Crime and Streetlamps

Analyzing Crime and Streetlamps in the City of Boston

As members of Boston University's Data Mechanics committee, we gathered and studied open source data provided by the City of Boston in order to create usable data related to crime and streetlamps in Boston. We expect the data to be relevant to Data Mechanics' future goals of creating innovative software solutions in order to make a more efficient and organized Boston.

BOSTON UNIVERSITY
CASCS591 L1 SPRING
2016 DATA MECHANICS

Nicholas Louie
(nlouie@bu.edu)
Thomas Hsu
(thsu8@bu.edu)
Department of
Computer Science
4/30/2016

TABLE OF CONTENTS

Table of Figures 1
1. Introduction 1
1.1 Data Mechanics 1
2. Analyzing Crime & Streetlamps 1
2.1 Introduction 1
2.2 Datasets 2
2.4 Implementation 3
2.5 Possibilities 4
2.6 Results 4
2.7 Limitations 5
4. Future Work 5

TABLE OF FIGURES

Figure 1 - All StreetLights in Boston 2
Figure 2 - Visualization of Crimes 3
Figure 3 - Number of Crimes vs. Distance to Nearest Streetlamp 4

1. INTRODUCTION

1.1 DATA MECHANICS

Boston, Massachusetts is one of the oldest metropolitan cities in the United States of America, and is an epicenter for higher education. We, as the Boston University Data Mechanics Committee want to study how to adapt mathematics and technology into useful software[1]. Today, technology is growing exponentially[2], and the City of Boston must learn to adapt and exploit new areas of technology. Data Mechanics wishes to be part of this technological shift, and create tools and software that will be part of Boston's physical and technological transition to the 21st century.

The amount of data available today is growing at a rapid rate, and data pertaining to Boston are not excluded from this comprehensibility infinite site of data. We obtain datasets from various public sources such as, City of Boston[3], Massachusetts Department of Transportation[4], Twitter[5], and Open Street Maps[6].

By taking advantage of the large amount of publically available data, Data Mechanics wants to model and quantify the City of Boston, in order to both create a more efficient Boston, and for the sake of academia.

2. ANALYZING CRIME & STREETLAMPS

2.1 INTRODUCTION

Our group was interested in study both crimes and streetlamps related to the city of Boston. As a part of our underlying goals, we asked a number of questions:

- What can we learn about crime in Boston given publically available data?
- Do crimes happen further away from streetlights?
- What other factors affect the crime incidence rate in Boston?
- Can we visualize the crimes and streetlamps in Boston?

[1] http://cs-people.bu.edu/lapets/591/s.php#!
[2] https://www-01.ibm.com/software/data/bigdata/what-is-big-data.html
[3] https://data.cityofboston.gov/
[4] https://www.massdot.state.ma.us/DevelopersData.aspx
[5] https://dev.twitter.com/
[6] https://www.openstreetmap.org

We wanted to know how streetlights relate to crimes in Boston and whether or not streetlamp placement has an effect on crime or if there are high-crime areas that may benefit from better lighting.

2.2 DATASETS

For our project, we used the following datasets from the City of Boston.

- Streetlight Locations[7]
- Crime Incident Reports[8]
- Employee Earnings Report (2012)[9]
- Employee Earnings Report (2013)[10]
- Employee Earnings Report (2014)[11]

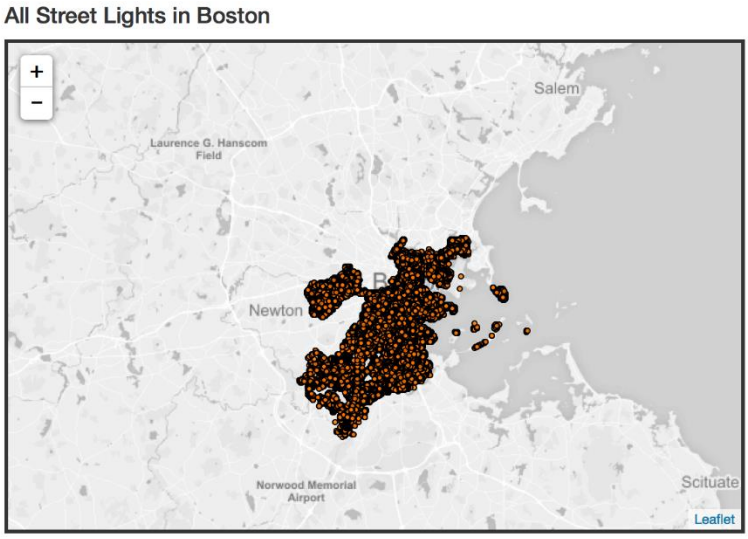

Figure 1 - All Streetlamps in Boston

[7] https://data.cityofboston.gov/Facilities/Streetlight-Locations/7hu5-gg2y
[8] https://data.cityofboston.gov/Public-Safety/Crime-Incident-Reports/7cdf-6fgx
[9] https://data.cityofboston.gov/Finance/Employee-Earnings-Report-2012/effb-uspk
[10] https://data.cityofboston.gov/Finance/Employee-Earnings-Report-2013/54s2-yxpg
[11] https://data.cityofboston.gov/Finance/Employee-Earnings-Report-2014/4swk-wcg8

Figure 2 - Visualization of Streetlamps by Street

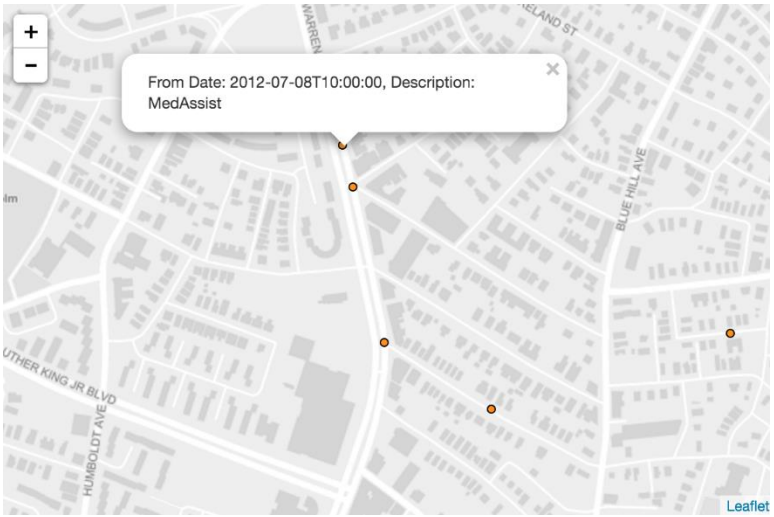

Figure 3 - Visualization of Crimes

2.4 IMPLEMENTATION

In order to implement our data manipulation, we computed the distance of each crime (at night) to the closest streetlamp using data provided from Boston's public datasets.
The data files were retrieved manually through a web browser which was considerably simpler than going through the Application Programming Interface (API)[12]. The files were transformed from the retrieved Javascript Object Notation (JSON)[13] data to text file for usage in Hadoop

[12] https://en.wikipedia.org/wiki/Application_programming_interface
[13] http://www.json.org/

cloud computing[14]. Originally the plan was to use Apache Hive[15] to perform SQL-like queries[16], but in creating an intermediate dataset, Hive stopped executing. So that step was relegated to a local machine, performed serially. The new intermediate set was then put back on the cluster. Data analysis was done on Hive.

2.5 POSSIBILITIES

We think solving this gives insight on how lighting plays a role in the incidences of crime and whether or not certain areas should include more lighting. By gathering this data, it could be used in creating a "danger" metric of certain areas in Boston where lighting may be insufficient. This data could also be used with Boston's LED Retrofit data as well[17]. By measuring LED Retrofit data, we could determine what areas most need retrofits and thus enjoy energy and environmental benefits.

It is possible that depending on the incidences of crime and lighting within a given area, a metric for 'danger' could be created to be used applications such as navigation in Microsoft's controversial "Avoid ghetto" feature[18].

2.6 RESULTS

By evaluating our distance to street light and number of crime data, most crimes are within several meters to a streetlight (note the use of a log scale). This may be due to inaccurate locations of crimes, but by viewing the streetlight visualization, it's interesting to see the precision of the coordinates of the streetlamps as they are viewable by street.

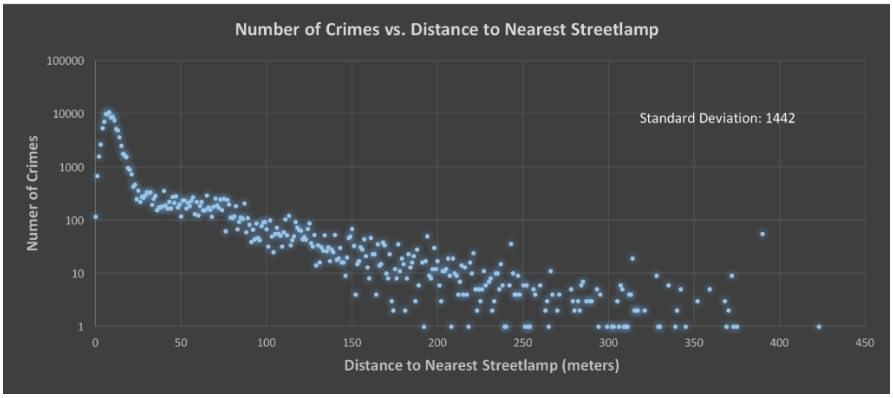

Figure 4 - Number of Crimes vs. Distance to Nearest Streetlamp

[14] https://en.wikipedia.org/wiki/Apache_Hadoop
[15] https://en.wikipedia.org/wiki/Apache_Hive
[16] https://en.wikipedia.org/wiki/SQL
[17] https://data.cityofboston.gov/Facilities/Street-Light-Retrofits-2014-/nk7w-bpnm
[18] http://www.cnet.com/news/the-joy-of-microsofts-avoid-ghetto-gps-patent/

2.7 LIMITATIONS

We encountered a number of limitations while manipulating the data and the analysis. We made several assumptions for simplicity as the computation would be extremely complex if we attempted to 'geofence' the entirety of indoor Boston, so we assumed that all crime occurs outdoors. We tried to filter only night crimes from 18:00 to 6:00. There were also a few nonsensical data points in the crimes set (Crimes occurring more than 8000 kilometers away from the nearest streetlamp), because not all reports in crime incidence were properly GPS tagged.

4. FUTURE WORK

Future work may include using other source of data such as salary of various departments, most notably, the Boston Police Department. It would be interesting to determine if the money going to police officer salary has a significant impact on crime. Other work could factor in the distance to streetlamps as a metric as part of a Data Mechanics danger rating application or function.