

### Abstract

Boston Public Schools issued a request to revamp their public school busing system. Given their many datasets we strove to combine them in interesting ways to help solve routing and timing issues, focusing on these sectors:

- bus yard locations
- average distances between of students' locations

### Datasets

- **BPS transportation challenge :**
- buses, students, schools

### Optimizing Bus Yard Locations

**Goal:** Find the optimal number of bus yards to maximize efficiency for bus routes

### Algorithm one: k-means

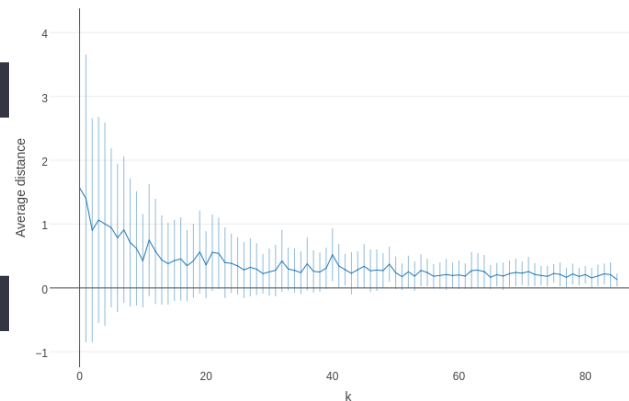
Use **k-means** to find the locations that minimize distances from schools to bus yards. To do so we analyzed how the different numbers of means affect the average distances from points to means

### Vincenty Distance

To calculate the distance between the mean and the schools we used **vincenty**

Shown below is an interactive graph that displays the average cost between all the points and the standard deviation for k means.

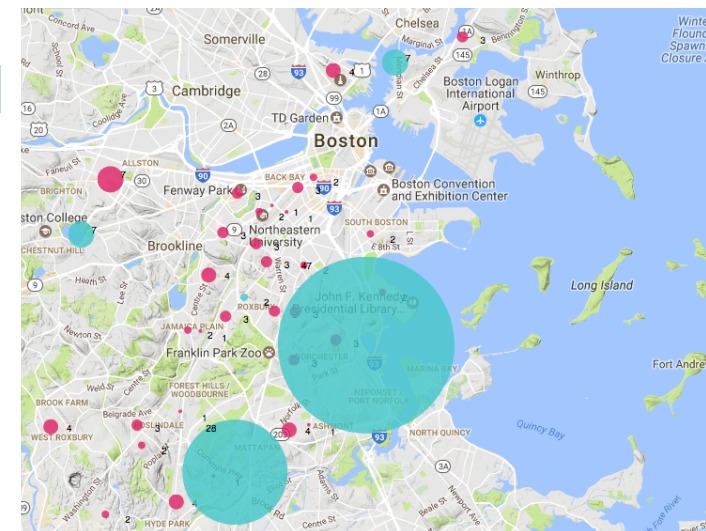
- **Plotly.js, d3, standard deviation**



### Optimal Bus Yards and Application

From the graph above, the optimal amount of means is consistently around half of the number of points in the dataset. We then took these means and mapped them to a map of Boston, with the circle sizes representing the number of schools they will service.

- **Google Maps** using **d3, CSS, HTML**



### Averaging Distances between all students

**Goal:** To produce useful data useful for buses to see how far they would have to travel between stops on average, in order for them to envision timing.

### Algorithm two: R-tree distances

Results: Used **vincenty** to generate a group of the closest 10 students to each student, and a group of students within a .5 mile radius of other students.

### Conclusions

These analysis will provide the foundation to enlighten further analysis in routing by using the optimal bus yard locations, the number of buses, and the distances between stops moving forward.