Optimization of Boston T-Stops for Access to Schools

CS504 Data Mechanics

Allen Qu

Introduction

Boston is home to many groups of people of different socioeconomic statuses. This project explores how different factors such as proximity of T-stops to schools affect education attainment as well as how education attainment affects poverty throughout neighborhoods in Boston. According to Boston Public Schools, public school transportation accounted for $110 million, or 11% of the city's budget (the 2nd highest in the nation). Currently, the city has public transportation infrastructure and train routes set up to serve the different neighborhoods. This project aims to examine how T-stops can be optimized to aid in access to education and reduce the cost of public school transportation.

Transformations and Statistical Analysis

To explore the relationship between education attainment and the distance from T-stops to schools, the following data was retrieved:

1. The location of T-stops from ArcGIS Hub
2. The locations of all public schools and colleges from Boston Open Data
3. Education attainment for people ages 25+ throughout the different neighborhoods of Boston from Boston Open Data

The T-stops and schools data were transformed to find the average distance from each T-stop in Boston to each school in a neighborhood. The education data was transformed to only include the percent of people with bachelor's degrees in the 2000s decade in each neighborhood and plotted against the average distance from a T-stop to a school in each neighborhood (Figure 1).

To explore the relationship between income and education attainment, additional data was retrieved from Boston Open Data on vulnerable populations within Boston. The data was transformed to find the percent of people who were low income in each neighborhood, and plotted against the percent of people with bachelor's degrees (Figure 2).

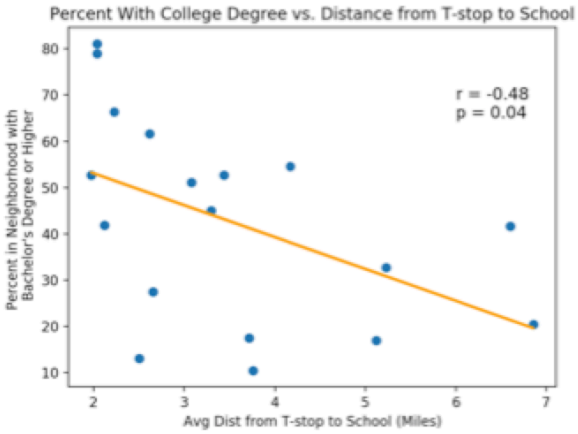

Figure 1: Relationship between Education Attainment and Distance from a T-stop to a school in a neighborhood

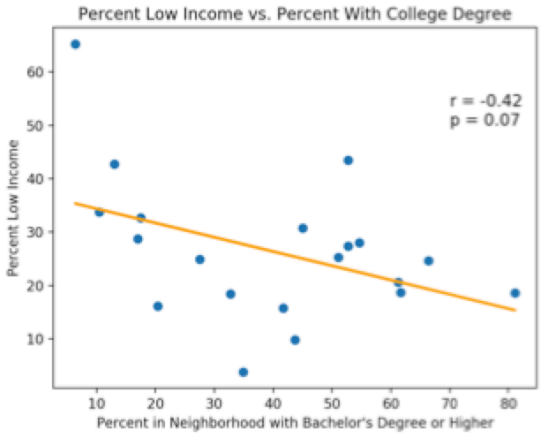

Figure 2: Relationship between Income and Education Attainment in a neighborhood

Techniques

A K-means algorithm was used to cluster the closest T-stops around the average coordinates of schools in each neighborhood. The optimization algorithm minimizes the Euclidean distance from each cluster of T-stops to the average coordinates of neighborhood schools. Since there were 18 neighborhoods in Boston that contained schools, the K-means algorithm gave an output of 18 centroid coordinates for optimized T-stops closest to the original average coordinates of schools.

Visualization

An interactive web map (Figure 3) was created with Python Flask and the JavaScript Leaflet library to showcase the 18 optimized T-stops from the K-means algorithm. Data is retrieved from a MongoDB database for the Flask web application with a RESTful web API. Each neighborhood is color coded with a specific shade of green representing the percent of people who have a bachelor's degree or higher. In the actual web map, users are able to float over one of the neighborhoods in Boston to get more information on the average distance from a T-stop to a school, percent of people with a bachelor's degree or higher, and the percent of people who are low income.

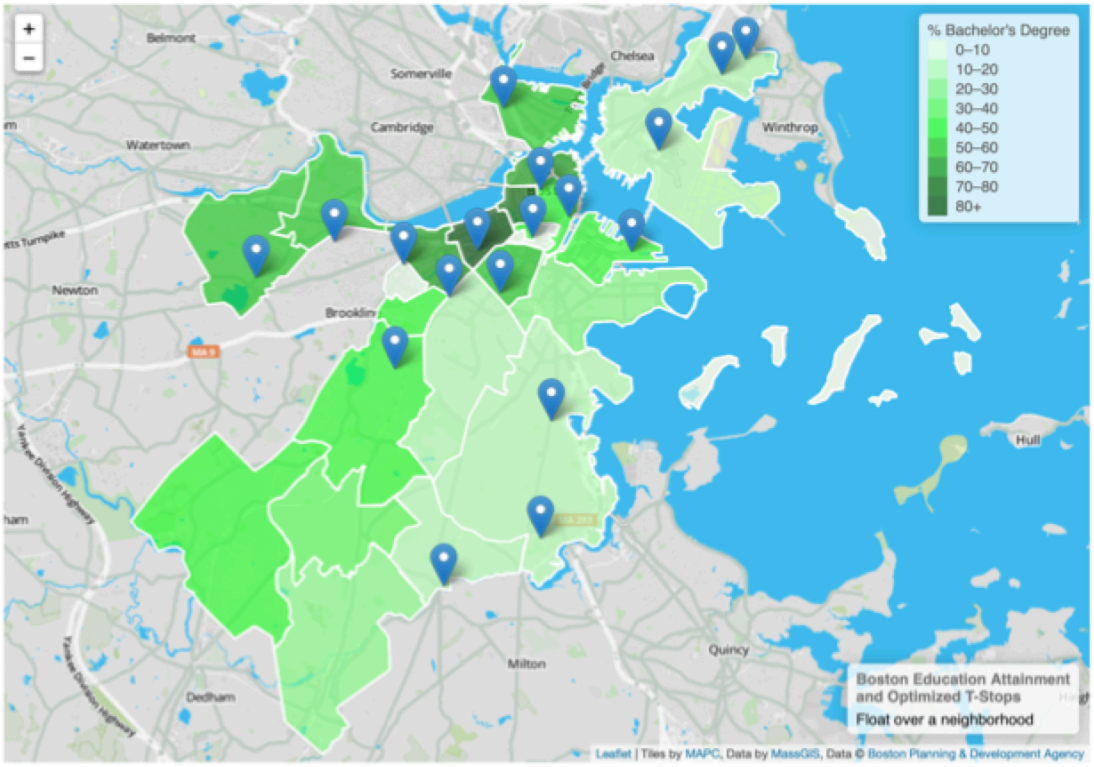

Figure 3: Interactive map showing the 18 optimized T-stops closest to the average coordinate of a school in each neighborhood that had schools. Float over a neighborhood to get more information on the average distance from a T-stop in Boston to a school in that neighborhood, percent of people with a bachelor's degree or higher, and percent of people who are low income.

Results

There is a moderate negative correlation of -0.48 between the average distance from a T-stop to a school in a neighborhood and the percent of people with a bachelor's degree or higher. This means that as the average distance from a T-stop to a school increases, the percent of people with a bachelor's degree decreases. Since the p value is 0.04, the results are significant. The correlation might be explained by the fact that many T-stops are concentrated around areas in the northern parts of Boston, where many of the neighborhoods have a large percentage of people with a bachelor's degree or higher.

Since many of the T-stops are concentrated around the northern parts of Boston, the K-means algorithm will be likely to cluster those T-stops together when computing an optimal T-stop to an average coordinate of a school in a neighborhood. Thus, many of the optimized T-stops seem to be clustered around the northern parts of Boston, where the neighborhoods with the most percentage of people with bachelor's degrees or higher are located.

There is a relatively weak negative correlation of -0.42 between the percent of people that are low income in a neighborhood and the percent of people with a bachelor's degree or higher. This means that the higher the percentage of people with a bachelor's degree in a neighborhood, the lower the percentage of people that are low income. However, the p value of 0.07 suggests that the results aren't significant. The percentage of people with a bachelor's degree or higher may not entirely predict the percentage of people with low income, and there may be other factors that affect the percentage of people in a neighborhood that are low income.

Future Work

The K-means algorithm used to find the optimal T-stops to schools in each neighborhood seems to give T-stops that are mostly located in neighborhoods where the percentage of people with bachelor's degrees or higher are already the highest in Boston. This is because the algorithm only optimizes the distance of current T-stop clusters to average neighborhood school coordinates, and doesn't take into consideration the percentage of people with a bachelor's degree or higher or the percentage of people who are low income. In the future, it might be interesting to come up with an algorithm that takes into consideration those two factors and see where the algorithm places optimal T- stops. More weight will be put on neighborhoods with lower percentages of people with bachelor's degrees or higher and people of low income for a T-stop to be placed. The project can also go further and investigate how these new optimal stops can reduce the hefty cost of Boston's public school transportation.