Carole Sung, Eric Chang, Ivan Wong

CS591

Professor Andrei Lapets

Optimal Locations for Government Development

Purpose

The purpose of this project was to extract data from the open-source Boston data portals, and recognize which subareas of the Greater Boston Area are in the greatest need for government development. In approaching this, we would aggregate points of government development such as police stations, streetlights, schools, and hospitals. With these, we would scale the amassed values to represent their respective "development scores" (this scale is currently arbitrary). We would then sum the development scores to provide the final scores for each neighborhood. With these, we then scale these development scores to the area's mean property value, and normalize the final values on a scale of 0-100. Higher normalized values indicate a greater need for government development, and lower normalized value indicate a lesser need.

Datasets

The datasets we've used for analyzation and the generation of our data are from:

- Property Assessments of 2016 from City of Boston data portal
- Cambridge schools from City of Cambridge data portal
- Boston Street Lights from Analyze Boston data portal
- Hospital Locations from Analyze Boston data portal
- Boston Public Schools from City of Boston data portal
- Boston Police Stations from Analyze Boston data portal

Algorithms and Transformations

aggpropValue - Aggregates properties based on zip code, and average the property values. Utilizes aggregation, projection, and union.

camSchoolsAgg - Aggregates schools in Cambridge based on coordinates, converting coordinates into zip codes, and then aggregating.

schoolsAgg - Aggregates Boston Public Schools based on zip code, utilizing aggregation and projection.

hospitalAgg - Aggregates hospitals in Boston based on zip code, utilizing aggregation and projection.

lightCoordinates - Aggregates streetlights in Boston based on coordinates, converting to zip code, utilizing aggregation and projection

policeAgg - Aggregates police stations in Boston based on zip code, utilizing aggregation and projection.

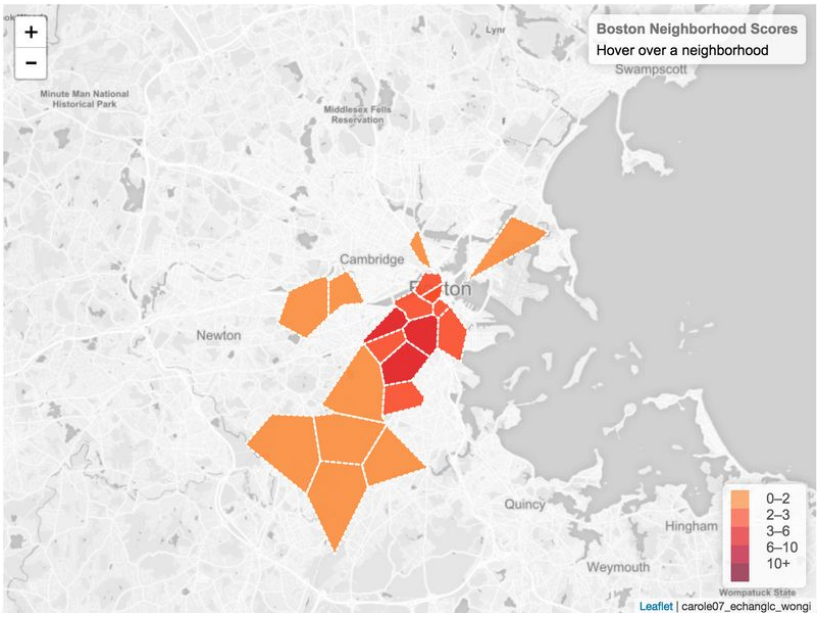
developmentScore - Aggregates all of the data. Projects into tuples of (zipcode, developmentCount), aggregates based on zipcode, then calculates development score by arbitrarily dividing the average property value by the number of buildings nearby, then normalizing the values to a score of 0-100.

Data/results generated

Zip Code	Development Score
02126	0.11
02130	1.48
02132	0.09
02113	100.00
02129	3.60
02128	7.61e-15
02127	8.74
02118	0.05
02120	5.74
02109	5.48
02111	22.33
02124	3.52
02125	8.04
02119	1.10
02134	2.07
02135	2.81
02116	0.41
02114	0.11
02121	1.41
02122	0.22
02115	21.96
02215	4.21
02131	0.89
02136	0.16

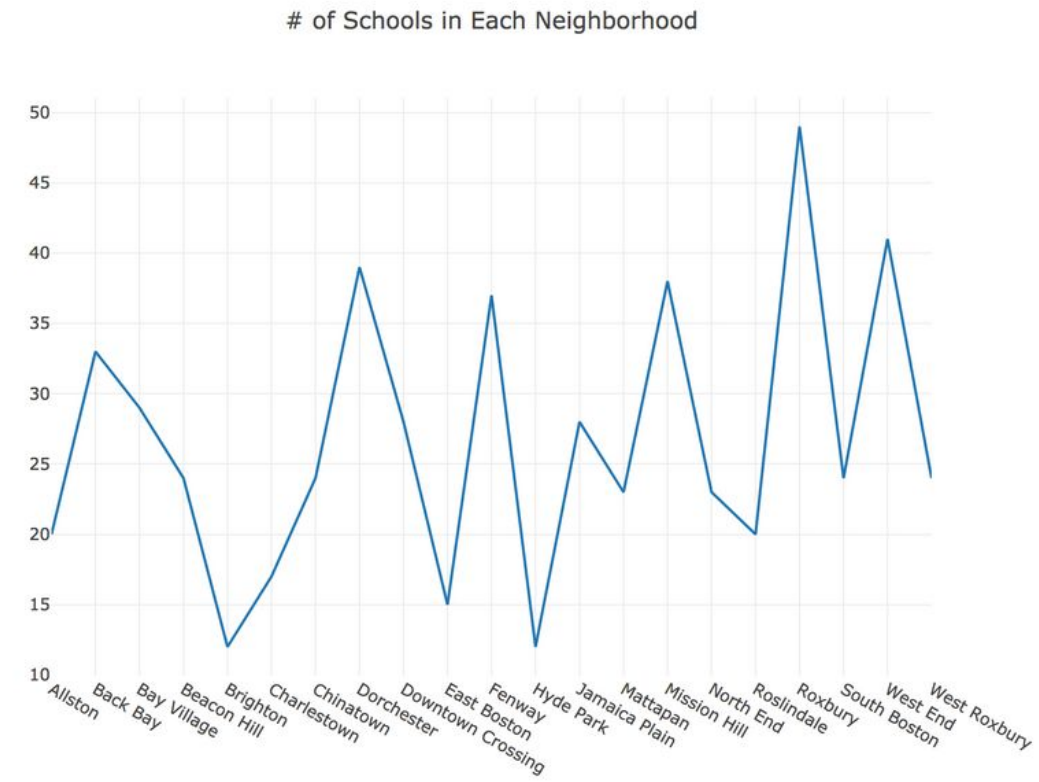
These results are based on the 2016 property value data because it consists of a larger dataset than the 2017 version. The development scores are based on a scale of 0-100 and rounded to the nearest hundredth decimal (except for the development score of 02128). It is possible that the development score generated for the area of zip code 02128 is significantly lower than the rest of the results is because the area includes the airport. We didn't take the airport or any public transportation into account when determining the arbitrary values of government property.

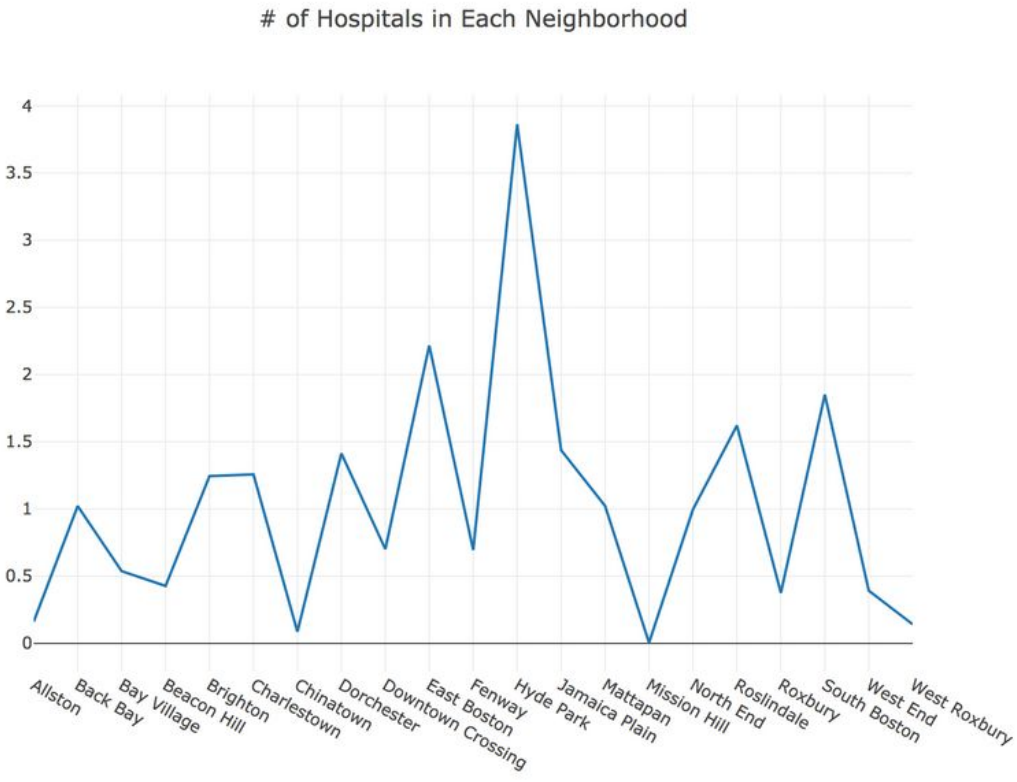
The following graph takes in the developmental scores for each neighborhood, and normalizes them to a scale of 0-15, then applying them as a density value to a map of Boston.

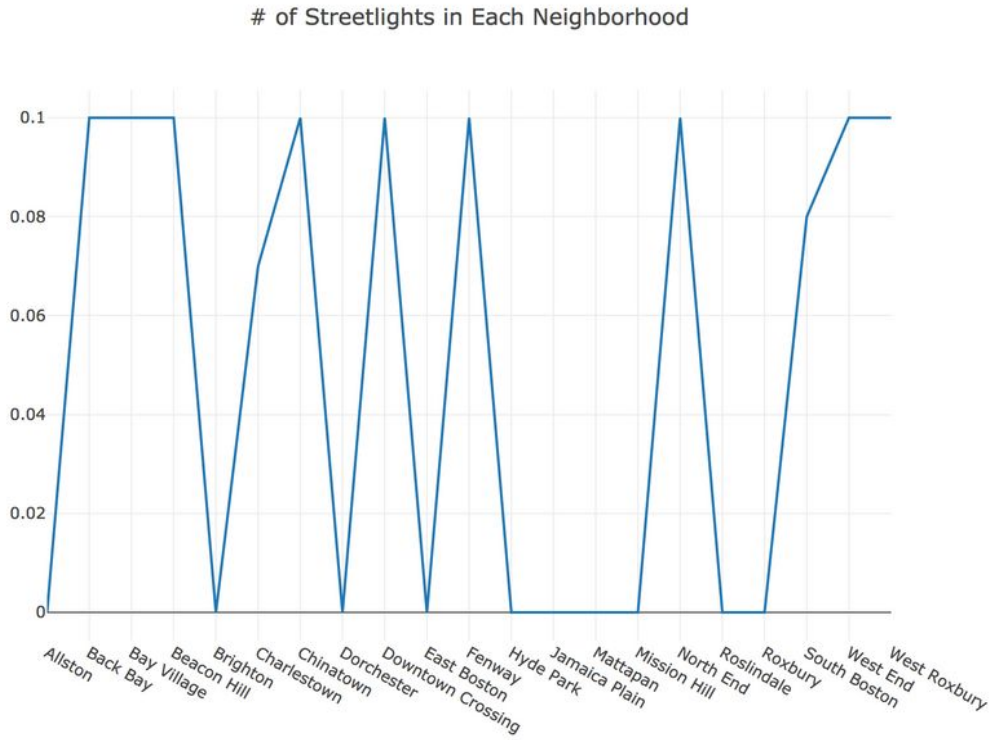


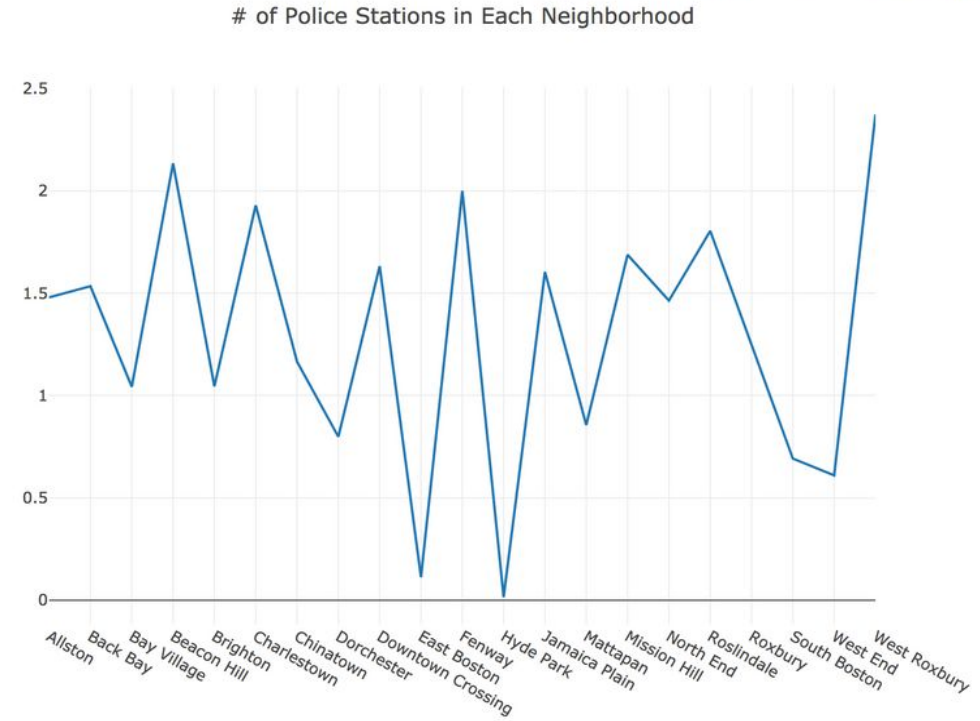
The following heatmap represents the number of hospitals, police stations, and schools in the Greater Boston Area.

These are graphs produced after scaling was applied, so values may not be whole integers / represent the actual number, but instead represents the value of each government structure for each neighborhood.






# of Streetlights in Each Neighborhood
0.1
0.08
0.06
0.04
0.02
0
Allston
Back Bay
Bay Village
Beacon Hill
Brighton
Charlestown
Chinatown
Dorchester
Downtown Crossing
East Boston
Fenway
Hyde Park
Jamaica Plain
Mattapan
Mission Hill
North End
Roslindale
Roxbury
South Boston
West End
West Roxbury


# of Police Stations in Each Neighborhood
2.5
2
1.5
1
0.5
0
Allston
Back Bay
Bay Village
Beacon Hill
Brighton
Charlestown
Chinatown
Dorchester
Downtown Crossing
East Boston
Fenway
Hyde Park
Jamaica Plain
Mattapan
Mission Hill
North End
Roslindale
Roxbury
South Boston
West End
West Roxbury

Conclusions

From the datasets provided, and through the transformations applied, we recognize that the areas of 02113, 02115, and 02111 are zip-code areas that have a disproportionately high mean property value, as compared with the level of government developmental structures. Some issues that may produce skewed results are the fact that storefronts and development outside of the included datasets (e.g. restaurants, paved roads, public structures) are not accounted for. This produces a result in which an area like 02113, which covers the North End, would have an outlier value of 100 due to its relatively higher property values, the algorithm not accounting for all of the stores, street development, etc, and the lack of government structures in the area.

Issues faced

The databases for several points of data went down during this process, so the new database for the property values lacked necessary details for the homes. Therefore, representing these values in our web app became impossible. We are still able to draw conclusions for zipcodes, but specific neighborhood conclusions were not possible.

Possibilities for the future

In the future we could hope to create a web app where users are able to interact with and scale the weights we applied to each point of development (e.g. hospitals, schools, streetlights, police stations), and see how the conclusions adjust accordingly.