Red Zone, Blue Zone: Discovering Parking Ticket Trends in New York City

Samuel S. ACKERMAN[1] and Dr. Rida E. MOUSTAFA

Department of Statistics, the George Washington University,

Washington, DC 20052, USA

July 2011

In this paper we study data on parking ticket violations in New York City in the month of March 2010 to review basic trends relating to the types of vehicles receiving tickets in various parts of the city. The variation in character of New York City neighborhoods—mostly residential in Staten Island, heavily commercial in Manhattan, for instance—contribute to differing traffic patterns and thus different types of violations; double parking, for example, is relatively less common in Staten Island than in the bustle of Manhattan. Using maps as well as clustering techniques, we will develop basic intuition about the geographic variation of passenger vehicle violations within the City.

Key Words: Data mining; clustering; clustergram; New York City; parking.

1. INTRODUCTION

Parking violations are a continual presence and source of annoyance in the lives of many residents of large modern cities, especially New York City. Especially with on-street parking relatively difficult to find in certain districts in the City, tickets can make commuting or performing errands by car particularly burdensome, restricting the hours and locations that parking is available; New York City's weekly alternate side parking policies for street cleaning sometimes lead apartment building doormen to temporarily double-park groups of residents' cars on the opposite side of the street to avoid tickets. They can also be an important source of revenue for cities: in the 2010 fiscal year, New York City's Department of Finance, the agency responsible for issuing parking ticket violations, collected nearly $605 million in revenue from these tickets.[2]

This paper aims to analyze New York City parking ticket data to present some patterns on ticket violation issuance, focusing on understanding the relationship between the location and type of violation. The academic research available on parking policy is relatively limited, and more so is the analysis of parking tickets specifically. A 2007

[1] Author correspondence may be directed to ackrmn@gwu.edu and shalash@gwu.edu

[2] "Mayor's Management Report," 181.

paper by Fisman and Miguel studied trends in parking tickets received by diplomats in New York City from 1997-2005, to see whether levels of corruption or attitude toward the U.S. in diplomats' home countries affected their rate of abuse of diplomatic privilege in evading parking tickets.[3] Their data originated from the New York City Department of Finance, as did the dataset used in this paper; however, they focus entirely on diplomatic vehicles, which comprise less than 0.1% of our dataset, and analyze time trends, while the dataset used in this paper is restricted to one month of tickets. Some other papers have applied economic modeling to parking policy, such as analyzing the cost of cars cruising while looking for street parking, optimal market allocation of parking spaces, and the equilibriums between off- and on-street parking.[4]

1.1 BACKGROUND ON TICKETS IN NEW YORK

New York City's major parking ticketing operations require an extensive infrastructure of courts and resources to resolve tickets; about 10 million tickets are issued each year and 1.2 million of them are contested in hearings, with only slightly more than half resulting in a violation paid.[5] To improve the system's efficiency and trim the administrative log, the City has instituted several key policies. For instance, a program introduced in 2005 allows drivers to plead guilty to a violation in return for a fine reduction of approximately 10 to 25 dollars, determined by violation and location, and thus skip a judicial hearing.[6] In March 2011, the City launched One-Click Hearings, a program through which ticket recipients could contest tickets online as well as submit supporting materials, such as written statements or photographs of the location, and thus avoid attending a hearing. The goal of the new program is to ease the burden on residents and small businesses whose employees would have to leave during work hours to attend hearings, rather than strictly to save the city money.[7]

A parallel version for commercial delivery vehicles, the Stipulated Fine program, was implemented in the mid-2000s. Parking in the City can be very expensive for delivery vehicles, since they are routinely ticketed during routes for violations such as double parking. The program reduces or eliminate fines for certain violations in exchange for the company waiving the right to dispute the fines; previously, a large backlog of court cases was created by companies disputing tickets, and this program was designed to cut City administrative costs as well as costs to the company.[8]

---

[3] Fisman and Miguel.
[4] Arnott and Inci, 3.
[5] Hernandez.
[6] McGinty and Blumenthal.
[7] Hernandez.
[8] See "Commercial Fleet Programs." York City Department of Finance. http://www.nyc.gov/html/dof/html/parking/park_commercial.shtml. Also see "Delivery Firms' Big Ticket Item: Parking Fines." Associated Press. September 1, 2006. http://www.msnbc.msn.com/id/14602712/

2. DATASET BACKGROUND

The dataset used in the analysis consists of the set of all 872,370 parking ticket violations issued in New York City in March 2010; each observation comprises one ticket issued. The data, originally collected by the New York City Department of Finance, are available through NYC Datamine, a website containing city datasets on topic such as facilities, social services, and event calendars. NYC Datamine was established as a project between the Mayor's Office and other City agencies to support the NYC BigApps initiative. In this project, launched in late 2009, software developers competed to create digital applications using these datasets. Some of the winning applications from 2010 included tools to help users find the best path to a nearby train stop, or to enable passengers to post live comments on taxi drivers to a forum.[9]

2.1 VARIABLES AND DATA PROCESSING

For each ticket, the following attributes were recorded: a unique summons number for the violation, the license plate of the car, the state to which the vehicle was registered, a three-letter code for the type of vehicle (passenger, commercial, specialty plates), the date the ticket was issued, a numeric violation code and a description (missing for many observations but completed using information from the Department of Finance Website), a dollar fine amount, and geographic data for where the violation occurred (borough, street name and number, intersection or landmark, an indicator for whether the violation occurred opposite/in front of, etc. of the location, and three street codes). Of these observations, 9,461 were discarded (about 1.09%) due to problems in key variables, primarily invalid or missing borough or address location, invalid violation codes, missing fine amounts, and missing license plate information; the final dataset consists of 862,909 observations. A substantial portion of the discarded observations were unusable as they were missing all four of these.

For the analysis, various manipulations were conducted. The variables describing the location where the ticket was issued were consolidated into an address string (ADDRESS), which was then checked for validity; for a small sample, latitude and longitude codes were obtained to display a distribution of locations. The violation codes (V_CODE)—of which there are more than 80 unique values[10]—were consolidated into six categories (VIOL_CAT): (1) parking during street cleaning hours, the single most common violation, comprising nearly 16% of tickets; (2) stopping, standing, or parking in illegal areas, or at certain hours; (3) parking in illegal ways or blocking access or traffic (e.g., double parking, parking the wrong way or at an angle); (4) parking beyond the time allowed by regulation or by the meter; and (5) parking without proper registration, documentation, or with damaged license plates, etc.

Other variables were generated to represent the day of the week the ticked was issued, as well as an indicator for whether the car was registered out of New York State. Using the license plate as a vehicle identifier, a count of the total tickets (PLATEFREQ)

---

[9] "Mayor Bloomberg Announces Winners of Inaugural NYC BigApps Competition."

[10] A nearly complete list may be found at http://www.nyc.gov/html/dof/html/parking/park_tickets_violations.shtml

issued to each vehicle was calculated; this count ignored observations with missing license plates, but included observations discarded for missing other key variables. Using the vehicle plate as an identifier, the license plate class (TYPE)—which includes codes for generic passenger and commercial vehicles, as well as specialty and custom plates for taxis, public transportation, and other vehicles—was made consistent for each vehicle plate. The plate class was then used in two ways to categorize vehicles: the variable VEH_TYPE identified the vehicle as a either a passenger, commercial, or other type; more than 95.75% of violations were identified as passenger or commercial.[11] The variable PCLASS identifies taxis or vehicles that have custom plate classes[12] (such as for a sports team or organization) as separate from the generic PAS or COM (passenger or commercial) plate classes. Lastly, the zip code of where the violation occurred was generated from the address information, and the latitude and longitude of the zip code were calculated.[13] Table 1 shows the attributes available in the final dataset:

Table 1

Attribute name	**Description**	**Values**
SUMMONS	Unique ticket summons identifier	Numeric
PLATE	License plate of vehicle	Character; unique to vehicle but multiple observation per vehicle are included
STATE	Code of state of vehicle registration; additional codes for diplomatic and government vehicles	
TYPE	Vehicle license plate class	
PCLASS	Vehicle class	"Commercial","Passenger", "Other","Taxi", "Custom"
VEH_TYPE	Vehicle type	"Commercial","Passenger", "Other"
V_CODE	Numeric code for violation category	Numeric, range from 6 to 99
VIOL_CAT	Category of violation (determined by author from violation code)	"Street clean", "Stop illegal area", "Blocking/improper park", "Overtime", "Missing/damaged reg"
DOISSUE	Date of ticket issue	
FINE	Dollar fine charged for ticket	Numeric

[11] The vehicle type category recorded in the original dataset is the New York State license plate class. Fortunately, nearly 94% of vehicles were already designated as either commercial or passenger, and several other classes were designated to one of these two. The other category includes vehicles with vanity plates that could not be confirmed as belonging to one of the categories, and vehicles such as taxis and tractors.
[12] Custom plate classes are separate from personalized or vanity plates, for which a vehicle owner can select a desired phrase or character combination for their license number.
[13] ZP4 software from Semaphore Corp. was used. Although zip codes were available for a sizable majority of addresses, addresses in Queens could not be coded as these addresses are listed under their neighborhood (e.g., Kew Gardens) rather than the borough. Zip code coordinates were derived from the R package *zipcode*.

BOROUGH	Borough in which violation occurred	"BN" (Brooklyn), "BX" (Bronx), "MH" (Manhattan), "QN" (Queens), "SI" (Staten Island)
ADDRESS	Location of violation	Text
PLATEFREQ	Number of tickets received by vehicle in March 2010	Numeric
NODATES	Number of separate days on which vehicle received tickets in March 2010	Numeric; range 1 to 31
OFFENDER	Indicator of 1 if PLATEFREQ>1	
DAYOFMONTH	Day of month on which violation occurred	Numeric; range 1 to 31
DAYOFWEEK	Day of the week on which violation occurred	Sunday-Saturday
WEEKEND	Indicator of 1 if violation occurred on Saturday or Sunday	
DIPLOMAT	Indicator of 1 if vehicle is diplomatic vehicle (license plate begins with "DIP")	
ZIP5	Five-digit zip code for ticket location.	
ZLATITUDE, ZLONGITUDE	Latitude and longitude of ZIP5	

The generated attributes PLATEFREQ and NODATES are month-specific totals for each vehicle; clearly their values would vary between months for a given vehicle if multi-month data were available, but in using them we implicitly assume that the relative monthly ticket frequencies give us some information that is intrinsic to a given vehicle and driver. For instance, while a certain passenger vehicle may receive two tickets in one month and none the next, if it instead received 20 parking tickets it is most likely that there is some intrinsic quality or usual behavior that we can attribute to this vehicle. Such a high ticket frequency may be indicative of someone who does particularly strive to avoid violations, or perhaps they regularly commute to an area where parking tickets are very common. These variables are treated as an intrinsic quality of the vehicle, even if the ticket was received on the first of the month.

2.2 DATASET SUMMARY

The data show that Manhattan accounts for close to half (about 43.8%) of tickets issued, followed by Brooklyn at 22.6%, which is sensible given that they are major business districts. This borough distribution can also be explained by the fact that tickets are issued disproportionately on weekdays; weighted by the frequency of each day in March 2010, 10.6% and 4.10% of tickets were issued on Saturday and Sunday, respectively, while each weekday accounted for between 15.8%-18.9% of tickets. Since on weekdays there is a larger flow to the city's business districts, tickets are more likely to be issued there than in Staten Island, a more residential borough; indeed, in Staten

Island a higher share of tickets are issued on the weekend than on any other borough. Below, Table 2 shows the distribution of tickets by borough:

Table 2

Borough	Brooklyn	Bronx	Manhattan	Queens	Staten Island
Share of Tickets	22.63%	12.33%	43.79%	19.89%	1.36%

Map 1

The *New York Times* map[14] above (Map 1) of the distribution of nearly 10 million parking tickets issued between July 2007 and June 2008 similarly shows a heavy concentration of tickets in Manhattan and especially below Central Park. The distribution within the other boroughs is not uniform either; each has at least one heavily-ticketed area around landmarks such as Fordham University in the Bronx. Table 3 below also shows the distribution of tickets by violation category.

Table 3

Violation Category	Stopping in illegal areas	Blocking traffic or improper parking	Missing or damaged registration or plates	Overtime parking	Parking during street cleaning
Share	29.47%	21.66%	16.50%	16.50%	15.87%

[14] Bloch and Cox. Blocks marked in red received 2,000 or more tickets over the period; those in yellow received between 500 and 1,999; those in gray received less than 500.

Examination of the distribution of individual violation codes by borough reveals a large variation in the numbers of each violation. For instance, violations having to do with commercial vehicles (such as unauthorized parking in a commercial zone) and municipal parking facilities (such as failing to display a muni receipt) happen almost exclusively in Manhattan; street cleaning violations, overall the most common, are issued rarely in Staten Island since cars are usually not parked there during the day. In Staten Island, the most common violation is for missing an inspection sticker, while it is only the sixth most common elsewhere; perhaps due to the higher concentration of vehicles in the more urban boroughs it is easier to issue street cleaning or no parking zone violations since most of the cars in a certain area may be committing the same violation; to issue an inspection sticker violation would require a closer look at an individual vehicle. The plots below show the distribution of violation categories (Fig. 1) and of a selection of the most common violations (Fig. 2) by borough.

Fig. 1: Violation category by borough

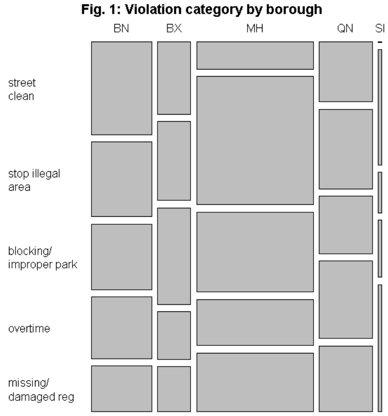
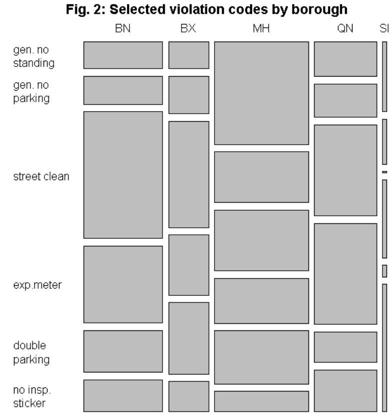
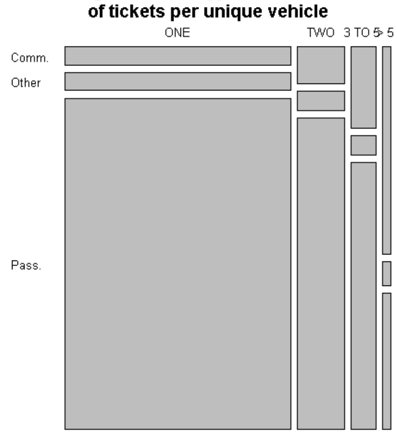


Fig. 2: Selected violation codes by borough



One can easily observe a high concentration of tickets among multiple offenders, frequently commercial delivery vehicles. More than 55% of all tickets in March were issued to vehicles that received two or more tickets in the whole month; the highest count was a vehicle that received 209 tickets. While almost 45% of tickets were issued to vehicles receiving a single ticket in the month, these vehicles accounted for about 74% of all unique vehicles in the sample (Fig. 3); furthermore, about 21% of tickets were issued to vehicles receiving 6 or more tickets in the month—these comprised only about 2.5% of all unique vehicles. About 73% of tickets were issued to passenger vehicles, though they accounted for 86% of unique vehicles; commercial vehicles accounted for 22% of tickets but only about 8.5% of vehicles.

Fig. 3: Vehicle type vs. number of tickets per unique vehicle



Because the dataset consists of only one month of observations, it omits vehicles that may have received tickets in other months but did not receive one in March. It is likely that many passenger vehicles receive parking tickets infrequently, while commercial delivery vehicles and repeat offender passenger vehicles could receive tickets in most months. Hence, there may be a bias toward a concentration of multiple offenders among the populatio of unique vehicles in a given month. From the mosaic plot below it is evident that at increasing levels of tickets per month, the vehicle is increasingly likely to be commercial rather than a passenger vehicle; the plot includes one observation per unique license plate, so the size of the higher ticket level category is not inflated by the few vehicles that are frequently ticketed.

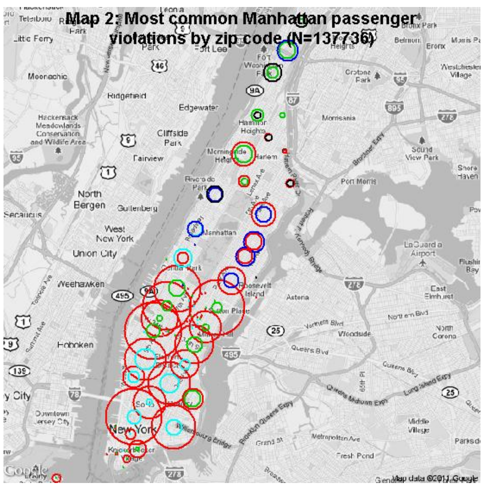
2.3 GEOGRAPHIC DISTRIBUTION OF TICKETS

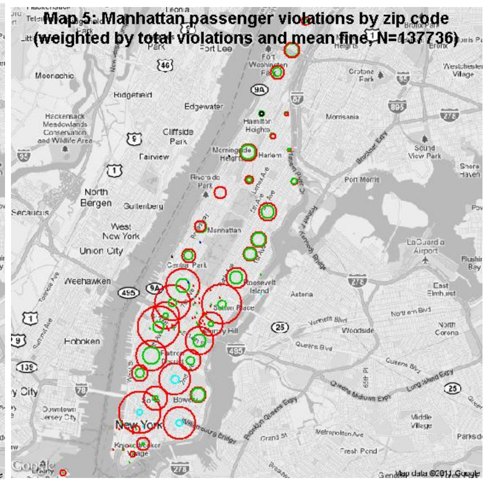
Examining the geographic distribution of violations by borough reveals interesting results. A driver in the city may be interested in knowing which areas are most risky to park in terms of the probability of receiving a ticket and the expected fine. To provide an answer that would be meaningful to the typical driver, in the following analysis the universe was restricted to violations by passenger vehicles (VEH_CLASS= ”PAS”). Using the five-digit zip code[15] derived from the address of the violation, and its latitude and longitude, two circles are plotted for each zip code, representing the two most-common violation categories for that zip code; the circles were scaled to be proportional to the overall share of borough violations represented by the circle. The colors represent the violation categories: street cleaning (black), general illegal

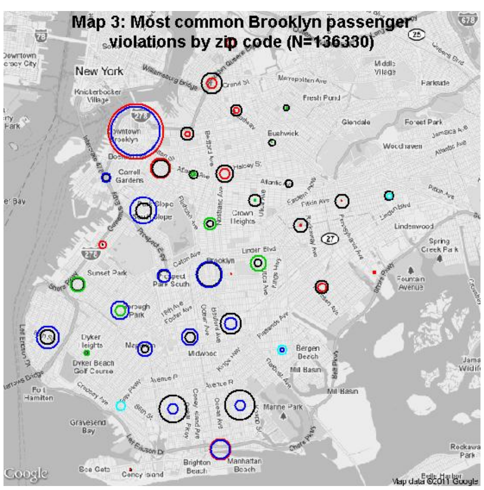
[15] Analysis was restricted to violations for which a zip code was identified; the identification rate was fairly high, for instance 82% in Manhattan. Due to slight variation in the rate of zip code identification between violations, each violation category from the subset of observations with zip code was randomly resampled from in order to return the distribution of violation within the subset to that of the borough as a whole. In each map title, ‘N’ represents the size of the sample. This was done by determining the largest sample size from the subset that would contain violations in the same ratios as the overall borough, regardless of whether the zip code was identified. The reduction in the size of the subset by resampling was about 17% in Manhattan, for instance.

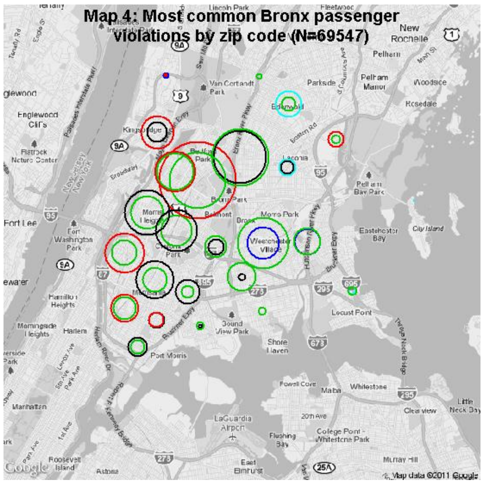
stopping/parking (red), blocking traffic (green), overtime (dark blue), and missing/damaged registration (turquoise).

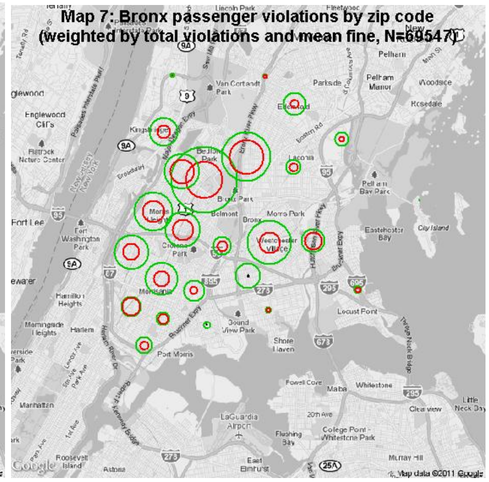
The maps reveal a surprising difference between the violation types that are likely in Manhattan, Brooklyn, and the Bronx (the three boroughs with the most tickets issued—excluding Queens, for which zip codes were not available). In Manhattan (Map 2), below Central Park, general illegal stopping/parking violations are issued at the highest rate (these also account for a large share of violations in Manhattan overall) with the missing/damaged registration as the second most common in lower Manhattan and blocking traffic as second in the area of midtown. On either side of Central Park, overtime and stopping/parking violations are most common, though the frequency of ticketing is much lower than in lower Manhattan.











In Brooklyn (Map 3), in contrast, tickets are most commonly issued for overtime and street cleaning violations in the lower part of the borough, and in the northwest areas (such as Brooklyn Heights, where there are several college campuses), stopping/parking, overtime, and street cleaning are the most common. In the major areas of the Bronx (Map 4), violations for street cleaning or stopping/parking are most common, usually followed by tickets for blocking traffic; general stopping/parking violations appear to be most common in where there is likely to be a high concentration of passenger vehicles, such as in the zip codes of 10458 (Fordham University) and 10451 (near Yankee Stadium).

Maps 2-4 display the geographic variance within boroughs in terms of ticket frequency. A driver who wishes to know the most risky areas to park would consider the fine associated with each likely violation. Maps 5-7 above display the same ticket frequency distribution as before, except that the violation frequencies are scaled by the mean fine in the borough overall for each violation category, and the two categories with the greatest weight are displayed. In this case, the concern shifts almost entirely across the three boroughs to blocking traffic and general stopping/parking violations. The mean fines for each borough are displayed below in Table 4.

As mentioned, the City charges more expensive fines in areas below 96th St. in Manhattan (near the top of Central Park) than elsewhere in the City. Among the most expensive violations are parking in front of a pedestrian ramp (code 67, \$165) and parking in a handicapped zone (code 27, \$180 including surcharge); these are assigned to the blocking traffic and illegal stopping categories, respectively. Many of the violations in these two categories are fined at \$115, while most of the rest are fined at \$65.

Table 4

Violation Category	Stopping in illegal areas	Blocking traffic or improper parking	Missing or damaged registration or plates	Overtime parking	Parking during street cleaning
Mean fine (Manhattan)	$110.10	$113.70	$65.60	$57.40	$55.20
Mean fine (Brooklyn)	$90.40	$113.70	$61.50	$35.40	$45.00
Mean fine (Bronx)	$88.40	$115.60	$60.40	$35.60	$45.00

Maps 5-7 show that although passenger vehicles are often at risk for street cleaning or overtime parking violations, that tickets for illegal stopping/parking and for blocking traffic are the most costly in terms of expense and likelihood across the three boroughs. In Manhattan (Map 5) below Central Park, the risk of receiving a general stopping/parking violation, weighted by the mean fine, is significantly higher than that of a ticket for a violation of blocking traffic. However, in upper Manhattan and the other two boroughs (Maps 6-7), the weighted risk of a ticket for blocking traffic is higher, as shown by the high proportion of large green rings; this is partiall accounted for by the higher average fine for blocking traffic violations in all three boroughs. It's possible that lower Manhattan, due to its more commercial nature, may have more curbs designated as restricted parking zones, such as in front of hotels and other businesses, and hence its greater weighted risk toward stopping/parking violations.

2.4 CLUSTERGRAM ANALYSIS

Using the intuition gained from the maps above, we will conduct some simple clustering analysis on the data observations. For the following, we restrict analysis to observations of passenger vehicles (VEH_TYPE= "PAS) as these are the most relevant to residential drivers, as opposed to companies. From this subset, a 2,000-member random sample was selected and k=2 to 15 clusters were formed[16] using four variables: VIOL_CAT, FINE, DAYOFWEEK, and PLATEFREQ. Fig. 4 below shows the average dissimilarity of each cluster for each level of k, and Fig. 5 shows the average of these average dissimilarities, weighted by the cluster size. It is clear that the cluster dissimilarities continue to decrease as the number of clusters increases, and begins to plateau around k=15.

The clusters are analyzed using an adaptation of the clustergram[17], a plot which is useful in applications of categorical variables, which displays the clustering

[16] The *daisy* package was used for distance matrix calculation due to its ability to handle categorical as well as numeric variables, and the PAM algorithm (package *cluster*) was used for cluster assignment.
[17] Schonlau (2002).

arrangements of observations over a series of different numbers of clusters. In these diagrams, the horizontal axis shows the number of clusters (k) at each stage; for each observation, the index number of its cluster assignment is plotted on the vertical axis against each value of k, and the points are connected by line segments, and the lines can be colored to represent the value of a particular variable. The clustergram allows us to inspect the stability of a particular cluster as k changes, as well as to see the composition of each cluster by a the values of a variable through the colors.

Fig. 4: Average cluster dissimilarities over k=2-15

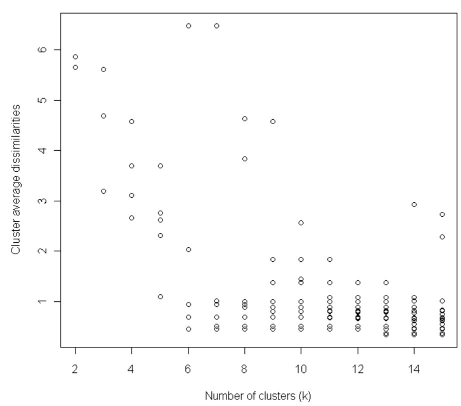
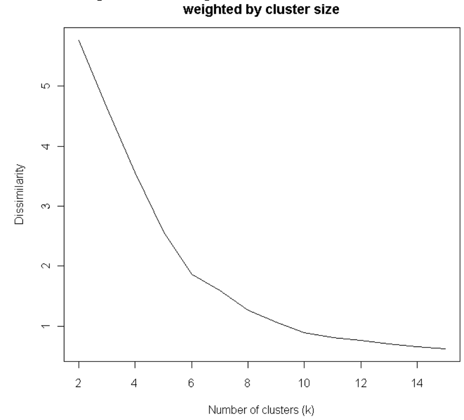


Fig. 5: Mean average cluster dissimilarities over k=2-15 weighted by cluster size



The clustergrams below show the cluster splites for k=2 to 15 for the 2,000-member sample described above, where the individual observations are colored by their values of a particular variable; in each, the clustering is identical, only the colors change. At k=15, there is an imbalance in the cluster sizes as we see that five clusters (numbers 3, 4, 5, 7, and 8) are significantly larger that the others, as is evidenced by the greater thickness of their bands. In addition, these five have remained very stable and distinct since k=6; most of the splitting occurs from the original cluster 8, which is separated from cluster 1 at k=8. This suggests there are five main groups of observations in the data; however, when the observations are colored by the value of BOROUGH (Fig. 6-7), we see that the clusters are not uniform by borough, rather that each contains observations from different boroughs. This further indicates that using the four attributes mentioned above to cluster observations does not predict the borough location; instead, similarity between observations is primarily determined by the violations themselves, not by those that may be more dependent on the driver who receives the ticket.

Fig. 6: Clustergram for N=2,000 sample, by borough

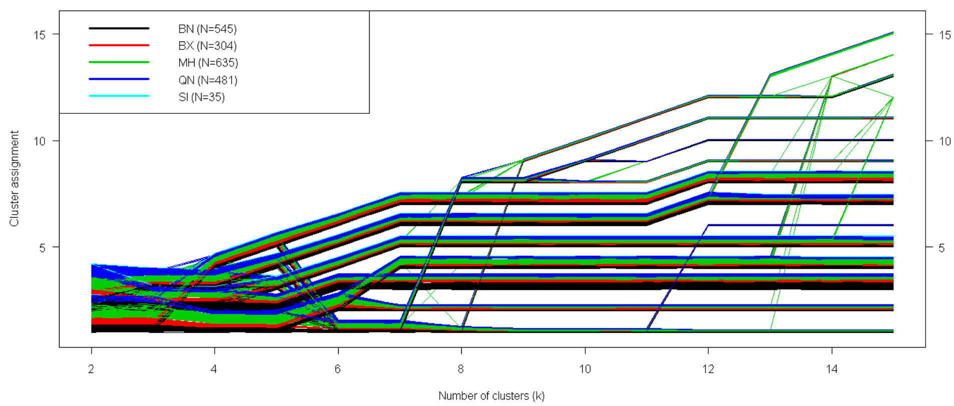
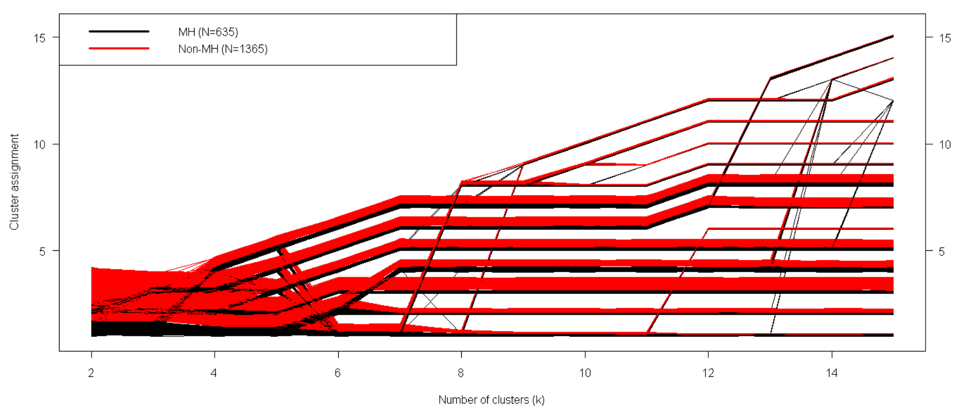


Fig. 7: Clustergram for 2,000 sample, by borough=MH



When the observations are clustered by the type of violation (VIOL_CAT), the clusters at k=15, and throughout the range of k, the clusters are nearly or completely uniform by the violation category (Fig. 8). One likely reason is that many violations that are grouped under the same violation category share other attributes, such as the fine amount; for instance, many violations under the general stopping/parking category are fined at $95 or $115, and many parking meter violations are fined at $35. Thus the information conveyed by the violation category is not entirely independent of the other clustering variables. Interestingly, in Fig. 8 it is clear that the five main clusters (3, 4, 5, 7, and 8) identified previously are each associated with one violation category: street cleaning, illegal stopping/parking, missing/damaged registration or equipment, overtime, and blocking traffic, respectively. Fig. 9 shows that for the most part, each of these five is also nearly uniformly associated with one fine level: $95 or higher for blocking traffic and $46-$65 for missing/damaged equipment; for the other three categories the association is not perfectly uniform due to differences in fines between Manhattan and other boroughs, such as street cleaning, which carries a fine of $65 below 96th St. and $45 elsewhere.

Fig. 8: Clustergram for 2,000 sample, by violation category

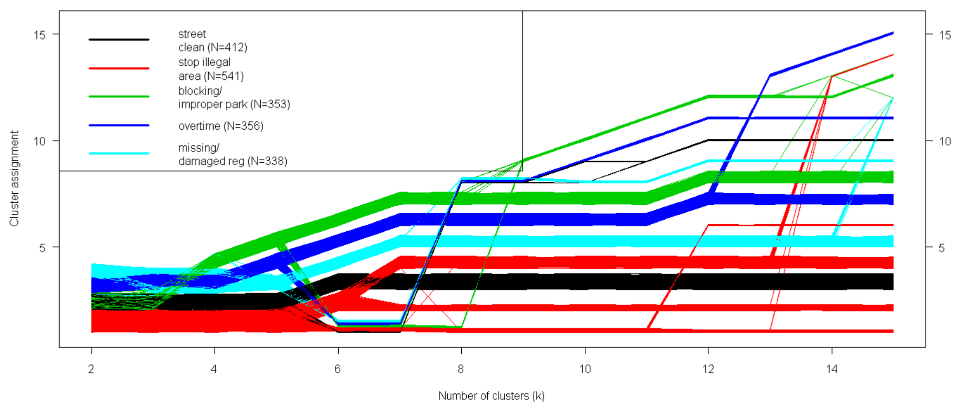
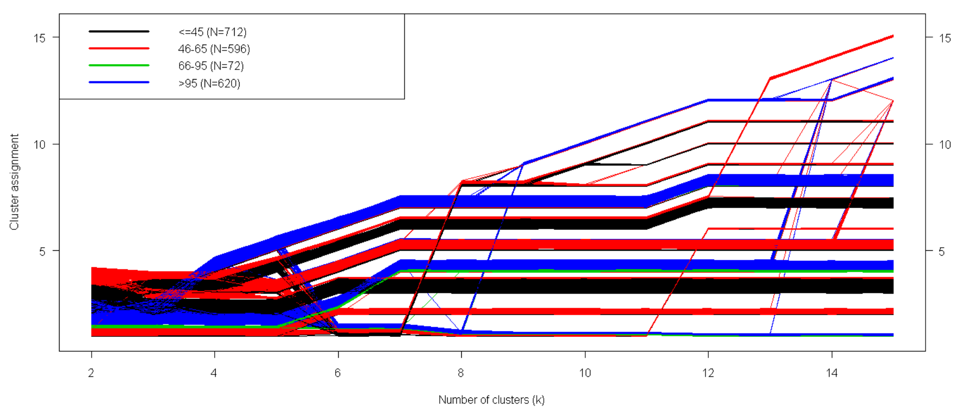


Fig. 9: Clustergram for 2,000 sample, by fine level



Who are the types of people who receive the most parking tickets? As shown in section 2.2 in Fig. 3, overall the ones who are frequent offenders are commercial vehicles. When analysis is restricted to passenger vehicles, however, we see that at low levels of ticket frequency (single offenders or those receiving between 2 and 4 tickets in the month) each violation category is fairly likely (Fig. 11). If passenger vehicles receive a lot of tickets (e.g., at least 34)[18], the violation is overwhelmingly likely to be for illegal stopping/parking or blocking traffic. This may be partially because fines like street cleaning can include a punishment of vehicle towing, which can be a much bigger hassle to deal with than just simply paying a fine. There is also no indication from the one months' worth of data that violations where PLATEFREQ are concentrated in either the earlier or later portion of the month.

[18] The boundary values for the PLATEFREQ division were created by a transformation, LN_PLATEFREQ=round(1+log(PLATEFREQ))

Fig. 10: Clustergram for 2,000 sample, by LN_PLATEFREQ

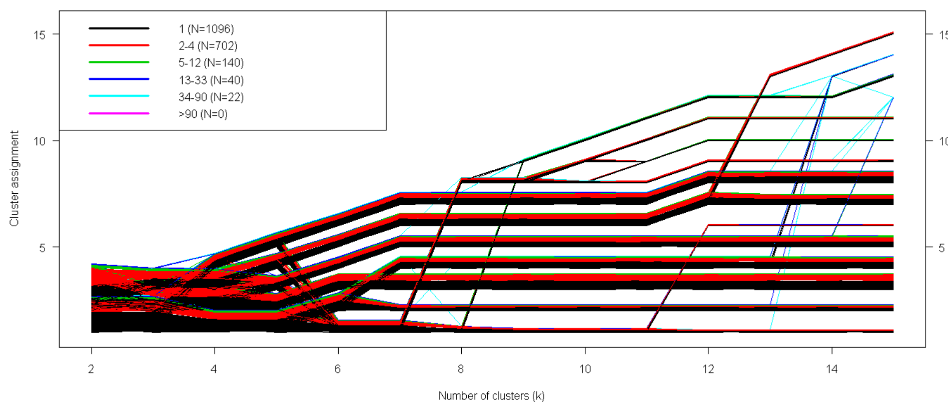
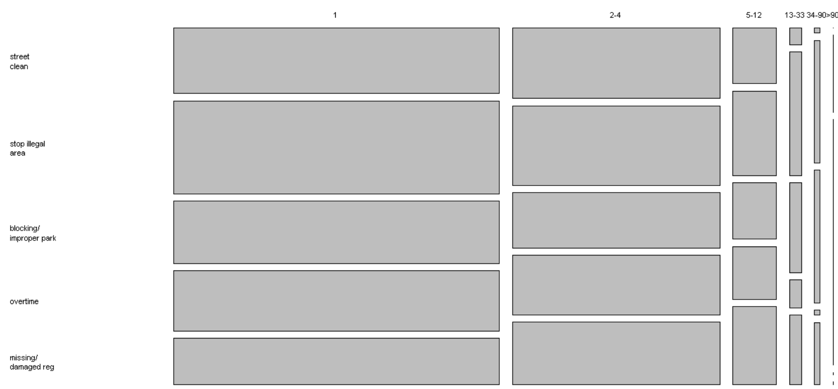


Fig. 11: Violation category by PLATEFREQ



As a driver in the city, one may also wish to know which days of the week are most risk in terms of receiving a specific violation. The clustergrams below (Figs. 12-13) show that the five main clusters are mixed by days of the week, except for Saturday and Sunday. Fig. 13 shows more clearly that at k=6, the weekend observations are split off from the five main clusters into a separate cluster number 1, along with two Friday tickets (these two are identical to each other in terms of the clustering attributes except for the fine). Together with Fig. 8, these clustergrams imply that while the violation category itself is the primary discriminating clustering attribute, while the day of the week is secondary once the initial split by violation type has been made. Certain violations are unlikely to occur on weekends, such as those involving parking meters and street cleaning (alternate side parking), which are not in effect on Sundays in the City.[19]

---

[19] NYCDOT. http://www.nyc.gov/html/dot/html/faqs/faqs_traffic.shtml

Fig. 12: Clustergram for 2,000 sample, by day of week

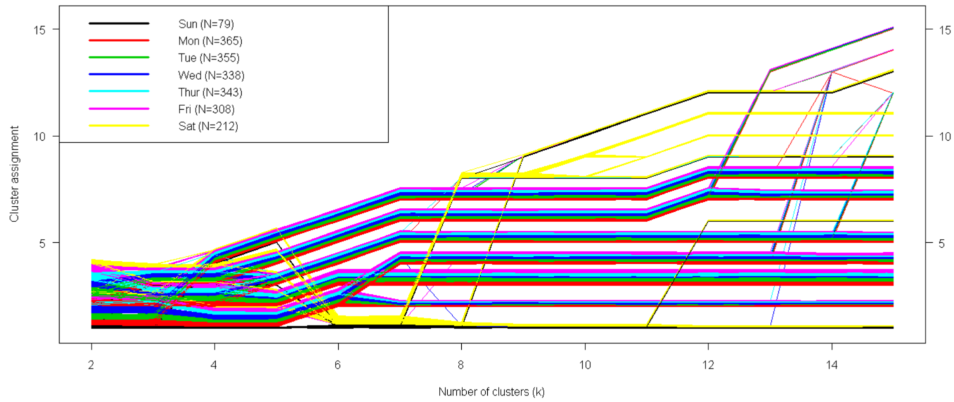
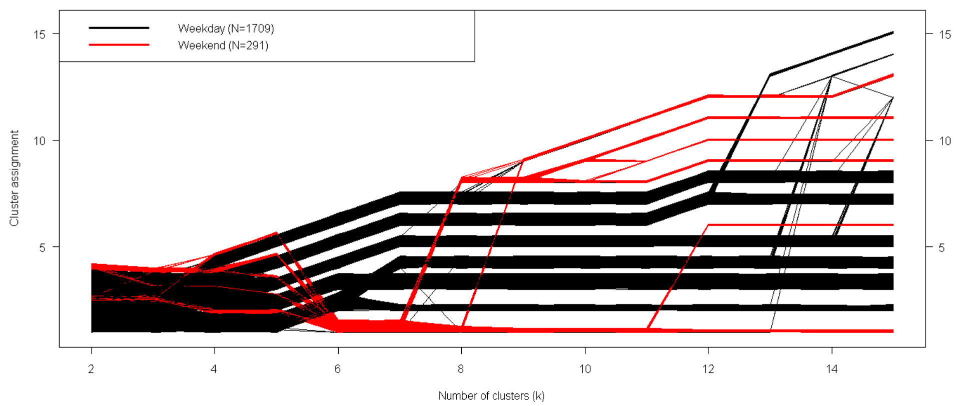


Fig. 13: Clustergram for 2,000 sample, by weekday/weekend



3. CONCLUSION

In a review of a month's worth of data on parking tickets in New York City, we have uncovered some basic facts about the types of violations that are most likely in certain times and places in the city. Tickets for stopping or parking in illegal zones, as well as usually violations for blocking traffic, such as double parking, tend to be the most common in areas with greater commercial concentration or traffic. Furthermore, when fine amounts are considered also, these two are the most risky for drivers in the City. These types of violations often occur as a result of a driver needing to stop temporarily in a busy area where parking is scarce; hence, they might try to park in a loading zone or double park for a few minutes to avoid having to search for a parking spot. The nature of the other three categories can be significantly different from these two because the other violations generally require a vehicle to be legally parked in the first place (such as at a meter). Finally, using clustering analysis and clustergrams, we demonstrated that the ticket observations can be separated into distinct groups based on the type of violation and the day of issuance; however, based on these attributes alone, the geographic location (here, the borough) cannot be identified with great certainty, especially once commerical vehicles, which operate mainly in areas such as downtown Manhattan, are removed from consideration.

Goals for further research in this data would include obtaining a panel dataset to study the hypothesis that officers ticket more at the end of a month to fill some monthly quota, for instance. Other interesting topics could involve studying attributes about the vehicle owner to a greater degree, such location of residence (the state of registration does not accurately capture this information necessarily) or income (perhaps using the posession of vanity plates as a proxy for this), or more details about the ticket, such as the hour it was issued. Some of these, however, are not available from the tickets themselves (such as demographics) or may require use of non-public databases to match license plates to owners, for instance. Lastly, in practical terms it would be desirable to apply the knowledge gained to assist with City operational or budgetary planning, such as allocating traffic enforcement officers to different regions.

SUPPLEMENTARY MATERIALS

Computer code: Perl, SAS, and R code used to clean dataset, create variables, and run analysis. The original dataset ("Parking Tickets") may be downloaded at http://www.nyc.gov/html/datamine/html/data/raw.shtml

ACKNOWLEDGEMENTS

I am greatly indebted to my professor Dr. Rida E. Moustafa, for his support and encouragement and for his time in discussing the various techniques employed in this paper. I would also like to thank all others who contributed helpful feedback for this paper, as well as the staff of NYCDatamine, in particular Albert Webber who assisted in communication with the Department of Finance. Lastly, I thank the Columbian College at the George Washington University for providing financial support to attend the 2011 Interface Symposium.

REFERENCES

Arnott, Richard and Eren Inci (2006), "An Integrated Model of Downtown Parking and Traffic Congestion," in *Journal of Urban Economics*, 60, 418-442.

Bloch, Matthew, and Amanda Cox. "A Year of Parking Tickets." *The New York Times*, 28 November 2008.
http://www.nytimes.com/interactive/2008/11/26/nyregion/20081128_PARKING.html

"Delivery Firms' Big Ticket Item: Parking Fines." Associated Press. September 1, 2006.
http://www.msnbc.msn.com/id/14602712/

Fisman, Raymond, and Edward Miguel (2007), "Corruption, Norms, and Legal Enforcement: Evidence from Diplomatic Parking Tickets," in *Journal of Political Economy*, 115 (6), 1020-1048.

Hernandez, Javier C. "Go Online, Not Downtown, to Fight a Parking Ticket." *The New York Times*, 21 March 2011.

McGinty, Jo Craven, and Ralph Blumenthal. "Adding to the City's Coffers, One Ticket at a Time." *The New York Times*, 28 November 2008.

McGinty, Jo Craven, and Ralph Blumenthal. "City Will Slash that Parking Fine, if You Ask." *The New York Times*, 1 January 2009.

New York City Department of Finance. "Got Tickets? Your Guide to Parking Ticket Hearings." http://www.nyc.gov/html/dof/html/pdf/07pdf/parking_guide.pdf

New York City Economic Development Corporation. "Mayor Bloomberg Announces Winners of Inaugural NYC BigApps Competition." February 5, 2010.
http://www.nycedc.com/PressRoom/PressReleases/Pages/NYCBigAppsCompetition.aspx

New York City Mayor's Office of Operations. "The Mayor's Management Report, Fiscal 2010." September 2010.
http://www.nyc.gov/html/ops/downloads/pdf/2010_mmr/0910_mmr.pdf

New York State Department of Motor Vehicles. "Vehicle Registration Class Codes."
http://www.nydmv.state.ny.us/regdup/rdpVehClass.htm

Schonlau, Matthias (2002), "The Clustergram: A Graph for Visualizing Hierarchical and Non-Hierarchical Cluster Analyses," in *The Stata Journal*, 2(4), 391-402.