NEW YORK DEPARTMENT OF SANITATION

Spatial Analysis of Complaints
New York City Department of Sanitation (DSNY)
311 Complaint Spatial Analysis Assessment

Sarah Williams, Director
Nick Klein, Research Assistant
EXECUTIVE SUMMARY

New York City rolled out its 311 call system in 2003 and according to its website it “is New York City’s phone number for government information and non-emergency services.” That means if New York citizens have a complaint, instead of calling 911 – they call 311. Calls requiring service are logged by location and time. People call about everything from dead birds and potholes to juvenile loitering and noise control. The 311 system has increased citizen knowledge about the ability to complain about municipal services, and therefore, has raised the numbers of complaints logged into their systems. This increase in data has captured the attention of municipal managers throughout the city.

Among the complaints that are logged, New Yorkers use the 311 call system to report missed trash collection to New York City’s Department of Sanitation (DSNY). Incorporation in the 311 program provides DSNY with the ability to create new statistical analysis techniques on their complaint data. Like all 311 calls, complaints are logged, and the department of interest inspects them for service and review. In the case of missed trash collection, New York City’s Department of Sanitation (DSNY) reviews the complaints.

The Department of Sanitation has always had a system for reviewing complaints, however, their incorporation into the 311 system has offered them a new opportunity to analyze their information spatially. This is because the data now comes with verifiable geographic location information in the form of X and Y coordinates. DSNY also benefits from 311’s advertising campaign because it has widened the amount of people who are aware that they can call 311 about trash collection. The combination of the geographic location information and the new 311 system has caused DSNY to review how they evaluate their complaint data. It has also created a heightened awareness about using the data to look at the spatial relationship between high complaint areas and socio-demographic factors.

This study looks at the spatial patterns of DSNY missed pick-up complaints. It also analyzes whether there is a relationship between spatially clustered complaints and neighborhoods with low density housing, people living in poverty, and minority status. We found:

1) High Complaint Rates and Lower Density Housing: Some Relationships
While there are areas in the city where complaints appear to cluster, it is hard to conclusively correlate those high rates to areas that have high numbers of one and two unit buildings. However, it should be noted that those areas with very high numbers of one and two family homes (almost anomalies) appear to have some correlation with high complaint rates. This means that while low density areas can account for some high complaints, they do not account for all the complaints.

2) Complaints and Neighborhood Demographics: Race & Income
Collection complaints do not appear to correlate with race and income. However, when Manhattan is excluded from the analysis it appears that high minority / high poverty neighborhoods complain less than areas that have low numbers of minorities and higher incomes levels. However, a statistically significant relationship between these demographic factors has yet to be tested and should be performed for further studies. This relationship might highlight a lack of awareness about 311 among high minority / high poverty neighborhoods or simply show that higher income neighborhoods tend to complain more.

3) Complaint Hot Spots: Places to Explore
Finally, we implemented a spatial clustering technique referred to as “hot spot” analysis to help determine areas with statistically significant numbers of complaints. We recommend that DSNY use the “hot spot” technique to determine areas which have statistically higher levels of complaints. Analyzing hot spots more closely will allow DSNY to better determine the local factors that could have an effect on complaint data. For example, further research might highlight increased construction in high complaint areas. This information would allow DSNY managers to address the problem and better plan for trash-pick up in those areas during the increased construction periods.
INTERPRETING THE GEOGRAPHIES

An exploration of different geographical aggregations was needed because analysis would be difficult using only point location information. Missed trash collection complaints are recorded with X and Y locations and when mapped they are represented as points. The number of points plotted can sometimes make it hard to visually and statistically interpret the data, which is highlighted in the map to the right. Determining ways to geographically translate point data into more meaningful representations usually means aggregating data into administrative and statistical boundaries. Depending on the aggregation level, the resulting maps may show how many complaints are in Community Board 3 or New York Department of Sanitation Section 12?

It was determined that 2000 Census block groups are the best geographic level for spatial analysis for this study. This is because block groups are the finest geographic scale that holds comparison socio-demographic data needed for spatial data analysis.

Data was also aggregated to DSNY administrative boundaries because knowing the number of complaints by administrative sections can provide an important management tool. However it was clear that DSNY administrative sections would not be suitable for detailed spatial analysis because their geographic areas are too large.

The maps in this section illustrate the results of different aggregations, they show: 1) Complaints mapped using the X and Y coordinates. 2) Complaints represented by DSNY administrative sections. 3) Complaints represented by 2000 Census block groups.

The aggregated maps show the actual number of complaints (counts) as well as complaints per households (rates). Visualizing statistical data by its population (rates), in this case number of households, helps correct for differences in boundary sizes and location densities. For example let's imagine how using counts versus rates might affect a comparison of complaints in Manhattan and Brooklyn. Manhattan is denser than Brooklyn, and therefore, you would expect to have more complaints in Manhattan because there are more people who could possibly complain. If you visualize a map of the number of complaints (counts) - Manhattan might always show higher complaints than Brooklyn simply because more people live there. However if you create a map that shows complaints per the number of households, you would be looking at the complaints based on all those whom possibly could complain. This helps to compare the two areas by showing the data in proportions. In other words you would be able to say 20% of households complain in Manhattan and 30% in Brooklyn. Versus a map that shows 500 people complained in Manhattan and 250 in Brooklyn. In this study rate maps helped to get a better understanding of how many complaints were logged given the total number of household that might be able to complain. This allowed us to compare how neighborhoods were doing against one another.

Location of 311 Calls

The map below shows the entire database of 311 complaints made during July 2004 through July 2007. The database includes over 88,000 complaints logged during the 3 years. Of the over 88,000 complaints, fewer than 1% could not be mapped. The amount of points plotted on this map makes it hard to visually interpret patterns. This data should to be aggregated using different spatial boundaries in order to understand intensities of complaints. Note that all the aggregated maps in this study use this complete data set. Therefore all maps represent complaints from July 2004 - July 2007.
These two maps show missed collection complaints aggregated using DSNY Section geography, which is an administrative unit used for managing trash collection. The left map shows the total numbers of complaints per each section. The right map illustrates complaints per households. Comparing these two maps reveals several areas in Manhattan, the Bronx, and Queens that have high numbers of complaints (counts) but low complaints per households (rates). These two maps illustrate the importance of making maps using proportions or rates.
Block Group Geography

These two maps show missed collection complaints by 2000 Census block group geography. Census block groups have a smaller geographical representation than DSNY sections, and therefore, allow geographic measurement at finer level of detail. Census block groups are the smallest Census geography which many socio-demographic variables can be obtained. While it is possible to get Census data by block (a boundary smaller than block group) it is limited to very basic data such as age and sex. Therefore, Census block groups are used for all of the analysis in this study.

The maps show that New York City’s outer boroughs appear to have more complaints than most of Manhattan. The left map represents the actual number of complaints aggregated to Census block group. The map on the right illustrates complaints per households. Notice that the rate maps help to highlight complaint intensities in the outer parts of Queens that might not have been identified because of their low densities.
SOCIO-DEMOGRAPHIC ANALYSIS

Understanding what causes collection trash pick-up complaints is a main concern for the New York City’s Department of Sanitation (DSNY). Of particular interest is whether there are neighborhood conditions that might cause higher complaints. Knowing conditions that cause complaints can help DSNY provide better services to those communities. In order to better understand the spatial relationship of missed collection complaints this study focused on analyzing variables DSNY thought caused complaints. In other words, statistically testing common or local knowledge regarding why people complain.

Neighborhoods made up of largely one and two buildings are often cited as having higher complaints. A spatial analysis of these communities helped determine if there is a statistically significant relationship between low density housing and complaints. An analysis of one and two unit buildings showed that there where some correlations, however, only in areas having levels higher levels of these low density homes.

DSNY also wants to ensure that they are providing equitable service. Therefore spatial analysis of complaints and socio-demographic information was performed to help determine if there is a relationship between demographics and service level. We found that complaints are not correlated with race or income, in other words on average complaints appear to come from a diverse set of communities.

Even though we found no correlation between high rates of complaints, income, and race we did find some interesting patterns in places where people do not complain. It appears that low income/minority communities often complain less. Understanding why these neighborhoods do not complain is just as important as knowing why communities complain. These communities might not be complaining because they do not know they can complain or because they have better services. It is important to determine why these communities are not complaining in order to understand whether they can be models for larger management strategies.

### Complainers 2007

The finding that there appears to be no correlation between complaints, race, and income is supported through our maps as well as data charting out these relationships. The table below shows twenty-five block groups the complained the most from January - July in 2007. The data is ordered showing the block groups with the highest complaints per household next to data about income and race. It is hard to discern a pattern because it appears the high rates of complaints come from a wide range of socio-demographic groups.

<table>
<thead>
<tr>
<th>Complaints per Household</th>
<th>Minority</th>
<th>Median Household Income</th>
<th>Borough</th>
</tr>
</thead>
<tbody>
<tr>
<td>6.90%</td>
<td>79%</td>
<td>$37,250</td>
<td>Bronx</td>
</tr>
<tr>
<td>5.26%</td>
<td>80%</td>
<td>$12,159</td>
<td>Bronx</td>
</tr>
<tr>
<td>4.76%</td>
<td>28%</td>
<td>$47,308</td>
<td>Brooklyn</td>
</tr>
<tr>
<td>4.92%</td>
<td>54%</td>
<td>$17,426</td>
<td>Brooklyn</td>
</tr>
<tr>
<td>4.31%</td>
<td>11%</td>
<td>$37,031</td>
<td>Brooklyn</td>
</tr>
<tr>
<td>8.43%</td>
<td>75%</td>
<td>$16,579</td>
<td>Brooklyn</td>
</tr>
<tr>
<td>5.33%</td>
<td>59%</td>
<td>$17,583</td>
<td>Brooklyn</td>
</tr>
<tr>
<td>4.97%</td>
<td>65%</td>
<td>$16,071</td>
<td>Brooklyn</td>
</tr>
<tr>
<td>8.49%</td>
<td>40%</td>
<td>$44,091</td>
<td>Brooklyn</td>
</tr>
<tr>
<td>9.62%</td>
<td>6%</td>
<td>$12,443</td>
<td>Brooklyn</td>
</tr>
<tr>
<td>7.69%</td>
<td>96%</td>
<td>$50,060</td>
<td>Brooklyn</td>
</tr>
<tr>
<td>16.05%</td>
<td>98%</td>
<td>$8,295</td>
<td>Brooklyn</td>
</tr>
<tr>
<td>6.84%</td>
<td>98%</td>
<td>$45,156</td>
<td>Brooklyn</td>
</tr>
<tr>
<td>4.63%</td>
<td>22%</td>
<td>$136,853</td>
<td>Manhattan</td>
</tr>
<tr>
<td>4.35%</td>
<td>40%</td>
<td>$29,625</td>
<td>Manhattan</td>
</tr>
<tr>
<td>5.26%</td>
<td>57%</td>
<td>$44,286</td>
<td>Queens</td>
</tr>
<tr>
<td>4.83%</td>
<td>100%</td>
<td>$37,708</td>
<td>Queens</td>
</tr>
<tr>
<td>5.34%</td>
<td>27%</td>
<td>$119,027</td>
<td>Queens</td>
</tr>
<tr>
<td>7.24%</td>
<td>25%</td>
<td>$40,772</td>
<td>Queens</td>
</tr>
<tr>
<td>5.14%</td>
<td>45%</td>
<td>$60,238</td>
<td>Queens</td>
</tr>
<tr>
<td>4.27%</td>
<td>86%</td>
<td>$33,750</td>
<td>Queens</td>
</tr>
<tr>
<td>5.66%</td>
<td>79%</td>
<td>$35,682</td>
<td>Queens</td>
</tr>
<tr>
<td>10.87%</td>
<td>88%</td>
<td>$18,958</td>
<td>Queens</td>
</tr>
<tr>
<td>5.32%</td>
<td>16%</td>
<td>$100,649</td>
<td>Queens</td>
</tr>
<tr>
<td>4.44%</td>
<td>44%</td>
<td>$53,125</td>
<td>Staten Island</td>
</tr>
</tbody>
</table>

25 Highest Collection Complaints per Household (January - July 2007) - Block Groups
An analysis of one and two unit buildings and missed collection complaints highlights what appears to be a small correlation between these variables. This relationship can be detected when visually comparing maps of these two variables. The map to the left shows distribution of buildings with one housing units (both single-detached and single-attached building types) and buildings with two housing units. The map to the right shows the distribution of 311 collection complaints per household at the block group geographic level. This visual comparison illustrates that most complaints come from the outer boroughs where there are larger numbers of one and two family buildings.
Further Investigation of Low Density Housing

Maps help to illustrate variable correlation, however, in order to confidently assess the measure of correlation a spatial analysis was performed on the two variables. The scatter plot below illustrates the results of the regression. The results appear to show that there is a correlation between these two variables when the number of one and two unit buildings is above the mean, but it is not clear there is a relationship when this number is at the mean or below.

In order to better interpret the relationship between the level of one and two unit buildings and complaints, the data was split into two sections. One section held the block groups that had over 500 one and two unit buildings and the other section held those block groups with below 500 one and two unit buildings. A regression analysis was performed on each section, the results are shown on the right. The graph on the top right shows the results for those block groups that have less than 500 one and two unit buildings, and the graph on the bottom shows those less than 500.

The graphs illustrate that the relationship between complaints and one and two unit buildings is only significant for those blocks groups that have very high levels of complaints. Notice how the top graph shows no trend line between the two variables while the bottom graph shows a weak linear trend. When these two variables are tested using the Pearson’s Correlation, the result is 0.801 (and is statistically significant at the 0.01 level). This demonstrates that while there appears to be a correlation between one and two unit buildings this statement only holds true for neighborhoods that have high numbers (over 500) one and two unit buildings.
Even though the analysis shows that there may be a relationship between housing characteristics and complaints. It also demonstrates that there are still many block groups that have high rates of complaints where housing characteristics may not be the cause. This is illustrated in the map below, which has black points on all the block groups that have over 500 one and two unit buildings and lays those points on top of the total number of missed collection complaints. Notice that there are many block groups with high numbers of one and two unit buildings that also have high rates of complaints. (This further illustrates the point made with the previous graphs.) However, you should also notice that there are still many block groups that have high levels of complaints and there are not high levels of one and two unit buildings. This tells us that housing characteristics are not the sole explanation for high complaint rates in many parts of the city.

Demographics: Income and Race

An exploration of social demographic variables and their relationship to complaints can be just as informative as an analysis of housing characteristics. If Manhattan is removed from the visual comparison of median household income and race, the results show that areas with high incomes and low levels of minorities, appear to complain more. This information tells us more about the complainers than the complaints themselves. It appears to suggest that low income/high minority neighborhoods complain less. Does this mean these communities have less to complain about or does it suggest that these neighborhoods do not know they can complain? One hypothesis might be that immigrant communities do not know they can complain, due to cultural barriers such as language, legal status, or social norms. Further quantitative and qualitative analysis should be performed to determine if there is a tendency for low income/high minority communities to complain less.

Understanding why communities do not complain can be just as important as understanding why they do complain. For example, a study that looks into why communities complain less might highlight the need for 311 advertising in alternative languages. An advertising campaign that reaches out to a larger cross section of New York City residents might help make the 311 systems more statistically valid and socially equitable. Knowing why communities do not complain might also highlight areas where service provision is working well. Understanding why service provision is successful could assist managers in developing strategies that replicate successes across the city. In other words low complaint areas might provide a model for city-wide service provision. It is recommended that any analysis of 311 data should look at causality in both low and high complaint areas.

The complexity of why people complain about service makes any analysis of complaints multi-variant. It is important that DSNY test variables that are considered common knowledge, this would provide statistically backing for their claims. It is also important to acknowledge that Manhattan has a very different character than the rest of the boroughs and it might be worth while to study it individually.

Some possible further investigations might include:

1] English Proficiency
Although the 311 system provides language assistance in 170 languages (August 2007), people with limited English skills may not be aware of the service.

2] Owner Occupancy
It is possible that owner occupancy may be correlated to 311 complaints. Owners may be more likely to call 311 to complain about missed collection.

3] Mixed Land Use
DSNY staff suggested that there might be a causal relationship between 311 complaints and the presence of commercial land uses in close proximity to residences.

4] New Construction
DSNY staff suggested that there is a possibility that new home construction could cause higher levels of complaints. This is because there is a lag between when a home is completed and when it is occupied, thereby causing a lag in services provision.
The following maps show a comparison of median household income and rate of missed collection complaints. If one removes Manhattan from the visual analysis, it appears that complaints are linked to higher income communities. For example notice the coincidence of complaints and higher than average incomes running through the mid-section of Queens and Staten Island. A significance test should be performed on these two variables.
Minority Neighborhoods and Complaints

The following maps show a comparison of low minority neighborhoods and rate of missed collection complaints. If one removes Manhattan from the visual analysis, it appears that complaints are linked to low-minority communities. Similar to the median household income maps on the previous pages notice the coincidence of complaints and low-minority neighborhoods running through the mid-section of Queens and Staten Island. This suggests that low-minority communities complaint more than high-minority communities. Do high-minority communities know they can complain? A significance test should be performed on these two variables.
HOT SPOT ANALYSIS

Hot spot analysis is a unique way to explore the spatial relationship of complaints on a city-wide level. It allows an analyst to determine whether high and low levels of complaints have a spatial significance given the complaint rates around those clusters. This technique was applied to an analysis of collection complaint data to determine areas that have significantly higher complaints, or hot spots. The results were mapped to show hot and cold complaint areas throughout New York City. This technique is commonly referred to as hot spot mappings.

The procedure used local G* statistics to test for locally significant spatial dependency (clustering) within the region. These tests were completed using an ArcGIS tool that performs a series of random permutations of the data. Each permutation compares the original values with the randomly permuted values of all locations except the test location. The results from the random permutations produce a value and an estimate of the statistical significance, all relative to the data and weight matrix.

The following two maps show the results of the hot spot analysis. The map on the right - shows “cold” spots - spatial clusters or regions where there are locally lower levels of collection complaints. The left map shows “hot” spots - areas where there are higher levels of collection complaints.
Localized Hot Spot Analysis

One major advantage of using hot spots analysis is that it allows you to highlight “problem” neighborhoods. Focusing on a hot spots allows one to identify a region or neighborhood that might be in need of further study. In our study, Lower Manhattan has proved to be an interesting anomaly. It appears to have significantly higher complaints than many of the communities that surround it, which means it is a hot spot. Given its distinct nature it was used to show how problem areas (hot spots) could be further explored.

Understanding why Lower Manhattan might be a hot spot should start by looking at why this area might be an anomaly. As mentioned earlier the New York City Department of Sanitation often cites housing and new construction as a cause for increased levels of complaints. Therefore a comparison of housing complaint rates might be helpful in getting a better understanding of this anomaly. Another causal condition may be the presence of new residential construction. Over the last few years lower Manhattan has had an increase in residential construction. It is be possible that there is a lag between residents moving to lower Manhattan and service provision.
When comparing complaint rates with total number of households in lower Manhattan, it is clear that the areas with higher complaints have less overall residents than the areas that surround it. This leads one to speculate that the lower levels of housing units might have something to do with higher levels of collection complaints. Perhaps there are fewer routes in this area or the extensive construction in the downtown area has caused interference with pick-ups. It is also possible that trash routes have not been established for some newly residential areas. Further investigation into the localized causes for complaints should be performed.
Hot Spots and Visualization

Identifying problem neighborhoods, such as this one in Lower Manhattan, allows for analysis techniques suitable for a more local scale. The following maps show just a few different types of analysis that can be performed on the local level.

As mentioned earlier representing complaints as individual points can make it hard to visually interpret. The following two maps show alternative ways of representing complaint information. The map below shows how many complaints are on each street segment. The more complaints there are the darker and thicker the street segment. While this method does not normalize data by the total number of households that could complain, it is helpful in understanding what streets might be more problematic for service crews. Streets with high levels of complaints can be surveyed for conditions, such as road blockages, that might have caused a missed trash collection. Knowing what caused the complaints can help the DSNY systematically handle problem area.

The map below illustrates areas that have high densities of complaints. Density and hot spot maps are often confused, it should be noted that density analysis is different from hot spot analysis. Hot spot analysis tells us whether high levels of complaints are significant given the complaints located around them. While density surfaces simply show the intensity of complaints, it does not explain whether that intensity is significant. While density surfaces do not measure significance, they are helpful in locating areas with high complaints.
Next Steps:

1] Use Hot Spots to Help Manage Complaints
Regarding the question of how 311 data should be used to improve DSNY service, we believe that while the data helps to highlight causal relationships between demographics and complaints, the analysis of the data would best serve the Department if they look at “hot spots” – those areas that have statistically higher numbers of complaints than the area that surrounds it. They should than look at commonalities between these “hot-spot” areas. In other words looking at what is similar about areas that have higher than normal levels of complaints. These commonalities might be demographic, but they might also be spatial. For example high complaint areas may lie on the border between two districts. This spatial relationship might help DSNY managers to determine management strategies that would help them deal with these locations. Further research should be done looking at the commonalities of hot spot locations.

2] Demographics of Low Complaint Areas : Do People Know they Can Complain?
Looking at the commonalities between high complaint areas is appropriate for service improvement. However, it should be noted that a detailed study of demographic factors might highlight caveats in using 311 data. For example it is possible the certain populations under complain because it might not be in their cultural background to complain? Understanding biases in the 311 data is just as important as understanding where high amounts of complaints are occurring. Further research should be performed to better understand low complaint areas.

Given the possible biases in all complaint data, using the hot spot analysis technique to study the commonalities between hot spot and cold spot neighborhoods might be more informative as a management tool.