Exploring Threats to Biking in Boston: An Introductory Analysis of BlueBikes, Sea Level Rise, and Crash Incidents in Boston, MA

Zack Armand

Alex Kramer

Northeastern University

DS5110: Data Management and Processing

Instructor*: John Rachlin*

June 22nd, 2023

Abstract

This report presents an introductory analysis on the threats to biking in Boston, focusing on two main sub categories: infrastructure and climate change. The goal was to gain meaningful insights into challenges cyclists in the city face, and threats that biking in Boston faces. In our Infrastructure analysis, we study the occurrences of bike crashes mapped across the city to investigate factors contributing to these incidents. This analysis led us to a need for more data, such as historical infrastructure data on: bike lanes, intersections, and the improvements into infrastructure. By understanding bicycle crashes across the city, we were able to high risk areas across the city to cyclists. These specific geographical hot spots need further research to understand why crashes are occuring there. Through this report we hope to provide an introductory analysis on the threats to biking in Boston.

Regarding climate change, we explored the impact of rising sea levels and the potential emissions saved through increased biking. By overlaying sea level rise data with Bluebike stations, and bike lanes we find what areas are most susceptible to the effects of climate change. Moreover we calculated the loss of biking in Boston, in different sea level rise scenarios. Finally we quantified the emissions reduced by Bluebike trips in 2022. This clearly highlights the environmental benefits of biking as a mode of transportation. By analyzing both infrastructure and climate change we were able to provide an introductory analysis on the complex dynamics between biking, safety, climate change, and sustainability.

Introduction

Biking has emerged as a popular sustainable mode of city transportation, offering a number of benefits. However in certain cities, the safety and viability of biking faces significant challenges. In many cities, those challenges range from infrastructure deficiencies to the impact of climate change. In this study we explore the threats to biking in Boston, specifically looking at infrastructure and the impact of climate change. This project aims to act as an introductory analysis on these threats to biking in Boston.

To guide this research we ask several key research questions: Where are bicycle crashes happening in Boston? Which Bluebiking stations and biking infrastructures are vulnerable to the impact of rising sea levels? How much emissions can be saved through Boston Bluebikes? By addressing these questions, we seek to provide insights in support of developing Boston as a more bike-friendly city. Overall, this report represents a small, yet important step in understanding and addressing threats to biking in Boston.

Data Sources

The data sources for this introductory analysis were mainly sourced and downloaded from Analyze Boston, an open data hub managed by the City of Boston Analytics Team. However, Bluebike Trip data and Bluebike Station data were downloaded through the Bluebike data portal. The Massachusetts municipal boundaries shapefile was downloaded from MassGIS, the Commonwealth of Massachusetts' open data portal. The data sources utilized in this project are generally broken down into two categories: tabular data and geospatial data. Tabular data includes: Bluebike Trip Data from 2017 and 2022, Bluebike Station data, Boston Police Crash Reports, and Boston Crash Fatality Reports. Geospatial data (in the form of shapefiles) include: Boston Neighborhood polygon, and all sea level rise (SLR) 10% storm event polygons (9 inch, 21 inch, and 36 inch rise), Massachusetts municipalities, and Boston Census Block Groups.

Methods

Bluebike trip data was originally formatted as a CSV for each month of the year. These files were combined into a file for 2017 and 2022 using Python. A single SQLite database of the tabular data was created using SQLite (version 3.36.0) and DBeaver (version 21.3.0). The main data cleaning task performed was properly filtering the data with a range of years down to just 2017 and 2022, creating additional views and tables for each dataset as needed. This allowed for more consistent and easier access once using Python. For example, the new table 'bluebike_trips_2017' captures all of the relevant information from the Bluebike trip data from 2017. One unique view created in SQL was 'bike_5001', which filters for trip information associated with the Bluebike bike ID 5001 for 2022. The years 2017 and 2022 were chosen for further analysis due to the completeness of data associated with each year, and to enable proper temporal comparison when needed.

Once the views and tables were established in the database, Python's SQLite3 library was used to establish a connection to the database within Jupyter Notebooks and draw in tables and views as needed. Utilizing Pandas, the data from the database was transformed into dataframes that were then further transformed, analyzed and visualized to perform an initial analysis on the data. Pandas dataframes with longitude and latitude data were often converted into a Geopandas geodataframe in order to spatially analyze the data.

Throughout this project, crash and fatality counts were only analyzed within the boundaries of Boston, due to limited public data on crash incidents or Vision Zero data from other municipalities. Crash data was grouped by hour of day, day of week, and season of the year in order to view and compare temporal trends in crashes. Season was broken down into hot and cold months. Hot represents April through September, and cold months represents October through March.

Bike crash counts per census block in Boston were calculated by spatially joining crash point data to census block polygons, and performing a count of the number of points within each block. These counts were then normalized by the area of each block group to quantify the relative density of crashes throughout Boston to visualize areas of high danger to cyclists. The area of the blocks were measured in square meters, so these areas were first converted to square kilometers in order to obtain the number of crashes per square kilometer in a block. These figures were then plotted to easily see the spatial variation in crash frequency.

The number of Bluebike stations impacted by various sea level rise scenarios was determined by counting the number of stations within each of the three SLR 10% storm polygons using a spatial join. Additionally, the total number of trips that started at each station in the system was calculated by summing up the count of rows after filtering the full trip dataframe for that station. Just start stations were chosen in order to not double-count trips - for example, a trip that started and ended at inundated stations would be double counted if start and end station counts were used. The number of impacted trips and stations due to SLR storm inundation were then calculated for each event, and these figures were compared to the total number of stations (rows in the stations dataframe) and total number of trips (rows in the trips dataframe). Inundated and unimpacted stations, sea level flooding, and bike lanes were then mapped. The station dot size was scaled to reflect the percent of trips originating from that station compared to the average number of trips originating at stations across all stations (trips originating at that station / average number of trips originating at all stations).

We also calculated total distance traveled by Bluebikes in 2017 and 2022, respectively, by calculating the distance between the start station and the end station using a Haversine function to calculate the distance between the two points based on their latitude and longitude data. However, this distance represents the absolute minimum distance traveled during a trip. Bluebikes does not provide high resolution trip data, only the start and end points of a trip. Therefore, the total distance traveled per year can be considered a bare minimum of distance traveled by Bluebike riders. To provide additional context, we referenced the Environmental Protection Agency's (EPA) statement that "the average passenger vehicle emits about 400 grams of CO2 per mile." [13] Once the total minimum trip distance was calculated, we then multiplied this mileage by 400 and converted from grams to metric tons to represent the total CO2 emissions saved by biking instead of driving.

Analysis

This paper examines two main threats to biking in Boston: Infrastructure and Climate Change. Under Infrastructure, this paper investigates bike lanes, crashes, fatalities, and Bluebike usage in Boston. Regarding Climate Change, the analysis focuses emissions and on the impact of rising sea levels on Bluebike stations and bike lanes. This section is adopting this structure in order to most effectively organize the information to be presented in a clear and concise manner.

Infrastructure

One of the prominent threats to biking in Boston is crashes involving bicycles. Using the City of Boston's public safety incident crash data, there were 390 bike crashes in 2017, and 242 bike crashes in 2022. Figure 1, examines the spatial distribution of crashes by census block level. By associating crash data with census blocks, we gain insights into areas where crashes are most prevalent. We are able to more clearly see areas where there exists a disproportionate number of crashes. An area of future analysis is utilizing this figure to further investigate why certain areas are hot spots for these crashes, although they do then to be clustered in certain areas of the city.

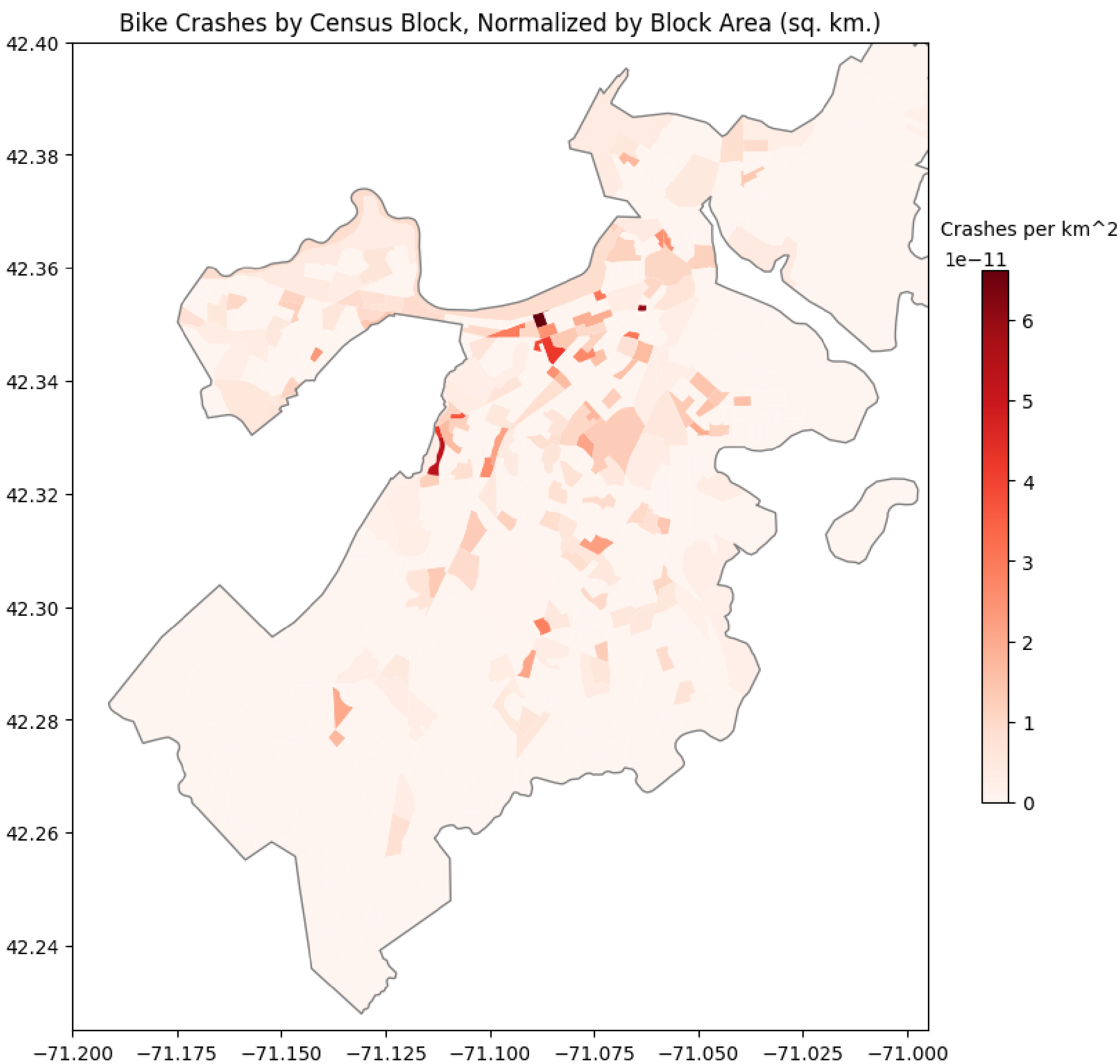

Figure 1: Bike Crashes by Census Block

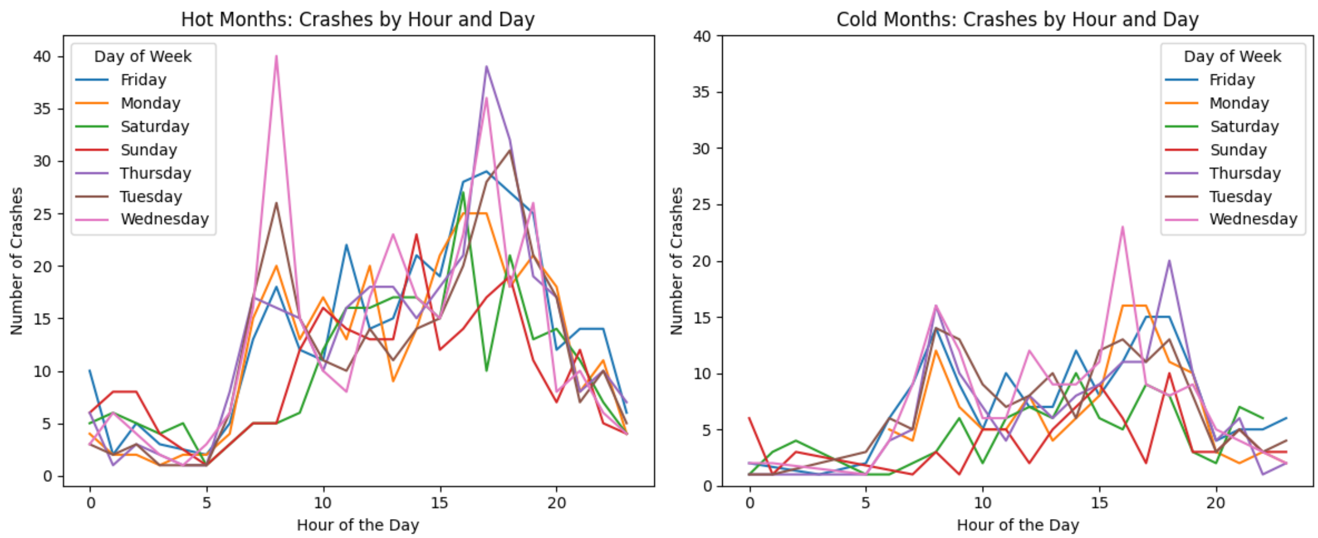

Figure 2: 2022 Bike Crashes by Season, Time, and Day of the Week

To further understand the trends and patterns of bicycle crashes, we analyzed the Bluebike trip data by hour, day of the week and season in Figure 2. Notably, we can see that more crashes occur in the hotter months, for both seasons we see that the most crashes occur during typical commuting hours. Presumably, there are more bicycle and car trips during these times, in part explaining some variation in crash count data.

Climate Change

Sea level rise due to climate change has the potential to impact biking within Boston. A 10 percent storm under the three SLR scenarios produced by Climate Ready Boston have varying impacts on the Bluebike network and the number of stations that would be inundated with water in such a storm. For example, the stations impacted due to flooding during a 10 percent storm under a 21-inch SLR can be seen in Figure 3.

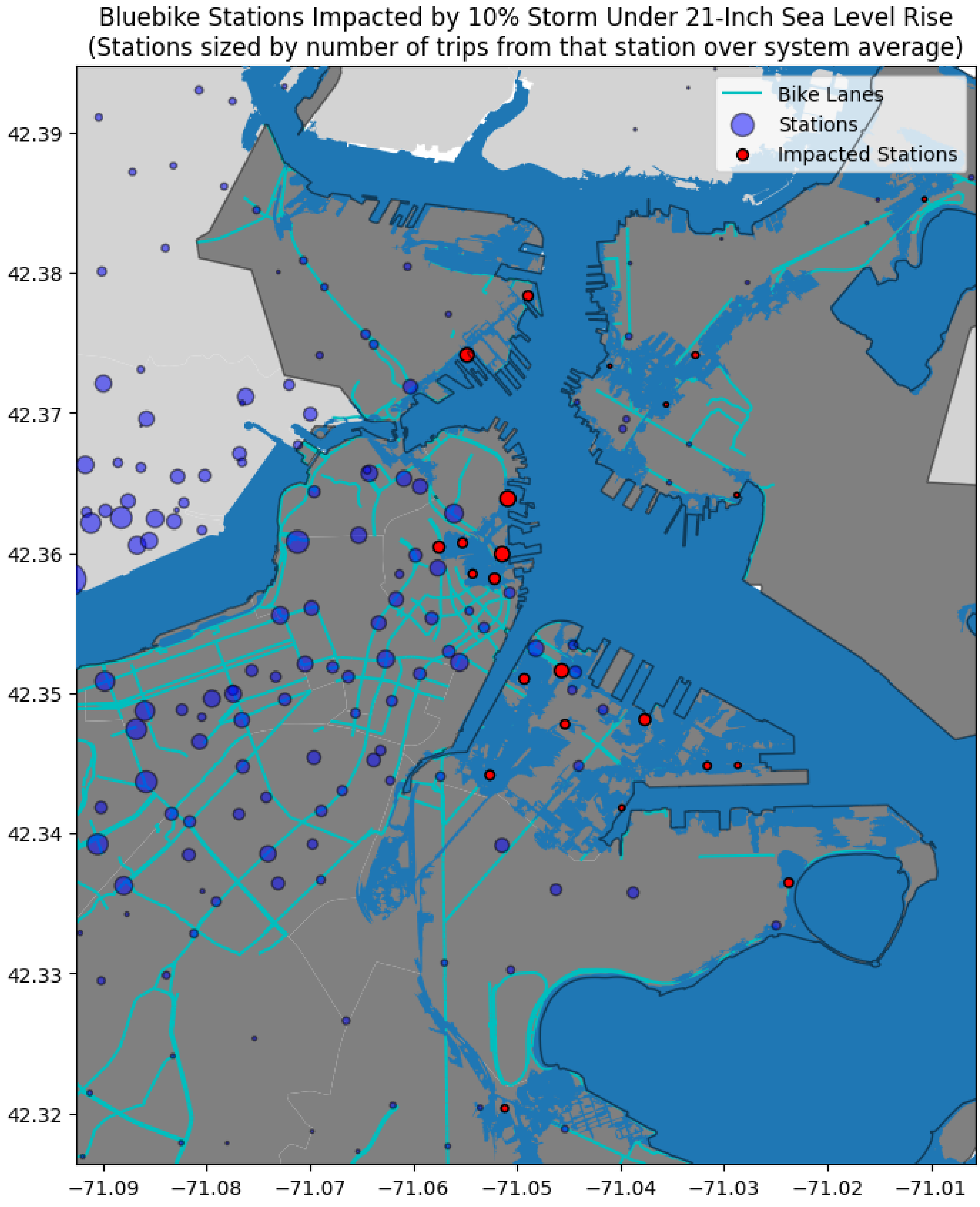

Figure 3: Impact of 10% storm under 21-inch sea level rise on Bluebike stations

The threat of sea level rise on the Bluebike network can be further quantified by examining the number of stations and rides impacted by such storms. The exact number of stations and rises impacted during storms under the three SLR scenarios can be seen in Table 1.

SLR Scenario	Stations Impacted	Percent of Total Stations	2022 Trips Lost	Percent of 2022 Total Trips
9 inch	8	1.75%	84,249	2.24%
21 inch	23	5.02%	175,858	4.68%
36 inch	56	12.23%	421,392	11.22%

Table 1: Bluebike Stations Impacted by 10% Storm Under Different SLR Scenarios

Out of the total 458 Bluebike stations in the Greater Boston Region, 8 would be inundated in a 9 inch sea level rise storm; 23 in a 21 inch SLR storm, and 56 in a 36 inch SLR storm. These stations are slightly busier than the average stations, with the number of trips originating from these stations being around 0.5-1% higher the percent of stations in the system they account for. Given that 36 inches of sea level rise are projected under all climate scenarios for the 2070's, impacting more than 10% of stations and trips in the Boston area poses a significant threat to biking in Boston.

While climate change poses a threat to biking in Boston, biking can provide a net positive environmental benefit. Replacing vehicle trips with bike trips can provide a prime opportunity for reducing greenhouse gas emissions, and according to the EPA, "car trips of under a mile add up to about 10 billion miles per year" nationwide [14]. Given that the average Bluebike trip in 2022 was around 1.19 miles, the trips taken by Bluebike users can be viewed as prime candidates for vehicle replacements. While it is difficult to quantify the exact ratio of vehicle trips replaced by biking, analyzing biking trips can present an optimistic view of the greenhouse gas reduction potential of cycling as a replacement for driving. Using the average emissions per mile of a car in the U.S. and assuming a 1:1 car:bike replacement ratio, we calculated the saved emissions by biking instead of driving a car. The total number of trips and mileage as well as metric tons CO2 emissions saved (based on the average grams of CO2 per mile of a car) can be seen in Table 2. These figures represent an absolute minimum, based on the linear distance between the start and end stations of each Bluebike trip.

	2017	**2022**	**Increase**
Total number of trips	1,313,774	3,757,281	2,443,507
Total trip distance (miles)	1,744,623	4,467,589	2,722,966
CO2 emissions saved (metric tons)	697	1,787	1,090

Table 2: Total emissions saved by biking

Conclusions

Our analysis aims to provide a valuable introductory analysis on the complex dynamics between biking, safety, climate change and sustainability. This introductory analysis was able to answer questions such as: Where are bicycle crashes happening in Boston? In figure 1, we can see where these crashes are happening but we lack the data to further investigate why we are seeing crashes in these areas. We were able to find that Bluebikes are offsetting a minimum of 1,787 CO2 emissions in 2022, with an upwards trend of more bikers every year. Finally, we found that the sea level rise will greatly affect biking in Boston, submerging 23 out of the current 251 Bluebike stations. These findings shed light on the threats to biking in Boston, and the complex dynamics between biking, safety, climate change and sustainability.

Limitations & Future Steps

Bluebike Data: One limitation for our analysis is that the Bluebike data does not provide an exact route taken by the cyclists. This information would have been valuable to compare it to the crash data, in order to see any overlaps and intersections between the two.

Geographic Scope: Our analysis focused just on Boston proper, and while this provides valuable insights into the local biking landscape, the findings may not be directly applicable to other regions. Further research could include multiple cities to take a more macro lens.

Bike Lanes Data: The available bike lane data may not fully represent Boston's bike lane network. Incorporating historical bike lane networks, including quantifying the different types of bike lanes and the number of bike lane miles over time, would give a more thorough view analysis between bike lanes and crashes.

Bike Crash Reports: The data provided in this dataset was from the Boston Police Department. This means that our crashes were only accounted for if a first responder (police) was sent to the site of the crash and relies on victim self-reporting.

Author Contributions

Author	**Contributions**
Zack Armand	- Initial Research - SQL Database - Climate change figures & report Methods, analysis, conclusions, references
Alex Kramer	- Initial Research - Infrastructure figures - Abstract, introduction, methods, analysis, conclusions

References

1. Bluebikes. “Blue Bikes Comprehensive Trip Histories”. https://s3.amazonaws.com/hubway-data/index.html

2. Bluebikes. “Bluebikes Stations.” https://s3.amazonaws.com/hubway-data/current_bluebikes_stations.csv

3. City of Boston. “2020 Census Block Groups in Boston.” Analyze Boston. Date generated: September 24, 2021. https://data.boston.gov/dataset/census-2020-block-groups

4. City of Boston. “9inch Sea Level Rise 10pct Annual Flood.” Analyze Boston. Date generated: September 22, 2017. Date modified: July 8, 2020. https://data.boston.gov/dataset/9inch-sea-level-rise-10pct-annual-flood

5. City of Boston. “21inch Sea Level Rise 10pct Annual Flood.” Analyze Boston. Date generated: September 22, 2017. Date modified: July 8, 2020. https://data.boston.gov/dataset/21inch-sea-level-rise-10pct-annual-flood

6. City of Boston. “36inch Sea Level Rise 10pct Annual Flood.” Analyze Boston. Date generated: September 22, 2017. Date modified: July 8, 2020. https://data.boston.gov/dataset/36inch-sea-level-rise-10pct-annual-flood

7. City of Boston. “Boston Neighborhoods.” Analyze Boston. Date generated: February 16, 2017. Date modified: December 7, 2020. https://data.boston.gov/dataset/boston-neighborhoods

8. City of Boston. “Existing Bike Network 2022.” Analyze Boston. Date created: June 3, 2022. https://data.boston.gov/dataset/existing-bike-network-2022

9. City of Boston. “Hydrography (Polygon).” Analyze Boston. Date generated: October 21, 2014. Date modified: February 24, 2023. https://data.boston.gov/dataset/hydrography-polygon

10. City of Boston. “Vision Zero Crash Records.” Analyze Boston. Date generated: May 1, 2018. Date modified: June 20, 2023. https://data.boston.gov/dataset/vision-zero-crash-records

11. City of Boston. “Vision Zero Fatality Records.” Analyze Boston. Date generated: July 23, 2018. Date modified: June 20, 2023. https://data.boston.gov/dataset/vision-zero-fatality-records

12. Commonwealth of Massachusetts. “MassGIS Data: Municipalities.” MassGIS. Date: April 2022. https://www.mass.gov/info-details/massgis-data-municipalities

13. United States Environmental Protection Agency (U.S. EPA). “Tailpipe Greenhouse Gas Emissions from a Typical Passenger Vehicle.” EPA-420-F-23-014. June 2023. https://www.epa.gov/system/files/documents/2023-05/420f23014.pdf

14. United States Environmental Protection Agency (U.S. EPA). "What If We Kept Our Cars Parked For Trips Less Than One Mile?" EPA-420-F-15-021. June 2015. https://nepis.epa.gov/Exe/ZyPdf.cgi?Dockey=P100MLPQ.pdf

Data Dictionary

- **Bluebike Trip 2017 & 2022 [1]:** These datasets contain Bluebike system data for each trip ridden in 2017 and 2022, structured as an individual CSV file for each month. This dataset contains a comprehensive trip history with columns such as trip duration, start time and date, stop time and date, start station, end station, bike ID and more.
- **Bluebike Station [2]:** This dataset contains Bluebike Station data including columns such as each station's Number, Name, Latitude, Longitude, Total Docks, and more. This data was collected and made available by Bluebikes.
- **Boston Crash Reports [10]:** This dataset includes information on crashes that required a public safety response. This data was collected from the Boston Police Department.
- **Boston Crash Fatalities [11]:** This dataset is a part of the Boston's Vision Zero program. It includes records on date, time, location, and type of fatality in Boston from 2015 - 2022. This data was collected from the Boston Police Department.
- **Boston Census Block Groups [3]**: This shapefile contains the polygon boundaries for 2020 Census block groups in Boston, developed by the U.S. Census Bureau but sourced by Analyze Boston.
- **Sea Level [4,5,6]:** These shapefiles outline the coastal and riverine flooding impacts of a 10 percent storm event (10-year flood) under three sea level rise scenarios: 9 inch rise, 21 inch rise and a 36 inch rise. These projections are part of Boston's Climate Ready assessments.
- **Boston Neighborhoods [7]:** The Boston Neighborhoods dataset is a collection of neighborhood boundaries, zip codes, and the 2010 Census tract boundaries. These boundaries are not considered official by the City of Boston, but provide a rough outline of the City of Boston.
- **Boston Bike Lane 2022 [8]:** This dataset provides a collection of current bicycle lanes within the Boston city limits. The dataset was obtained from the City of Boston Transportation department.
- **Boston Hydrography [9]:** This dataset is the geospatial boundaries of the various waterbody features within the City of Boston. The shapefile was obtained from AnalyzeBoston.
- **Massachusetts Municipalities [12]:** This dataset, obtained from MassGIS, is a shapefile of all municipal boundaries within Massachusetts.