Boston Biking

A data mechanics project focused on improving the lives of bikers in Boston.

Created as a part of CS504: Data Mechanics.

Members:

Dennis Wu (wud@bu.edu, U10469253)
Nuan Huang (nhuang54@bu.edu, U92881129)
Tony Kim (tkixi@bu.edu, U65555103)

Introduction

Background:

Dennis, Nuan, and Tony are all Boston residents and avid bikers. Biking is a healthy and environmentally friendly way to get to different places. So, in order to promote biking and make Boston a better biking environment for everyone, we decided to focus on finding issues with current-day biking (like safety or availability of bikes) and finding areas for the city to improve on.

Problem/Topic:

With our project, we focused on dangerous areas for bikers and bike availability.

In order to locate precarious areas for bikers, we tried to search for factors that could contribute to bike accidents or potential correlations between bike accidents and the city of Boston. To try and solve this problem, we utilized pandas in order to try and draw conclusions about our bike and accident data. We looked at aspects of Boston biking like the number of streetlights and accident-prone regions.

To maximize bike availability for the Boston population, we looked to find the best places along Boston's biking paths to place bike rental stations. Using our datasets in addition to the Z3 Theorem Prover, we created a constraint satisfaction problem that, when solved, provides optimal solutions to our bike station location problem.

Data Sets:

Boston Bike Lanes

Description:
This data set contains information on all of the bike lanes in Boston, such as the jurisdiction, street, and length of the bike lane.

Use:

- We used this data set to check for biking accidents along these lanes.
- Also, this data set was used to choose locations along these bike paths for possible bike rental stations.

Origin:
BostonGIS

Boston Collisions

Description:
This data set, provided by Vision Zero Boston, contains information on transportation accidents in Boston, like the time, date, location, and type of incident.

Use:

- Using this data set, we examined only accidents that contained bikes.
- This helped us determine safe/dangerous areas for bikers.
- It was also used to compare with other statistics to find correlations for biking accidents.

Origin:
Vision Zero Boston

Boston Traffic Signal Locations

Description:
Analyze Boston provides this data set, which contains information on all of Boston's traffic signals.

Use:

- This was one of the data sets that we used to find a correlation between accidents and Boston's streets.

Origin:
Analyze Boston

Streetlights Collisions (Transformation)

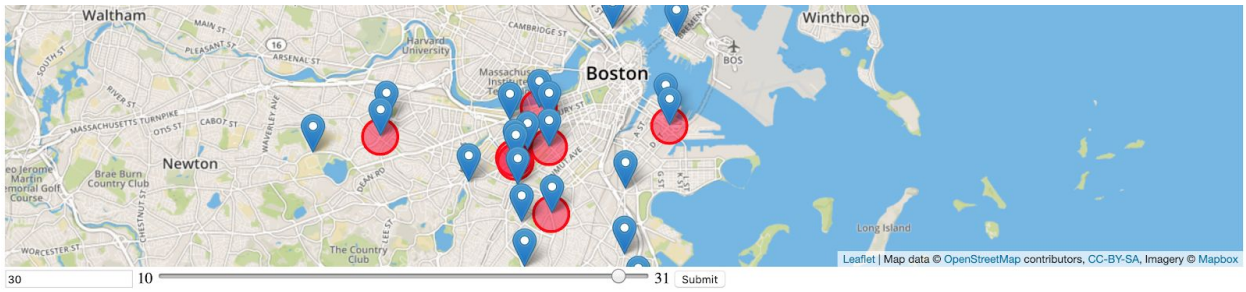
Description:

This transformation used our streetlights and collisions data set to create a new data set. We narrowed down the collisions to focus only on bike collisions. Then, we aggregate every bike collision at a particular intersection, to help analyze if the intersection is prone to danger or not. We also use the OpenCage API to obtain the latitude and longitudes of each intersection.

Algorithms / Analysis / Visualizations

Finding Minimal Stations (Z3 Theorem Solver) - findMinimalStations.py

[Visualization available on Project 3 website].



30 biking paths, represented by blue markers on the map.

Red circles under the markers represent streets that should have a hubway station on them in order to minimize amount of stations for an distance that was deemed walkable by our team.

Goal

With this study, we focused to improve the availability of bikes around Boston by finding the best locations for bike stations along the biking paths.

Tools

Boston Bike Lanes data set, Z3 Solver, OpenCage API, leafletJS.

Method:

Using the Z3 solver, we add constraints that only require 1 station for each bike lane and constraints that allow 1 station to cover two streets if they are close enough. The Z3 solver then provides us with a solution the constraints, which we tightne until we receive a minimal solution (the least amount of total stations).

Limitations / Challenges

We had to readjust the Boston Bikes using the OpenCage API to include lat/long values.

Interactibility
We allow the user to adjust the amount of biking paths the algorithm takes in. There is, however, an upper limit due to us being restricted by the OpenCage API which limits the amount of daily requests for LAT/LONG values for each street.

Results
Vary depending on amount of biking paths we take into account.

Pairwise Correlation (pandas) - pairwiseCorrelation.py

Goal
With this analysis of our data, we sought to find out whether or not there was a correlation between streetlights on a street and collisions on that street.

Tools
Streetlights Collisions transformations, pandas (Python Data Analysis Library), OpenCage API.

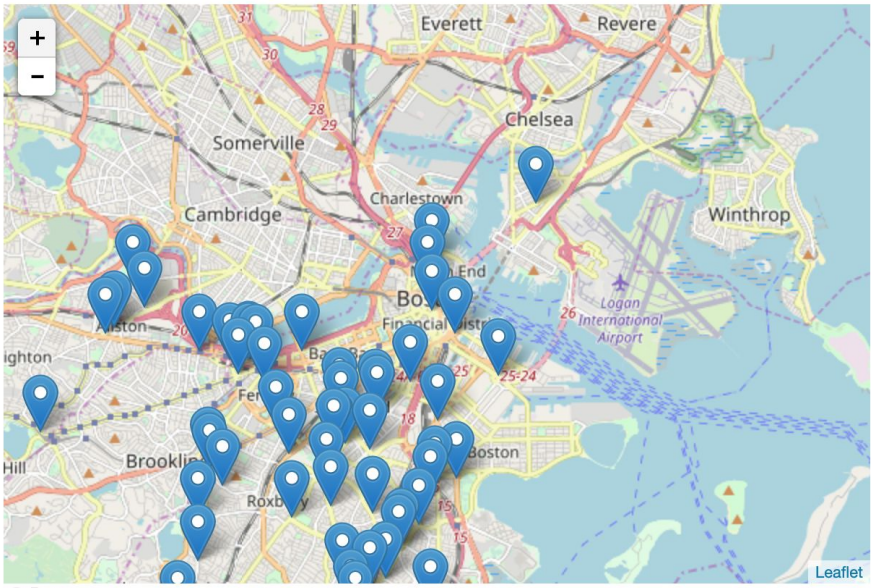
Method
In pairwiseCorrelation.py, we utilized a dataset from a new transformation (streetlights_collsions.py). We took the number of bike accidents that occur on a particular street and accumulated the present number of streetlights on that particular street during the bike accident. After aggregating the two values with the street as the key, we were able to discover the total number of streetlights that were present over all the bike accidents on any particular street that had any bike collisions.

Limitations / Challenges
We don't take into account length of street or distance a streetlight covers (information not in our data set).

Results
From the value, we saw a positive pairwise correlation between the two values (correlation coefficient = +0.9575). However, due to our limitations, we cannot reasonably conclude that streetlights contribute extensively to bike collisions. We are also limited by the OpenCage API that helps us compute the correlation as it is limited to only 2500 requests a day.

Collision Visualization

[Visualization available on Project 3 website].



- January
- February
- March
- April
- May
- June
- July
- August
- September
- October
- November
- December

Goal

With this visualization, we hope to answer questions on where most of the collisions are happening and in what areas we should focus on preventing collisions. In addition, we can see if the date has an effect on accidents.

Tools

Boston Collisions data set, OpenCage API.

Interactibility

This filters all non-bike collisions and based on the months the user selects, it'll add every bike collision that has occured on the specified month since 2015.

Future Work

Our team plans to continue our research into Boston biking by expanding our research into causes of collisions using the clusters we locate in our visualization. There were many possibilities that we haven't touched upon yet, like precipitation (rain, snow, hail) and temperature affecting the safety of bikers. It may also help to receive input from other bikers themselves to have a more holistic biker-centric solution.

In additions, we plan to search for more correlations between unsafe biking areas and Boston's roads. We had initially considered an idea to find optimal placement of streetlights around Boston's biking paths as well to keep Boston's bikers safe at night.