AN APPLICATION OF GALE AND SHAPLEY TO THE CITY OF NEW YORK

ALFRED GALICHON, OCTAVIA GHELFI

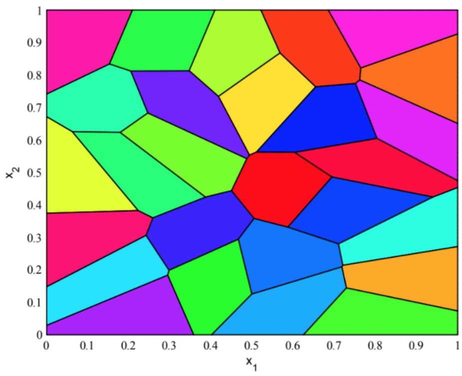
Abstract

We study two possible solutions to a semi-discrete allocation problem by comparing two different spatial allocations: the first one, the efficient allocation, is obtained through the minimization of the total distance traveled by students to reach their school; the second one, the fair allocation, is the result of an application of a generalized version of the Gale and Shapley algorithm. We develop a notion of distance between spatial allocations that allows us to establish a comparison and subsequently study in which cases the two allocations coincide. We then apply the generalized Gale and Shapley algorithm to the many-to-one allocation of kindergarten and elementary school students to public schools in the New York City area and observe that the resulting allocations are stable, but unequal.
[1]

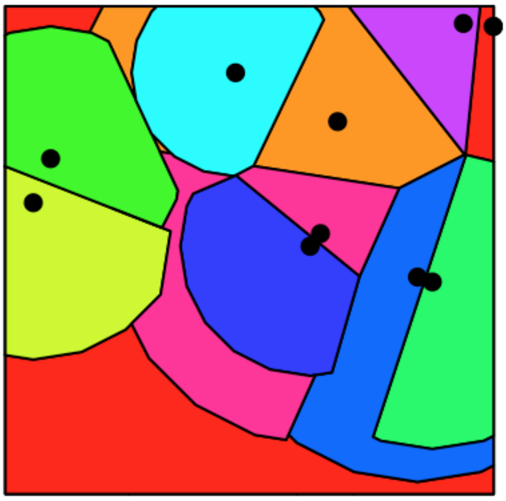
[1] *Date:* June 11th, 2019. Support from NSF grant DMS-1716489 is gratefully acknowledged. Any comments and suggestions are welcome.

1. Introduction

Consider a one-to-many matching problem, with the "many" side of the market consists of infinitely many points. As an example consider a city of dimensions $[0,1] \times [0,1]$ and let the city be populated by students $\mathcal{X}$ with a distribution $\mathcal{P}$ over the territory. For now assume that $\mathcal{P}$ is a uniform distribution, so that the number of students in a given area of the city will simply be the Lebesgue measure of that area. Let $\mathcal{Y} = \{y_1, ..., y_m\}$ be a discrete set of points representing schools in the area, and indicate with $q_j > 0$ the capacity of center j. If we assume $\sum_j q_j = 1$ we will have that each student must be assigned to a school. An allocation mechanism is a map $T : \mathcal{X} \rightrightarrows \mathcal{Y}$ such that student $x \in \mathcal{X}$ is allocated to school $T(x) \in \mathcal{Y}$. The map should satisfy $P\{x\,|T(x) = y_j\} = q_j$, which describes that the number of students that attend school j is q_j. In 1984 Aurenhammer et al. proposed an allocation on the plane based on power diagrams or Laguerre diagrams. We will call such solution the efficient allocation. It is called efficient because such allocation minimizes the sum of total distances traveled by each student to get to the school they are assigned to. The allocation is in fact the solution to a Monge-Kantorovich problem with cost being equal to the Euclidean distance. The resulting cells are convex polytopes as depicted in the following graph.



In their 2006 paper Hoffman et al. propose a different solution to the spatial allocation problem which they call the fair allocation. The solution is the result of a generalized Gale and Shapley algorithm applied to the spatial setting, where preferences both for students and schools are defined by their reciprocal Euclidean distance. In other words it is costly for a school and a student that are far away to match with each other, and this cost is split between them. The allocation is obtained by growing circles around each school until conflicts arise. The geometry of the resulting cells is more complex than the one of the efficient allocation, as cells need not be convex and need not be connected.



In the first case the allocation coincides with a matching in a transferable utilities regime, while the second one is the resulting allocation of a matching algorithm with non transferable utilities. In this paper we study the intermediate case of exponential transferable utilities, which allows us to develop a metric of distance between allocations. We describe in detail the two extreme allocations, which we will call efficient and fair, in the one and two dimensional cases. We will then propose a discretized version of this algorithm, that we can apply to the school student allocation in New York City, using real data on population density and school location. We show some results from both the efficient and fair allocation, applied to kindergartens and elementary schools in the five boroughs of New York. In particular, we show that the fair allocaiton is not really fair. The match resulting from absence of blocking pairs assigns some lucky students to nearby schools, while assigns others to schools in a different borough, that requires them to travel an unreasonably long distance.

1.1. **Exponential Transferable Utility.** Consider the equilibrium in the Exponential Transferable Utility problem

$$\begin{cases} (PF): & \mu \in \mathcal{M}(P,Q) \\ (DF): & \exp\left(\frac{u(x)-\alpha(x,y)}{\tau}\right) + \exp\left(\frac{v(y)-\gamma(y,x)}{\tau}\right) \geq 2 \\ (NC): & (x,y) \in Supp(\pi) \implies \exp\left(\frac{u(x)-\alpha(x,y)}{\tau}\right) + \exp\left(\frac{v(y)-\gamma(y,x)}{\tau}\right) = 2 \end{cases}$$

Notice that when $\tau \to \infty$ we have optimal transport (TU matching) and when $\tau \to 0$ we have Gale and Shapley (NTU matching).

We in particular consider the case when $\alpha(x,y) = \gamma(y,x) = -d^2(x,y)$ where d^2 is the square of the Euclidean distance. The ETU problem then becomes

$$\begin{cases} (PF): & \mu \in \mathcal{M}(P,Q) \\ (DF): & \exp\left(\frac{u(x)+d^2(x,y)}{\tau}\right) + \exp\left(\frac{v(y)+d^2(x,y)}{\tau}\right) \geq 2 \\ (NC): & (x,y) \in Supp(\pi) \implies \exp\left(\frac{u(x)+d^2(x,y)}{\tau}\right) + \exp\left(\frac{v(y)+d^2(x,y)}{\tau}\right) = 2 \end{cases}$$

$$\begin{cases} (PF): & \mu \in \mathcal{M}(P,Q) \\ (DF): & \exp\left(\frac{u(x)}{\tau}\right) + \exp\left(\frac{v(y)}{\tau}\right) \geq 2\exp\left(-\frac{d(x,y)}{\tau}\right) \\ (NC): & (x,y) \in Supp(\pi) \implies \exp\left(\frac{u(x)}{\tau}\right) + \exp\left(\frac{v(y)}{\tau}\right) = 2\exp\left(-\frac{d(x,y)}{\tau}\right) \end{cases}$$

Let $U(x) = \exp\left(\frac{u(x)}{\tau}\right)$, $V(y) = \exp\left(\frac{v(y)}{\tau}\right)$ and let $\phi(x,y) = 2\exp\left(-\frac{d(x,y)}{\tau}\right)$, so that the problem becomes

$$\begin{cases} (PF): & \mu \in \mathcal{M}(P,Q) \\ (DF): & U(x) + V(y) \geq \phi(x,y) \\ (NC): & (x,y) \in Supp(\pi) \implies U(x) + V(y) = \phi(x,y) \end{cases}$$

and one can easily notice that (U,V) that solves the above problem is also the solution to the Monge-Kantorovich problem

$$\min_{U,V} \int U(x)dP(x) + \int V(y)dQ(y)$$
$$s.t.\ U(x) + V(y) \geq 2\exp\left(-\frac{d^2(x,y)}{\tau}\right)$$

Hence the ETU version of the problem with alignment of preferences can be also formulated as an optimization problem. The equilibrium matching in the ETU case solves

$$\max_{\pi \in \mathcal{M}(P,Q)} \int 2\exp\left(-\frac{d^2(x,y)}{\tau}\right) d\mu(x,y)$$

and in particular in the TU case ($\tau \to \infty$) it solves

$$\max_{\pi \in \mathcal{M}(P,Q)} \int \left(-d^2(x,y)\right) d\mu(x,y)$$

in the NTU case ($\tau \to 0$) the solution is a coupling $\mu \in \mathcal{M}(P,Q)$ such that $d(x,y)$ is minimized in lexicographic order, minimizing those with smallest distances first.

2. One Dimensional Case

More in general assume that students are uniformly distributed on the interval $[0,1]$, and there are j schools $\mathcal{Y} = \{y_1, ..., y_j\}$. Site j is located at point $y_j \in [0,1]$ and has capacity q_j. Let the surplus function be

$$\phi_\lambda(x,y) = \frac{\exp\left(-\lambda \left|x-y\right|^2\right) - 1}{\lambda}$$

Assume that the utility function of a student located at x and going to school y_j is

$$\phi_\lambda(x,y) - v_j$$

where v_j is the price charged by site j. The resulting assignment maximizes

$$\max_{\pi \in \mathcal{M}(P,Q)} \mathbb{E}_\pi \left[\frac{\exp\left(-\lambda \left|x-y\right|^2\right)}{\lambda}\right]$$

When $\lambda \to \infty$ the matching tends to the Gale and Shapley matching; in particular the matching will minimize the value of $\left|x-y\right|^2$ in lexicographic order, starting from the lowest one.

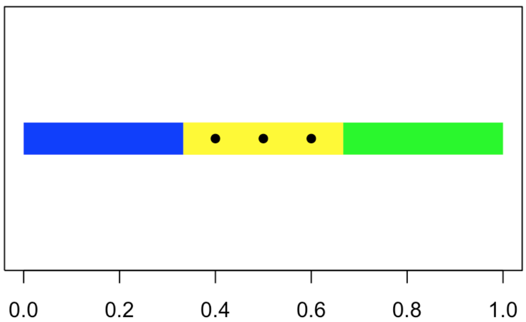
When $\lambda \to 0$ this is the matching that minimizes $\mathbb{E}_\pi\left[\left|x-y\right|^2\right]$ which is the Positive Assortative Matching (PAM) matching, or the matching obtained solving the maximization problem with transferable utilities.

When $\lambda \to -\infty$ the matching will tend to the one that minimizes the value of $\left|x-y\right|^2$ in lexicographic order starting from the highest value. We will call this the bottle neck matching.

In other words we can interpret λ as the degree at which students care about having to pay a fee versus the distance they have to travel. When λ is close to 0 it means that individuals care way more about the value of the transfer than they do about the distance they have to travel. Therefore it is easy to induce the PAM allocation. When $\lambda \to \infty$ students care infinitely more about the distance they travel than they do about the price they pay, so it is impossible to obtain an allocation that is different than the one obtained through Gale and Shapley, which we can interpret as the one that will naturally arise from the market without the intervention of a planner. Lastly when $\lambda \to -\infty$ people care about the distance they travel enormously more than they do about a transfer but the ones that will be at an advantage are the ones that are further away from the centers.

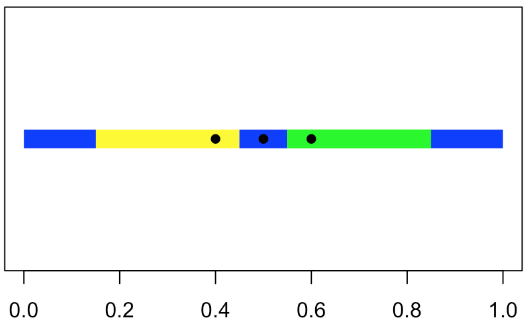
Consider the following example. Let $\mathcal{X} = [0, 1]$ and $\mathcal{P}$ be a uniform distribution. $\mathcal{P}$ being uniform implies that the mass of students that live in the section $[0.1, 0.2]$ of $\mathcal{X}$ is equal to the Lebsegue measure of the section, that is 0.1. Let the schools be located at $\mathcal{Y} = \{0.4,\ 0.5,\ 0.6\}$ and let them have capacity $\frac{1}{3}$ each. The efficient allocation is the PAM allocation, that is

$$T(x) = \begin{cases} 0.4 & if\ x \in \left[0, \frac{1}{3}\right) \\ 0.5 & if\ x \in \left[\frac{1}{3}, \frac{2}{3}\right) \\ 0.6 & if\ x \in \left[\frac{2}{3}, 1\right] \end{cases}$$



The fair allocation instead is the following

$$T(x) = \begin{cases} 0.4 & if\ x \in [0.15, 0.45) \\ 0.5 & if\ x \in [0, 0.15) \cup [0.45, 0.55) \cup [0.85, 1) \\ 0.6 & if\ x \in [0.55, 0.85) \end{cases}$$



Now suppose the centers can charge a price v_k, and let $v_1 = v_3$ and without loss of generality set them both equal to 0 so that the only non zero price is v_2. Suppose $x \in \left[0, \frac{1}{2}\right]$. The ranking for x in this case will be either $1 \succ 2 \succ 3$ or $2 \succ 1 \succ 3$, in either case x will prefer 1 over 3. In particular the utility of going to 1 will be

$$U(x, y_1) = \frac{\exp\left(-\lambda \left|x - y_1\right|^2\right)}{\lambda}$$

The utility of going to center 2 will be

$$U(x, y_2) = \frac{\exp\left(-\lambda \left|x - y_2\right|^2\right)}{\lambda} - v_2$$

x is indifferent between going to center 1 and center 2 when

$$\frac{\exp\left(-\lambda \left|x - y_1\right|^2\right)}{\lambda} = \frac{\exp\left(-\lambda \left|x - y_2\right|^2\right)}{\lambda} - v_2$$

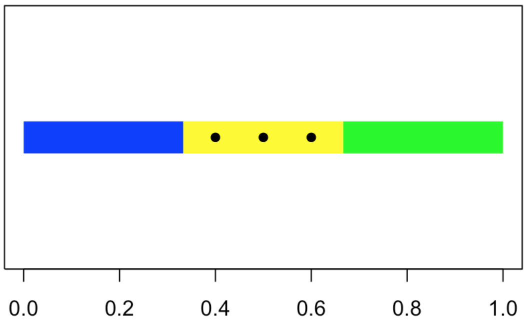
or in other words when

$$v_2 = \frac{\exp\left(-\lambda \left|x - y_2\right|^2\right) - \exp\left(-\lambda \left|x - y_1\right|^2\right)}{\lambda}$$

x will choose 2 over 1 if and only if $\phi_\lambda(x, y_2) - v_2 \geq \phi_\lambda(x, y_1)$. Therefore the equilibrium price will be set so that

$$\int_0^{\frac{1}{2}} \mathbb{I}\left\{\phi_\lambda(x, y_2) - \phi_\lambda(x, y_1) \geq v_2\right\} dt = \frac{1}{6}$$

A v_2 that satisfies the above expression will induce the PAM allocation as depicted in the image below



2.1. **Critical Point.** An interesting question is what is the value of λ at which it becomes welfare maximizing to send $x = 0$ to y_2 and $x = \frac{1}{3}$ to y_1. Or in other words, what is the value of λ such that the assignment the maximizes

$$\max_{\pi \in \mathcal{M}(P,Q)} \mathbb{E}_\pi \left[\frac{\exp\left(-\lambda \left|x-y\right|^2\right)}{\lambda} \right]$$

is the fair allocation rather than the PAM one.

It will be better to have $x = 0$ go to y_2 and $x = \frac{1}{3}$ go to y_1 when the surplus generated in such a way is greater or equal than the one generated in the PAM allocation - where we have $x = 0$ going to y_1 and $x = \frac{1}{3}$ going to y_2. That translates into

$$\frac{\exp\left(-\lambda \left|y_2\right|^2\right)}{\lambda} + \frac{\exp\left(-\lambda \left|y_1 - \frac{1}{3}\right|^2\right)}{\lambda} \geq \frac{\exp\left(-\lambda \left|y_1\right|^2\right)}{\lambda} + \frac{\exp\left(-\lambda \left|y_2 - \frac{1}{3}\right|^2\right)}{\lambda}$$

where again the left handside is the surplus generated by 0 and $\frac{1}{3}$ in the "fair" allocation, and the right handside is the surplus generated in the PAM allocation. Now plugging in the values of $y_1 = 0.4$ and $y_2 = 0.5$ as in our example and simplifying we obtain

$$\exp\left(-\lambda \left|0.5\right|^2\right) + \exp\left(-\lambda \left|0.4 - \frac{1}{3}\right|^2\right) \geq \exp\left(-\lambda \left|0.4\right|^2\right) + \exp\left(-\lambda \left|0.5 - \frac{1}{3}\right|^2\right)$$

$$\exp\left(-\frac{1}{4}\lambda\right) + \exp\left(-\frac{1}{225}\lambda\right) \geq \exp\left(-\frac{4}{25}\lambda\right) + \exp\left(-\frac{1}{36}\lambda\right)$$

That has a numerical solution of $\lambda \approx 7.23355$. So for values of $\lambda \geq 7.23355$ the solution to the optimization problem is the fair allocation.

2.2. **Inducing Positive Assortative Matching with Transfers.** Without prices both $x = 0$ and $x = \frac{1}{3}$ will prefer to go to y_1 . In order to induce $x = \frac{1}{3}$ to go to y_2 instead we will have to compensate him for at least

$$\phi_\lambda(\frac{1}{3}, y_1) - \phi_\lambda(\frac{1}{3}, y_2)$$

Thus the subsidy to go to y_2 should be $v_2 \geq \phi_\lambda(\frac{1}{3}, y_1) - \phi_\lambda(\frac{1}{3}, y_2)$. However we want at the same time that the subsidy is not as high as to induce $x = 0$ to go to y_2, so

$$v_2 < \phi_\lambda(0, y_1) - \phi_\lambda(0, y_2)$$

With the values of y_1 and y_2 that we have the two inequalities are

$$\frac{\exp\left(-\lambda\left|y_1\right|^2\right)}{\lambda} - \frac{\exp\left(-\lambda\left|y_2\right|^2\right)}{\lambda} \geq v_2$$

$$\frac{\exp\left(-\lambda\left|y_1 - \frac{1}{3}\right|^2\right)}{\lambda} - \frac{\exp\left(-\lambda\left|y_2 - \frac{1}{3}\right|^2\right)}{\lambda} \leq v_2$$

Notice how it is possible to induce the above allocation only as long as

$$\frac{\exp\left(-\lambda\left|y_1\right|^2\right)}{\lambda} - \frac{\exp\left(-\lambda\left|y_2\right|^2\right)}{\lambda} \geq \frac{\exp\left(-\lambda\left|y_1 - \frac{1}{3}\right|^2\right)}{\lambda} - \frac{\exp\left(-\lambda\left|y_2 - \frac{1}{3}\right|^2\right)}{\lambda}$$

which leads to $\lambda \approx 7.23355$ as above. So we conclude that for $\lambda \geq 7.23355$ it is not only not optimal to induce PAM, it is also impossible to induce it through transfers. Intuitively, when students care significantly more about the distance they travel rather than the price they pay or the subsidy they get to go to a certain school, it's impossible to induce an allocation different than the one that naturally arises in the market, but that's not a problem as it wouldn't be optimal to induce PAM anyway.

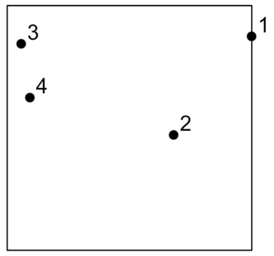
3. Two Dimensional Case

We will now consider the problem of finding both the efficient and the fair allocation in the two dimensional case, that is when we assume that sites are located on the surface of a city, and the continuum of agents represented by the surface itself are facing the decision of what site to go to. In principle we can assume the shape of the city and the distribution of agents on the surface of the city to be of whatever sorts. However, for the sake of a clearer exposition we will assume that the city is a $[0,1] \times [0,1]$ square and that the distribution of agents on it is uniform, so that the mass of agents on an area $A \subset [0,1] \times [0,1]$ will be given by the Lebesgue measure $\mathcal{L}(A)$ of the area. We will initially assume that the sites coordinates are exogenously given, and will determine from there the fair and efficient allocation.

3.1. **Generalized Gale and Shapley Algorithm.** The algorithm that I propose here is a generalization of the Gale and Shapley algorithm with preferences completely described by distance. Agents on both sides of the market want to minimize the distance they travel to reach their counterpart. Let's now consider the algorithm that will lead us to the fair allocation in the two dimensional case. The algorithm will be as follows

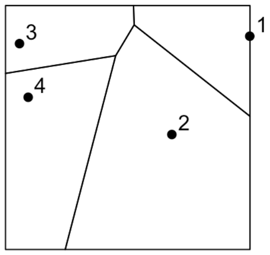
(1) Proposal Phase: students - represented by the continuum of points on the surface of the area - will apply for their favorite school that hasn't rejected them yet.
(2) Disposal Phase: schools will tentatively accept students up to capacity. If a school receives more applications than it can serve, it will reject their least favorite students.
(3) Update Phase: students that have been rejected in the previous step apply to their favorite school that has not rejected them yet.

Example 1. Continuous case. Consider the following application of the algorithm. As mentioned before we will consider a square surface of $[0,1] \times [0,1]$. And consider 4 sites randomly positioned on the territory.

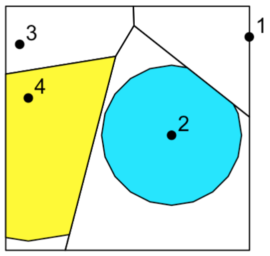


Iteration 1:

Proposal Phase: Students will apply to the closest school. This results in the Voronoi allocation. The lines in the graph represent students that are equidistant between two centers.

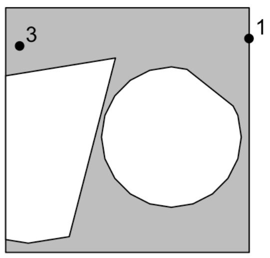


Disposal Phase: Schools that have an excess of demand choose the students that they prefer to admit, up to capacity.

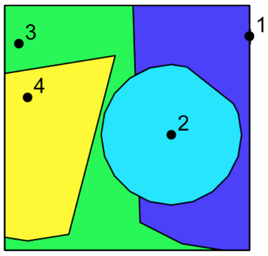


Schools that have an excess of supply tentatively accept all students that have applied. Consider now the rejected territories in grey in the graph.

At the second iteration of the algorithm the students living in these territories will apply again to their favorite school among the ones that have not rejected them at a previous step. Again the schools with an excess of demand will choose among their favorite students, that is the closest ones. Some other students will be rejected and will need to reapply, starting the third iteration of the algorithm and so on, until convergence.



The final allocation in this particular case is the following



3.2. **Efficient Allocation.** We want now to answer a different question, that is which is the allocation that minimizes the total distance traveled by students and respects the capacity constraints of schools. In other words, we want to find an allocation ψ such that

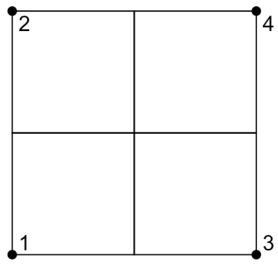
$$\int_{\mathcal{X}} d(x,\psi(x))dx = \min_{\psi} \int_{\mathcal{X}} d(x,y)dx$$
$$s.t. \quad \int_{\mathcal{X}} \mathbb{I}\left\{x\,|\psi(x)=y\right\} = q_y \quad \forall y \in \mathcal{Y}$$

We can obtain this allocation by imposing a system of transfers from students to schools that will make students go to the prescribed school, similarly to the example discussed in the one dimensional case. Here is an example of the algorithm

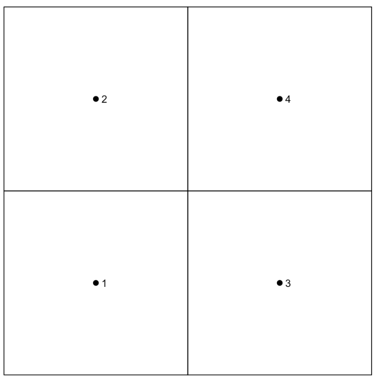
We initially have the Voronoi tasselation, and then imposing an increasingly large transfer t that students have to pay to go to a certain school boundaries will shift parallelly until we obtain the final allocation

Notice that similarly to the one dimensional case certain centers end up being outside of the territory they serve.

3.3. **When Fair and Efficient Allocation Coincide.** Now an interesting question is whether ex-ante we could place centers in such a way that the fair allocation obtained through the Gale and Shapley algorithm and the efficient allocation coincide. The answer is yes, and the location of the sites on the territory will depend on the shape of the territory itself. So for now we can focus on the square territory $[0,1]\times[0,1]$ and try and answer that question for the case at hand. Consider for instance the location of the following points.



It's easy to see that if we calculate the efficient allocation and the fair allocation the two coincide, and as a matter of fact the algorithm for the fair allocation stops after the first iteration. However it is interesting to notice, that there are infinite other possible locations for 4 points on the surface of the city such that the fair and the efficient allocation coincide, and some are better than others. In particular consider the following possible locations that all give rise to the same allocation.



Notice that in the last one the total distance travelled by students is less than the one travelled by students in the second graph, which is in turn less than what they have to travel in the first graph.

3.4. **Bottleneck Algorithm for Inequality Minimization.** Let a labeling of pairs be defined in the following way: $l : \mathcal{X} \times \mathcal{Y} \to \mathbb{N}$, such that if $d(x, y) < d(x, y')$ then $l(x, y) < l(x', y)$. The labeling gives higher ranking to the pair that is most desirable, in the sense that involves the least distance traveled. We know adapt the bottleneck algorithm to this setting, to develop an algorithm of spatial allocation that minimizes inequality.

Consider a labeling $\mathcal{L} = \{1, ..., L\}$ with $1 \succ 2 \succ ... \succ L$.

<u>Threshold strategy:</u>

Algorithm 1: Threshold Algorithm

```
Initialization:
Set l_0 = 1 and l_1 = L
while L* = {l : l_0 < l < l_1} ≠ ∅ do
    Choose l* the median of L*;
    if there exists a feasible matching such that no assigned couple has a
     label above l* then
        l_1 = l*;
    else
        l_0 = l*;
    end
end
```

The smallest value $l^\star$ for which the corresponding problem allows for a perfect match is the value of the Bottleneck problem.

Example 2. Let $I = \{1, 2, 3, 4\}$ and $J = \{a, b, c, d\}$. Consider the following labeling

l	Pairs
1	$(3, a)$
2	$(2, a), (4, a), (1, b)$
3	$(1, c), (1, d), (4, d)$
4	$(3, d)$
5	$(2, c), (4, b)$
6	$(4, c)$
7	$(2, b)$
8	$(2, d), (3, c), (1, a)$
9	$(3, b)$

(1) $l_0 = 1$ and $l_1 = 9$. $L^\star = \{l : 1 < l < 9\} \neq \emptyset$. The median of $L^\star$ is $l^\star = 5$.
Is there a feasible match such that no couple with labeling $l > 5$ is assigned? Yes

$$\{(3, a), (1, b), (2, c), (4, d)\}$$

Therefore $l_1 = 5$

(2) $l_0 = 1$ and $l_1 = 5$, $L^\star = \{l : 1 < l < 5\} \neq \emptyset$. The median of $L^\star$ is $l^\star = 3$.
Is there a feasible match such that no couple with labeling $l > 3$ is assigned? No, we should assign both b and c to 1 which is not possible
Therefore $l_0 = 3$

(3) $l_0 = 3$ and $l_1 = 5$, $L^\star = \{l : 3 < l < 5\} \neq \emptyset$. The median of $L^\star$ is $l^\star = 4$.
Is there a feasible match such that no couple with labeling $l > 4$ is assigned? No, again in any allocation with $l \leq 4$, we need to have both b and c both assigned to 1, which is not possible.
Therefore $l_0 = 4$

(4) $l_0 = 4$ and $l_1 = 5$, $L^\star = \{l : 4 < l < 5\} = \emptyset$. Therefore the algorithm has converged.

The smallest value $l^\star$ for which the corresponding problem allows for a perfect match is the value of the Bottleneck problem. In this case $l^\star = 5$.

4. Discretization of the Problem

Now consider a discretized version of the problem, where the "many" side of the market is composed of a finite number of agents. I consider the problem of matching different students to different schools. The type of a student and a school is fully determined by their geographical location. Let x be the vector of geographical coordinates of a student of type x, $(\text{long}_x, \text{lat}_x)$, and let y be the vector of geographical coordinates of school y $(\text{long}_y, \text{lat}_y)$. Let d_{xy} be the Euclidean distance between points x and y, formally

$$d_{xy} = \left[\left(\text{long}_x - \text{long}_y\right)^2 + (\text{lat}_x - \text{lat}_y)^2\right]^{\frac{1}{2}}$$

Let p_x be the mass of students at geographical coordinates x and q_y the number of students that a school at location y can accept. Again we study, as in the one-dimensional case, both the fair and efficient allocation in this setting. The efficient allocation is the feasible matching μ that minimizes total distance traveled by the students, subject to the capacity constraints of the schools and students. Formally

$$\begin{aligned} \min_{\mu} & \sum_{xy} d_{xy}\mu_{xy} \\ s.t. & \sum_{y} \mu_{xy} \leq 1 \\ & \sum_{x} \mu_{xy} \leq q_y \end{aligned}$$

On the other hand the stable allocation, obtained through the generalized Gale and Shapley, is an allocation with absence of blocking pairs. That is, let $\mu(x) \in \mathcal{Y}$ be the school x is matched with and let $\mu(y) \subset \mathcal{X}$ be the set of students that are assigned to school y. The allocation μ is stable if

$$d_{xy} < d_{x\mu(x)} \implies \nexists x' \in \mu(y) \text{ s.t.} d_{x'y} > d_{xy}$$

and

$$\exists x \notin \mu(y) \ and \ x' \in \mu(y) \text{ s.t. } d_{xy} < d_{x'y} \implies d_{x\mu(x)} < d_{xy}$$

In words, the allocation is stable if for every school that x would prefer to its current assignment $\mu(x)$, the school is matched only to students that live closer to the school than x, and if for every student that y would prefer to any of its current assigned students, x is not interested in a match with y because they are assigned to a closer school. Notice that we can reformulate the stability concept in a cleaner way as

$$d_{xy} < d_{x\mu(x)} \implies \max_{x' \in \mu(y)} d_{x'y} < d_{xy}$$

and

$$\exists x \notin \mu(y) \text{ s.t. } d_{xy} < \max_{x' \in \mu(y)} d_{x'y} \implies d_{x\mu(x)} < d_{xy}$$

The quantity $\max_{x \in \mu(y)} d_{xy}$ is the radius of school y.

The following Figures 4.1 and 4.2 display an example of efficient (distance minimizing) and fair allocation in the two dimensional case. The first one is a linear programming problem computed using Gurobi, the second one using a Gale and Shapley generalized algorithm. Notice that by using the monotone transformation proposed in Section 2 as the objective function rather than the distance, we obtain "intermediate" allocations that are neither fair, nor efficient in the classical sense.

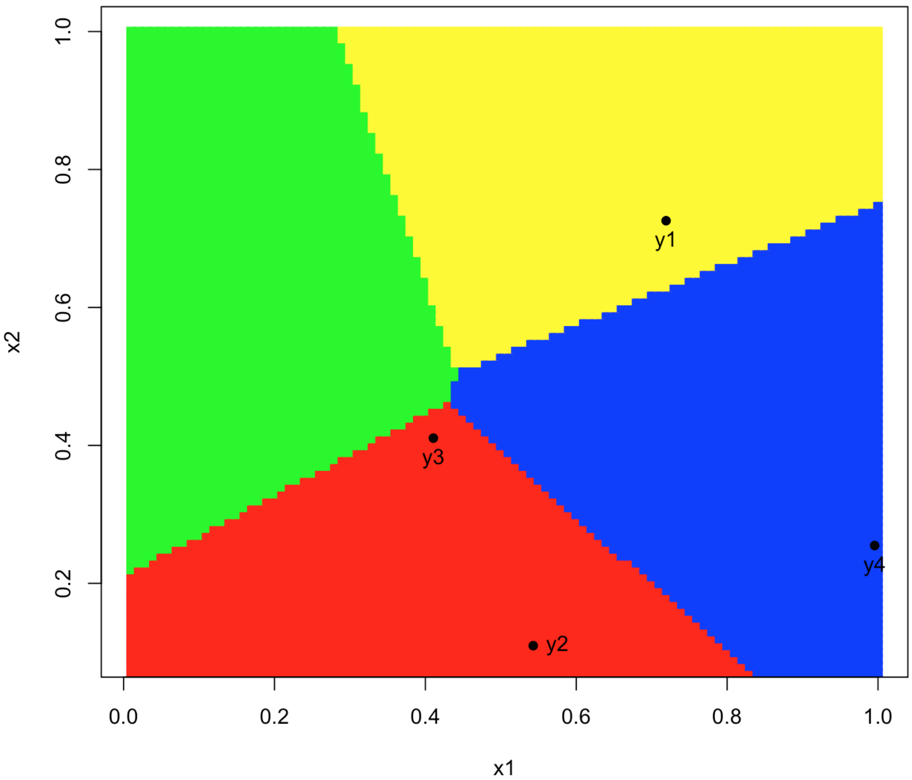


FIGURE 4.1. Efficient Allocation

5. DATA

The data are open data from the NYC Department of Education, publicly available, merged with US Census data about New York City tracts and population demographics in each tract. The data include 2166 census tracts in the New York City area. The sample of schools from the NYCDOE includes the locations of 1701 public schools, that consist of 256 kindergartens, 905 elementary, 602 middle and 564 high schools. Some schools may belong to more than one category, for example if they offer schooling for grades 1st to 12th. School locations are shown in the following maps, together with demographics of each tract. In order to estimate supply, I obtain information about school capacity for each school in the database from the NYC Department of Education, Enrollment, Capacity and Utilization Report for the 2016-2017 school year. For the 50 schools observations that do not have data on their capacity I use their current enrollment instead. In order to estimate demand, I use Census data for each tract to derive information about the number of kids in the age range of interest. I take the number of potential students in each Census tract to be the total number of kids in the age of interest in that census tract. Since I will be using distance as the only driver of utility in school search

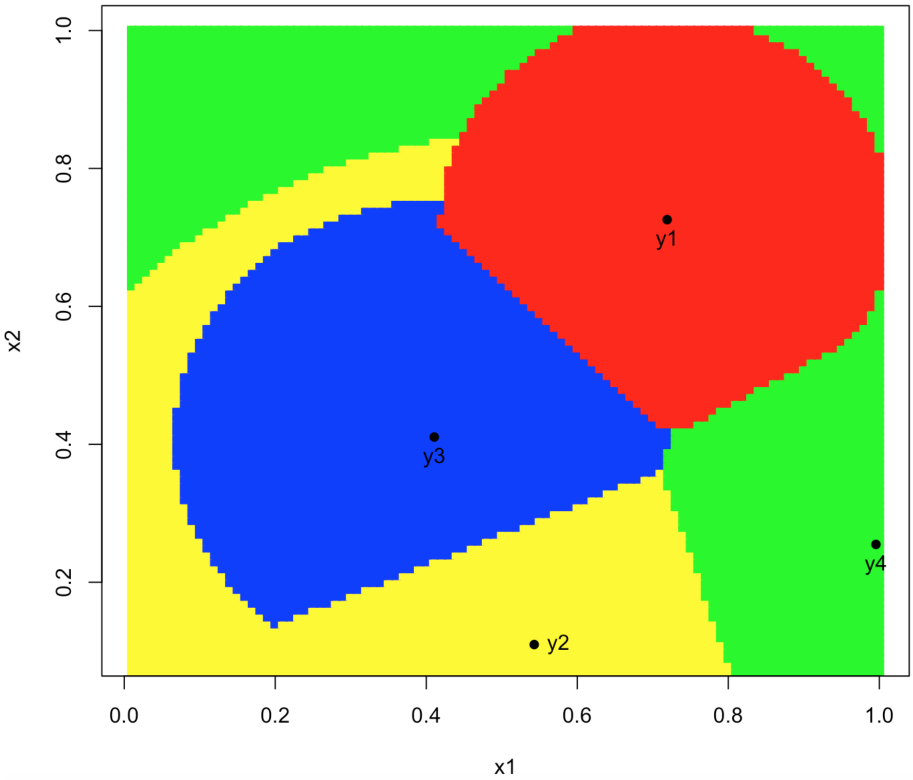


FIGURE 4.2. Fair Allocation (Generalized Gale and Shapley)

for both the school and the student, and since information on the specific address within a tract of a child in the age range of interest is not available, I approximate the distance that a student has to travel to go to the school as the distance between the centroid of the tract and the school.

6. Application to New York City

Suppose that both students and schools care about the distance the student has to travel to the school. This can be somewhat of a reasonable metric of utility for the student, and can be justified for schools by saying that schools need to provie yellow bus services to students, and prefer therefore to have their students live close to the school in order to save on time of bus services they need to provide. Since data on the specific location of students within the census tract is not available, I will approximate the distance a student needs to travel to the school with the distance from the centroid of the census tract to the school. In this sense, all students living in a given census tract will be regarded as equal. Formally, let μ be a matching and let q_i be the capacity of school i. In order for it to be a feasible allocation it needs to satisfy the following

$$\mu_{ij} \in \{0,1\}$$
$$\sum_j \mu_{ij} \leq q_i$$
$$\sum_i \mu_{ij} \leq 1$$

It is important to note that there are more kids in school age in New York City than schools have capacity. Since the NYC public school system cannot satisfy the demand, some of the students remain unmatched at the end of the algorithm, so that effectively the allocation that we obtain satisfies the following:

$$\mu_{ij} \in \{0,1\}$$
$$\sum_j \mu_{ij} = q_i$$
$$\sum_i \mu_{ij} \leq 1$$

Students that are not allocated to any school, are assumed to go to a private school.

Assume that both students and schools care about the distance the student has to travel to the school. This seems an intuitive approximate metric of utility for the student, and can be justified for schools saying that schools need to provide yellow bus services to students, and prefer therefore, to have their students live close to the school in order to save on time of bus service they need to provide.

The assignment algorithm proceeds as follows:

(1) Proposal Phase:

Students apply to the school that is closest to their home.

(2) Disposal Phase:

Schools with an excess of supply temptatively accept all students.

Schools with an excess of demand reject some students in order to respect their capacity constraint. Rejection is based, whenever possible, on distance. Ties are broken randomly.

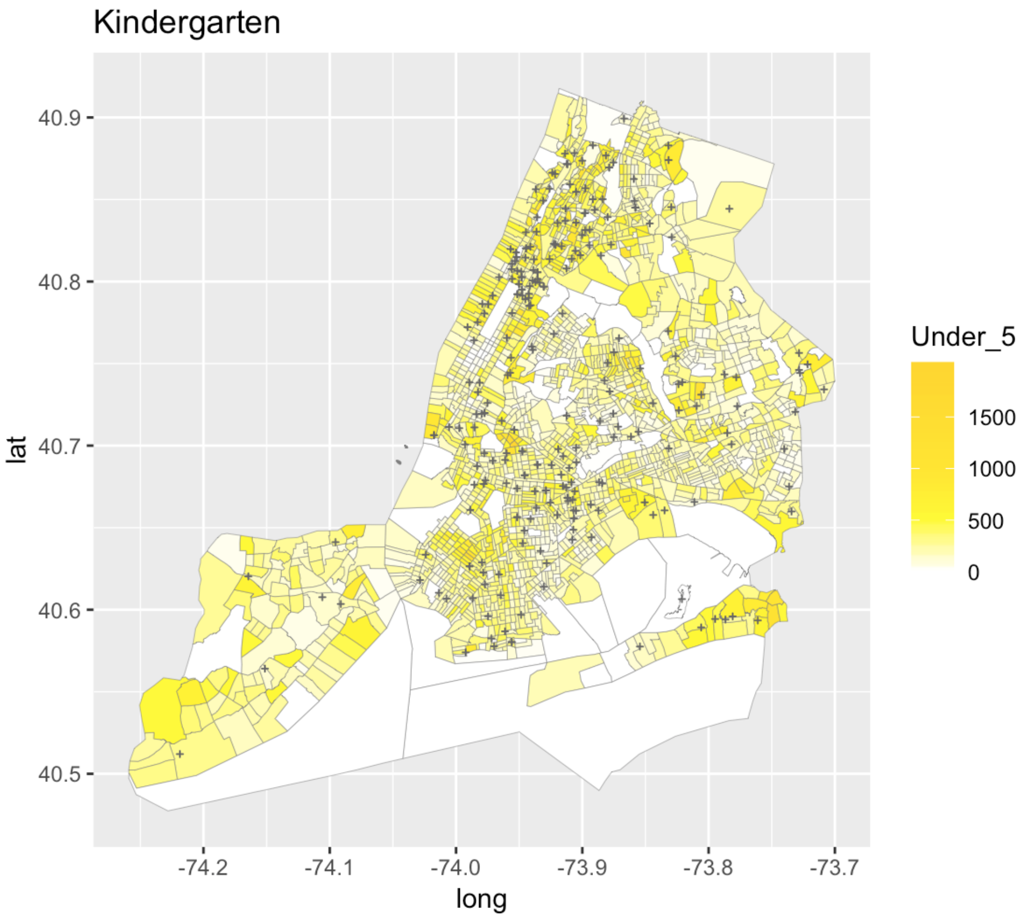
(3) Update Phase:

Rejected students apply to their favorite school among the ones that have not rejected them.

Notice that since all students in a tract are assumed to be equally distant to each school, in the first iteration of the proposal phase of the algorithm all students in a tract apply to the same school. If a school has an excess of demand, it accepts first students that live closest. If the school needs to decide rejection or approval of students that are equally distant to the school the choice is assumed to be random with equal probability to each student of being accepted. This implies that some schools might reject part of the students from a census tract while accepting others, giving rise to unequal treatment. It also means that kids that live in the same Census tract might be going to different schools, traveling different distances.

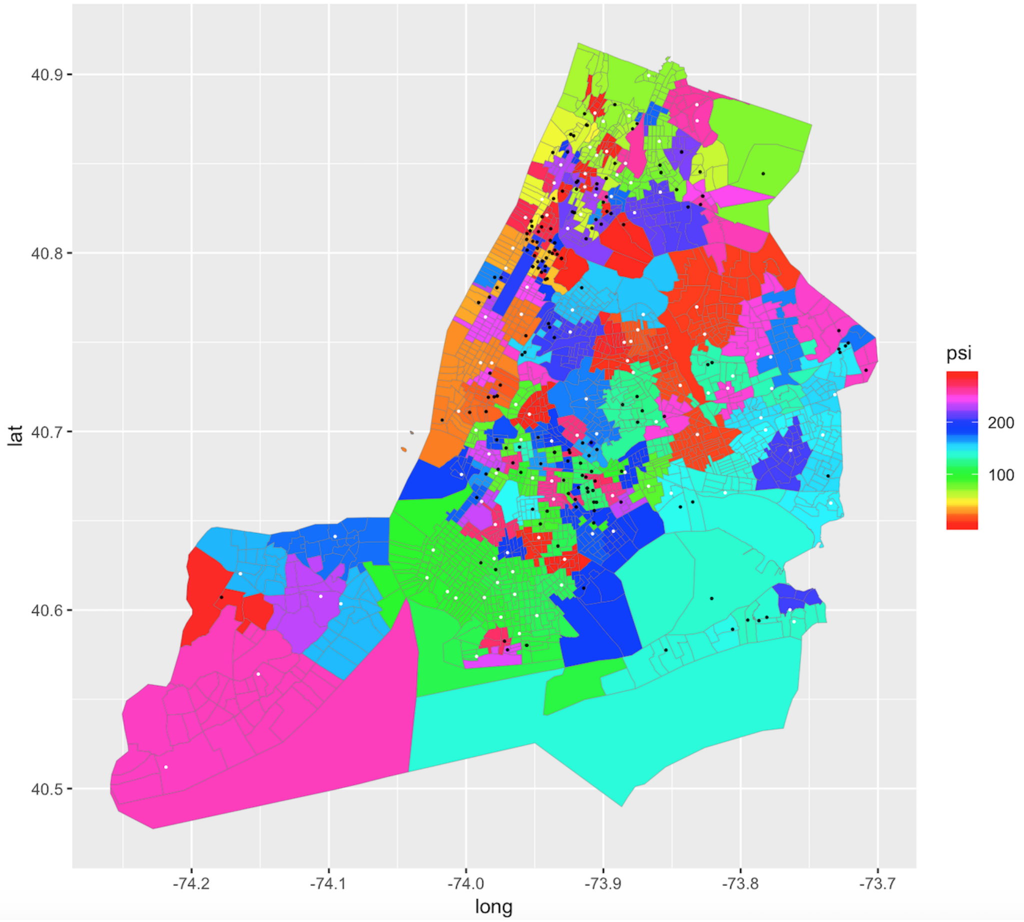
We can now use the algorithm to apply it to the New York City public school system. First of all, we need to take into account the uneven distribution of the population in the NYC territory, which makes it unrealistic to assume that we can use a Lebesgue measure of the territory. I will consider the allocations for students in the NYC territory for kindergarten, elementary school, middle school and high school. First of all I present some maps to show the density and demographics of population in the NYC territory and the location of the schools that interest that type of population. The following maps show the density of population by census tract. The grey crossess indicate the locations of kindergartens, elementary, middle and high schools respectively.

6.1. **Kindergarten Allocation.** We can now move on to find the generalized Gale and Shapley allocation. First of all consider a map of the distribution of children under the age of 5 in each Census tract.



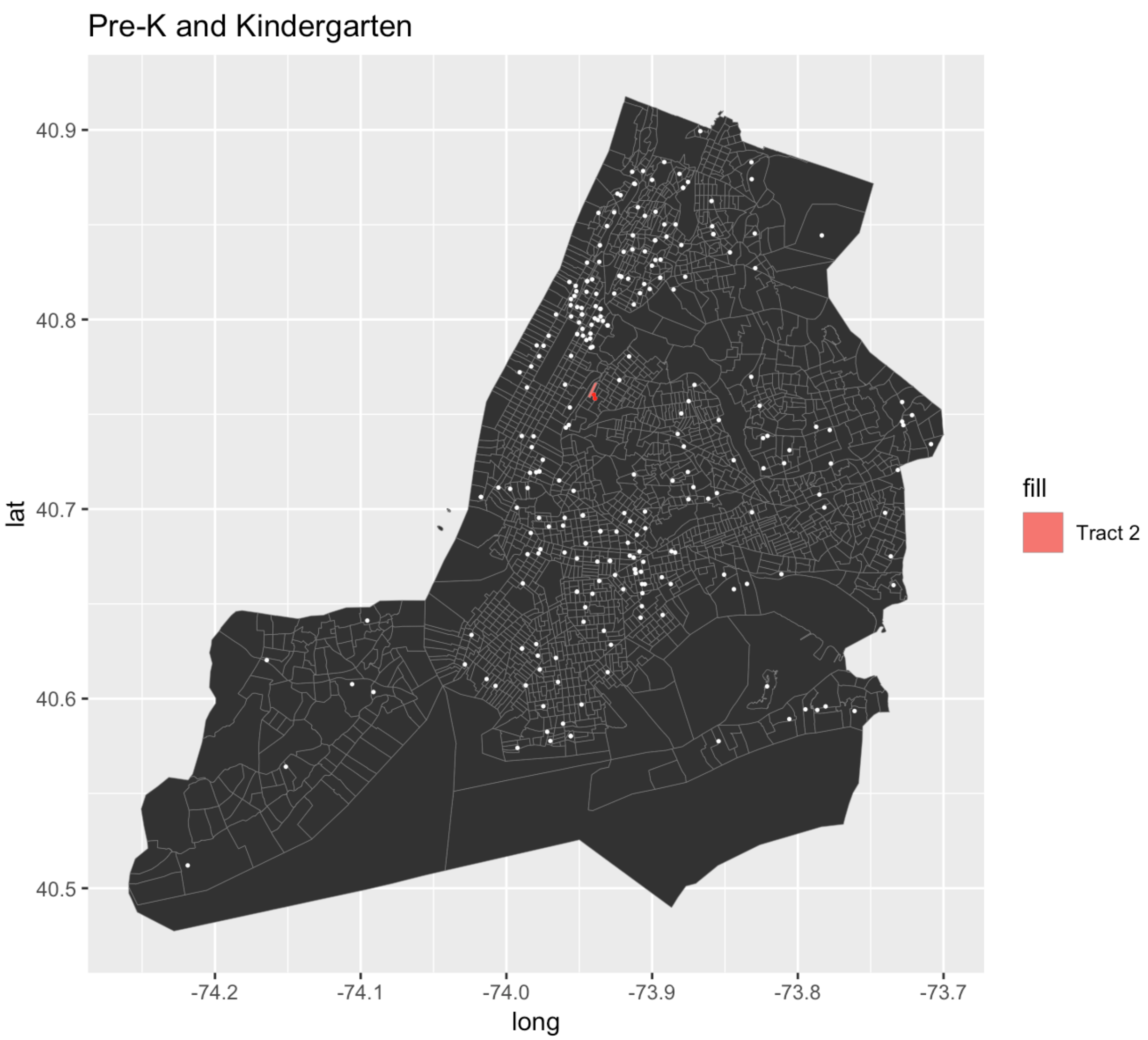
In the first iteration of the algorithm, all kids under 5 years of age are assumed to apply to the closest kindergarten. If the kindergarten can accept all applicants while respecting its capacity constraint, it does so. If it cannot, it rejects a number of students equal to $r_i^{(1)} = a_i^{(1)} - q_i$, where $a_i^{(1)}$ is the number of applicants to school i at iteration 1 and q_i is the capacity of school i. The set of students who are rejected will be the set of students who live furthest to the school. In case of ties, the choice is randomly assigned with equal probability to every student.

It is interesting to see the first iteration of the algorithm because it provides insight into who are the schools that present an excess of supply versus the schools that present an excess of demand. Tracts of the same color are tracts whose students apply to the same school in the first roung. In the following figure the black dots present schools with an excess of demand at the first stage, and the white dots represent schools with an excess of supply in the first phase. It is interesting to see that in certain areas where schools are concentrated - like the area North of central park - schools present an excess of supply. In other areas where on the other hand there is a scarcity of public schools, school face an excess of demand.

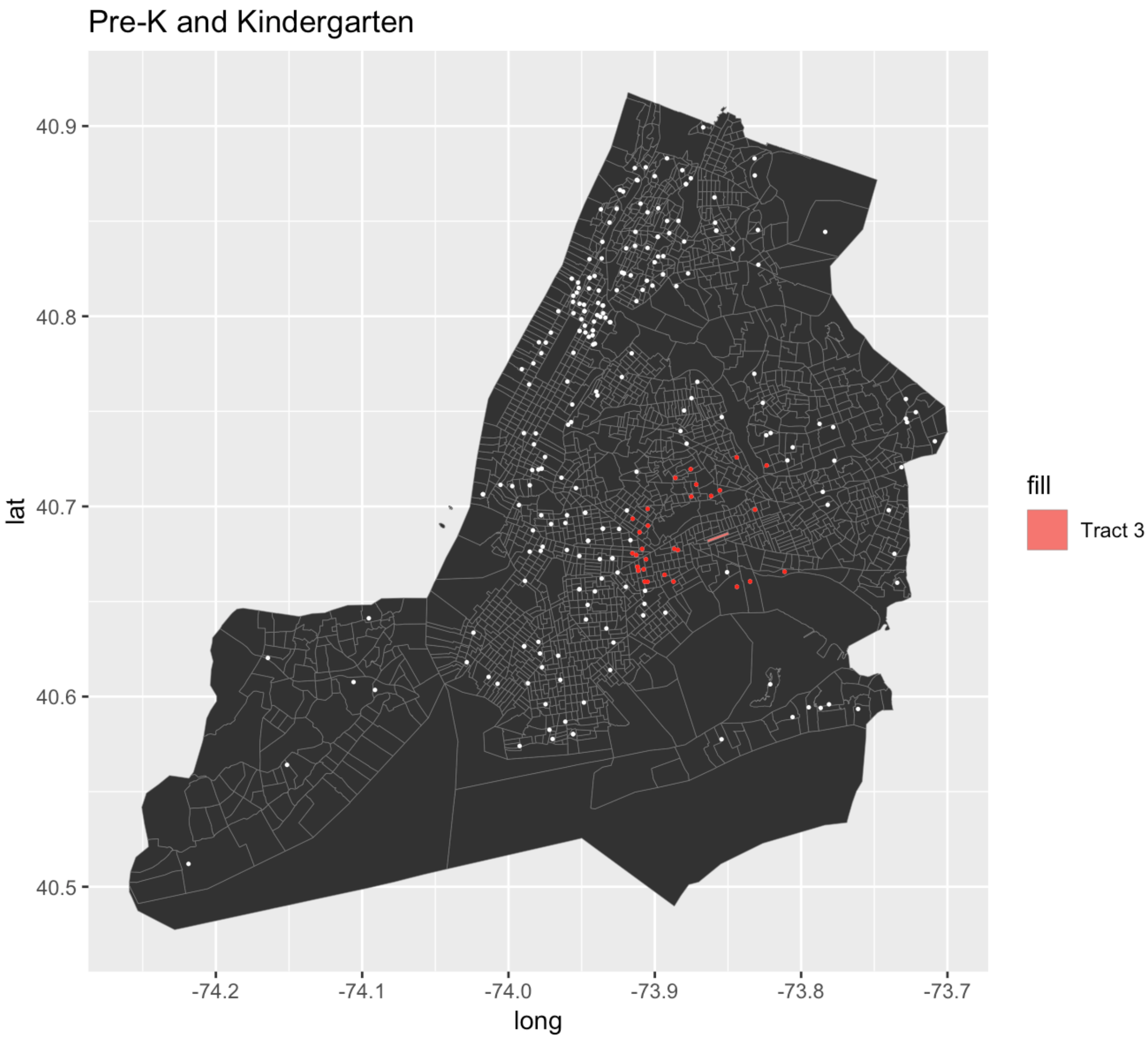


Running the full algorithm until convergence I obtain the following allocation.

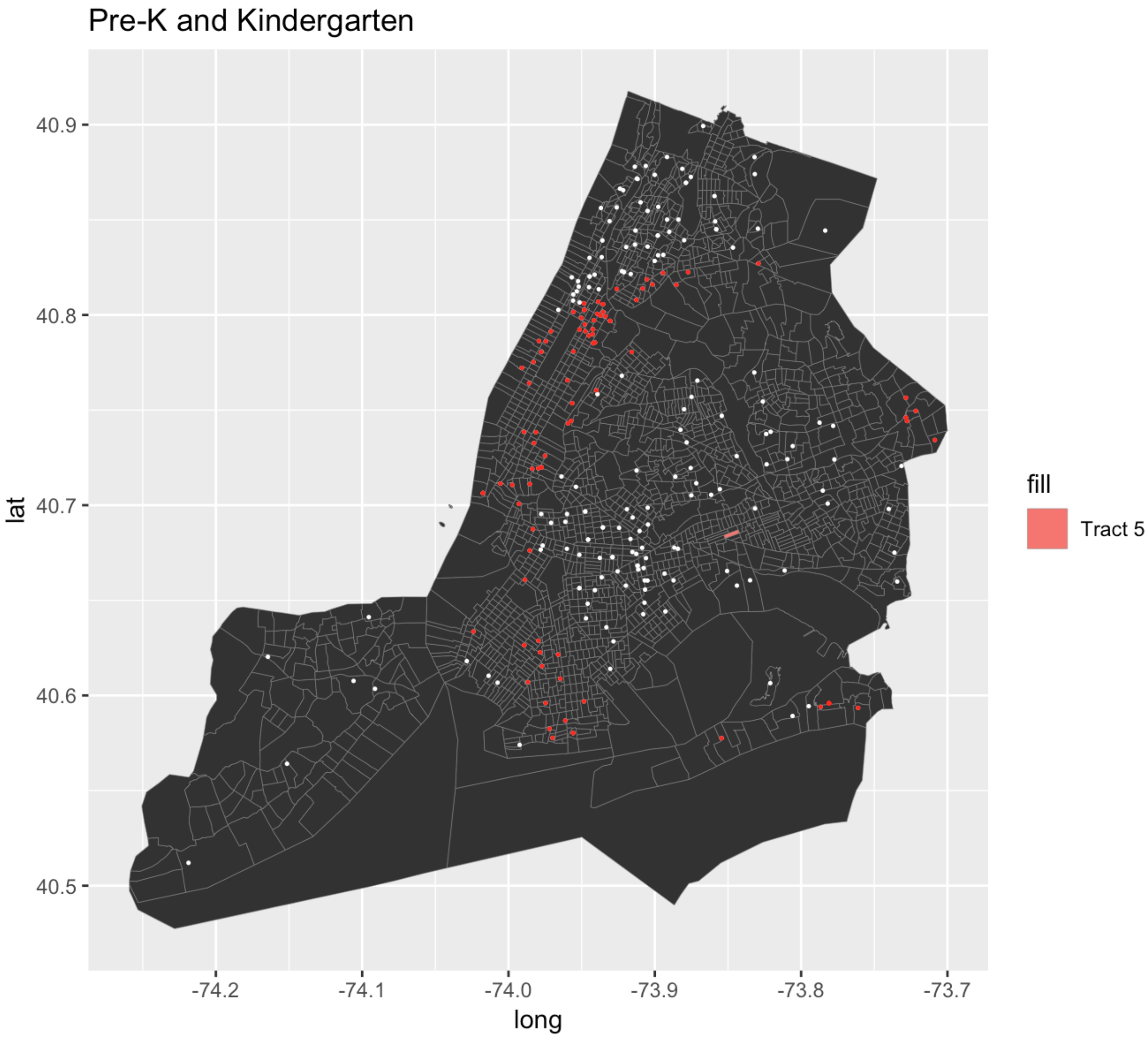
Notice how the allocation can be very unequal, both from the point of view of the students and the point of view of the schools. For instance consider students that live in Census tract 2: they are assigned to two schools, very close to the tract itself. This means that they need to travel a very short distance to the schools, and since the schools are very close to each other, they students that live in the same tract are treated quite equally.



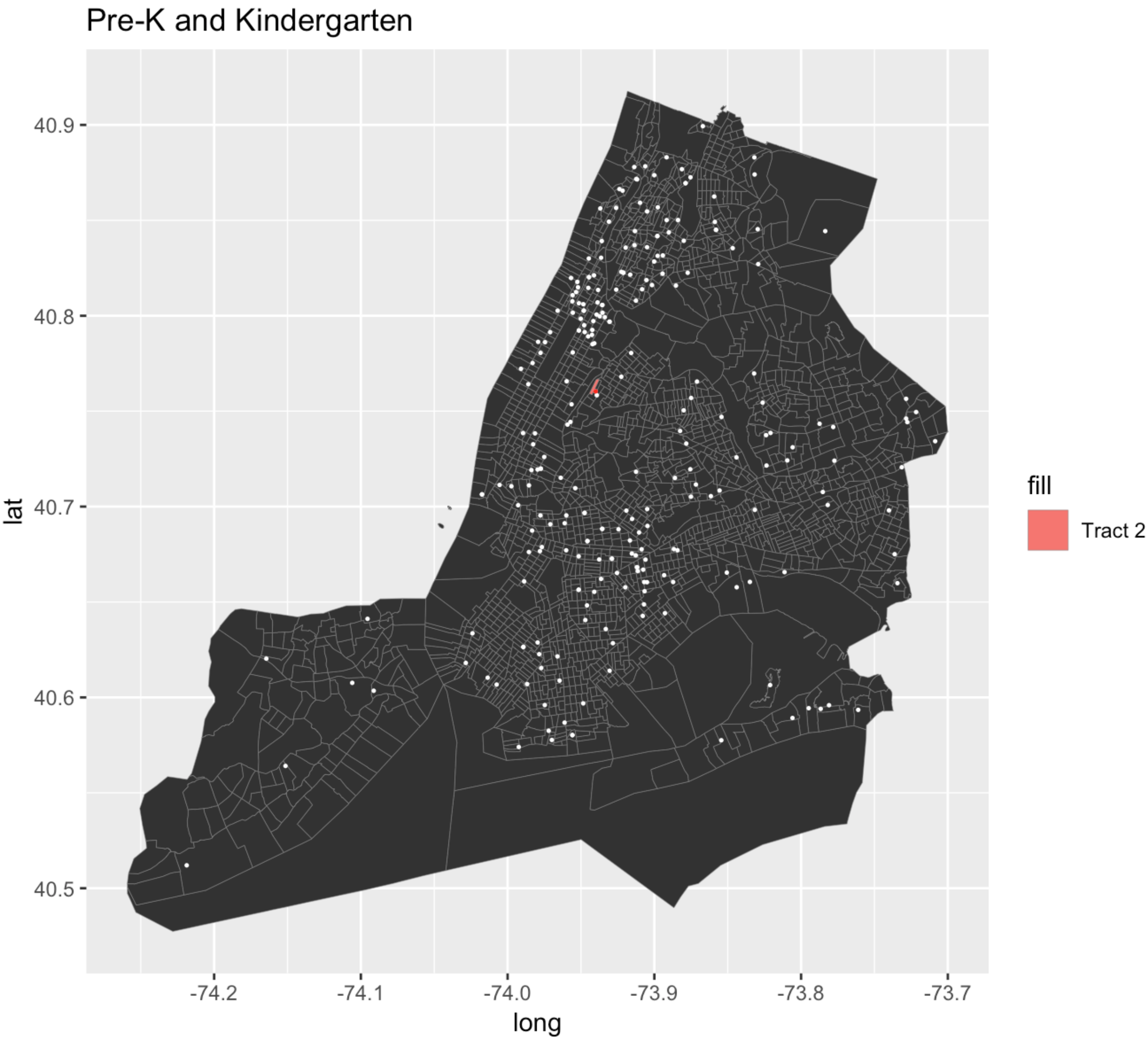
On the other hand, students of tract 3 are assigned to 24 different schools and they need to travel a moderate distance to the school. Also inequality among students within the same tract is higher than in the previous case. Some students that live in tract 3 will be randomly accepted by a closer school and other students will have to travel a longer distance.

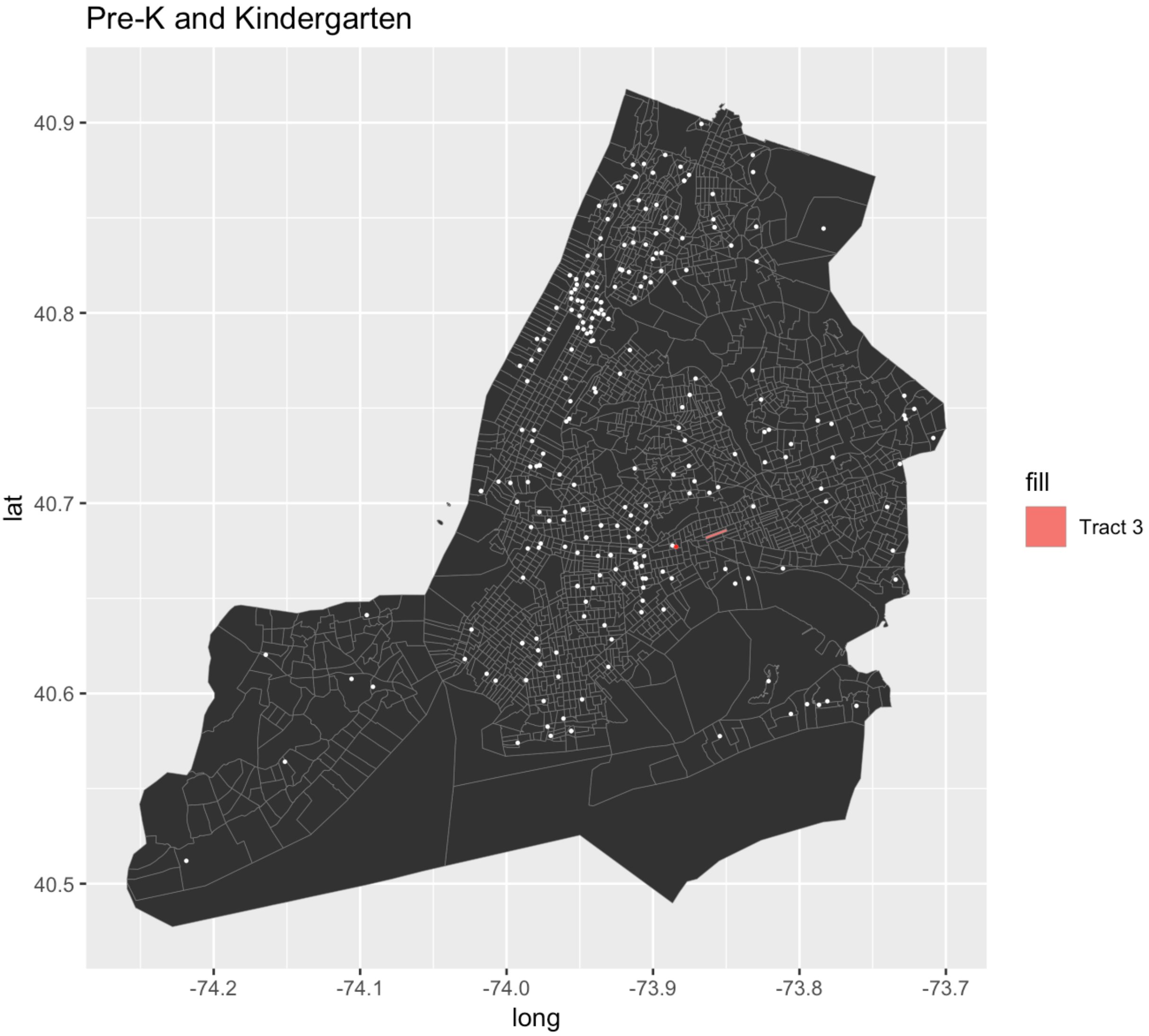


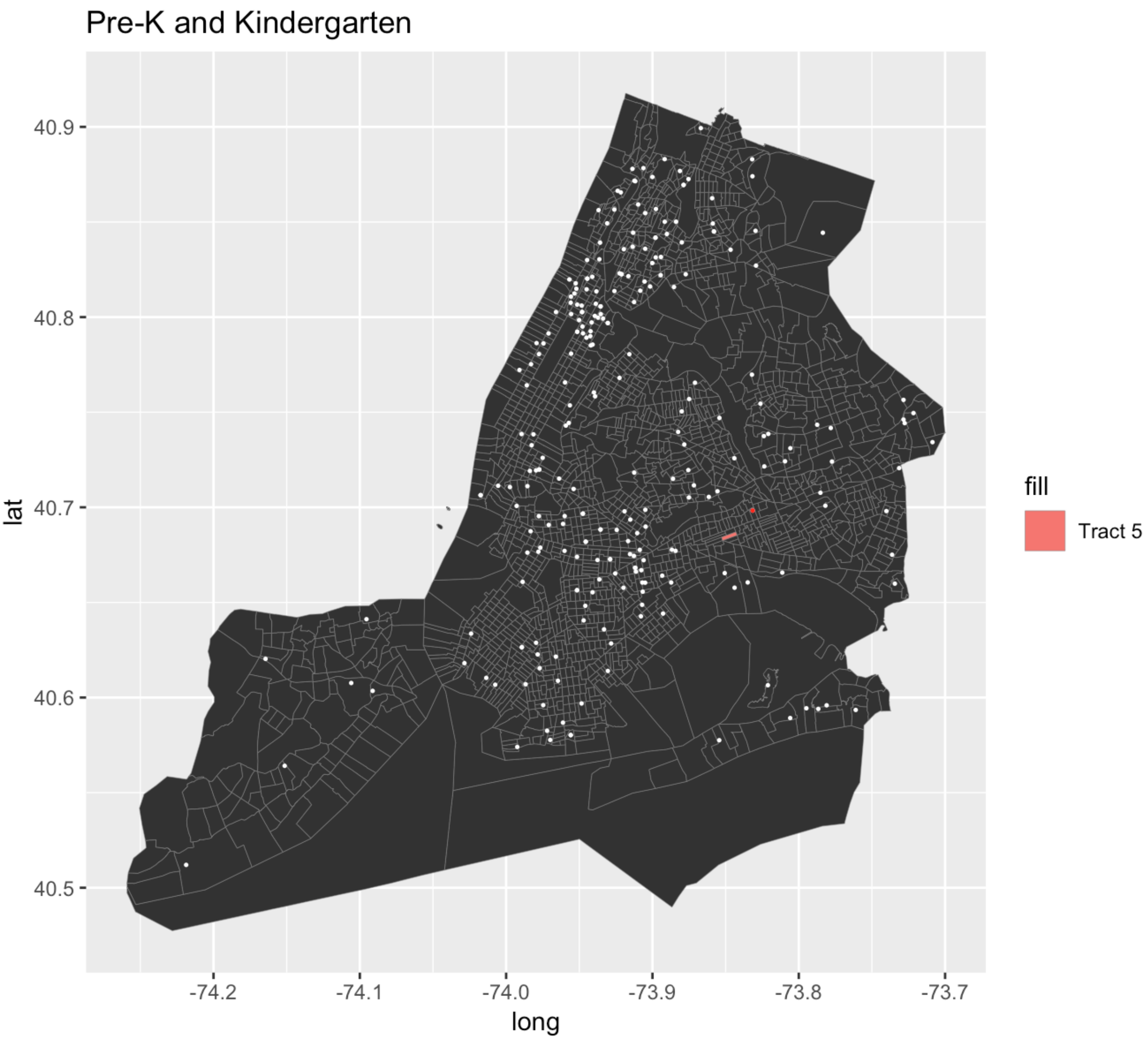
Finally, as an extreme example, students of tract 5 are assigned to 78 different schools. The tract is located in East Queens, close to Long Island. Some of the schools it is assigned to through the generalized Gale and Shapley allocation mechanism are located all the way in East Harlem, the Upper West side, and downtown Brooklyn. Some of them at a distance of approximately 90 km from the center of the tract.



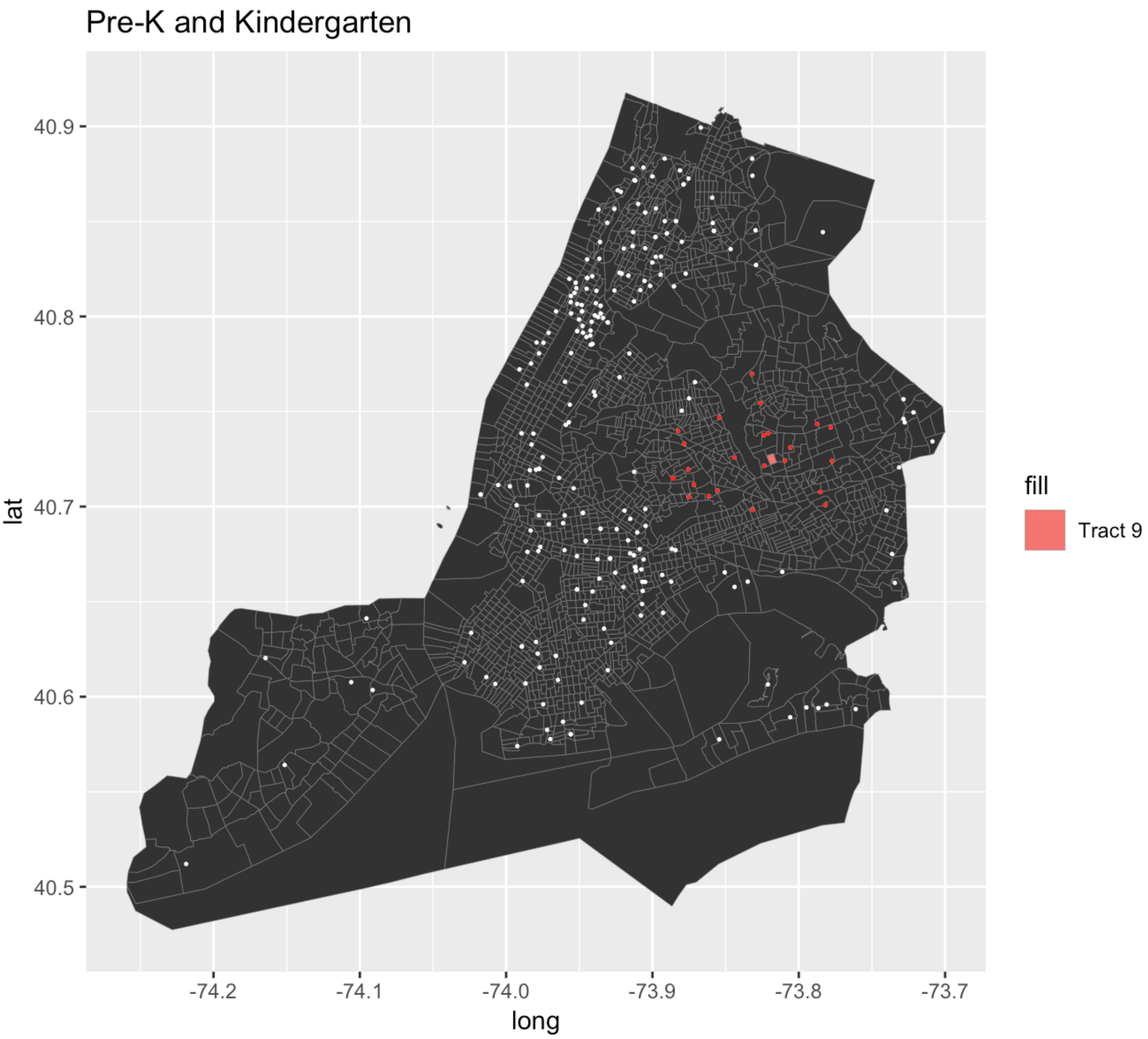
6.2. **Increasing school capacity.** One natural question that might arise is what happens if we increase school capacity to satisfy the potential demand. For the following simulation of Gale and Shapley, I substitue the actual school capacity as found in the data with the school capacity that is necessary to have a spot in a kindergarten for every kid in New York City. I assume that the capacity of each school is uniform and equal to $k = \frac{n}{K}$ where n is the number of children under the age of 5 in New York City and K is the number of kindergartens in New York City. Results are much better in terms of inequality. Looking again at the same census tracts as before, we can see that they are now all fully assigned to a single school, which is especially for tracts 3 and 5 much closer to what they were assigned to without full and uniform capacity.



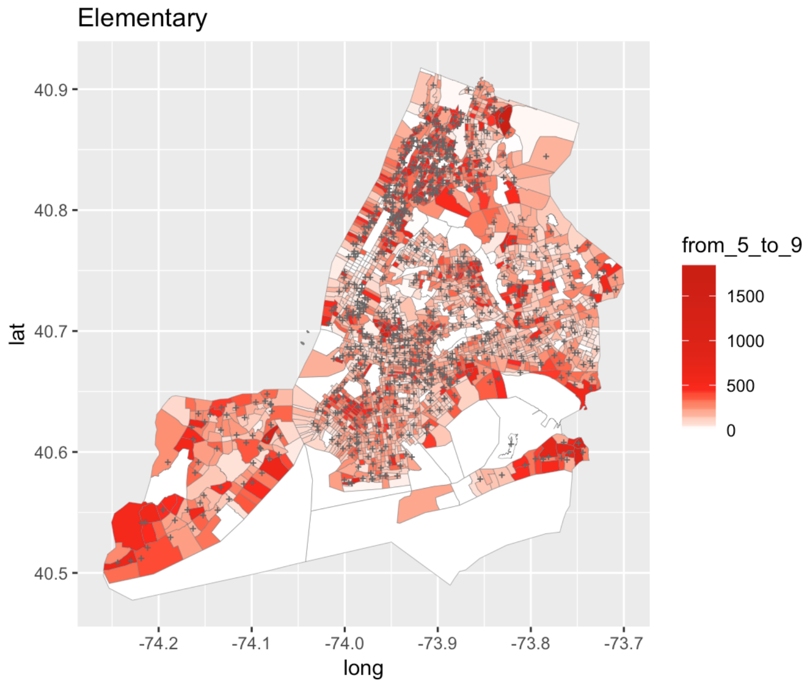

Pre-K and Kindergarten
lat
40.9
40.8
40.7
40.6
40.5
-74.2
-74.1
-74.0
-73.9
-73.8
-73.7
long
fill
Tract 3


Pre-K and Kindergarten
40.9
40.8
40.7
40.6
40.5
lat
-74.2
-74.1
-74.0
-73.9
-73.8
-73.7
long
fill
Tract 5

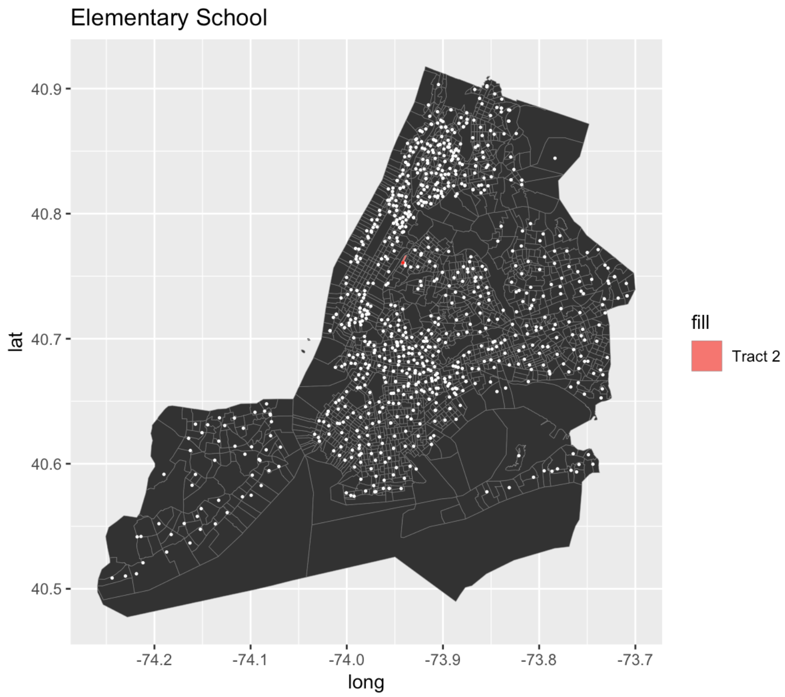
It is important to notice that however, even with full and uniform capacity, certain census tracts are assigned to multiple schools, some of which might be far. As an example consider the assignment of census tract 9.

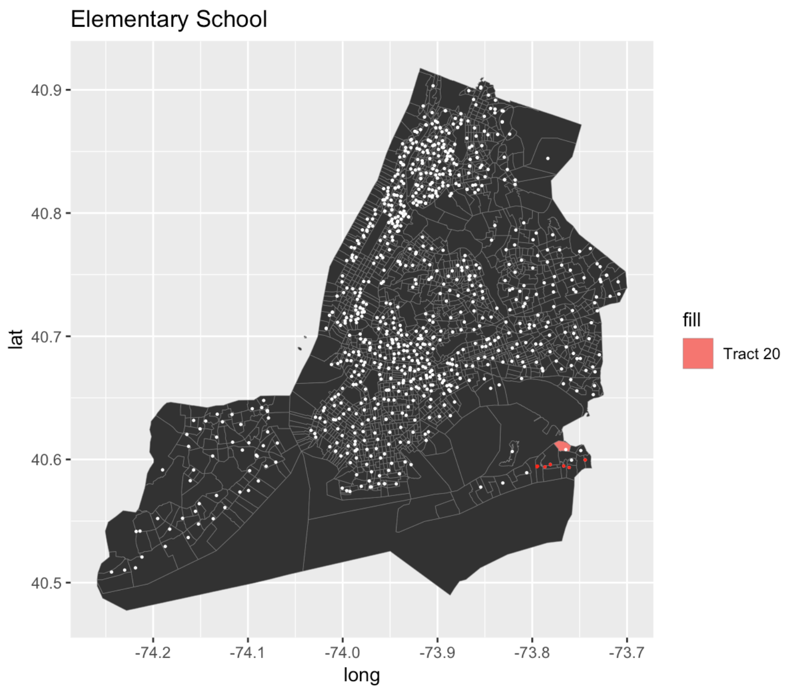


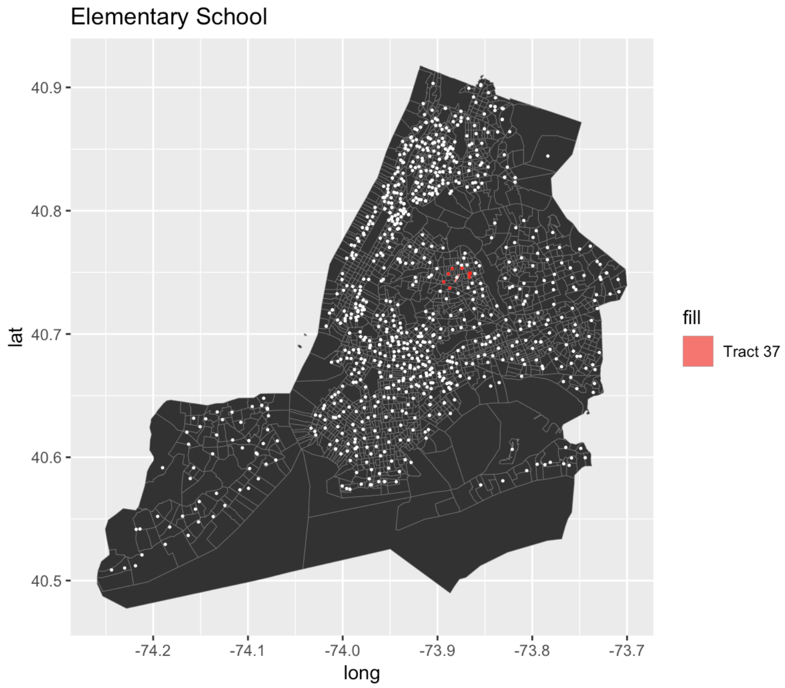
6.3. **Elementary School Allocation.** Once again, let us start by considering the number of children with ages between 6 and 10 per census tract.



And we can now procede to applying the generalized Gale and Shapley algorithm using the actual school capacities. Here are the allocations of some of the tracts. Again, white dots are all the elementary schools in NYC, and the red dots are the schools that the students of the tract under consideration (also in red) are assigned to. Again in certain cases the allocation is very close to the tract, as in the case of Tract 2. In other cases, the tract is assigned to several schools, and the distance traveled is longer. The degree of inequality crucially depends on whether the supply is significantly less than the demand or not. One of the stark characteristics of the fair allocation is that the school a tract is assigned to must not necessarily be the school that is in the tract's territory, as in the case of Tract 37 (last image).


Elementary School
40.9
40.8
40.7
40.6
40.5
lat
-74.2
-74.1
-74.0
-73.9
-73.8
-73.7
long
fill
Tract 2


Elementary School
40.9
40.8
40.7
40.6
40.5
lat
-74.2
-74.1
-74.0
-73.9
-73.8
-73.7
long
fill
Tract 20



7. Conclusions

In conclusion, we explored some geographical applications of different matching algorithms, ranging from a linear optimization problem that yields that distance minimizing allocation, the a generalized version of the Gale and Shapley algorithm, that yields the so-called “fair” allocation. We adapt the algorithm by Gale and Shapley to fit a many-to-one matching problem, in a discretized version of Hoffman (2006) and we present some examples of the matching in a one-dimensional and two-dimensional setting. The simulations provide an intuition of how such algorithm, while providing a stable allocation, can yield unequal results. Going forward, we are interested in formalizing a notion of inequality that allows us to compare the degree of inequality delivered in the stable allocation, when compared to the one delivered by an efficient allocation.

References

Aurenhammer, F., and H. Edelsbrunner (1984): ““An Optimal Algorithm for Constructing the Weighted Voronoi Diagram in the Plane,” *Pattern Recognition*, 17(2), 251–257.

Dupuy, A., A. Galichon, S. Jaffe, and S. Kominers (2017): “Taxation in Matching Markets,” *International Economic Review*.

Gale, D., and L. Shapley (1962): “College admissions and stability of marriage,” *The American Mathematical Monthly*, 69(1), 9–15.

Galichon, A., and Y.-W. Hsieh (2018): “Aggregate Stable Matching with Money Burning,” *Working paper*.

Galichon, A., S. Kominers, and S. Weber (2019): “Costly Concessions: An Empirical Framework for Matching with Imperfectly Transferable Utility,” *Journal of Political Economy*, 127(6), 2875–2925.

Hatfield, J. W., and P. R. Milgrom (2008): “Matching with Contracts,” *American Economic Review*, 98(3), 1189–1194.

Hoffman, C., A. H. Y. P. (2006): “A Stable Marriage of Poisson and Lebesgue,” *The Annals of Probability*, 34(4), 1241–1272.

Kelso, A. S., and V. P. Crawford (1982): “Job Matching, Coalition Formation, and Gross Substitutes,” *Econometrica*, 50(6), 1483–1504.