

First insights into Boston as a 15 minute city

Darren Liu, Katie Lee, Khaled Sulayman, Sandra Zhen

Boston University; Boston New Urban Mechanics Team 2; CS506: Computational Tools for Data Science

I. Introduction

In response to COVID-19, Boston, along with 96 other cities in the C40, have committed to a “green and just recovery.” Among many things, cities in the C40 have committed to making sure that all their residents have safe and green methods to access fundamental resources and amenities so that their cities are more resilient to future health crises. Such an idea is encapsulated by the “15 minute city” or a city where every resident has access to essential services and amenities within a 15 minute walk or bike ride. The 15 minute city is not a new concept as cities like Melbourne and Paris have already implemented measures to help promote equitable access. Boston is looking to start building a 15 minute city and thus, we are working with Boston’s New Urban Mechanics team to help start such development. We want to identify parcels that don’t have access to amenities within a 15 minute walk/bike within the city of Boston and identify specifically who and where is Boston lacking. Ideally, we would like to identify zip codes or general areas of Boston that need further development of essential services to help advise future city planning. We have four key questions to be answered:

1. What amenities in Boston belong to each of the six categories of essential amenities necessary for a 15-minute city?
2. What percentage of parcels are within 15 minutes of essential amenities?
3. Which areas of the city are underserved in terms of a lack of essential amenities, and which amenities are they missing?
4. What are the demographics of the underserved (non-15-minute) communities?

II. Methodology

To begin, we sorted amenities into six different categories: healthcare, commercial, social, food access, education, and recreation. The categorization of amenities is fairly intuitive. Healthcare encapsulates all of the hospitals and pharmacies and education includes schools and universities. Food access amenities consist of parcels that provide access to bulk food and does not include restaurants, cafes, etc. The latter do not tend to promote general food access and are placed under different categories. Recreational amenities promote general leisure and sport, which includes places like parks and gyms. Social amenities encourage interactions between people and encompass places like cafes, bars, and places of worship. Commercial amenities are the general “catch-all” for other amenities that are not strictly residential.

For reference, “spatial analysis” refers to how we dealt with determining which amenities are close to each parcel. We performed two different types of spatial analysis which will be referred to as “simple” and “complex”. For simple analysis, we first layered a grid on top of a map of Boston. For every residential parcel, a circle was defined with the parcel at center and radius being the distance that we estimated a person could walk within 15 minutes. We did not take into account streets, the Charles, or general spatial obstacles; it was assumed that a person could walk in a straight line from the residential parcel to each amenity. We then checked which types of amenities were present within that circle.

For complex analysis, we imported street data into a network graph with street segment lengths as weights. For every street intersection, we made an ego graph with each of the connections being other street intersections that are within 15 minutes. We then took each residential parcel and amenity and assigned it to a street segment. We then assumed each parcel and amenity to be close to either end of the street and thus assigned each parcel and amenity to an intersection. Using the corresponding ego graphs, we then checked each connection in the ego graph for a given residential parcel to determine whether it was within 15 minutes of any given amenity.

III. GitHub Repository and Code Description

All code can be accessed [here](#).

Dependencies

- ast
- Folium
- GeoPandas
- Jupyter notebook
- NetworkX
- NumPy
- os
- pandas
- pyproj
- Shapely

code

The preprocessing folder contains all the code that was used to clean the separate raw datasets. These programs span businesses, greenspaces, healthcare, schools, etc. The programs are a combination of both python and jupyter notebook files. These files output cleaned CSV files to datasets_clean.

The maps folder contains html files that are used for figures regarding our general spatial analysis.

The demographics folder contains the program we used for our demographic analysis of the general Boston area and the specific zip codes of interest. The maps folder within the demographics directory contains html files that visualize our demographics analysis.

merge.py is used to merge all of the cleaned datasets together, distance.py performs the simple analysis, and distance_crawl.py performs the complex spatial analysis.

dataset_ignore

The dataset_ignore folder is the folder that would have contained the raw datasets. We intentionally did not push these datasets to Github as they are generally large. Refer to the preprocessing code and the initial revised proposal for raw datasets.

datasets_clean

Each of the cleaned datasets (that is the output of the preprocessing code) are within this folder and are generally self-explanatory. For instance, “pharmacies_sanitized.csv” contains cleaned data about pharmacies in Boston. Each of the cleaned datasets generally contain name, address, category, zip code, latitude, and longitude features.

The final merged dataset is called final_merge.csv in this directory. The features on it are name, zip, lat, lon, category, and address. Most of these are self-explanatory. lat and lon correspond to latitude and longitude respectively, address refers to street address, and category is amenity category. Each of the amenity datasets was standardized and concatenated to create this dataset.

IV. Simple Analysis

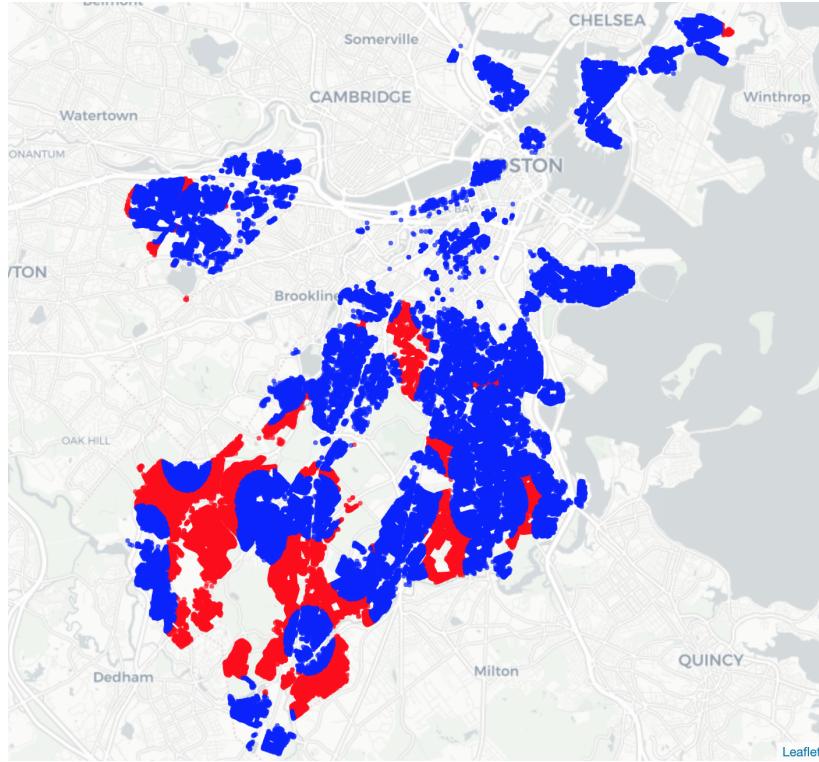


Figure 1: Visualization of Boston as a 15 minute city utilizing simple spatial analysis. Well resourced parcels, or parcels that are 15 minutes within at least 1 of each type of amenity, is visualized in blue. Under resourced parcels are colored red.

Throughout this analysis, “under-resourced” parcel refers to a parcel that is missing at least one type of amenity within its radius and “well-resourced” parcel refers to a parcel that is not missing any amenity types within its radius. Through our simple spatial analysis, most of the parcels in the main Boston area are well-resourced (Figure 1). Most of the under-resourced parcels are further away from the main Boston area. Looking at the coverage of each individual amenity, most amenities covered most of Boston. Social amenities, however, were sparse and most of the under-resourced parcels in Figure 1 are due to the lack of a social amenity (Figure 2). The social amenity was very lacking and was considerably smaller than the other amenities. Talks with Nayeli and Kat also revealed that we were missing some residential parcels in the main Boston area. In the complex analysis, we added the missing residential parcels and refurbished the social amenity.

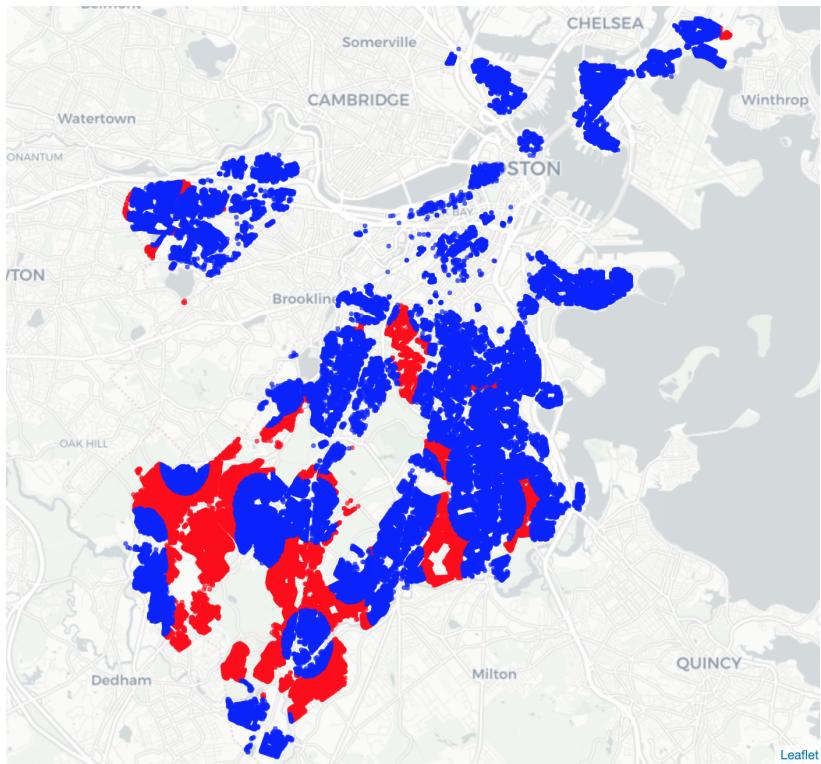


Figure 2: Visualization of which residential parcels are within 15 minutes to a social amenity utilizing simple spatial analysis. Figure 1 is strikingly similar to Figure 2, indicating that most of the under resourced parcels are likely due to missing a social amenity.

V. Complex Spatial Analysis

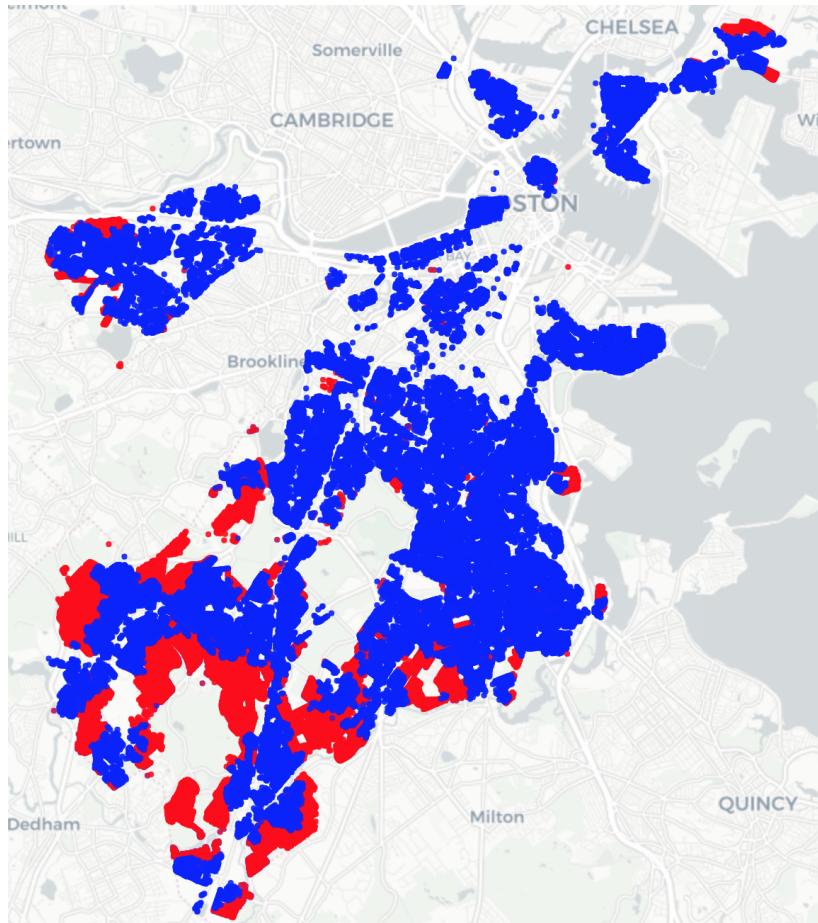


Figure 3: Visualization of Boston as a 15 minute city utilizing our complex spatial analysis. Well resourced parcels, or parcels that are 15 minutes within at least 1 of each type of amenity, is visualized in blue. Under resourced parcels are colored red.

In general, we can see that our results from our simple analysis and complex analysis are consistent, strengthening our confidence in our results. Most of the underserved parcels are in south Boston, which make sense as they are further from the more metropolitan parts of Boston. 78.42% of all residential parcels are 15 minutes within each type of amenity. In general, no one single amenity type is significantly more lacking than any other amenity type (Figure 4). Commercial covers all of Boston, but that is somewhat expected, because commercial ended up being our “catch-all” for any parcel that didn’t clearly fit under any other amenity type. The amount of parcels in commercial is quite large as a result and thus commercial likely covers most of Boston just by sheer number and density of parcels.

Looking at each of the individual zip codes, we wanted to identify zip codes that contained a large proportion of under resourced parcels (Figure 5). We first filtered for zip codes with less than 1000 total parcels as most of these zip codes are not entirely in Boston or are a result of weird zoning (e.g., Boston Commons has its own zip code with a parcel size of 1). As a result, we were able to identify four major zip codes that are lacking a disproportionate amount of essential amenities, and had a significantly smaller average number of amenities per parcel. These zip codes correspond to Mattapan, Hyde Park, West Roxbury, and Roslindale. Overall for these zip codes, no single amenity is lacking significantly

more than any other amenity. For these underserved zip codes, amenities tend to be lacking across all categories in similar amounts rather than any amenity type being particularly lacking. It should be noted, however, that social and health visually appear to lack a bit more than the other amenities—this appearance does not correspond to a significant difference. Such a result indicates that further development in these communities should be focused on introducing a wide range of amenities rather than focusing on introducing any particular type of amenity.

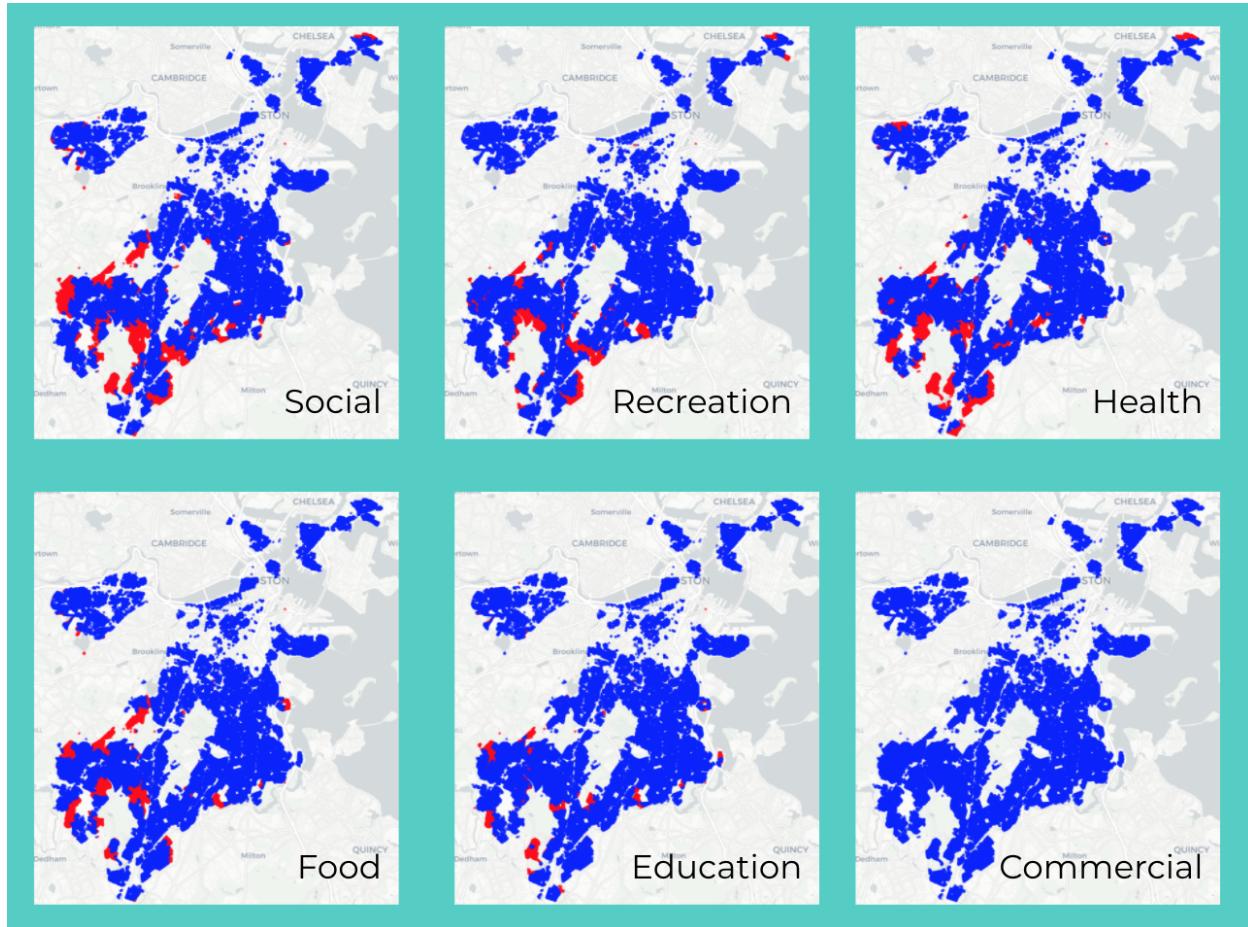


Figure 4: Each map represents how much of Boston is within 15 minutes of a given amenity utilizing the complex spatial algorithm. Parcels colored blue are within 15 minutes of walking of that particular amenity, and those colored red are not.

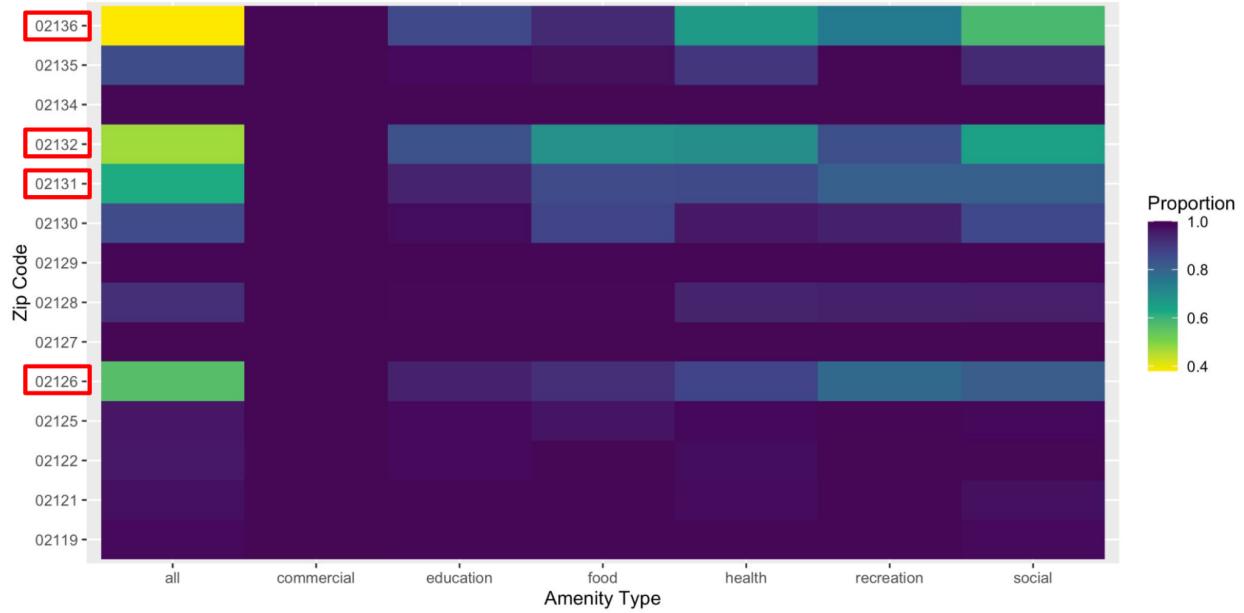


Figure 4: Heatmap showing the proportion of parcels within a zip code that contain a specific amenity. Each cell represents the proportion of parcels within a zip code that contain that amenity. Darker blue cells indicate that a large proportion of parcels within that zipcode contain the specified amenity. Zip codes were filtered to be greater than 1000 parcels. Four major zip codes were identified as majorly lacking amenities (boxed in red): Hyde Park, West Roxbury, Roslindale, and Mattapan.

VI. Demographic Analysis

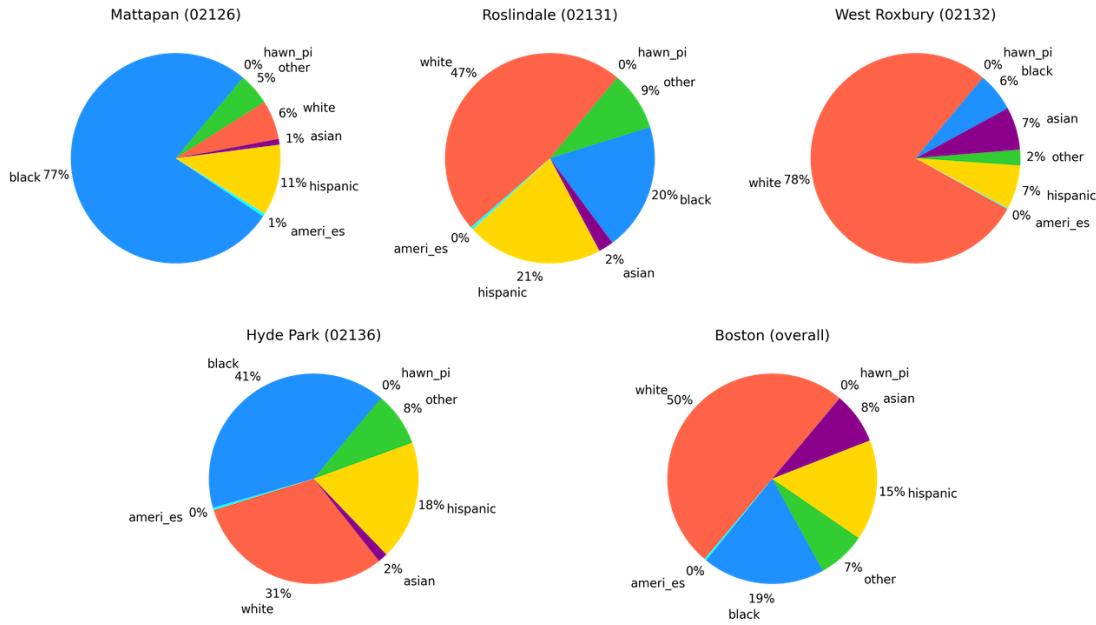


Figure 5: Race demographic pie charts for underserved zip codes in Boston. Data acquired from U.S. Census. Bottom right chart depicts aggregate racial makeup of Boston as a whole.

During our analysis, we were able to identify the above zip codes as notably underserved in at least one amenity category. This result may correspond to locations and community types of each zip code—suburban communities with a higher prevalence of cars and less density in general would result in a larger number of residential parcels without access to amenities by foot. However, it is worth noting that Mattapan does not quite follow this trend, and so may need to be examined for lack of resources.

Looking at the demographic distribution of each zip code, it is interesting to see that only two have a non-white race as the majority of the population. Particularly, in comparison to the demographic makeup of Boston as a whole, neighborhoods like West Roxbury are more white-concentrated than Boston as a whole. In general, there does not appear to be a strong correlation between minority population and amenity access in a particular zip code.

That said, it is important to note that Hyde Park, a black-majority area, is the only region that is underserved in both education and healthcare, two of the most critical amenity categories. Roslindale and Hyde Park both have notably dense Hispanic populations, with the latter being one of the most underserved areas. We also found that asian populations were not very high in these areas, as compared to the rest of Boston. In general, Hawaiian/Pacific Islander and Native American/American Indian are not very populous in the Boston area, and it is difficult to make deterministic inferences on ‘multi-race’ or ‘other’.



Figure 6: Age demographics for Mattapan, Roslindale, West Roxbury, and Hyde Park (left-to-right, top-to-bottom). Age groups denominated in 5-year increments, ending with ages 85 and up. Bar height indicates population density of that particular age group. Data acquired from Census.

On the other hand, the age data seems much more uniform among the under-resourced areas, with a high concentration of young and middle-aged adults living in these underserved areas. All of these areas seem to have especially few senior inhabitants, and the population of children to young adults seems to vary slightly.

VII. Limitations

We looked at whether at least one of each type of parcel was within each residential amenity. However, a more interesting analysis would be to look at the density/quality of each amenity for a given parcel. For instance, a parcel with only one bodega vs a parcel with 4 high quality supermarkets are treated equally under our algorithm even though the first parcel is more underserved than the second. As a quick aside, we expect the computational complexity of such a problem to be immense. Our current complex analysis takes many hours to run and we expect this time to exponentially increase when looking at density.

The main limitation of the simple analysis is how we assume that a person can walk in a straight line from their residential to the amenity in question. In reality, people are subject to obstacles such as rivers and streets that prevent the most optimal movement. The simple analysis model was built mostly to help build confidence in our complex analysis.

Our complex spatial analysis, while more nuanced than the simple analysis, is much more computationally heavy. We expect this to be a result of the sheer amount of parcels. For future analysis, we believe that there should be a focus on the quality of parcels rather than the quantity. For instance, there are a much higher amount of commercial parcels than any other type of amenity. Commercial amenities are much less interesting as they clearly cover all of Boston, but add unnecessary computational complexity to our programs. Ideally, we could identify and remove amenities that are redundant and/or remove amenities that are not important. Such a task is much easier said than done though and would require a deep dive into all of our datasets.

VIII. Discussion of Essential Questions

- 1. What amenities in Boston belong to each of the six categories of essential amenities necessary for a 15-minute city?**

The specifics of classification are discussed in the beginning of the Methodology section.

- 2. What percentage of parcels are within 15 minutes of essential amenities?**

Through our analysis, we found out that 78.42% of all parcels are within 15 minutes of essential amenities

- 3. Which areas of the city are underserved in terms of a lack of essential amenities, and which amenities are they missing?**

The four major zipcodes of interest that we pointed out are Hyde Park, West Roxbury, Roslindale, and Mattapan. Overall, these areas are not necessarily lacking a specific amenity. Rather, they are missing a wide range of amenities, albeit social and recreation tend to be slightly more lacking than the other. We advise that further development in these underserved communities focus on providing amenities across the board. If such broad development is logically difficult, we suggest focusing on introducing more social and recreation amenities. Examples of these amenities are churches and clubs (social) as well as gyms and parks (recreation).

- 4. What are the demographics of the underserved (non-15-minute) communities?**

While we were able to find that these zip codes have, on average, a significantly lower Asian population, these communities have very different racial distributions in general. Additionally, the median age of the under-resourced communities was found to be significantly higher than that of the well-resourced communities.