Eliminating Child Care Deserts in New York State: A Budget-Minimizing Siting and Expansion Model

Chunghyun Han[1] Jiahe Ling[1] Ginni Wang[1] Shiying Yan[1]
ch4005@columbia.edu jl7264@columbia.edu xw3088@columbia.edu sy3320@columbia.edu

[1]*Department of Industrial Engineering and Operations Research, Columbia University*

1 Introduction

Access to reliable and affordable child care enables parents to participate in the labor market and supports children's early development. Yet many communities face persistent "child care deserts", where licensed supply falls far short of need. Nationally, 14.4 million children (about 67.8 percent of all U.S. children) have all available parents in the workforce, and many families curtail work or leave jobs because care is inaccessible or unaffordable (Lovejoy, 2024). Prior mapping shows that 51 percent of U.S. residents live in a child care desert, and the share rises to nearly 64 percent in New York State (CAP, 2020). These figures underscore the urgency of targeted, statewide planning.

Previous research provides building blocks for such planning. Equal-access models jointly choose locations and capacities to reduce inequality in accessibility; fairness-aware covering models encode equity directly in the objective; and multiobjective daycare models trade off coverage, distance, and cost (Blanco & Gázquez, 2023; Li et al., 2017; Luna Rojas & Flores de la Mota, 2022; Ofsted, 2025). Together, these strands motivate an approach that integrates siting with capacity decisions under explicit equity goals.

This study develops an integer optimization model that minimizes the total budget required to eliminate child care deserts across New York State ZIP codes. The model combines population by age, income, parental employment, and existing capacity to identify where to build new facilities and where to expand current ones. It incorporates expansion cost tiers, minimum-distance rules between facilities, and age-specific capacity for children ages 0 to 5. The contribution is a policy-ready framework that turns equity goals into binding coverage guarantees, co-optimizes siting and expansion under realistic constraints, and delivers implementable, costed plans for allocating limited public resources.

2 Background

2.1 Literature Review

Improving childcare access sits at the intersection of facility location, capacity planning, and equity. Recent studies provide complementary tools: two-step equal-access designs that separate siting from sizing, fairness-aware covering models that formalize equity, and daycare planning that trades off coverage and cost.

Li et al. (2017) frame planning as minimizing inequality in a gravity-based accessibility index and propose a sequential method that first selects new sites with a genetic algorithm and then optimizes capacities with quadratic programming within bounds. Their experiments show that changing locations reduces inequality more than adjusting capacities, which supports planning that prioritizes siting before sizing. Ofsted (2025) adapts this approach at national scale for England: a two-step workflow chooses where provision should be and then assigns places, with case studies reporting a 56 percent reduction in accessibility variation in Lincolnshire and a 61 percent reduction in Sandwell and Walsall. From the equity side, Blanco and Gázquez (2023) embed fairness measures into maximal covering using ordered weighted averaging and alpha fairness, and provide second-order cone and linear reformulations that make the models computationally tractable. Within childcare specifically, Luna Rojas and Flores de la Mota (2022) develop a multiobjective location–allocation model for Mexico City that minimizes operating cost and travel distance while maximizing covered demand under budget, capacity, and service-radius constraints; in their test case, coverage is fully achieved and the Pareto analysis typically recommends opening five mostly indirect-service centers within budget.

Taken together, these works motivate designs that equalize access through siting, calibrate capacity, and make equity explicit. The present formulation builds directly on this logic by turning equal-access goals into hard coverage guarantees at the zipcode level, embedding an age 0 to 5 coverage policy with explicit assignment limits, allowing realistic segmented expansions at existing sites with increasing marginal costs, and preventing over-clustering through minimum distance rules. Unifying these elements in a single integer programming yields policy-ready plans that decide where to build, how large to build, how much to expand, and how many 0 to 5 places to assign while certifying that no zipcode remains a childcare desert.

2.2 Data

This study integrates five sources: ZIP level child population by age with 1,646 observations, parental employment rates with 1,375 observations, average individual income with 1,534 observations, regulated child care facilities with 15,604 records, and candidate facility locations with 215,400 points. The workflow begins with a full join of population, employment, and income to form the demographic panel, followed by median imputation for missing numeric fields to retain coverage across ZIP codes. Median imputation was employed because the distributions of variables with missing values were skewed. Population measures are standardized by approximating ages 10 to 12 as three fifths of the reported 10 to 14 total under a uniform distribution assumption, constructing ages 0 to 12 as the sum of under five, ages 5 to 9, and the derived ages 10 to 12, and rounding counts to integers. ZIP codes are tagged as high demand when the employment rate is at least 0.60 or average income is at most 60,000 dollars, otherwise they are tagged as normal (Kaya, 2025).

Facility records are then aligned to the demographic panel and restricted to valid coordinates. Two capacity measures are constructed. Under five capacity equals infant plus toddler plus preschool plus five twelfths of undifferentiated children capacity, assuming even age distribution within that field. Total ages 0 to 12 capacity equals infant plus toddler plus preschool plus school age plus children. After restricting to ZIP codes present in the demographic panel, 15,000 facilities remain and 14 are removed because their ZIP codes are not in the demographic data. Capacities are aggregated to the ZIP level and merged into the demographic panel, with missing aggregates set to zero to represent areas with no licensed supply.

Candidate locations are screened in two stages. First, a proximity filter removes candidates within 0.06 miles (Euclidean distance) of any existing facility, leaving 207,250 candidates and removing 8,150. Second, candidates are restricted to ZIP codes present in the demographic panel, leaving 174,459 and removing 32,791 because their ZIP codes are not in the demographic data.

Figure 1 illustrates the geographic distribution of child-care deserts and under-5 care compliance across New York State, comparing conditions before and after facility expansion and new construction.

3 Methodology

3.1 General Assumptions

Demand is deterministic and measured at the ZIP code level for ages 0 to 12 and for ages 0 to 5. A ZIP is classified as high demand when at least 60 percent of parents are employed or the average income is 60,000 dollars or less per year, otherwise it is normal demand (Kaya, 2025). A ZIP is a child care desert if available licensed slots for ages two weeks to 12 years are at most one half of the local child population in high demand areas, or at most one third in normal demand areas (Kaya, 2025). State policy for young children requires that available slots for ages 0 to 5 be at least two thirds of the local 0 to 5 population after investments (Kaya, 2025). New facilities are available only in three sizes with location-invariant parameters: Small costs 65,000 dollars with capacity 100 and a maximum of 50 slots for ages 0 to 5, Medium costs 95,000 dollars with capacity 200 and a maximum of 100 slots for ages 0 to 5, Large costs 115,000 dollars with capacity 400 and a maximum of 200 slots for ages 0 to 5 (Kaya, 2025). Each added 0 to 5 slot carries an additional 100 dollars regardless of whether it is created by construction or expansion (Kaya, 2025). The models are static and single period, costs and populations are known and constant, and capacity within ages 0 to 12 is perfectly usable across ages except for the explicit cap on 0 to 5 slots at new facilities.

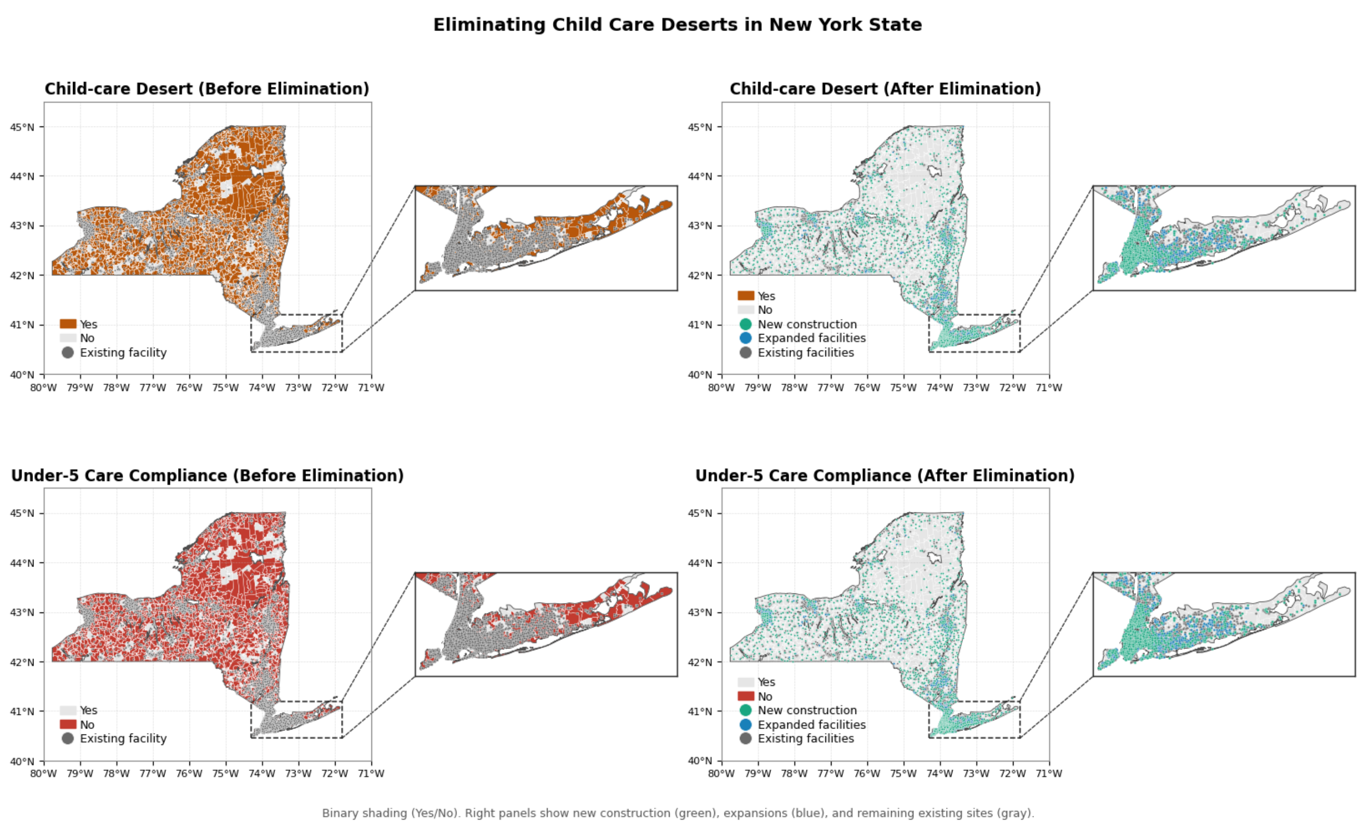

Figure 1: Elimination of child-care deserts and improvement in under-5 care compliance in New York State.

3.2 Baseline Budget Allocation Model

The Baseline Budget Allocation Model imposes no siting or distance constraints and allows any number of new facilities to be built in any ZIP. Existing facilities can expand by at most the smaller of 120 percent of current capacity and 500 slots, and added 0 to 5 slots cannot exceed total added slots at a facility. The expansion cost follows the objective exactly and has two parts. First, a pro-rata activation component equal to 20,000 dollars multiplied by the share of added slots up to current capacity, which ramps from zero when there is no expansion to 20,000 dollars when the number of added slots reaches current capacity and then remains capped. Second, a marginal component of 200 dollars per slot for additions beyond current capacity. Costs are separable across facilities and linear in the decision variables apart from this capped ramp, and feasibility and coverage are evaluated at the ZIP level.

In the baseline model, notation follows Table 1, where I indexes ZIP codes and F_i denotes the set of licensed facilities located in ZIP i. New construction uses discrete sizes $s \in S = \{\text{Small}, \text{Medium}, \text{Large}\}$ with build cost K_s, total capacity C_s, and an age specific limit $C_s^{(0\text{–}5)}$ for children ages 0–5. Population is summarized by P_i for ages 0–12 and $P_i^{(0\text{–}5)}$ for ages 0–5. Equity requirements are encoded by the desert factor α_i, equal to $\frac{1}{2}$ in high demand ZIP codes and $\frac{1}{3}$ otherwise. Existing supply at each facility $f \in F_i$ is represented by n_f total licensed slots for ages 0–12 and $n_f^{(0\text{–}5)}$ licensed slots dedicated to ages 0–5.

The decision variables allocate resources on two margins. At existing facilities, x_f denotes the total number of added slots and $x_f^{(0\text{–}5)}$ the added slots reserved for ages 0–5. At the ZIP level, $y_{i,s}$ gives the number of new facilities of size s built in ZIP i and $z_{i,s}$ gives the number of ages 0–5 slots assigned within those new facilities. A binary trigger $u_f \in 0, 1$ indicates whether any expansion occurs at facility f, and v_f records the portion of expansion above current capacity through $v_f = \max\{0, x_f - n_f\}$, so $x_f - v_f$ is the expansion up to the current capacity level.

$$
\begin{aligned}
\min\ & \sum_{i\in I}\sum_{s\in S}\left(K_s\, y_{i,s} + 100\, z_{i,s}\right) + \sum_{i\in I}\sum_{f\in F_i}\left(100\, x_f^{(0\text{–}5)} + 20{,}000\, u_f\, \frac{x_f - v_f}{n_f} + 200\, v_f\right)\\
s.t.\ & \sum_{f\in F_i}(n_f + x_f) + \sum_{s\in S} C_s\, y_{i,s} \ \ge\ \alpha_i P_i \qquad \forall i \in I\\
& \sum_{f\in F_i}\left(n_f^{(0\text{–}5)} + x_f^{(0\text{–}5)}\right) + \sum_{s\in S} z_{i,s} \ \ge\ \tfrac{2}{3}\, P_i^{(0\text{–}5)} \qquad \forall i \in I\\
& 0 \le z_{i,s} \le C_s^{(0\text{–}5)}\, y_{i,s} \qquad \forall i \in I,\ \forall s \in S\\
& 0 \le x_f \le 1.2\, n_f, \quad 0 \le x_f \le 500 \qquad \forall f \in F\\
& 0 \le x_f^{(0\text{–}5)} \le x_f \qquad \forall f \in F\\
& x_f \le 1.2\, n_f\, u_f, \quad x_f \le 500\, u_f \qquad \forall f \in F\\
& v_f \ge x_f - n_f, \quad v_f \le x_f \qquad \forall f \in F\\
& y_{i,s},\, z_{i,s},\, x_f,\, x_f^{(0\text{–}5)} \in \mathbb{Z}_{\ge 0}, \quad u_f \in \{0,1\}, \quad v_f \in \mathbb{Z}_{\ge 0}
\end{aligned}
$$

The baseline objective has three parts. New construction incurs a fixed cost K_s per facility of size s and a linear 0–5 assignment cost $100\, z_{i,s}$, representing equipment and staffing premia for younger children at new sites. Additions at existing facilities carry a linear 0–5 cost $100\, x_f^{(0\text{–}5)}$ plus a two–piece expansion charge: an activation overhead $20{,}000\, u_f\, (x_f - v_f)/n_f$ that is paid only if $u_f = 1$ (i.e., $x_f > 0$) and is amortized evenly across the first n_f added slots, and a marginal capital cost $200\, v_f$ on slots beyond current capacity, where $v_f = \max\{0, x_f - n_f\}$. The assumptions are that per–slot costs for ages 0–5 are linear and identical at new and existing sites, fixed build costs depend only on the chosen size, and deeper expansions become more expensive on average once $x_f > n_f$. This yields a convex, piecewise-linear cost in x_f with a kink at $x_f = n_f$, which tends to favor shallow, distributed expansions unless the size-specific build option (K_s, C_s) is more economical under the coverage constraints.

The objective minimizes total public cost by summing fixed build and 0–5 assignment terms $\sum_{i\in I}\sum_{s\in S}(K_s y_{i,s} + 100\, z_{i,s})$ and expansion terms $\sum_{i\in I}\sum_{f\in F_i}\left(100\, x_f^{(0\text{–}5)} + 20{,}000\, u_f\, (x_f - v_f)/n_f + 200\, v_f\right)$, where the proportional charge $20{,}000\, u_f\, (x_f - v_f)/n_f$ distributes a one time activation cost across the first n_f added slots at f and 200, v_f prices the portion that exceeds the current capacity. Coverage constraints require the total capacity in each ZIP to meet equity and age specific targets, namely $\sum_{f\in F_i}(n_f + x_f) + \sum_{s\in S} C_s y_{i,s} \ge \alpha_i P_i$ and $\sum_{f\in F_i}(n_f^{(0\text{–}5)} + x_f^{(0\text{–}5)}) + \sum_{s\in S} z_{i,s} \ge \frac{2}{3} P_i^{(0\text{–}5)}$. Feasibility at new builds is enforced by $0 \le z_{i,s} \le C_s^{(0\text{–}5)} y_{i,s}$. Expansion magnitudes satisfy $0 \le x_f \le 1.2\, n_f$ and $0 \le x_f \le 500$, with $0 \le x_f^{(0\text{–}5)} \le x_f$. The trigger and the above capacity portion are linked by $x_f \le 1.2\, n_f\, u_f$, $x_f \le 500\, u_f$, $v_f \ge x_f - n_f$, and $v_f \le x_f$. Integrality and nonnegativity hold with $y_{i,s}, z_{i,s}, x_f, x_f^{(0\text{–}5)}, v_f \in \mathbb{Z}_{\ge 0}$ and $u_f \in \{0, 1\}$.

Table 1: Sets, Parameters, and Decision Variables

Baseline Budget Allocation Model	
Sets and Parameters	
I	Set of ZIP codes.
S	Sizes {Small, Medium, Large}.
F_i	Facilities located in ZIP i.
α_i	Desert threshold: $1/2$ if ZIP i high demand; else $1/3$.
P_i	Population age 0–12 in ZIP i.
$P_i^{(0\text{–}5)}$	Population age 0–5 in ZIP i.
n_f	Number of existing slots at facility f.
$n_f^{(0\text{–}5)}$	Number of existing slots for age 0–5 at facility f.
K_s	Build cost (USD) for s that $K_s \in \{65{,}000,\ 95{,}000,\ 115{,}000\}$.
C_s	Capacity for s that $C_s \in \{100,\ 200,\ 400\}$.
$C_s^{(0\text{–}5)}$	Max slots for age 0–5 for s that $C_s^{(0\text{–}5)} \in \{50,\ 100,\ 200\}$.
Decision variables	
x_f	Number of slots expanded at facility f.
$x_f^{(0\text{–}5)}$	Number of slots for age 0–5 expanded at facility f.
$y_{i,s}$	Number of new facilities of size s built in ZIP i.
$z_{i,s}$	Number of slots for age 0–5 in new facilities of size s in ZIP i.
$u_f \in \{0,1\}$	1 iff facility f expands ($x_f > 0$); else 0.
$v_f \ge 0,\ v_f \in \mathbb{Z}$	Number of slots above capacity: $v_f = \max\{0,\ x_f - n_f\}$.

Integrated Facility Location and Capacity Expansion Model	
Sets and Parameters	
I	Set of ZIP codes.
J_i^{new}	Candidate new facility locations in ZIP i.
J_i^{existing}	Existing facility locations in ZIP i.
J_i	All facilities in ZIP i ($J_i = J_i^{\text{existing}} \cup J_i^{\text{new}}$).
S	Sizes {Small, Medium, Large}.
α_i	Desert threshold: $1/2$ if ZIP i high demand; else $1/3$.
P_i	Population age 0–12 in ZIP i.
$P_i^{(0\text{–}5)}$	Population age 0–5 in ZIP i.
$n_{i,j}$	Number of existing slots at facility j in ZIP i.
$n_{i,j}^{(0\text{–}5)}$	Number of existing slots for age 0–5 at facility j in ZIP i.
K_s	Build cost (USD) for s that $K_s \in \{65{,}000,\ 95{,}000,\ 115{,}000\}$.
C_s	Capacity for s that $C_s \in \{100,\ 200,\ 400\}$.
$C_s^{(0\text{–}5)}$	Max slots for age 0–5 for s that $C_s^{(0\text{–}5)} \in \{50,\ 100,\ 200\}$.
$d_{i,j,z,k}$	Distance between facilities (i, j) and (z, k).
$\mathcal{P}^{NN}$	Index set of new facility and new facility pairs with $(i, j) \prec (z, k)$ and $d_{i,j,z,k} < 0.06$.
$\mathcal{P}^{NE}$	Index set of new facility and existing facility pairs with $d_{i,j,z,k} < 0.06$.
Decision variables	
$t_{i,j}^{(1)},\ t_{i,j}^{(2)},\ t_{i,j}^{(3)}$	Counts of segment (1)–(3) expansion units at existing facility j in ZIP i.
$x_{i,j}$	Number of slots expanded at existing facility j in ZIP i.
$x_{i,j}^{(0\text{–}5)}$	Number of slots for age 0–5 expanded at existing facility j in ZIP i.
$y_{i,j,s} \in \{0,1\}$	1 iff a new facility j of size s is built in ZIP i; else 0.
$z_{i,j,s}$	Number of slots for age 0–5 assigned to new facility j of size s in ZIP i.

3.3 Integrated Facility Location and Capacity Expansion Model

The Integrated Facility Location and Capacity Expansion Model restricts construction to predefined candidate sites and permits at most one size choice at any selected site. Spatial dispersion is enforced by distance rules: two candidate sites within 0.06 miles cannot both be opened, and any candidate within 0.06 miles (Euclidean distance) of an existing provider is not eligible. Age zero to five assignments in new facilities must respect the size specific caps, and age zero to five additions at an existing facility cannot exceed its total added slots. Expansion at existing facilities is tiered and capacity bounded: the first tier covers ten percent of current capacity, the second tier five percent, and the third tier five percent, yielding a twenty percent cap per site. Expansion costs increase by tier and have two components. A pro rata activation charge of twenty thousand dollars is allocated across the added slots up to current capacity. Marginal costs apply to each tier at two hundred dollars per slot in tier one, four hundred dollars per slot in tier two, and one thousand dollars per slot in tier three.

$$
\begin{aligned}
\min\ & \sum_{i\in I}\sum_{j\in J_i^{\text{new}}}\sum_{s\in S}\big(K_s y_{i,j,s} + 100\, z_{i,j,s}\big) + \sum_{i\in I}\sum_{j\in J_i^{\text{existing}}}\Big(100\, x_{i,j}^{(0\text{–}5)} + \big(\tfrac{20{,}000}{n_{i,j}}+200\big)t_{i,j}^{(1)} + \big(\tfrac{20{,}000}{n_{i,j}}+400\big)t_{i,j}^{(2)} + \big(\tfrac{20{,}000}{n_{i,j}}+1000\big)t_{i,j}^{(3)}\Big)\\
s.t.\ & \sum_{j\in J_i^{\text{existing}}}\big(n_{i,j}+x_{i,j}\big) + \sum_{j\in J_i^{\text{new}}}\sum_{s\in S} C_s y_{i,j,s} \ \ge\ \alpha_i P_i \qquad \forall i\in I\\
& \sum_{j\in J_i^{\text{existing}}}\big(n_{i,j}^{(0\text{–}5)} + x_{i,j}^{(0\text{–}5)}\big) + \sum_{j\in J_i^{\text{new}}}\sum_{s\in S} z_{i,j,s} \ \ge\ \tfrac{2}{3}P_i^{(0\text{–}5)} \qquad \forall i\in I\\
& 0\le z_{i,j,s}\le C_s^{(0\text{–}5)} y_{i,j,s} \qquad \forall i\in I,\ \forall j\in J_i^{\text{new}},\ \forall s\in S\\
& \sum_{s\in S} y_{i,j,s}\le 1 \qquad \forall i\in I,\ \forall j\in J_i^{\text{new}}\\
& x_{i,j} = t_{i,j}^{(1)} + t_{i,j}^{(2)} + t_{i,j}^{(3)} \qquad \forall i\in I,\ \forall j\in J_i^{\text{existing}}\\
& 0\le t_{i,j}^{(1)}\le 0.10\, n_{i,j},\quad 0\le t_{i,j}^{(2)}\le 0.05\, n_{i,j},\quad 0\le t_{i,j}^{(3)}\le 0.05\, n_{i,j} \qquad \forall i\in I,\ \forall j\in J_i^{\text{existing}}\\
& 0\le x_{i,j}\le 0.20\, n_{i,j},\quad 0\le x_{i,j}^{(0\text{–}5)}\le x_{i,j} \qquad \forall i\in I,\ \forall j\in J_i^{\text{existing}}\\
& \sum_{s\in S} y_{i,j,s} + \sum_{s\in S} y_{z,k,s}\le 1\ \ \forall (i,j,z,k)\in\mathcal{P}^{NN}\ ;\ \sum_{s\in S} y_{i,j,s} = 0\ \ \forall (i,j,z,k)\in\mathcal{P}^{NE}\\
& y_{i,j,s}\in\{0,1\},\quad z_{i,j,s}, x_{i,j}, x_{i,j}^{(0\text{–}5)}\in\mathbb{Z}_{\ge 0},\quad t_{i,j}^{(1)}, t_{i,j}^{(2)}, t_{i,j}^{(3)}\ge 0
\end{aligned}
$$

In the integrated model, notation follows Table 1. The index set I enumerates ZIP codes. For each $i\in I$, J_i^{existing}, and J_i^{new}, denote existing and candidate sites, with $J_i = J_i^{\text{existing}}\cup J_i^{\text{new}}$. Facilities can be built in discrete sizes $s\in S = \{\text{Small}, \text{Medium}, \text{Large}\}$, each with a build cost K_s, capacity C_s, and a maximum allocation $C_s^{(0\text{–}5)}$ for ages 0–5. Populations are summarized by P_i for ages 0–12 and $P_i^{(0\text{–}5)}$ for ages 0–5. The parameter α_i encodes the child care desert threshold, with $\alpha_i = \frac{1}{2}$ in high demand ZIP codes and $\alpha_i = \frac{1}{3}$ otherwise. Existing capacity at facility $j\in J_i^{\text{existing}}$ is $n_{i,j}$ for ages 0–12, and $n_{i,j}^{(0\text{–}5)}$ for ages 0–5. Spatial frictions are represented by distances $d_{i,j,z,k}$ between facilities (i,j) and (z,k) and by the conflict index sets $\mathcal{P}^{NN}$ and $\mathcal{P}^{NE}$, which collect new–new and new–existing facility pairs with $d_{i,j,z,k} < 0.06$.

The decision variables comprise binary build choices $y_{i,j,s}$ indicating whether candidate $j\in J_i^{\text{new}}$ is opened at size s in ZIP i, and allocations $z_{i,j,s}$ for the number of 0–5 slots at that new site, with $0\le z_{i,j,s}\le C_s^{(0\text{–}5)} y_{i,j,s}$ and a one size per site restriction $\sum_{s\in S} y_{i,j,s}\le 1$. Capacity expansion at existing facilities is modeled by $x_{i,j}$ total added slots and $x_{i,j}^{(0\text{–}5)}$ added 0–5 slots, with $0\le x_{i,j}^{(0\text{–}5)}\le x_{i,j}$. The variable $x_{i,j}$ is decomposed into segments $t_{i,j}^{(1)}$, $t_{i,j}^{(2)}$, and $t_{i,j}^{(3)}$ that satisfy $x_{i,j} = t_{i,j}^{(1)} + t_{i,j}^{(2)} + t_{i,j}^{(3)}$, with segment caps

$t_{i,j}^{(1)} \leq 0.10\, n_{i,j}$, $t_{i,j}^{(2)} \leq 0.05, n_{i,j}$, and $t_{i,j}^{(3)} \leq 0.05\, n_{i,j}$, which jointly imply $0 \leq x_{i,j} \leq 0.20\, n_{i,j}$. Increasing segment coefficients in the objective induce an ordered use of $t_{i,j}^{(1)}$, $t_{i,j}^{(2)}$, and $t_{i,j}^{(3)}$, and the notation supports constraints that guaranty no ZIP remains a child care desert, that at least $\frac{2}{3}P_i^{(0\text{–}5)}$ is served for ages 0–5, that age specific limits are respected in new builds, and that distance based conflicts prevent over-clustering.

The model minimizes total public cost by jointly selecting new facilities and expanding existing capacity. The objective includes fixed build costs $K_s y_{i,j,s}$ and per slot assignment costs $100\, z_{i,j,s}$ for ages 0–5 at new sites, and it prices expansions at existing sites through $100/x_{i,j}^{(0\text{–}5)}$ plus piecewise charges $\left(\frac{20{,}000}{n_{i,j}} + 200\right)t_{i,j}^{(1)}$, $\left(\frac{20{,}000}{n_{i,j}} + 400\right)t_{i,j}^{(2)}$, and $\left(\frac{20{,}000}{n_{i,j}} + 1000\right)t_{i,j}^{(3)}$. Because these coefficients are increasing, optimal solutions exhaust $t_{i,j}^{(1)}$ before $t_{i,j}^{(2)}$ and then $t_{i,j}^{(3)}$ within their segment limits. Coverage is enforced through two requirements. First, the no desert condition requires total capacity in each ZIP to exceed $\alpha_i P_i$. Second, the age policy for 0–5 requires at least $\frac{2}{3}P_i^{(0\text{–}5)}$ places. Feasibility at new builds is maintained by $0 \leq z_{i,j,s} \leq C_s^{(0\text{–}5)} y_{i,j,s}$ and by a one size rule $\sum_{s\in S} y_{i,j,s} \leq 1$. Expansions at existing sites satisfy $x_{i,j} = t_{i,j}^{(1)} + t_{i,j}^{(2)} + t_{i,j}^{(3)}$, with $0 \leq t_{i,j}^{(1)} \leq 0.10\, n_{i,j}$, $0 \leq t_{i,j}^{(2)} \leq 0.05\, n_{i,j}$, $0 \leq t_{i,j}^{(3)} \leq 0.05\, n_{i,j}$, and $0 \leq x_{i,j} \leq 0.20\, n_{i,j}$, while age targeted additions obey $0 \leq x_{i,j}^{(0\text{–}5)} \leq x_{i,j}$. Spatial dispersion is imposed by distance constraints, where for all $(i, j, z, k) \in \mathcal{P}^{NN}$ at most one of the candidates can be selected, and for all $(i, j, z, k) \in \mathcal{P}^{NE}$ no new site may be chosen near an existing provider. Integrality and nonnegativity close the model, with $y_{i,j,s} \in \{0, 1\}$, $z_{i,j,s}\, x_{i,j}\, x_{i,j}^{(0\text{–}5)} \in \mathbb{Z}_{\geq 0}$, and $t_{i,j}^{(1)}, t_{i,j}^{(2)}, t_{i,j}^{(3)} \geq 0$.

4 Results and Discussions

4.1 Objective Overview

The optimization in *Baseline Budget Allocation Model* produced an objective value of 255,138,373.93 dollars. This amount represents the minimum total cost, under the model assumptions, required to eliminate all childcare deserts in New York State while satisfying the specified 0–5 policy requirements. When location constraints and realistic facility-expansion limits are introduced in *Integrated Facility Location and Capacity Expansion Model*, the objective rises to 423,942,065.87 dollars, an increase of 168,803,691.94 dollars (66.16 percent). The increase reflects additional expenditures associated with siting restrictions and practical limits on capacity expansion.

Furthermore, this study verified solution validity with three integrity checks. First, targeted investment: new construction and expansions are limited to areas previously identified as childcare deserts or to tracts where 0–5 demand remained unmet. Second, post-expansion verification: after simulated additions, no tract remained classified as a childcare desert and all regions met the minimum 0–5 capacity requirement. Third, cost ordering: the solution consistently deploys lower-cost capacity options before higher-cost alternatives, confirming correct prioritization of expenditures. Collectively, these checks indicate that the solutions are feasible, meet policy criteria, and respect the model's cost structure.

4.2 Cost Analysis

4.2.1 Baseline Budget Allocation Model

The *Baseline Budget Allocation Model* delivers 813,087 new childcare slots statewide, with 391,300 from new construction and 421,787 from expansions. At least one new facility is added in 86 percent of ZIP codes, which indicates broad geographic coverage rather than a concentration of investment in a small set of tracts. The distribution of additions across ZIP codes is right skewed. The median ZIP receives 135 new slots, while the mean is 450.5. This indicates that the model directs larger absolute increases to high demand areas while still delivering smaller but meaningful gains to the majority of localities.

Construction accounts for 391,300 slots at a total cost of 157,225,900 dollars, an average of 401.8 dollars per slot. Of these, 160,609 are infant slots, equal to 41.0 percent of construction capacity. Spending is split between 61,625,749 dollars for ages 0 to 5 and 95,600,151 dollars for ages 5 to 12. Expansion contributes 421,787 slots at a total cost of 97,912,474 dollars, an average of 232.1 dollars per slot. Of these, 385,882 are infant slots, equal to 91.5 percent of expansion capacity. Spending is concentrated on ages 0 to 5 at 91,450,976 dollars, with 6,461,498 dollars directed to ages 5 to 12.

These figures imply distinct roles for the two instruments. Expansion is the primary lever for infant capacity, since more than 91 percent of added seats and more than 93 percent of spending are directed to ages 0 to 5. Construction provides a more balanced age mix, with about 40 percent of cost directed to ages 0 to 5 and a higher average cost per slot. Consistent with this pattern, ZIP codes such as 10001, 12211, 12566, 12569, and 12570 allocate the entirety of expansion spending to infant slots, while ZIP codes with lower infant shares rely more on construction to broaden overall capacity.

4.2.2 Integrated Facility Location and Capacity Expansion Model

The *Integrated Facility Location and Capacity Expansion Model* delivers 1,150,870 new slots statewide, composed of 1,110,800 from new construction and 40,070 from expansions. Construction totals 403,874,400 dollars for an average of 363.6 dollars per slot, with 507,994 infant slots, equal to 45.7 percent of construction capacity. Construction spending allocates 179,750,124 dollars to ages 0 to 5, equal to 44.5 percent, and 224,124,277 dollars to ages 5 to 12, equal to 55.5 percent. Expansion totals 20,067,960 dollars for an average of 500.8 dollars per slot, with 38,918 infant slots, equal to 97.1 percent of expansion capacity. Expansion spending allocates 19,554,566 dollars to ages 0 to 5, equal to 97.4 percent, and 513,394 dollars to ages 5 to 12, equal to 2.6 percent.

These outcomes indicate distinct roles for the two instruments. Expansions are a targeted mechanism for rapidly adding infant capacity, capturing nearly all expansion slots and spending in the 0 to 5 group, while exhibiting a higher marginal cost per slot. Construction supplies a broader age mix at a lower average cost per slot in this model, consistent with the idea that new builds scale mixed capacity more efficiently once siting and size can be optimized jointly. ZIP level patterns reinforce this interpretation. For example, 10001, 12154, and 12144 allocate all expansion costs to infant slots, whereas ZIP codes with lower infant shares rely more on construction, suggesting local constraints or strategic emphasis on new facilities when broader capacity is required.

4.3 Model Comparison

Adding siting, size choice, distance separation, and segmented expansion limits raises cost and shifts the mix toward construction. The objective increases from 255,138,373.93 dollars in the *Baseline Budget Allocation Model* to 423,942,065.87 dollars in the *Integrated Facility Location and Capacity Expansion Model*, a rise of 168,803,691.94 dollars or 66.16 percent. Average cost per slot increases from 313.8 dollars to 368.4 dollars.

Spending composition moves from 61.6 percent construction and 38.4 percent expansion in the baseline to 95.3 percent construction and 4.7 percent expansion in the integrated model. Unit costs diverge by instrument. Construction becomes cheaper per slot in the integrated model, 363.6 dollars versus 401.8 dollars, reflecting better matching of size and location to demand. Expansion becomes more expensive, 500.8 dollars

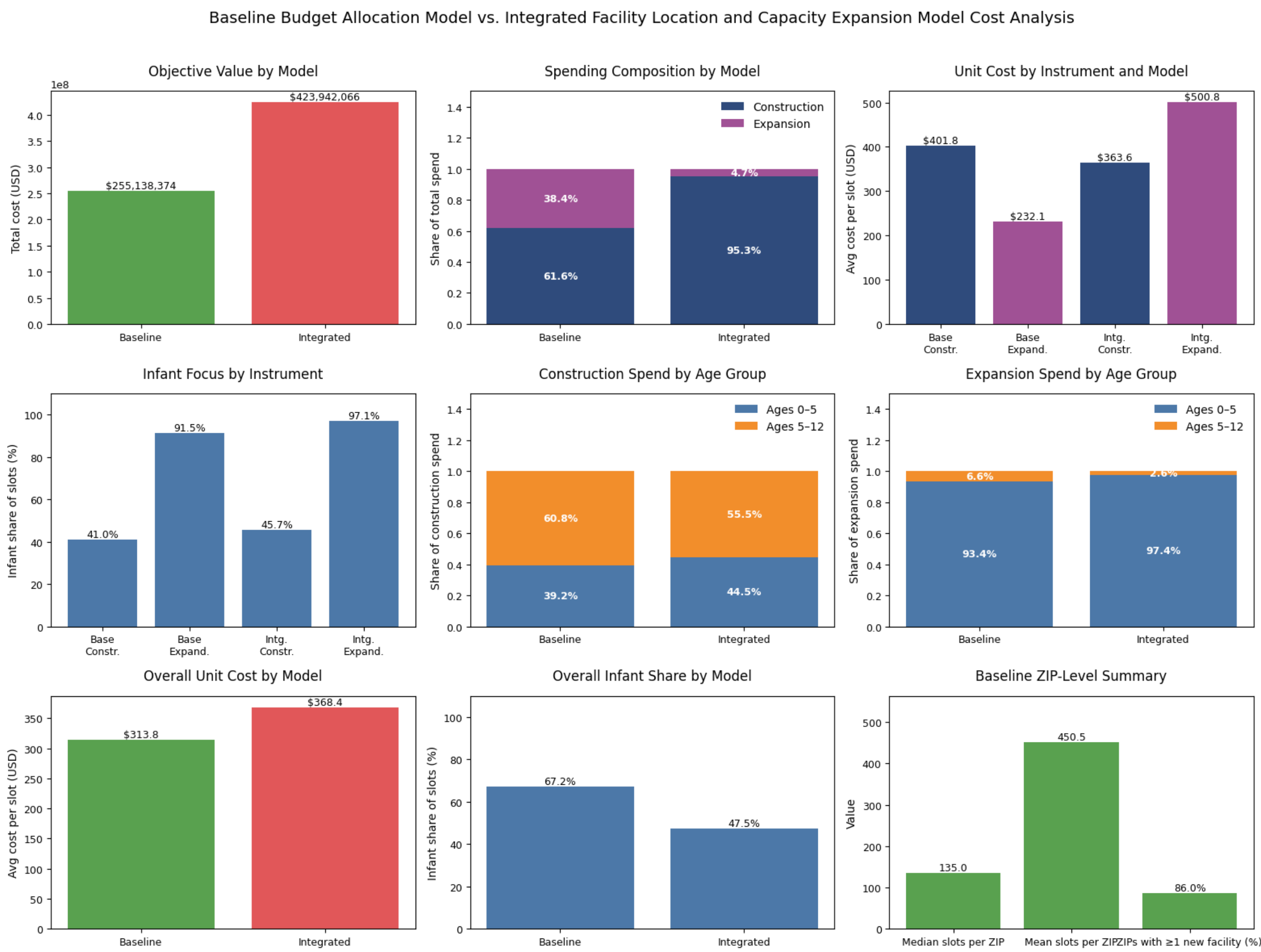

Figure 2: Baseline Budget Allocation Model vs. Integrated Facility Location and Capacity Expansion Model Cost Comparison

versus 232.1 dollars, due to segment caps and reduced scope for large, low cost increments at incumbents.

Age mix adjusts accordingly. The baseline delivers 67.2 percent of slots for ages 0 to 5 overall, driven by infant focused expansions. The integrated model delivers 47.5 percent infant slots overall because capacity is added mainly through construction, which carries a 45.7 percent infant share. At the ZIP level both models direct expansions to infant shortfalls, while areas needing broader capacity rely more on construction. Both models meet the desert and 0 to 5 coverage requirements, but the integrated formulation trades higher total cost for implementability and a more balanced capacity profile.

Figure 2 compares total cost, spending composition, and unit-cost metrics between the baseline and integrated models, highlighting differences in construction vs. expansion spending and the infant-focused share of capacity.

5 Conclusion

This study develops and evaluates two policy-ready optimization frameworks to eliminate child care deserts across New York State while meeting an explicit access guarantee for children ages 0 to 5. The *Baseline Budget Allocation Model* demonstrates that deserts can be removed at a minimum cost of 255,138,373.93 dollars by combining new construction and facility expansion, yielding 813,087 additional slots with broad geographic reach. The *Integrated Facility Location and Capacity Expansion Model* adds real-world feasibility through chosen-site construction, distance separation, and segmented expansion limits. These features raise the minimum cost to 423,942,065.87 dollars but deliver 1,150,870 slots and a capacity mix that is more spatially coherent and implementable. Across both models, expansions are the high-leverage mechanism for rapidly increasing infant capacity, whereas construction supplies a balanced age mix and scales coverage where broader capacity is required. All solutions pass integrity checks on targeting, post-intervention coverage, and cost ordering, confirming feasibility and internal consistency.

The results provide several policy insights. First, eliminating deserts at scale is achievable under transparent rules that tie investments to local population needs and under-5 requirements. Second, the choice between expansion and construction should be context specific: expansions efficiently relieve infant shortfalls at incumbent sites, while construction is preferable when mixed capacity and spatial dispersion are priorities. Third, siting and size choices materially affect budget needs and unit costs, indicating that implementability constraints should be included early in statewide planning to avoid underestimating resources.

The analysis has limitations that suggest directions for future work. The models are static and deterministic, assume separable costs, use Euclidean distance as a proxy for proximity, and do not capture price responses, staffing constraints, or within-ZIP travel frictions. Extending the framework to incorporate demand uncertainty, phased multi-year investments, workforce and operating capacity, and richer accessibility metrics would sharpen cost forecasts and equity outcomes. Embedding distributional objectives, such as explicit fairness or travel-time thresholds, and conducting scenario-based sensitivity analysis on key parameters would further support robust public budgeting and siting decisions.

Code and Data Availability

All code and data required to reproduce this work are publicly available in the GitHub repository: `github.com/JiaheLing/ChildCareDeserts`.

References

Blanco, V., & Gázquez, R. (2023). Fairness in maximal covering location problems. *Computers & Operations Research*, *157*, 106287. https://doi.org/10.1016/j.cor.2023.106287

CAP. (2020). America's child care deserts in 2018 [Center for american progress].

Kaya, Y. B. (2025, September). Project 1: Eliminating child care deserts in new york state [Instructor: Y. B. Kaya. PDF: IEOR4004_Fall25_Project1.pdf].

Li, X., Wang, F., & Yi, H. (2017). A two-step approach to planning new facilities towards equal accessibility. *Environment and Planning B: Urban Analytics and City Science*, *44*(6), 994–1011. https://doi.org/10.1177/0265813516657083

Lovejoy, A. (2024). States are taking action to address the child care crisis. *Center for American Progress*.

Luna Rojas, V. M., & Flores de la Mota, I. (2022). A multiobjective location-allocation model for day-care facilities planning. *Proceedings of the 34th European Modeling & Simulation Symposium (EMSS 2022)*, 040. https://doi.org/10.46354/i3m.2022.emss.040

Ofsted. (2025, May 14). Main findings: How childcare could be optimised across local areas [GOV.UK.]. https://www.gov.uk/government/publications/how-childcare-could-be-optimised-across-local-areas/main-findings-how-childcare-could-be-optimised-across-local-areas

A Appendix: Visualizations for The Problem of Budgeting

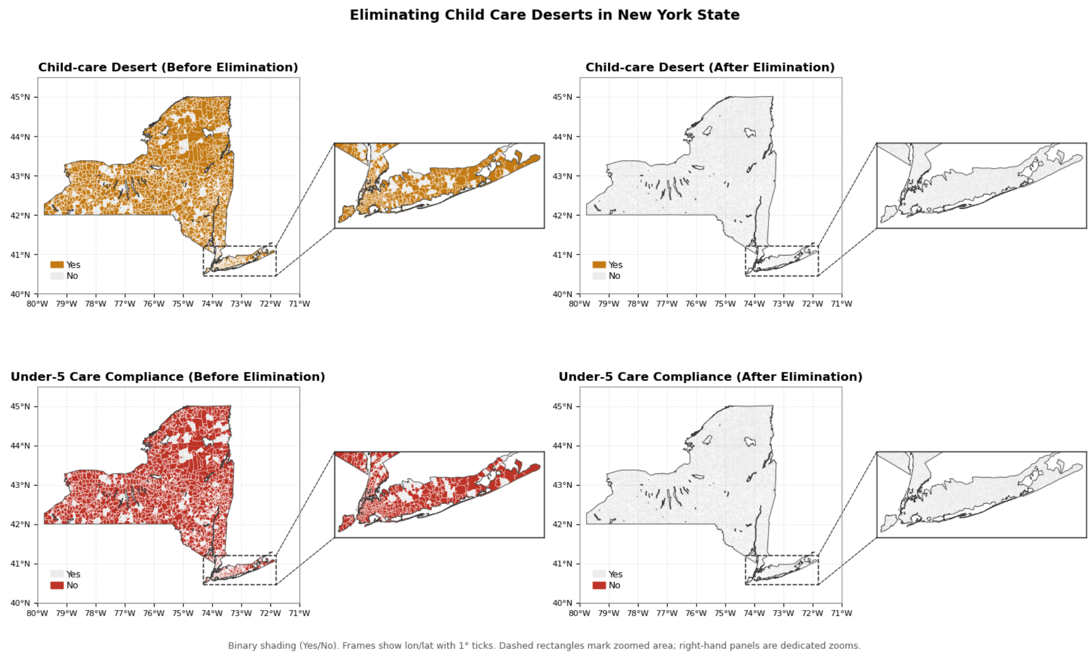

Figure 3: Child-care deserts and under-5 compliance (before & after).

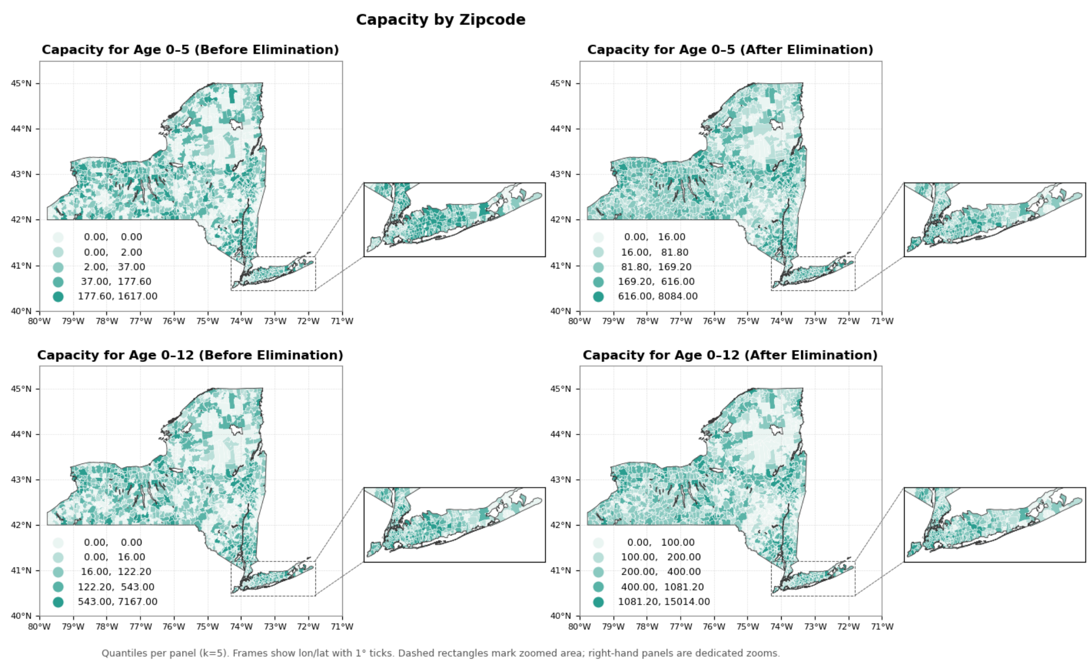

Figure 4: ZIP-level capacity (quantiles) for ages 0–5 and 0–12 (before & after).

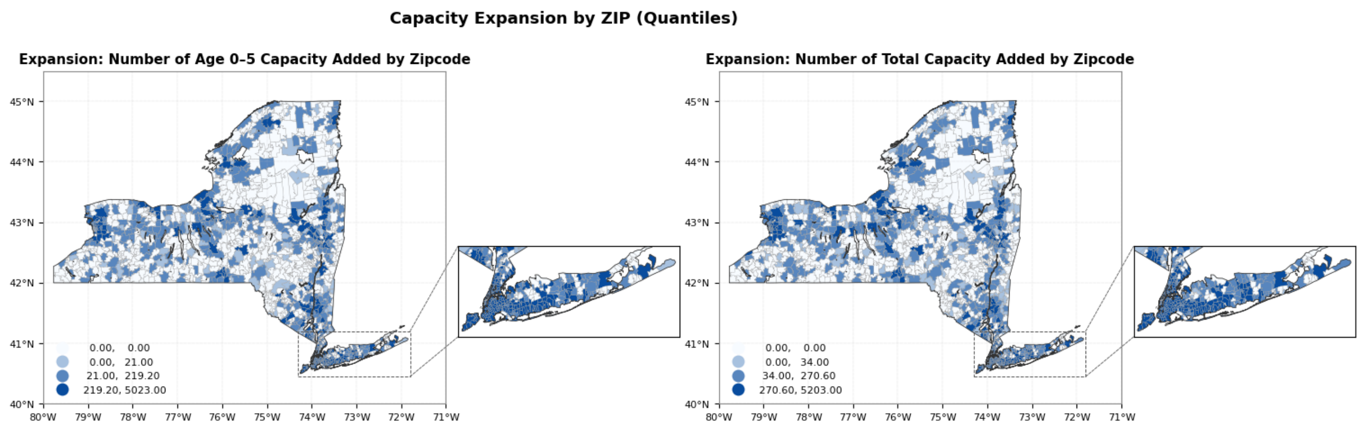

Figure 5: Expansion: number of capacity slots added by ZIP (quantiles).

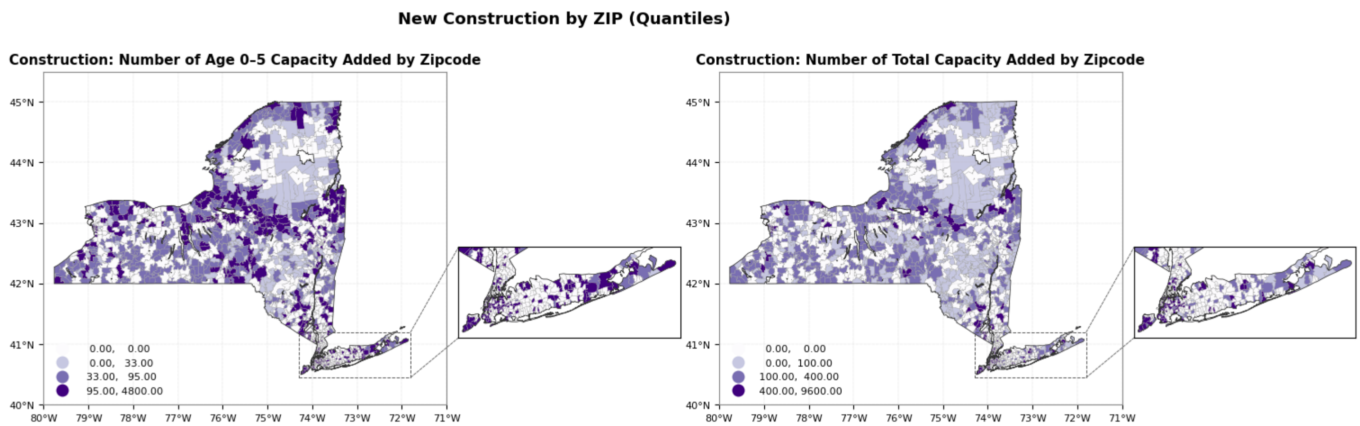

Figure 6: Construction: number of capacity slots added by ZIP (quantiles).

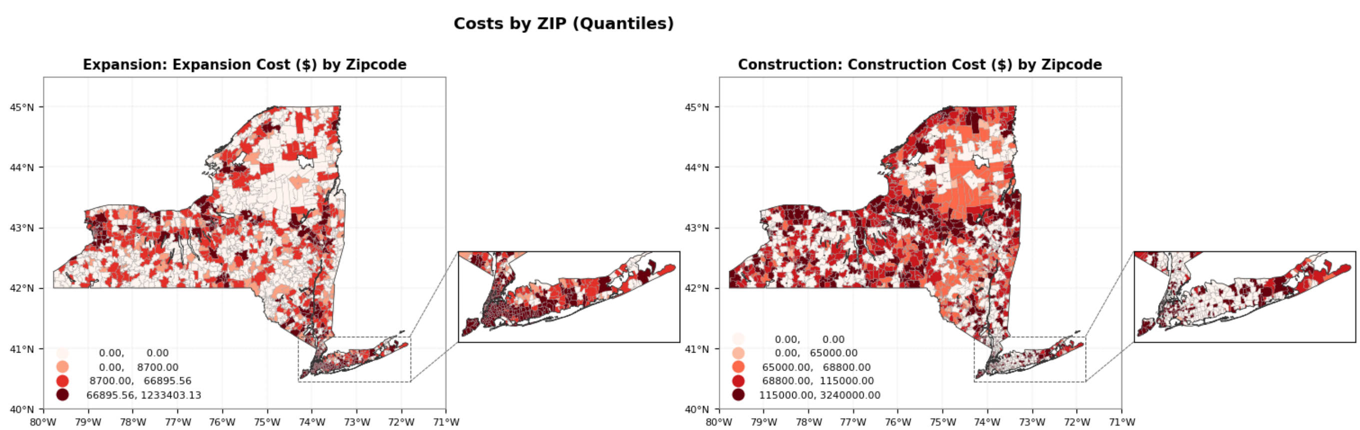

Figure 7: Costs by ZIP (quantiles): expansion and construction cost distributions.

B Appendix: Visualizations for The Problem of Realistic Capacity Expansion and Location

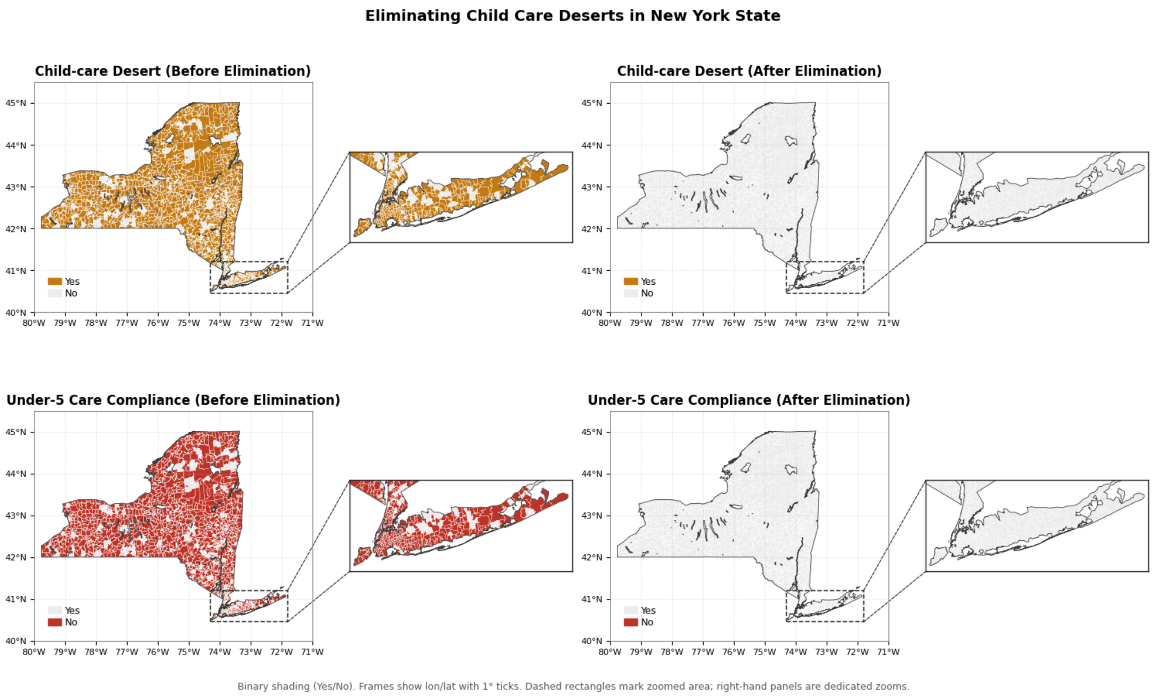

Figure 8: Child-care deserts and under-5 compliance (before & after).

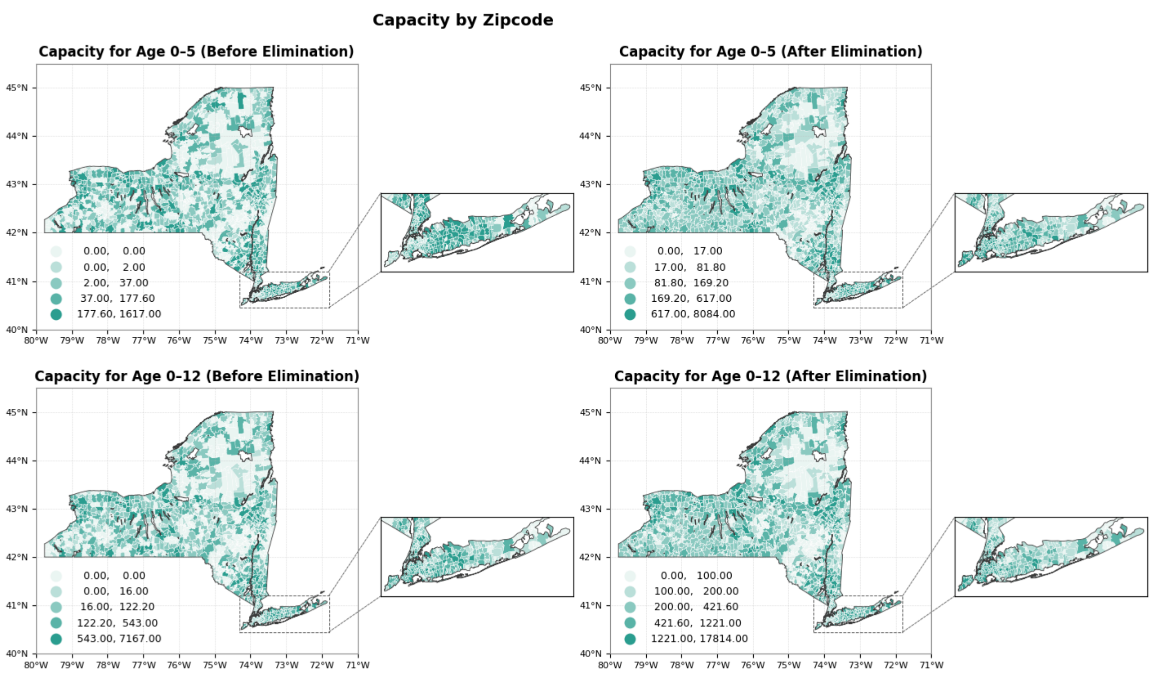

Figure 9: ZIP-level capacity (quantiles) for ages 0–5 and 0–12 (before & after).

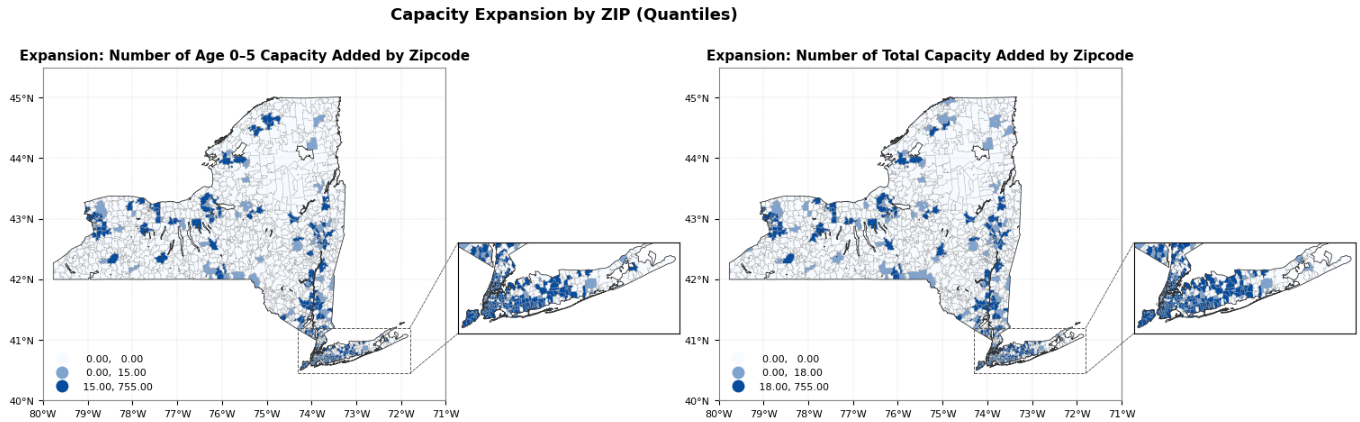

Figure 10: Expansion: number of capacity slots added by ZIP (quantiles).

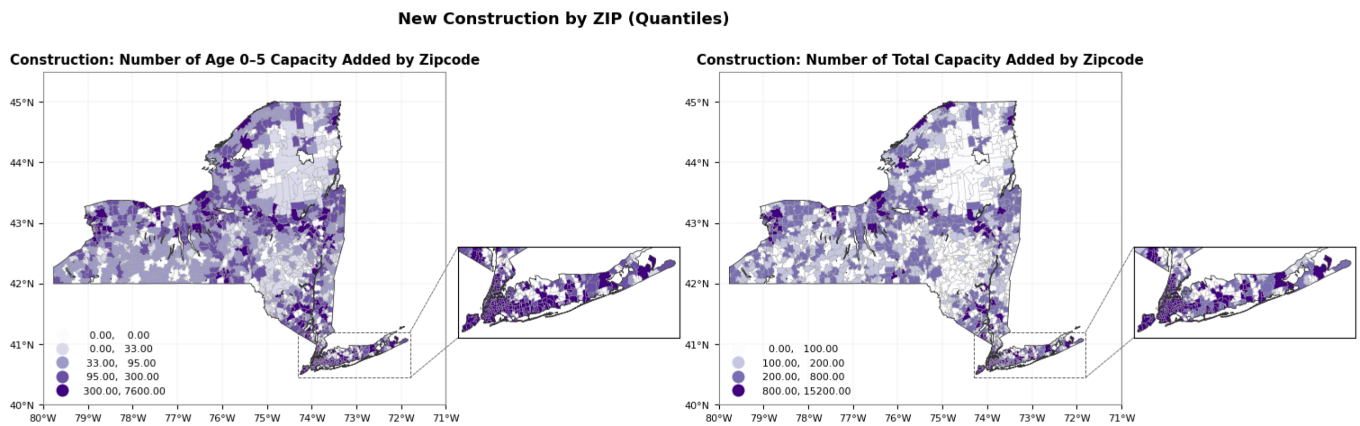

Figure 11: Construction: number of capacity slots added by ZIP (quantiles).

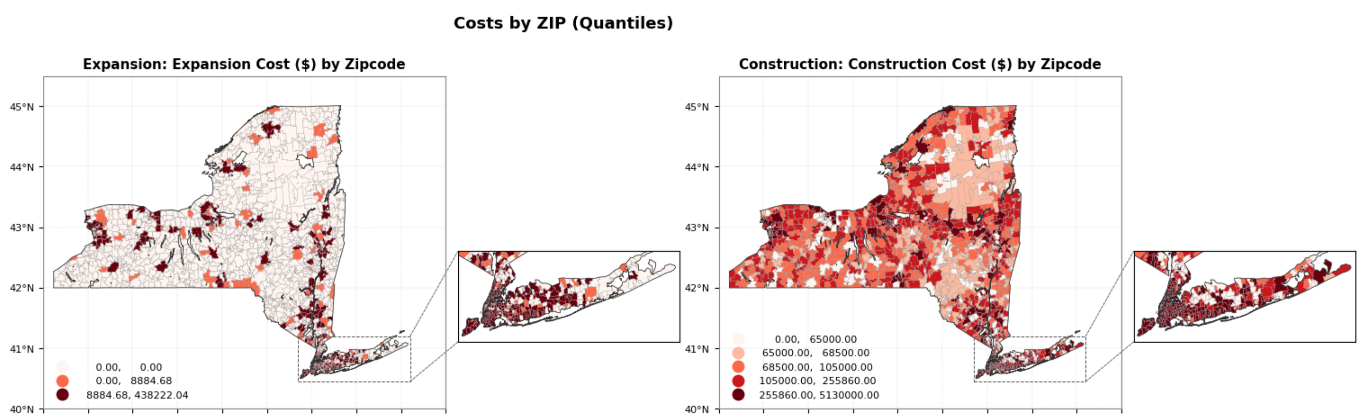

Figure 12: Costs by ZIP (quantiles): expansion and construction cost distributions.

C Appendix: Equity in Child Care Access

[Assumptions] The appendix model operates at the ZIP code level with known populations by age group and known facility characteristics. New construction is allowed in any ZIP and may use only three standardized sizes with location-invariant costs and capacities. Each new facility has a hard cap on the number of slots that can be reserved for ages 0 to 5. Expansions at existing facilities are limited to the smaller of 120 percent of current capacity and 500 slots, and added 0 to 5 slots cannot exceed total added slots at a site. Expansion cost has two parts. A pro rata activation charge of 20,000 dollars is spread across the first tranche of added slots up to the current capacity, and a marginal charge of 200 dollars applies to each slot added beyond the current capacity. Assigning an additional 0 to 5 slot, whether through construction or expansion, adds 100 dollars. Decisions are integer valued, costs are separable across facilities, and the planning horizon is single period.

Table 2: Parameters and Decision Variables

Parameters	
I	Set of ZIP codes.
S	Facility sizes $S = \{\text{Small, Medium, Large}\}$.
F_i	Set of existing facilities in ZIP i.
α_i	Desert threshold factor: $1/2$ if ZIP i is high demand, else $1/3$.
P_i	Population ages 0–12 in ZIP i.
$P_i^{(0\text{–}5)}$	Population ages 0–5 in ZIP i.
n_f	Current number of slots at facility f.
$n_f^{(0\text{–}5)}$	Current number of 0–5 slots at facility f.
K_s	Build cost of size s, $K_s = \{65{,}000,\ 95{,}000,\ 115{,}000\}$.
C_s	Capacity of size s, $C_s = \{100,\ 200,\ 400\}$.
$C_s^{(0\text{–}5)}$	Maximum 0–5 slots for size s, $C_s^{(0\text{–}5)} = \{50,\ 100,\ 200\}$.
Decision variables	
x_f	Added slots at existing facility f.
$x_f^{(0\text{–}5)}$	Added 0–5 slots at existing facility f.
$y_{i,s}$	Number of new facilities of size s built in ZIP i.
$z_{i,s}$	Number of 0–5 slots in new facilities of size s in ZIP i.
$u_f \in \{0, 1\}$	Indicator that facility f expands.
v_f	Slots added above current capacity at facility f, $v_f = \max\{0,\ x_f - n_f\}$.

[Objective and constraints] The objective maximizes a weighted access index that combines the post-investment coverage rate for ages 0 to 5 with weight two thirds and the coverage rate for ages 5 to 12 with weight one third. New facilities contribute 0 to 5 slots directly through the assignment decision, and the remaining capacity contributes to ages 5 to 12. The budget constraint limits total spending to 100 million dollars and accounts for construction costs, the 100-dollar per 0 to 5 slot assignment in new builds, the 100-dollar per expanded 0 to 5 slot at incumbents, the pro rata 20,000-dollar activation charge tied to added slots up to current capacity, and the 200-dollar marginal cost for slots beyond current capacity. A fairness constraint bounds differences in total coverage across ZIP codes by at most 0.1. Coverage requirements enforce no desert status for overall capacity and require at least two thirds coverage for ages 0 to 5. Additional constraints cap 0 to 5 assignments in new facilities, limit expansions at existing sites, link expansions to the activation indicator, and impose nonnegativity and integrality on all decisions.

$$\max \sum_{i\in I}\left[\frac{2}{3}\cdot\frac{\sum_{f\in F_i}(n_f^{(0–5)}+x_f^{(0–5)})+\sum_{s\in S} z_{i,s}}{P_i^{(0–5)}}+\frac{1}{3}\cdot\frac{\sum_{f\in F_i}(n_f-n_f^{(0–5)}+x_f-x_f^{(0–5)})+\sum_{s\in S}(C_s-z_{i,s})\,y_{i,s}}{P_i-P_i^{(0–5)}}\right]$$

$$\text{s.t.}\quad \sum_{i\in I}\sum_{s\in S}\left(K_s y_{i,s}+100\,z_{i,s}\right)+\sum_{i\in I}\sum_{f\in F_i}\left(100\,x_f^{(0–5)}+20{,}000\,u_f\frac{x_f-v_f}{n_f}+200\,v_f\right)\le 10^8 \quad \text{(Budget cap)}$$

$$\left|\frac{\sum_{f\in F_i}(n_f+x_f)+\sum_{s\in S}C_s y_{i,s}}{P_i}-\frac{\sum_{f\in F_j}(n_f+x_f)+\sum_{s\in S}C_s y_{j,s}}{P_j}\right|\le 0.1,\ \forall i,j\in I \quad \text{(Equity bound)}$$

$$\sum_{f\in F_i}(n_f+x_f)+\sum_{s\in S}C_s y_{i,s}\ge \alpha_i P_i,\ \forall i\in I \quad \text{(Desert rule)}$$

$$\sum_{f\in F_i}\left(n_f^{(0–5)}+x_f^{(0–5)}\right)+\sum_{s\in S}z_{i,s}\ge \tfrac{2}{3}P_i^{(0–5)},\ \forall i\in I \quad \text{(Under 5 coverage)}$$

$$0\le z_{i,s}\le C_s^{(0–5)}y_{i,s},\ \forall i\in I,\ \forall s\in S \quad \text{(New-build cap)}$$

$$0\le x_f\le \min\{1.2\,n_f,500\},\ \ 0\le x_f^{(0–5)}\le x_f,\ \forall f\in\bigcup_i F_i \quad \text{(Expansion cap)}$$

$$x_f\le 1.2\,n_f\,u_f,\ \ x_f\le 500\,u_f,\ \ v_f\ge x_f-n_f,\ \ v_f\le x_f,\ \forall f \quad \text{(Linking)}$$

$$y_{i,s},z_{i,s},x_f,x_f^{(0–5)},v_f\in\mathbb{Z}_{\ge 0},\ \ u_f\in\{0,1\} \quad \text{(Integrality)}$$