No Adults Allowed! Unsupervised Learning Applied to Gerrymandered School Districts

Divya Siddarth and Amber Thomas

Motivation

With more than 80% of public school students attending the school assigned to them by district, it is clear that school district boundaries play a critical role in determining the educational opportunities and resources provided to these students. Unfortunately, school district lines are constantly redrawn, and historical and anecdotal evidence suggests that these boundaries are often artificially manipulated, or **gerrymandered**, into irregular shapes (1). Often, these manipulations directly result in deliberately exclusionary zoning processes that create **artificial economic disparity** or **racial segregation** between school districts. Further, while congressional gerrymandering (in which county boundaries are drawn to benefit a particular political party over another) has been a much studied phenomenon, school district gerrymandering is far less well-documented. However, categorizing and identifying these districts, and potential demographic and economic markers of school district gerrymandering, is an important first step towards **equalizing funding and other resources** across districts instead of having pockets of underserved students.

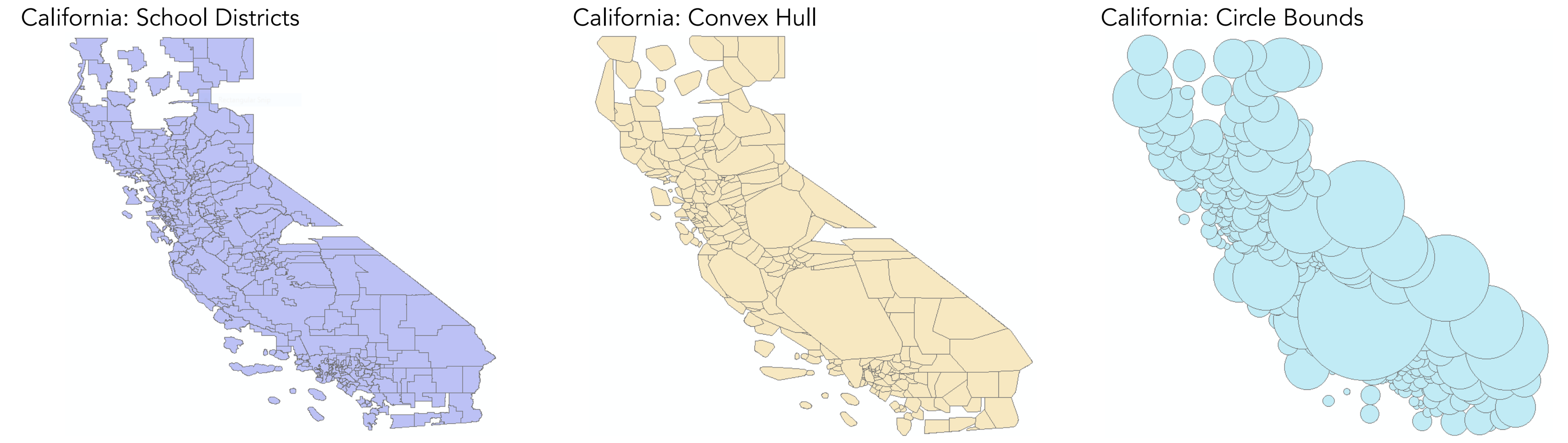
Data

The dataset used to capture school district boundaries was provided by the Department of Education's TIGER/Line database (2), and was formatted as shapefiles, a common industry standard for representing spatial data in points, lines and polygons. Relevant raw features included geographic id and name, all other features were calculated using ArcGIS software. Census datasets were used for demographic data, divided along census tract lines, and specific raw demographic features included the percentage of population below poverty, percentage racial breakdowns, and mean household income. Measures of polygon irregularity were calculated from the raw shapefile data.

Features

Feature vertices were generated from the boundary data of school districts by rasterizing the boundary shapes, and area and perimeter were calculated. To create a measure of irregularities in shape, minimum bounding geometries were then layered on top of the original school districts – both convex hulls and circle bounds were generated for each district.

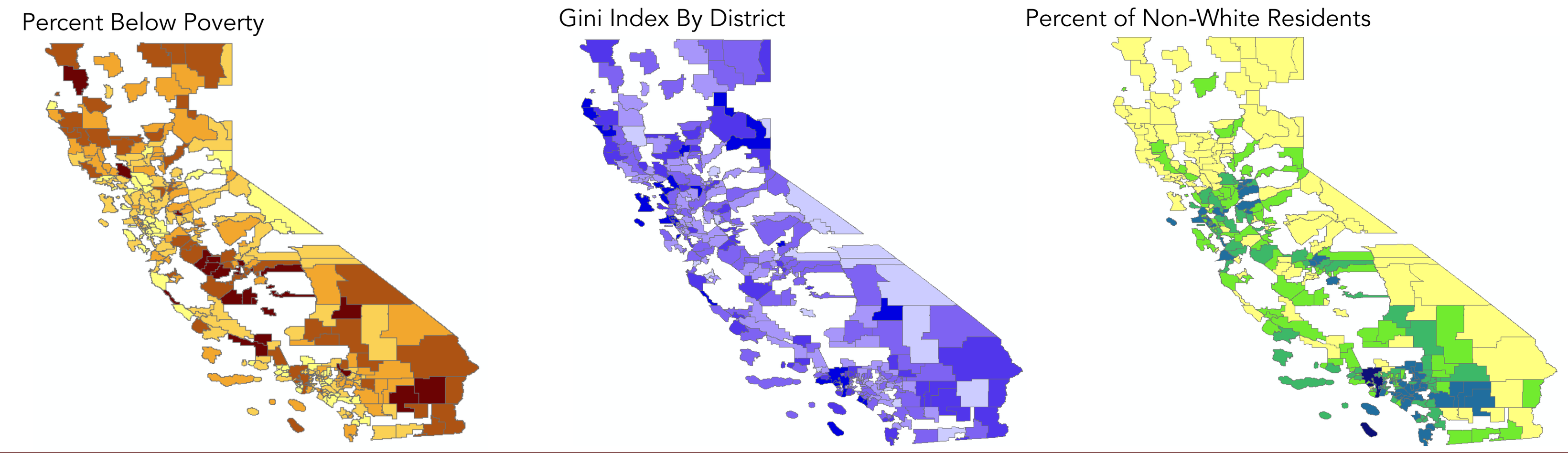
California: School Districts — California: Convex Hull — California: Circle Bounds



Features, contd.

Demographic data was combined with geographic data. The geospatial coordinates of the data were projected onto a flattened plane, then the two datasets were overlaid in ArcMAP, and the GEO_id of the census tract was spatially correlated with its school district in which it fell.

Percent Below Poverty — Gini Index By District — Percent of Non-White Residents



Models

Clustering Analysis:
We used Ward's method of hierarchical clustering on a small sample of the data to determine the number of clusters needed. We then proceeded with the more efficient k-means algorithm, creating 9 clusters, measuring and minimizing the Euclidean distance between our 15-feature vector and centroids.

Polygon Irregularity Measurements:
Indentation was measured by the Schwartzberg index and the Polsby-Popper index, which were used to compare perimeter and area of the school district to perimeter and area of the corresponding minimum bound circle. Dispersion was measured as a function of the Reock index and the Convex Hull index, which were used to compare the area and perimeter of the school district to the corresponding minimum bound convex hull.

Results

We performed k-means clustering on our data; the number of iterations totaled 143 to meet a convergence factor of 0.0001. Cluster centers are displayed below. We notice qualitative correlations between, particularly, mean income of a cluster and the polygon irregularity indices, with higher mean incomes corresponding to less irregular polygons.

	1	2	3	4	5	6	7	8	9
Schwartzberg	0.429	0.415	0.399	0.410	0.420	0.443	0.439	0.280	0.280
Polsby	0.656	0.640	0.622	0.635	0.646	0.670	0.666	0.475	0.609
Reock	0.549	0.553	0.550	0.549	0.545	0.543	0.540	0.531	0.541
ConvexHull	0.211	0.211	0.203	0.209	0.206	0.213	0.210	0.126	0.186
% White	88.8	91.1	73.3	84.5	86.3	90.0	90.4	80.4	82.9
% Black	3.1	1.3	16.1	3.7	5.8	3.6	2.8	1.0	2.9
% Asian	2.6	0.5	0.6	6.1	0.7	0.8	1.4	12.7	9.5
% Other	5.5	7.1	10	5.7	7.2	5.6	5.4	5.9	4.7
Mean Income	8.5E+04	8.7E+04	4.4E+04	1.1E+05	5.3E+04	6.2E+04	7.2E+04	2.6E+05	1.5E+05
Std. Dev. Income	6262	28548	4554	7455	4917	5677	6344	23596	9241
% Below Poverty	8	12	27	6	18	13	10	3	5

Discussion

In this project, we look at the practice of school district gerrymandering, focusing on the use of geospatial and clustering techniques to examine boundary anomalies. In fact, the present algorithm evolved from several failed clustering attempts, the first of which involved generating vertices for along the district boundary, creating a function of distance from center with respect to angle, and using this measure for clustering. We eventually used an updated algorithm that involved the minimum bounding geometry described. We found that many of our cluster centers were quite close together, excluding clusters eight and nine, which were more dissimilar to the other cluster centers. However, our preliminary analysis of polygon irregularities yielded troubling results regarding the consistent use of unintuitive school district bounds, which we would like to explore further. Further, throughout the course of this project, we noticed a severe lack of information regarding school district gerrymandering, and particularly a lack of previous computational work on the topic, and this is a lack we hope to fill with our future work.

Future

Our next step will be to create a model from our cluster analysis that can predict whether or not school districts, and more intriguingly, school attendance zones (smaller zones within districts), were subject to gerrymandering. Perhaps inclusion of features such as dependence on government assistance, or type of district (urban vs rural), would be included. It would also be interesting to examine how the existence of gerrymandering in particular districts, predicted by our model, relates to those school district's budgets and earmarks from federal and state government. From that point, we could delve more deeply into resource distribution at the governmental level as it relates to gerrymandering.

References

1. Wright, J. Skelly. "Public School Desegregation: Legal Remedies for De Facto Segregation." New York University Law Review 40.2 (1965): 285-310.
2. *National Center for Education Statistics (2015). Education Geographic and Demographic Information: School District Boundaries. <https://nces.ed.gov/programs/edge/geographicDistrictBoundary.aspx>*
3. Richards, Meredith and Kori Stroub. "An Accident of Geography? Assessing the Gerrymandering of School Attendance Zones." Teachers College Record 117.7 (2015): p 1-32.