Public Facilities, Crimes and Average Income

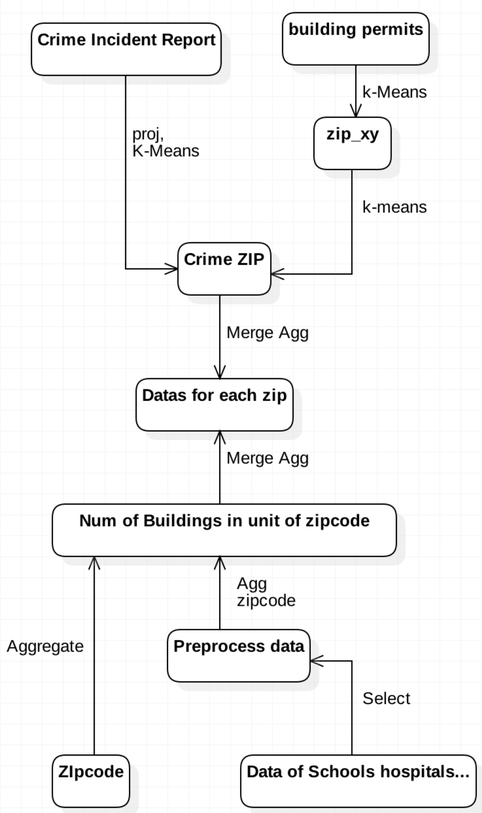
Yehui Huang, Yingqiao Xiong, Hongyu Zhou, Chang Gao

BOSTON UNIVERSITY

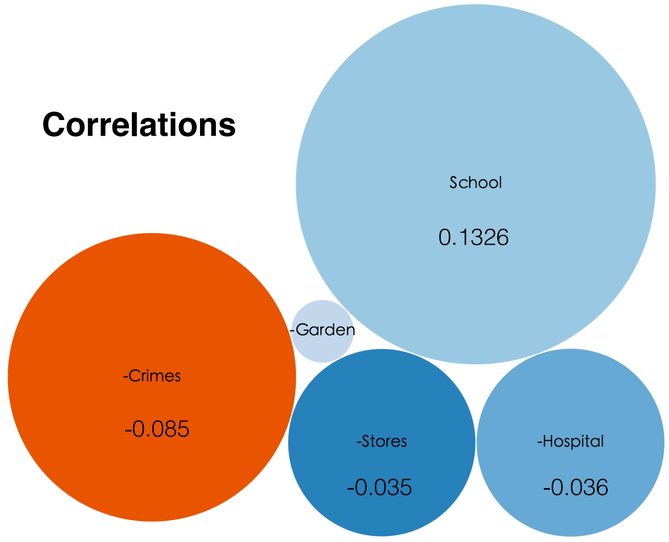
Introduction:

In this project, we attempt to analyze the relationships between average income of residents in certain neighborhoods and number of public facilities, crimes in the same neighborhoods throughout the greater Boston area. In order to analyze the relationship, we use the calculate the numbers of the hospitals, public gardens, schools, stores and crimes in each neighborhood, and then perform MapReduce algorithms on the data to derive the intended format and content. We also operate machine learning algorithms, such as k-means and linear regression to process data such that the correlation between each factor and income can be derived, and the numerical coefficient can be estimated.

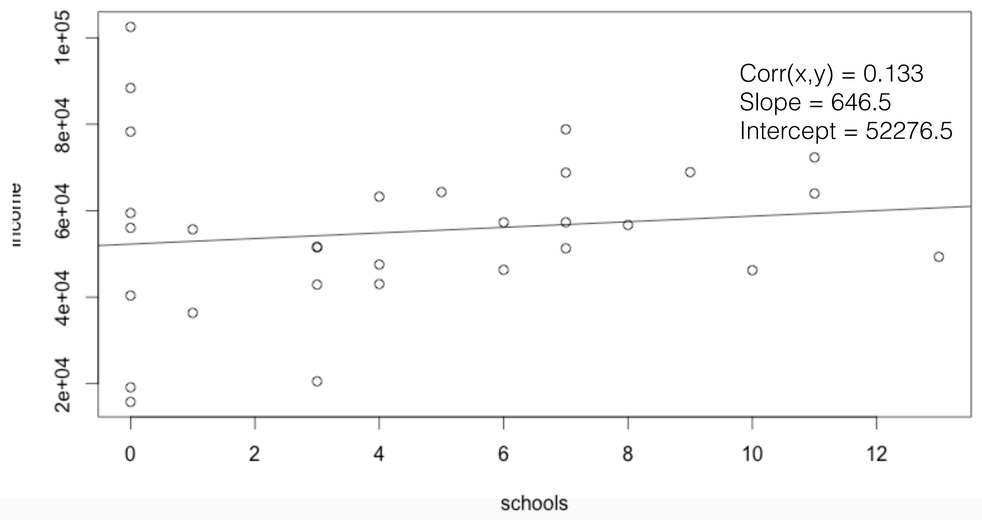
Process



Correlations

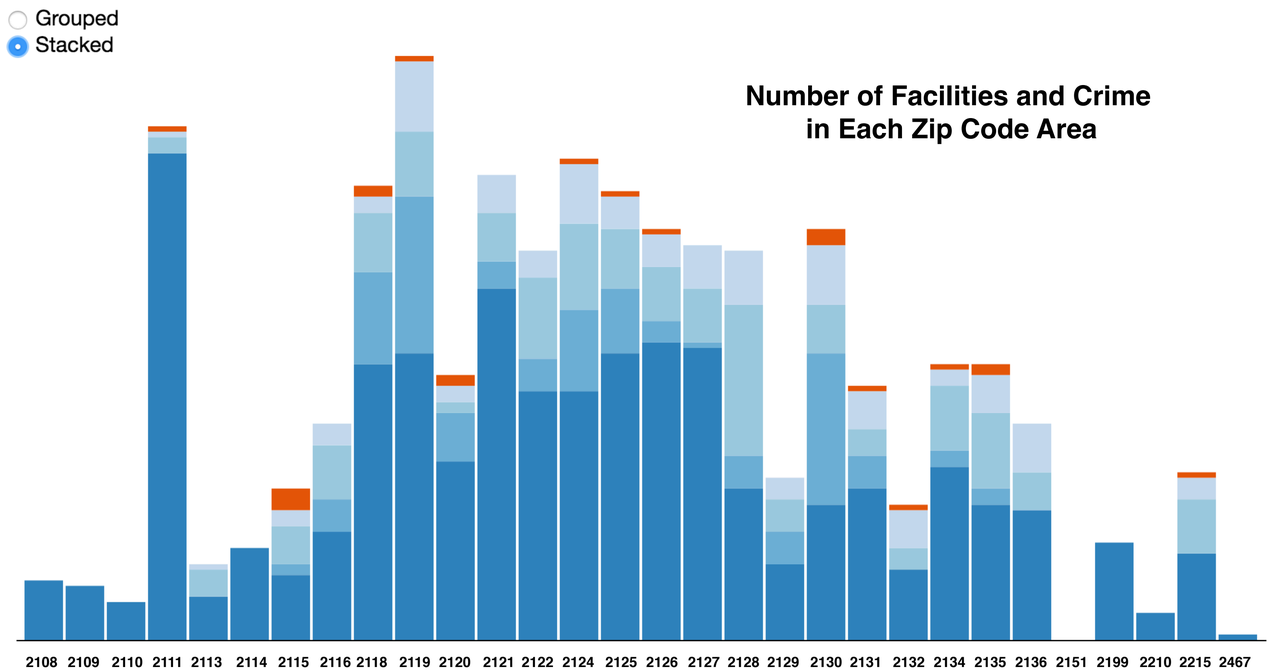


Linear Regression



K-Means Algorithm

To Generate optimal crime incidents in each zip code, we use the K-means algorithm.This algorithm uses the building permits to generate optimal clustering point for each zip code area. We then compare each crime incident point to each clustering point, and add them to the closest zip code point. Thus, a set of optimal crime incident point in their related zip code area is derived.



Conclusions and Possible Extensions

The correlations between income and each of the possible factors are vague, as most of the correlation coefficients fall in the range of [-0.1, 0.1]. The linear regression model does not fit well for the data sets, as the coefficients are noticeably close to 0. Thus, we shall consider other factors which could have affected the average income in certain areas, and also consider the correlations between many of the independent variables. In addition, the possible extension will be to provide suggestions to City of Boston regarding the issues of urban property management, such as the planning of constructions of schools and gardens.