Measuring Child-Friendliness in Boston Neighborhoods

By Ji Eun Yang and Robin (Hin Lok) Liu

CS 591 Data Mechanics Fall 2016

Boston University

Introduction:

In the past decade, city-living has become a very popular and economically favorable trend in the US. The City of Boston has grown and attracted a diverse population from all over the world. From 2000 to 2010, Boston's overall population increased by 4.8% to 617,594, and the total number of households increased from 251,212 to 252,699. Despite Boston's general trend of growth, there is a vital demographic that it has been losing: families with children. The number of household families with children dramatically fell 11%, now standing at 16.8%, falling behind cities like New York, Vancouver, and Tokyo. According to the Boston Globe, "there is something a little bit bloodless about a city dominated by empty-nesters and young hipsters, a kind of shell without a core. Compared to young professionals just out of college and retiring baby boomers with pricey pied-a-terres, families with children tend to be more stable, community-focused, and civically engaged." (bostonglobe.com). Most families with children avoid cities like Boston due to the high cost of living and "uneven" public schools. Although Boston is great at catering to young professionals, it lacks the needs that families with children need. Therefore, to address this problem, we created a method to calculate child friendliness in Boston neighborhoods. These calculations can highlight areas that do better in catering to families children, and especially areas that can do much better.

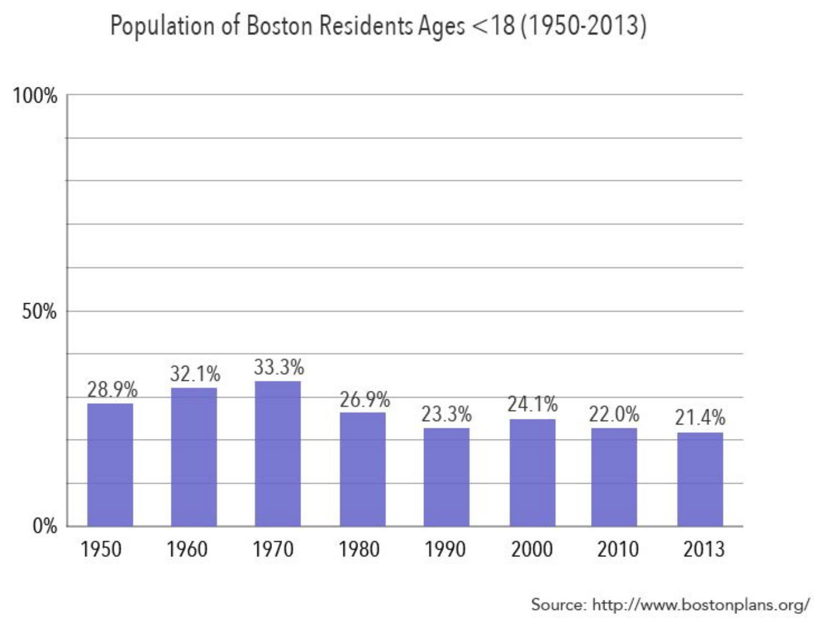

Figure 1: A graph that represents the gradual decline of the children population in Boston

Method:

We will accumulate various metrics on certain characteristics of all the neighborhoods in Boston to calculate a score that measures child-friendliness. We believe that these four characteristics are strong determinants of child-friendliness: education, safety, health, and housing costs. To measure scores for each of these characteristics, we collected data sets from the City of Boston site in the following areas:

- Schools: We collected the name (SCH_NAME) and coordinates (Location) of each school from the Boston Public Schools (School Year 2012-2013) dataset.
- Crime: We collected the offense type (OFFENSE CODE GROUP) and the coordinates (Location) of each entry in the Crime Incident Reports (August 2015 - To Date).
- Hospitals: We collected the name (NAME), neighborhood (NEIGH), and coordinates (Location) of every hospital in the Hospital Locations dataset.

- Property: We collected the zipcode (ZIPCODE), property type (LU or Land Use), total assessed value for the property (AV_TOTAL) and coordinates (LONGITUDE, LATITUDE) of all the properties from the Property Assessment 2016 dataset.

To measure an education score, for each neighborhood we counted the number of schools using a threshold radius of 3 km. We did the same counting method for various crimes within the threshold radius for each neighborhood. As for hospitals, we used an optimization algorithm to find the minimum distance hospital. For property values, we first filtered out properties that did not fall under the category of residential use. Then, for each neighborhood, we added up the total property values, then divided by the number of properties per neighborhood to find the average residential property value per neighborhood. After gathering all of the individual characteristic metrics, we bring them together in our scoring algorithm discussed

Algorithm:

After calculating the individual category scores for each neighborhood, we used the following algorithm to calculate a relative score:

$$Score \mathrel{+}= hospital_{min\ dist} * 0.25$$

$$Score \mathrel{+}= school_{count} * 0.25$$

$$Score \mathrel{-}= crime_{count} * 0.25$$

$$Score \mathrel{-}= property_{avg\ val} * 0.25 .$$

The calculated scores ranged from -1000 to -8000, so we scaled them to be a more appropriate value, while remaining proportional among all of the neighborhood scores. We used the following:

$$Score_{new} = ((-1)/Score) * 1000000.$$

For our interactive map visualization, algorithm, we calculated school_count , hospital_count , crime_count , property_count by finding the number of each category (school/hospital/crime/property) within a 3 km distance of the marker placed on the map. We use the following equation :

Score = school_count * 0.25 + hospital_count * 0.25 - property_count * 0.25 - crime_count * 0.25

Then we scale the score, due to there being a much larger amount of crimes and properties compared to schools and hospitals :

Score = -1 / score * 10000

To display the nearest hospital , school and their respective addresses, we use Google Maps' Distance Matrix Service. We find the closest hospital and school to the marker placed on the interactive map and make one API call for each category. The result is the address of the hospital/school. By calculating the closest hospital/school and passing in only one query, we circumvent Google Maps' API 25 query limit.

Visuals:

For the visuals of this project, we decided that we first needed a map of Boston with all of the neighborhoods and its corresponding scores to depict which areas of Boston have higher scores than others. In light of this, we developed a choropleth graph using Leaflet (Figure 2), where the darker colors are correlated with higher scores, and the lighter colors are correlated with lower scores. We made the map interactive so that when users hover over a neighborhood the scores are displayed on the top right corner. In order to recreate the general areas for each neighborhood,

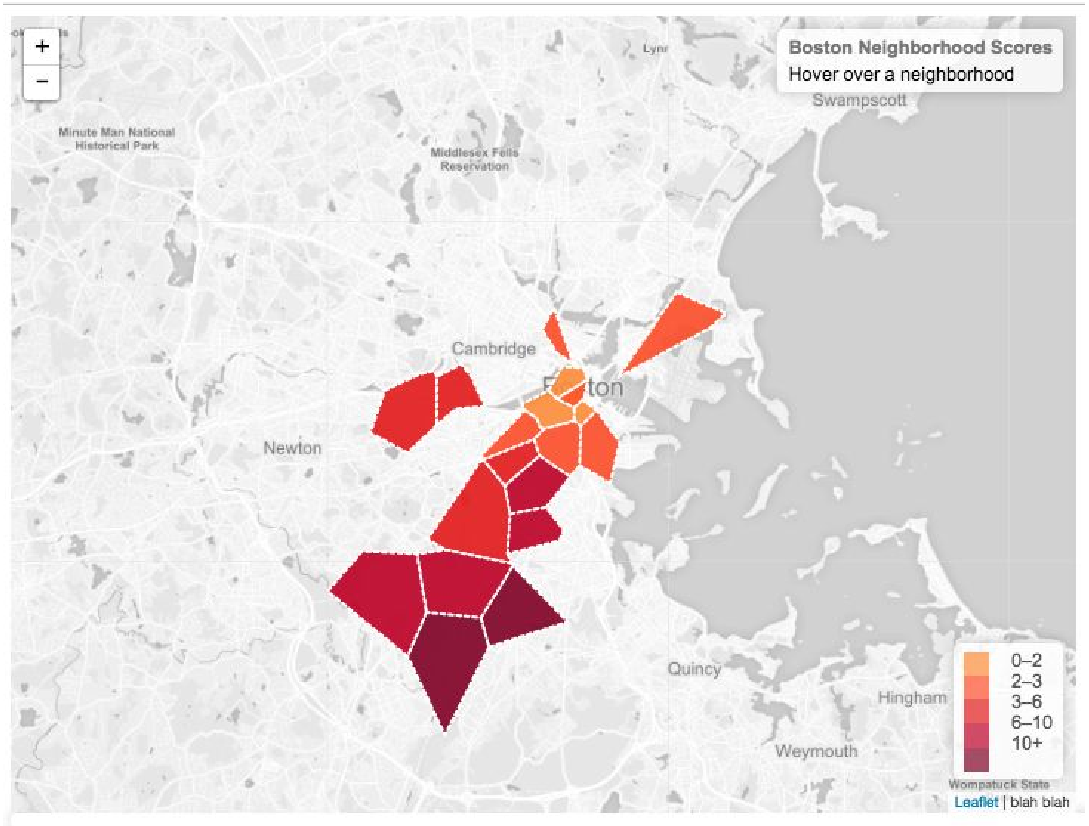

Figure 2. Choropleth graph depicting neighborhoods and their scores.

We also developed an interactive map using the Google Maps API so that users could get more specific data regarding a specific location in Boston (Figure 3). The user is able to click on any point on the map, and the results will display the closest hospital and school, with corresponding distances as well as the score of that specific location. This way, users are not restricted by a general neighborhood as a whole, but can still get a score based on a distinct location in the city.

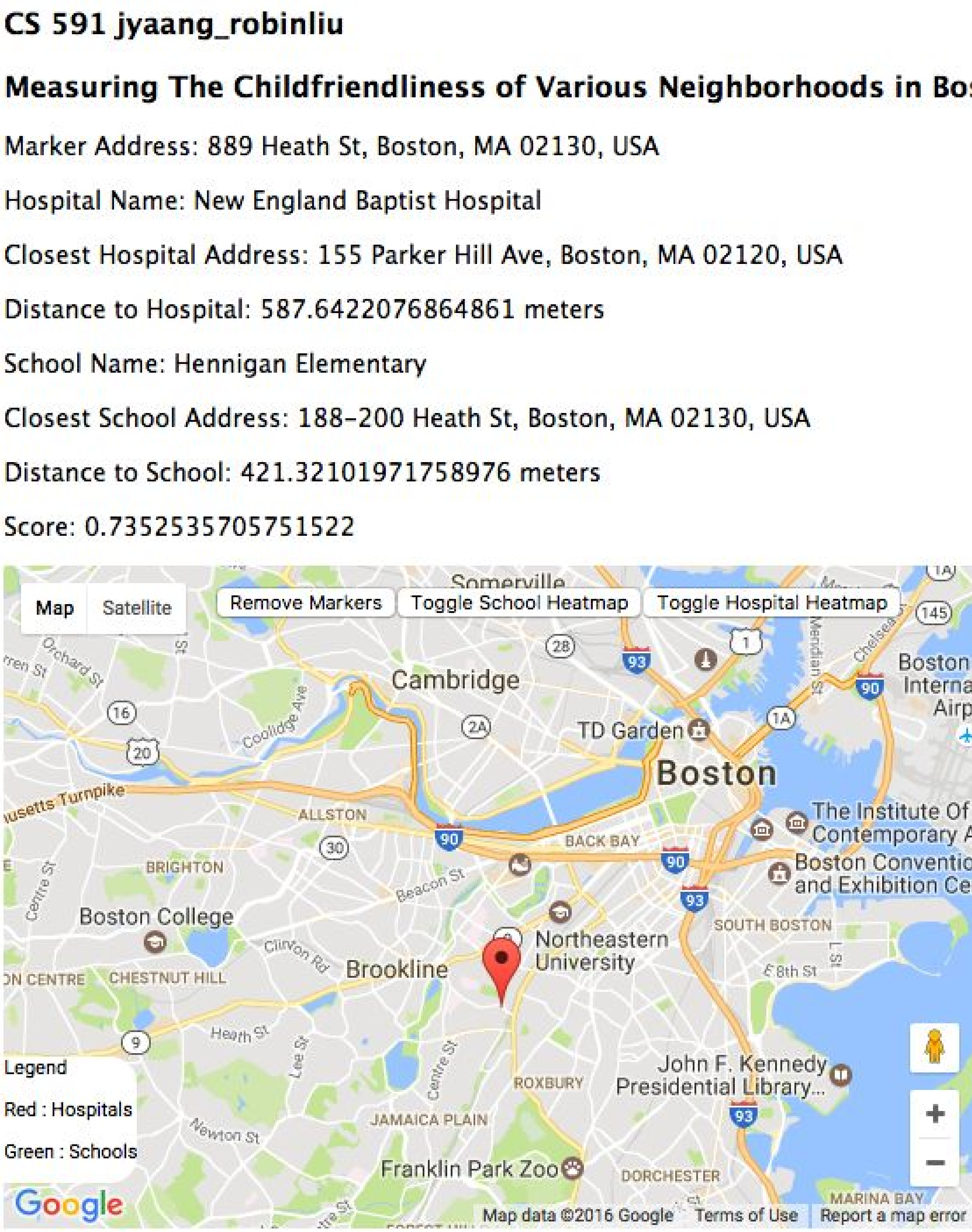

Figure 3. Interactive Map to calculate score of any point in the city.

Lastly, we developed various graphs to display neighborhoods scores in specific categories, like number of hospitals, number of schools, number of crimes, and average housing costs (Figure 4). This way, if there is a specific area that users are more concerned about, such as crimes, the user can see how a given neighborhood compares to other neighborhoods in terms of specifically crimes.

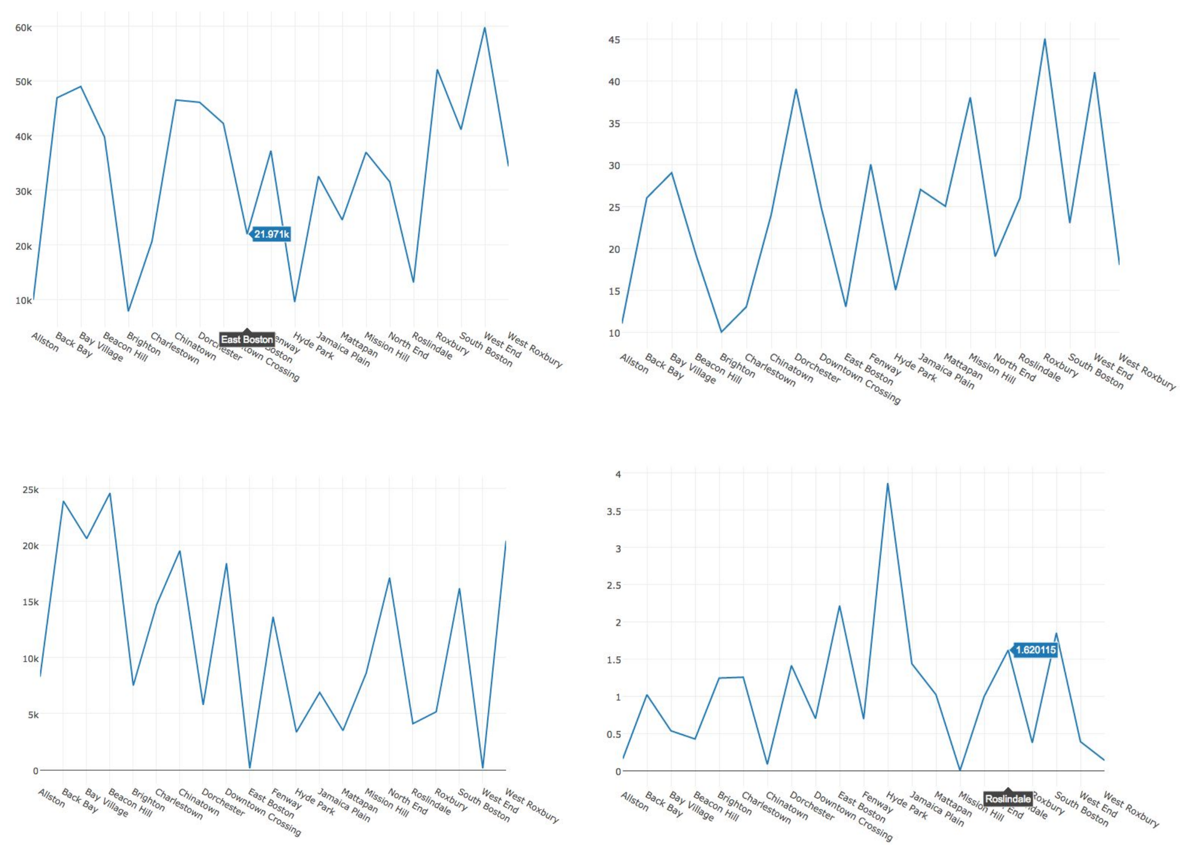

Figure 4. Top left: Number of Crimes, Top right: Number of schools, Bottom left: Average residential property value, Bottom right: Number of hospitals.

Conclusion:

In conclusion, we saw that average property values had the biggest impact on our scores. When we originally scored the neighborhoods without the property values, areas like Back Bay and Downtown Crossing had higher scores due to the concentration of schools and hospitals in those areas. However, when we included property values into the mix, the scores flipped, and areas like Hyde Park and Mattapan with lower property values scored higher.

We tried to add recreational centers to the mix; however, when we tried to implement the locations of the recreational centers, each location was so far away, it had little to no unique effect per score calculation. Therefore, this is a huge indication that recreational centers in the City of Boston are sparse and can perhaps be something that can be emphasized when considering child-friendliness in Boston.

Future Work:

In the future, this project could be extended to include even more factors other than schools, hospitals, crimes, and property values. There is an extended number of various factors that can contribute to the value of child-friendliness that could make this algorithm more definitive. Example datasets that we also considered include park and recreational centers, local food inspections, and sex offender locations.

An area that we could further improve is our use of the dataset on schools. When calculating the scores for education, we limited the value to the number of schools in each neighborhood. However, instead of focusing primarily on the quantity, we could also have incorporated the quality of education based on district test scores.

A major problem we faced was also the magnitude of data we were dealing with. Especially the datasets on property values and crimes where there were hundreds of thousands of entries to parse through, we had some difficulty in terms of performance and efficiency. Hence this was an area we could definitely improve on in the future.

Resources:

Census/demographic:

http://www.census.gov/quickfacts/table/PST045215/2507000

https://suburbanstats.org/population/massachusetts/how-many-people-live-in-boston

http://www.bphc.org/healthdata/health-of-boston-report/Documents/HOB-2014-2015/1_Demographics_HOB%202014-2015.pdf

http://www.bphc.org/healthdata/Documents/HBC_Final_103113_ForWeb.pdf

Zip Code geometry long/lat:

https://fusiontables.google.com/DataSource?docid=1Lae-86jeUDLmA6-8APDDqazlTOy1GsTXh28DAkw#rows:id=1