Adam Elass

Streetlights and Crime: Characterizing T Stop Safety at Night

Objective:

The objective of this project was to analyze the safety of T stops at night for a few given lines and to see if a correlation exists between the number of streetlights and the number of crimes around a given station.

Datasets:

The datasets used for this project were Crime Incident Reports from the City of Boston (2012-2015), Streetlight Locations for the City of Boston, and MBTA T Stops data.

New Datasets Created:

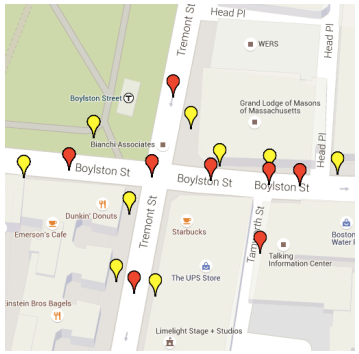
In order to properly do my analysis, I needed to create a few new datasets from the original three mentioned above. One such dataset, for example, paired each crime with all the streetlights within a given distance from it.

Assumptions Made:

Because the streetlight dataset only contained information about the locations of the streetlights, I assumed that streetlights were turned on at sunset and turned off at sunrise.

Approach:

I initially filtered out crimes that didn't happen between sunset and 1AM, since I assumed the streetlights wouldn't be on before sunset and most T stations seem to close around 1AM. Next, I selected all crimes that happened within a given distance from a station, and then selected all streetlights that were within a given distance from a certain crime.



Crimes in red, streetlights in yellow

Regarding correlation, I went through the stations and created tuples in the form (num_streetlights, num_crimes) and then calculated the Pearson correlation coefficient to see if there was a correlation between the number of streetlights around a station and the number of crimes around the station.

As for my scoring metric used for characterizing the safety of the stops, I extracted the number of crimes at each hour between 9PM and 1AM for a given station and then calculated a weighted sum of the crimes, assigning higher weights to the earlier hours. The higher the sum, the less "safe" a station, as defined by this metric.

Results:

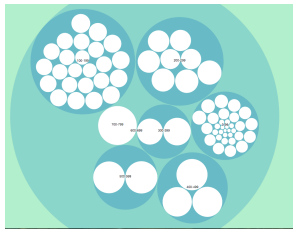
The results for the correlation were as follows:

$$\rho = 0.406289747918$$

$$p - value = 0.000361666922126$$

This tells us that there exists a positive correlation between the number of streetlights and the number of crimes. Although this initially may seem counterintuitive, it makes sense if we realize there are many other indicators of crime rates, and these other indicators may also contribute to the number of streetlights as well (e.g., population size).

As for the scoring part, the weighted sums were in the range 0-799, with the lowest score being 5 and corresponding to Boston Univ East and the highest score being 765 and corresponding to North Station. There were fewer stations with higher scores, and many stations with relatively low scores (in the range 0-199).



Scoring results represented in a circle packing visualization.