Tiffany Lo

CS591 – Data Mechanics

May 4, 2016

# Final Report

Although Boston's relatively small size, walkability, bike friendliness, and convenient public transportation system provides for a myriad of means to get to a food establishment without a car, parking availability is a major factor for many when deciding where to eat. Street meter parking is a key source of parking for many businesses, so assessing the number of parking meters and the frequency of parking tickets received in proximity to a food establishment is a helpful metric to gauge the parking availability of a location. Another interesting consideration is to assess the strength of the effect local food establishments may have on an area's parking violations.

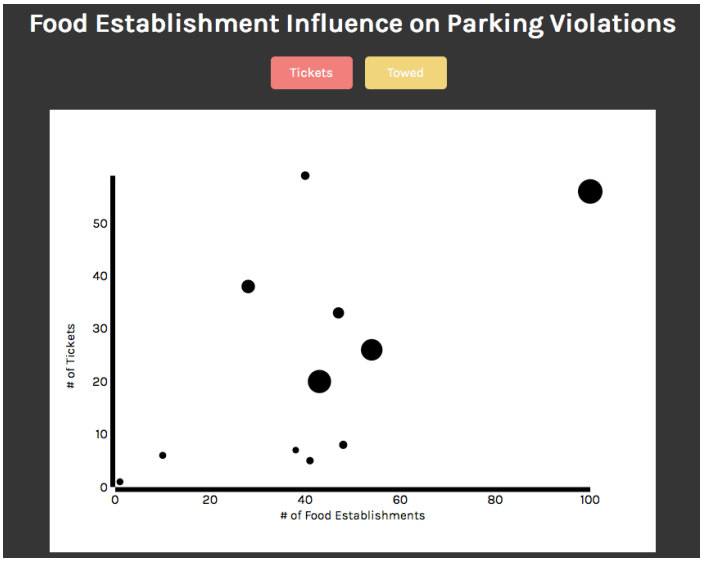
A total of five datasets were used in this project. The City of Boston data portal provided the "Active Food Establishments," "Crime Incidents (INCIDENT_TYPE_DESCRIPTION=towed)," and "Parking Tickets" datasets. These were retrieved using the Socrata API. As for the parking meters dataset, it was a GeoJSON file found on the BostonOpenData ARCGIS site. The fifth dataset also came from the City of Boston data portal, however it was a downloadable set of shapefiles delineating Boston neighborhoods. These shapefiles had to be converted to GeoJSON before they were usable in Leaflet.js visualization applications. The four primary datasets had varying fields describing their location and for the tickets and tow incidents, time of occurrence. In order to standardize these common fields for aggregation, three-letter street address acronyms were standardized to two-letter forms, and then used to request full address information from either the OpenStreetMap Nomatim or Google V3 API. All of Boston's zipcodes begin with zero, which is omitted when the zipcode is treated as an integer rather than a string. Prepending zipcode results that are four digits long rather than five before updating the field in the database ensures that the

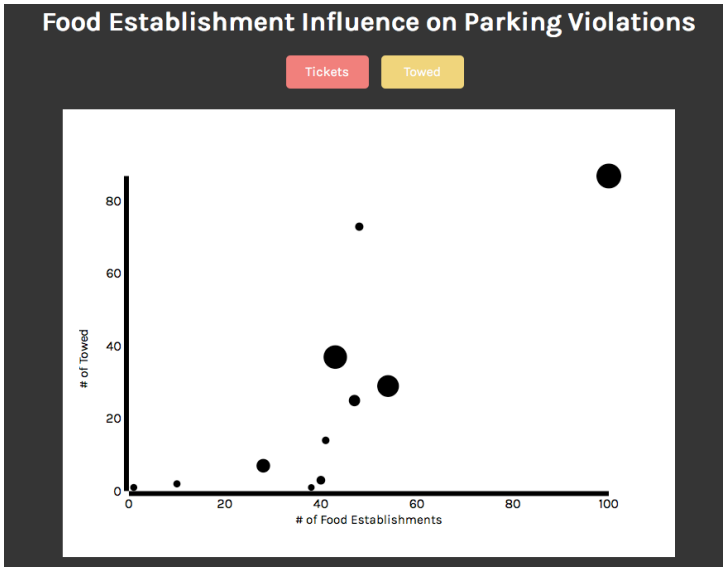
zipcodes are all consistent. Longitude and latitude coordinates were stored as well for records without that information already.

Once the data was cleaned and all the necessary fields filled, various forms of aggregation based on the day and time of the week an incident occurred, location of meter/incidence subdivided by zipcode, and available meters based on a 0.7 mile walking radius. The process of aggregation required further manipulation of the data points and identified large disparities in in how location, time, and breadth of data contribute to each dataset.

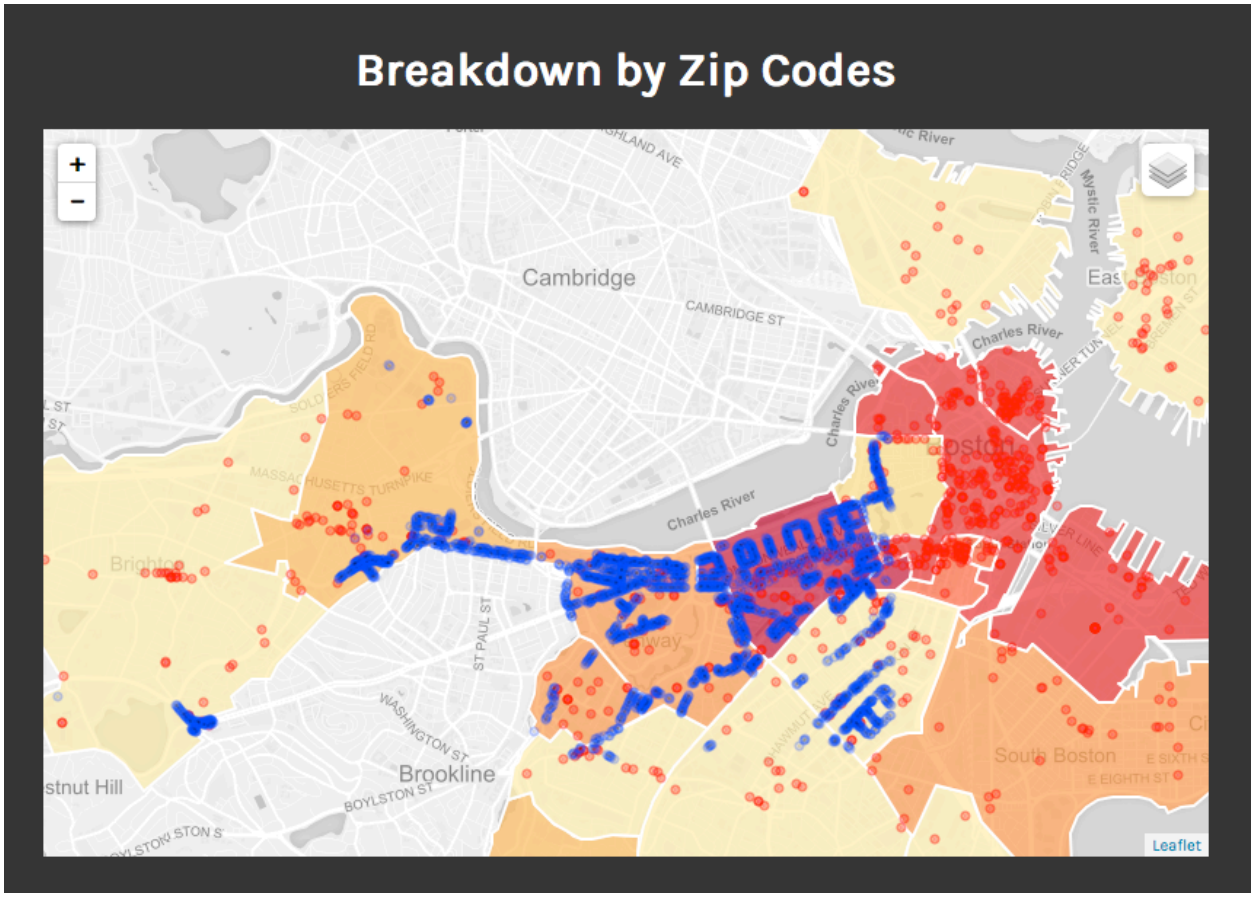
Regardless, these result outputs were put to use in further calculations, analysis, and visualizations. Calculating p-values for the null hypotheses that a) More food establishments in an area would not lead to more tickets given in an area, b) More food establishments in an area would not lead to more tow violations happening in an area. For null hypothesis a), given that the p-value was 0.00, there is strong evidence that disproves it, meaning that more food establishments in an area does not lessen the number of tickets given. The disproving of null hypothesis a) does not conflict with the common sense logic that zip codes with more food establishments are likely more busy and therefore have more parking activity that may lead to tickets. For hypothesis b), given that the p-value was 0.63, there is strong evidence that proves it, meaning more food establishments in an area would not lead to more tow violations happening. The proving of null hypothesis b) encourages further investigation into why cars are being towed and how busy areas with many more food establishments do not see more tow violations than areas with fewer. The visualization below compares number of food establishments to number of tickets (left) and tow violations (right) with the added dimension of circle radius based on number of parking meters in the zip code area. Within the range of 40-60 food establishments, a varied number of tickets and tow violations can be seen in both of the graphs. The few points on the graph make it difficult to identify strong trends as

there are only a handful of zip codes within the City of Boston, though the assumption that areas with many restaurants are busy in terms of car traffic as well seems to stand.



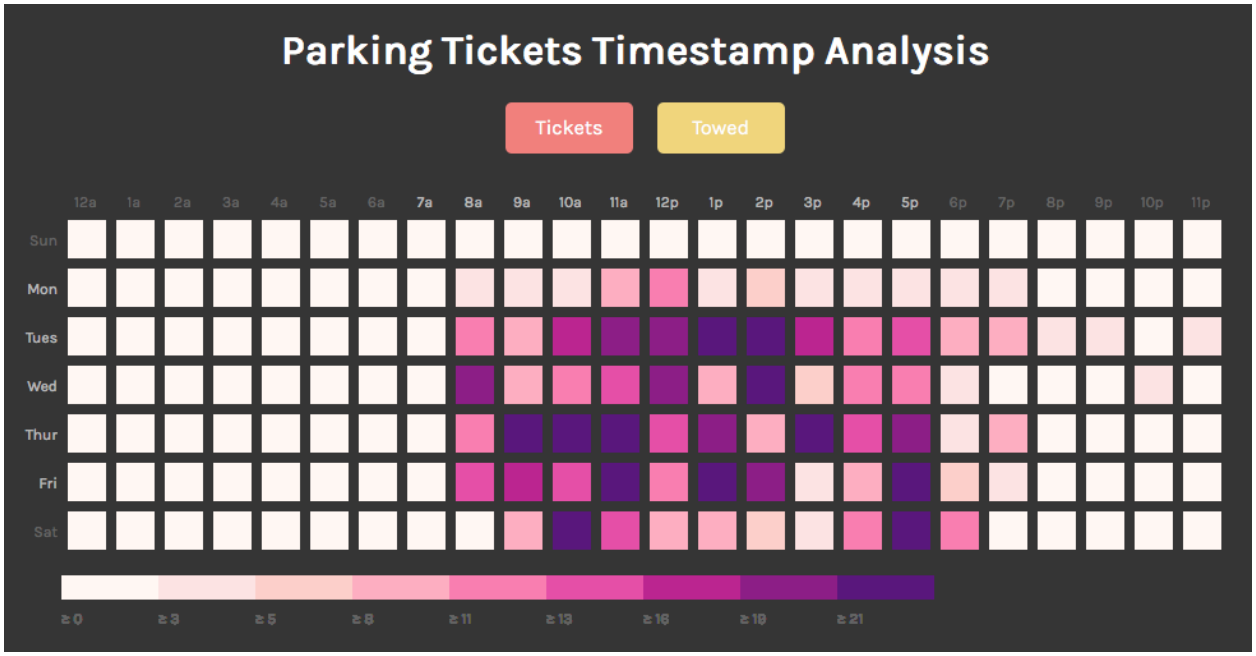

Mapping the zip code areas within the city with meters (blue) and food establishments (red) provided a means to better understand the scope of the datasets. Although the Parking Meters dataset was the largest, it is only highly concentrated in several areas and missing in many of the



outer zip code areas. A majority of the street meter parking spots line down Commonwealth Avenue and other main streets, but the busy downtown areas have no recorded meters. This peculiarity certainly raises concerns about the completeness of the dataset. Areas vary from yellow to red based on their tickets and tow violation scores. The score is summed based on their occurrence as normalized by the areas with the most tickets and towing violations. Red regions are arguably hubs within the city of Boston and contain many food establishments as well. This aligns with our disproving the null hypothesis a) in the earlier calculations.

Another angle to study the relationship between food establishments and parking violations was seeing if there was an increase in tickets or tow violations during peak meal times when patrons would visit a food establishment. As with previous analyses, the confounding variable of meal times coinciding with peak traffic hours and the shifts for serving parking tickets should be factored in this analysis. The heatmap below illustrates that parking tickets are given during the workday from Monday to Saturday, with high concentrations on Thursday and Friday around 9-11am, Tuesday and Wednesday from 12-2pm, and Friday and Saturday at 5pm.

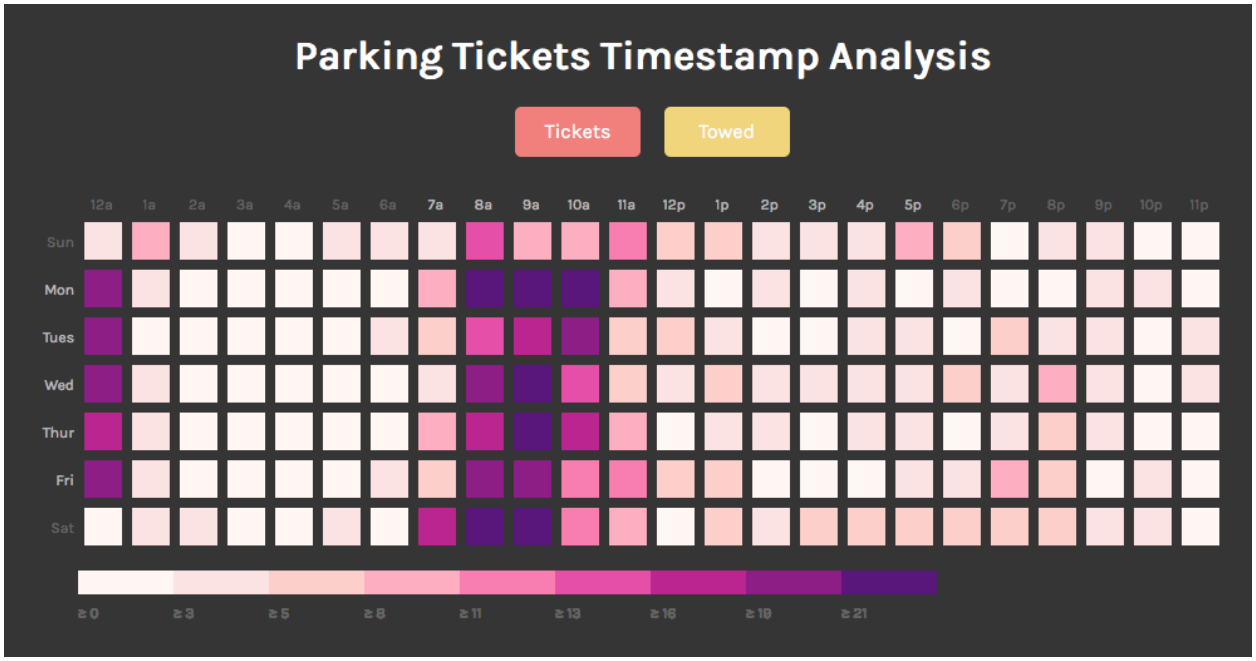
Tickets

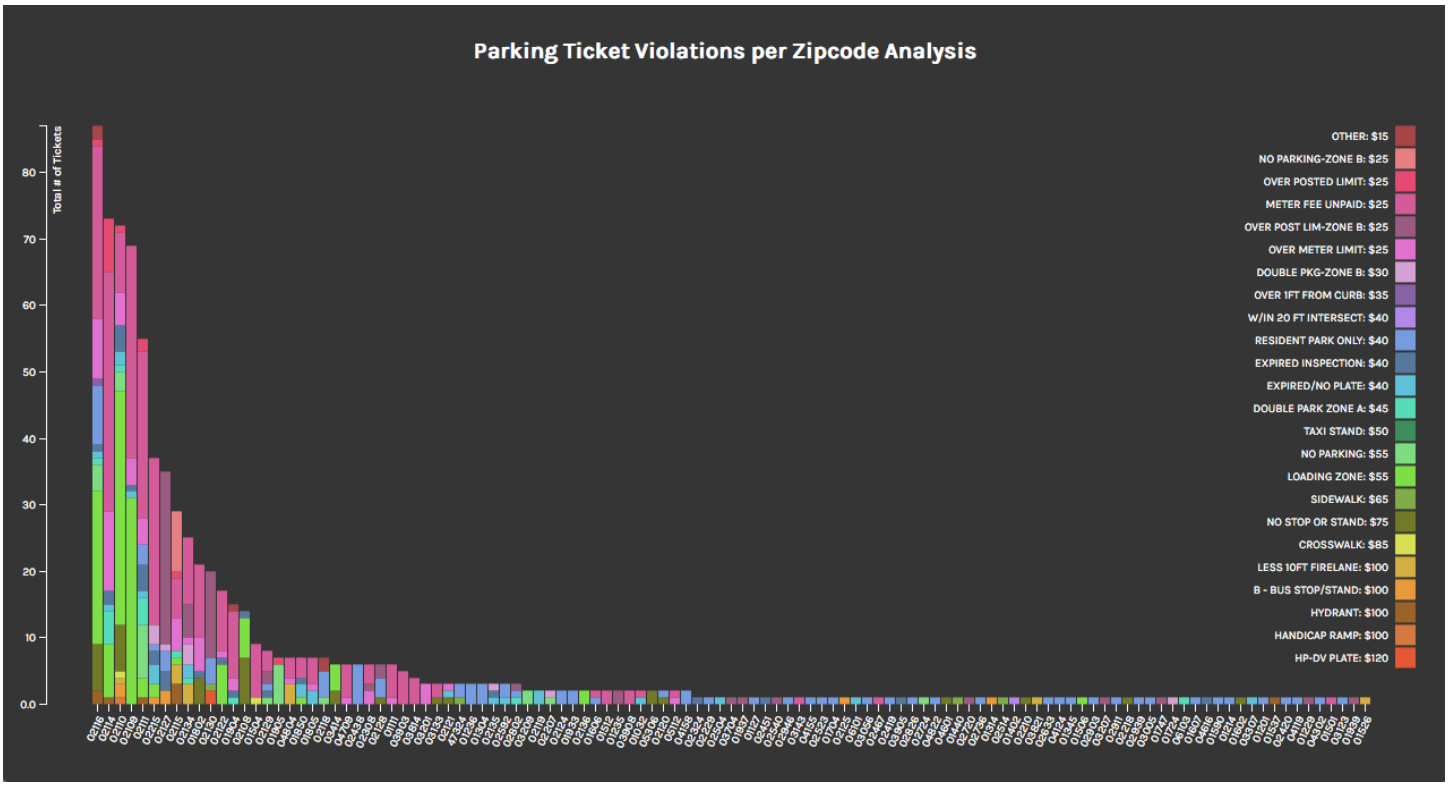


Arguably, those times of highest ticket frequency are popular mealtimes, however if there was a strong trend for parking violations at mealtimes due to patrons visiting food establishments, the trends would be more consistent throughout the entirety of the week.

The relationship between mealtime hours and tow violations is even weaker than that of parking tickets. It seems that cars are towed at consistently the same times in the day, mainly in the morning hours. Since it is unlikely that lots of cars are towed every night at 12am, the dark shaded squares at 12am from Monday and Friday most likely reflect that many timestamps for tow violations were taken as merely datestamps in the system. The lack of time information defaults the datetime object to 00:00:00 or 12am.

Towed






Based on learning that tows happen on a regular schedule and that as the null hypothesis b) found has little relation with food establishment prevalence, the next step was to focus closer on how parking tickets are distributed. Rather than looking at parking tickets per zip code area at a macro scale, the graph above breaks down the number of tickets in an area according to the type of violation. The legend and bars organize the violations based on their fine amounts as a barometer for the seriousness of the violations. Most tickets are either minor or medium level offences, though it is important to note that "Meter Fee Unpaid" and "Loading Zone" bars are the most prominent of most zip codes. These are both common parking infractions, though they may be related to restaurant patron parking as well. Another interesting note from this is the distribution of tickets concentrating in downtown areas, but also having a tail of one-ticket zip codes that simply indicates errors in the geocoding to zip code process.

To answer the more specific goal of this project, a metric for food establishments was created using number of meters and tickets in its proximity. It was challenging to determine a reasonable walking distance radius that could distinguish the closely clustered food establishments and parking meters whilst reflecting real-life expectations of convenience when parking on the street for a meal. Through experiments with distances from 0.3 to 1 mile to study the range of scores food establishments in differing areas received, 0.4 miles was found as the ideal range that went beyond parking right outside of an establishment while keeping it within grouping most food establishments with a cluster of meters. Using MongoDB's geospatial indexing, the queries to find surrounding meters and tickets was very expedient to complete for the 1,000 food establishments in the dataset. Scores still ranged wildly and could be refined, but the system was simple, assigning one point for each meter and deducting 0.3 points for each ticket violation. Meters are single occurrences that benefit the restaurant's parking availability and tickets are repeated occurrences that build up to indicate a parking concern for the restaurant. This metric assigned scores ranging from the top scorer, McDonald's at Kenmore Square with 585.1 points and the lowest scorers, food establishments in the Chinatown and outer edges of Downtown with -66.6 points. The aforementioned clusters of parking meters in the Back Bay and Fenway area along with the lack of meters in the downtown area in the dataset are evident in these disparate scores.

Overall, the findings reflect a basic understanding that areas with many restaurants are areas where many parking tickets and tows are issued as well. This can be explained due to the overall increased traffic in busy areas downtown, especially with the observation that there are no strong direct links between the food establishments themselves and the parking violations. Concerns regarding the incompleteness of the datasets' coverage of the city, need for controlling the data points against traffic, population, land area, and system schedules, errors in keying timestamps and location coordinates, and difficulty of classifying clusters within such a small area hindered the

analysis process significantly. Very little can be firmly concluded as a result of these factors. The inquiry into how parking resources are serving the public and businesses is very important and as the demographics of Boston areas continue evolving, areas that have minimal parking availability but increasing number of food establishments may require that additional parking be added to the city. As an extension or refinement of this project, the overall accessibility of a restaurant including MBTA stop data could be interesting and improve the score of downtown eateries. If there was a good method to scrape information from Yelp regarding food establishments' designated parking, that information could vastly improve the scoring as well.