Optimizing Green Line T Stops

Steve Jarvis, Cristina Estupiñán

Overview

The Green Line seems to stop too frequently. Presuming the best location for a stop is where one doesn't already exist and people would most use one, we think analyzing the popularity of stops against the distance to alternatives can help decide which existing stops are most valuable and which can be omitted. To make such a determination, we use a variant of the k-means clustering algorithm that also takes into account a given weight for each point. The weight, in this case, is a function of the stop's popularity and the proximity to the nearest alternative.

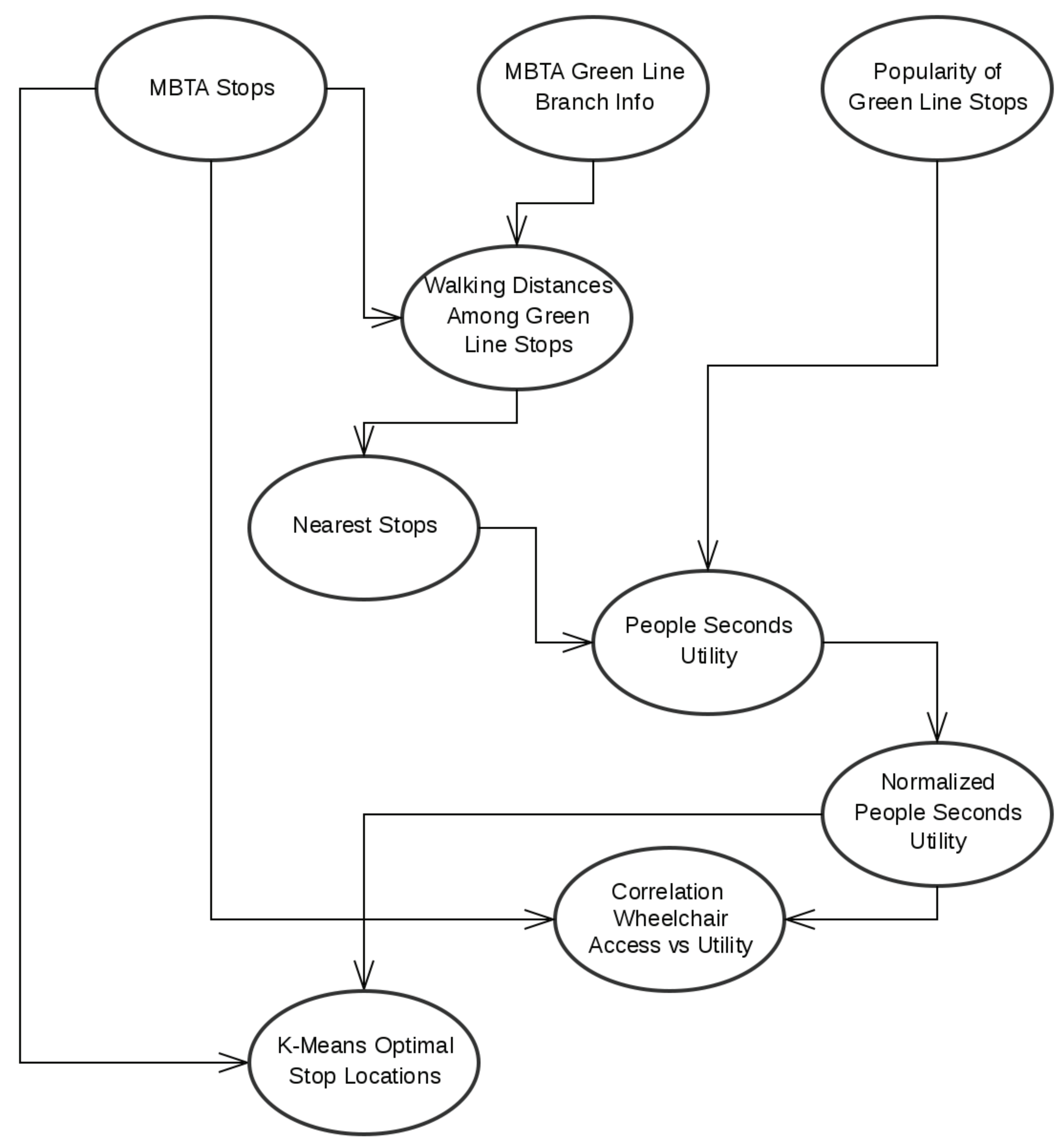

Figure 1: Visual of our dataset flow. We begin with three datasets provided by the MBTA and use these to dervie the other datasets.

Correlation of Score and Access

We want to find a correlation between our utility and the wheelchair accessibility of each stop. Our utility is a function of the popularity and easy alternative stops; the higher the utility, the better. We calculate our correlation coefficient as 0.395 and our p-value as 0.0009996. Our low p-value indicates that we have a significant correlation between a high utility and wheelchair accessibility.

Utility Measurement

Our utility measurement is a score that uses both the popularity of each stop and the walking time to the next nearest stop. A high score is good; this means that the stop is valued and saves the greatest amount of time for the collective commuting group. The measurement is scaled from 1 to 1000.

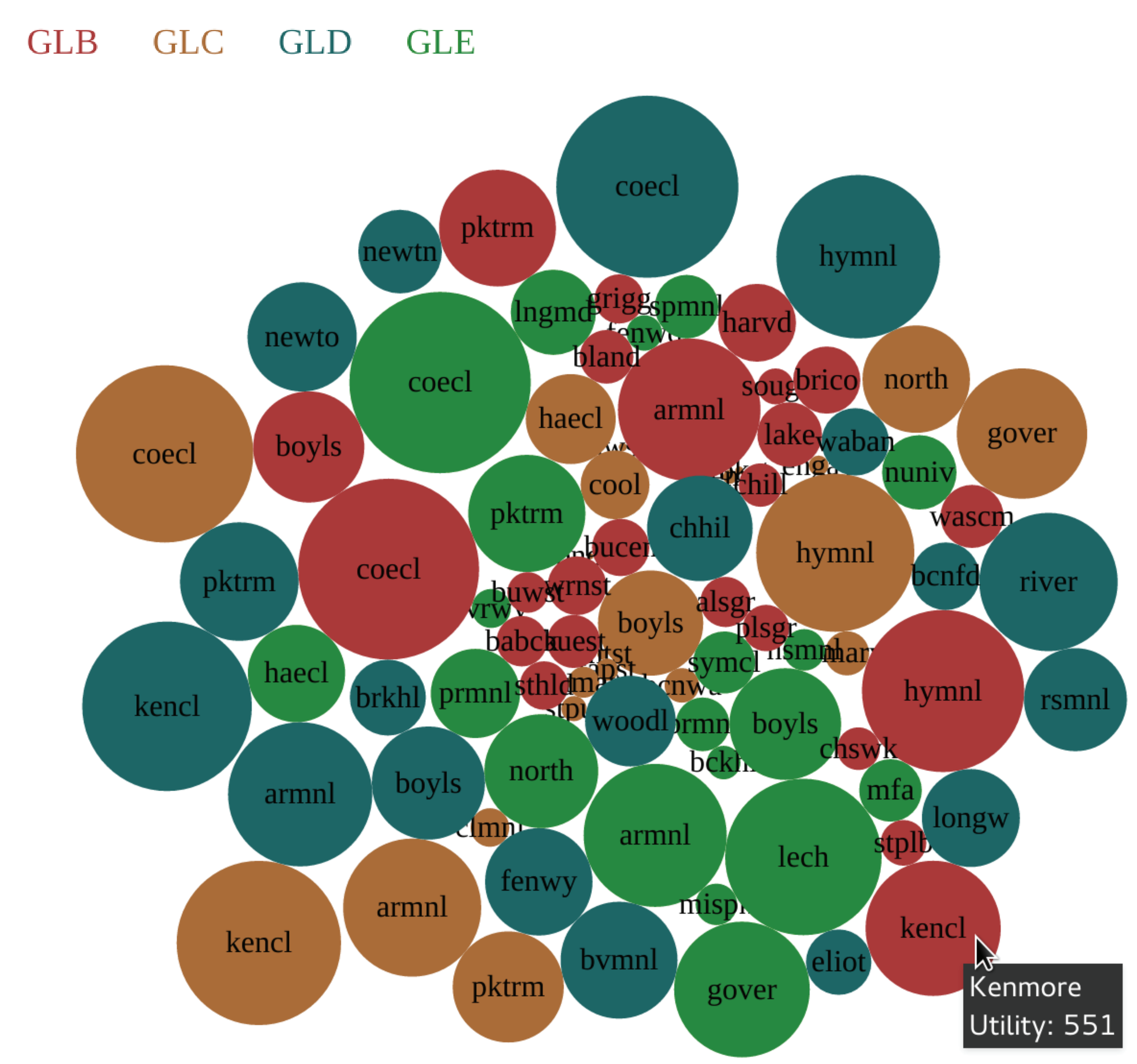

Figure 2: Visualization of our utility score. The size of the bubble relates to how large the utility score is. Stops are counted seperately for each branch they serve.

Weighted *K*-Means Algorithm

To generate optimal stops, we use the weighted *k*-means clustering algorithm. This algorithm divides the dataset into *k* clusters that optimize some measurement. In our case, the measurement is the distance between exisiting stops and the *k* means. We use the utility measurement to give a weight to each stop which would influence the position of each mean. Once the algorithm is finished running, we have a set of optimal stops for each branch.

Optimal Stops

In 2014, the MBTA proposed removing four stops from the B branch: BU West, St Paul St, Pleasant St., and Babcock St., replacing them with one new stop just west of BU West and a second between Babcock St. and Pleasant St. Our algorithm agreed and, if we were to use it to improve the B branch by just one stop, would remove the Pleasant St. and Babcock St. stops, replacing them with one in between.

Figure 3: The current stops on the MBTA Green Line.

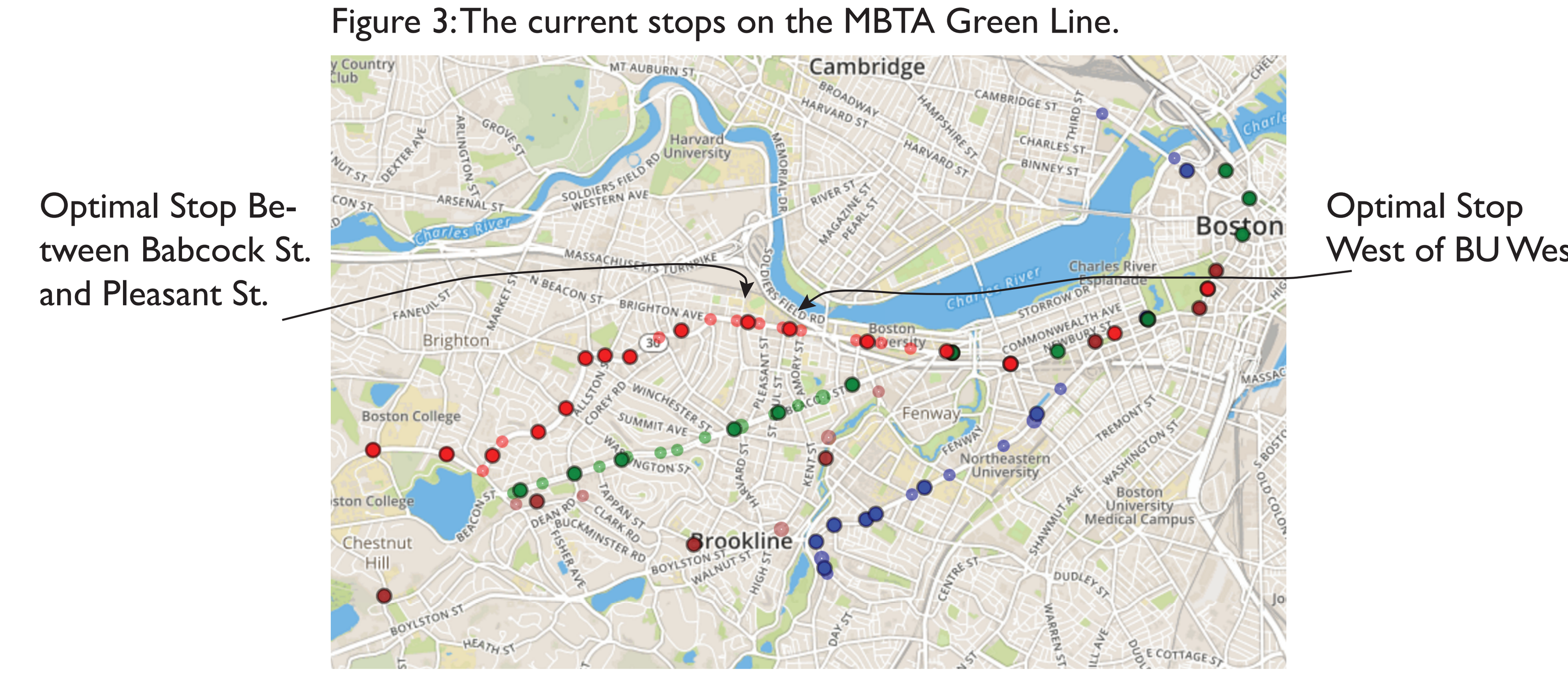

Figure 4: Our *k*-means generated stops when *k* is equal to 16, 13, 15, and 9, for the respective branches.

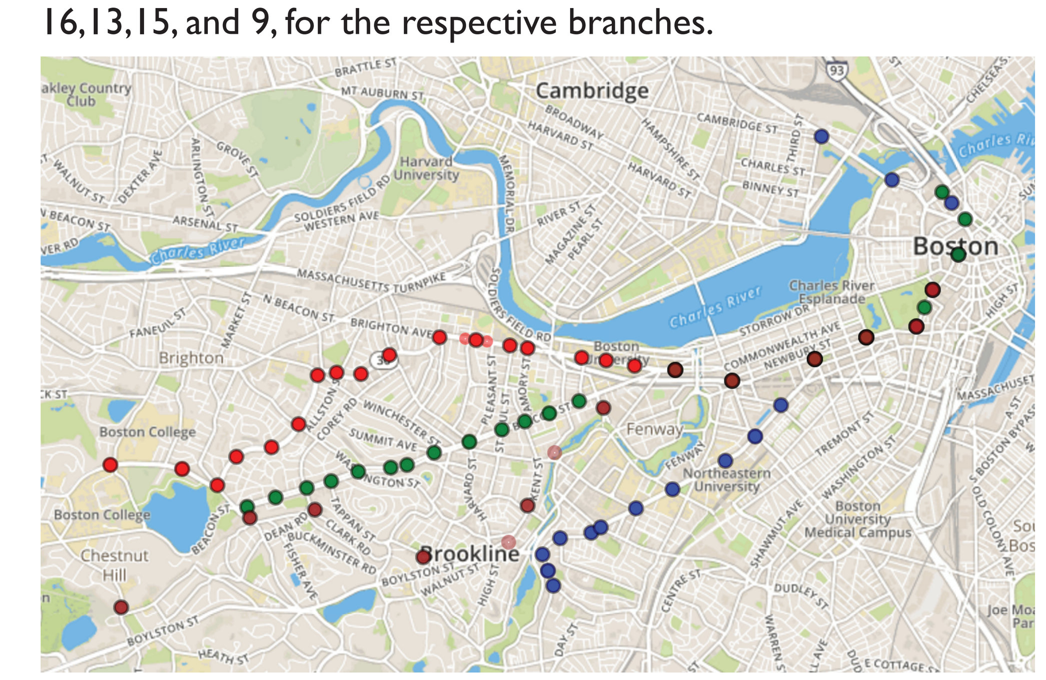

Figure 5: Our *k*-means generated stops when *k* is one fewer than the number of current stops on each branch. This eliminates one stop on each branch.