Mapping voter preference
Spatial clusters in Taiwan’s 2014 local elections

In November 2014, Taiwan’s 23 million people elected a range of local leaders: governors, mayors, and councilors. These elections were a victory for the Democratic Progressive Party (DPP), historically the opposition party in Taiwanese politics. If this trend continues, the DPP may win the 2016 presidential and legislative elections. Because the DPP is pro-independence, its election would upset relations in the Taiwan Strait.

This project examines voter turnout for the DPP, considering whether, where, and why any statistically unusual clusters of DPP support exist in Taiwan.

Methodology
I worked with districts (admin level 3), the lowest level for which I could find data. I built a spreadsheet using election results from the Taiwan Electoral Database (選舉資料庫網點), and joined it to a shapefile provided by the Taiwanese Data Sharing Platform (台灣資料開放平台). While nine elections took place in 2014, I studied only the elections for county councilors (like a U.S. state legislature). Since a relatively large slate of candidates were elected (220), those races were more suitable to study patterns of party support than the smaller mayoral elections. I studied only the four counties shown here due to time constraints.

To find clusters, I used GeoDA, an open source program from Arizona State University. I calculated the "local Moran’s I" for each district, a test of spatial association (Anselin, 1995). Briefly, this test determined whether the level of DPP support (measured by % of votes for DPP candidates) in each district was more similar to surrounding districts than would be expected by chance. After computing 9,999 permutations (a robust setting), the test identified three significant areas: a cluster of districts with high support for the DPP, another cluster with low support, and an outlier—a district with high support surrounded by low support.

I tested the impact of demographic factors on support for the DPP.

Methodology
I collected demographic data from the 2010 Taiwanese Census, and used Stata to run an "ordinary least squares" (OLS) regression. Since the Breusch-Pagan test for heteroskedasticity returned a high value of 20.9, I used robust standard errors. I considered using a spatial regression model to account for spatial auto-correlation, but the Lagrange multipliers for both lag and error were low enough (2.04 and 1.78) that OLS is the preferred model (Anselin, 1996).

The variable "Use of Taiwanese" is a relative measure, in which the percent population who speak Mandarin is subtracted from the percent population who speak "Taiwanese" (Hokkio). Results
The results of the regression are shown to the right. Most variables had no statistically significant impact. The DPP is sometimes considered an ethnic Taiwanese party, and was thought to have benefited from young voter support in the most recent election. However, this regression does not support those interpretations.

Many factors were not captured by this model. Citizens may have cast their votes due to preference for the DPP position on certain issues, but I had no data on these preferences. I also could not show the influence of patronage networks and graft (vote buying is not unusual in Taiwan).

However, one variable was significant: the percentage of candidates in each race who belonged to the DPP. This variable had a strong, statistically significant, positive correlation to the percentage of votes the DPP received in a district. This is something of a tautology; it’s reasonable to assume that if there are more DPP candidates to vote for, then the DPP as a whole will get more votes. More significantly, the uneven levels of DPP candidacy across districts may indicate that the party is choosing its targets—that it has prioritized some districts. The districts with many DPP candidates might also receive more party resources, like advertising funds and visits from party officials. Finally, since citizens signatures must be secured to put a candidate on the ballot in the first place, a high ratio of DPP candidates might show strong support in that district.

Visualizing trends
The two small maps on the right show a) the DPP’s share of candidates and b) the DPP’s share of votes. The larger map combines these two, showing where the trends do and do not correlate.

Conclusion
The grouping and fracturing of political support, like the ebb and flow of ideas which it represents, is a complex problem. In this case, geography played a smaller role than did numerous human factors. These factors are generally hard to measure, but GIS provides some tools for their depiction, aiding our understanding in a small but significant way.

If not geography, then what?
I tested the impact of demographic factors on support for the DPP.

Methodology
I collected demographic data from the 2010 Taiwanese Census, and used Stata to run an "ordinary least squares" (OLS) regression. Since the Breusch-Pagan test for heteroskedasticity returned a high value of 20.9, I used robust standard errors. I considered using a spatial regression model to account for spatial auto-correlation, but the Lagrange multipliers for both lag and error were low enough (2.04 and 1.78) that OLS is the preferred model (Anselin, 1996).

The variable "Use of Taiwanese" is a relative measure, in which the percent population who speak Mandarin is subtracted from the percent population who speak "Taiwanese" (Hokkio). Results
The results of the regression are shown to the right. Most variables had no statistically significant impact. The DPP is sometimes considered an ethnic Taiwanese party, and was thought to have benefited from young voter support in the most recent election. However, this regression does not support those interpretations.

Many factors were not captured by this model. Citizens may have cast their votes due to preference for the DPP position on certain issues, but I had no data on these preferences. I also could not show the influence of patronage networks and graft (vote buying is not unusual in Taiwan).

However, one variable was significant: the percentage of candidates in each race who belonged to the DPP. This variable had a strong, statistically significant, positive correlation to the percentage of votes the DPP received in a district. This is something of a tautology; it’s reasonable to assume that if there are more DPP candidates to vote for, then the DPP as a whole will get more votes. More significantly, the uneven levels of DPP candidacy across districts may indicate that the party is choosing its targets—that it has prioritized some districts. The districts with many DPP candidates might also receive more party resources, like advertising funds and visits from party officials. Finally, since citizens signatures must be secured to put a candidate on the ballot in the first place, a high ratio of DPP candidates might show strong support in that district.

Visualizing trends
The two small maps on the right show a) the DPP’s share of candidates and b) the DPP’s share of votes. The larger map combines these two, showing where the trends do and do not correlate.

Conclusion
The grouping and fracturing of political support, like the ebb and flow of ideas which it represents, is a complex problem. In this case, geography played a smaller role than did numerous human factors. These factors are generally hard to measure, but GIS provides some tools for their depiction, aiding our understanding in a small but significant way.

If not geography, then what?
I tested the impact of demographic factors on support for the DPP.

Methodology
I collected demographic data from the 2010 Taiwanese Census, and used Stata to run an "ordinary least squares" (OLS) regression. Since the Breusch-Pagan test for heteroskedasticity returned a high value of 20.9, I used robust standard errors. I considered using a spatial regression model to account for spatial auto-correlation, but the Lagrange multipliers for both lag and error were low enough (2.04 and 1.78) that OLS is the preferred model (Anselin, 1996).

The variable "Use of Taiwanese" is a relative measure, in which the percent population who speak Mandarin is subtracted from the percent population who speak "Taiwanese" (Hokkio). Results
The results of the regression are shown to the right. Most variables had no statistically significant impact. The DPP is sometimes considered an ethnic Taiwanese party, and was thought to have benefited from young voter support in the most recent election. However, this regression does not support those interpretations.

Many factors were not captured by this model. Citizens may have cast their votes due to preference for the DPP position on certain issues, but I had no data on these preferences. I also could not show the influence of patronage networks and graft (vote buying is not unusual in Taiwan).

However, one variable was significant: the percentage of candidates in each race who belonged to the DPP. This variable had a strong, statistically significant, positive correlation to the percentage of votes the DPP received in a district. This is something of a tautology; it’s reasonable to assume that if there are more DPP candidates to vote for, then the DPP as a whole will get more votes. More significantly, the uneven levels of DPP candidacy across districts may indicate that the party is choosing its targets—that it has prioritized some districts. The districts with many DPP candidates might also receive more party resources, like advertising funds and visits from party officials. Finally, since citizens signatures must be secured to put a candidate on the ballot in the first place, a high ratio of DPP candidates might show strong support in that district.

Visualizing trends
The two small maps on the right show a) the DPP’s share of candidates and b) the DPP’s share of votes. The larger map combines these two, showing where the trends do and do not correlate.

Conclusion
The grouping and fracturing of political support, like the ebb and flow of ideas which it represents, is a complex problem. In this case, geography played a smaller role than did numerous human factors. These factors are generally hard to measure, but GIS provides some tools for their depiction, aiding our understanding in a small but significant way.