107全國大學校院數位人文大數據學生競賽

臺中選舉分析

張鈞政 周亭錡 陳怡廷 曾秀雯 徐嘉璟
江曉嵐 陳嘉葳 楊舜翔 白祐誠 黃敬恆 鄧喻安

棄保

總體層次資料

變異性不足

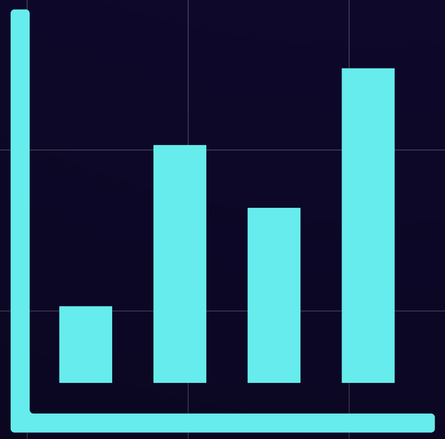
變天

村里整併

出乎意料

總體層次資料

又老又窮！



總體層次資料

拼經濟！

芝加哥學派

選區總體資料

＋

人口學變項

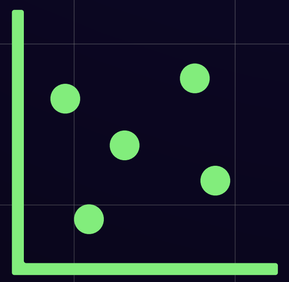
個體層次資料

→ 釐清脈絡因素支持誰？

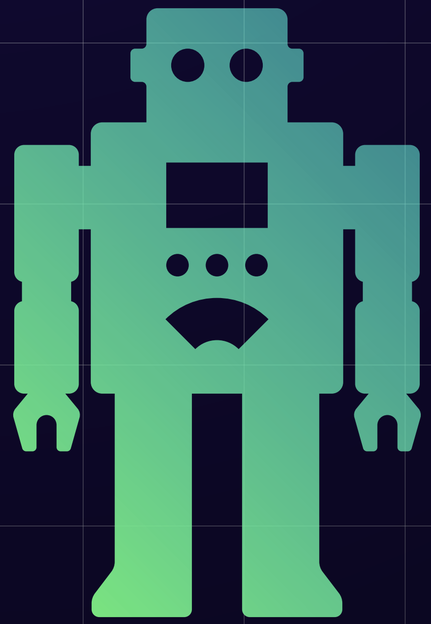
①

如何研究？

升格後五年度選舉資料
2010、2012、2014、2016、2018



人口與社會變項
人口密度、性比例、老化指數、
大專以上畢業比例、所得中位數

機器  學習

了解選舉情形和社會變項之間的關係

如何研究？②

以 **催票率** 區域內政黨總得票數 / 區域內總選舉權人數 取代 **得票率** 區域內政黨總得票數 / 區域內總有效票數

納入投票率後，找出「基本盤」。

大綱

① 2018大選
& 近五年現況

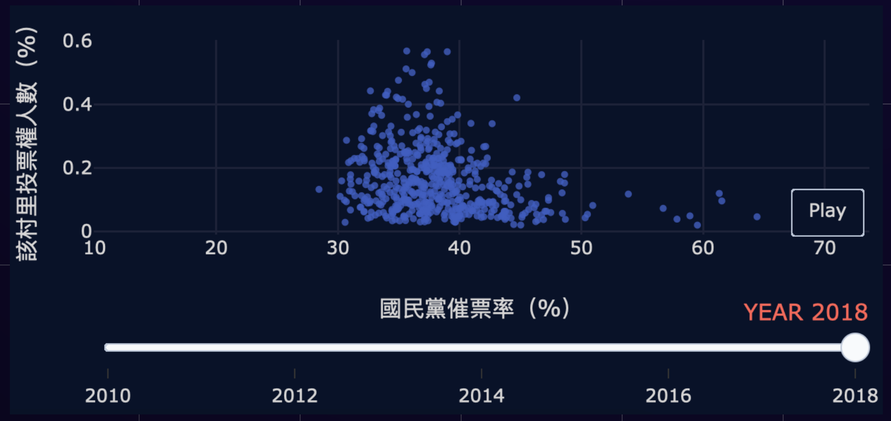
② 空間分析
& 變項散佈圖

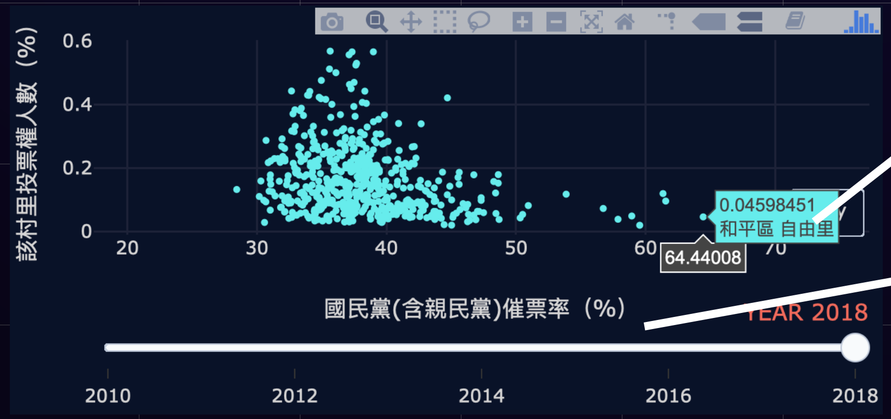
③ 以機器學習
檢驗變項

④ 泛藍泛綠
基本盤呈現

① 2018大選& 近五年現況

以互動式圖表呈現五年催票率之變化





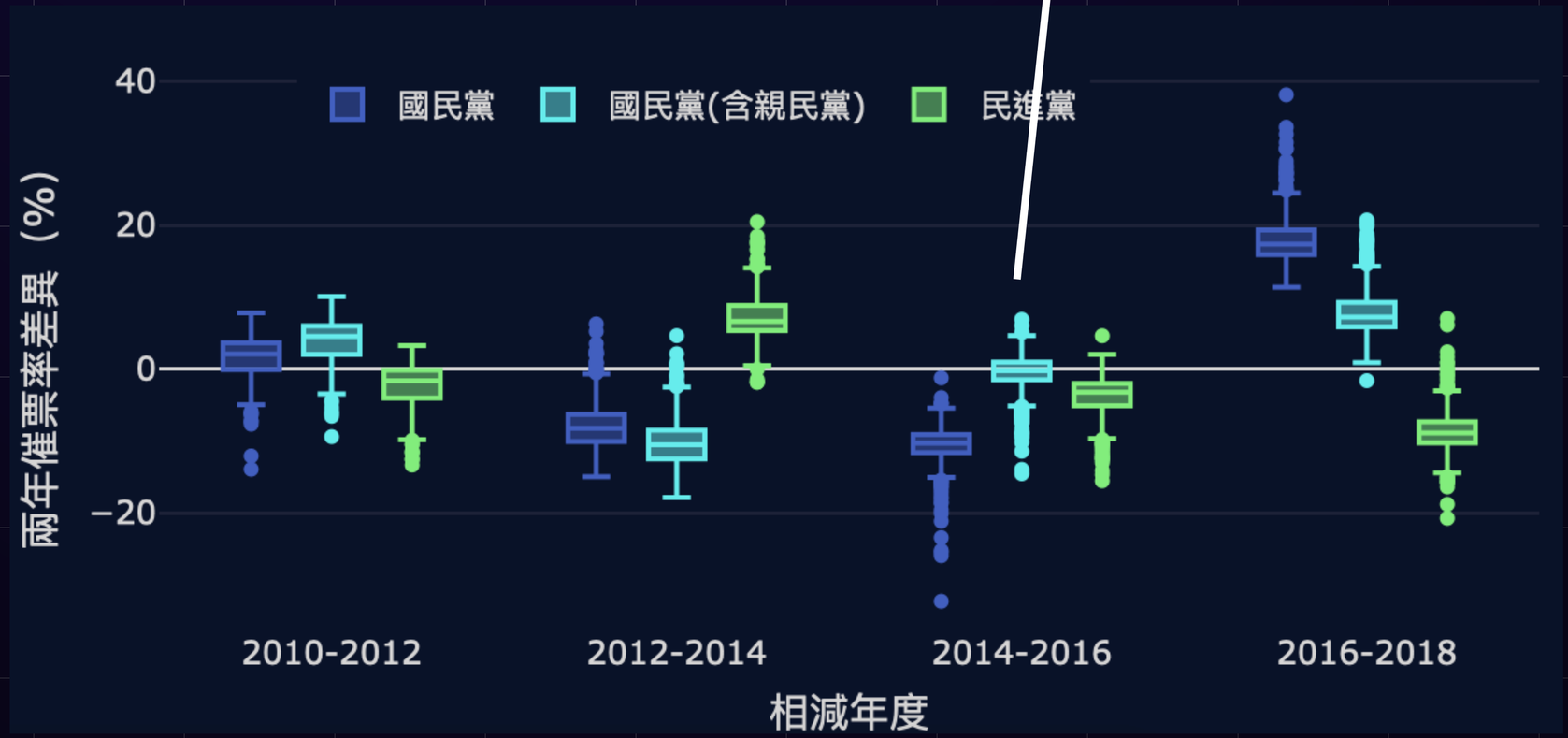
可點選顯示村里資料

泛藍（國民黨＋親民黨）

① 2018大選& 近五年現況

盒形圖呈現每兩年催票率變化

避免競合關係干擾，
將國民黨與親民黨視作
「泛藍」



② 空間分析 & 變項散佈圖

Moran’s I

了解相鄰的地區是否有相像行為

橫軸：與周邊區域的相關程度

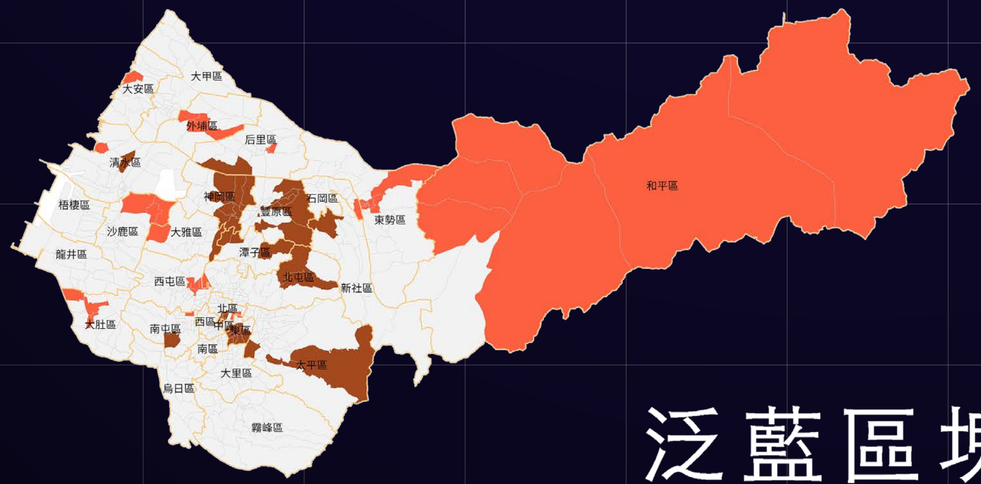
縱軸：催票率高低

縱軸變項高於平均值，
且與周邊區域有空間正相關

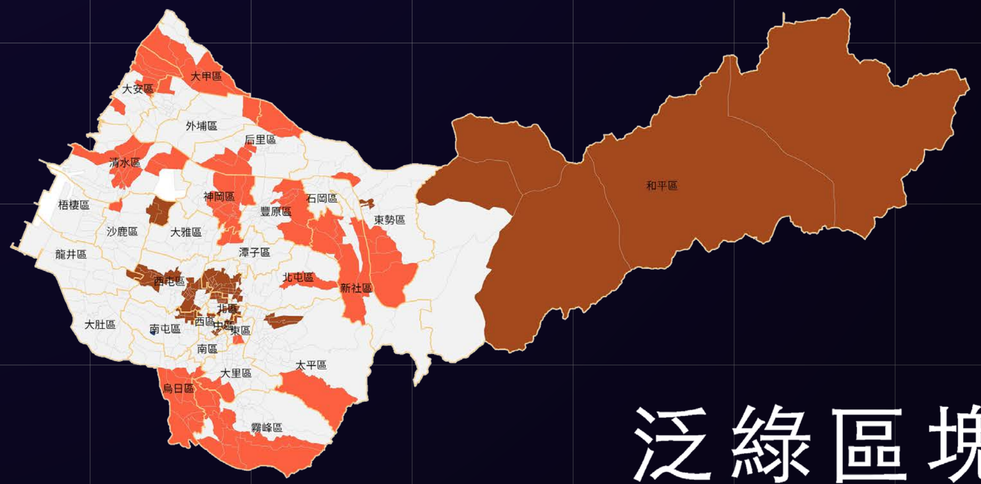
② 空間分析 & 變項散佈圖

2010年LISA

普通區　high-high　low-high　low-low



泛藍區塊

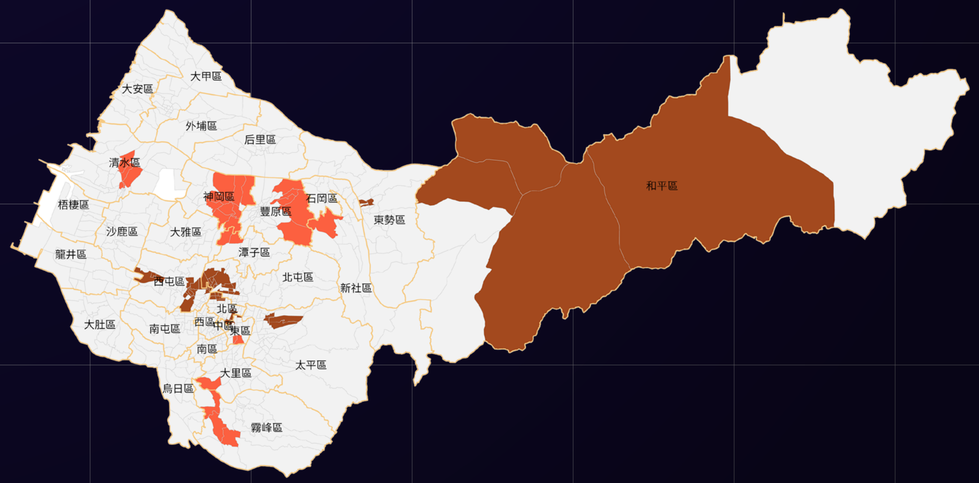


泛綠區塊

② 空間分析 & 變項散佈圖

熱區、冷區

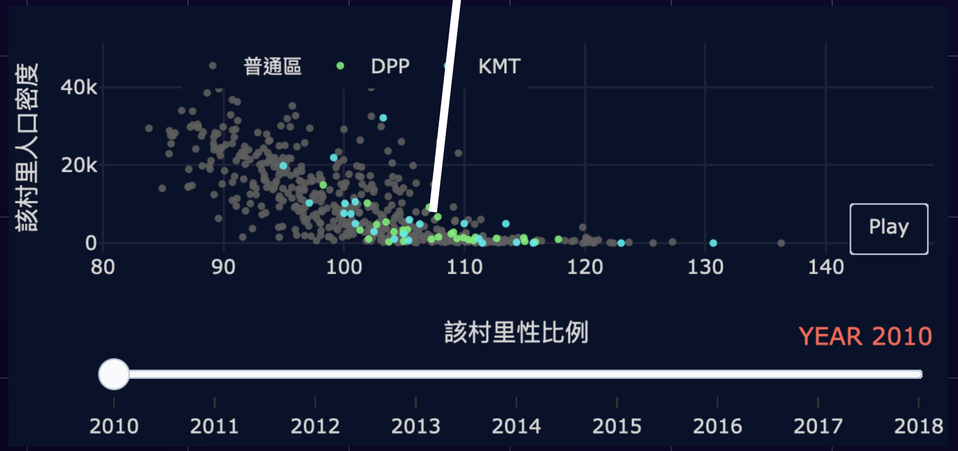
普通區　high-high　low-high　low-low



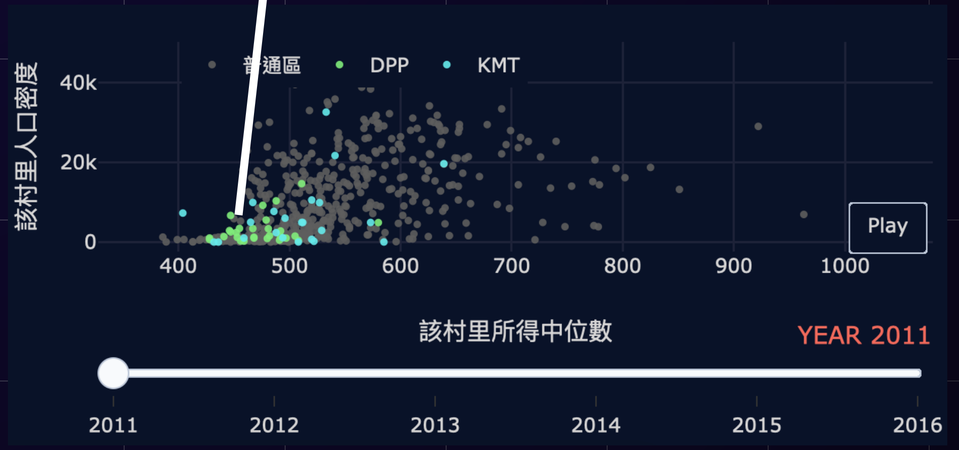
將熱區圖與其他村里之人口與社會變項做對照

② 空間分析 & 變項散佈圖

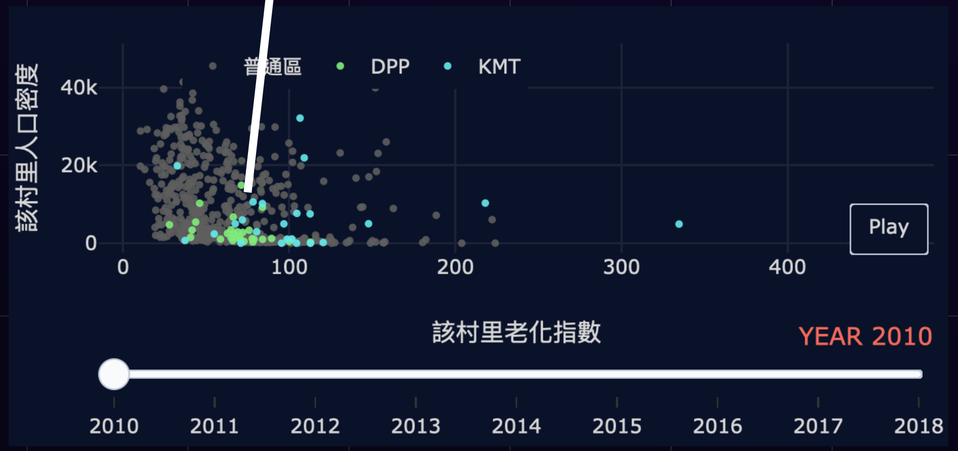
性比例較低



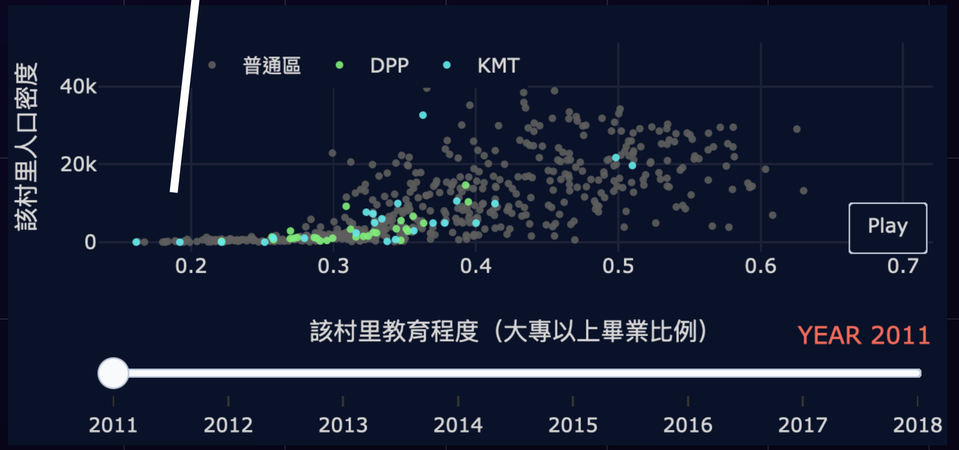
所得中位數較低



無明顯差異



大專以上畢業比例較低



③ 以機器學習檢驗變項

置入變數後以隨機森林（random forest）製作模型

模型一（ 精準度：0.9681）：

2010年國民黨催票率 + 2010年到2012年國民黨催票率的變化率 + 2010年民進黨催票率 +
2010年到2012年國民黨催票率的變化率 +2010年人口密度 +
2010年到2014年人口密度的變化率 +2010年性比例 + 2010年到2014年性比例的變化率 +
2010年老化指數+ 2010年到2014年老化指數的變化率 +2010年大專生以上比例 +
2010年到2014年大專生以上比例的變化率 +2010年收入中位數 +
2010年到2014年收入中位數的變化率

去預測 2014年該村里的催票率是藍大於綠抑或是綠大於藍

③ 以機器學習檢驗變項

置入變數後以隨機森林（random forest）製作模型

模型二（精準度：0.9521）：

2012年國民黨催票率 + 2012年到2014年國民黨催票率的變化率 +2012年民進黨催票率 + 2012年到2014年國民黨催票率的變化率 + 2012年人口密度 + 2010年到2016年人口密度的變化率 +2012年性比例 + 2010年到2016年性比例的變化率 + 2012年老化指數+ 2010年到2016年老化指數的變化率 +2012年大專生以上比例 + 2010年到2016年大專生以上比例的變化率 + 2012年收入中位數 + 2012年到2016年收入中位數的變化率

去預測 2016年該村里的催票率是藍大於綠抑或是綠大於藍

③ 以機器學習檢驗變項

置入變數後以隨機森林（random forest）製作模型

模型三（精準度：0.9521）：

2014年國民黨催票率 + 2014年到2016年國民黨催票率的變化率 +2014年民進黨催票率 +
2014年到2016年國民黨催票率的變化率 +2014年人口密度 +
2014年到2018年人口密度的變化率 +2014年性比例 + 2014年到2018年性比例的變化率 +
2014年老化指數+ 2014年到2018年老化指數的變化率 + 2014年大專生以上比例 +
2014年到2018年大專生以上比例的變化率 +2014年收入中位數 +
2014年到2018年收入中位數的變化率

去預測 2018年該村里的催票率是藍大於綠抑或是綠大於藍

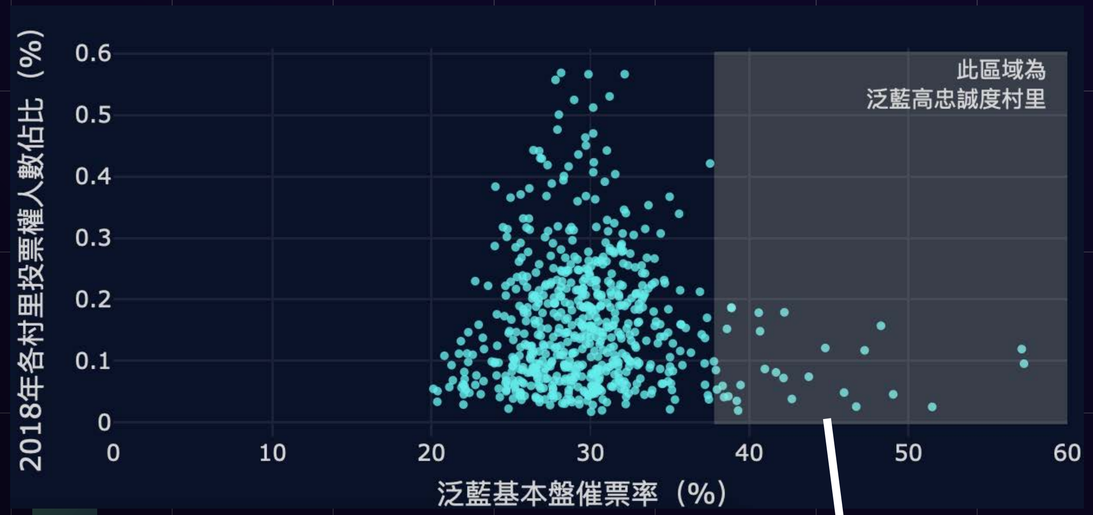
③ 以機器學習檢驗變項

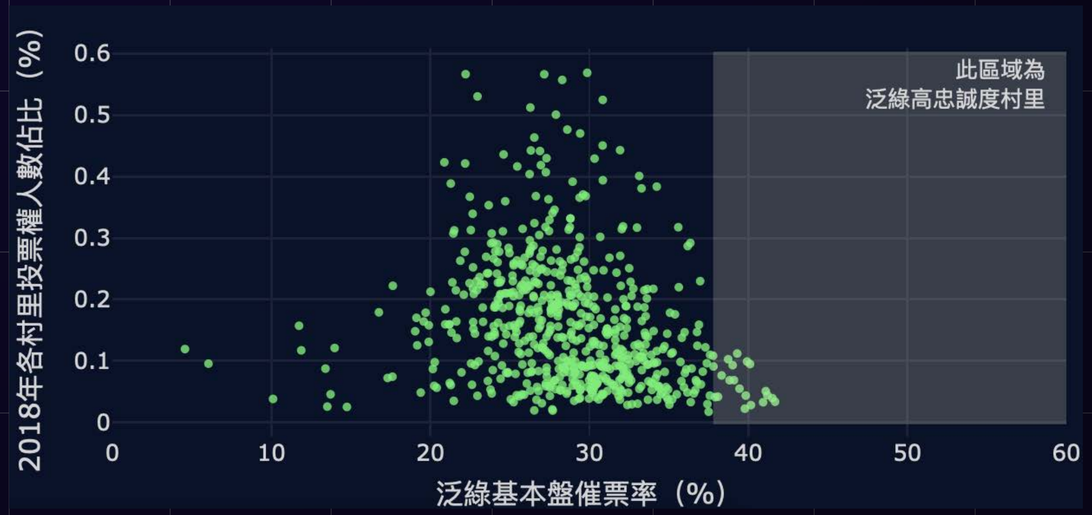
結論：

1. 以上三組模組的準確度都有高達95%以上的精準度，但因以上三組模組並不相同，因此並非就找到可以拿來預測未來選戰的最好模組。
2. 模組交叉套入資料精準度會大降，有趣的是前兩組模組交叉使用的精準度仍然很高，但最後一組模組卻無法與前兩組交叉使用，若使用將會導致精準度急遽下降。
3. 在此，五項人口與社會變項雖皆有高度的解釋力，但其權重會因時間而變，這也是在現在進行選舉預測時，多採用個體資料（民意調查）為主流之原因。
4. 所以本研究並不打算以社會變項去預測下次的選舉結果，而是嘗試以機器學習的方式提供一個新的處理總體資料的選項，讓往後選舉預測的途徑可以更多元。

④ 泛藍泛綠基本盤呈現

五次選舉中最高投票率75.76%*50%=催票率0.3788
高於此值的村里稱為高忠誠度村里。

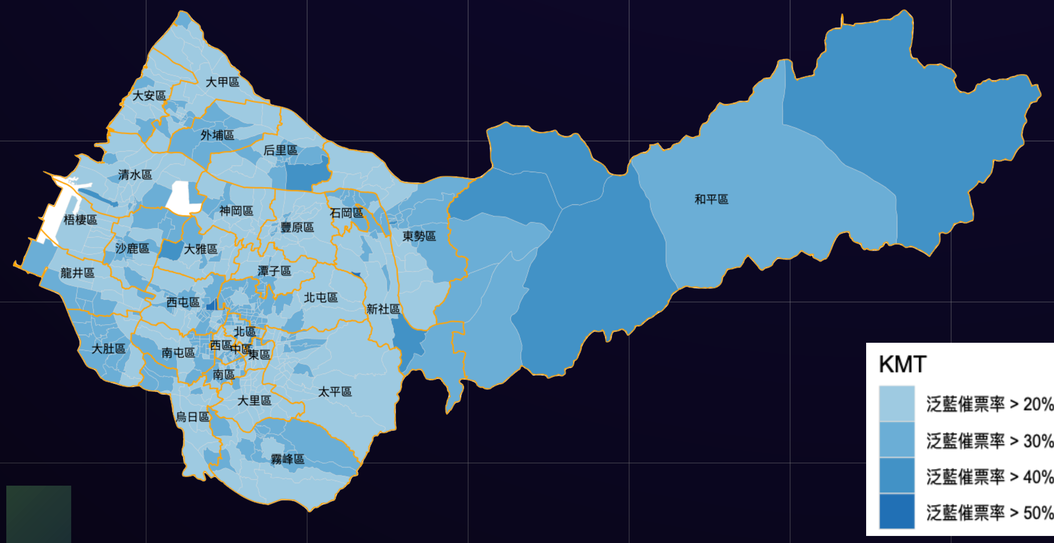




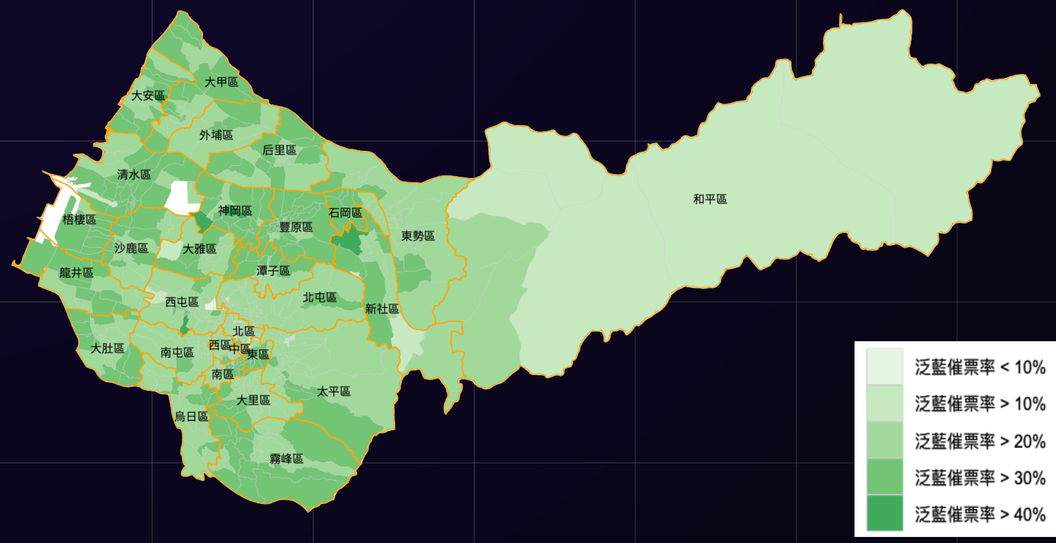
泛藍高忠誠度村里高於泛綠者

④ 泛藍泛綠基本盤呈現

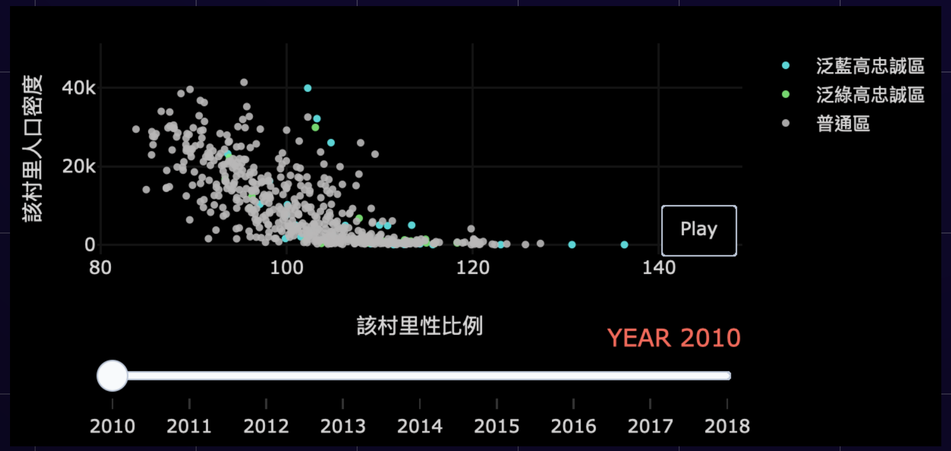
泛藍基本盤地圖

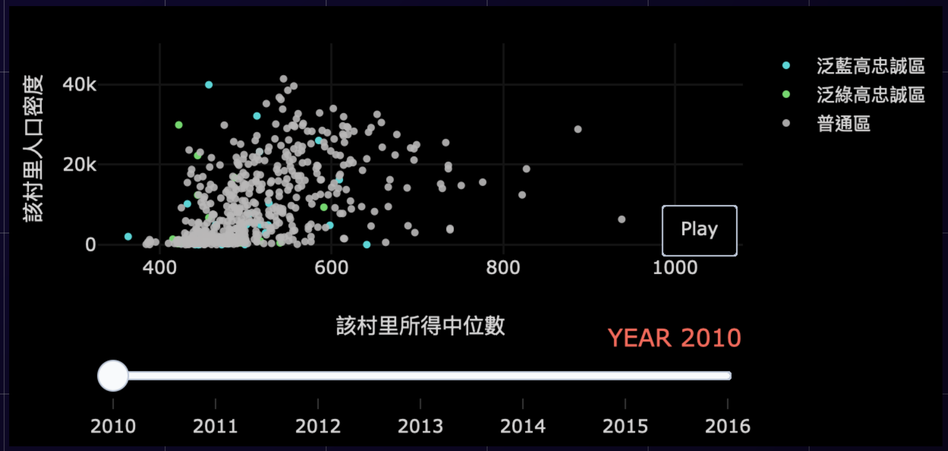


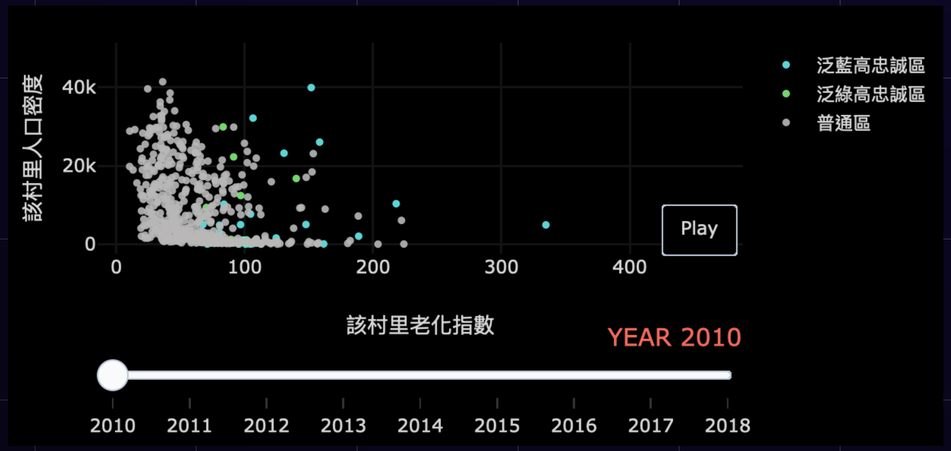
泛綠基本盤地圖



④ 泛藍泛綠基本盤呈現


該村里人口密度
40k
20k
0
80
100
120
140
Play
泛藍高忠誠區
泛綠高忠誠區
普通區
該村里性比例
YEAR 2010
2010
2011
2012
2013
2014
2015
2016
2017
2018


該村里人口密度
40k
20k
0
400
600
800
1000
Play
泛藍高忠誠區
泛綠高忠誠區
普通區
該村里所得中位數
YEAR 2010
2010
2011
2012
2013
2014
2015
2016


該村里人口密度
40k
20k
0
0
100
200
300
400
Play
泛藍高忠誠區
泛綠高忠誠區
普通區
該村里老化指數
YEAR 2010
2010
2011
2012
2013
2014
2015
2016
2017
2018

該村里人口密度
40k
20k
0
0.2
0.4
0.6
0.8
Play
泛藍高忠誠區
泛綠高忠誠區
普通區
該村里教育程度（大專以上畢業比例）
YEAR 2010
2010
2011
2012
2013
2014
2015
2016
2017
2018