

# *The Rent Is Too Damn Loud*

*Martin Schiff*

2019-02-24

*Ain't nothing goin' on now  
But the rent, ah*

— James Brown

*Pretty girls, pretty boys  
Have you ever heard your mommy shout  
Noise annoys*

— Pete Shelley

## *Introduction*

ANYONE THAT HAS LIVED IN New York City (or visited for any length of time) will confirm two of its foundational truths:

1. Housing is costly.
2. It is not, on the whole, a quiet place.

My line of work is acoustics, but on the side I also dabble in rent anxiety and in listening to my upstairs neighbor, who fancies himself an Electronic Dance Musician. Both keep me up at night, wondering how escalating rent and noise dissatisfaction may interrelate, and if one could be predictive of the other. Are there parts of town that combine affordable rent and low risk of noise aggravation?<sup>1</sup> Can past complaints combine with housing trends to predict when and where noise complaints are most likely?

For this capstone project in the HarvardX Data Science course series, I examine over nine years of New York City noise complaint records. A mashup with neighborhood rent allows comparison on a local level, with sufficient breadth to train a machine learning model for projection of future complaints.

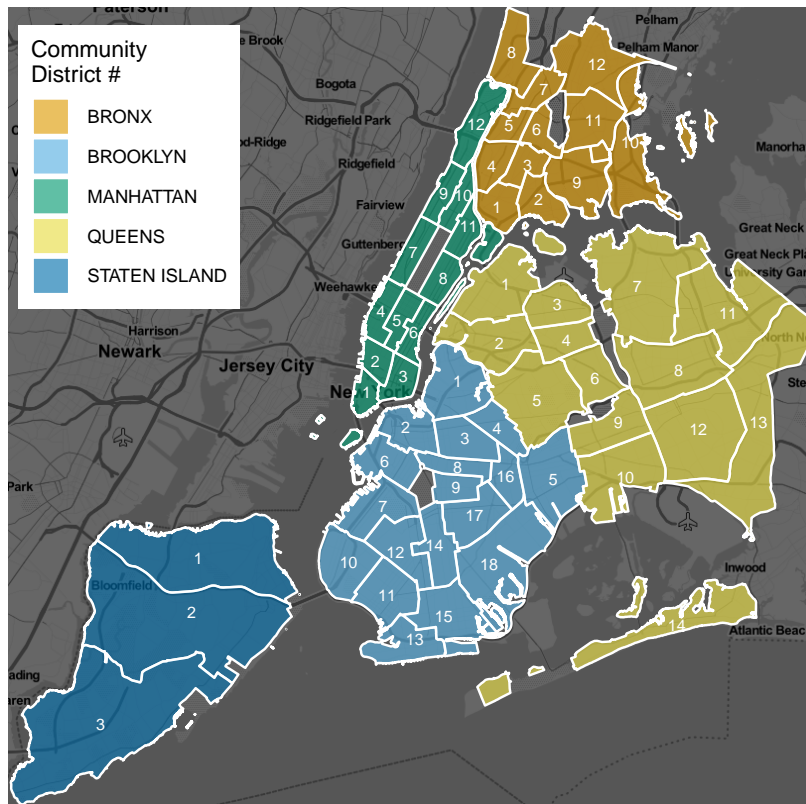
<sup>1</sup> This, the holy grail of the working acoustician.

## *Data Overview*

New York City has a 311 telephone hotline for access to city services, enhanced in recent years with web and app portals. One of its many

functions is to collect and assign noise complaints to city agencies for enforcement.<sup>2</sup> As part of its Open Data initiative, the city provides access to the 311 database from 2010 to the present, updated nightly. Each call (or online submission) is logged with the time, nature of the complaint, and geovalidated location information on the site of the issue.

New York City's 59 community districts are the smallest political/administrative units in the city, each comprising a few neighborhoods in one of the five boroughs. This analysis also makes use of the 2010 census population counts for each community district (CD) and their boundary definitions.



<sup>2</sup> Generally, complaints about loud music, parties, and the like are routed to the police, while fixed noise sources (building machinery, nightclubs, etc.) with actionable limits are referred to the city's Department of Environmental Protection.

Figure 1: New York City's 59 community districts.

Finally, real estate database company Streteasy publishes aggregate sales and rental listing data for the New York City market. Monthly median rents back to 2010 are available for most neighborhoods, as well as on a borough- and city-wide basis.

Since local neighborhood boundaries are rather colloquially defined<sup>3</sup>, and are in any case not included with the Streteasy figures, I had to manually match listing neighborhoods to the larger community districts.

These asking rents only circumstantially characterize wider housing costs. They represent what one might pay if forced to move in a given month—perhaps due to noise complaints.

<sup>3</sup> Not least by those in the real estate business, who can be inclined to smear the desirable-neighborhood lipstick onto a pig of an apartment.

### *Related Work*

If the ins and outs of New York City noise complaints are of interest, a few related projects on the same dataset are also worth checking out. Each certainly provided useful background for this study.

- This New Yorker piece examines trends in the 2013-2014 data, mapping down to the narrower census neighborhood tabulation area.
- An interactive project from design firm Carto explores the 2015 data down to the even narrower census tract.
- A 2015 social science study examined relationships between ethnic diversity and the likelihood of disputes, finding that appeals to authority may be more likely in areas with “fuzzy” racial boundaries.

### *Methods & Analysis*

FOR THIS ANALYSIS, I reviewed the 311 database and rent data separately, using insights from each to guide their combination.

#### *Noise Complaints, Individually*

To gather the 311 data, I used the Socrata Open Data API via R package *Rsocrata* to access all records with a permutation of “noise” in the complaint, discarding records about, for instance, malfunctioning street lights or poor taxi service.<sup>4</sup> This still yields nearly three million records, so I limited the fields to the timestamp, description, latitude/longitude, community district, and the channel used (phone/app/online), discarding entries that lacked usable or meaningful location.

<sup>4</sup> The city collects these complaints because the taxi dispatch also tends to discard them.

Before reading too much into this, it’s important to recognize what these complaints can tell us and what they can’t. The location information represents the target of the complaint, not who lodged the complaint or where they were located, which is not public. There is also not necessarily any connection to actual noise level, and in fact even the number of complaints may not necessarily indicate severity, since one unhappy person can lodge dozens or hundreds of submissions.

As an example, a heatmap comparison of complaint density between 2015 and 2018 appears to reveal a significant new hot spot near St. John’s University<sup>5</sup> in Queens community district 8. Is there some new (or newly-significant) source of noise that has riled up the neighbors?

<sup>5</sup> It almost looks like... a Red Storm.

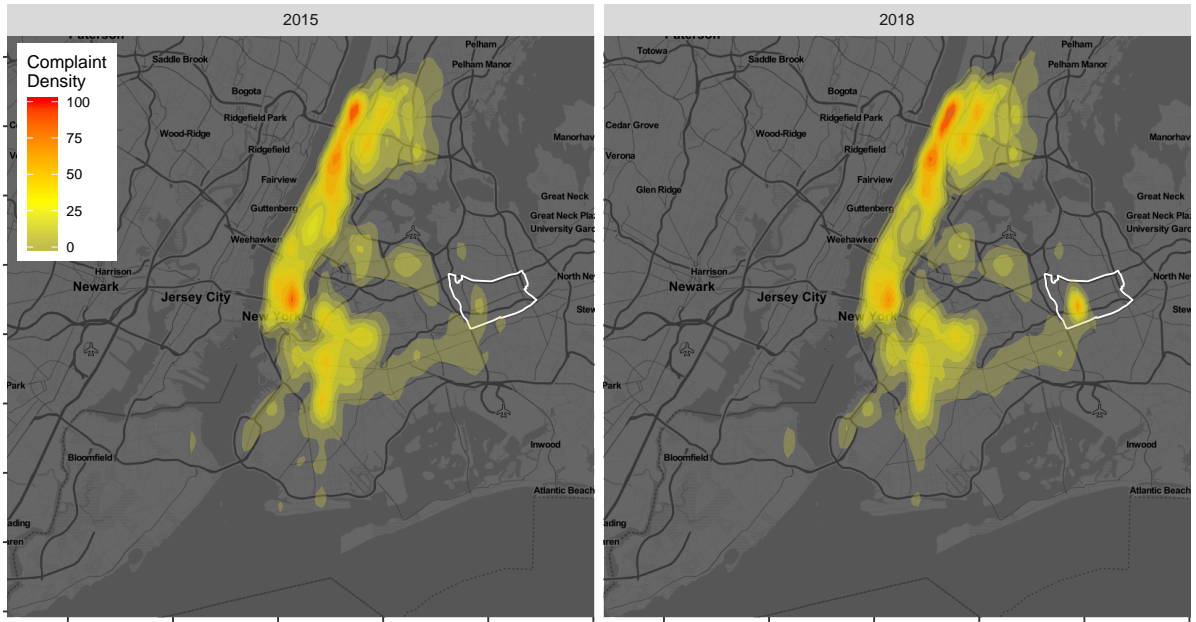


Figure 2: Complaint density in 2015 and 2018, highlighting Queens CD8.

Drilling down inside this district, 40 percent of the 2018 complaints were about only two locations...

Table 1: The top two complaint targets in Queens CD8, 2018.

| Longitude | Latitude | Complaints | % of Total |
|-----------|----------|------------|------------|
| -73.8038  | 40.7187  | 3243       | 28         |
| -73.8078  | 40.7203  | 1382       | 12         |

...and the complaints on those two locations had overwhelmingly to do with alarms and horns.

| Descriptor          | Complaints |
|---------------------|------------|
| Car/Truck Horn      | 2773       |
| Noise: Alarms (NR3) | 1840       |
| Engine Idling       | 11         |
| Car/Truck Music     | 1          |

The first location, it turns out, is the EMS ambulance garage at Queens General Hospital. The second is an FDNY firehouse. Both were active long before 2018, or 2015, or in the case of the firehouse, 1928, so it's possible that one or more people simply started having a bigger problem with the sirens.

A common trajectory: the noise was always there, but new neighbors have different expectations. There is unfortunately no way to extract whether this example describes new neighbors or old ones—nor many people complaining a little or one person complaining a lot.

### Noise Complaints, in Aggregate

To advance the analysis, I folded the data into a total number of noise complaints for each month in each community district. Because the districts vary, I normalized a complaint rate per 1000 residents based on the 2010 census.

For a bigger picture, we can also look to this population-normalized rate on a borough and citywide basis. The borough-level monthly complaint rate shows three immediate trends: a distinct seasonality within each year across all boroughs, a consistent ordering to complaint rates among the five boroughs, and a general increase in complaints with time.

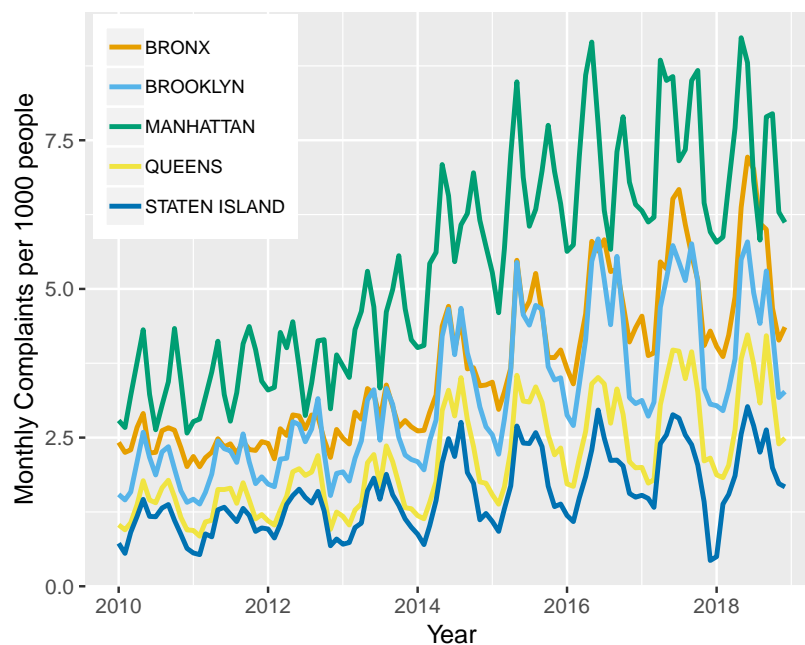


Figure 3: Monthly complaint rates by borough.

The seasonality within each year has an intuitive pattern. Complaint rates are higher in warmer months, since more people are outside or opening their windows, and gatherings and music may spill out onto sidewalks or patios. The pronounced dip in the middle of each summer and winter holiday may have to do with people leaving town on vacation when school is out.<sup>6</sup> The ordering of the boroughs follows from the heatmaps above: Manhattan leads the way in noise complaints, whether or not it leads in noise, followed by the Bronx, Brooklyn, Queens, and finally suburban Staten Island.

The general increase in rate over the years is less intuitive. It's unlikely that the city soundscape has changed very significantly over the period, and in fact advances in technology often make things

<sup>6</sup> Perhaps a greater number of tourists rush in to take their place, but they come here for the noise, not to file complaints about it.

quieter.<sup>7</sup> Changing awareness or expectations about noise may be a factor, particularly as formerly-modest parts of the city become wealthier.

The nature of the 311 system itself may also come into play, as awareness of its availability grew—and when its web and mobile channels began accepting complaints, it may have further raised utilization among a tech-native demographic.

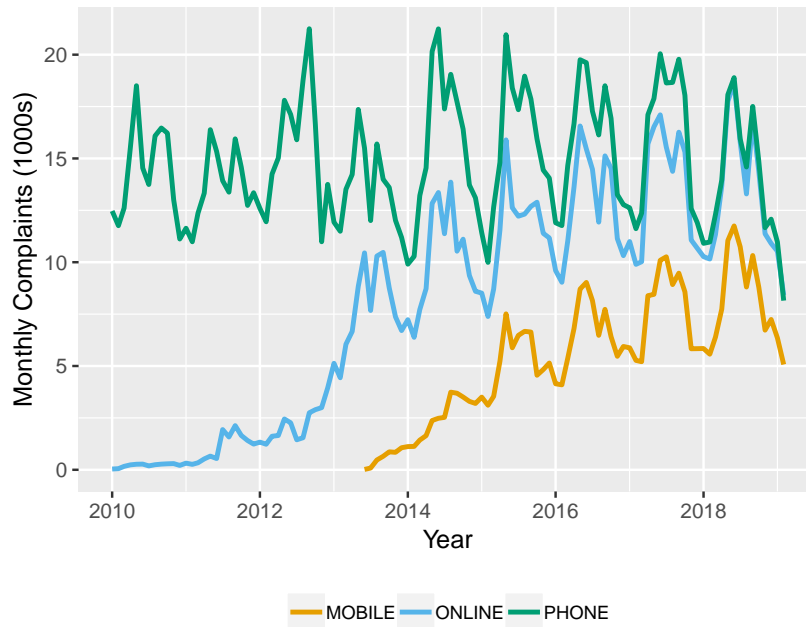


Figure 4: Monthly complaints by channel: folks that disavow the phone when hiring a car or ordering pad thai probably complain the same way.

### Median Rents

Rolling in the median rent benchmark first requires some conversion from the neighborhood-level figures to wider community districts. As an approximation, I assigned each district the average of the rents for its constituent neighborhoods, as verified at the NYC Community Portal.

Some neighborhoods span more than one district, so their rents factor more than once. Where rent values are missing for all constituent neighborhoods in a district, the borough-wide median filled in for that month.<sup>8</sup>

Manhattan and Brooklyn represent the most active markets, with many data points and less variability than the remaining boroughs. Again, rents exhibit a general increase over the period, whether from general inflation or recovery from the real estate bubble just prior,

<sup>8</sup> Staten Island, as in all things, is an outlier. It only offers borough-wide rents, and none at all for early years, which I simply excluded. Name a specific neighborhood there, anyhow. There's the mall, and the expressway, and at least two bridges, right? Who can really be sure.

and seasonality with higher values in summer (more demand) than in winter.

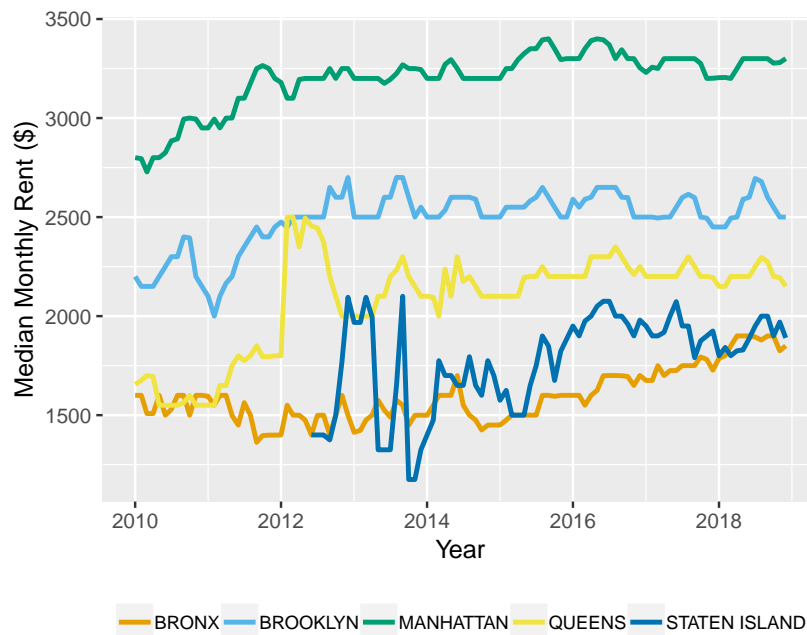


Figure 5: Monthly median rent by borough.

### *Complaint Rate vs. Rent*

Having assembled a complaint rate and a median rent benchmark for each month across every community district, I started by looking for a relationship between the two. Pulling back to the borough level for a bigger picture, a scatter of all nine years of complaint/rent pairs illustrates some trends.

While there are differences between the boroughs, a positive (but not exactly linear) correlation is apparent between complaint rate and rent within each borough. While this seems intuitive—higher rents may reduce tolerance for noise, while simultaneously drawing the noisy nightlife and crowds that flock to up-and-coming neighborhoods—it’s also a case of correlation not proving causation. As discussed above, both the complaint rate and the rents show seasonality and growth with time that may or may not stem from the same causes.

Despite this apparent relationship between complaints and rent in the wider borough view, will the same hold when filtered down to individual community districts? Taking Manhattan’s 12 community districts as an example, it does.



Figure 6: Monthly complaint rate and rent by borough.



Figure 7: Monthly complaint rate and rent in Manhattan's 12 community districts.

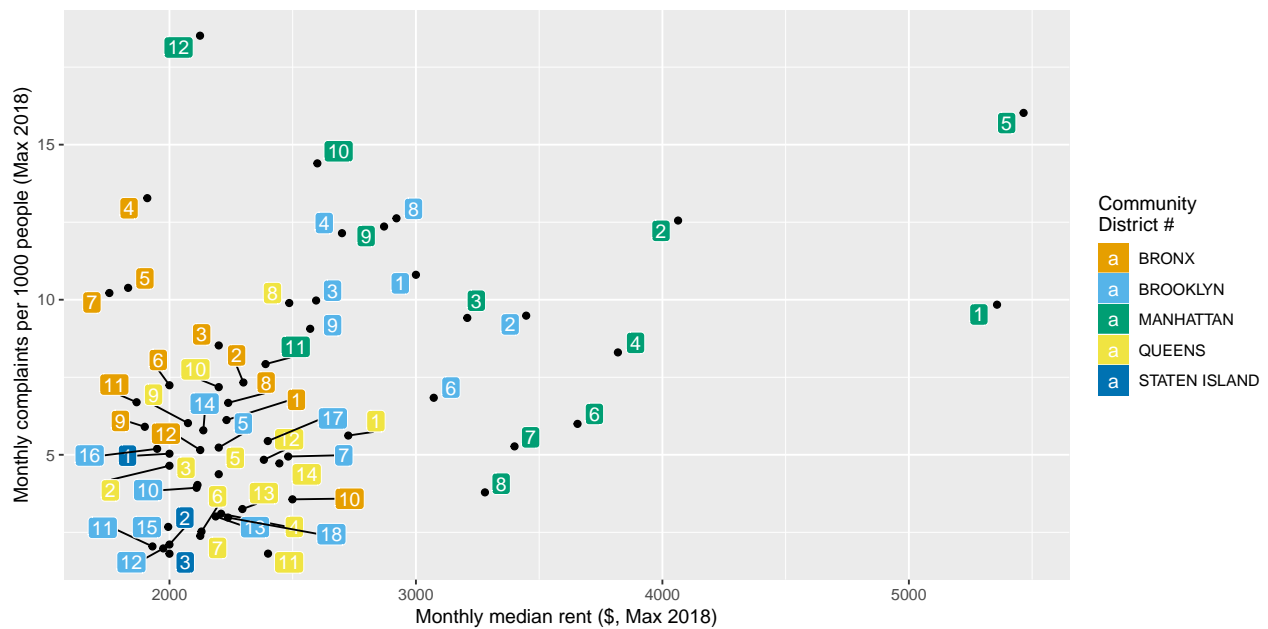


Again, while each individual district shows its own trend, there is a general positive correlation between monthly rent and monthly noise complaint rate.<sup>9</sup> The remaining boroughs show similar relationships, though at a narrower range of rents.

Let's try to address the ultimate question that follows: what are the districts that combine affordable rent and a low complaint rate? Taking this on for the most recent year of data, I tagged each district with its highest monthly rent and highest monthly complaint rate for 2018, producing a worst-case condition<sup>10</sup> for comparison.

<sup>9</sup> Each scatter describes a personality of sorts. Inwood, CD12 on the left? Sensible but complains a lot. Midtown, CD5 on the right? Extravagant and just as prone. Upper East Side, CD8 in the middle? Actually pretty laid back.

<sup>10</sup> Everyone is complaining, and everything is expensive. Welcome to New York.



As the prior plots illustrated, most districts in Manhattan are high-rent, high-complaint territory, though the Upper East Side (CD8) proves a surprisingly good value. Setting aside Staten Island,<sup>11</sup> Brooklyn districts 11 (Bensonhurst, Gravesend) and 12 (Kensington, Borough Park) appear to combine low rent and low complaint rates even after taking the worst-case values for both.

<sup>11</sup> As one does.

Of course, we can't have nice things here for very long, and a good value very quickly becomes the same old grind once people find out and move there.<sup>12</sup> A model for predicting future noise complaint rates could prove pretty useful for the noise-sensitive and rent-averse, not to mention for the city agencies tasked with responding to noise complaints.

<sup>12</sup> "Nobody goes there anymore. It's too crowded." — Yogi Berra

### *Modeling Approach*

Based on these findings, I decided to try to predict the monthly complaint rate in an arbitrary month using the borough, community district, calendar month, and median rent as predictive features in a machine learning model. Since the unknown month's median rent isn't going to be available in real time, I lagged the rent values by one month, effectively predicting one month's noise complaint rate using the prior month's known rent figures. The district and borough account for local variation, and the calendar month should help track the seasonality.

Trial and error revealed random forest regression to be a suitable machine learning model, which I optimized against root-mean-square error (RMSE) using the `Rborist` and `caret` R packages. Including the `doParallel` package improved the execution time on a dual-core CPU. The complaint and rent figures from 2010 through 2017 formed the training set, optimized using 25 bootstrap samples and the default 500 trees. The 2018 through present months became the test set for final evaluation of the model.

### *Results*

THE MODEL TRAINING on the 2010-2017 data concluded with optimal `Rborist` tuning parameters of 22 randomly-selected predictors and a minimum node size of 7. The resulting overall mean absolute error (MAE) in the 2018+ monthly noise complaint rates was 0.81 complaints per 1000 people.

Predicted and actual complaint rates for an example district (Brooklyn CD6, where I live) show a reasonable correspondence, and a local MAE of 0.52 complaints per 1000 people.

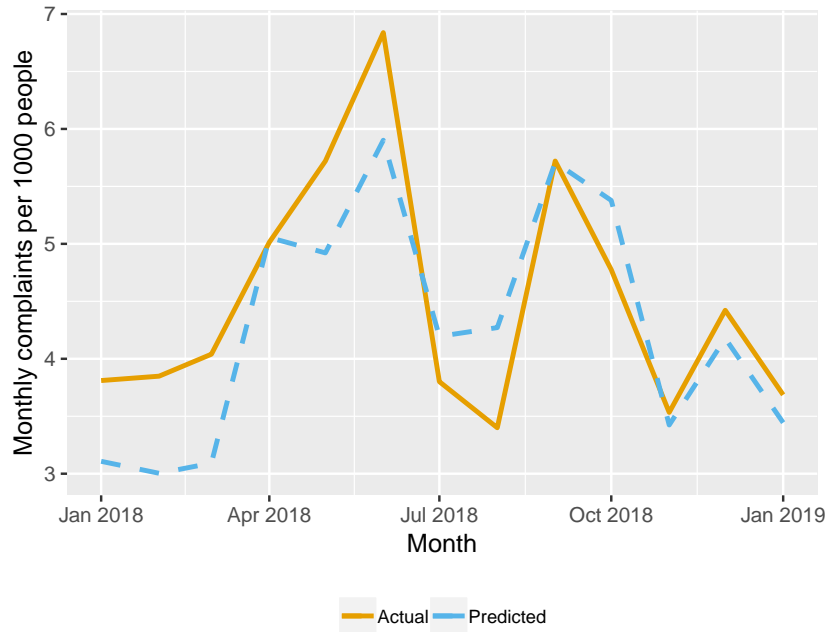
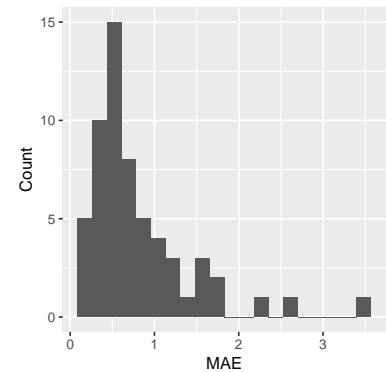


Figure 8: Noise complaint rate prediction in the 2018+ test set for Brooklyn CD6 (MAE = 0.52 ).

Looking into the distribution of mean absolute prediction error for the 59 individual community districts, the median value is 0.6 complaints per 1000 people, and 75 percent of the values are less than 1. The top three outliers are

| Community District | MAE |
|--------------------|-----|
| CD8 Queens         | 3.5 |
| CD12 Manhattan     | 2.6 |
| CD5 Manhattan      | 2.4 |

Interestingly, the highest error outlier at 3.5 complaints per 1000 people is Queens CD8—the very district that became extremely unhappy about its fire trucks and ambulances in 2018. The next two are Manhattan CD12 and CD5, both identified earlier as having unusually high complaint rates compared to other districts. It’s possible that these districts are skewed by some outlying factor<sup>13</sup> that makes projection of complaints less straightforward.



<sup>13</sup> Again, we don’t have access to the complaint source information, but this outlying factor may have a name and a well-trod speed dial.

## Conclusion

I LEFT MORE THAN A FEW interesting stones unturned in this analysis. It's likely that the prediction model leaves room for improvement with more sophisticated or intensive algorithms, or with connection to unexplored sociogeographic datasets. There is a wealth of data available about New York City's community districts, and there's no reason a finer resolution at the neighborhood level or lower wouldn't be possible with a little more finessing.

There may also be more useful outcomes to model than the complaint rate in each month; why not try to predict at the individual complaint level? Having an idea of the complaint descriptor or the agency with jurisdiction ahead of time might make the 311 intake system a little more efficient.

Still, the relationship between noise complaints and housing cost turned out to be interesting, revealing that while complaints do increase with rent, some parts of New York City just complain more than others, whether because of their actual soundscape or because of other factors (such as the diversity study mentioned in the Related Work). I expect that this geographic insight will prove valuable the next time I have to find a new apartment.<sup>14</sup>

I did attempt to predict each individual noise complaint. Unfortunately, the noise complaint landscape overwhelmingly tilts toward Party/Music gripes, which seem like the most likely outcome at any time of day or night in any corner of New York City. No model could do much better than simply predicting that every 311 noise complaint was likely to be about someone else's loud music.

<sup>14</sup> In the time it has taken you to read this report, my rent went up twice.

## Acknowledgments

Background map tiles by Stamen Design, under CC BY 3.0. Data by OpenStreetMap, under ODbL.

Document set in *Tufte Handout* style from R package *tufte* by JJ Allaire and Yihui Xie.