Green Spaces and Crime

From a climate and public health perspective, the argument for green spaces in cities is undeniable. However, a long-term justification is often a hard sell for a city’s short-term budget. Especially those which are difficult to quantify. It is therefore becoming increasingly important for climate and public health advocates to find justifications for green spaces, occurring in real-time. The most important real-time benefits will be those that offset the greatest challenges to cities, socially and economically. Much of a city’s budget and reputation depends on its crime. If it could be shown that green spaces correspond to any kind of reduction in crime, or perhaps even predictions for crime occurrences, the green space implementation rate could be significantly increased. And Al Gore would be jubilant.

To illustrate this relationship, crime data can be used in conjunction with the coordinates of city parks. For a city like Baltimore whose crime is under national scrutiny, crime statistics are fairly easy to come by. Baltimore City has an open access website, where crime data is available. Each crime is recorded with a date, time stamp, zone, neighborhood, coordinates, weapon, and incidents in a neat csv file. Ideally, crime would be analyzed for all green spaces within the city. However, it is difficult to find a database of all communal “green” coordinates. Instead, crime will be cross-referenced with a database of coordinates for official city parks. There are 18 parks in Baltimore, according to Baltimore City’s Recreation and Parks website. A data frame was created using coordinates chosen from Google Maps.

The crime data for Baltimore City exists in one data frame for 2014 and 2015. As 2015 is not yet over and contains skews from The Uprising, 2014 data seemed appropriate for a full analysis. A new data frame was created, containing 2014 crimes with additional columns that separated out latitude and longitude. In order to do this, all null values had to be dropped. Using the two data frames that contain latitude and longitude values, a function could be run which finds the minimum distance between the crime and any of the park coordinates for each crime. The function was created using a formula that converts the distance between coordinates to Euclidean distance. For each crime, the function iterates over all park coordinates, returning the minimum distance between the crime and a set of park coordinates.

The new data frame to be analyzed contains rows filled with details of crimes and their distance to the closest park. With this information, a few types of analysis can be done. The obvious first step is plotting crime versus distance from a park. This data can be filtered further, for time of day, type of crime, month, or season. Furthermore, an additional column for park can be added to the data frame, allowing for further classification. The data will be regressed and strength of relationships can be revealed. In doing so, we can determine what kinds of similar patterns are seen across all parks.

Depending on the patterns evident, probabilities for crime at certain proximities from parks can be created. This can be analyzed for a variety of filters, due to the detail of the crime data. Ultimately, it would be ideal to create some kind of predictive system for crime as a function of proximity to parks.