Emily Brown
Service Learning Response 1
Big Data for Cities
October 20, 2021

Both the city walk and the in-class workshop influenced how I have been thinking about and planning to analyze my data. Prior to the city walk, I thought that the status of a violation (whether it was "open" or "closed") was an indication of the current status of the property. For my virtual city walk, I decided to explore the Inspectional Services Neighborhood Statistical Area with the most violations marked "open". Within this area, I explored the street with the most open violations, which was Blue Hill Avenue. Along this street, there were many properties with serious violations with an "open" status, including violations with the description "unsafe and dangerous". There were also many properties with several "open" violations for things like improper storage of trash and overgrown weeds. I expected to find that properties with "open" violations still had the problems persisting. However, this was not the case. Many of the properties with "open" violations seem to have remedied the problem, at least at the time of the Google street view images. Several of the properties also changed hands and were renovated. I also found that there were several properties along Blue Hill Avenue that appeared to have serious issues such as broken and boarded up windows and other visible, external damage. However, these properties did not appear anywhere in the dataset.

From the city walk, I learned that the amount of "open" violations did not necessarily indicate the current status of the property. I also learned that the length of "open" violations did not either, and that there may be real "problem properties" that are missing from the dataset. This made me wonder what the driver for neighbors or tenants reporting "problem properties" was, if properties with visible, external issues were not reported. It also made me wonder what the best way to gain insights to the current condition of a property or area is from the data. I started to

think that the best way to measure a property with chronic issues may not be whether there is a violation marked "open" but whether there are repeated violations over time for the same issue. To test this theory, myself and my group mates looked at the property in the city with the most violations, regardless of status, and found a property with over 800 violations. Most of these violations were for "overgrown weeds". From Google street view, we could actually see an image from a few years ago with an incredible amount of overgrown weeds, which were cut back when the property changed hands. However, this is just one example, so I'd like to do further analysis and observation to see whether the total number of violations is a better indication than the amount or length of open violations.

As a result of the in-class workshop, I began to think more about what types of other datasets from the City of Boston I could join with the Violations dataset to learn more about the properties, areas, and trends within the city. In order to learn about landlords that have the most properties with chronic problems, and potentially learn about absentee landlords, I had the idea of joining information on property owners to the dataset. Upon talking to the workshop visitors, they thought this was an interesting idea and suggested joining the assessing information from the City of Boston. I was able to do this during the workshop and began exploring trends in the data based on property owners. I also began thinking about how to incorporate COVID-19 data into the dataset, and during the workshop, created a new variable for whether a violation was reported before or after the first COVID-19 case in the city. I hope to be able to join the violations that occurred after the first COVID-19 case in the city with other COVID-19 data to observe and analyze trends.

I hope to be able to join the assessing and violations data with datasets from other organizations, including from places like the Metropolitan Area Planning Council as suggested

by the workshop visitors. From conversations during the workshop, I also began to think about how properties may be intentionally neglected until an owner can sell or renovate a property. I knew from previous research that property owners often let a property deteriorate in hopes that tenants will leave or that they can evict tenants in order to renovate and increase rental prices or renovate and sell the property. I wondered whether a decline in violations at a particular location, especially following a change in the owner, could be an indicator of gentrification.

Overall, both the city walk and the in-class workshop made me think more deeply about what latent constructs I want to explore and which variables are the best indicators of these latent constructs. It also made me think about which external datasets I could join with the violations dataset to do the most useful and interesting analysis. I look forward to doing more data analysis, comparing my findings to another city walk, and talking them over with future workshop guests.