Identifying Communities of Interest in Yolo County Using Ballot Voting Patterns

A Cluster Analysis-Based Approach

Bapu Vaitla

10/13/2021

Contents

Note: this document is a summary of the analysis. The complete HTML document, which includes scripts to fully reproduce the results, is in progress. Please write to <bapu.vaitla@gmail.com> if you would like the code immediately.

1. Introduction

As a complement to public testimony, this paper performs a quantitative analysis of communities of interest (COIs) in Yolo County. I use cluster analysis techniques on ballot voting data over the last five general elections (2012 through 2020) to identify communities of political interest.

The strengths of this approach:

- Voting data is likely to capture a wide range of social and economic interests. When voters make a ballot measure choice, they are implicitly weighing trade-offs across their various interests and identities.
- While the ARC is prohibited from considering party registration or the results of partisan elections, the use of ballot voting data is allowable. Ballot voting data offers a look at the political views of a wide range of Yolo County residents, not only the small minority that submit public comments.
- Ballot voting data may even be preferable to partisan registration or vote data as a source of information about political interests. A sufficiently large and diverse set of ballot measures—which indeed has been the norm in recent California general elections—gives insight into political values in multiple dimensions, not simply a single conservative-progressive dimension.

The weaknesses of this approach:

- Voter turnout, while relatively high in recent elections, is far from universal (66% of eligible Yolo County residents voted in the 2020 election). The interests of non-voters are thus not captured in the analysis, and these may be the residents of Yolo County who are most socially and economically vulnerable.
- No guidance for quantitative COI analysis exists in California or Yolo County election law, nor is there a social scientific consensus on the optimal way to carry out such analysis. In addition, there are thus far no well-defined methods for weighing or combining community testimony and voting-based quantitative data.
- While the last five general elections have featured ballot measures spanning many policy arenas, these measures are not exhaustive of all issues important to communities. Furthermore, partly because of the increasing cost of California ballot campaigns, ballot measures may reflect interest group influence more than community values.

2. Data

I use Yolo County election and geographic data from the last five general elections (2012 through 2020) obtained from the Statewide Database. Specifically, I use consolidated precinct (SRPREC)[1] ballot measure Statement of Votes (SOV) data and associated geographic files.

[1]Consolidated precincts are created by the Statewide Database to reconcile registration and voting precincts. These precinct boundaries change from election to election, as do the total number of precincts.

The scripts in the full HTML report provide code for downloading and formatting this data. Descriptions of the ballot measures for each election are accessible by clicking on the relevant year: 2012, 2014, 2016, 2018, 2020.

3. Methods

Cluster analysis divides observations—in this case, the consolidated precincts of Yolo County—into homogeneous and distinct groups. "Homogeneous" implies minimizing *within*-cluster variance in ballot measure voting patterns. "Distinct" implies maximizing *across*-cluster variance in ballot measure voting patterns. The result is to create clusters of similar voters within the County. A model-based clustering algorithm is used.[2]

Note that **the cluster analysis is intended to inform the map-drawing process, not to serve as a method for drawing the maps themselves.** I explicitly do not constrain the results to follow the redistricting criteria of contiguity, equal population size, or the target number of voting districts. Satisfying these criteria would require arbitrary decisions that considerably weaken the clustering results. In this paper, I wish to present an initial look at the clustering of voting patterns before such trade-offs are imposed.[3]

4. Results

Election-by-election results are presented and discussed below. Precincts with the same color belong to the same voting cluster.[4] The number of clusters varies by election year, reflecting the underlying ballot voting data. The plots next to each map show the distance between clusters, with the clusters represented as ellipses encircling the gray precinct markers. Less distance between the ellipses on the graph indicates greater political similarity.

Note that the x-axis and y-axis on the cluster plots represent aggregate measures of the various initiatives on the ballot in a given year. However, this paper does not explicitly link these axes to specific sets of political values; that is, a "higher" score on either the x-axis or y-axis does not necessarily imply a political orientation.[5] Identifying the political values represented by each cluster could easily be done by looking at how each axis relates to each ballot measure. However, identifying what each cluster represents politically may obscure the more fundamental fact of their clustering, and thus is not done in this paper.

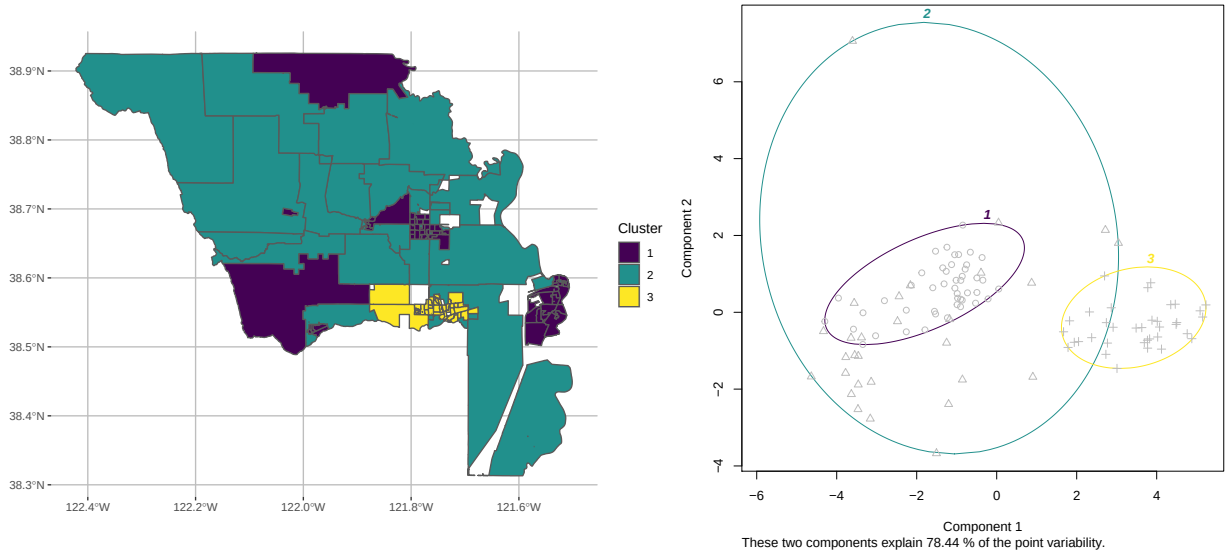
[2] I implement this algorithm in R's *mclust* package. Estimation is performed using an expectation maximization algorithm initialized by hierarchical model-based agglomerative clustering, and based on parameterized finite Gaussian mixture models. Optimal models are selected using the Bayesian information criterion.

[3] In addition, applying the contiguity and equal population size criteria demands a computationally intensive search of the enormous space of possible maps. This is beyond the scope of the present analysis.

[4] Uncolored precincts in the maps lack ballot voting data for that year.

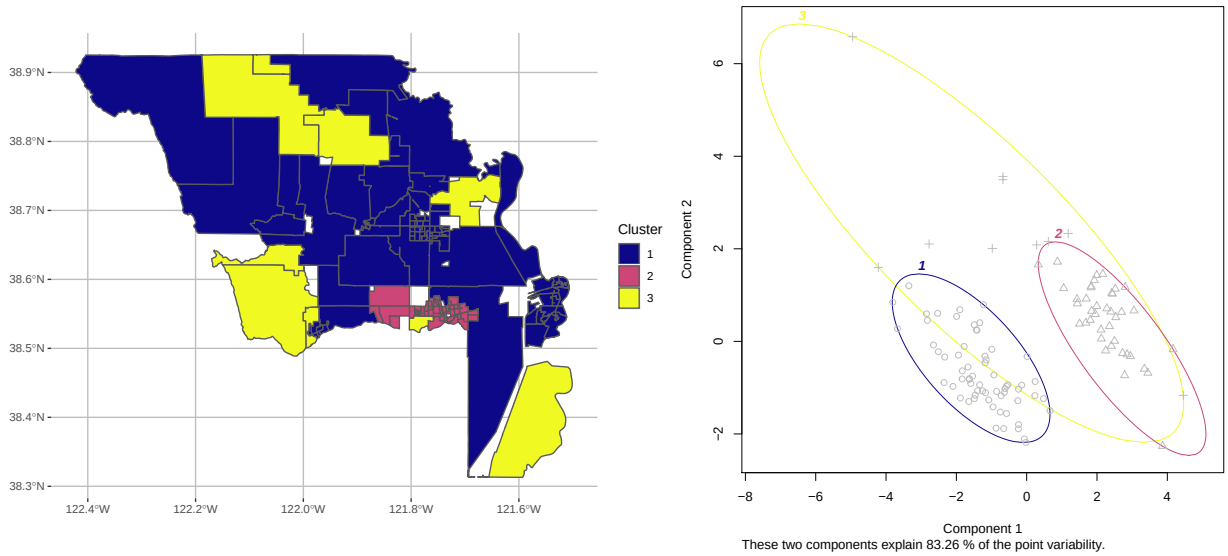
[5] These cluster plots map each precinct (shown by gray markers) in a two-dimensional principal component space. The 2D space is a reduction of the multidimensional ballot measure space for each election. As the statements below the cluster plots indicate, around 80% of variance in ballot measure voting is explained by these two components.

2012 general election



Three clusters are obtained from the 2012 voting data. Cluster 1 (dark purple) includes most of Woodland, West Sacramento, and Winters, as well as a few rural precincts. Cluster 2 (teal) includes Davis and areas to the immediate west. Cluster 3 (yellow) encompasses most of the rural areas of the County. The clusterplot shows that clusters 1 and 3 are quite distinct, and cluster 2 overlaps to a greater extent with cluster 1 than cluster 3. Overall, the 2012 map suggests that Davis is a distinct voting cluster from the rest of the County.

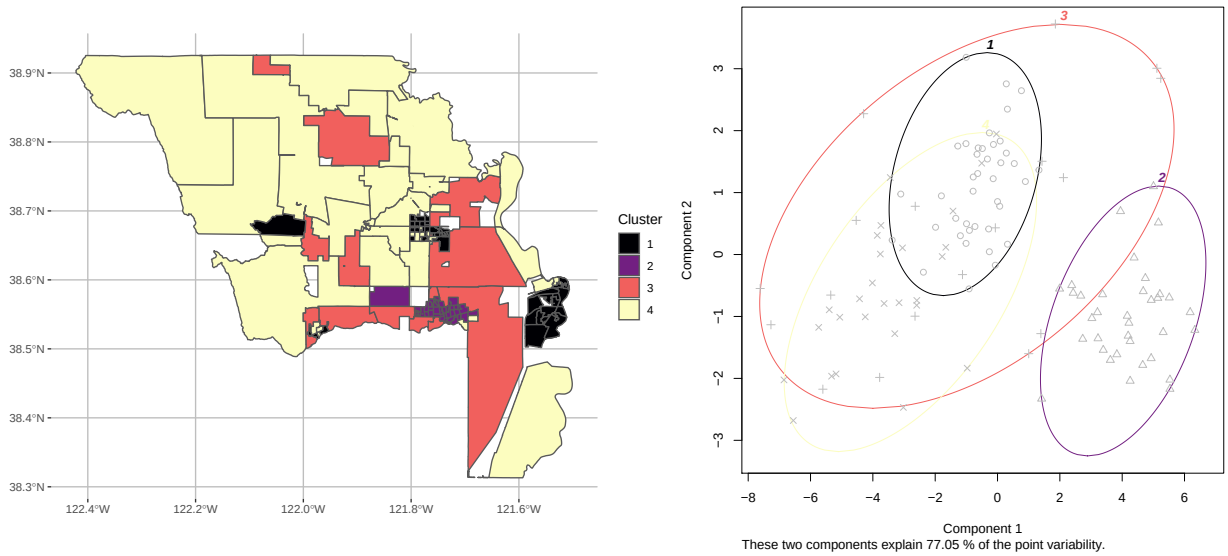
2014 general election



The 2014 voting data gives similar results. (Note that the colors have been intentionally altered from map to map to make clear that a cluster on one map should not be taken to represent the same "voting values type" as another year's map). Woodland, West Sacramento, and Winters are all in Cluster 1 (dark blue). Davis and areas west, represented by Cluster 2

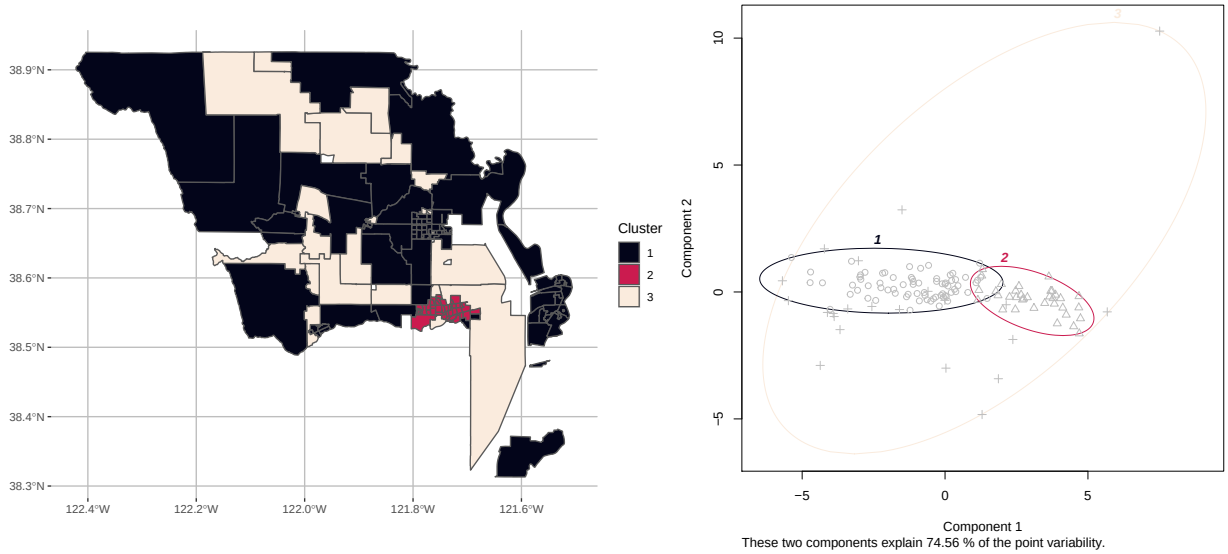
(pink), are still distinct. However, Clarksburg and the surrounding rural areas, as well as north-central parts of the County, are now in their own voting cluster (3; yellow). In this map, the two urban clusters are non-overlapping in the cluster plot, while the rural cluster spans both of the other two clusters.

2016 general election



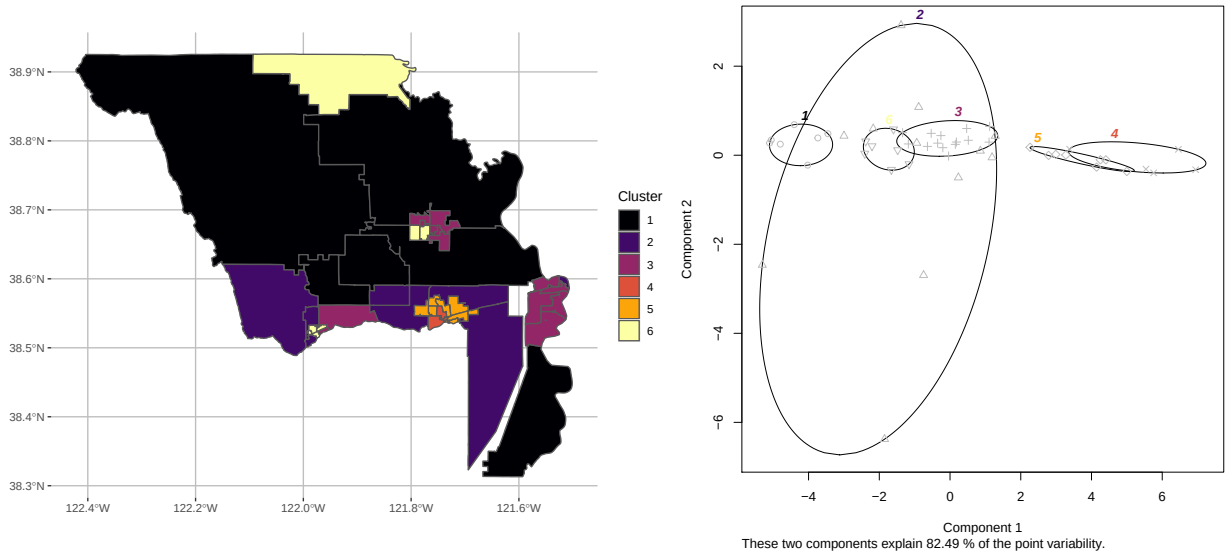
The 2016 data suggests the presence of four voting clusters. Again, most of Woodland and West Sacramento are in the same cluster (1; black); Winters (as well as northern parts of West Sacramento), however, is divided between cluster 1 and the largely rural cluster 4 (light yellow). Davis is once again in a single cluster (2; violet). The most pronounced difference in this map from previous years is the presence of two distinct rural clusters (3, salmon, and 4, light yellow). The cluster plot suggests, however, that the two rural clusters are largely overlapping. The Davis cluster overlaps slightly with rural cluster 3—perhaps reflecting geographic proximity—but not with clusters 1 or 4.

2018 general election



The 2018 map returns to a three-cluster grouping. Cluster 1 (black) includes all of Woodland, West Sacramento, Winters, and many rural areas; cluster 2 (red) nearly all of Davis, and cluster 3 (eggshell) the remainder of the rural areas. The cluster plot suggests slightly more overlap than in previous elections. The presence of the rural cluster 3 data points (indicated with pluses) show greater overlap with cluster 1 than the Davis cluster 3.

2020 general election



The 2020 data deviates somewhat from the pattern of the previous four general elections. Note first that the total number of consolidated precincts was almost halved in this election, as reflected by the larger precinct shapes. Cluster 1 (black) is almost entirely rural, with the exception of the northeast corner of West Sacramento (the Lighthouse neighborhood). In the previous maps, rural areas were either divided into two clusters (2016, 2018)—which may

partly be an artifact of the greater number of clusters in these years relative to 2020—or the dominant rural cluster included most of Woodland (2012, 2014). While Cluster 2 (dark purple) of the 2020 map also appears to be a largely rural district in the southern part of the County, the total votes are dominated by the precincts immediately surrounding Davis. Cluster 3 (plum) includes nearly all of West Sacramento, much of Woodland, and some rural areas around Winters. Cluster 4 (dark orange) is Downtown Davis and the precinct covered by the University of California. The 2020 map is also the only map that divides Davis; cluster 5 (light orange) represents the rest of the city, with this cluster politically closer to the Davis-peripheral cluster 3. We see, however, that clusters 4 and 5 are closely aligned. Cluster 6 (light yellow) encompasses nearly all of Winters and the southwest quadrant of Woodland. The Davis clusters are most distant from the rural cluster 1. The Winters/southwest Woodland cluster 6 and West Sacramento/Woodland cluster 3 are closely aligned.

5. Discussion

Taken together, the maps offer several messages.

First, the 2020 election offered a considerably more complex picture than previous elections. There are two reasons why weighting the 2020 election data more heavily in the current redistricting process might be desirable. One, recent elections are more likely to reflect more accurate present-day political communities. Two, although turnout was far from universal, the 2020 election featured extraordinarily high numbers of votes, with approximately 66% of those eligible in Yolo County casting ballots, about nine percentage points higher than in the 2016 presidential election.

Second, the 2020 map suggests that the Woodland/West Sacramento/Winters political community may be becoming more heterogeneous, with West Sacramento and the eastern/northeastern parts of Woodland as one cluster and Winters/southwest Woodland as another.

Third, nearly all of West Sacramento is in the same cluster in every map. The exceptions are the Riverside/CHP and Riverpoint neighborhoods in 2012; those neighborhoods and part of Broderick/Bryte in 2014; and the the Lighthouse neighborhood in 2020.

Fourth, Davis is clearly a distinct political community. Dividing the city into two clusters—as the equal population requirement necessitates—will inevitably split a COI. The 2020 map suggests that the best way to divide the city may be by creating a "core," student-dominated district (although, following the Board of Supervisors' guidance, this district must include some unincorporated areas) and a "peripheral" district that encompasses the unincorporated areas surrounding Davis.

Fifth, the various maps give different messages on the question of one versus two rural clusters. The rural areas are presented as relatively homogeneous in two maps (2012 and 2020) and more heterogeneous in the three others (2014, 2016, 2018) that contain at least two rural voting clusters.