Relationship Between Crime and Service Requests

Arjun Lamba (arjunlam@bu.edu)

INTRODUCTION

The goal of this project was to look at how the cleanliness of an area is related to the number of crimes that are committed in that area. To determine how well maintained an area is I looked at the number of 311 services requests, the number of pothole repair requests and the number of mayor hotline requests.

DATA

Five data sets were taken from the City of Boston Data Portal. One data set gave the locations of the crimes committed between 2015-2016. Four data sets dealt with the maintenance of Boston (i.e. 311, pothole repairs, mayor hotline and development) and they gave the locations of buildings that are part of the neighborhood development project and locations of the service requests along with the type of requests.

METHODS

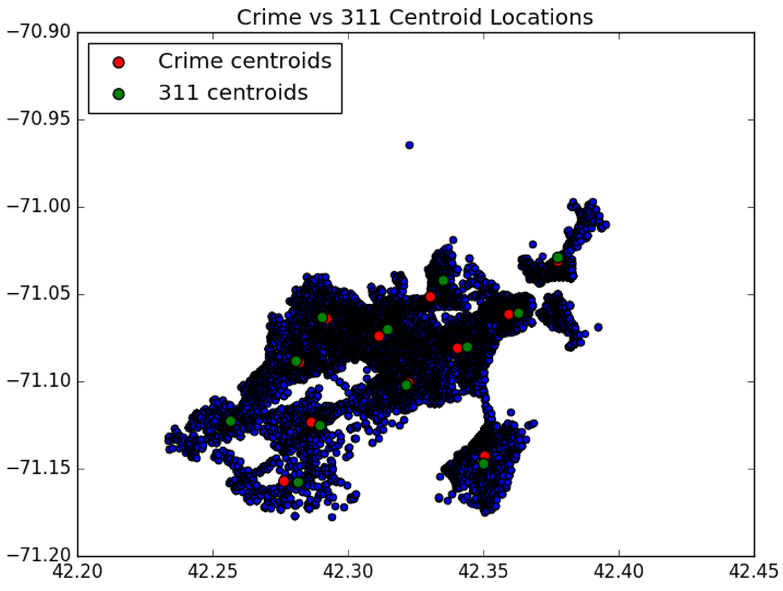
The K-means algorithm was used and basic statistics such as average, correlation and p-value were computed. Histograms of the x and y coordinates of crime locations and 311 service requests locations were also used.

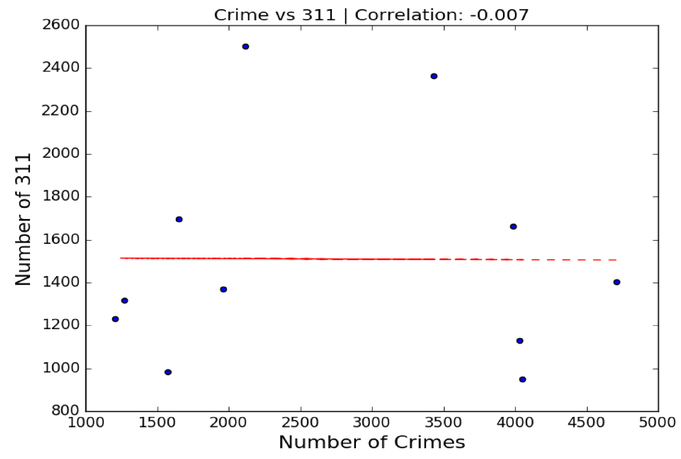
GROUPING BY ZIP CODES

Crime and the service request were aggregated according of the zip codes they occurred in. Then only zip codes that had greater than zero crimes were considered. The correlation coefficient and p-values were computed for crime verses 311, crime vs. hotline, crime vs. potholes and crime vs. all the requests added up.

CENTERS OF CRIME AND SERVICE REQUESTS

The k-means algorithm was run on crime and 311 requests locations. This was done to find where the crimes and 311 requests were grouping. This allowed me to see if the centroids had similar locations which would mean that crime and 311 requests occurrence are related.





RESULTS

When grouping the crimes and service requests according to zip codes and then calculating the correlation the correlations are all negative and between 0 and -0.5. So the more service requests an area generates that less crime that area is generating. The centroid locations of crime and 311 requests that were generated by the k-means algorithm were very close and for some k cluster numbers the centroids were overlapping.

CONCLUSIONS AND FUTURE WORK

The negative and near zero correlation coefficients means that there is either no relationship and whenever a relationship does exist it tends to be inverse (more requests equals less crime). This was in contrast to the k-means results where the clusters of the crimes vs. 311 requests were shown to be in very similar locations suggesting that areas that produce the most 311 requests also produce the most crimes. These two analysis also show that how we aggregate things will have a major impact on the relationships of the data sets. In the future I would like to calculate the gap-statistics for the k-means and analyze the location histograms.