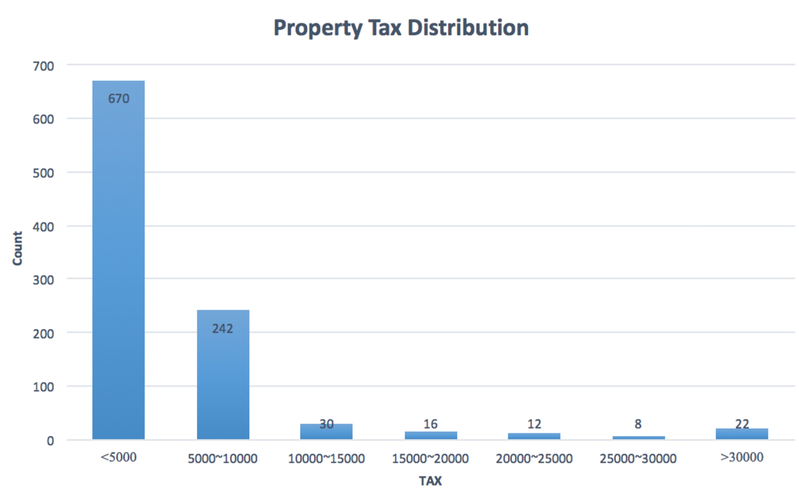
Boston Property Value

Introduction

We aimed to determine how dining and access to schools can have an impact on the value of real estate properties. Using data provided by City of Boston, we determined the correlation between property tax (a measure of value) and distance to the nearest restaurants and schools.



Data Source

We used 3 datasets including restaurant establishments, properties, and schools. Restaurant dataset contains information on the location and type of establishments. Property dataset includes location and taxes paid by the owner/s. School dataset contains information on the location and level of education the school offers.

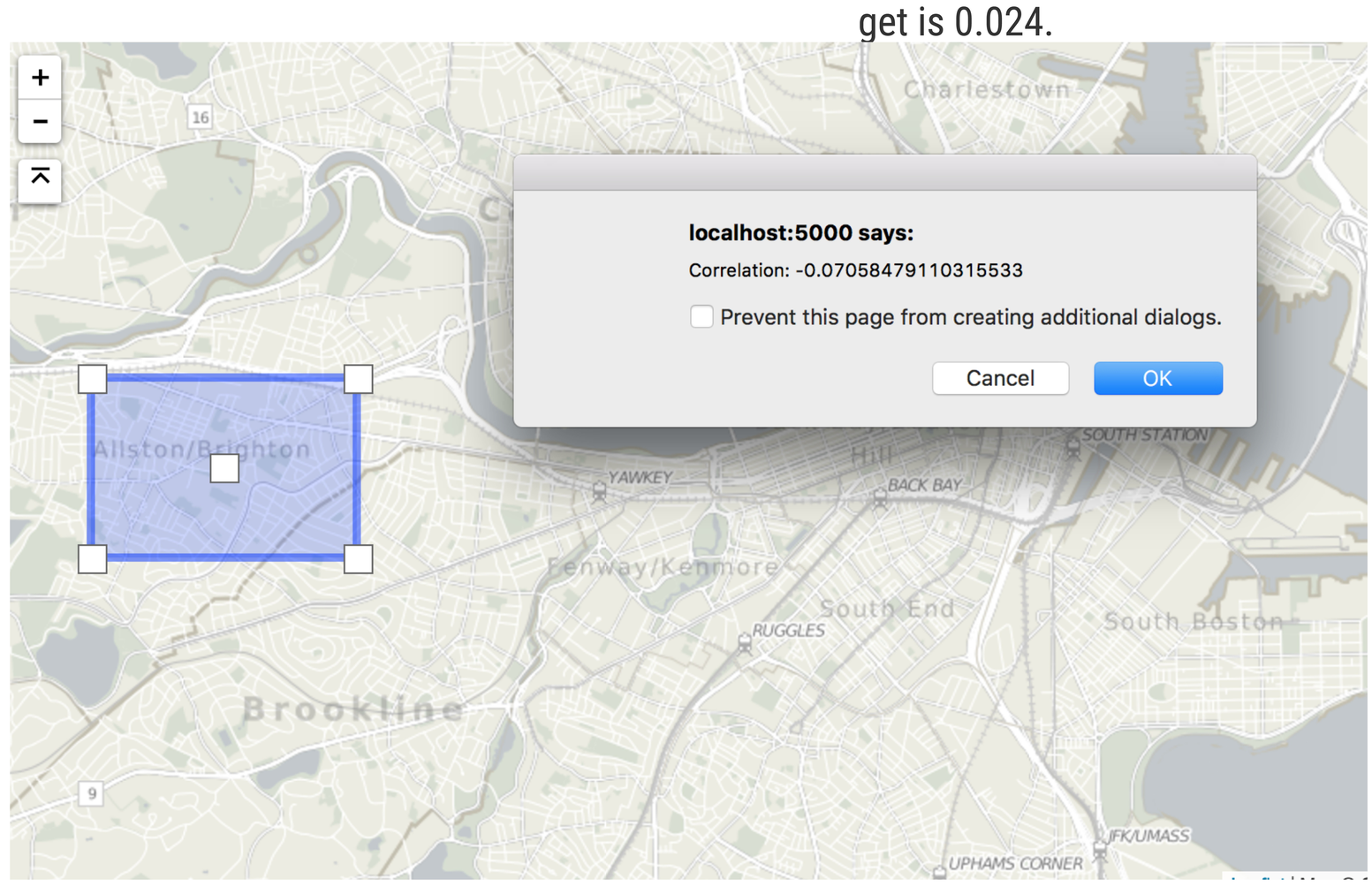
Method

The two main methods we used are K-means clustering algorithm and statistical analysis. We combined the coordinates of restaurants and schools. We want to measure the distance between the centroids of these restaurants and schools to properties. We also made changes to the K-means algorithm. Coordinates are floats so when the algorithm calculates the coordinates of centroids, they change by a very unnoticeable difference every iteration. Then we added a loop to check if the difference between previous centroids and centroids from the current iteration is less than 0.001 then stop the algorithm.
Another method we used was correlation analysis. We used the distance from a property to a closest mean of restaurants and schools and the property tax paid by the owner/s. We also checked the P-value with different number of permutations.

Future work

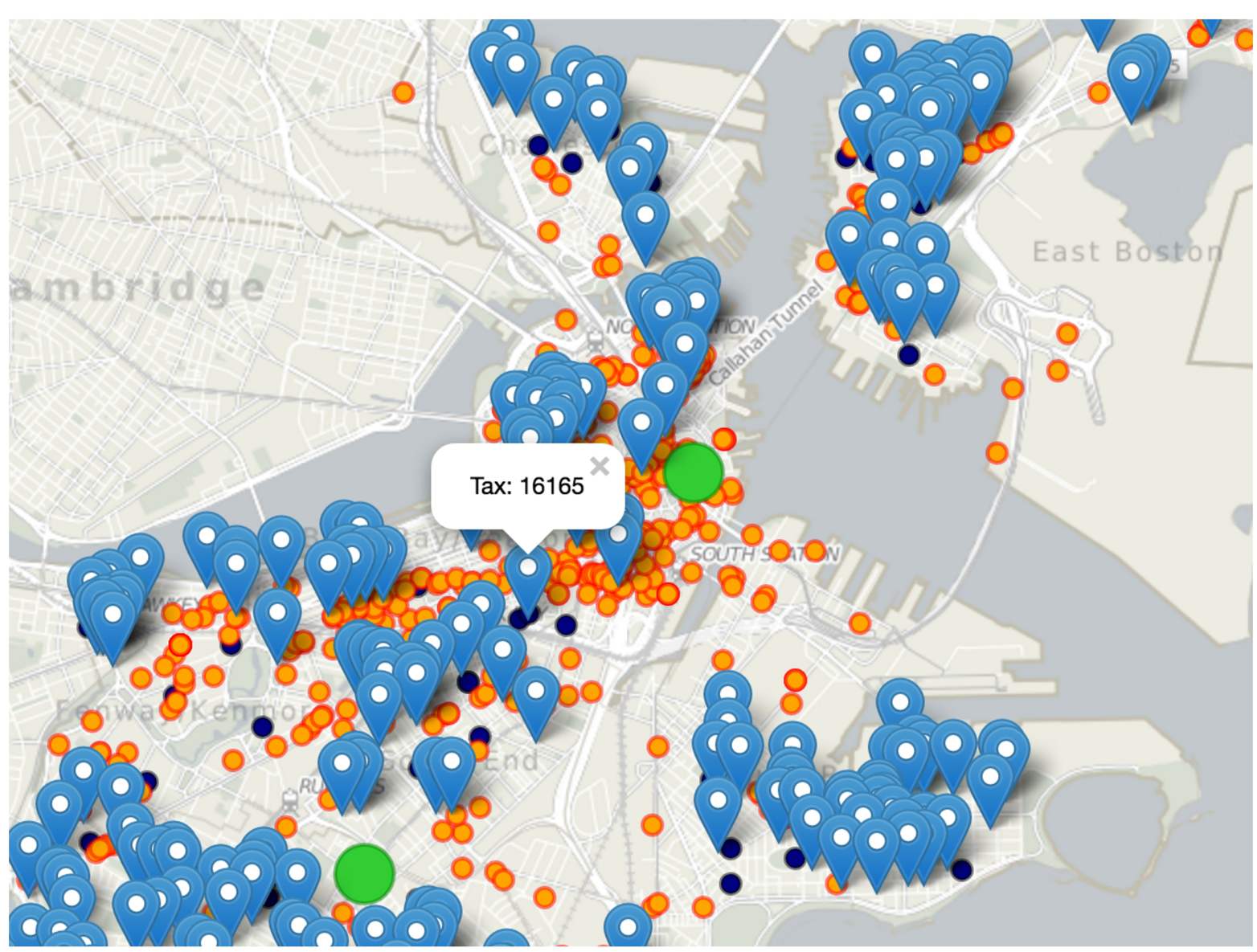
We should try to find more factors that affect real estate value in Boston. Perhaps quality of the property, lot size or other factors are correlated with property value. We could separate them into groups depending on ranking or scores.

This chart below is a screenshot of an interactive visualization that allows users to select an area and the dialog box would show the correlation coefficient between distance and value.



Results

Our K-means algorithm produces 4 centroids. They are located in Back Bay, Allston, South Boston and Weld Hills The correlation between property value and distance to the nearest centroids 0.145. We also calculated the P-value for this correlation coefficient. The best p value we get is 0.024.



Blue Marker: Properties Orange: Restaurants Green Circle: centers Navy: School