

# MATLAB 的大數據分析與應用

指導老師：張光遠

小組組員：

00757046 黃子軒

00757026 邵安祺

00757142 謝宛蓉

00757123 林湘羚

## 壹、研究目的與動機

盛夏將至，登革熱疫情都會在每年七至八月逐漸猖狂，並在九至十月達至高峰，本次大數據分析主要探討造成台灣登革熱疫情爆發的影響因子，包括氣象因子（如溫度、濕度）。另外，藉由各式統計病例數來次要探討登革熱的歷年嚴重程度、活躍月份以及確診族群關係，試著從統計圖中歸納出登革熱的全貌，凸顯防治登革熱的重要，並為台灣登革熱疫情盡一份心力。

## 貳、研究方法

首先，從「政府資料開放平臺」取得登革熱相關資料進行分析，內容包含案例年齡、所在區域以及性別等資料。再利用爬蟲技術（使用語言：python），從「中央氣象局」歷年每月氣象取得所需的各年各月的溫度與濕度。由於資料完整性，我們取 2009 年 1 月至 2020 年 4 月用以統計。將兩者進行交叉分析，並且觀察分析後的結果。

## 參、相關作品調查比較

【行政院衛生署】氣候變遷下台灣登革熱空間時間分布預測模型建立研究  
<https://www.cdc.gov.tw/uploads/files/6b7a57f1-c4e1-4d31-9c0b-9b6836258bf2.pdf>

【行政院科技部】登革熱的台灣經驗

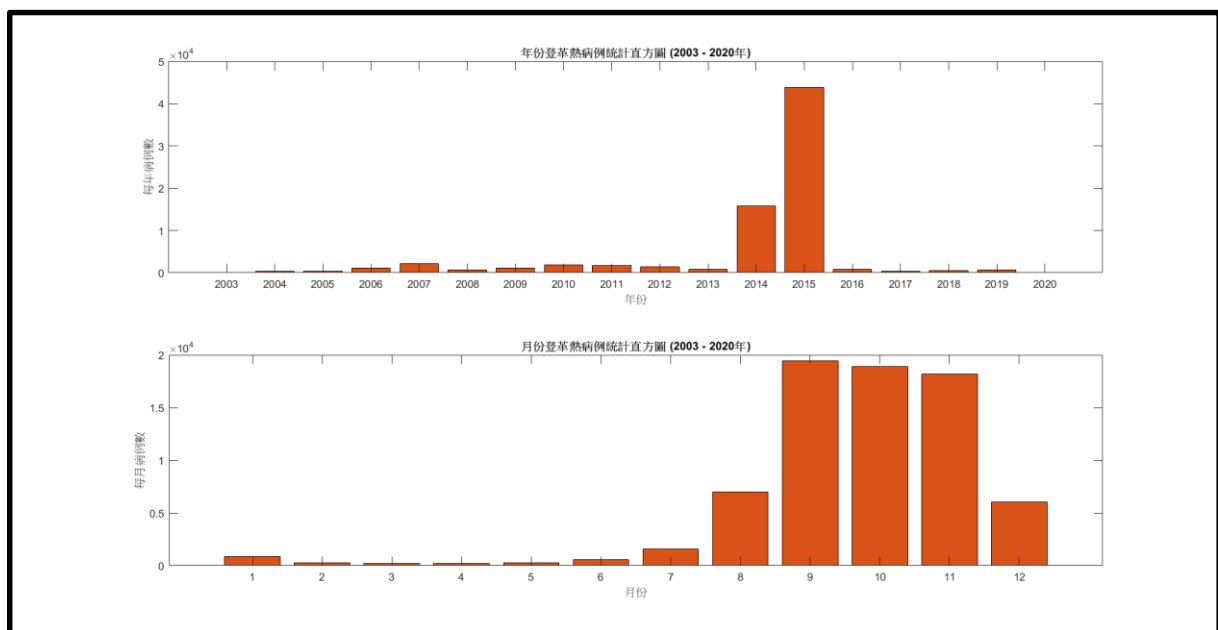
<https://www-ws.pthg.gov.tw/Upload/2015pthg/82/ckfile/a967348c-df12-4399-855f-2c679cbedc32.pdf>

【大數聚】從大數據看登革熱事件(2015 年)

<https://group.dailyview.tw/article/detail/418>

## 肆、圖表呈現

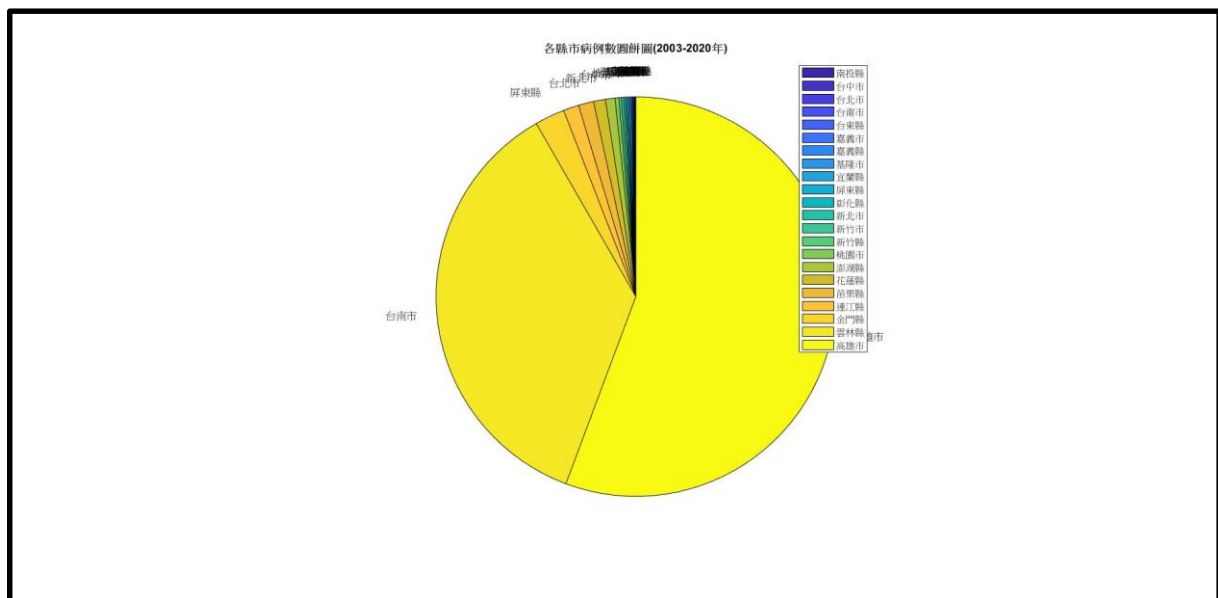
總病例時間分佈



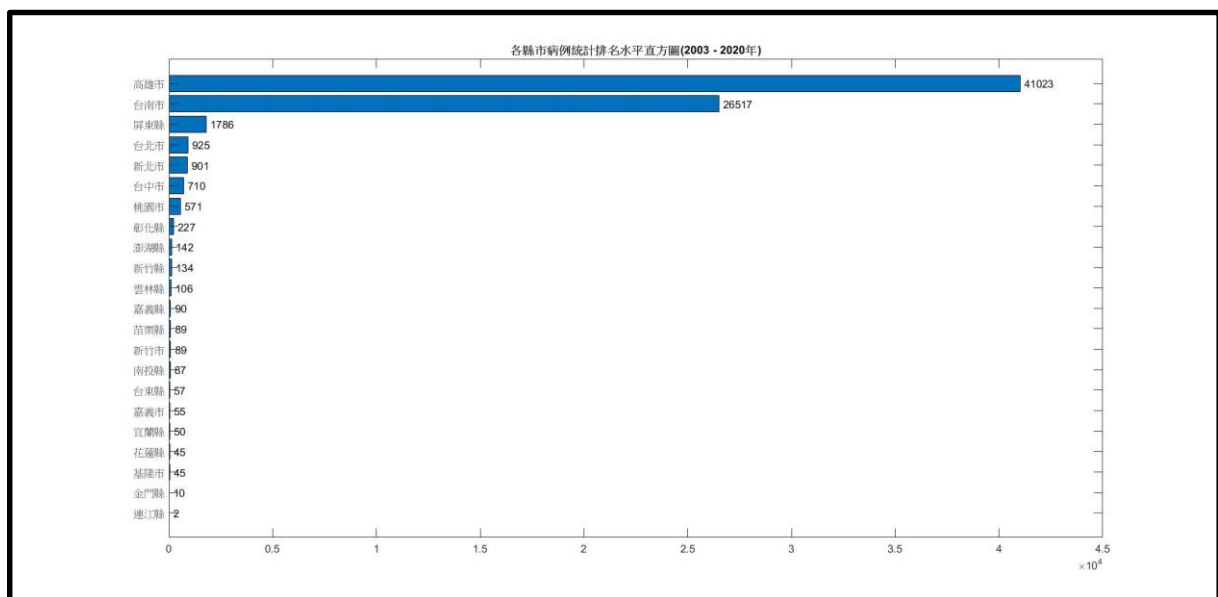
(圖 1)年份登革熱病例統計直方圖 (2003 - 2020 年)及月份登革熱病例統計直方圖 (2003 - 2020 年)

從(圖 1)中，可簡易看出登革熱大爆發的年份，與登革熱盛行的月份以及高峰、低峰期。

## 總病例地區分佈



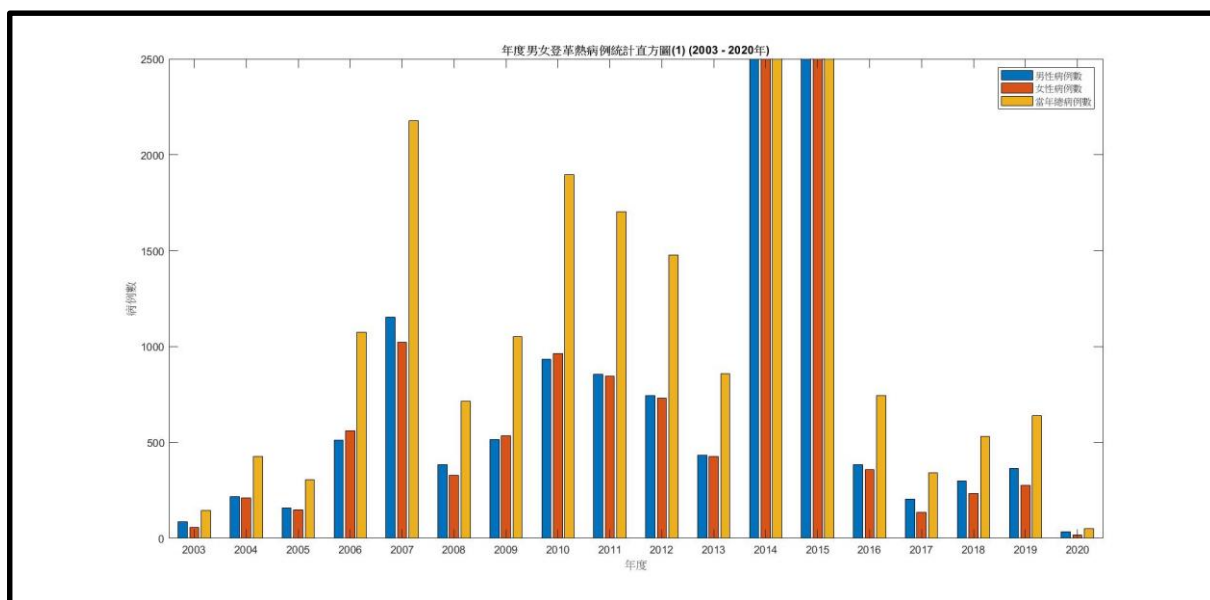
(圖 2)各縣市病例數圖餅圖(2003 - 2020 年)



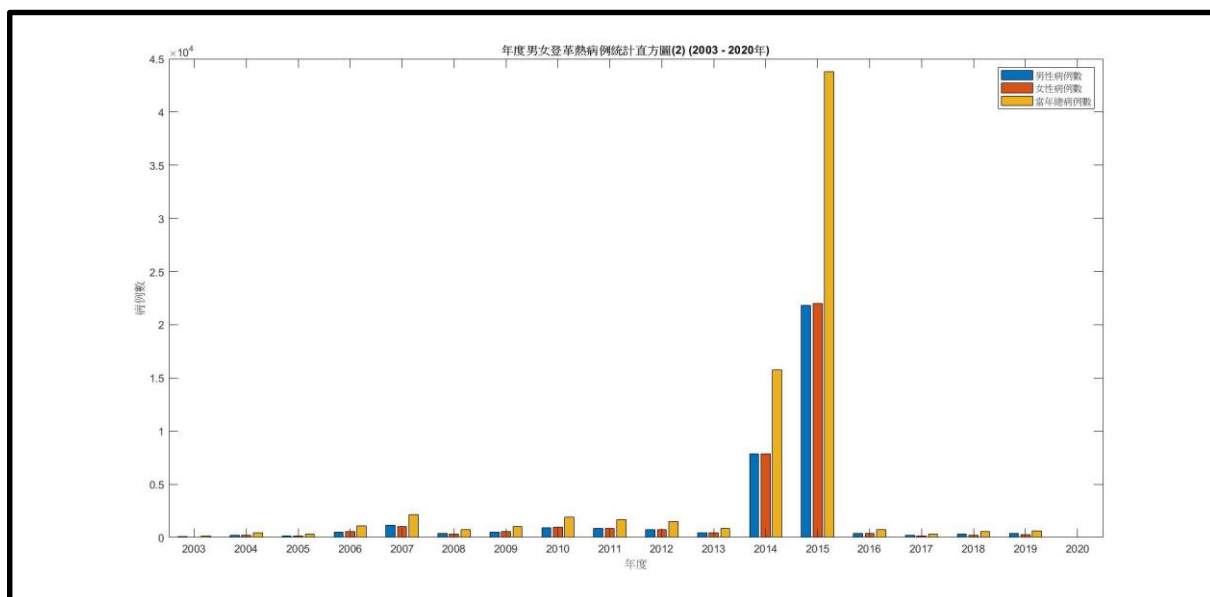
(圖 3)各縣市病例統計排名水平直方圖

從(圖 2)、(圖 3)中，縣市累積病例數前三名由「高雄市」、「台南市」與「屏東縣」包下，可見 2014、2015 年登革熱疫情的爆發尤為嚴重，查詢相關資料後發現，可能與高雄意外氣爆又偏逢連夜雨有最大關係；台南年初缺水造成民眾的儲水行動，夏季又遭逢蘇迪勒和杜鵑兩大超級颱風肆虐，故也和積水孳生媒蚊有最大關係；屏東地處台灣最南端，夏季颱風環伺，屏東大概率首當其衝，且鄉村型的縣市常有更多孳生源，是病媒蚊喜好的環境。

## 總病例性別分布

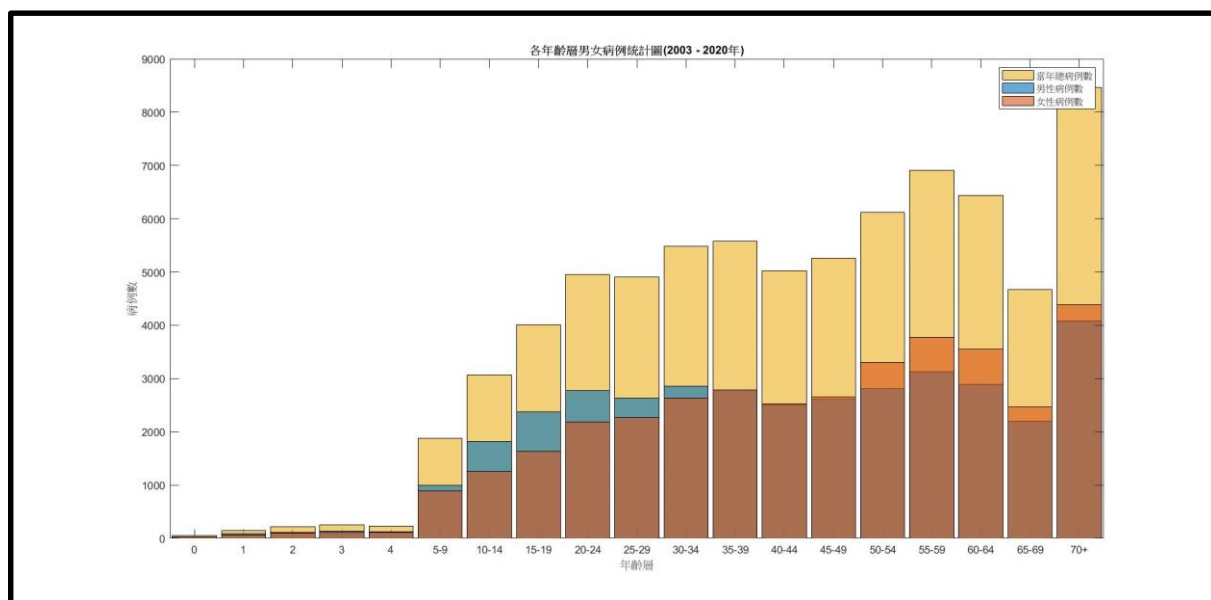


(圖 4-1)年度男女登革熱病例統計直方圖(細節)(2003 - 2020 年)



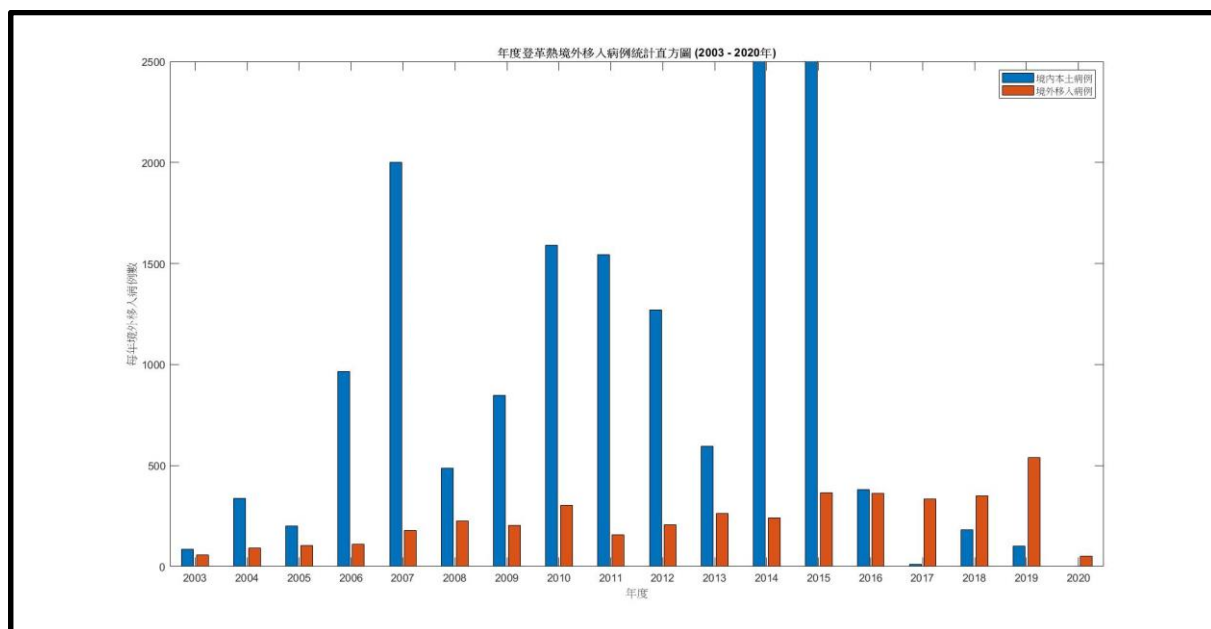
(圖 4-2)年度男女登革熱病例統計直方圖(全圖)(2003 - 2020 年)

## 總病例年齡層分佈



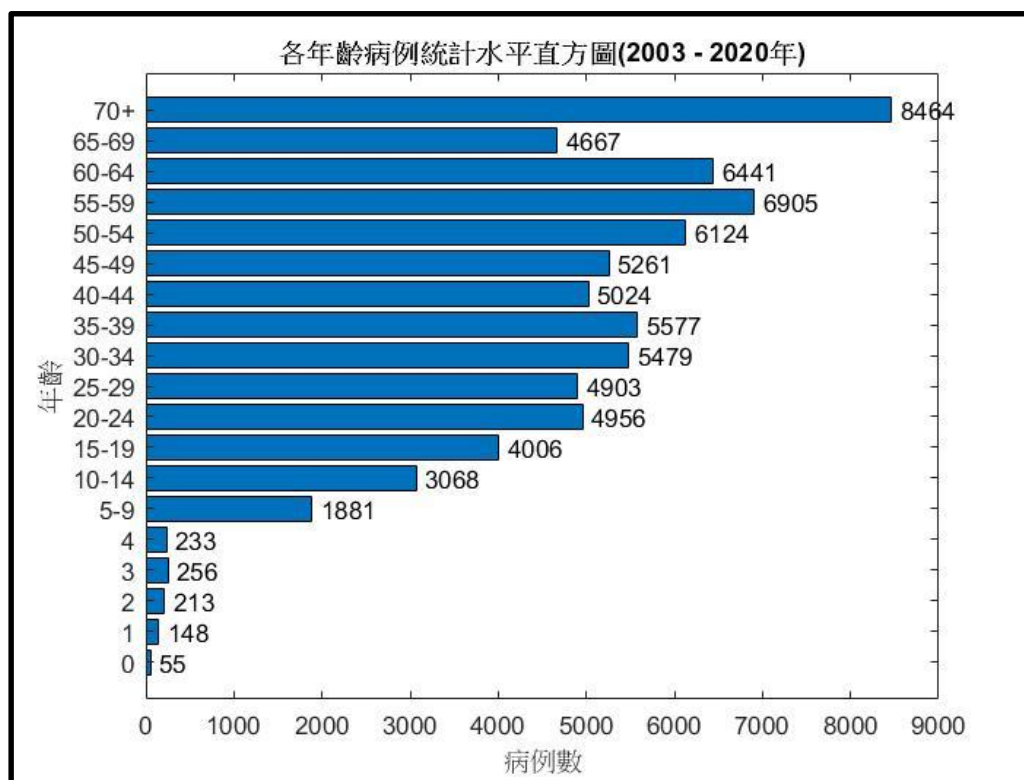
(圖 5)各年齡層男女病例統計圖(2003 - 2020 年)

從(圖 4)、(圖 5)，可觀察到各年度以及各年齡層的男女病例數量差異不大，但高齡層相對於低齡層感染人數較多。



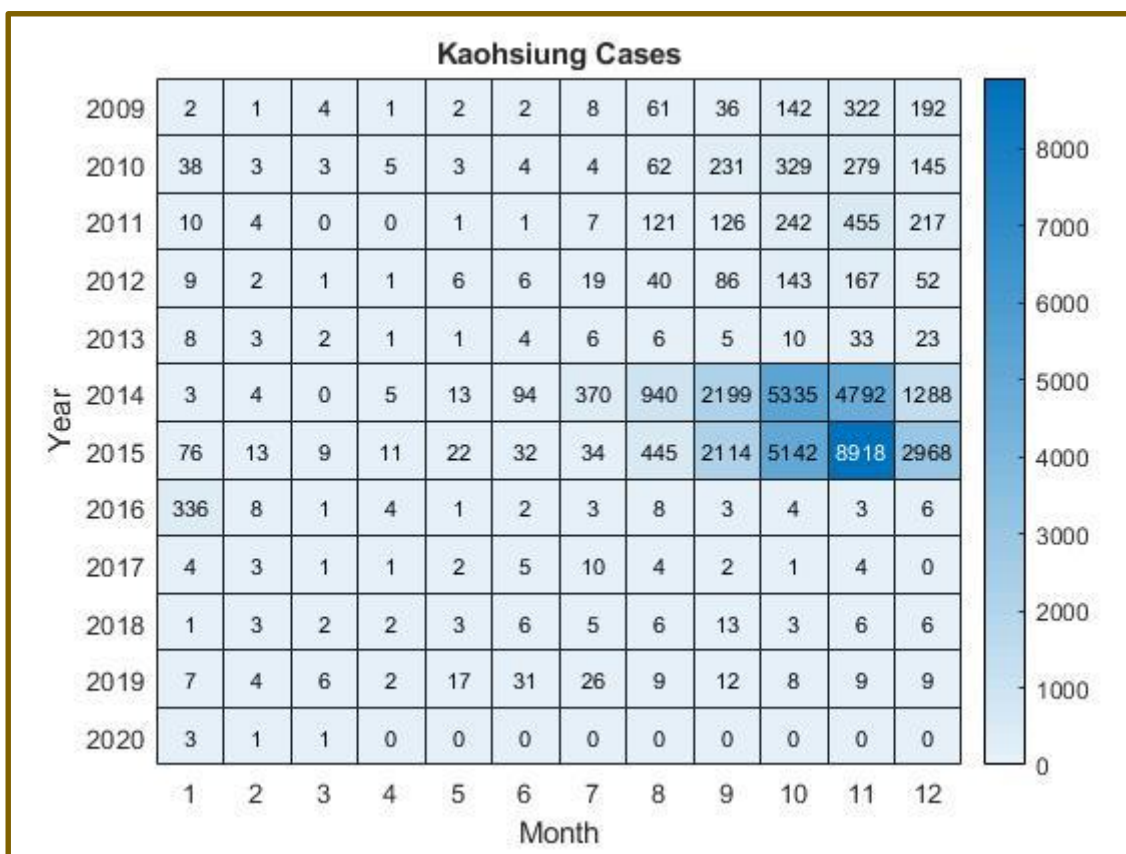
(圖 6)年度登革熱境外移入病例統計直方圖(2003 - 2020 年)

從上圖表中，不難發現一有趣現象——至 2017 年起，境內本土病例就開始低於境外移入病例，我們推斷和國家蚊媒傳染病防治研究中心給予高雄市府協助有關，除了戶外布放「誘蚊產卵器」，藉當地居民協助首次嘗試將「蚊媒誘殺桶」布放室內，經監測後效果顯著。

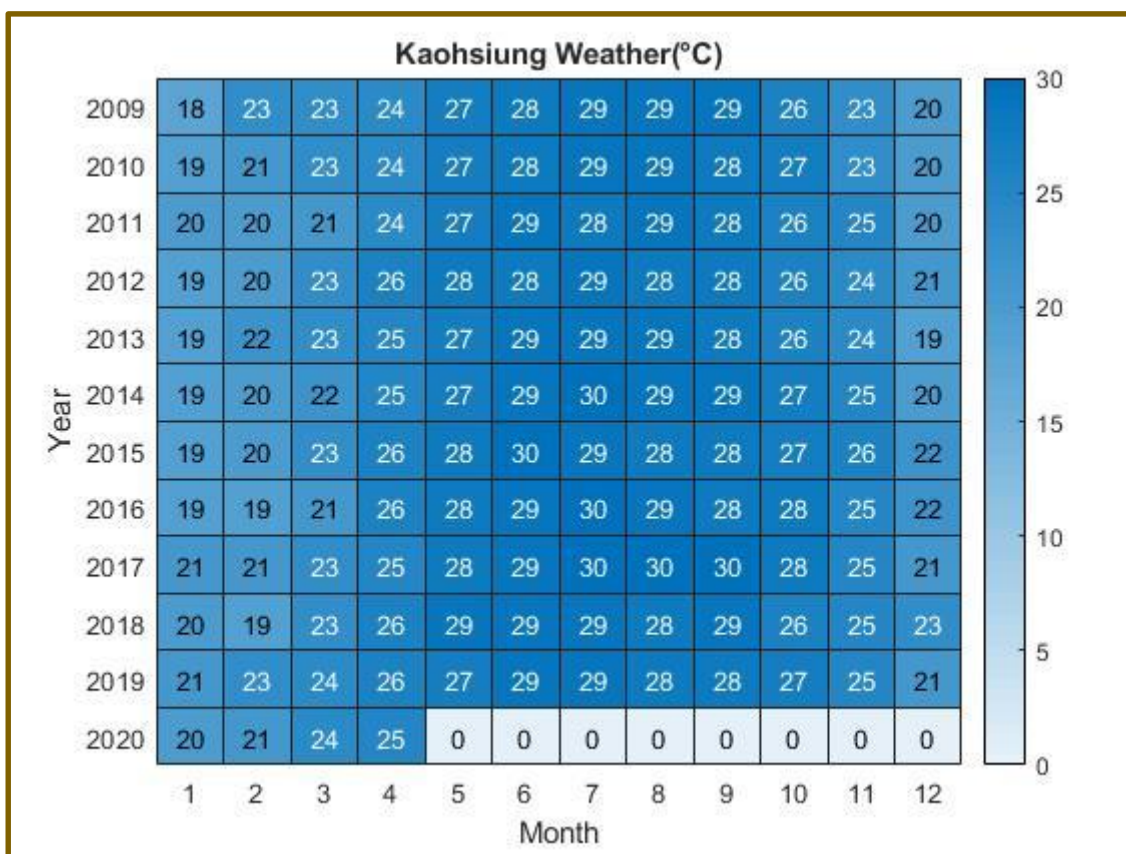


(圖 7)各年齡病例統計水平直方圖(2003 - 2020 年)

從上圖表中，可觀察到 50 歲以上高齡人口的累積病例數正式突破六千例以上，大體上，台灣幼年人口的罹患率(於 2016 年 5 月修訂)相對低於成年人口很多，症狀與病程亦較成年人口輕微且不太典型，老年人口的罹患率雖然較低，但台灣感染族群以老年人居多，我們推斷是由於老年人常聚集且免疫力不如成年人所致。



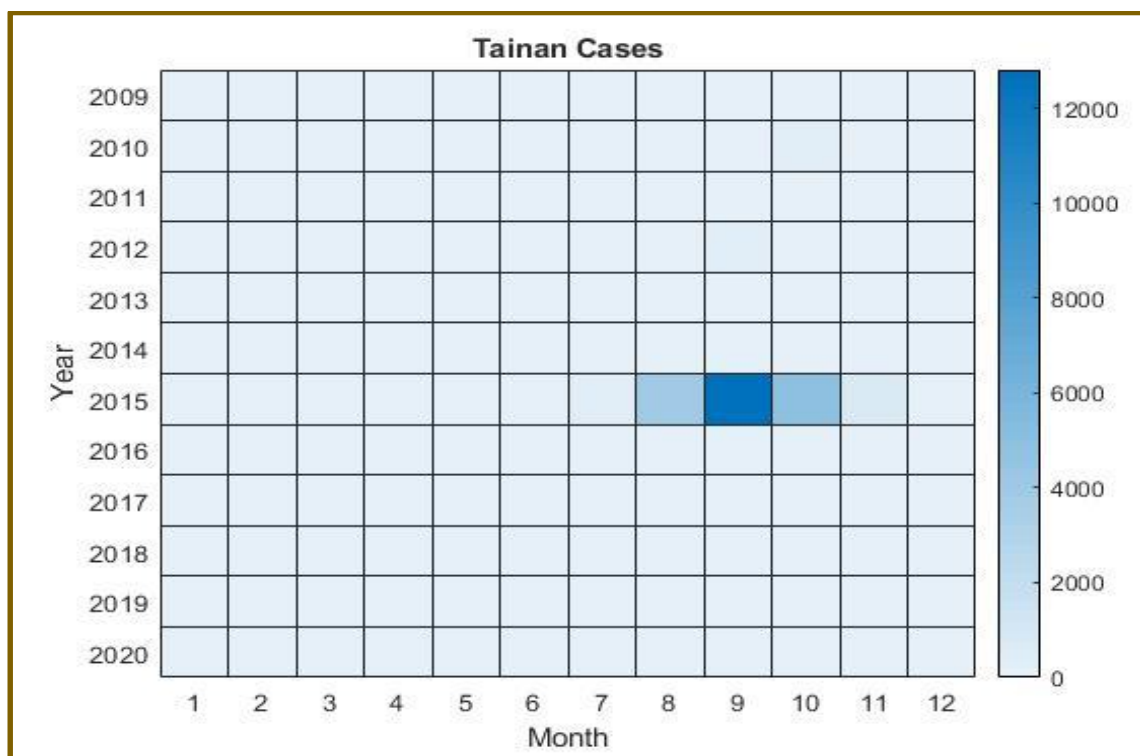
(圖 8)Kaohsiung Cases



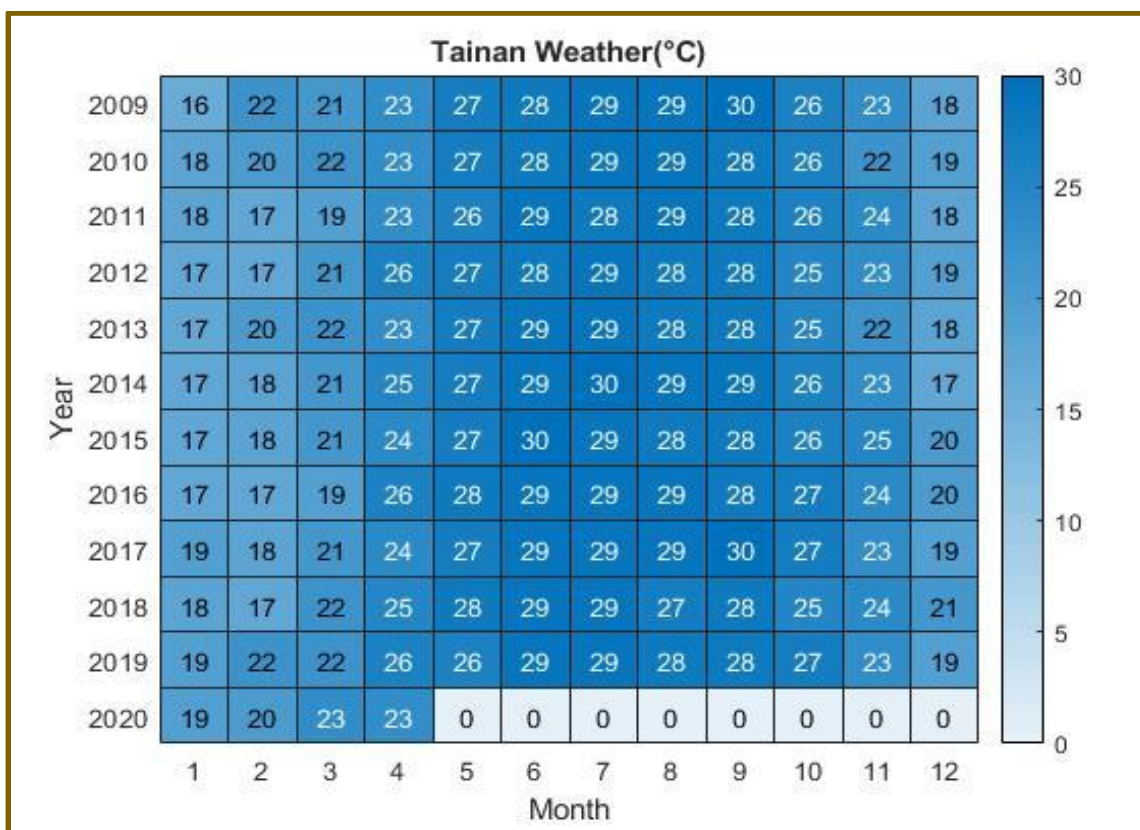
(圖 9)Kaohsiung Weather(°C)



以上兩張圖表分別是高雄 2009~2020 每月的登革熱 case 數跟溫度。登革熱高峰期集中在 8~12 月。



(圖 10)Tainan Weather(°C)



(圖 11)Tainan Weather(°C)

同上，以上兩張圖表分別是臺南 2009~2020 每月的登革熱 case 數跟溫度。跟高雄一樣顛峰期在 8~12 月。

綜合以上四張圖表，得知確診人數和溫度是正相關的。2014、2015 的夏天比其它年平均多一度，而造成 10~11 月的確診人數多於其他年。除此之外，2015 再加上聖嬰現象的影響，而冬天高於 20 度的天數變多，冬天也比其它年溫暖，讓蚊子趁暖而入，繁衍數量劇增，導致感染人數也暴增，成為歷史上確診人數最多的一年。

## 伍、大數據分析結果與實際情形對比

以下為我們推測 2014 高雄疫情大爆炸的可能原因：

1. **人為因素**：回顧 2014 年，因為**高雄發生氣爆**，又遇連日大雨，氣爆過後許多廢棄空間積水，應是登革熱大爆發主因。
2. **環境因素**：熱帶性氣候、暖冬影響，導致**高雄 2 月就有病例**(登革熱大多 5 月才開始流行)。
3. **病毒傳播**：傳播登革熱的凶手之一「**埃及斑蚊**」，主要分布在嘉義布袋以南的各縣市，埃及斑蚊又比白線斑蚊傳播登革熱更有效率，因此過往台灣登革熱的重災區本來就主要集中於南部，而高雄、台南尤為嚴重。

以下為我們推測 2015 台南疫情大爆炸的可能原因：

1. **環境因素**：年初**水情吃緊**，促進民眾儲水行為，提供病媒蚊良好的孳生環境。入夏後雨勢不斷，加上父親節的**蘇迪勒颱風**及中秋節的**杜鵑颱風**，豐沛雨量增強病媒蚊的大量繁殖，亦增加防疫工作的困難度。
2. **病毒特性**：當年流行的登革熱病毒為第二型，過去臺南市較少流行，部份症狀也為**非典型**，如嘔吐及腹瀉，較難即時診斷。

3. **人為因素**：數個公共場所如跳蚤市場、花卉市場等處，查獲時疫情已悶燒多時，受限**結構複雜、孳清不易**等問題，導致關聯個案如同火苗般延燒各處，亦造成疫情跨區分布。面對急速擴大的疫情，**防疫人力、物資缺乏**，無法有效控制疫情。

參考資料：

<http://www.mmh.org.tw/taitam/ic/ic/ic05-001.htm>

## 陸、結論

在登革熱大數據分析完後，在眾多圖表中不難發現「高雄市」與「台南市」在 2014、2015 年突破天際般的病例數，起初困擾著我們無法將分析後的數據更完美且均勻地呈現，但反倒引起了我們的好奇心，為何兩市有如此「突出」的表現？於是我們將箭頭指向第二大數據——氣象數據，透過分析後圖表的呈現，我們發現 2014、2015 的夏天確實些微地比其它年高，這正好可以解釋為何這兩年的病例數為何這麼高。

## 柒、參考資料

【政府資料開放平台】地區年齡性別統計表-登革熱

<https://data.gov.tw/dataset/6879>

【交通部中央氣象局】每月氣象

<https://www.cwb.gov.tw/V8/C/C/Statistics/monthlydata.html>