Optimizing Opportunity: An Algorithmic Approach to Redistricting for Fairer School Funding

Jordan Abbott
New America Foundation, Education Funding Equity Initiative
abbott@newamerica.org

Abstract

School district boundaries in the United States are not just lines on a map; they are mechanisms that perpetuate deep-seated educational inequities by directly linking school funding to local property wealth. We present a computational framework for optimizing district boundaries to improve resource equity while reducing racial and economic segregation. This study designs a novel two-stage algorithmic process that uses maximally compact plan initialization from spatial clustering and iterative refinement using Markov Chain Monte Carlo (MCMC) optimization. This hybrid approach can reduce required iterations by more than 90%, relative to traditional methods, and allows systematic variation of different numbers of districts. Hard constraints including contiguity, minimum enrollment thresholds, and infrastructure capacity limits are still enforced. Optimization targets three Theil indices measuring property tax capacity disparities, racial segregation, and economic segregation. Across 42 states, results show average state-level improvements of 66.6% in tax-base equality, 47.6% reduction in racial segregation, and 65.0% decrease in economic segregation.

1 Introduction

Public school district borders determine both educational access and taxing jurisdictions, directly impacting resources available to students. Because a significant portion of school district funding is derived from local property taxes, these boundaries create systematic disparities that correlate with racial and economic segregation. While states compensate through progressive funding formulas, high-wealth districts can more easily raise additional local revenue (Kenyon and Munteanu, 2021). As a result, these school systems can easily increase their budgets beyond what lower-wealth districts can match. Nationwide, property taxes reportedly comprised 65% of local revenues in 2021, but can be much higher in some states (Common Core of Data (CCD), n.d). Given the connection between property values and neighborhood affordability, the students that lose out tend to be those from low-income backgrounds.

The problem also has a troubling racial dimension. This funding system is layered on top of generations of policies and government practices that have created and entrenched racial and economic segregation in housing markets (Kuhn et al., 2018). At different phases of America's past and present, this has included redlining and racially discriminatory mortgage lending; court enforcement of racially restrictive covenants; government-funded construction of segregated housing developments; exclusionary zoning policies; and unfair property assessment, among other forms of discrimination. These factors have shaped both the racial composition of neighborhoods and the property values in the taxing jurisdictions from which school districts raise local dollars.

The result is a map of highly segregated residential communities that demonstrate stark economic divides (Reardon and Weathers, 2024). Left unmitigated, the legacies of discriminatory policies have the potential to intersect to create, shape, and enforce new patterns of segregation (Reardon and Owens, 2014). This problem is further reified by the ways in which school district borders function: both as geographic areas that are home to district students, and as the taxing jurisdictions that yield their local funding (Stadler and Abbott, 2024). Because these boundaries determine the students served by a district and the local funding available for its schools, they function to separate students from resources—and from each other.

Furthermore, students from low-income backgrounds, and with other needs and challenges have demonstrably higher funding needs than students from high-income families (Jackson et al., 2015; Jackson and Mackevicius, 2024). While state aid

Proceedings of the Artificial Intelligence in Measurement and Education Conference (AIME-Con) – Volume 2: Works in Progress, pages 25–33

can and should be used as a tool to provide additional support to students with higher needs, this goal is undermined when these funds are eaten up in pursuit of achieving basic funding parity with high wealth districts. This burdens state education budgets with compensating for existing inequity, rather than achieving equity (Gartner, 2023). Given the high degree of alignment between segregated school districts, patterns of residential segregation in the communities they serve, and funding divides, one option is to consider drawing better school district borders.

2 Theoretical Development

This paper presents the first national-scale computational framework for school redistricting by adapting methods from legislative redistricting. We draw on MCMC methods used to efficiently explore high-dimensional solution spaces for boundary realignment (McCartan and Imai, 2023; Fifield et al., 2014), but introduce additional constraints not often considered in legislative contexts, including property tax capacity, infrastructure limitations, and multi-dimensional segregation measures.

Our work builds upon a small but growing body of literature that has explored educational boundary optimization from different but complementary perspectives. The framework designed by Gillani (2023) found that intradistrict segregation could be reduced while maintaining travel times. However, this approach is not designed to address the interdistrict dynamics where nearly two-thirds of all racial segregation occurs (Owens, 2016). Simko (2024) advanced this line of inquiry through a detailed case study of New Jersey, highlighting the importance of crossing district lines to integrate school systems. These valuable analyses centered on demographic integration, with logistical constraints like student capacity and travel times, while keeping the number of districts constant. Our research extends this conversation by shifting both the geographic scale and the central objective. We also expand the solution space, exploring configurations that vary district counts from 25% to 175% of current levels, in doing so, significantly expanding potential equity gains.[1]

This study adopts fiscal equity as a central optimization goal, a dimension not observed in prior demographic-focused studies. The primary contribution of this approach is the integration of parcel-level property tax assessment data, which allows us to directly model the tax base of each potential district. To complement this, we also incorporate Small Area Income and Poverty Estimates (SAIPE), creating a multi-dimensional economic profile of each proposed district. By constructing boundaries around both fiscal equity and demographic balance, our model is designed to create districts with equitable and sustainable local revenue capacity, a vital consideration given that property taxes constitute, on average, 40% of all district funding (CCD n.d.).

Although changing school district boundaries can be politically challenging, policymakers may be unaware of the extent of existing divides or the degree to which they can be mitigated. Further, there is compelling evidence of efficiency savings from consolidating districts (Duncombe and Yinger, 2007; Dodson and Garrett, 2008). In light of this, several states, including Arkansas, Pennsylvania, and New Jersey have recently undertaken efforts to examine the feasibility of district mergers and other boundary changes. This algorithmic approach provides an objective framework to supply legislators with evidence of the potential benefits of redistricting, including fiscal savings, deconcentrated poverty, and integrated school systems.

3 Problem Formulation

Optimization of school district boundaries represents a high-dimensional combinatorial problem where geographic units are assigned to districts while satisfying multiple objectives and constraints. Unlike legislative redistricting, school redistricting must simultaneously consider property tax capacity, demographic integration, and infrastructure capacity. This section formalizes the mathematical framework underlying our optimization approach.

3.1 Multi-Objective Optimization Framework

Census tracts serve as atomic geographic units that must be assigned to districts. Let $G = \{1, 2, ..., n\}$ represent the set of tracts in a state, and let $D = \{1, 2, ..., k\}$ represent the set of districts, where k varies systematically. Each tract $i \in G$ must be assigned to exactly one district $d \in D$, creating a partition of the geographic space.

Optimization seeks to minimize an objective

[1] For some, the immediate reaction may be that such changes are unrealistic. However, our core approach posits that considering drastic changes to district counts is necessary to unlock the full potential for equity gains.

function combining three equity dimensions:

$$f(D) = w_1 T_{\text{val}}(D) + w_2 T_{\text{racial}}(D) + w_3 T_{\text{econ}}(D) \tag{1}$$

where T_{val} measures disparities in per-pupil property tax capacity, T_{racial} captures multigroup racial segregation, and T_{econ} quantifies economic segregation based on binary poverty status. Default weights are set to $w_1 = 3.0$, $w_2 = 1.0$, and $w_3 = 1.0$, reflecting a priority on addressing disparities in property tax capacity while maintaining focus on integration objectives.

3.2 Mathematical Formulation of Theil Indices

3.2.1 Theil's T for Disparities in Property Tax Capacity

The Theil T-index captures inequality in assessed property values per pupil across districts:

$$T_{\text{funding}} = \sum_i \left(p_i \cdot \frac{x_i}{\mu} \cdot \log\left(\frac{x_i}{\mu}\right)\right) \tag{2}$$

where:

- $p_i = n_i/N$, the proportion of total students in district i
- n_i = number of children in district i
- N = total children in the state
- x_i = assessed property value per pupil in district i
- μ = state mean assessed value per pupil

This formula directly measures the capacity to raise local revenue, as property assessments form the tax base for school funding. The index equals zero when all districts have identical per-pupil property values, and increases with greater inequality.

3.2.2 Multigroup Theil's H for Racial Segregation

For racial integration, we employ the multigroup entropy-based Theil H-index:

$$T_{\text{racial}} = (E_{\text{state}} - E_{\text{weighted}})/E_{\text{state}} \tag{3}$$

where:

- $E_{\text{state}} = -\sum_r(\pi_r \cdot \log(\pi_r))$, the entropy of racial composition at state level
- π_r = proportion of racial group r in total state enrollment
- $E_{\text{weighted}} = \sum_i(p_i \cdot E_i)$, the enrollment-weighted average of district entropies
- $E_i = -\sum_r(\pi_{ir} \cdot \log(\pi_{ir}))$, the entropy within district i
- π_{ir} = proportion of group r in district i

The index ranges from 0 (perfect integration, where every district mirrors state demographics) to 1 (complete segregation). This multigroup formulation avoids the limitations of binary indices and captures the full complexity of racial composition.

3.2.3 Binary Theil's H for Economic Segregation

Economic segregation uses a similar entropy-based approach with two groups:

$$T_{\text{economic}} = (E_{\text{state}} - E_{\text{weighted}})/E_{\text{state}} \tag{4}$$

Applied to binary poverty status, as defined by SAIPE, this measure captures the concentration of economic disadvantage across districts. The binary formulation is appropriate, given the policy relevance of poverty thresholds for federal program eligibility.

3.3 Constraint Specifications

1. **Geographic Contiguity:** Each district d must form a connected component under rook adjacency (shared edges, not just vertices).

2. **Minimum Population Threshold:** Each district must contain at least m children, where $m = 0.5 \times \min$ (current district resident school-aged population in each state).

3. **Infrastructure Capacity:** For each district d: $\sum$(children in tracts assigned to d) $\leq 1.25 \times$ $\sum$(capacity d).

4 Data Architecture

4.1 Geographic Foundation

Census tracts serve as the geographic units for our simulation approach, providing a standardized nationwide framework with sufficient granularity to capture local-level variation. We use 2020 census tract boundaries from the Census Bureau, approximately 80,000 tracts. These polygons define our building blocks for spatial optimization.

Tract adjacency relationships are established using rook contiguity, including only shared tract boundaries rather than vertices. The resulting adjacency matrix forms the foundation for contiguity constraint checking and move generation during optimization. Disconnected components (e.g., islands, water boundaries) are connected via minimum distance stitching between nearest tract centroids to ensure graph connectivity.

4.2 Property Valuation Data

Property valuation data is provided by the Center for Geospatial Solutions at the Lincoln Institute of Land Policy, pre-aggregated at the relevant geographic units of analysis. This dataset provides total assessed property values for each unit from the most recent year available, which directly determine local education revenue capacity. Unlike market values or sale prices, assessed values reflect the actual tax base available to districts.[2]

4.3 Demographic Data

4.3.1 Demographic Composition

Demographic data come from two primary sources. Racial and ethnic composition for five categories (White non-Hispanic, Black, Native American, Asian, and Hispanic/Latino) are obtained from the American Community Survey 5-year estimates (2018–22) for the population ages 5–17. Economic status, specifically child poverty rates, are derived from the Census Bureau's Small Area Income and Poverty Estimates (SAIPE) program. This approach captures all school-age children regardless of enrollment status.

Our methodology is designed to be robust to the noise introduced to tract-level counts by the Census's Disclosure Avoidance System (DAS). As documented by Kenny and et al. (2021), this noise is non-systematic for census tracts. By aggregating multiple tracts to form each simulated district, our models leverage the law of large numbers, substantially diminishing the effects of random error at our scale of analysis.

4.3.2 Enrollment Capacity Estimation

School capacity constraints are derived from the historical maximum enrollment for each school from the past decade, bounding practical capacity without new construction.[3] This sets an upper bound for the number of children that can be assigned to a simulated district, ensuring that redistricting respects facility constraints.

4.4 Data Integration

4.4.1 School District Mapping

Existing school district boundaries require careful processing due to the complex structure of American educational governance. We include only districts with assessment, demographic, and poverty status data for more than 75 percent geographic coverage by the underlying census tracts.[4] Similarly, states with less than 75 percent geographic coverage are excluded from our analysis. Those included have on average, 92.6 percent coverage.

Where elementary and secondary districts overlap, elementary districts are assigned to their corresponding unified or secondary district to avoid double-counting. In cases where only elementary or secondary districts exist, we use those boundaries and their corresponding data directly. This process yields approximately 10,500 school districts with sufficient demographic, property assessment, and spatial data for analysis.

5 Algorithmic Framework

The optimization framework employs a two-stage approach: spatial clustering for initialization, simulated annealing for refinement, followed by systematic variation across district counts. This section details the technical implementation of each stage and the mechanisms for constraint enforcement.

5.1 Stage 1: SKATER Initialization

Optimization begins with SKATER (Spatial 'K'luster Analysis by Tree Edge Removal), which generates geographically coherent initial district configurations through constrained graph partitioning (AssunÇão et al., 2006). By starting from maximally compact configurations rather than random assignments, we reduce the required iterations for convergence from hundreds of thousands (typical in redistricting literature) to approximately 2.5

[2] Our model optimizes for the potential tax base (assessed property value) rather than actual tax revenues, as we do not incorporate current tax rates. The assumption is that a large-scale reorganization would likely necessitate a recalibration of tax rates, making the underlying tax base the more stable and relevant metric for long-term fiscal capacity.

[3] Enrollment data are used only as a proxy for existing infrastructure capacity. We geocode schools to census tracts and use the maximum historical enrollment for each school to estimate available seats. School-age children counts from the census are used for demographic and population metrics. We multiply the capacity estimate by 1.25 as a conservative buffer, recognizing that not all resident children enroll in public schools.

[4] Coverage here refers to the spatial overlap of census tracts containing the necessary data (property assessment, demographic, and poverty) with existing school district boundaries. A small number of states and some rural areas exhibit higher rates of missing data and are thus excluded.

times the number of census tracts in each state. This reduction in computational expense allows us to complete the first school system redistricting analysis that is national in scope.

SKATER constructs a minimum spanning tree from the tract adjacency graph using edge weights based on scaled geographic coordinates. Each resulting partition forms a contiguous district, eliminating the need for post-hoc contiguity repair that plagues random initialization approaches often used in MCMC optimization.

5.1.1 Capacity Repair Mechanism

When SKATER produces initial configurations that violate the capacity constraint, a repair mechanism attempts to restore feasibility before optimization. For each violating district, tracts are evaluated for reassignment to neighboring districts with available capacity. The repair process attempts up to 50 chained explorations of 2,000 iterations each.

5.1.2 Status Quo Fallback Strategy

In cases where the SKATER initialization does not satisfy our constraints, and the chained repair strategy is unable to resolve the issue, we implement a fallback initialization approach. The system instead starts the optimization engine from the configuration of census tracts most similar to the status quo configuration of school districts. Because census tracts are not conterminous with existing school districts, we assign each tract to its geographic majority overlap district. This fallback mechanism ensures that the optimization can proceed, resolving the issue of invalid starting points.

5.2 Stage 2: MCMC Optimization

Following SKATER initialization, MCMC-based simulated annealing refines boundaries to minimize our previously defined objective function, optimizing on our criteria while maintaining all constraints. At each iteration, the algorithm selects a tract for potential reassignment. Border tracts are identified and preferentially selected, as they represent the only tracts that can change districts while maintaining contiguity. The selected tract is proposed for reassignment to a randomly chosen adjacent district.

The algorithm starts with a high temperature, its willingness to accept worse solutions, and it gradually becomes more selective over time, reducing this acceptance rate by 1 percent after each step, using the cooling formula $T(t+1) = T(t) \times 0.99$.

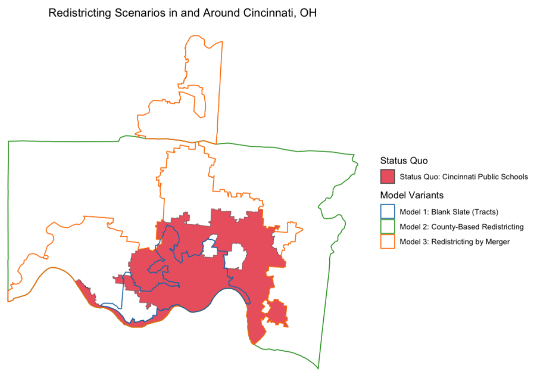

Figure 1: Example plans generated by each model in and around Cincinnati.

Beginning with acceptance to any move ($T_0 = 1.0$), the process continues until it has made 2.5 times as many successful changes as there are geographic units in the redistricting plan. Each move validates compliance with the constraints outlined in section 3.3.

5.3 Three Model Variants

The algorithmic framework is applied to two distinct redistricting models and a programmatic county-level merge, each offering different trade-offs between optimization flexibility and implementation feasibility.

5.3.1 Model 1: Blank-Slate Redistricting (Tract-Level Optimization)

This model uses census tracts as atomic units, providing maximum flexibility to create optimal boundaries. This model can completely reconfigure districts without regard to existing boundaries. It establishes the theoretical frontier for equity improvements.

5.3.2 Model 2: County-Based Redistricting

This model implements a programmatic, county-based consolidation, assigning all tracts within each county to a single district. The constraints of section 3.3 are relaxed for illustrative purposes. This simulation serves as a baseline to show what simple administrative consolidation achieves versus algorithmic optimization.

5.3.3 Model 3: Redistricting by Merger (Optimized District Consolidation)

This model uses existing school districts as atomic units, preserving current boundaries while allowing mergers. The same SKATER-optimization framework operates on a district adjacency graph rather than tract-level data. This provides more politically feasible solutions that maintain district identities while still pooling resources.

Both optimization models undergo systematic variation of current district counts. Each produces tract-to-district assignments with complete Theil index calculations, enabling direct comparison of equity impacts. The tract-level model demonstrates maximum theoretical improvements. The consolidation model balances feasibility with equity gains.[5] The county benchmark validates the value of optimization over simple administrative boundaries. Figure 2 compares the per pupil tax base equity improvements in Maryland.

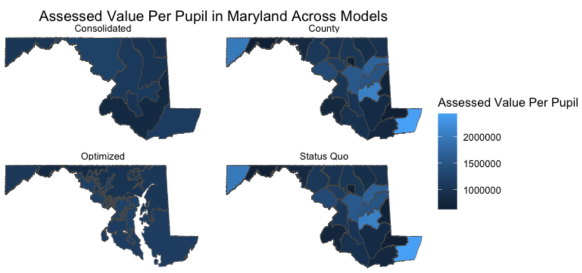

Figure 2: Per pupil tax base equity improvements in Maryland across models.

6 Output Specification

Each configuration for each model includes complete unit-to-district assignment vectors alongside a comprehensive metric suite: the three Theil indices (tax capacity, race, and poverty status), Polsby-Popper compactness scores, and additional metrics. This structure enables systematic comparison across varied numbers of districts, revealing that changing district count can improve resource distribution. The format remains consistent across all three models, which facilitates direct comparison of their relative performance.

6.1 Pareto Frontier Construction

Rather than selecting a single "optimal" solution, we identify the set of Pareto-efficient configurations that represent different trade-offs among competing objectives. A proposal is included only if no other configuration performs better on all objectives simultaneously. The Pareto selection evaluates across four criteria: minimizing the three Theil indices and maximizing geographic compactness (Polsby-Popper).

The frontier reveals critical trade-offs that cannot be resolved through optimization alone. Some configurations achieve significant funding equity but maintain racial segregation, while others integrate diverse populations at the cost of funding disparities.

6.2 Configuration Selection

While the Pareto frontier presents all efficient options, practical implementation requires selecting a single configuration. Normalized scores are combined using policy-determined weights that reflect our optimization metrics and compactness. The default weights prioritize tax base equity while maintaining focus on integration objectives and geographic coherence. The weighted score for each configuration equals the sum of each normalized measure multiplied by its corresponding weight. The configuration with the lowest weighted score is selected as the recommended plan for each state. This selection is performed only among Pareto-efficient configurations, ensuring the chosen plan is not dominated by any alternative.

The framework's key strength is its flexibility to accommodate different policy priorities. Stakeholders can adjust weights to explore how different priorities affect optimal configurations. This approach transforms a complex multi-objective optimization problem into a structured decision process. Rather than claiming to identify a single "best" solution, we provide a menu of high-quality options and a transparent mechanism for selection based on explicit policy priorities.

7 Results and Conclusion

Our strategic boundary optimization across 42 states reveals substantial potential for improving educational equity. Figure 3 demonstrates that the three models demonstrate that purposeful redrawing of district lines can significantly reduce property tax disparities in every state while simultaneously decreasing racial and economic segregation between districts.[6] As referenced in Table 1, the

[5] A potential critique is that aggregating smaller geographic units (like tracts or districts) into larger ones will mathematically reduce measured segregation by definition, as it averages over local variations. While this is true, the magnitude of the reduction achieved through our optimization far exceeds what would be expected from simple aggregation, demonstrating the value of purposeful boundary drawing.

[6] Figures detailing Theil improvements across states for racial and economic segregation can be found in Appendix A.

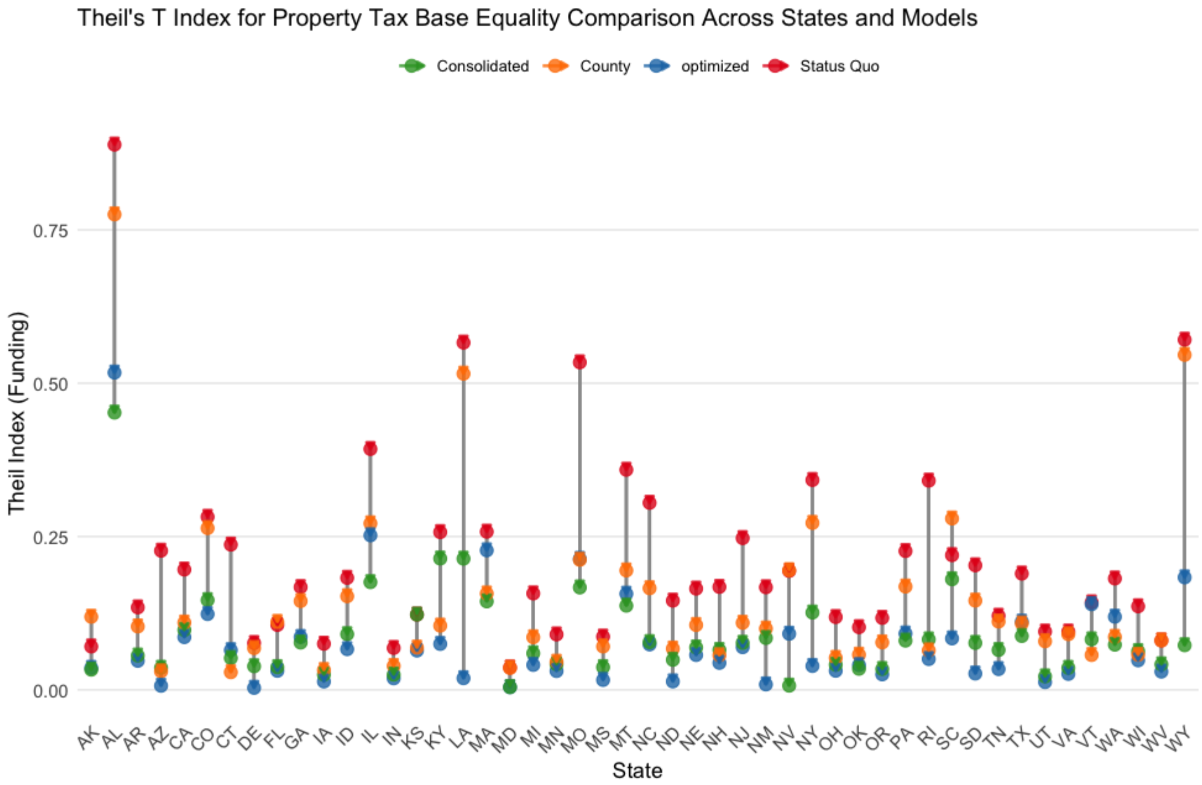

Figure 3: Theil Index improvement for property tax disparities across 42 states. This figure spans the full width of the page to show detail across all states.

Model	Property Tax Equity Improvement	Racial Integration Improvement	Economic Integration Improvement
Blank-Slate Redistricting	66.6%	47.6%	65.0%
County-Based Redistricting	39.0%	40.7%	57.2%
Redistricting by Merger	63.0%	48.2%	54.6%

Table 1: Equity Improvements by Redistricting Model. *Note:* Applied to 42 states. County-based results reflect 37 states that would see boundary changes.

Blank-Slate approach achieves average improvements of 66.6% in tax-base equality, 47.6% reduction in racial segregation, and 65.0% decrease in economic segregation. Most notably, these improvements can be achieved purely through boundary changes, without any student or family having to move. The Merger model offers a potentially more politically feasible alternative, keeping district identities intact while still delivering meaningful improvements.

While our local search approach cannot guarantee global optimality, the framework provides policymakers with concrete evidence of redistricting's potential benefits. The primary barrier to implementation remains political feasibility, as communities maintain strong attachments to existing districts. Additionally, using rook contiguity rather than actual road networks and historical enrollment maximums for capacity estimates may not capture all practical constraints like transportation barriers or current infrastructure conditions. Future work could address these limitations by incorporating dynamic demographic modeling, actual transportation networks, and mechanisms to predict post-redistricting property value adjustments.

Despite these constraints, this study establishes the first national-scale framework for school redistricting. By demonstrating that significant equity improvements are technically achievable across diverse state contexts, we provide an objective foundation for policy discussions about using boundary change as a tool for educational equity. The flexibility of our multi-objective optimization approach allows stakeholders to explore trade-offs transparently, transforming a complex challenge into a structured decision process grounded in empirical evidence.

References

R. M. AssunÇão, M. C. Neves, G. Câmara, and C. Da Costa Freitas. 2006. Efficient regionalization techniques for socio-economic geographical units using minimum spanning trees. *International Journal of Geographical Information Science*, 20(7):797–811.

Marvin E. Dodson and T. A. Garrett. 2008. Inefficient education spending in public school districts: A case for consolidation? *Contemporary Economic Policy*, 22(2):270–280.

William Duncombe and John Yinger. 2007. Does school district consolidation cut costs? *Education Finance and Policy*, 2(4):341–375.

Benjamin Fifield, Kosuke Imai, and James Olmsted. 2014. Automated redistricting simulation using markov chain monte carlo. pages 593–604.

Matthew Gartner. 2023. State education budgets and equity. *Education Policy Analysis Archives*.

Beeferman D. Vega-Pourheydarian C. Overney C. Van Hentenryck P. Roy D. Gillani, N. 2023. Redrawing attendance boundaries to promote racial and ethnic diversity in elementary schools. *Educational Researcher*, 52(6):348–364.

C. Kirabo Jackson, Rucker C. Johnson, and Claudia Persico. 2015. The effects of school spending on educational and economic outcomes: Evidence from school finance reforms. *NBER Working Paper Series*.

C. Kirabo Jackson and Claire Mackevicius. 2024. What impacts can we expect from school spending policy? evidence from evaluations in the united states. *American Economic Journal: Applied Economics*, pages 412–446.

C. Kenny and et al. 2021. The impact of the u.s. census disclosure avoidance system on redistricting and voting rights analysis. *Science Advances*.

Daphne A. Kenyon and Bethany Munteanu. 2021. State-by-state disparities in school district revenue. In *Lincoln Institute of Land Policy*.

Moritz Kuhn, Moritz Schularick, and Ulrike I. Steins. 2018. The historical process of racial and economic segregation. *Working Paper*.

Cory McCartan and Kosuke Imai. 2023. Sequential monte carlo for sampling balanced and compact redistricting plans. *Annals of Applied Statistics*, 17(4):3300–3323.

Ann Owens. 2016. Inequality in children's contexts: Income segregation of households with and without children. *American Sociological Review*, 81(3):456–488.

Sean F. Reardon and Ann Owens. 2014. 60 years after brown: Trends and consequences of school segregation. *Annual Review of Sociology*, 40:199–218.

Sean F. Reardon and Ericka Weathers. 2024. Reardon, s. f., weathers, e. s., fahle, e. m., jang, h., kalogrides, d. *American Sociological Review*, 89(06):971–1010.

Tyler Simko. 2024. School desegregation by redrawing district boundaries. *Nature Scientific Reports*.

Zahava Stadler and Jordan Abbott. 2024. Crossing the line: Segregation and resource inequality between america's school districts. *New America*.

A State-Level Theil Improvements

This appendix provides additional visualizations of the state-level improvements for racial and economic segregation metrics, complementing Figure 3 in the main text.

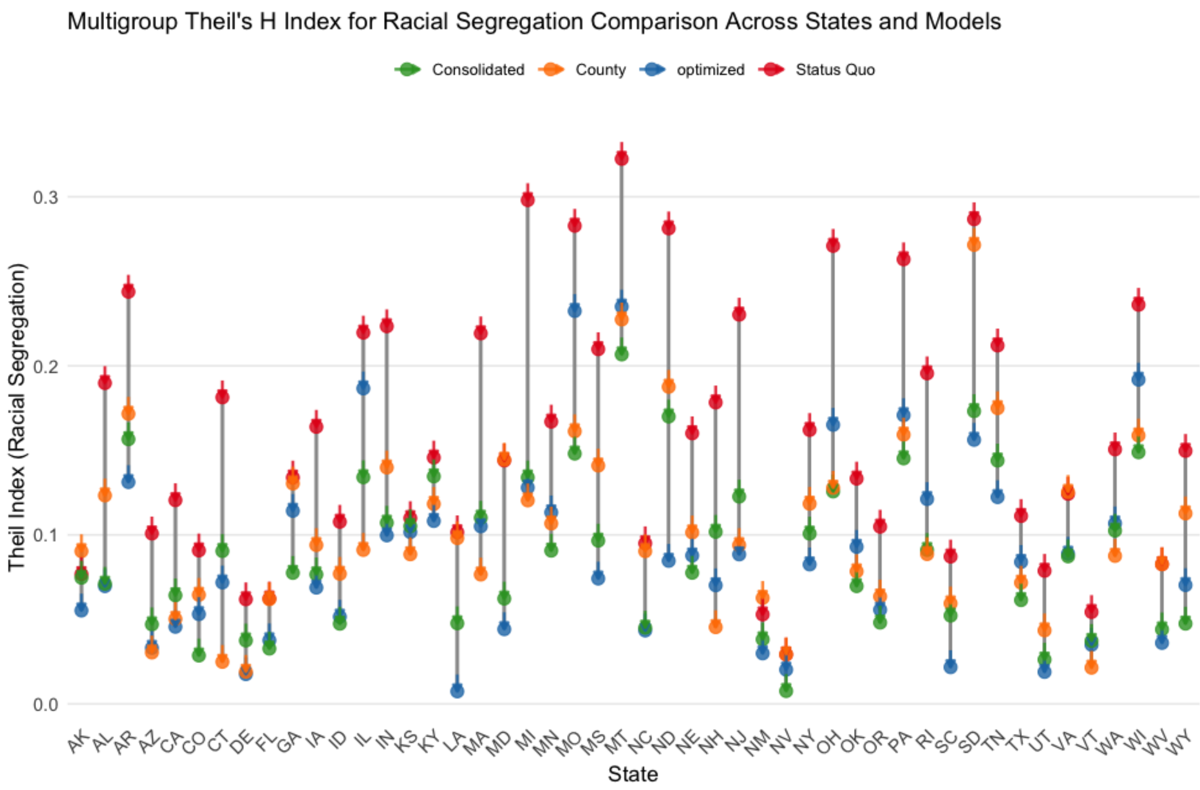

Figure 4: Theil Index improvement for racial segregation across 42 states.

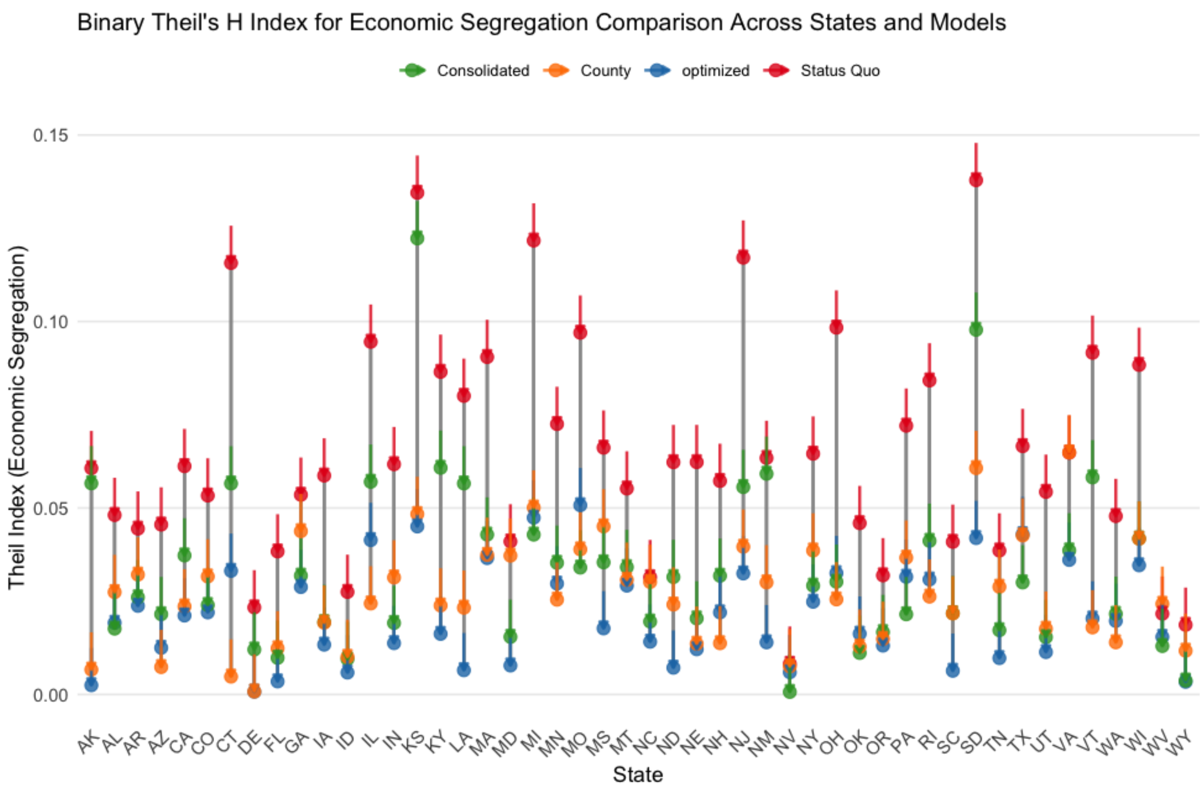

Figure 5: Theil Index improvement for economic segregation across 42 states.