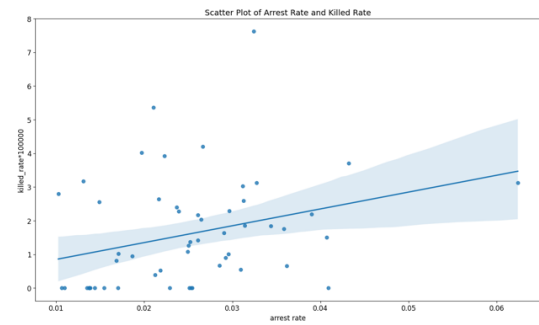
(三) 加州警察執法地圖



圖九、搜查率與逮捕率散布圖

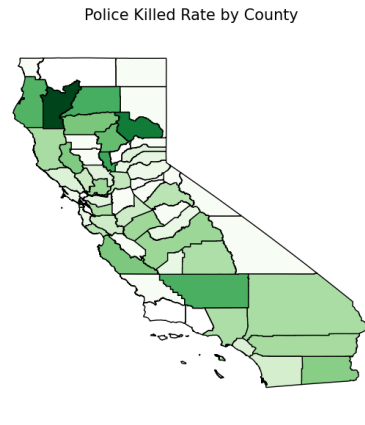


圖十、搜查率與警暴率散布圖

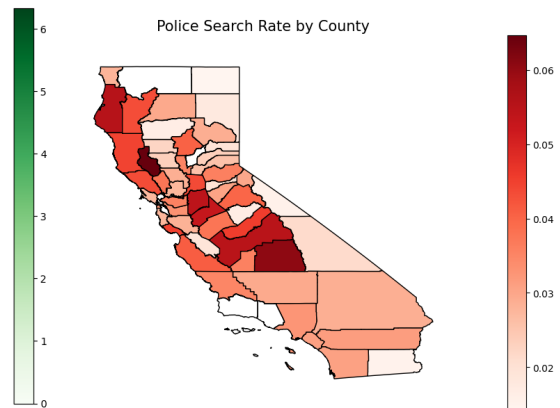


圖十一、逮捕率與警暴率散布圖

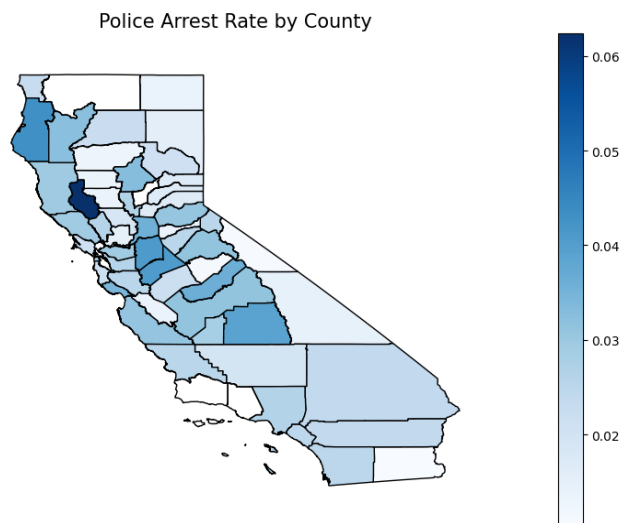
搜查率與逮捕率跟警暴率之間皆呈低度正相關。這個結果是合理的，因為搜查率、逮捕率的資料源頭是因交通事件而攔查後實施搜查或逮捕的紀錄，而警暴率的資料源頭是各郡警察暴力致死的數量統計。兩者的數據範圍略有差距，因此較難看出兩者的關係。



圖十二、加州警暴率地圖（郡）



圖十二、加州搜查率地圖（郡）



圖十三、加州逮捕率地圖（郡）

搜查率、逮捕率最高的郡都落在湖郡（Lake County, CA），警暴率最高者則落在三一郡（Trinity County, CA）。經查，該郡人口偏少，警暴率居冠的背後原因應該與母體數量過低有關（因為第二名的 Plumas County 人口也不多）。然而，警暴率第四名的 Kern County 人口排名高居加州第 11，可以合理推測此郡的高警暴率有參考價值。

六、回饋回覆與小研究

（一）關於更改逮捕率定義的建議

有同學建議說，將逮捕率定義改成「經搜查後逮捕人數／攔停總人數」，因為這樣更能反映其他變量的差距。我們聽取建議後立刻著手修改，卻發現因酒駕被攔停者，各種族被逮捕的人數皆超過被搜查人數，換言之，有些案例是不經搜查直接逮捕。

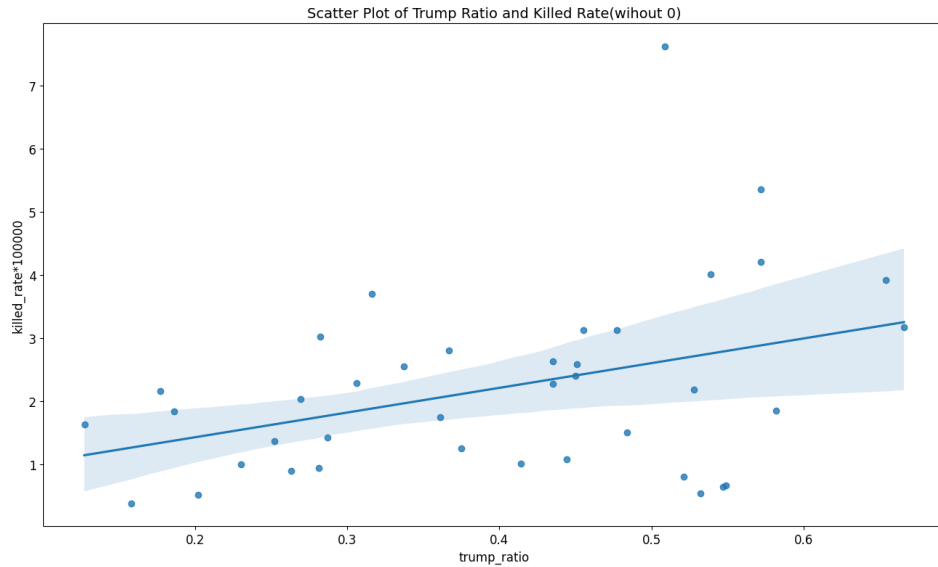
表五、更改逮捕率定義

| violation | driver_race | search_rate | arrest_rate | directarrest_rate |
|-----------|-------------|-------------|-------------|-------------------|
| DUI | Asian | 7.98% | 104% | 0.31% |
| DUI | Black | 12.1% | 102% | 0.29% |
| DUI | Hispanic | 18% | 102% | 0.28% |
| DUI | Other | 8.13% | 103% | 0.25% |
| DUI | White | 12.7% | 104% | 0.55% |

從上表可見，在同為因酒駕而被攔檢的情況下，亞裔及白人有最高的逮捕率，其中又以白人直接逮捕率(direct arrest rate)最高，是一個神奇的小發現。然而，由於存在直接逮捕的情形，若改採行同學建議的定義去做進一步的計算與分析，會有超過 100%的問題，不便於部分統計的進行。

（二）Back the Blue：政治信念對數據的影響

有建議說，加州幅員遼闊、人口眾多，每個郡的政治傾向也略有差異。例如，不支持槍枝管制、較保守的郡會傾向支持警察擁有較大的執法裁量權。Back the Blue 意譯為中文是「警察的命也是命」，即是由此而來。而經由思考過後我們發現確實有其道理，但首先我們遇到的問題是要怎麼定義「保守的郡/保守黨」。美國的兩大黨為民主黨與共和黨，在 2020 年總統大選分別推出拜登與川普作為候選人，而因為普遍而言川普的選民在許多價值上都被歸類為保守派（例如：反墮胎 pro-life、反移民 anti-immigrant），所以我們決定用 2020 川普在加州各個郡的得票率數據來分析，換言之，川普得票率較高的郡，就會被認定是「較保守的郡」。



圖十四、川普得票率與警暴率散布圖

循著上述建議，我們想知道「政治立場」與「警察執法」之間有無統計上的關聯性，結果發現兩組數據間的相關係數 r 為 0.3838，t-test 的 p 值為 0.0000000588，屬於低度相關但差異高度顯著的類型。這種情況在統計分析上應該不屬於常見的類型，因為雖然兩個變量之間相關不大，但從數據來看這種弱相關應該是存在的，而非純屬偶然。然而，這種相關的實際強度非常小，基本上可以忽略。或者，也可能是因為樣本數據量非常大，即使變量之間相關很弱，也可以被檢測出來顯著不等於 0。

綜上所述，相關係數 r 很小代表兩變量之間存在的關聯性非常微弱，即使 p 值顯著，統計上也無法直接推導出有實際的預測價值。因此，「政治立場」與「警察執法」之間統計上相關分析能得出的結果有限，需要考慮其他更複雜的統計模型來揭示變量之間的複雜關係。

（三）驗證是否有顯著差異的建議

有同學說，直接從圖表看，很難判斷有色人種的逮捕/搜查率之間是否有顯著差異，建議我們可以分析看看統計上是否有顯著差異。我們認為十分有道理，因此著手進行統計與分析。經過計算，「有色人種 vs. 白人的逮捕率」t 檢定的 p 值是 0.8719，「有色人種 vs. 白人的搜查率」t 檢定的 p 值是 0.7601，雖然從圖九、十中可見，有色人種在相同違規事項下的搜查率與逮捕率大多略高於白人，但是我們利用 Scipy 中的 `stats.ttest_ind(a,b)` 處理過後，發現其 p -value 分別為 0.76 及 0.87，因此推論兩者並無顯著差異。

綜上所述，我們認為此建議很有道理，推論出「有色人種 vs. 白人的逮捕率、搜查率」之間雖然前者高於後者有可見的差異，但統計上並不顯著。

七、分工表

魏子鈞：資料搜索（資料二、三、四）、資料前處理（資料三、四）、資料視覺化（加州警察執法地圖）、製作 PPT、整理回覆同學回饋並調整研究方法、製作書面報告

徐子淵：資料搜索（資料一）、資料前處理（資料一、二）、資料視覺化（統計圖、年曆、相關性分析、顯著性分析）、製作 PPT、調整研究方法、HTML 網頁製作、製作書面報告