

Allocation of Pre-Kindergarten Seats in New York City

Ravi Shroff, Richard Dunks, Jeongki Lim, Haozhe Wang and Miguel Castro

{rs4606, rad416, jl2684, hw1067, mac1077}@nyu.edu

Center for Urban Science and Progress, 1 Metrotech Center, 19th floor, Brooklyn, NY, 11201

Abstract

We consider the problem of identifying locations in New York City that are currently underserved with respect to access to pre-Kindergarten programs. We use two public datasets; the spatial distribution of four-year-olds, and the distribution and seating capacities of pre-Kindergarten programs in public schools and community based organizations. We implement a random allocation algorithm to identify and map underserved locations, then see how these locations change as capacity is added in a random fashion. Our model incorporates travel distance, and we measure the sensitivity of our results to variations in this parameter. We provide evidence that as the pre-Kindergarten capacity in our model increases, the effectiveness of this capacity - as measured by the number of unused seats - decreases, to the extent that when the total capacity in the city equals the number of children, almost 20,000 seats remain unused.

The implementation of a “Universal” full day pre-Kindergarten program has become an important topic in New York City and is a central concern of Mayor Bill De Blasio’s administration. Universal pre-Kindergarten (pre-K) means that every four-year-old child should have free access to 180 full days of instruction, centered around state pre-K standards. Studies have shown numerous benefits from pre-K instruction, including increased cognitive abilities, higher test scores in the short term, and access to higher paying jobs in the long term (Heuvel 2013). Besides ongoing debates on the financing of Universal pre-K in New York, there is a shortage of pre-K slots to accommodate every applicant. There is also no clear guideline to assess the need for pre-K programs in different school districts and neighborhoods so as to most effectively add new slots. The purpose of this paper is to use publicly available data to model the underserved regions of New York City with regards to pre-K accessibility. We display our results visually and examine the effects of changing model parameters to simulate additional capacity and variable travel distance. Finally, we discuss alternative approaches and extensions to our model.

Currently, public pre-K programs in New York City are offered in public schools and community based organizations (CBOs). Programs are half day (two hours and 30 minutes of instruction) or full day (six hours and 20 minutes

of instruction). Application processes differ for programs in public schools versus those in CBOs; whereas admission to CBO programs is first come, first served, admission to public school programs considers all applications submitted before the deadline and decides placement based on a list of Admissions Priorities. We now present the existing figures on pre-K demand and capacity as given by the Mayor’s Office, NYC Department of Education (DOE), and others (“Office of the Mayor” 2014).

- **73,250 children** require access to full day pre-Kindergarten programs. This number is derived by taking the 81,748 children enrolled in Kindergarten and subtracting the estimated 8498 children who will enroll in private full day pre-K programs. As explained below, this figure is substantially different than the estimated population of 105,000 four-year-olds in New York City based on census data.
- **58,528 pre-K seats** are currently available. This includes 26,364 half day seats and 32,164 full day seats. Of all currently available seats, 23,671 are in public schools and 34,857 are in CBOs, although the full day seats are almost evenly distributed between public schools and CBOs.
- hence, roughly **41,000 full day seats** need to be added.

The planned implementation of Universal pre-K will occur over a two year period.

- in the 2014-2015 school year, **23,640 full day seats** will be added, split almost evenly between half day conversions and new seats.
- in the 2015-2016 school year, **17,446 full day seats** will be added, with the vast majority being half day conversions.

There are ongoing discussions about where new full day seats *can* be located. The administration of NYC is searching for space not only in existing public schools and CBOs, but also in other city-owned buildings. There are independent discussions of implementation details such as teacher training, program standardization and support for speakers of English as a second language and students with disabilities. We seek to contribute to the question of where new seats *should* be located by determining which locations in the city have low pre-K accessibility, using only public data on population and existing capacity.

Data

We use two publicly available datasets in our model; the spatial distribution of the four-year-old population and the locations and capacities of pre-Kindergarten facilities in New York City.

The distribution of four-year-olds was derived from the 2012 five-year aggregated American Community Survey (ACS). The ACS groups the population into various age bands for reporting purposes. To estimate the number of four-year-olds living in a particular census tract, the population under age five was assumed to be evenly distributed among children aged zero to four, and divided by five to provide the population of four-year-olds. Note that this method gives a total four year old population of roughly 105,000 in NYC, significantly higher than the population estimated by the Mayor’s office and DOE. We will use “child” and “four-year-old” interchangeably.

The locations of pre-K facilities in public schools were gathered from the PEDIcities open data portal (PED 2013), then capacity information was extracted from NYC’s open data portal (DOE 2013). Of the 1,406 pre-K sites listed by the DOE, we removed 29 sites which no longer have Pre-K seats for the 2013-2014 school year.

The locations of the roughly 850 pre-K facilities in CBOs were obtained from both the PEDIcities open data portal and a report from the Mayor’s office (“Department of Education” 2014). The six sites which appear to no longer host pre-K seats in the PEDIcities dataset have been removed from the 855 listed. We were unable to obtain precise CBO capacity information, except for the programs at two sites, Baychester Academy and the Staten Island School of Civic Leadership, with respective reported capacities of 36 and 18 students. Subtracting these 54 seats from the total CBO capacity of 34,857 (listed in the Mayor’s report) gives an estimated 41 seats per facility. Hence, each CBO for which the actual capacity is unknown has been assigned a capacity of 41.

For the purposes of this analysis, we subdivided the city into 2930 equal regular hexagons of side length roughly 350 meters. This created a common spatial unit of analysis, allowing data to be aggregated as necessary to compare capacity within various political, administrative, and other delineations. The size was chosen to provide compatibility with transit time data, to be incorporated into future work. The population of four-year-olds in a particular hexagon was determined by the percent overlap between the hexagon each census tract as in (Maantay, Maroko, and Herrmann 2007). The census tracts used here are those released by the NYC Department of City Planning (DCP). These are clipped to the shoreline and provide a more reliable estimate of the area of the census tract in which population actually resides than tract boundaries provided by the Census Bureau. The pre-K capacities were summed for all sites located within a hexagon, with separate figures for public school and CBO capacities. No distinction was made between part time and full time seats, thus a site with 18 morning seats and 18 afternoon seats would have a capacity of 36. This is consistent with how DOE reports capacity figures in their annual report. Note that conversion of part time pre-K seats to full

time may actually decrease listed capacity. Finally, each hexagon was also assigned to the school district with which it overlapped most.

Algorithm

Our allocation algorithm (Algorithm 1, below) uses Monte Carlo methods to determine areas of New York City underserved in terms of pre-Kindergarten program access. The input consists of the geographic distribution of four-year-olds and current public school and CBO pre-K capacities for each of the 2930 hexagons described above. We simulated the ability of parents to take their children to nearby pre-K seats by creating for each hexagon H a list of “nearby” hexagons, defined to be those hexagons overlapping a disc of fixed radius (the “travel distance”) centered at the given hexagon. We assume that although a child in H may attend any CBO pre-K program in any nearby hexagon, he or she may only attend those public school pre-K programs in nearby hexagons in the same school district as H . This assumption is consistent with public school pre-K admissions criteria, which give strong preference to an applicant whose residence is in the same school district as the target school.

Each hexagon H has resident population of children P_H , current number of assigned public school and CBO students S_H^{ps} and S_H^{cbo} respectively, public school and CBO capacities C_H^{ps} and C_H^{cbo} respectively, a list of nearby hexagons $\mathbf{F}_H = \{F_1, \dots, F_{j_H}\}$ in the same school district as H , and a list of all nearby hexagons $\mathbf{G}_H = \{G_1, \dots, G_{k_H}\}$ (so $\mathbf{F}_H \subseteq \mathbf{G}_H$). Initially, S_H^{ps} and S_H^{cbo} are set to zero and P_H is set to P_H^0 , the initial resident population.

We say any hexagon K is a *non-full neighbor* of H if $K \in \mathbf{F}_H$ and $S_K^{ps} < C_K^{ps}$ or $K \in \mathbf{G}_H$ and $S_K^{cbo} < C_K^{cbo}$. We say H is *usable* if it has at least one child ($P_H > 0$), and at least one non-full neighbor.

The algorithm then works as follows. While there exists at least one usable hexagon, randomly choose such a usable H and randomly choose a non-full nearby hexagon K from either \mathbf{F}_H or \mathbf{G}_H . Decrement H ’s resident population P_H by 1 and if K was chosen from \mathbf{F}_H , increment K ’s assigned public school students S_K^{ps} by 1, otherwise increment K ’s assigned CBO students S_K^{cbo} by 1.

When the algorithm terminates, for each hexagon H compute the output statistic P_H/P_H^0 , the percentage of resident children in H that were unable to be allocated to a pre-K spot. We chose to run the algorithm six times due to computational limitations and average the output statistics. Finally, we visualized the averaged output statistic by creating a map where each hexagon was shaded from black to light grey based on the output statistic (black means all children were unallocated and light grey means all children were allocated to pre-K seats). Hexagons that did not possess children to begin with are white.

Algorithm 1

Input: Hexagons $\mathcal{H} = \{H_n\}_{n=1}^{2930}$
 $H_n = \{P_{H_n}, S_{H_n}^{ps}, S_{H_n}^{cbo}, C_{H_n}^{ps}, C_{H_n}^{cbo}, \mathbf{F}_{H_n}, \mathbf{G}_{H_n}\}$
Initialize parameters:
 $P_{H_n} = P_{H_n}^0 \{\forall n\}$
 $S_{H_n}^{ps} = 0 \{\forall n\}$
 $S_{H_n}^{cbo} = 0 \{\forall n\}$
 $\mathcal{X} = [H \in \mathcal{H} \text{ s.t. } H \text{ usable}]$
while length(\mathcal{X}) $\neq 0$ **do**
 randomly choose $H \in \mathcal{X}$
 randomly choose non-full neighbor K of H from either \mathbf{F}_H or \mathbf{G}_H
 if K chosen from \mathbf{F}_H **then**
 $S_K^{ps} = S_K^{ps} + 1$
 else
 $S_K^{cbo} = S_K^{cbo} + 1$
 end if
 $P_H = P_H - 1$
 $\mathcal{X} = [H \in \mathcal{H} \text{ s.t. } H \text{ usable}]$
end while
Output: Pairs $\{(n, o_n)\}_{n=1}^{2930}$
 where $o_n = \frac{P_{H_n}}{P_{H_n}^0}$ if $P_{H_n}^0 \neq 0$, and $o_n = -1$ otherwise

Results

We ran our algorithm multiple times, varying two parameters, the total capacity and travel distance. Our first map displays the results of our algorithm where we have set the travel distance to two kilometers, roughly how far one may travel by foot or public transport in New York City in 15 minutes. As explained above, this map shades each hexagon from light grey to black based on the percentage of resident children that weren't allocated a pre-K seat (black means a high percentage of children remained unallocated). The school district boundaries are overlaid. Note that in our model there are a total of 105,376 children and 55,558 available seats. Also note that while we have chosen to map the percentages of unallocated children rather than the total number, the two maps are virtually identical.

Figure 1 allows the identification of spatial clusters of hexagons with poor accessibility to pre-K programs. For those familiar with the geography of New York City, many of the “underserved” clusters are in historically low-income areas or areas requiring long transit times to the rest of the city. The results of this simulation yield 49,849 unallocated children, or merely 31 unused pre-Kindergarten seats. Note that the number of unused seats are computed according to

$$\text{unused seats} = \text{total seats} - \text{allocated children},$$

where “allocated children” is the difference of the total number of children and the number of children unallocated. We next examine the sensitivity of these numbers to varying travel times. In Table 1, we increase travel distance by increments of 200 meters to a maximum travel distance of 3000 meters (roughly how far one may travel in 20 minutes by foot or public transit). We notice that the number of unallocated children and unused seats varies very little, since most



Figure 1: Travel distance of 2000 meters, no added capacity

seats are already occupied, and we expect the number of unused seats to generally decrease as travel distance increases.

Travel distance	Unallocated children (stdev)	Unused seats
2000m	49,849 ($\sigma = 10.29$)	31
2200m	49,587 ($\sigma = 15.39$)	39
2400m	49,851 ($\sigma = 14.27$)	33
2600m	49,844 ($\sigma = 15.84$)	26
2800m	49,840 ($\sigma = 17.92$)	22
3000m	49,846 ($\sigma = 12.80$)	28

Table 1: Varying travel distance, no added capacity

Next, we run the algorithm where we again set travel distance to 2000 meters, but now randomly add 25,000 seats across the city. We pick 250 hexagons at random, increment public school capacity by 100 in 125 of these hexagons, and CBO capacity by 100 in the other 125 hexagons. The resulting map appears in Figure 2. We note that many areas of the map are significantly improved in terms of percentage of children unallocated, though the large clusters in the center and top of the map remain. In this simulation there are 30,937 children unallocated and 6,119 unused seats. Hence roughly 24.4 percent of the added capacity remains unused, even though over 29 percent of children remain unallocated.

In Table 2 we vary travel distance again and see how the number of unallocated children and unused seats varies. We see a slight reduction in both figures (unused seats decrease by about 11 percent) as travel distance increases from 2 kilometers to 3 kilometers. This suggests that a modest increase in travel distance is not enough to significantly re-

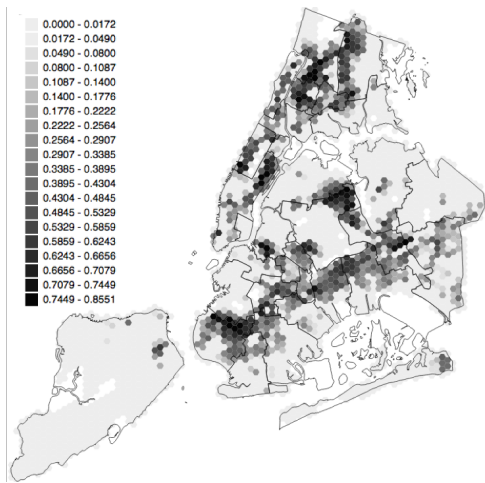


Figure 2: Travel distance of 2000 meters, 25,000 seats added capacity

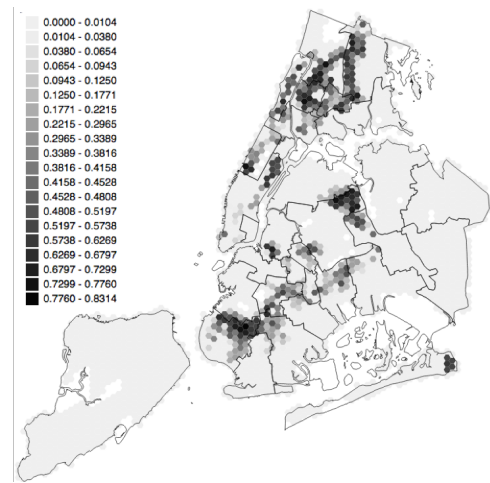


Figure 3: Travel distance of 2000 meters, 50,000 seats added capacity

duce the inefficiencies created by randomly adding 25,000 pre-Kindergarten seats.

Travel distance	Unallocated children (stdev)	Unused seats
2000m	30,937 ($\sigma = 27.27$)	6,119
2200m	30,674 ($\sigma = 40.57$)	5,856
2400m	30,723 ($\sigma = 26.32$)	5,905
2600m	30,364 ($\sigma = 53.84$)	5,546
2800m	30,226 ($\sigma = 13.98$)	5,408
3000m	30,261 ($\sigma = 68.01$)	5,443

Table 2: Varying travel distance, 25,000 seats added capacity

For our final run, the travel distance is again 2000 meters, but we vary capacity by randomly adding 50,000 seats across the city, so the total number of seats is now 105,558, which is 182 more than the 105,376 children in the city. We pick 500 hexagons at random, incrementing public school capacity by 100 in half of these hexagons, and CBO capacity by 100 in the other half. The resulting map appears in Figure 3. Although there is significant improvement in this map as compared to Figures 1 and 2, there are still a number of underserved spatial hexagon clusters. In this simulation there are 19,681 children unallocated and 19,681 unused seats (we ignore the 182 extra seats that would always remain unused). Hence almost 40 percent of the added capacity remains unused, even though 20 percent of children remain unallocated and there are more seats than children.

In Table 3 we vary travel distance and notice that as dis-

tance increases, there is again only a modest reduction in the number of unallocated children and unused seats. This again implies that an increase of 50 percent in travel distance is not nearly enough to reduce the total number of unallocated children to zero, given the 50,000 seats added to the original capacity.

Travel distance	Unallocated children (stdev)	Unused seats
2000m	19,681 ($\sigma = 62.43$)	19,681
2200m	19,200 ($\sigma = 43.95$)	19,200
2400m	19,184 ($\sigma = 31.95$)	19,184
2600m	18,514 ($\sigma = 59.99$)	18,514
2800m	18,232 ($\sigma = 25.99$)	18,232
3000m	18,242 ($\sigma = 57.33$)	18,242

Table 3: Varying travel distance, 50,000 seats added capacity

Discussion and Conclusion

Any strategy for adding pre-K capacity to minimize unallocated children should be measured by its performance against the random assignment benchmark we have constructed in this paper. However, several aspects of this problem make it challenging with respect to implementing a straightforward optimization solution.

First is the difficulty of accurately measuring assumptions and input parameters. For example, it may be incorrect to assume parents seek to minimize the distance between their

home and choice of pre-K facility; they make take their child to a program near their place of work. It is also difficult to measure the capacity of individual CBO programs (the DOE did not have this information readily available) and even the estimated number of children who require access to pre-K in the first place. According to census data, we estimate there are 105,000 four-year-olds but the city reported only 73,250 require a public pre-K program. Even incorporating the city's estimate of 8,498 private pre-K attendees, there is still a difference of over 20,000 children.

Second is the necessary flexibility and robustness of a solution. The selection of locations for additional pre-Kindergarten facilities may occur at different points in time, and change due to policy (e.g. it is politically expedient to place facilities in a certain neighborhood) or unforeseen circumstances (such as a CBO not passing a health inspection). A solution that is very sensitive to small changes in input parameters would not be the best choice to implement in New York.

Third is the issue of scale. With the number of children and existing pre-K locations already in the tens of thousands, an optimization solution may not be computationally cost-effective when parameter measurements are also uncertain.

Additional complications when building a more detailed model of pre-K accessibility include uncertainty in how the spatial distribution and total population of four-year-olds will vary with time, as well as potential switches from private pre-K programs to public pre-K. We would like to incorporate more accurate transit time data to account for differences between bus, subway, and walking time, rather than the simple distance-based estimation in our model. Finally, we plan to incorporate a differential analysis of half day versus full day programs, as children occupying the same CBO for different morning and afternoon programs may be displaced if that CBO only offers full day programs.

We do anticipate that the optimal strategy to assign children to programs will incorporate both linear optimization techniques (to model public school program assignment) and stochastic aspects to model CBO program selection. Note that the question under consideration in this paper is closely related to the facility location problem from operations research, specifically the single-source capacitated facility location problem (Holmberg, Rönnqvist, and Yuan 1999). Although we have chosen reasonable input parameters for our model based on public data, with more accurate, not necessarily public data (for example, specific CBO capacities and parent preference information) we could better model the areas of New York City that should be targeted for increased pre-K capacity.

Acknowledgements

The authors would like to acknowledge Dr. Sharad Goel and Professors Manuela Veloso and Theodoros Damoulas for their assistance and stimulating conversations regarding this work.

References

- “Department of Education”. 2014. Pre-kindergarten directory 2014-2015. Technical report.
2013. Enrollment capacity and utilization reports district 1-32. <https://data.cityofnewyork.us/Education/Enrollment-Capacity-and-Utilization-Reports-Distri/my4g-bvvs>.
- Heuvel, K. 2013. Oklahoma is schooling the nation on early education. *The Washington Post*.
- Holmberg, K.; Rönnqvist, M.; and Yuan, D. 1999. An exact algorithm for the capacitated facility location problems with single sourcing. *European Journal of Operational Research* 113(3):544–559.
- Maantay, J.; Maroko, A.; and Herrmann, C. 2007. Mapping population distribution in the urban environment: The cadastral-based expert dasymetric system (ceds). *Cartography and Geographic Information Science* 34(2):77–102.
- “Office of the Mayor”. 2014. Ready to launch: New York city’s implementation plan for free, high-quality, full-day universal pre-kindergarten. Technical report.
2013. Universal pre-k locations. <http://nycdata.pediacities.com/dataset/universal-pre-k-locations/resource/88c86aff-c965-4491-9b04-3b1ddab36009>.