How Income Per Capita Affects Alternative Methods of Transportation

Liam DeBeasi, Simon Nichols, Arman Sanentz, Molly Shopper
CS591: Data Mechanics – Fall 2016

Abstract: The City of Boston offers different methods of alternative transportation, ranging from the MBTA subway and bus systems, all the way to Hubway bike stops. These systems exist all over Boston, connecting different suburbs and neighborhoods together. Each area has its own characteristics and features, with varying income per capita. We wanted to know if every address in Boston has at least one form of alternative transportation within walking distance, and if not, where is more transit needed? Does this information relate in any way to income per capita?

Method: To determine if every address has at least one form of alternative transportation, we first created a grid along coarse (~1.1km/0.7 miles) longitudinal and latitudinal lines, and added up the total number of bus/subway/Hubway stops in these blocks. We then assigned these values to each of the houses in these blocks. At this point, every address had a numerical value associated with it, indicating the number of transit stops that are in walking distance from it. From here, we built our constraint satisfaction problem mentioned above: does every house have at least one form of alternative transit within walking distance, and if not, is more transit needed? There are houses without transit within walking distance. More stops are needed. To determine where, we first stored the addresses that have zero transit stops in walking distance separately from the master list of addresses. We ran the k-means algorithm on this set to determine the placement of new transit stop locations, maximizing the number of houses formerly without transit covered by the stops. When placing one of these optimal stops, we operated under the assumption that no transportation system is better than the other for a given location; in other words, all transportation methods were treated equally. Therefore, the methods of transportation for our optimal transit stops are not specified, and can be Hubway, subway, or bus stops. Next, we looked at income per capita data for 15 neighborhoods in Boston compiled over the course of the past five years. We summed up the addresses in those neighborhoods which have zero transit stops within walking distance and plotted them against the income per capita of said neighborhoods. From here, we calculated an R^2 value of the distribution, to determine if any correlation exists.

The Data Sets:

- Master Address List (taken from Property Assessment)[1]
- Hubway Stops[2]
- MBTA Bus Stops[3]
- MBTA Subway Stops[4]
- Income Data Per Capita (This dataset was later updated to add new income data)[5]

From these data sets, we combined some of them into new data sets:

- **addressValue.py -** This file combines all bus stop and Hubway data with a given address. We see how many of these stops/stations are present, relative to a given address.
- **transit.py -** This file combines all Hubway and bus stop data together into a single database.
- **neighborhoodZipCodes.py -** This file uses the addresses and income repos to get a list of zip codes, assign them to a neighborhood name, and then assign the zip codes/neighborhoods an income.
- **constraintSatisfaction.py -** This file creates a repo of all the houses without any buses, subways, or Hubway stops.
- **kmeans.py -** This file uses the repo from constraintSatisfaction.py and uses a k-means algorithm to determine where optimal stops (of any of the three transport systems) should be placed.

Results: After running constraintSatisfaction.py, we were presented with 1709 addresses that do not have transit stops nearby. Therefore, not every address in Boston has at least one form of alternative transportation within walking distance. Next, we looked to placing our optimal transit stops via k-means. In order to determine the number of clusters (i.e. transit stops) that both minimizes error (the sum of distances to their closest cluster center) and maximizes the number of clusters until increasing the number of clusters does not improve accuracy, we selected four clusters to place.

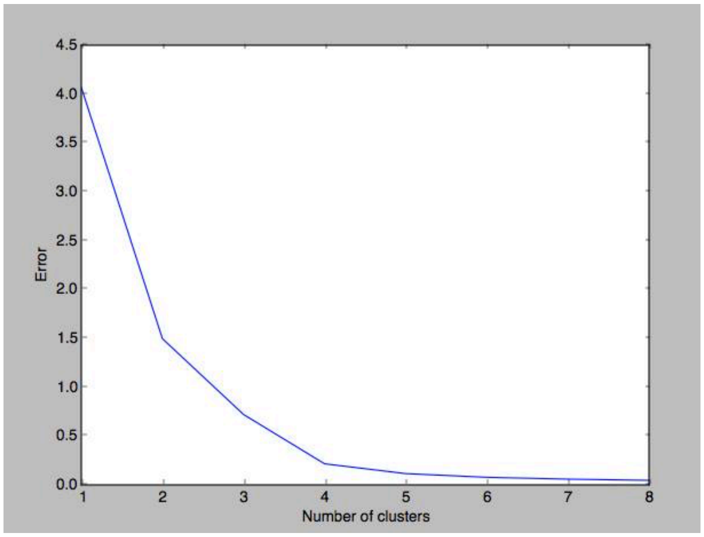


Figure 1: Determining the Ideal Number of Clusters to Use

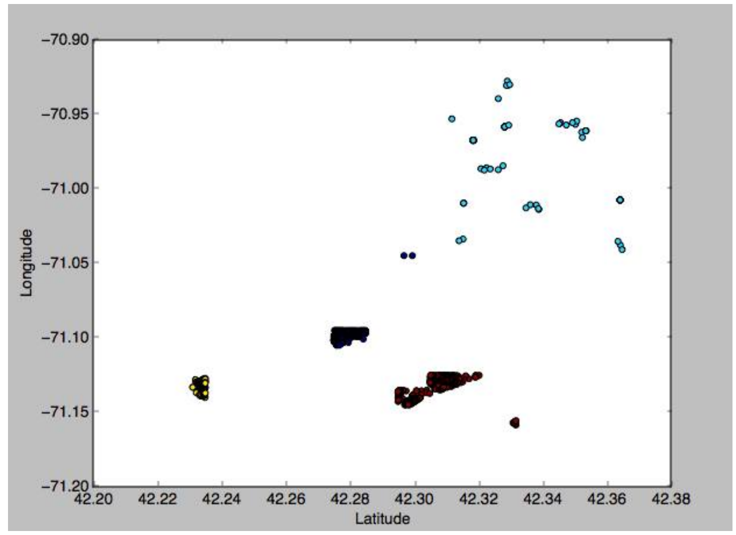


Figure 2: k-means Clusters

After running kmeans.py, we were presented with 4 clusters of latitudes and longitudes, and their respective centers (i.e. new stops), as shown in Figure 2. Using Leaflet.js, shown in Figure 3, we plotted existing transit stops and our proposed new stops on a map. Close-ups of each optimal stop can be seen in Figures 4-7.

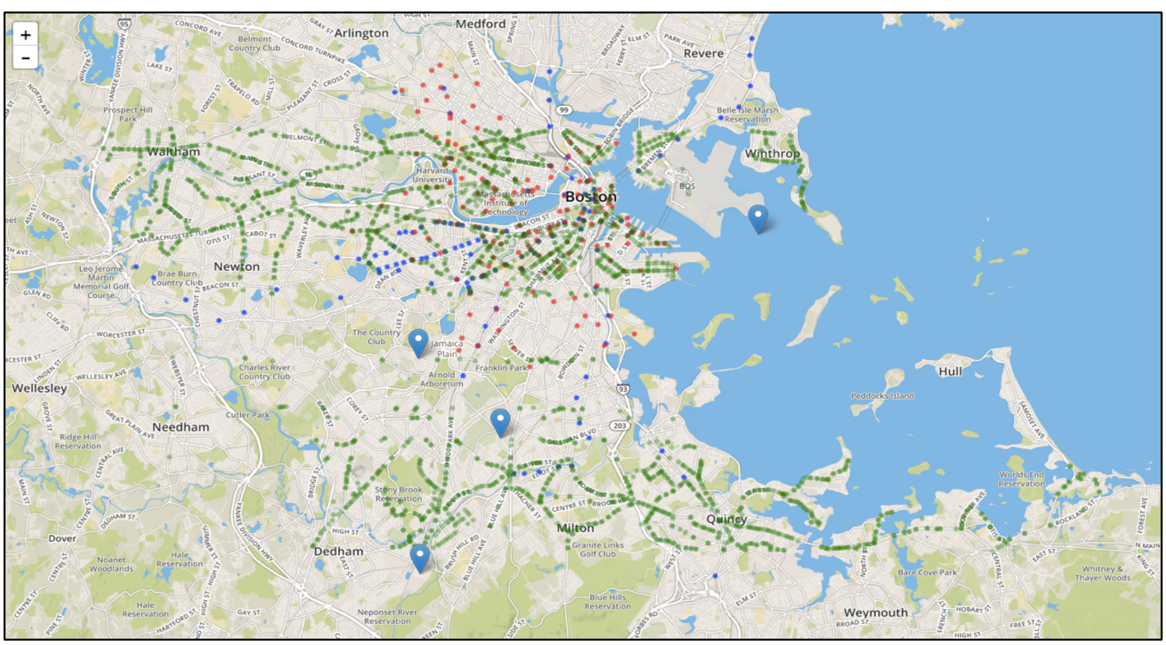


Figure 3: Existing Transit Stop Locations & Four Proposed Optimal Transit Stop Locations

Existing MBTA Subway Stop Existing MBTA Bus Stop Existing Hubway Stop Proposed Optimal Transit Stop

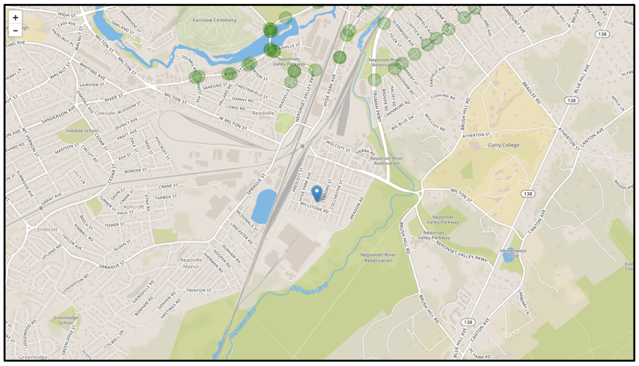


Figure 4: Location 1 - 53 Millstone Rd. -Readville

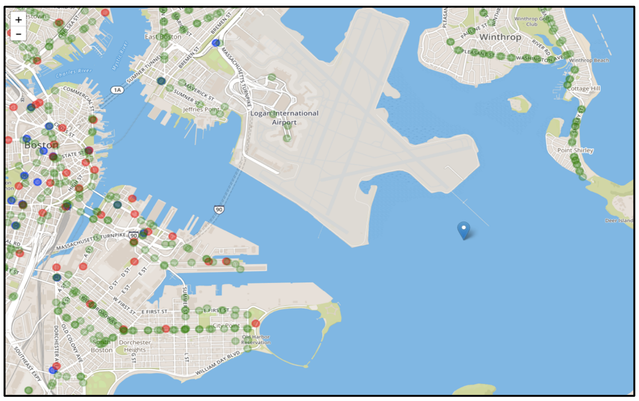


Figure 5: Location 2 - Boston Harbor

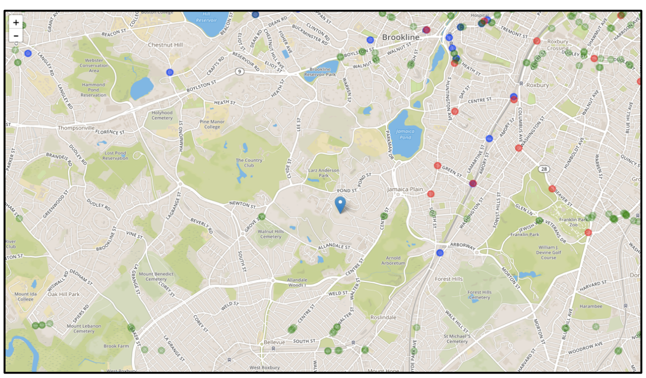


Figure 6: Location 3 - Showa Park - Jamaica Plain

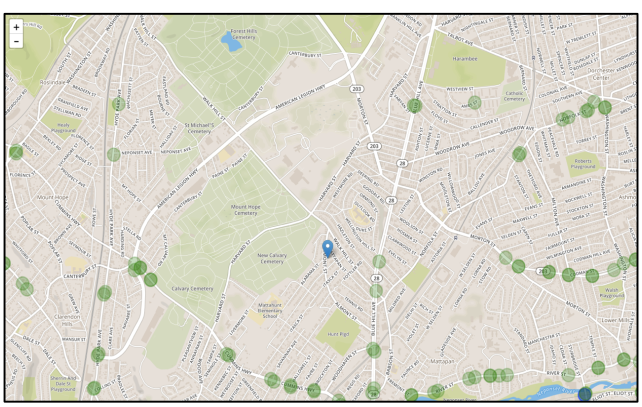


Figure 7: Location 4 - 182 Almont St. - Mattapan

- **Location 1:** Millstone Rd. near Curry College in Readville. This location is the farthest from the center of Boston, and MBTA Bus Service is the closest.
- **Location 2:** Boston Harbor. The algorithm disregarded the actual geography of the area, thus placing it in the water. The placement would make sense here, as it proposes a central location for South Boston and Winthrop addresses.
- **Location 3:** Showa Park in Jamaica Plain. This area is primarily served by numerous bus routes, and the closest subway stop is Forest Hills.
- **Location 4:** Almont St. in Mattapan. This location is far from the Ashmont-Mattapan high speed rail line, and no Hubway Stops are nearby.

In Figure 8, we see the distances of these optimal stops from existing transit stops. It is quite visible that the optimal stops are closest to already existing bus stops.

Optimal Stop Locations	Closest Existing Transit Stop	Distance (miles)
Location 1	**Bus Stop:** Readville St @ Norton St.	0.63 miles
Location 2	**Bus Stop:** Tafts Ave @ Mugford St. (Winthrop)	1.18 miles
Location 3	**Bus Stop:** Centre St opp Rambler Rd. (Jamaica Plain)	0.43 miles
Location 4	**Bus Stop:** Blue Hill Ave @ Fessenden St. (Mattapan)	0.25 miles

Figure 8: Distance Information for Proposed Optimal Transit Stop Locations

In addition, we created an interactive web service using Node, jQuery, and MongoDB, as well as the Google Geocoding and Maps Autocomplete APIs. To use the service, simply type in an address in the greater Boston area to receive the number of Hubway and MBTA subway and bus stops within walking distance from that address. During testing, we found that the numbers of stops were quite higher than expected for nearby areas. Upon review, we determined that this was due to floating point precision with latitude and longitude. This same issue was present throughout our research. Specifically, the different rounding methods that Python and JavaScript use accounted for the discrepancies in the results. This issue would require more time than allotted to resolve. Screenshots of this service in action can be seen in Figures 9 and 10.

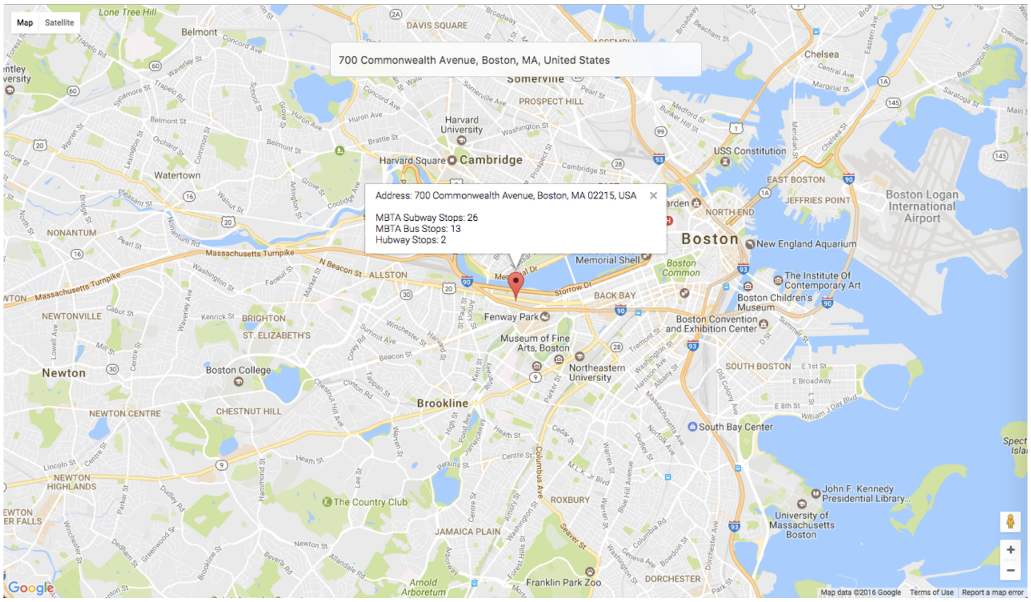


Figure 9: Results from searching 700 Commonwealth Ave. into the Service

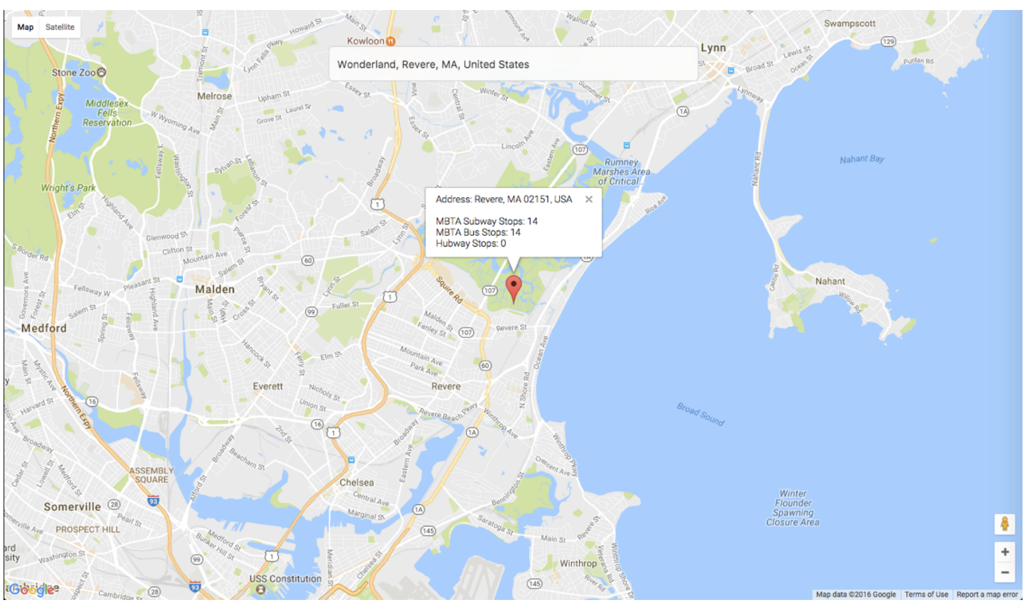
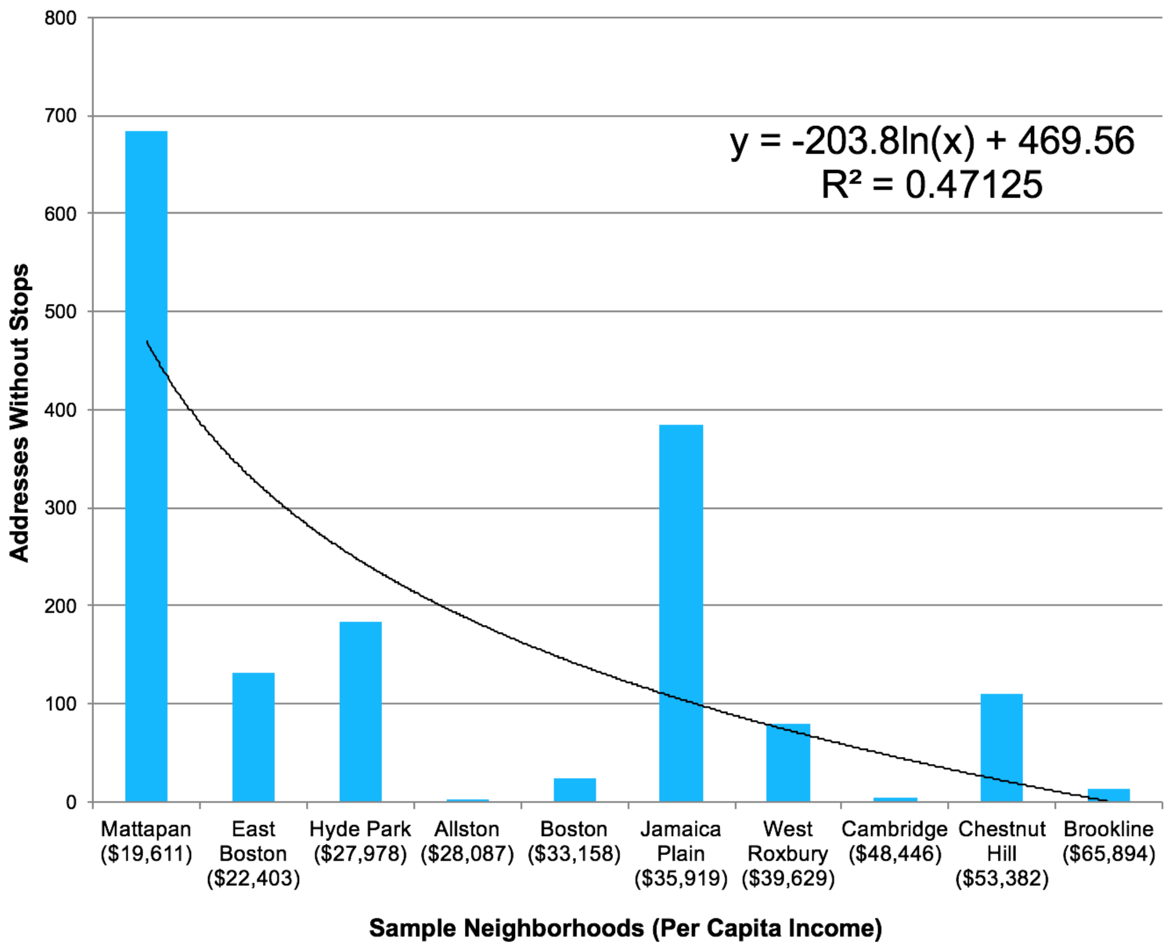


Figure 10: Results from searching Wonderland (subway stop) into the Service



s

Figure 11: Addresses Without A Transit Stop Within Walking Distance

In Figure 11, we plotted the per capita income of the neighborhood against addresses without income. This illustrates that it would be very difficult to determine the income just from the number of houses. While this does not indicate correlation specifically, it does show that they are not entirely related. This does not mean that lower income neighborhoods do have more or fewer stops because of the income, but just that it would be hard to determine the value of one from the value of the other. Although there is little (but not zero) correlation between income and lack of accessible public transit, this may be explained by distance from Boston and the ubiquity of the bus (the T alone may be much more likely to serve higher income communities). Regardless, we were still able to determine four optimal locations for stops of any sort.

Future Work: While our work took steps in the right direction, there is still a lot of work to be done. Immediately, taking a closer look at each type of stop individually would elucidate a lot of information about those individual means, and perhaps indicate a weighting system for each stop (how much each type of stop should count in our comprehensive analysis). For example, does the T only serve the higher income areas, and are Hubway stops only in Boston and Cambridge proper, while buses serve more of the greater Boston area? With this, we could also determine which locations need which transit the most (in this report we just determined *where* transit is needed, and not *what*). In these analyses, each should have their own R^2 and p-value, to properly determine correlation between the data. Then, adding the Commuter Rail may indicate if there is a similar discrepancy between income (or lack thereof) and alternative transit. Of course, many other factors would need to be taken into account, such as who is taking the T to where and why. Are those riding the T of a lower income, or are they just commuters? We would also need to look at the surrounding area, distances from other stops, etc.

Conclusion: Our project provided an excellent foundation to really build upon in trying to answer the incredibly difficult political question of whether or not income has affected the development of Massachusetts transit; delving into the details can only aid in finding a solution. The bus system might be widespread, but is the T? And are the prices of the bus system fair to lower income communities? In answering some questions--are there gaps in Greater Boston transit--we've only dipped our toes into an ocean of other questions.

1. https://data.cityofboston.gov/Permitting/Property-Assessment-2016/i7w8-ure5
2. http://bostonopendata.boston.opendata.arcgis.com/datasets/ee7474e2a0aa45cbbdfe0b747a5eb032_4
3. http://bostonopendata.boston.opendata.arcgis.com/datasets/f1a43ad3c46b4ac89b74cdaba393ccac_4
4. https://www.mbta.com/uploadedfiles/MBTA_GTFS.zip
5. https://github.com/Data-Mechanics/course-2016-spr-proj/blob/222e4fd34ad436932b58c44cdf8b31c2e9da27c4/jlam17_mckay678/data/Boston_IncomePerCapita.json