11/2022 - DS 4200 - ‹Group 8/Bo Sung Kim, Ethan Reres, Zhiheng Zhao›

Data Report: ‹Sewer and Drainage›

‹Bo Sung Kim, Ethan Reres, Zhiheng Zhao›

Project Summary

This project pertains to the sewer systems within the city of Boston, and more specifically in neighborhoods along the Massachusetts Avenue area. Sewer systems are a vital aspect of cities and towns that allow communities to stay healthy and safe throughout their day-to-day lives. As a proactive measure, the goal of an interactive data visualization will be placed, in order to keep public health in its best shape for as long as possible.

In order to keep the focus of proactive measures towards community health and safety, both public and private sewer systems in Boston need to be accounted for. Additionally, past blockage and weather data will be utilized in order to better prepare for a future mishap or trend. The visualization's intent is to effectively assist community organizations in bolstering their current sewer systems, and finding the areas that need the most help.

Data

Our main data is the overflow that occurred in Boston from 2021 to 2022, and the data that represents the location of the lead service. For the additional data, Bo Sung found the 311 service request from 2012 to 2022 data, Zhiheng found the Boston sewer network map in 1905, and Ehtan found the rainfall in Boston neighbors.

The overflow data can help analyze the accident-prone areas during 2021. Lead Service data can show the overview of the Boston sewer system. However, the data gives each lead services' longitude and latitude but do not show how the lead services are connected to each other.

The 311 service request data is good for analyzing incidents. If we analyze the data by collecting 10 years of data only related to the Boston Water & Sewer Commission, we can identify the area that has frequent accidents or predict the period when the accidents became frequent.

The Boston sewer network map in 1905 is a good source to analyze how the sewer system works. We need coordinate data to design our own map but it is difficult to find the coordinate of the sewer system in Boston. Therefore, we can analyze and use it even if the data type is an image. The rainfall data shows the amount of rainfall by the neighbors in Boston. This data helps which water facilities should be prepared in advance.

Clean-up & Challenges

The most difficult part of the clean up of the data was to select and organize the data we needed for the final project. In the case of 311 Service Requests, there was a big data size as there was 10 years of data. We selected the data only related to the Boston Water & Sewer Commission which is needed for our final work, and we selected the column which we think needed in our final project such as open date of request, close date of request, reason, and department to treat this problem. While observing the data, there were cases

where the data had empty space, so we thought about how to treat this. However, after selecting the columns which we needed, there was no empty space data. In the case of the network map data, we need the coordinate data to create our own map in python or tableau; however, we cannot find it. We only have the image data.

Data Type

In the SSO Location 1 Year data, SSO location is a categorical data type. The neighborhood, spill type are categorical data types. Start date, start time, end date and end time are ordinal data types. Cause and corrective action are categorical data types and x and y are quantitative data types. In the lead service dataset, lead service address is a categorical data type. X and y are quantitative data types.

In the 311 Service Requests 2012-2022 which we cleaned up the dataset from 311, we have 10 columns. The open_dt means open date of the service request and this is an ordinal data type. The closed_dt means closed date of the service request and this is an ordinal data type. The case status means that the case is still open or closed or overdue. The case_status is a categorical dataset. The closure_reason means the reason if the case is closed and this is a categorical data type. The case title is a categorical data type. The subject, reason and departments are categorical data types. The longitude and latitude are a quantitative data type.

The Boston map data is just image data. We cannot say any data type in the image. The rainfall data, the region is a categorical data type and the rainfall is a quantitative data type.

Links

Overflow:
https://www.bwsc.org/environment-education/maproom/combined-sewer-overflow-map-and-public-notification

Lead Service: https://www.bwsc.org/environment-education/maproom/lead-service-map

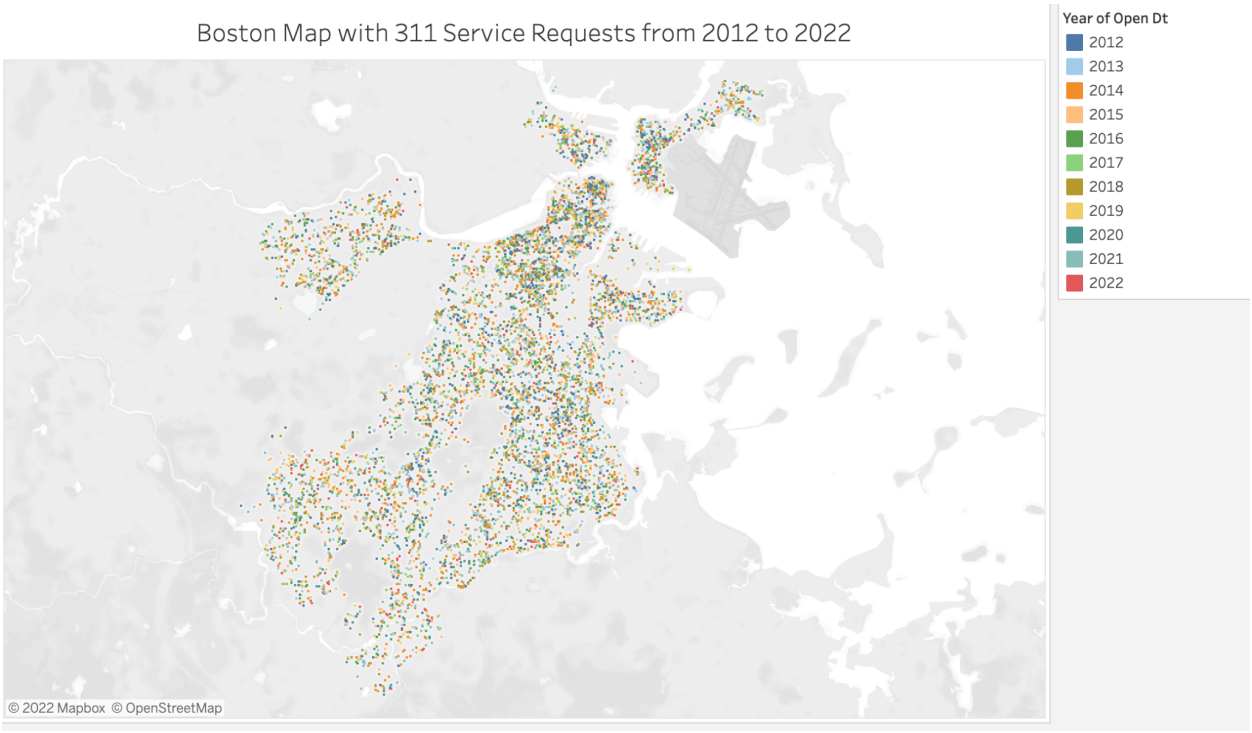
311 Service Requests: https://data.boston.gov/dataset/311-service-requests

Boston Sewer Map in 1905:

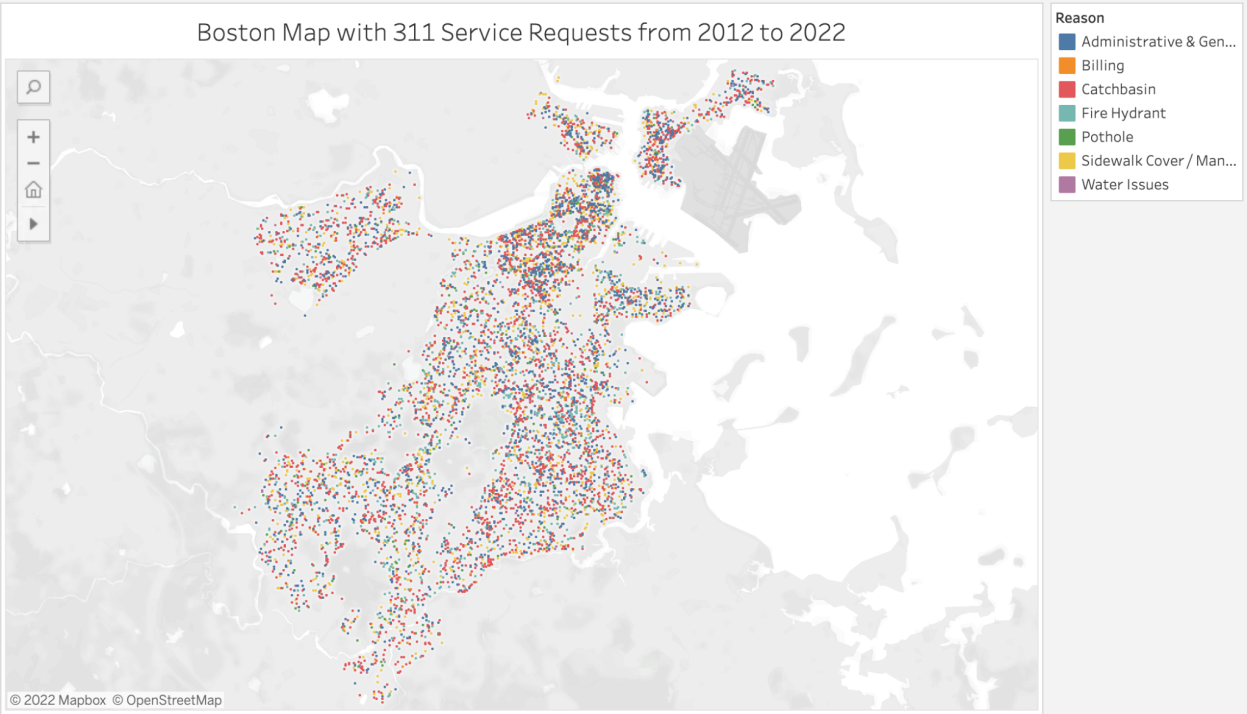
https://www.atlasobscura.com/articles/boston-historic-sewer-map-patriots-lost

Rainfall: https://data.boston.gov/dataset/rainfall-data

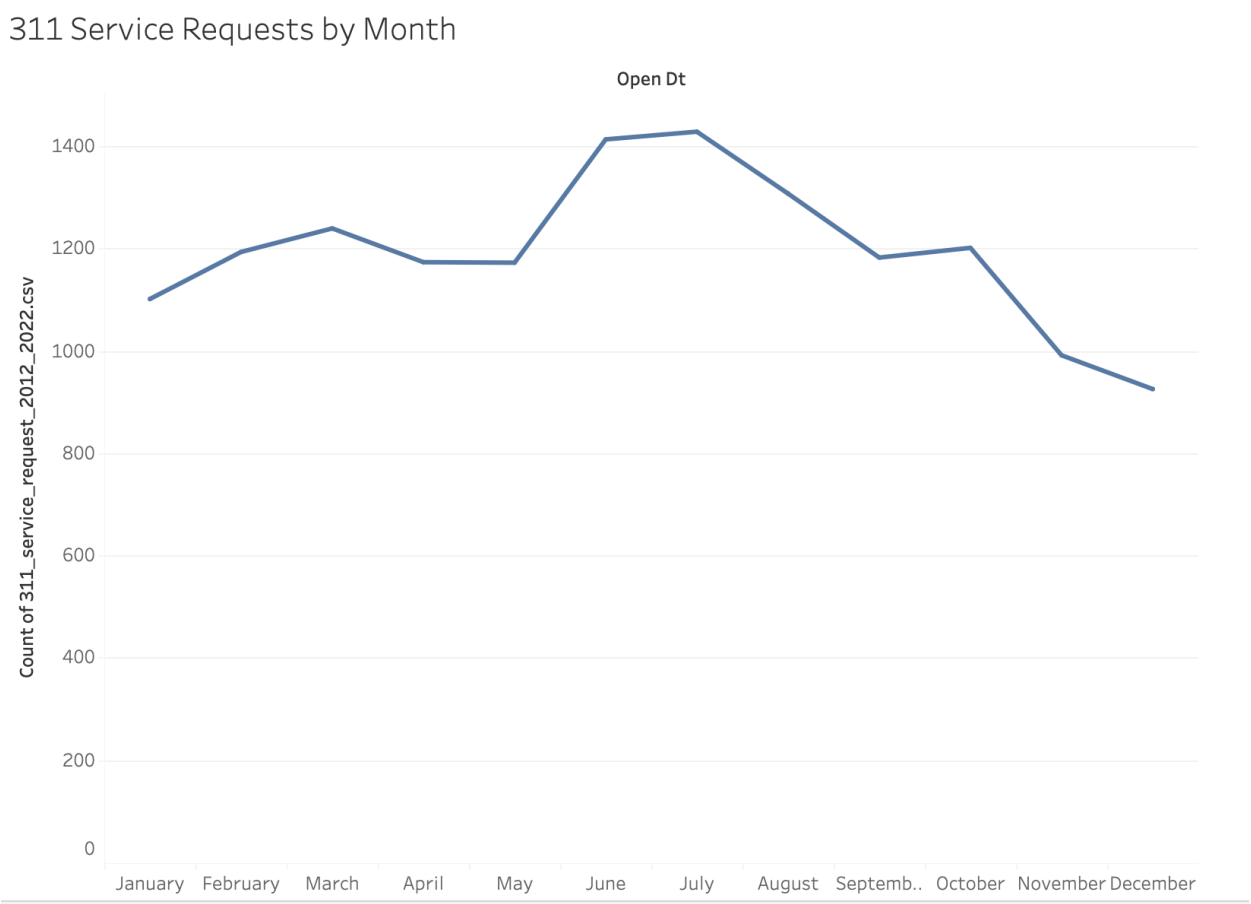
Observations & Insights



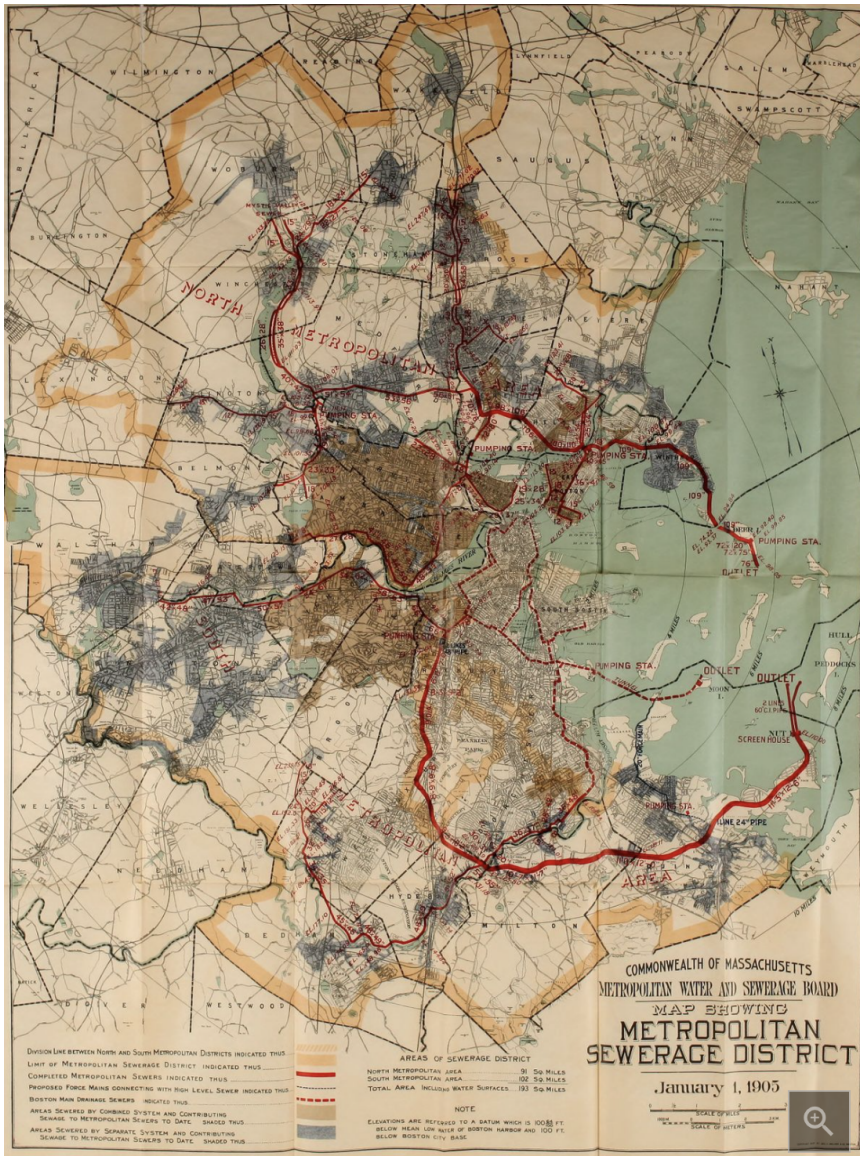
The map above represents the 311 Service Requests from 2012 to 2022 in the Boston map. If the viewer clicks the year in the legend, located on the right side of the map, the map shows the specific year's data. The map shows that the 311 service request areas are concentrated in the Back Bay, Beacon Hill, Bay Village and South End. This trend is similar in other years.



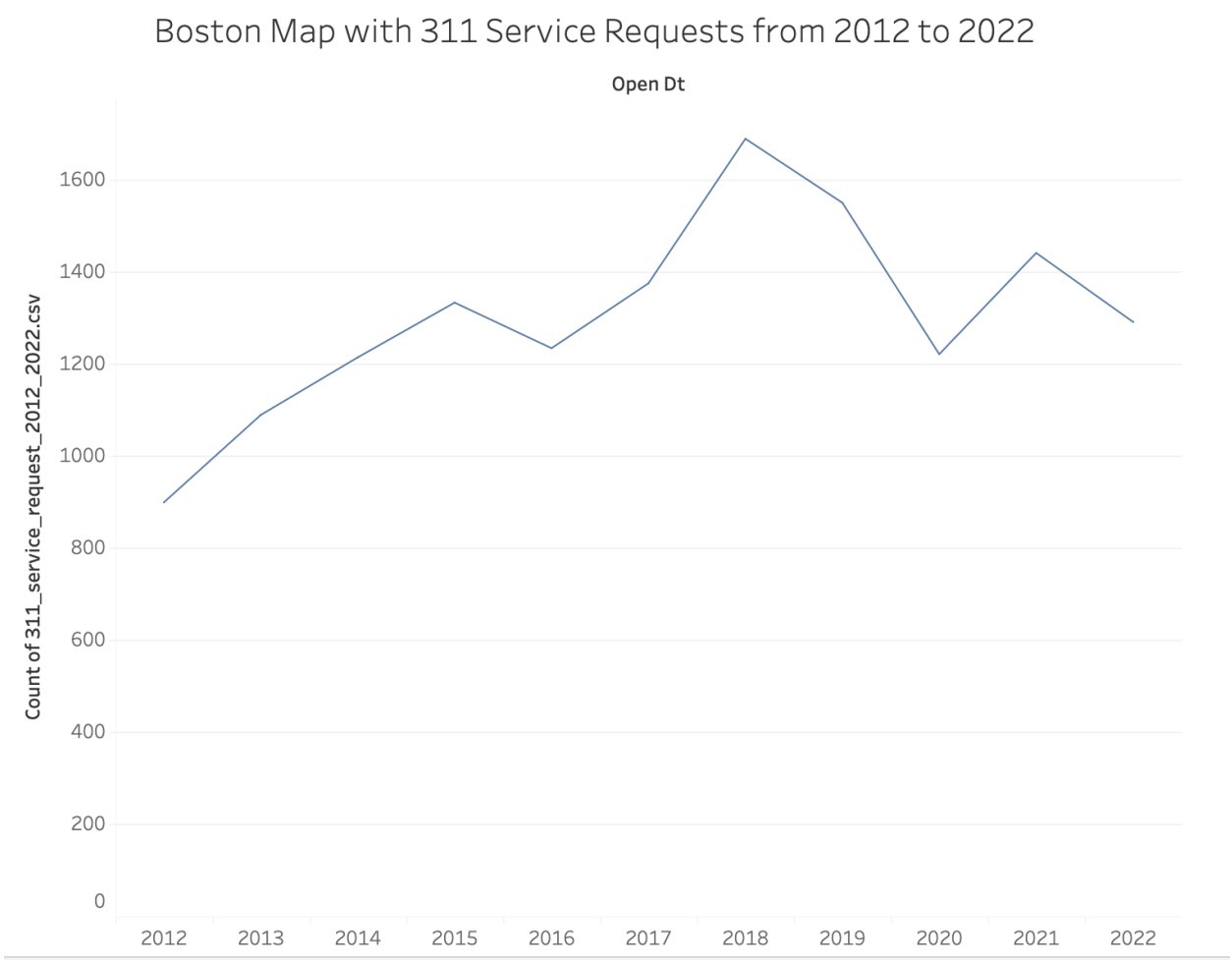
This is a map showing the locations for different requesting reasons for 311 service requests from 2012 to 2022. Each color represents a requesting reason. If the viewer clicks the reason in the legend, located on the right side of the map, the map shows the specific reason's data. We can see from the map visualization that the locations for each kind of requesting reason are evenly distributed across Boston.



This graph represents the number of 311 service requests per month. From this, it can be seen that there are mainly many requests from May to June. This trend shows in the different years. The results we expected are similar to the results of rainfall, but it is different from the trend of rainfall. The peculiarity is that there are the most requests in May and June, not in the winter when there is a lot of rainfall.



This data is a sewer network map of 1905 Boston. We could not find the latest data, or the data contained locations, but we could find a sewer network map of the past. As Boston is a coastal city, it can be seen that the sewer system runs from the sea to the city. From this data, we can see how the entire network of sewer system was organized and worked



This is a line chart showing the change of number of 311 service requests from 2012 to 2022. From this line chart we can see that the number of requests was roughly increasing from 2012 to 2018. Then decreased from 2018 to 2020 and roughly remained the same after that. The peak was 2018.

Task Analysis

Index (ID #)	“Domain” Task	Analytic Task (Low-level, “Query”)	Search Task (Mid-level)	Analyze Task (High-level)
1	Examine overflow in sewer system, where the most frequent accidents occur	Find Extremum	Locate	Present
2	Analyze rainfall in Boston neighborhoods to find out the relationship between rainfall and sewer overflow	Correlate	Look up	Discover

Conclusion

As this project is developing, several findings as well as challenges are arising. We are able to find and specify sewer accidents with spill data, service requests, and neighborhood data, as provided by the Boston Water and Sewer Commission. However, finding reliable water retention data, and connecting our findings from public sewer systems to private sewer systems are proving to be an issue. Private data is not as uniform compared to public, and so special care will have to be taken in order to port over the same functionality in our visualization for both types of data.

Our next steps involve finishing up data analysis for private sewers, water retention, and weather data, and then deciding on a goal for an effective interactive data visualization. The map should have aspects that delineate a neighborhood’s accident

history, sewer quality, and other miscellaneous data that should help professionals decide where to focus their work on.

Part 1:

As part of your Assignment 7b submission document, include a page(s) which lists links to each of your chosen additional data sets (and which team member contributed the data) as well as a brief (few sentence) description of the data and why it is relevant for your final project.

1. Bo Sung: 311 Service Request: https://data.boston.gov/dataset/311-service-requests
   311 Service Request is good to analyze the accident related to the Boston water & sewer commission. There are many types of accidents such as general requests, sidewalk manhole problems and water overflow. We can analyze service requests and find out which areas need improvement by looking at accident-prone areas.
2. Zhiheng: Sewer Network Boston Map in 1905:
   https://www.atlasobscura.com/articles/boston-historic-sewer-map-patriots-lost
   The data is great to show the overview of the sewer system in Boston. Although it is data from the past, it is valuable data considering that recent data cannot be found and the sewer system does not change much.
3. Ethan: https://data.boston.gov/dataset/rainfall-data
   a. https://www.bwsc.org/environment-education/rainfall-garden
   b. Rainfall data is a valuable aspect of nature and weather data. It will be added in to account for future trends, and what neighborhoods usually receive more rainfall than others. There may be a correlation between rainfall amount and sewer quality or accident frequency.