City of Boston 15-minute city

Research of the features that can make Boston a “15-minute city” where essential services and resources are all 15 minutes from each other via walking distance

New Urban Mechanics Team 1

Jieying Liu, Lianzhu Zhang, Mingcheng Xu, Soowhan Park

April 30, 2021

Submitted for CS506 - Spring 2021

I. Background and Motivation

From project proposal:

We would like to research the features that can make Boston a "15-minute city" where essential services and resources are all 15 minutes from each other via train, bus, or walk. Essential services within the city include: hospitals, healthcare providers, grocers, supermarkets, parks/green spaces, etc. We will target dense urban areas for the 15-minute city concept, this will be done by identifying parcels of land in the city with the highest concentration of essential services and then finding the shortest paths between them.

II. Introduction of Data

Originally, we had 170,731 small parcels, found and defined in many common Boston zoning viewers. In the beginning, we were given separate datasets, each representing all the services/businesses, grocers, and green spaces in Boston. To create the ultimate dataset that combines all the information, we have conducted research on 5 different categories of essential services: *supermarket, grocery, hospital, health care and green space* (i.e. public park). Statistically, we have found valid addresses for 68 supermarkets, 1,876 groceries, 156 hospitals, 2,493 health cares and 1,027 green spaces in the City of Boston.

To be more specific, supermarkets include the large-sized stores such as Target and Costco, whereas the grocery includes the smaller scale of stores such as 7-Eleven and convenience stores. For hospitals, it includes the large scale of health clinic centers such as Massachusetts General Hospital, whereas healthcare providers include the small scale of health providers such as personal health-clinic offices. Lastly, green space includes any scales of parks and gardens. Note that the number of hospitals may be too big because some hospitals have different buildings and these buildings are counted separately.

III. Research Method

Because the distance in a city usually refers to the road distance instead of a straight line, our primary design was to use Google Map's Distance Matrix API (DMA) for accurate calculation of the distance between each parcel and the essential services. However, DMA is not a free service itself and has a usage cap each month, so we need to come up with solutions to avoid going beyond that limit while still completing our task. That comes in two ways: 1) reduce the number of parcels we have and 2) reduce the number of essential services we search for each parcel.

a) To simplify the computation as well as to avoid going beyond the free usage cap of Google distance matrix API, we have successfully combined these small parcels into 13,121 big parcels based on their geographical location. For example, if two houses are on different small parcels but their distance is less than 50 meters, they are considered to be within the same bigger parcel. We choose 50 meters as the parameter because it is a

relatively short distance that can be covered in seconds. For different purposes, the parameter can be changed.
An example of merged small parcels is shown below. These small parcels are in the same block, and combining them does not affect the statistical value of our research.

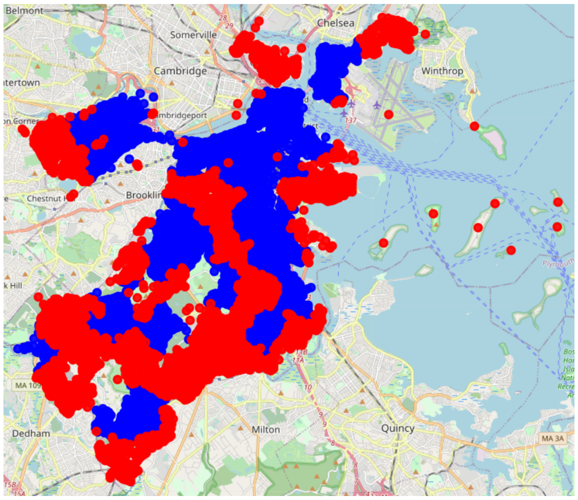
```
'27 -37 CHESTNUT ST #103, CHARLESTOWN,02129',
'27 -37 CHESTNUT ST #107, CHARLESTOWN,02129',
'27 -37 CHESTNUT ST #109, CHARLESTOWN,02129',
'27 -37 CHESTNUT ST #102, CHARLESTOWN,02129',
'27 -37 CHESTNUT ST #104, CHARLESTOWN,02129',
'27 -37 CHESTNUT ST #106, CHARLESTOWN,02129',
'27 -37 CHESTNUT ST #108, CHARLESTOWN,02129',
'27 -37 CHESTNUT ST #110, CHARLESTOWN,02129',
```

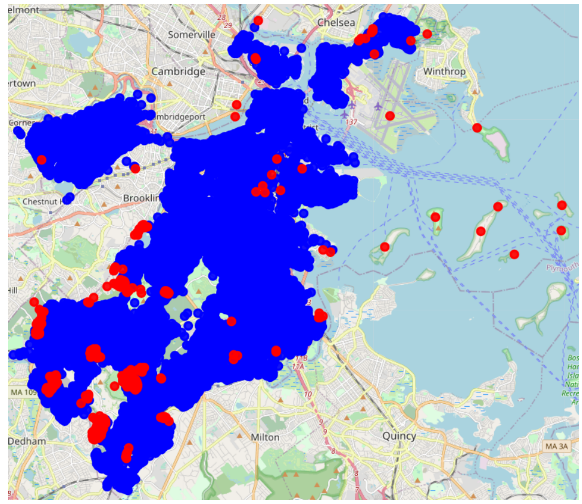
b) Using Euclidean distance, each big parcel is matched with several closest essential amenities of each kind. This redundancy compensates for the fact that the one with the smallest Euclidean distance may not be the one with the shortest road distance. Note that when using the second closest essential amenity to calculate distance for those parcels that do not satisfy a 15-min range, the number of parcels covered increase by 744 in total. Compared to the total number of parcels we have found, this is a strong indication that the matching of each pair and its first closest services has already been very effective.

IV. Visualization and Analysis

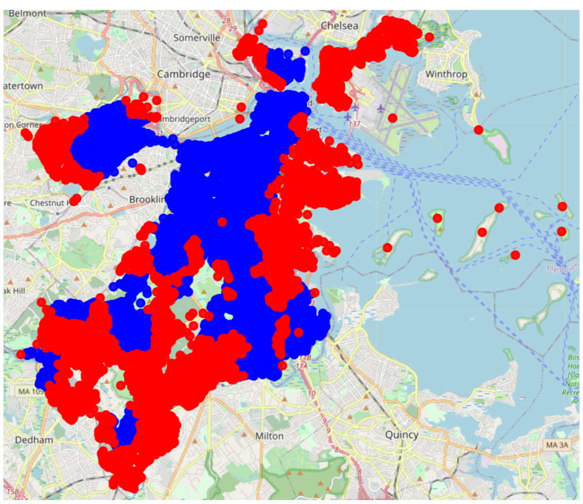
For all figures below, the red represents the parcels that are not covered in the 15-min walking range for the indicated essential service while blue stands for the opposite.

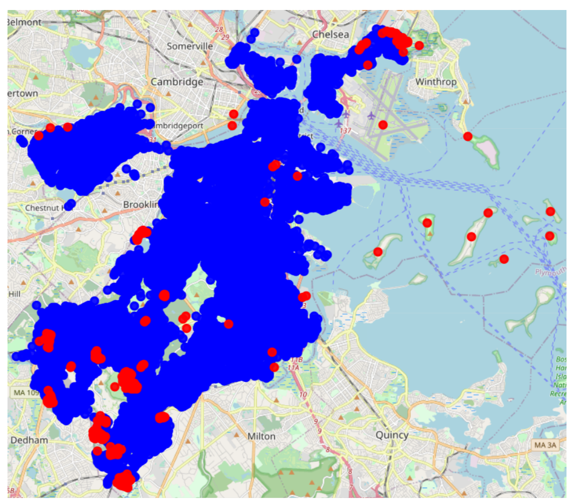
a) What percentage of residents are 15 minutes within essential amenities in a parcel of land?



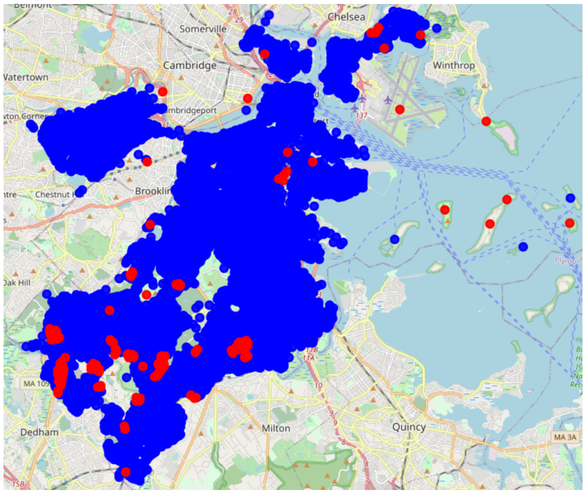


Supermarket; 46% parcels not covered. Grocery; 2.5% parcels not covered.





Hospital; 46% parcels not covered. Healthcare; 2.3% parcels not covered.

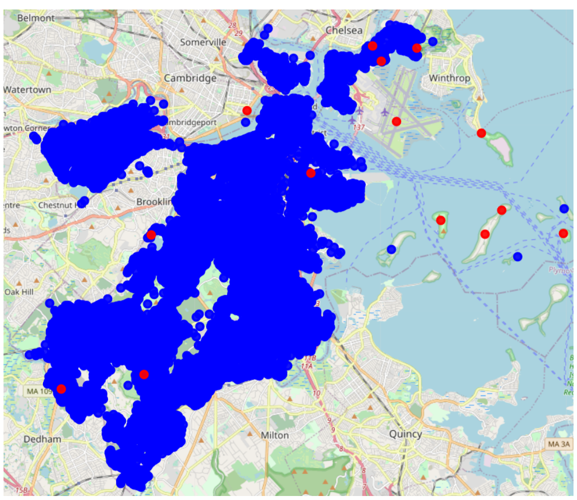


Parks and green space; 1.5% parcels not covered.

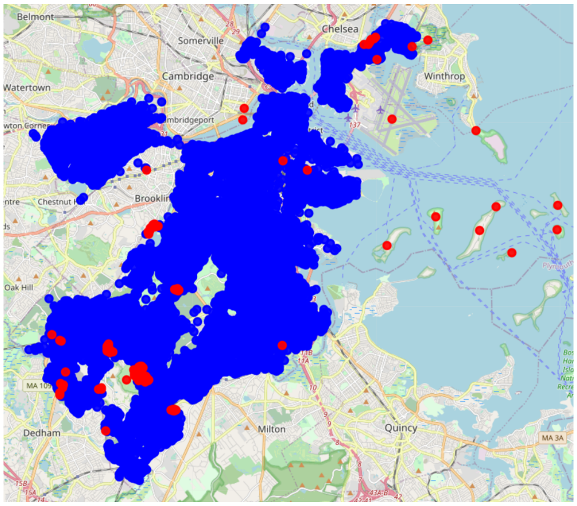
The percentages of parcels that lack each kind of essential amenity (supermarkets, groceries, hospitals, health cares and green spaces) are 0.468, 0.0298, 0.472, 0.0316, 0.039.

Furthermore,

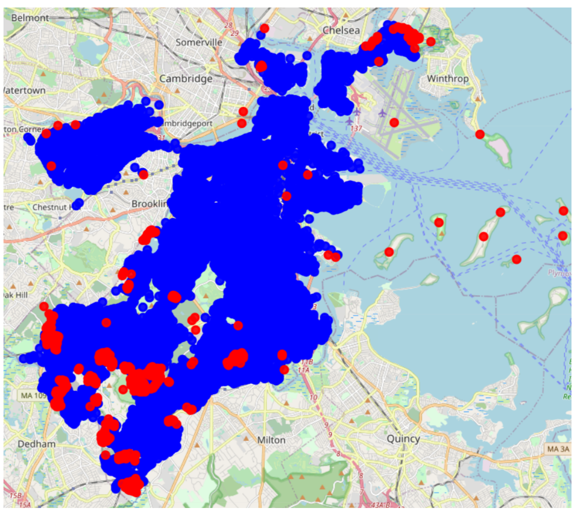
0.1% of the parcels are not covered by at least 1 essential service,



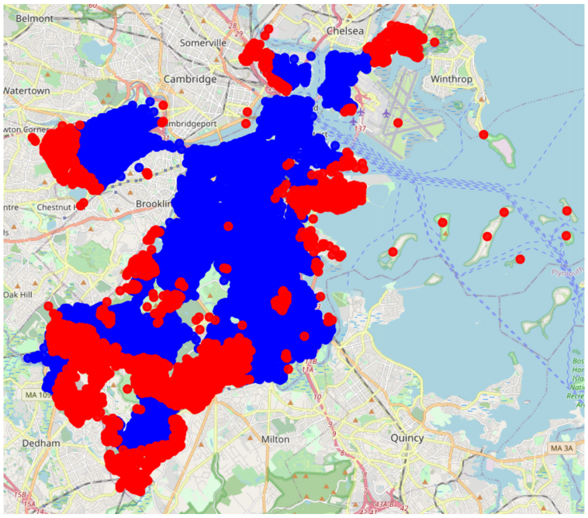
0.98% of the parcels are not covered by at least 2 essential services,



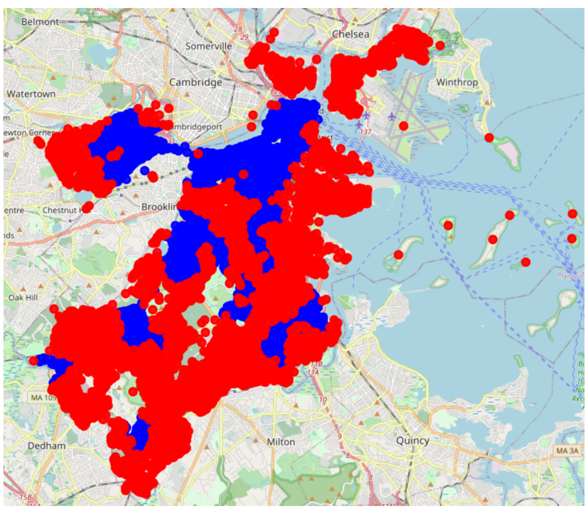
3.7% of the parcels are not covered by at least 3 essential services,



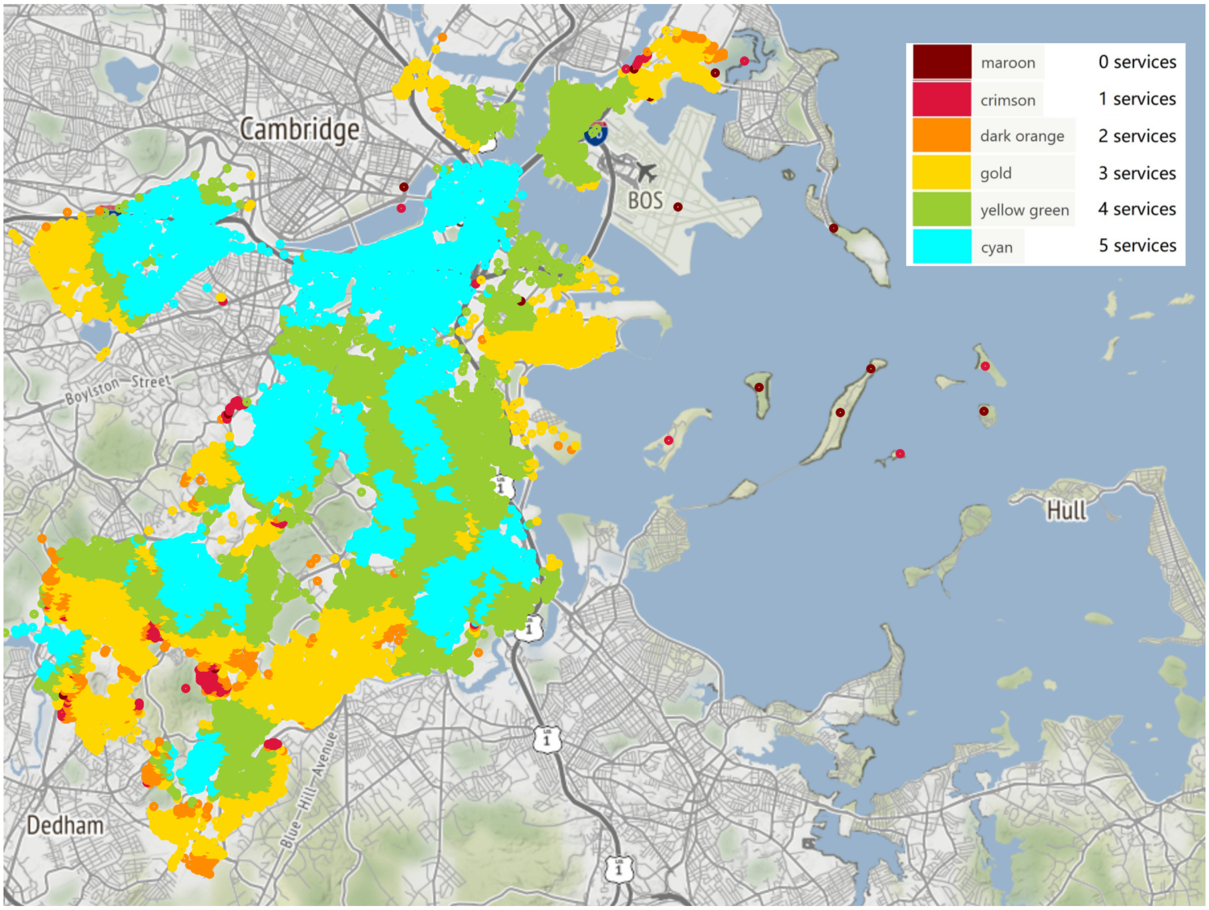
29.8% of the parcels are not covered by at least 4 essential services,



and 63.9% of the parcels are not covered by all 5 essential services.



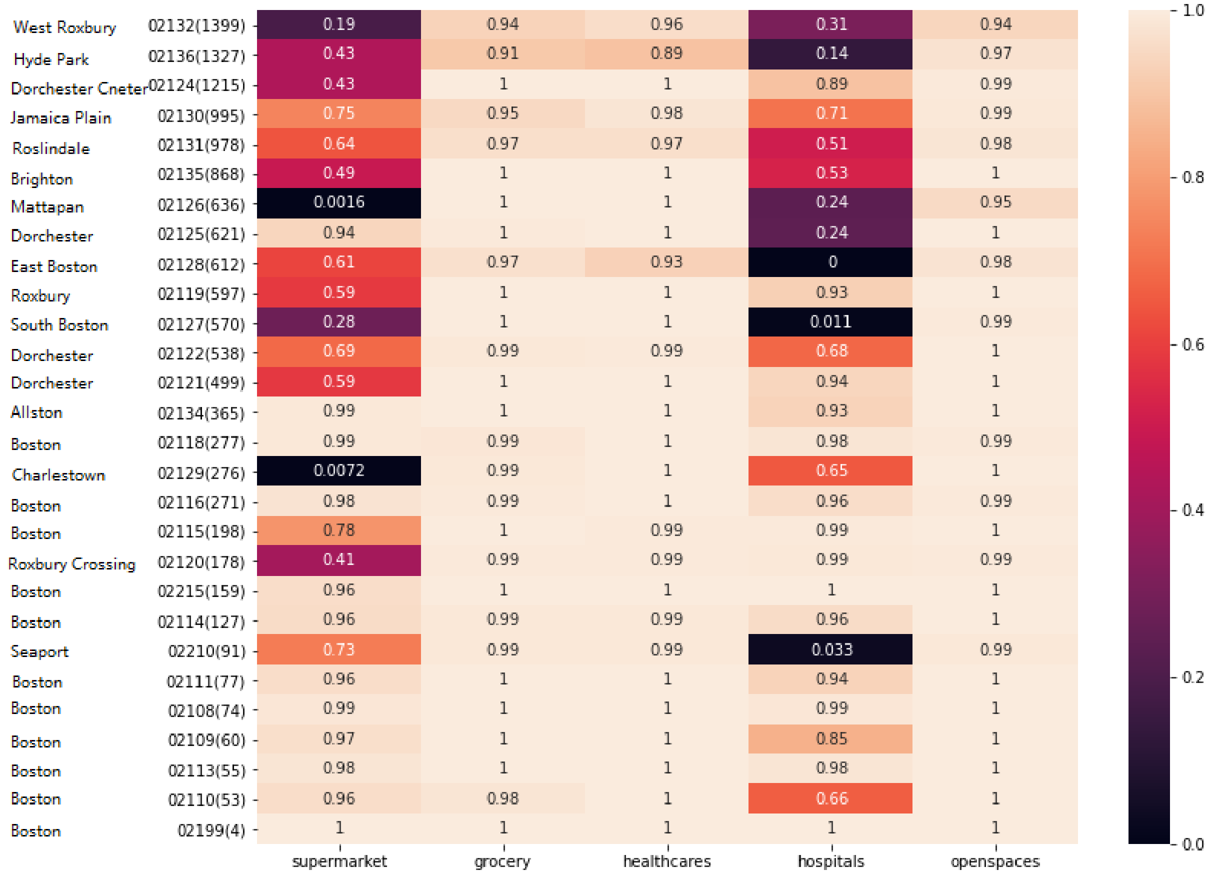
Here is an overview of the coverage situation



b) Which areas of the city are underserved in terms of a lack of essential amenities?

Area analysis is done in terms of zip codes. For our dataset, we have 29 zip codes: [02136, 02111, 02131, 02116, 02129, 02132, 02127, 02126, 02122, 02113, 02215, 02118, 02124, 02125, 02114, 02119, 02121, 02130, 02110, 02128, 02199, 02115, 02120, 02134, 02135, 02210, 02109, 02108, 02133].

The heatmap of coverage rate of each service for each zip code is shown below, sorted in descending order of zip code size i.e. number of parcels included in each zip code. X axis represents services, Y axis is in the format of "zip code (number of parcels included in this zip code)".



It can be seen clearly that the most underserved services are supermarkets and hospitals. In this context, we can list the zip codes that are underserved:

supermarket-underserved areas:

** zip code: city, served population/total population, underserved population*

02120: Roxbury Crossing, 6224/15181, 8957 people have no access to supermarkets

02126: Mattapan, 41/25562, 25521 people have no access to supermarkets

02127: South Boston, 8903/31799, 22896 people have no access to supermarkets

02129: Charlestown, 118/16439, 16321 people have no access to supermarkets

02132: West Roxbury, 4914/25861, 20947 people have no access to supermarkets

hospital-underserved areas:

** zip code: city, served population/total population, underserved population*

02125: Dorchester, 7991/33295, 25304 people have no access to hospital

02126: Mattapan, 6135/25562, 19427 people have no access to hospital

02127: South Boston, 350/31799, 31449 people have no access to hospital

02128: East Boston, 0/40508, 40508 people have no access to hospital

02132: West Roxbury, 8017/25861, 17944 people have no access to hospital

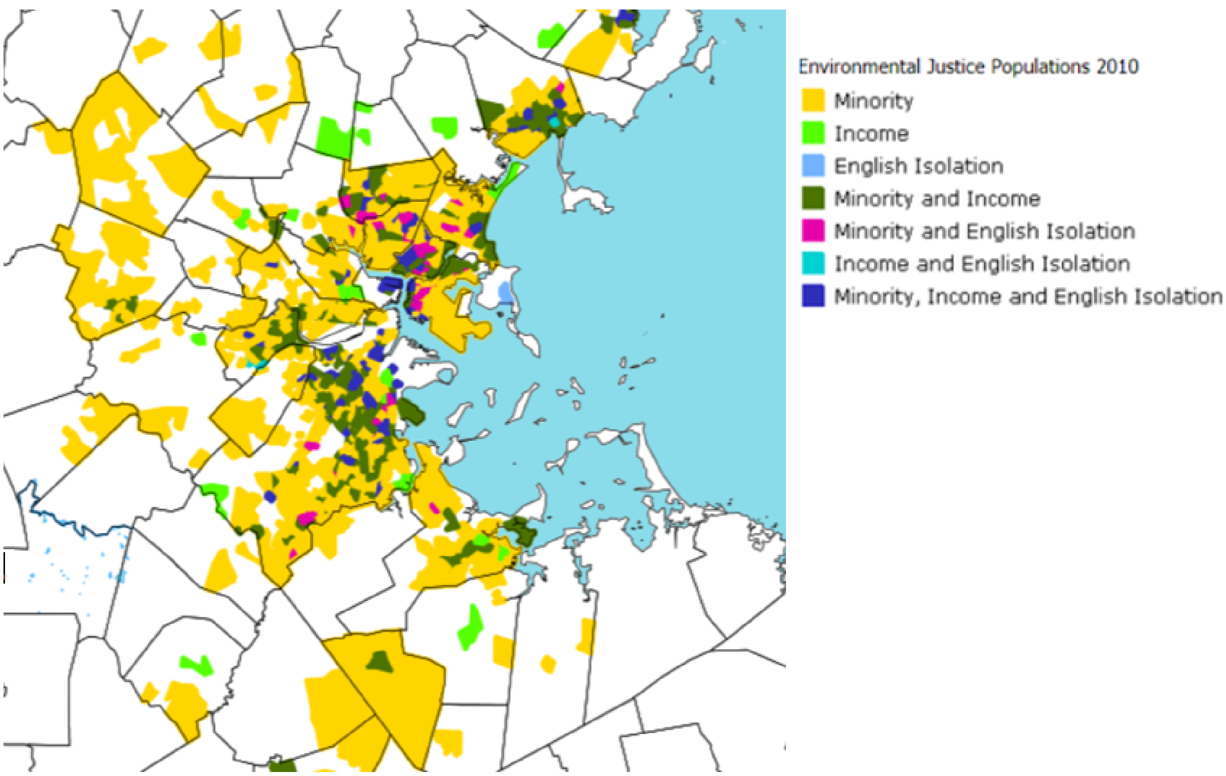
02136: Hyde Park, 3988/28488, 24500 people have no access to hospital

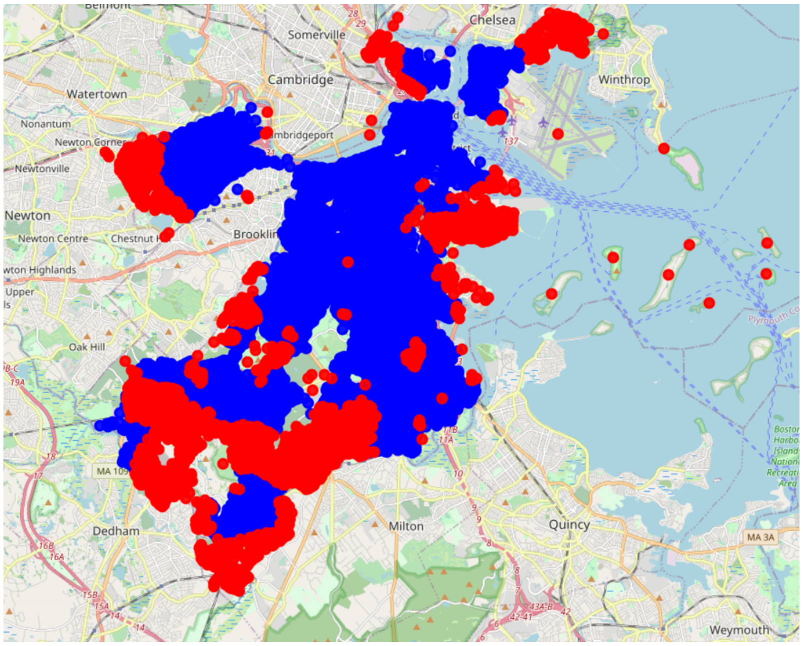
02210: Seaport, 69/2090, 2021 people have no access to hospital

c) demographic analysis:

The demographic map we use are from Environmental Justice Demographics: http://maps.massgis.state.ma.us/map_ol/ej.php

A screenshot of that map is given below:





Given that we have a visualization for the map with the number of services accessible from each parcel, we can analyze the result by comparing it to the demographic map. Assume that we define parcels to be insufficient of essential amenities if those parcels have access to less than 4 essential services.

The first image above shows the demographic information, and the second image marks area with less than 4 essential services with red. Note that the demographic image contains an area that is larger than Boston city.

As we can see from the first image, the city of Boston is very diverse. Most of the areas are marked with "minority" (yellow), "minority and income" (dark green), "minority and english isolation" (purple) and "minority, income and english isolation" (dark blue).

From the second image, we can notice that the most underserved areas are located in the outer region of Boston. Also, there's a huge cluster of underserved parcels in the southern part of Boston. Although those underserved areas are covered by a wide range of minorities, it's hard to conclude that there's a concrete relation or pattern between underserved regions and minorities. First of all, significant portions of underserved areas do not consist of minorities(areas that are not colored in the first image). Second, many of the well-served areas, colored blue in the second graph, are in fact marked minority (and even income and english isolation) in the demographic graph. Therefore, we couldn't find a solid relation between distribution of essential services and certain groups of people who were considered to be minority.

In general, the number of access to services does not depend on an individual's language, income or race, but depends on their location. According to the first image, most neighborhoods in

Boston have a diverse population, and we can infer that there may be a correlation between population density and number of services provided instead of factors like race and income.

Due to the design of the project, we only considered services that are within the governance area of Boston, excluding the essential services in different districts such as Brookline. Thus, most of the underserved areas in the outer regions of Boston are potentially miscalculated. If we conduct the same research on a much larger scale that incorporates the regions such as Brookline or Newton, then the percentage of areas that are underserved in terms of essential services may decrease.

V. Limitations and Possible Plans

As we mentioned above, the free usage capacity of Google Map's Distance Matrix API is limited, so we had to refine the scope of our project. For instance, our initial plan was to research other essential services with the following purposes: recreation, education, and social. However, we would need to pay the extra API requests after exceeding the free limits, thus we had to cut off the features that have lower priority.

VI. Reflection on Procedures of our research

Four raw datasets are collected for this research: the Boston Master businesses and establishments dataset, the grocers dataset from MAPC, the green space dataset, and all parcels in the City of Boston dataset. The first 3 datasets are preprocessed first and combined together into one dataset containing 5 kinds of essential services: Supermarket, Grocery, Hospital, Health care and Open Spaces.

To be more specific, supermarkets include bigger scale shops like Target where you can buy a wide variety of goods, while groceries refer to smaller scale shops such as Seven-Eleven or other convenience stores that are limited to the basic but essential goods. Hospitals include the large scale medical centers such as Massachusetts General Hospital, whereas healthcare providers are personal clinics specializing in certain fields such as heart disease.

In our dataset, there are lots of hospitals because many hospitals have different buildings that are counted separately. We could have combined those buildings into 1, but we needed separate latitude and longitude for each building to locate a place and calculate the pair distance. One integrated hospital with a single coordinate instead of one for each building will lead to inaccuracy and potentially exclude some parcels that should be counted as within a certain range.

For the "health care" part, the data involves medical specialists like physical therapists and cardiologists. It is possible that a parcel is close to a physical therapist but not to a cardiologist, while another parcel does the opposite, but they are both counted as within 15-min range under the "health care" category. In general, this category suggests whether some places are covered by at least any kind of medical care. It is possible to divide health care providers into more concrete small categories for different diseases, which can be done by future researchers if they have background knowledge in the medical field.

The "open space" category includes both parks (which are large areas) and green spaces (which are relatively small). When searching for green spaces, a portion of green spaces in the dataset failed to be located using Python codes. We did not manually search for those addresses as our test turns out that some of them lack essential information and cannot be found online. So, they are omitted instead. If we are provided with a more accurate dataset that contains the valid address and coordinates of each green space, we will be able to include more existing green spaces in our research. But according to our final result, this may be unnecessary: only 1.5% area does not have access to any open space in our dataset. This may suggest that Boston is a green city and most citizens have access to an open space in 15 minutes.

We considered adding a "recreation" category, but it overlaps with the functionality of the open spaces. Since open spaces are already covered in almost all parcels in a 15-min range, adding more recreational places will not shift the outcome.

Additionally, the Boston Master businesses dataset and establishments dataset contain some information that is inconsistent with itself. The information in the dataset is collected by the previous team from 3 sources: Google, Bing, and Yelp. For the same entity, different sources can possibly give different labels. For example, by later examination, we discover that the East Boston Neighborhood Health Center is labeled as "hospital" by Yelp but "healthcare" by Bing (and we do not know which one is true). Our code uses Bing as priority, so that's why our final outcome for the east boston area is underserved by hospitals. This is not known before we finally examine our results, and possibly is not the only flaw in the dataset. However, none of us are familiar with East Boston, as are many other regions, and it is impossible for us to find these by humans, given it is such a massive and noisy dataset. Although this dataset is the only source we can rely on to find information about the services, we believe that this dataset has no problem for the most of contents. For further research, if accurate statistics from the government or any official source is provided, the outcome can definitely be more precise.

Last but not the least, after preprocessing, we use Google Map API to find the coordinates of each parcel. There are 170731 of them. The Distance Matrix from Google Map is the only tool we find that can give an accurate road distance between 2 places instead of the straight line distance (which in our opinion cannot be used to evaluate the 15-min range because travelling in a city is not straightforward in most cases).

However, the problem arises as this tool is not free. Indeed, Google is generous enough to give each account a free usage of 200 dollars per month. But this Distance Matrix is not intended for such massive use: given so many parcels, the free usage will quickly run short before we get any useful result. Therefore, we need to minimize the number of searches (each search looks for the distance between one parcel and one instance of a service). First, we combine small parcels into big parcels: close buildings (e.g., within 50 meters, which takes seconds on feet) are considered to be in one big parcel. undoubtedly, this sacrifices some accuracy, but as a trade-off, it gives hope for us to finish this project.

The second step is to pair each parcel with the closest essential services of each kind using euclidean distance and search for the real road distance using the Distance Matrix. In later review, it turns out that Manhattan distance may be a better estimation for pairing. But such a flaw is offset by redundancy: we match more than 1 closet essential service to each parcel and search for the second option if the first is out of the 15-min range. As it turns out, the euclidean distance approximation still does a great job.

In conclusion, the constraint on Google Map API usage is the most important factor that limits our research. We have to carefully consider how many searches to make, and by the time the report is being made, we have almost run out of the free usage we have. Therefore, further modification of the result is hard, and we are unable to add new features like "school".

For future teams that may want to work on the same project, if they want to get a more precise result (say, focus on 170731 small parcels instead of 13121 bigger ones), we suggest that they either get enough funds or dedicate more time for this project (so the free usage can reset every month). Several thousands dollars should be enough (as long as your API key is not banned by Google due to bulk loading data, which once happened to us), or it may take more than half a year to finish the task without paying.

For those who want to use our result, we have put the information of big parcels, their pairing with essential services and the distances into several datasets, and all can be found in our Spark repository, including instructions.

Special thanks to our professor, clients and program manager, whose supports are essential to our research!