

Bike Lane Data Report

Project: Bike Lane Blocks (DS)

Team Member: Mingdao Che

Yucheng Zhu

Zhengyang Tang

Introduction:

This project is designed for analyzing the bike blockages over the years. According to the dataset given by Boston City and other supporters, we mainly focus on: the reasons that lead to bike blocks, block changes over years, the correlation between bike lane block and crashes and bike rider counts with bike lane blocks. Actually, we have done some of the problems which will be shown in the following sections, and there still remains some challenges for us to overcome and we may need some more data support to get a better solution because of the blockage data in Boston 311 dataset is limited.

311 Request Service data(Problem 1-4):

In the 311 dataset from 2011 to 2019, each row is a reported case corresponding to public utilities.

To extract these useful data from all data related with the bike lanes, we keep searching “bike lane”, and find these blockage report mainly belong to “Parking Enforcement”, “Request for Snow Plowing”, “Request for Street cleaning” etc. We also considered the “repaint bike lanes” as a kind of blockage. There will be pieces of data related to the bike block on next page.

Once we determined which rows are reports for bike blockage, we can start finding relations between bike lane blockage and bike lane crashes, the number of bike riders, seasons, nearby businesses and neighborhoods. Useful columns for the analysis are **open_dt**(the time when bike lane blockages occur), **ontime**(whether bike lane blockages are handled on time), **case_title**(case details), **closure_reason**(case details), **type**(case details), **neighborhood**, **latitude**, **longitude**.

With **open_dt** associated with **dispatch_ts** and **mode_type** in the bike crash dataset, we can determine the relationships between bike lane blockage and bike crashes; with **latitude** and **longitude** associated with Google Map API, we can determine the relationship between bike

lane blockage and nearby businesses; with **open_dt**, we can determine the relationship between bike lane blockage and seasons; with **open_dt** associated with the bike count dataset, we can determine the relationships between bike lane blockage and bike riders; finally, with **open_dt** and **neighborhood**, we can determine the relationship between bike lane blockage and neighborhoods.

Problem 1: What businesses have the highest correlation with complaints about bike lane blockages on 311?

By classifying Boston and Cambridge's 311 data, we found the highest correlation businesses with bike lane blockage is **illegal parking**. And we may misunderstand the “business” here, and we'll continue working on the correlation.

Problem 2: Which neighborhoods in Boston have the highest number of 311 calls/tickets related to bike lane/parking blockages?

Back Bay has the highest number of 311 calls related to bike lane blockages.

To get the result, we first extracted rows for bike blockage from the 311 calls dataset, and put the corresponding data in the “neighborhood” column in a dictionary, where keys are the neighborhoods' names and values are how many times this neighborhood appeared in bike blockage data. At last, we got the dictionary like this:

{'Boston': 72, 'Back Bay': 110, 'Jamaica Plain': 23, 'Roslindale': 5, 'Allston / Brighton': 43, 'Dorchester': 14, 'Hyde Park': 5, 'Fenway / Kenmore / Audubon Circle / Longwood': 63, 'South Boston / South Boston Waterfront': 33, 'Charlestown': 6, 'West Roxbury': 1, 'Roxbury': 30, 'South End': 35, 'Greater Mattapan': 5, 'Downtown / Financial District': 42, 'East Boston': 6, 'Mission Hill': 4, 'Beacon Hill': 1, 'South Boston': 1} and a plot for this:

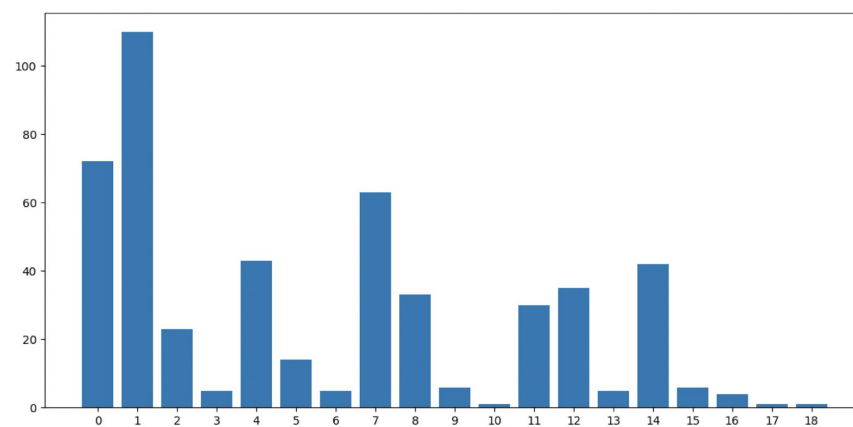


Figure.1 The calls of neighborhoods

where X coordinates are the index of the dictionary keys and Y coordinates are how many times this neighborhood appeared in bike blockage data.

Thusly we got the conclusion that Back Bay has the highest number of 311 calls related to bike lane blockages.

Problem 3: What is the breakdown of reasons cited for bike lane blockages? e.g. construction, double-parked vehicles, etc.

When we tried to find the specific reason for every block issue, we first using Boston's 311 data. But the difficult part is the breakdown reason for the block is hard to define.

Then we used Cambridge's 311 data, [Cambridge, MA](#), which is already filtered and all the data belong to bike lane block, and there are 1342 messages in total, so it's easier to classify the reasons.

We use 8 classes to define the reasons: Snow, Repair, Garbage, Uber/Lyft, Tree/Branch, illegal parking, Blank and Others. For the first 6 classes, we set some keywords for them separately, for example, we set a list: ['car', 'vehicle', 'park', 'truck', 'SUV', 'bus', 'shuttle', 'driver', 'delivery', 'taxi'] to extract "illegal parking".

"Blank" here means there are 226 rows' description are blank. "Others" here means the data we cannot tell the class based on the description: 234 rows' description don't have the keywords we set for other classes.

Following is the histogram for the classes.

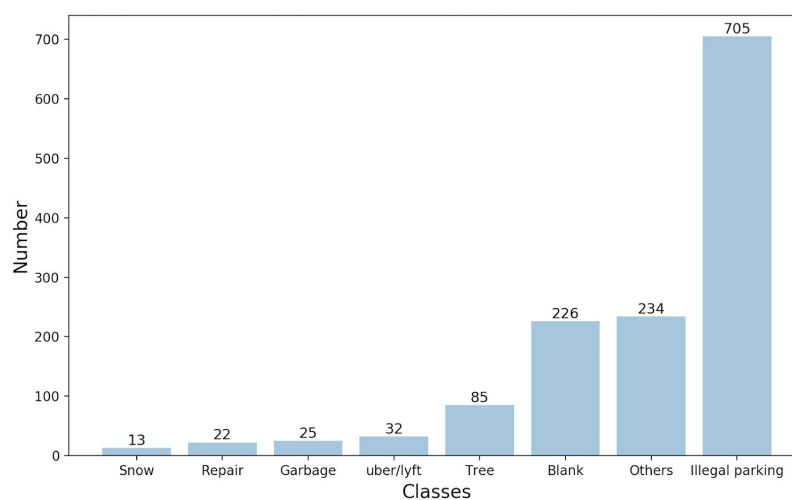


Figure.2 Bike Lane Blockage Classes

**Problem 4: Have 311 calls related to this issue changed over the past 5 years?
Seasonally?**

There have been some changes in 311 calls related to this issue over the past few years. As you can see in Figure 2 shown below, the number of bike lane blockage remains relatively low through 2011-2015, and suddenly start increasing from 2015 to 2018, and the decreased a little in 2019, but mostly because the data for 2019 is not complete since this year is not finished yet, so bike lane issue has been increasing fast these years, which is shown in Figure.3,

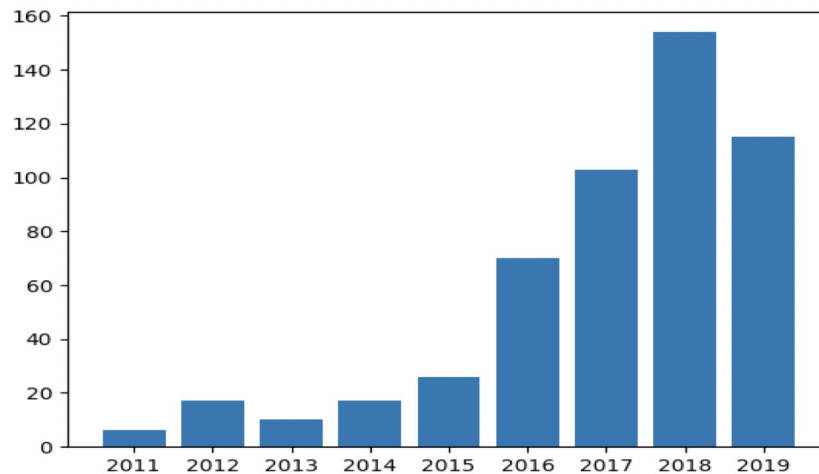


Figure.3 Bike Lane Issues from 2011 to 2019

We also count the on the time-solving rate (number of issues solved on time/number of issues) of bike lane blockage, finding that although the number of issues keeps increasing, the percentage of issues solved on time is not decreasing, in fact, it even increased a little.

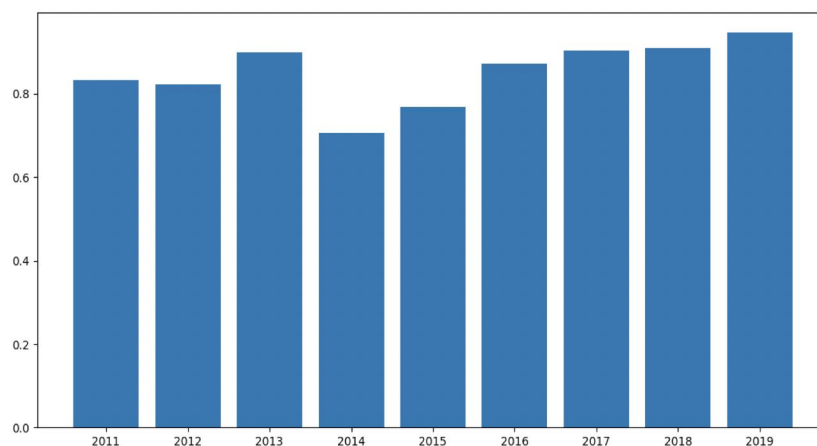


Figure.4 Time-solving Rate from 2011 to 2019

We also try to find the relations of issues and months (or seasons), and as figure 4 and figure 5 shows, it turns out that fewer blockages happen in winter and early spring (November to February), and mostly happens at mid of summer (June), but the on time-solving rate doesn't seem to be different through every month.

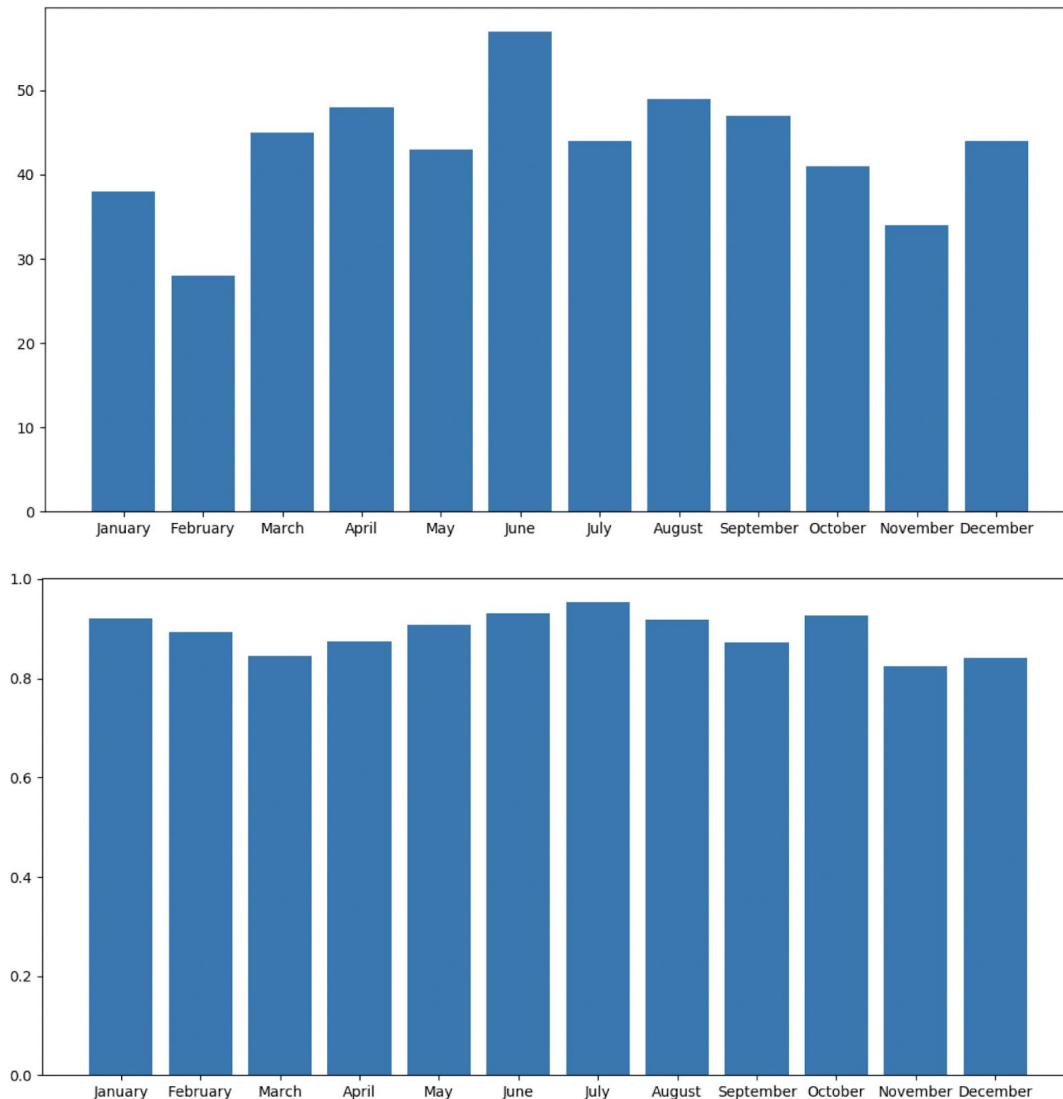


Figure.5 Time-solving Rate Seasonally Changes

To get these results, we first extracted rows for bike blockage in 311 data, and split the “open_dt” column into year and month, then threw them to two dictionaries. The year dictionary has keys of years and values of the number of issues that happened and on time-solving rate in that year, and the month dictionary has keys of months and values of the number of issues that happened and on time-solving rate in that month.

Bike Crash Data(Problem 5):

The data posted by the Boston government is with some common information about crashes that happened from 2015 to 2019. This dataset contains some columns named **mode_type**, **dispatch_ts**, **location_type**, **street**, **xstreet1**, **xstreet2**, **x_cord**, **y_cord**, **lat** and **long**.

For this project, **dispatch_ts**, **mode_type**, **lat** and **long** provide our team with basic information about the crash. The **dispatch_ts** is the dispatch time of the police which can be considered as the time when the crash happened. The **mode_type** is the type of crash and 'bike' means a bike crash. The **location_type** means whether this crash happened in Intersection or somewhere else. The **lat** and **long** offer geographical information about the crash. The **street**, **xstreet1**, **xstreet2**, **x_cord**, and **y_cord** are not that useful which may not be included in this project.

When obtaining **dispatch_ts**, **mode_type**, **lat** and **long**, we can establish the correlation between the crashes and bike lane blockages over the years.

The 311 Boston Data is from 2011 till now, but the crash data is from 2015 till now. So our team designed to utilize the data beginning on 01/01/2015 to analyze the correlation between bike blockages and crashes. As the first concern, we visualized the number of crashes by years. The trend is shown in Figure.6,

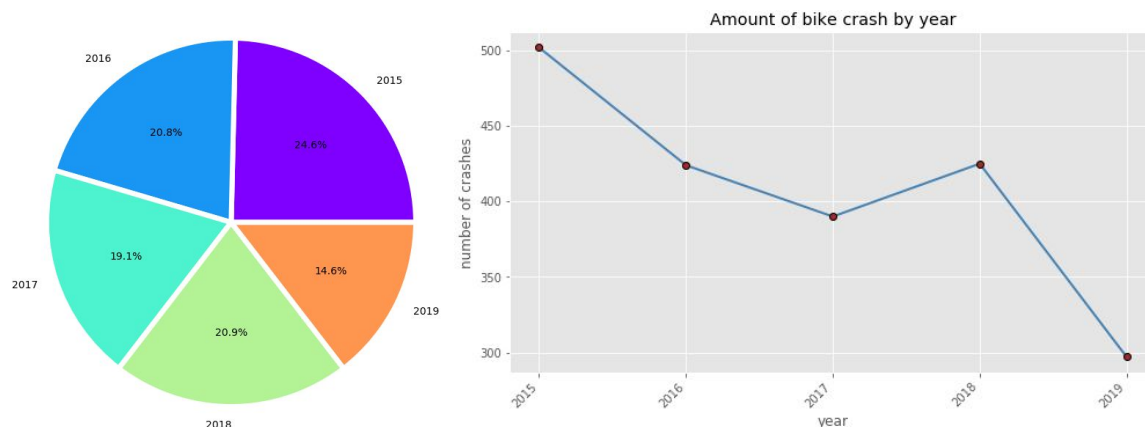


Figure.6 Percentage and Trend of crashes over years

According to the Figure., the highest number of crashes is in 2015 and the lowest is in 2019, and we can find there exists a decrease in crashes in Boston through the five years via the line chart, which is absolutely a good situation for the residents in Boston.

To observe it more clearly, We concentrate on the 2015 crash data to plot the data on Google Map in Figure.7,

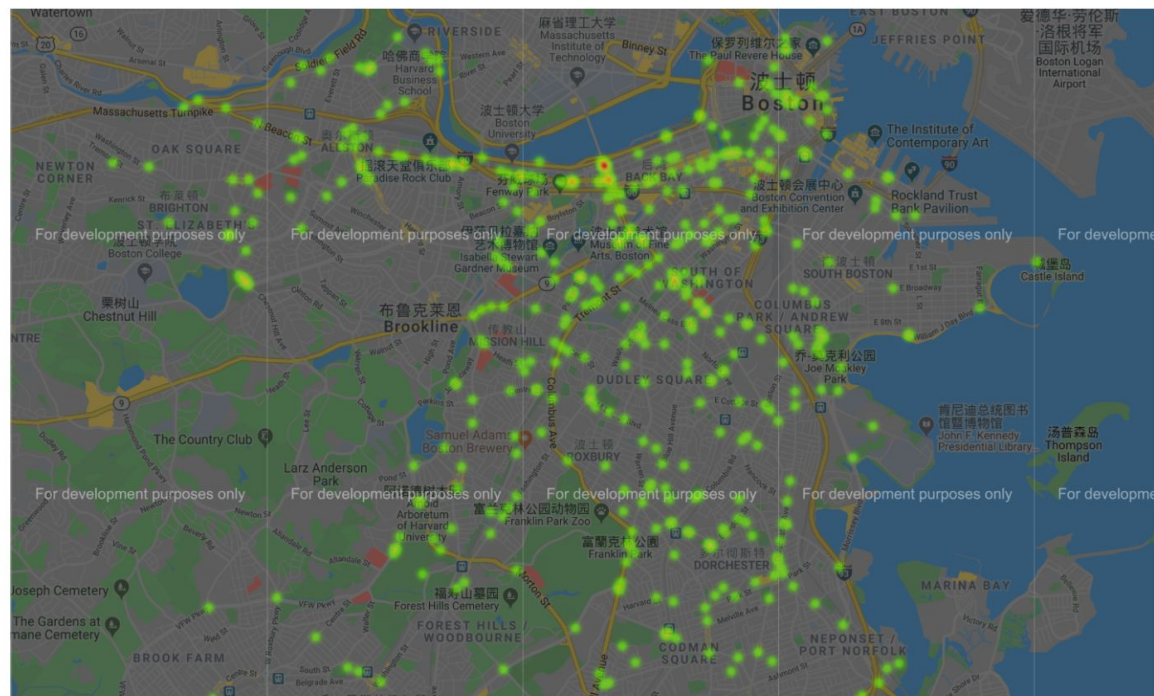


Figure.7 Heatmap of bike crashes of Boston in 2015

The main bike lane blockages are surrounding the commonwealth. After primary recognition of the data then we move to establish the correlation between the crashes and blocks.

Problem 5: What is the correlation of bike lane closures to reported bike crashes?

The crash dataset contains 20590 records of the crashes that happened in Boston from 2015 till now and 2038 of them are related to bicycles. By computing the date and coordination, we can figure out the relationship between the blocks and crashes. One thing that needs to notice is that the date in the block dataset is with the time zone of US Eastern, but the date in crashes is based on None TZ. As a result, our team needs to set the TZ to None when processing the data.

However, when we implemented the code for this question, the obstacle stopped our steps which is the size of blocks in Boston is too limited to build the relationship between blocks and crashes. That is we use a dictionary whose keys are the index of the blocks and each key has a list of crashes' indexes, but we can not find a single list when respected to the location and the date. The problem is shown in Figure.8,

0 : []	0 : []	0 : []
1 : []	1 : []	1 : []
2 : []	2 : []	2 : []
3 : [1645]	3 : []	3 : []
4 : []	4 : []	4 : []
5 : []	5 : []	5 : []
6 : [2386, 3011, 4795, 7350]	6 : []	6 : []
7 : []	7 : [3, 12, 26, 30, 104, 109, 139, 179, 181, 187, 190]	7 : []
8 : []	8 : []	8 : []
9 : []	9 : []	9 : []
10 : []	10 : []	10 : []
11 : []	11 : []	11 : []
12 : []	12 : []	12 : []
13 : []	13 : []	13 : []
14 : []	14 : []	14 : []
15 : []	15 : []	15 : []
16 : []	16 : []	16 : []
17 : []	17 : []	17 : []
18 : []	18 : []	18 : []
19 : []	19 : []	19 : []
20 : []	20 : []	20 : []
21 : []	21 : [3, 12, 26, 30, 104, 109, 139, 179, 181, 187, 190, 80, 886, 901, 903, 905, 909, 919, 938, 947, 948, 949, 976, 1113, 1114, 1119, 1125, 1126, 1128, 1140, 1157, 1159, 116, 1298, 1304, 1308, 1312, 1313, 1326, 1328, 1329, 1335, 134, 1478, 1493, 1496, 1503, 1506, 1508, 1509, 1530, 1536, 154, 1679, 1683, 1697, 1699, 1716, 1718, 1732, 1736, 1738, 173, 1839, 1850, 1862, 1865, 1873, 1874, 1875, 1878, 1884, 188, 1990, 1992, 2002, 2009, 2011, 2025, 2035, 2040, 2043, 204, 2141, 2152, 2158, 2189, 2198, 2199, 2207, 2210, 2215, 222, 2360, 2367, 2376, 2386, 2400, 2406, 2407, 2412, 2420, 242, 22 : [], 23 : [3, 12, 26, 30, 104, 109, 139, 179, 181, 187, 190], 24 : [3, 12, 26, 30, 104, 109, 139, 179, 181, 187, 190,	21 : []
22 : []	22 : []	22 : []
23 : []	23 : []	23 : []
24 : []	24 : []	24 : []
25 : []	25 : []	25 : []
26 : []	26 : []	26 : []
27 : []	27 : []	27 : []
28 : []	28 : []	28 : []
29 : []	29 : []	29 : []
30 : [4095]	30 : []	30 : []

Figure.8 Dict 1 with location

Dict two with date

Dict 3 with both

Even though we assumed the same location means that location ± 0.0001 of their coordinate, we can not find the obvious correlation between them based on the 311 data, and next step we plan to add more data from Cambridge to do further analysis.

Bike Count Data(Problem 7):

This dataset provides very detailed information about bike counts. Our team can easily find out the number of bicycles in a specific street or neighborhood. Corresponding to the 311 location data, it provides an opportunity for our team to build a direct correlation between the number of bike riders + 311 calls for lane closures.

Problem 7: Is there a direct correlation between the number of bike riders + 311 calls for lane closures? Seasonally?

For this question, the only thing we worried about is that there is not enough block data to build the correlation. If we meet this requirement, we can work out the correlation in the future.

Google Trend (Problem 6):

Problem 6: How have google searches of bike lane closures in Boston changed over the years?

One mainly usage of Google Trend is to show the changes in Google searches over the years. As there do not exist enough “bike lane block” or “bike lane closure” searches to plot. We scaled the searches plot to “Bike Lane” in Figure.9,

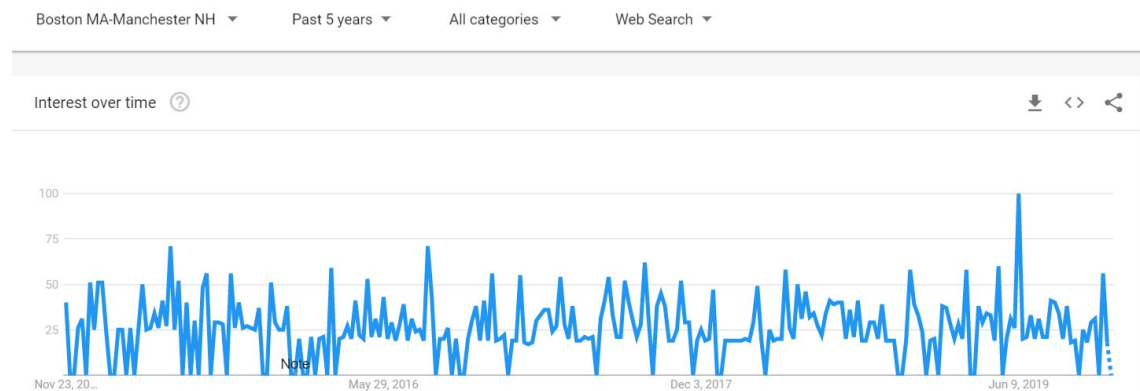


Figure.9 Bike Lane Searches over the past 5 Years

Question:

1. About question 1: We may misunderstand the “business” here, and shall we continue working on the correlation of blockage with business such as school, restaurant, and laundry?

Github Link:

https://github.com/mdche001/CS506_Bike_Lane