

Waterfront Visitation and Equity Study: Groundtruthing Follow-Up

Author: Ford Fishman

Introduction

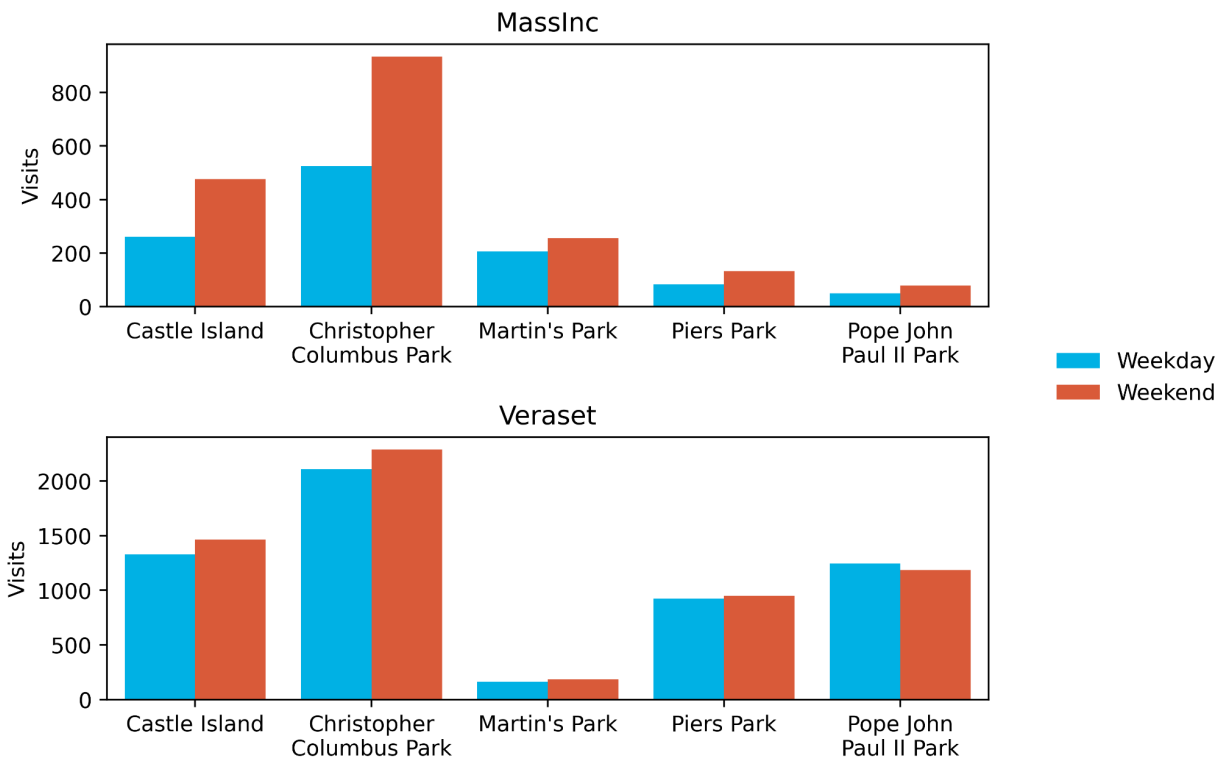
As a follow-up to my first [Waterfront Visitation and Equity Study Project Report](#), I have conducted several more analyses comparing the MassInc and Veraset datasets. For more background information on my analysis, please refer to [my first report](#). This follow-up contains four additional comparisons of the datasets: [Overall visitation by weekday and weekend](#), [repeat visitation](#), [out-of-state visitation](#), and [visit home cities and neighborhoods](#). Additionally, to conclude the groundtruthing, I have included an [inter-rater reliability analysis](#) to quantify the degree of agreement between the MassInc and Veraset data.

Methods

To start, I will recap the data acquisition methods used in the [prior report](#). The Veraset data is hosted on a geospatial data platform called Tectonix. To access the Veraset data, I met with Tectonix team members for hands-on instruction on the platform. I created a dashboard consisting of polygons of the five parks corresponding to the MassInc data. Due to some abnormally high visitor and visit counts for Christopher Columbus Park for August 2022, a more natural analog for August 2023, I solely exported August 2021, as the next closest analog. The one exception to this is the home location data, for which data from the entirety of 2021 was exported.

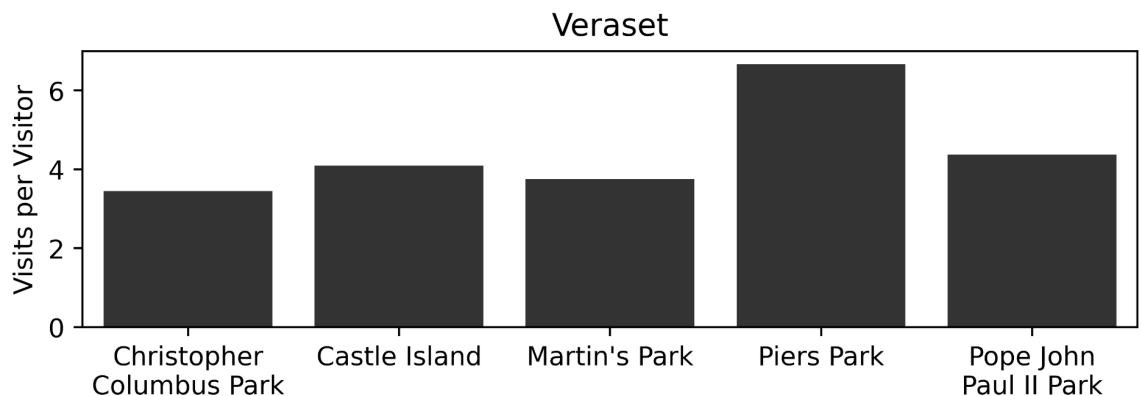
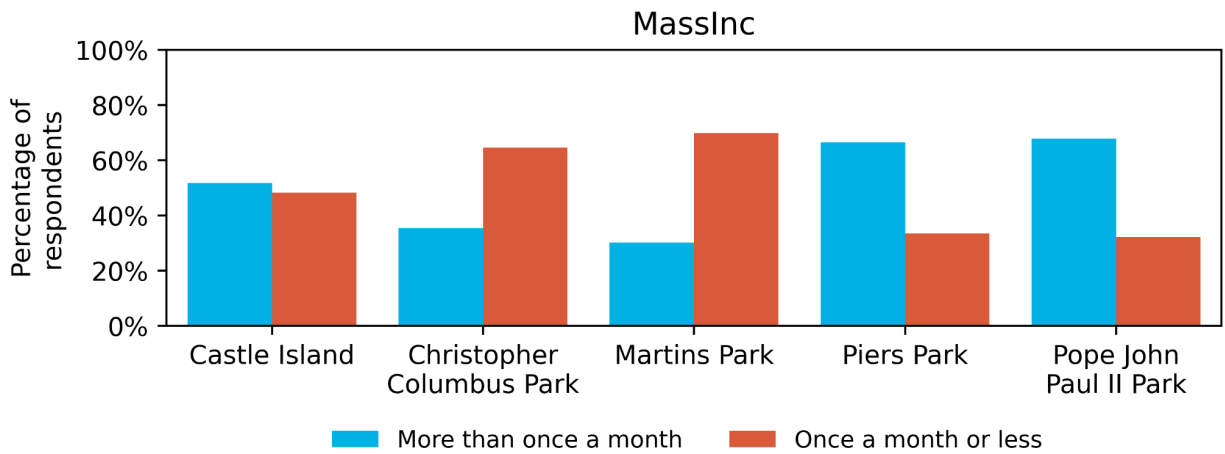
The records I exported from Tectonix included the number of visits and the number of visitors to each park. For most of the analyses below, visits were used to compare to the MassInc foot traffic and survey data. Analysis was conducted in a Python notebook, which you can access [here](#).

Overall visitation by weekday and weekend



In both datasets, Christopher Columbus Park had the highest visitation on both weekdays and weekends, followed by Castle Island. The two datasets did not align, however, with the remaining parks. In MassInc, the three parks had similar numbers of visits, with Pope John Paul II Park being the lowest, followed by Piers Park, and then Martin's Park. Veraset, on the other hand, showed the opposite trend, with Martin's Park having the fewest visits, followed by Piers Park, and then Pope John Paul II Park. Almost all parks had more visits on weekend days than on weekdays. The only exception was in Pope John Paul II Park, where there was higher weekday than weekend visitation.

Repeat visitation



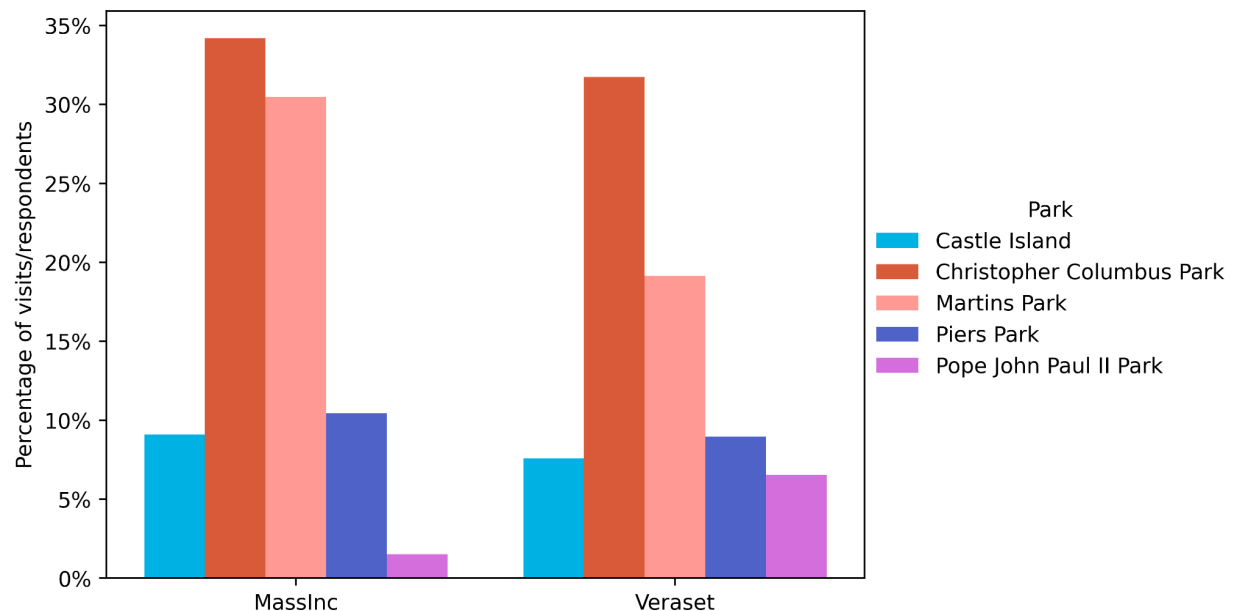
For MassInc, repeat visitation was inferred directly from the survey data. Respondents were grouped by how often they said they came to each park. I further aggregated these responses into two groups: respondents coming more than once a month and respondents coming once a month or less often. With Veraset, a proxy used for repeat visitation was the number of visits per visitor. For instance, four visits per visitor implies that each visitor came to that park on average four times.

The two datasets had some agreement when considering repeat visitations. They did not align on which parks have the highest and lowest repeat visitation: MassInc had Pope John Paul II Park with the highest percentage of repeat visitors and Martin's Park with the lowest, while Veraset had Piers Park with by far the most repeat visitors and Christopher Columbus Park with the lowest. However, in both datasets, Piers Park and Pope John Paul II Park were the two most revisited parks, and Christopher Columbus Park and Martin's Park were the least revisited, with Castle Island being in the middle.

Home location

Veraset records have associated census block information to help determine the home location of each visit. MassInc respondents provided a zipcode for their home location, which I used to determine their home state and city. Any record or respondent in either dataset with Boston as a home city was also assigned a home neighborhood. For the Veraset home location data, the data was taken from the entirety of 2021, instead of only August 2021, as was the case for all previous Veraset groundtruthing data. For the following analyses, the word visit will be used to refer to both Veraset visits and MassInc respondents, though respondent may be used to differentiate between the two.

Out-of-state visitation

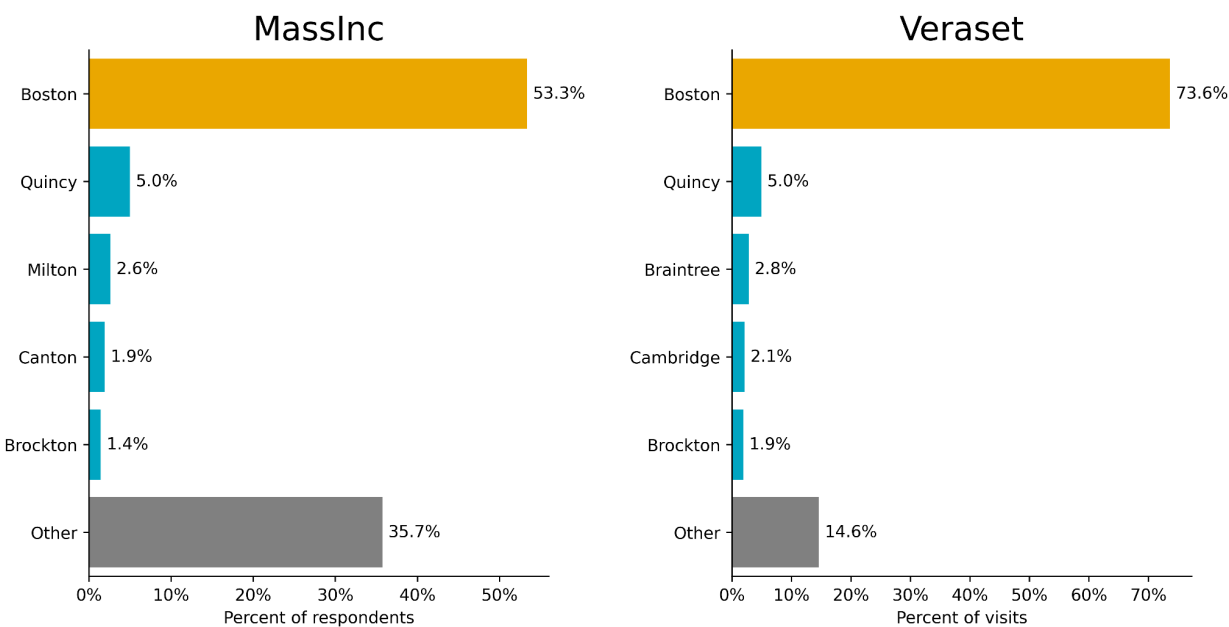


The two datasets strongly aligned on the percentage of out-of-state respondents and visits. All parks had a minority of their visits coming from out of state. In both datasets, Christopher Columbus Park had the highest percentage of out-of-state visits (> 30%), followed by Martin's Park, Piers Park, Castle Island, and lastly Pope John Paul II Park. The largest differences between the datasets were with Martin's Park, where the out-of-state percentage in MassInc was much higher than in Veraset, and with Pope John Paul II Park, where the opposite was the case.

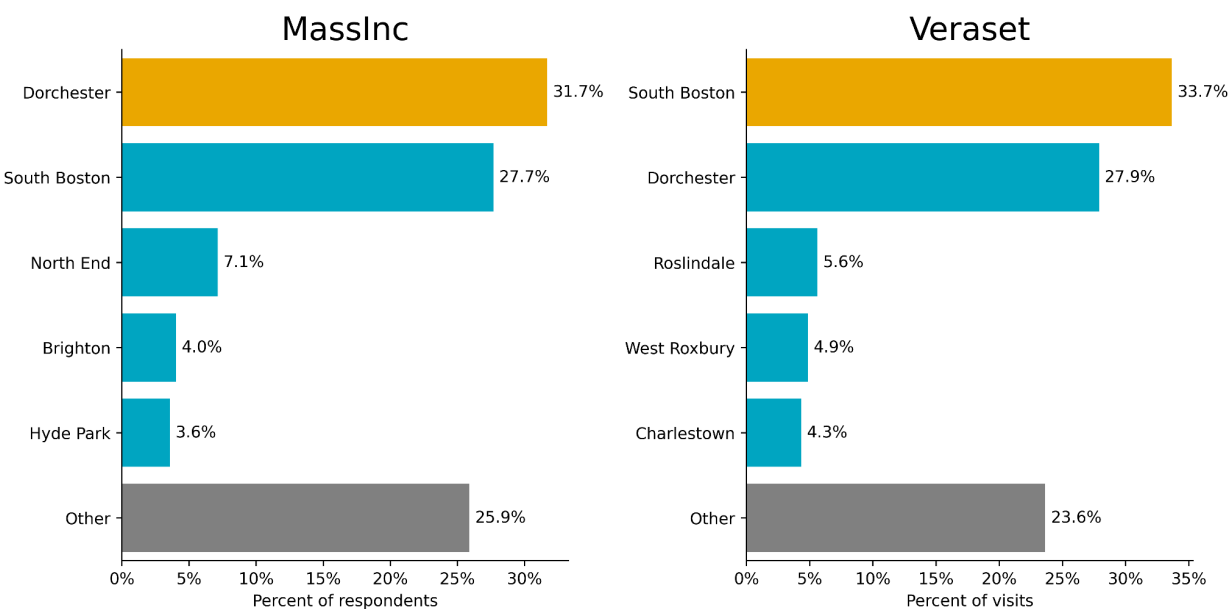
Visit home cities and neighborhoods

Castle Island

City



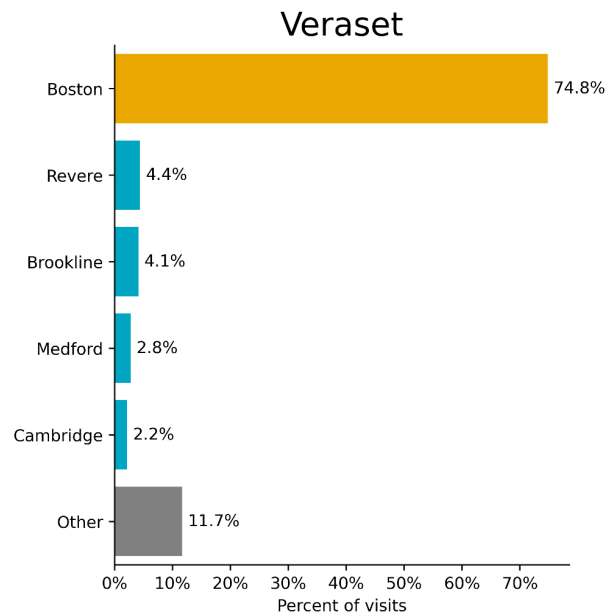
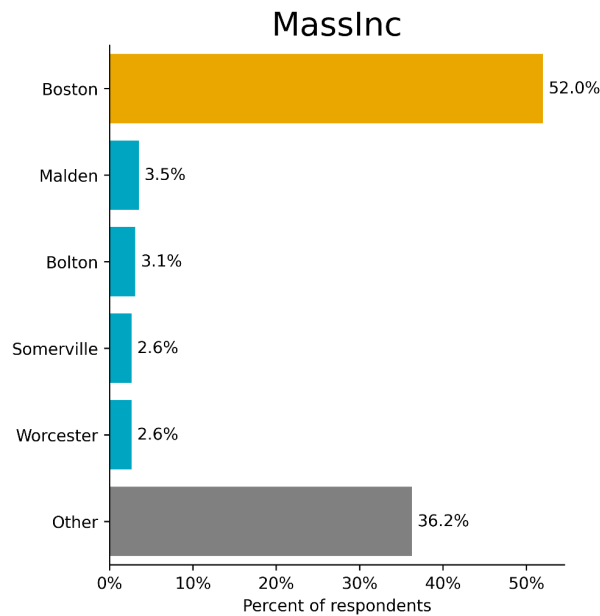
Neighborhood



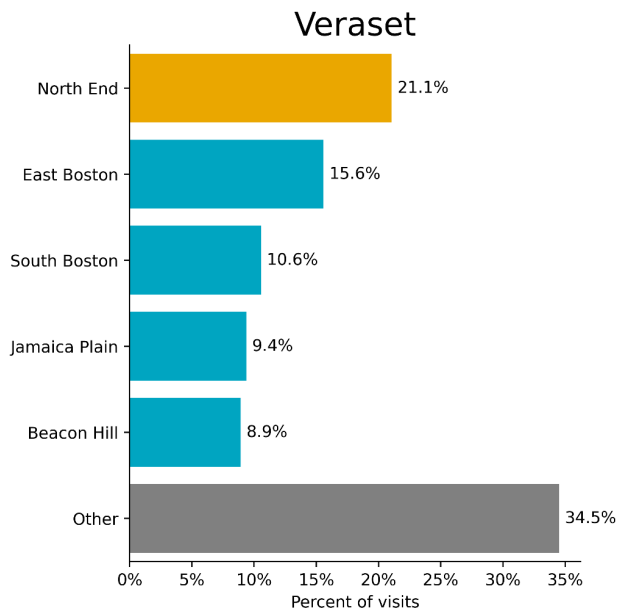
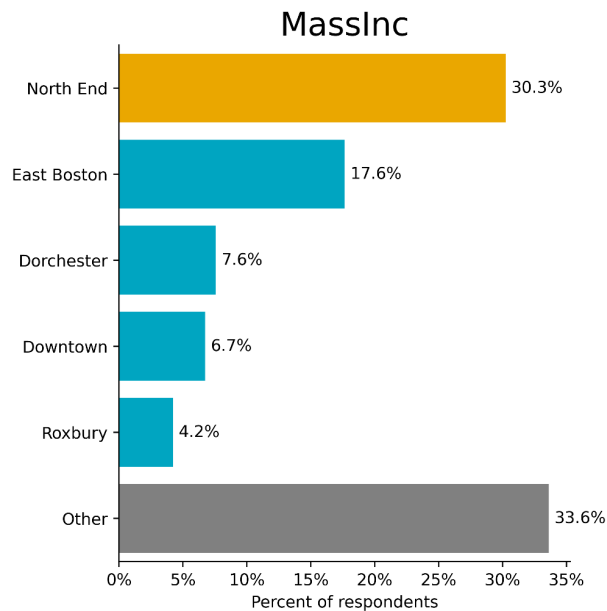
Both datasets showed that Castle Island visitors primarily come from Boston, and those who came from Boston are primarily from Dorchester and South Boston. However, in the MassInc dataset, Dorchester was the home neighborhood of more respondents than South Boston, while the opposite was true of Veraset visits. Both showed that Quincy was the second most common home city. For both city and neighborhood, the “Other” category was more common with MassInc than with Veraset.

Christopher Columbus Park

City



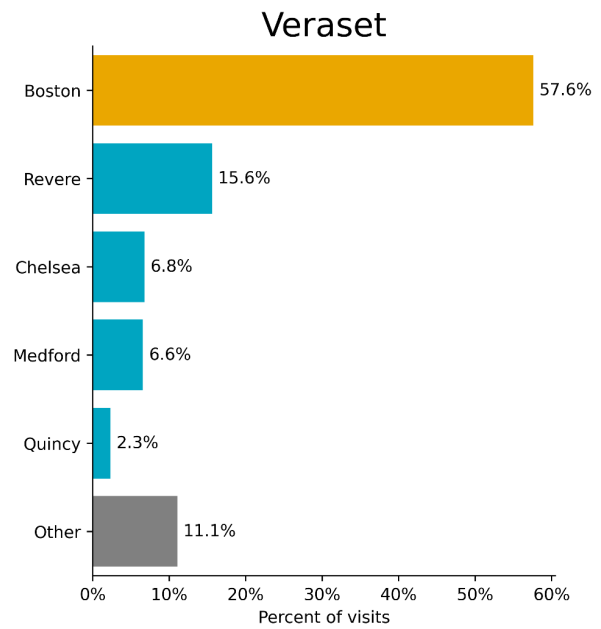
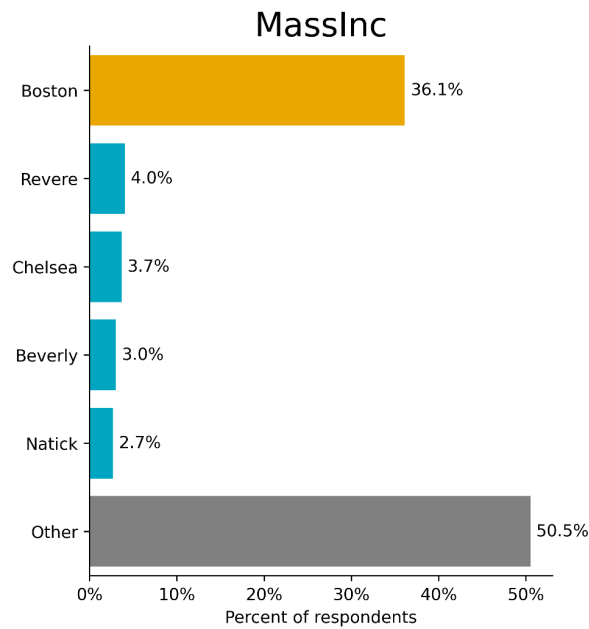
Neighborhood



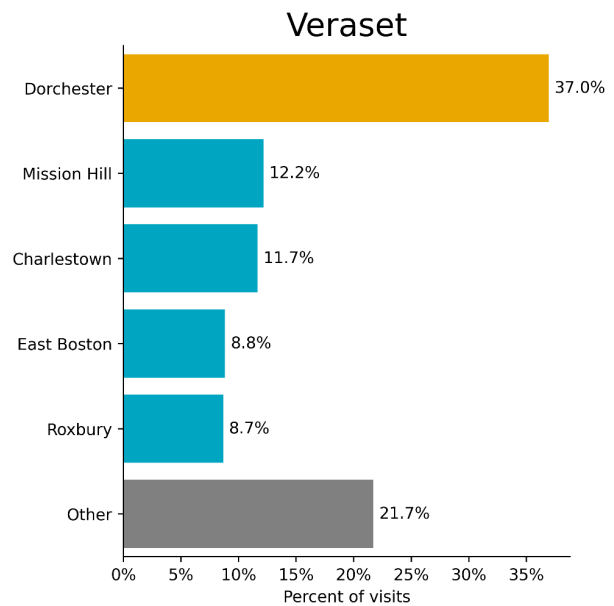
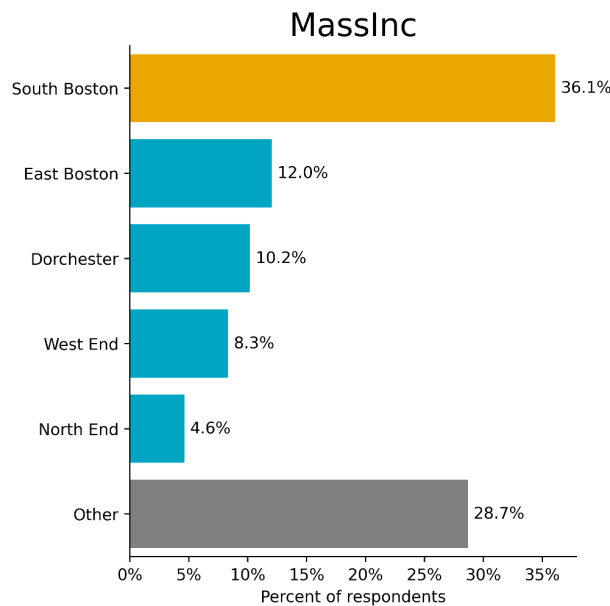
Both datasets showed that primarily, visits to Christopher Columbus Parks came from Boston, and Boston visits were most commonly from the North End and East Boston, though many other home neighborhoods were common.

Martin's Park

City



Neighborhood

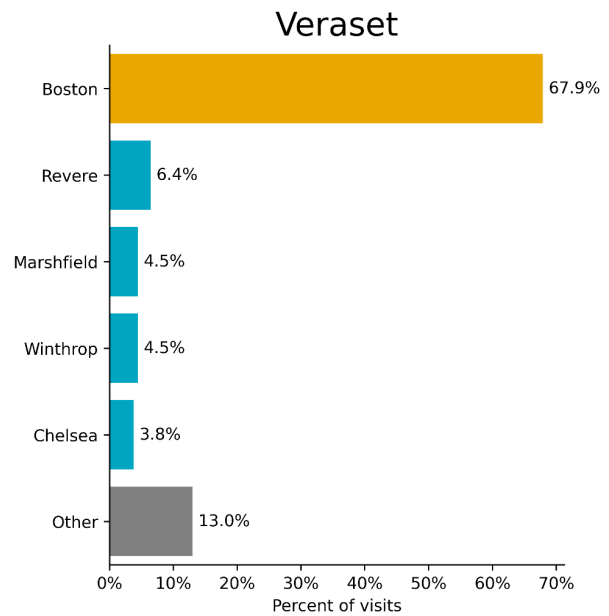
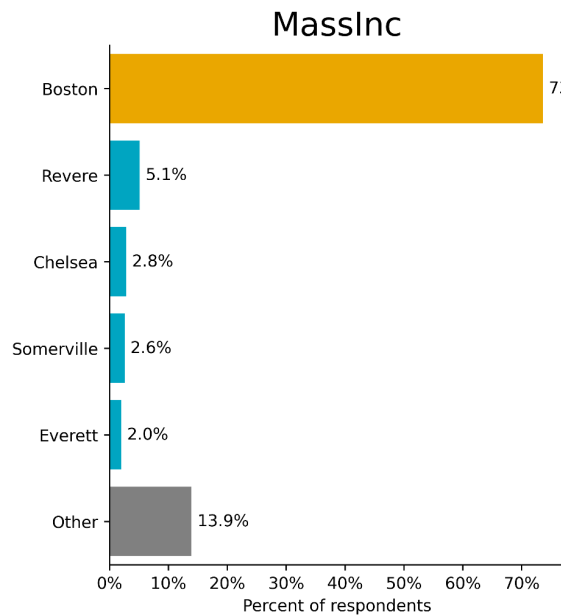


For Martin's Park, MassInc and Veraset both showed that Boston was the most common home city, but less than 50% of the MassInc respondents came from Boston, which was not the case for Veraset visits. Both showed that Revere and Chelsea were the home cities for many visits. MassInc visits came from a wider variety of cities than did Veraset, with over half of all visits coming from a city that was not in the top five most common. In terms of the home

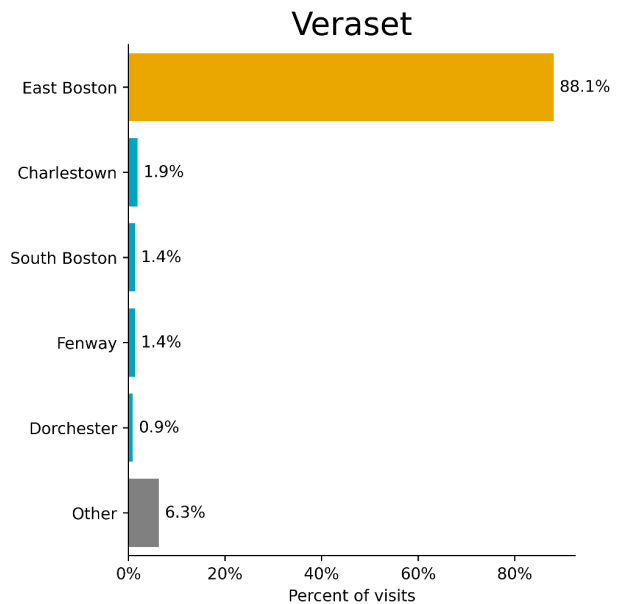
neighborhood of Boston visits, the two datasets drastically differed, with South Boston being the most common for MassInc and Dorchester being the most common for Veraset, and East Boston being the only commonality in the top five between the datasets.

Piers Park

City



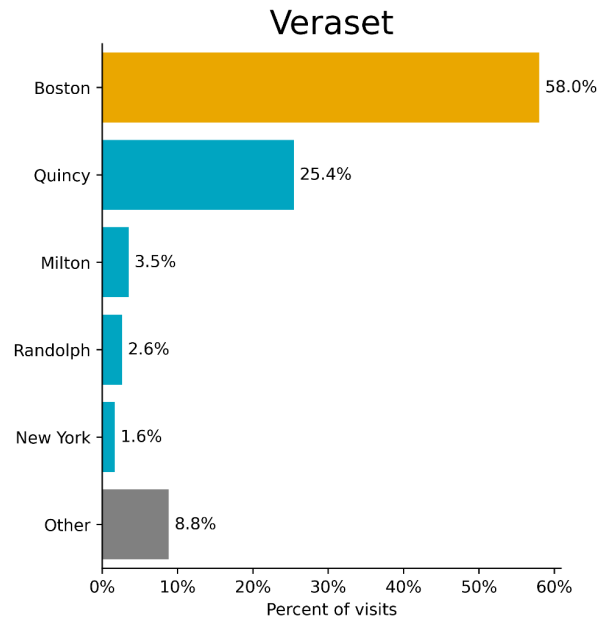
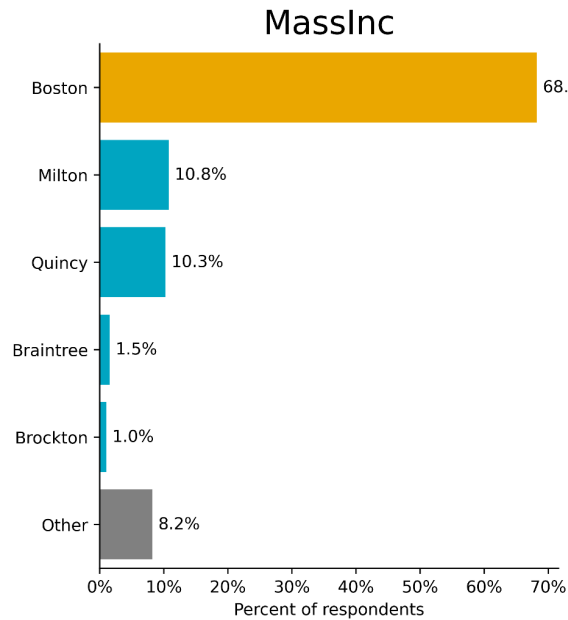
Neighborhood



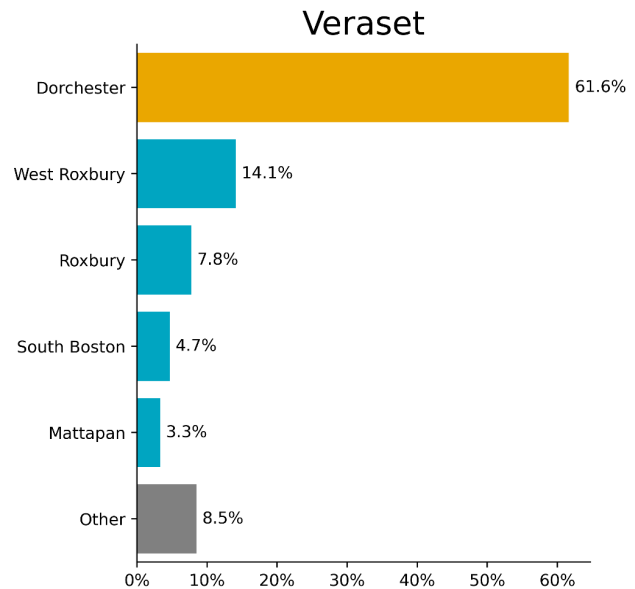
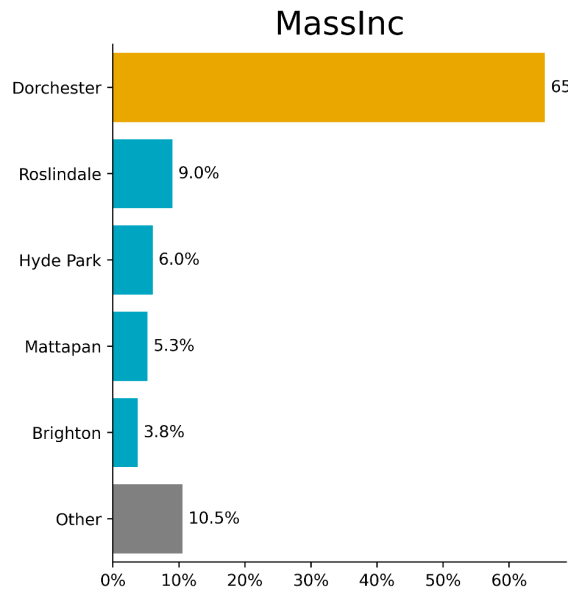
With Piers Park, both datasets were highly aligned, with Boston being the most common home city, followed by Revere, and with the overwhelming majority of Boston visits coming from East Boston. Both also showed that Dorchester and South Boston were other common home neighborhoods.

Pope John Paul II Park

City



Neighborhood



The two datasets showed moderate agreement when it came to the home location of Pope John Paul II Park. Boston again was the home city of most visits. Quincy and Milton were the next most common home cities, but Veraset showed that Quincy was much more common than Milton, which was not the case with MassInc. In terms of home neighborhoods of Boston visits,

both showed that Dorchester was the most common home neighborhood, but with little in common apart from that, with Mattapan being the only other shared top five neighborhood.

Inter-rater reliability analysis

To quantify the overall degree of agreement between the Veraset and MassInc data, I conducted an inter-rater reliability (IRR) analysis. This analysis consisted of posing hypotheses for each dataset to respond to, such as “The home city of most visits in all parks is Boston”. The full hypothesis data, including a list of all hypotheses, can be found [here](#). Each hypothesis was determined to be true or false for the two datasets. This allowed me to calculate two metrics summarizing their agreement: percent agreement and Cohen’s kappa. Percent agreement is simply the percentage of hypotheses where both datasets provide the same answer. Cohen’s kappa is a metric that is commonly used in IRR that considers how likely the percent agreement is given the total number of true and false responses. For more information on Cohen’s kappa, please see [this article](#).

The percent agreement between the two datasets on the hypotheses was 61.0%. This is a moderate degree of agreement between the two. Though not a high degree of agreement, both datasets have areas where they strongly align, as well as other areas with moderate agreement, and others with limited agreement. This is not surprising, given that both mobility data and survey data have inherent biases. Additionally, the datasets were not collected in the same years, making even a moderate degree of agreement impressive.

Cohen’s kappa was 0.076 for the two datasets. This value indicates minimal agreement between the datasets. However, Cohen’s kappa is a rather strict metric that generally assumes that random agreement is common, so therefore a high degree of agreement is necessary to generate a large kappa value. Therefore, Cohen’s kappa is included here for completeness, but percent agreement can be used as the primary statistic for this IRR.