homework iii

Antonio Neto, Lina Brilliantova, Vladislav Djuraskovic

October 22, 2021

Contents

Introduction

The primary source of government information and non emergency services is the New York City's 311. The objective is to provide easy access to all New York government services and information in a single place. To do this, they help Agencies improve service delivery by managing, filtering and redirecting occurencies from all New York City, and consequently, allowing them to focus on their core missions and manage their workload efficiently. Also, they provide insights to improve City government through accurate, consistent measurement and analysis of service delivery.

The objective of this work is to describe and visualize the dataset NYC311. The dataset contains almost 17 million individual service requests for the New York City 311 service from 2003 to 2015. Each request includes information about the service request type, location, description and additional details. These service requests are received by the NYC 311 and forwarded to the relevant agencies according to the type of service requested, such as the New York City Police Department, and many other. The agency responds to the request, addresses it and then close the ticket.

In this work we will firstly presents a description for each column of the data and then we do some exploratory data analyses using visual analytic. Finally, we will present some insights that we found by analyzing the data.

The head of the table

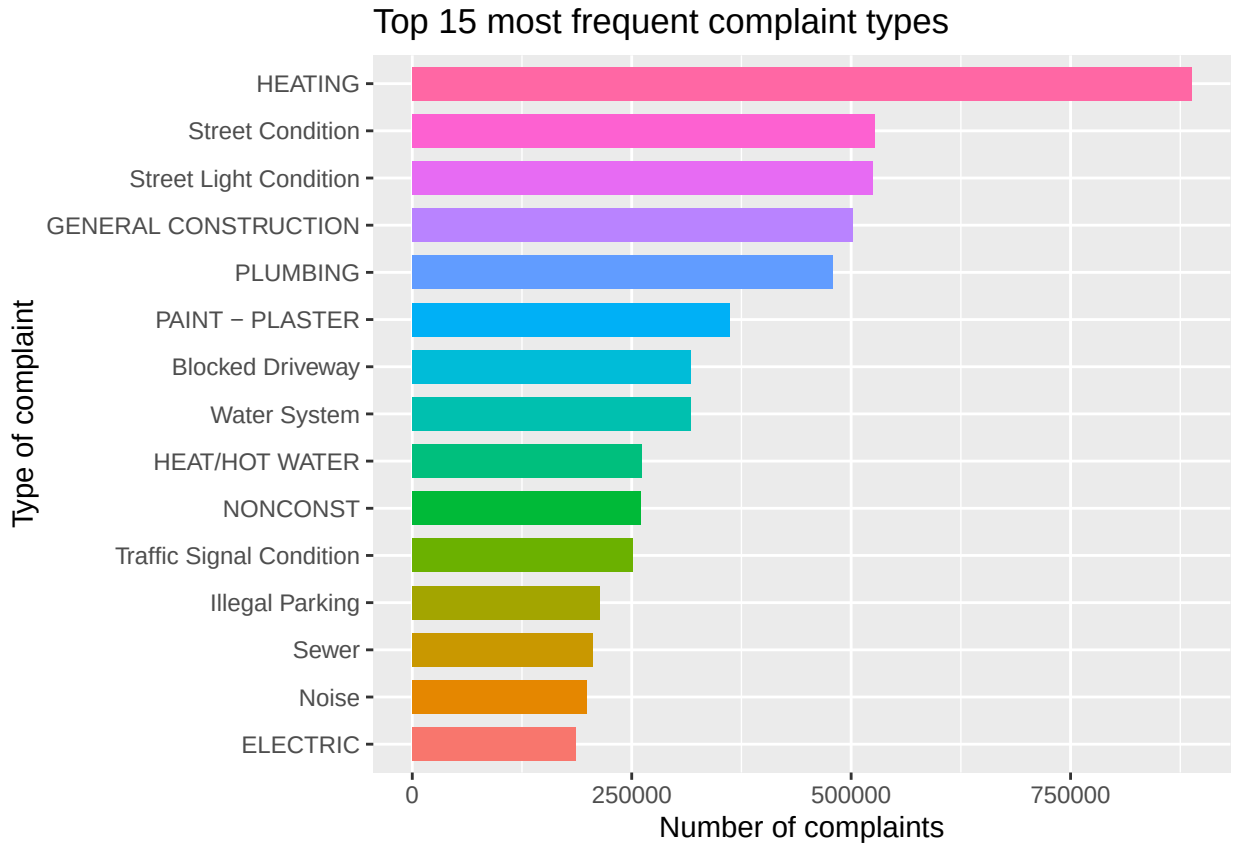
Here we produce a table of just some relevant columns of data.

Agency	AgencyName	ComplaintType	Descriptor	Status	Borough	IncidentZip	CreatedDate	ClosedDate	Month	Year	hour
NYPD	New York City Police Department	Vending	In Prohibited Area	Closed	BRONX	10465	2015-04-14 02:14:40	2015-04-14 03:03:22	4	2015	2
NYPD	New York City Police Department	Blocked Driveway	No Access	Open	BROOKLYN	11234	2015-04-14 02:10:12	NA	4	2015	2
NYPD	New York City Police Department	Noise - Street/Sidewalk	Loud Music/Party	Open	BROOKLYN	11204	2015-04-14 02:03:01	NA	4	2015	2
NYPD	New York City Police Department	Noise - Street/Sidewalk	Loud Talking	Assigned	BROOKLYN	11211	2015-04-14 02:02:40	NA	4	2015	2
NYPD	New York City Police Department	Noise - Street/Sidewalk	Loud Talking	Closed	MANHATTAN	10025	2015-04-14 02:00:04	2015-04-14 02:47:33	4	2015	2
NYPD	New York City Police Department	Noise - Street/Sidewalk	Loud Talking	Closed	BROOKLYN	11205	2015-04-14 01:52:15	2015-04-14 02:11:10	4	2015	1

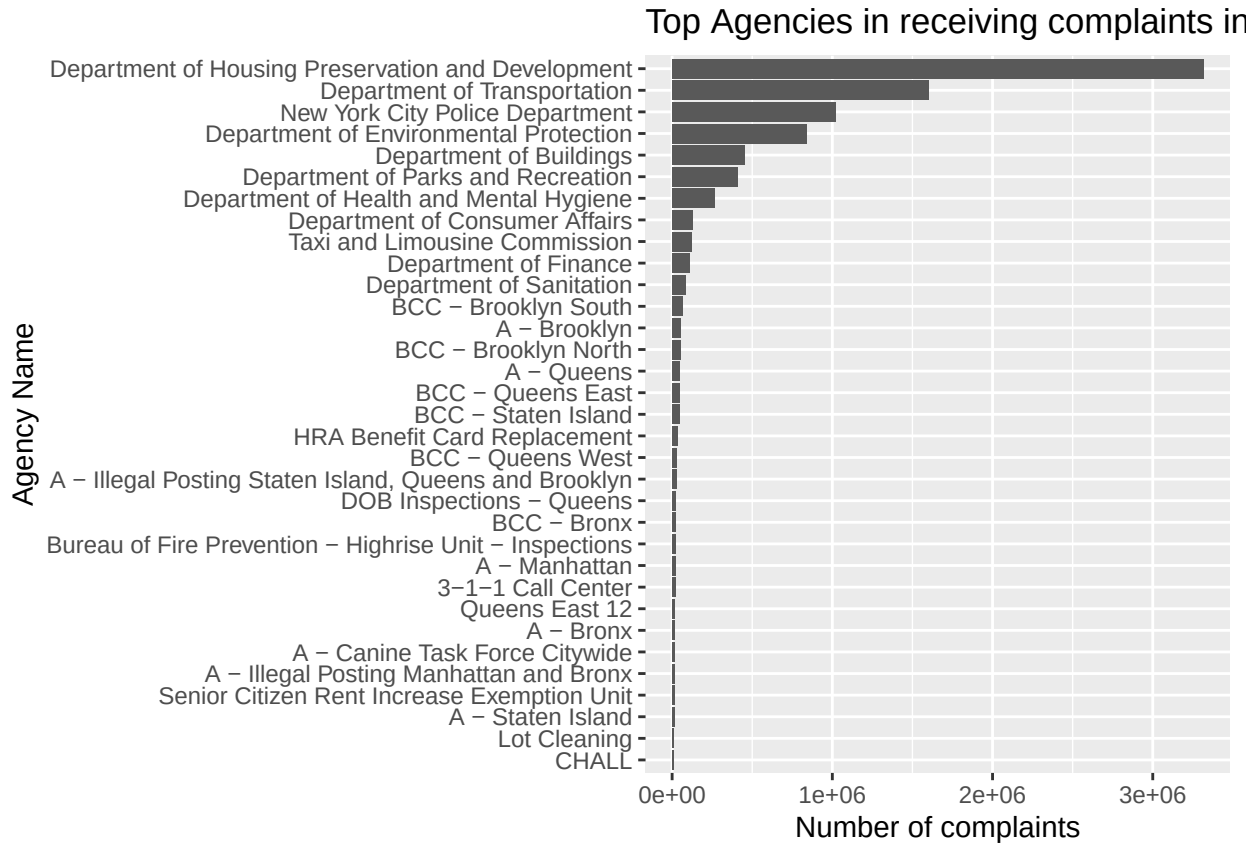
Exploration

Most popular requested services

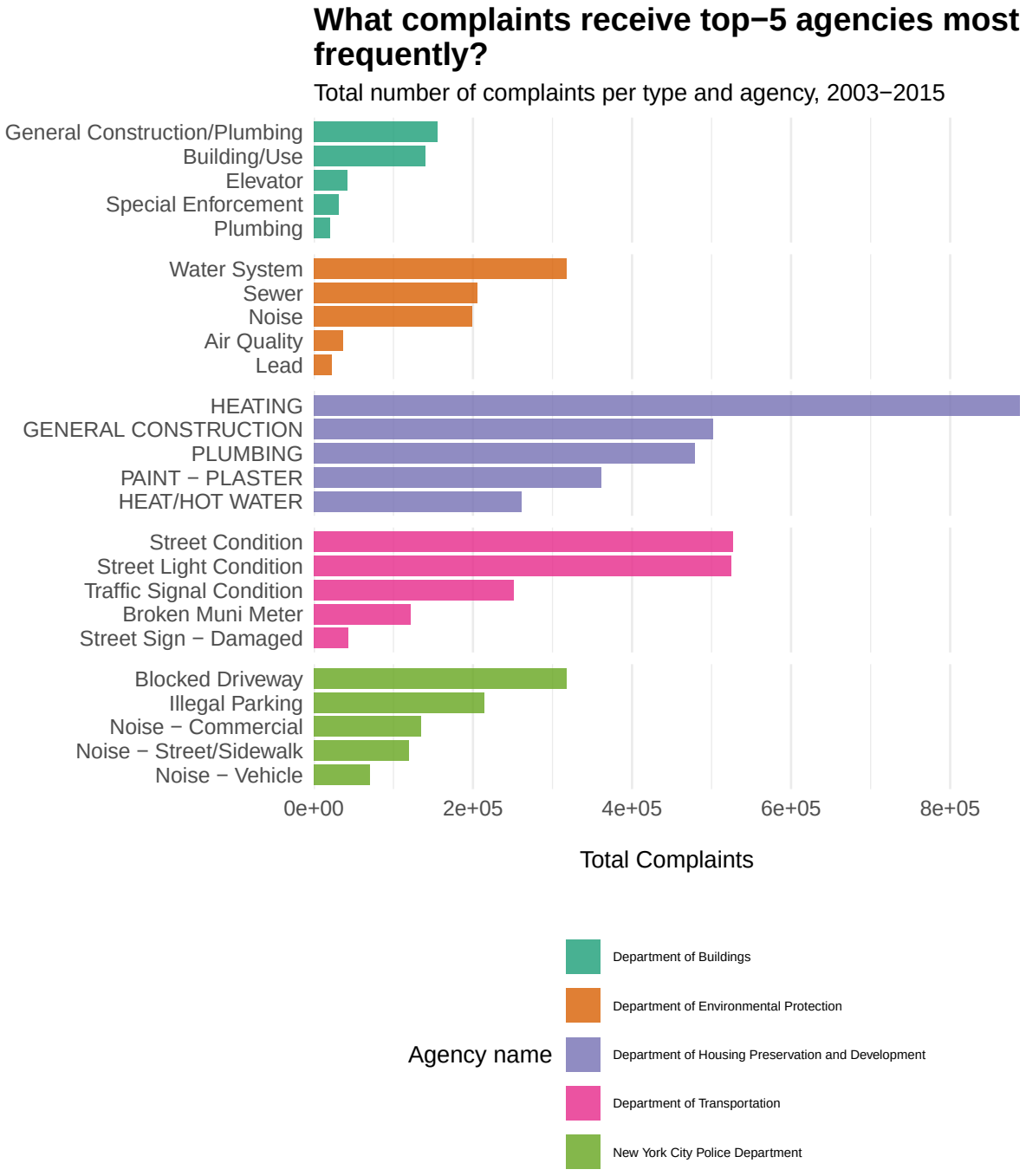
In the next graph we present the most requested services followed by a graph that is presenting the most requested agencies.



	variable_name	variable_description
1	AddressType	is it a street or an intersection
2	Agency	abbreviation for the agency name
3	AgencyName	full agency name
4	Borough	town
5	BridgeHighwayDirection	bridge highway direction
6	BridgeHighwayName	bridge highway name
7	BridgeHighwaySegment	bridge highway segment
9	ClosedDate	time of case dismissal
10	CommunityBoard	community board
11	ComplaintType	type of complaint
12	CreatedDate	time of a complaint
13	CrossStreet1	cross street 1
14	CrossStreet2	cross street 2
17	Descriptor	consequences of a complaint
18	DueDate	due date
19	FacilityType	type of an object where incident happened if it happened in the object
20	FerryDirection	ferry direction
21	FerryTerminalName	ferry terminal name
22	GarageLotName	garage lot name
23	IncidentAddress	adress where incident happened
24	IncidentZip	postal code
25	IntersectionStreet1	if incident happen at intersection what is the name of the first street
26	IntersectionStreet2	if incident happen at intersection what is the name of the second street
27	Landmark	object or something else that is close to place of an incident
28	Latitude	latitude
30	Location	location
32	LocationType	street, building, school or something else..
33	Longitude	longitude
34	ParkBorough	hood where park is
35	ParkFacilityName	Park name if there is one
37	ResolutionActionUpdatedDate	unclear
38	RoadRamp	road ramp
39	SchoolAddress	school address
40	SchoolCity	city where school is located
41	SchoolCode	school code
42	SchoolName	School Name if incident happened there
43	SchoolNotFound	school not found
44	SchoolNumber	School Number if incident happened there
45	SchoolorCitywideComplaint	schoolor complaint
46	SchoolPhoneNumber	school phone number
47	SchoolRegion	school region
48	SchoolState	school state
50	Status	status of an incident
51	StreetName	name of the street where incident happened
52	TaxiCompanyBorough	town for taxy company
53	TaxiPickUpLocation	where taxi picked up costumer
56	UniqueKey	id of a complaint
59	VehicleType	type of a vehicle
60	XCoordinate(StatePlane)	x coordinate
61	YCoordinate(StatePlane)	y coordinate



Complaints are distributed non-uniformly among the responsible agencies. There is a group of agencies that receive substantially larger number of requests: Department of Housing Preservation and Development (HPD), Department of Transportation (DOT), New York City Police Departments (NYPD), Department of Environmental protection (DEP), Department of Buildings (DOB), and Department of Parks and Recreation (DPR). Even among those 6-top agencies, the number of requests varies significantly: e.g. HDP is twofold different from DOT, and threefold of NYPD and DEP, with the absolute difference covering millions of requests. This makes HPD a highly demanding agency according to the data. It looks like the number of complaints per agency follows an exponential decay distribution with few highly demanded top agencies and a long tail of rarely used agencies.



Similar dynamics is observed in variability of requests within an agency, where few popular requests coexists with a long list of all kinds of other unpopular types. Interestingly, despite DOT is a second most popular agency, its top-5 complaint types combined are responsible for smaller number of requests compared to the rest of the agencies. This observation might imply more even spread of complaint types within that agency.

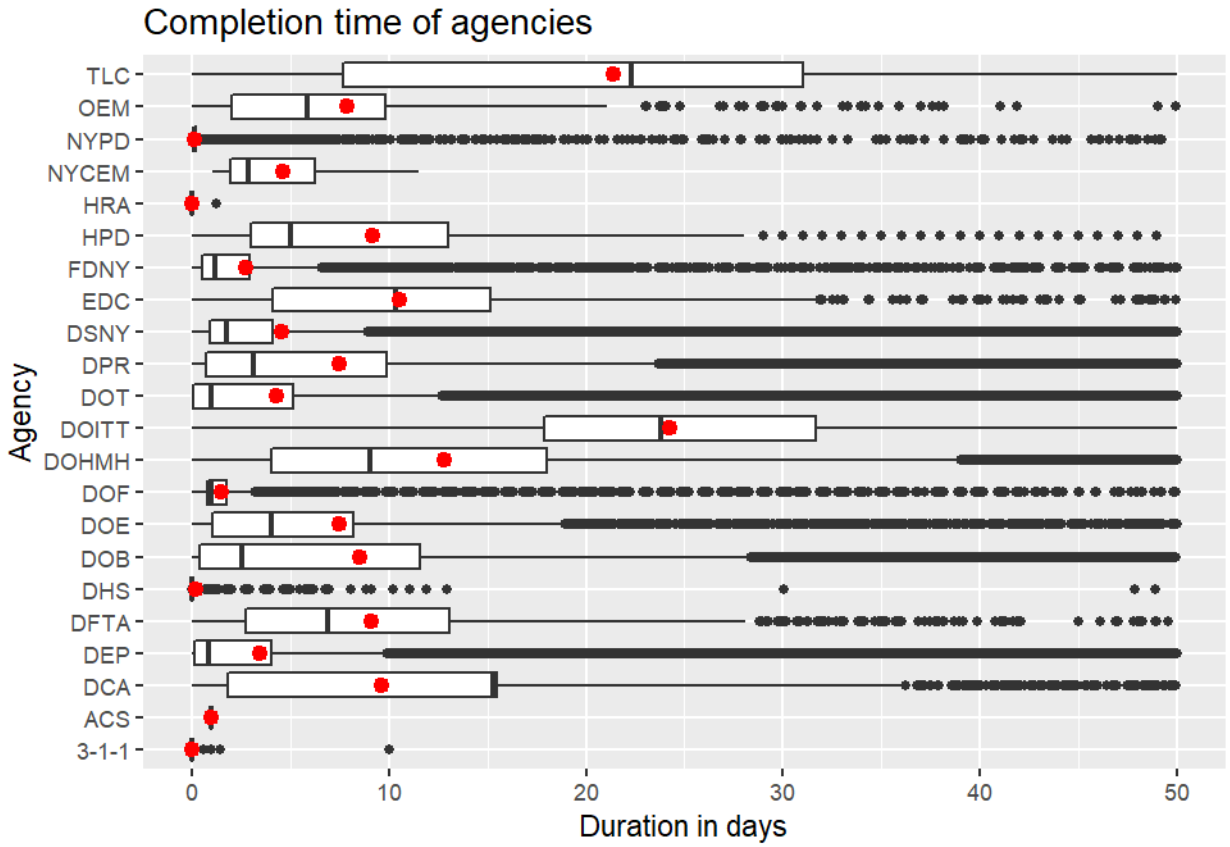
Analysis of completion time

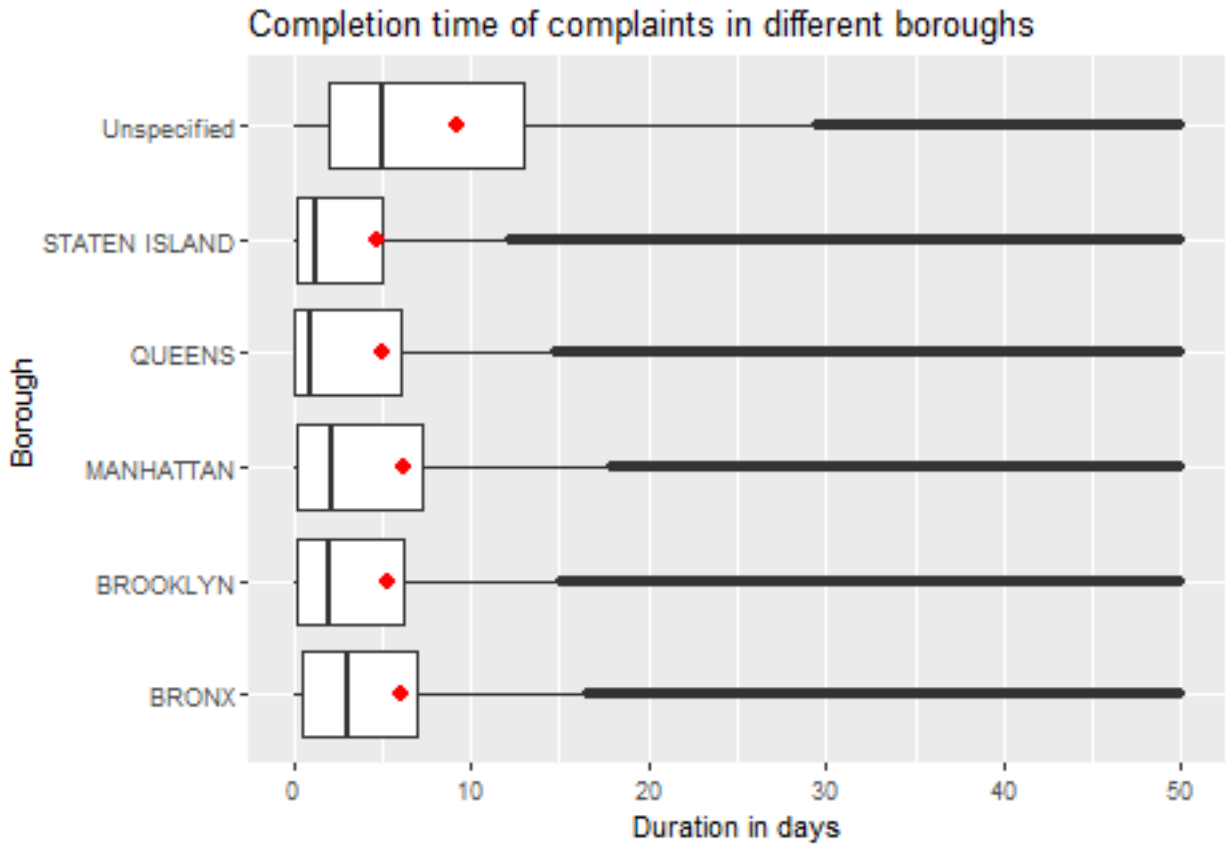
The following graphs presents the distribution of the time that each agency take to close an occurrence. We can observe that TLC and DOT are the agencies that took more time to close its issues, with an average

greater than 20 days to do it. Next we can observe that most agencies took between 5 to 10 days to close the occurrences. Finally, we observe that 5 agencies presents very fast responses, mostly less than a day.

As we completed this analysis for only cases with 'Closed' status, it is interesting to check the relationship between the completion rate (% of cases that are closed compared to 'Open', 'Pending' and 'Assigned') and the average completion time (this analysis will be made in the next subsection).

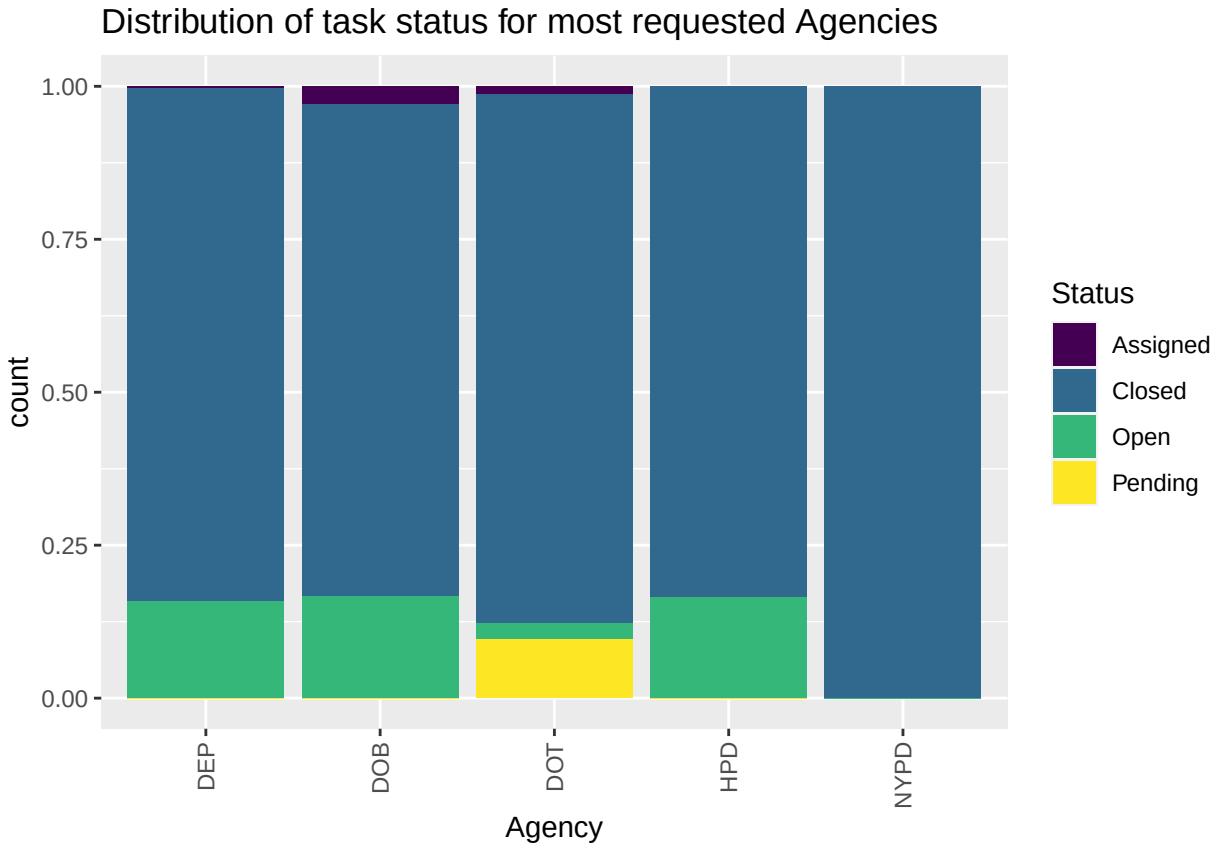
The difference between completion time between boroughs is minimal, which might suggest a high level of centralization in completing requests.





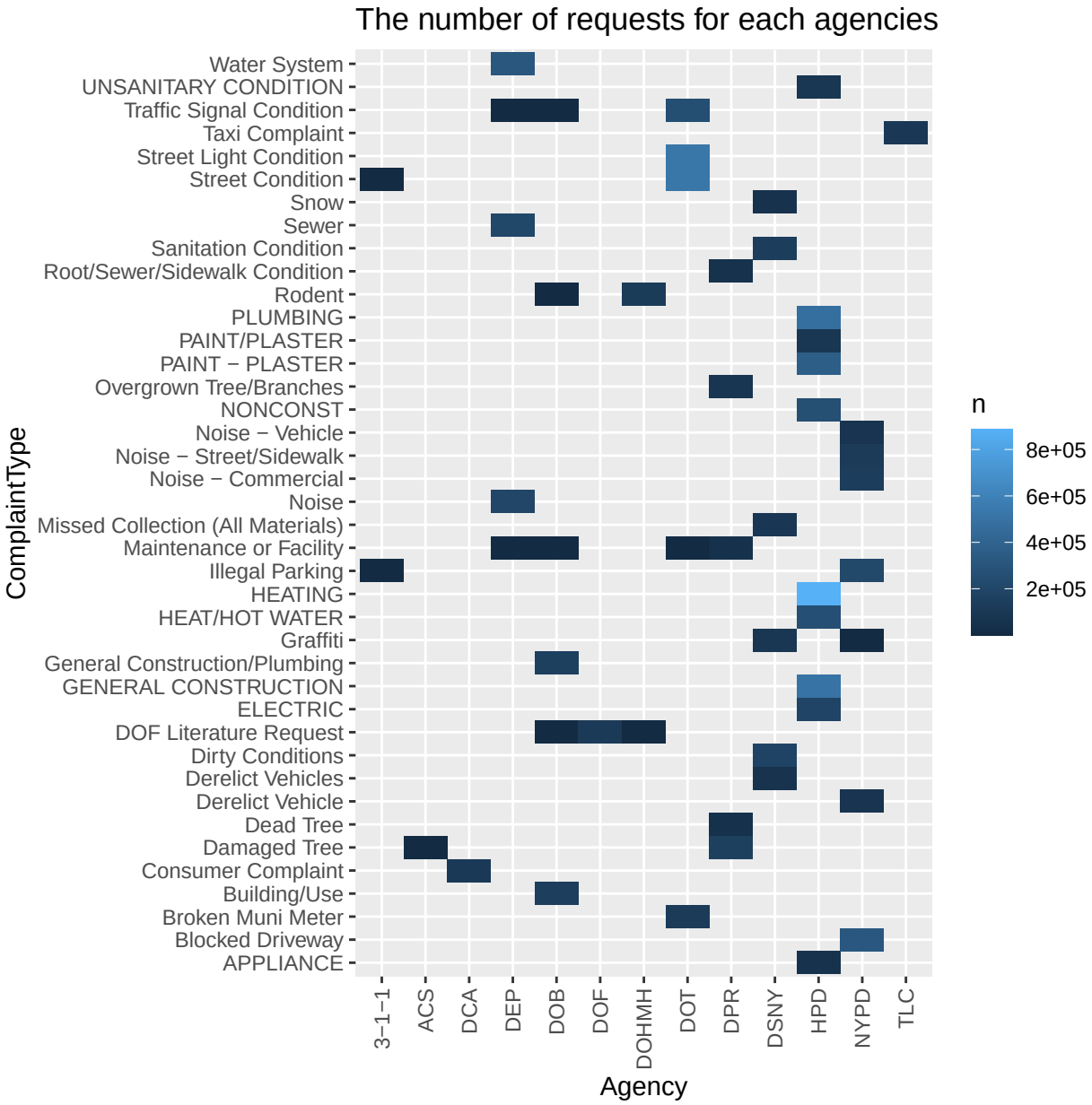
Analysis of completion rate

The following plot present the different of task status distribution over top agencies. We can see that DOT presents a lot of pending task compared with the other agency, while NYPD is the agency that has the most closed tasks over the top agencies.

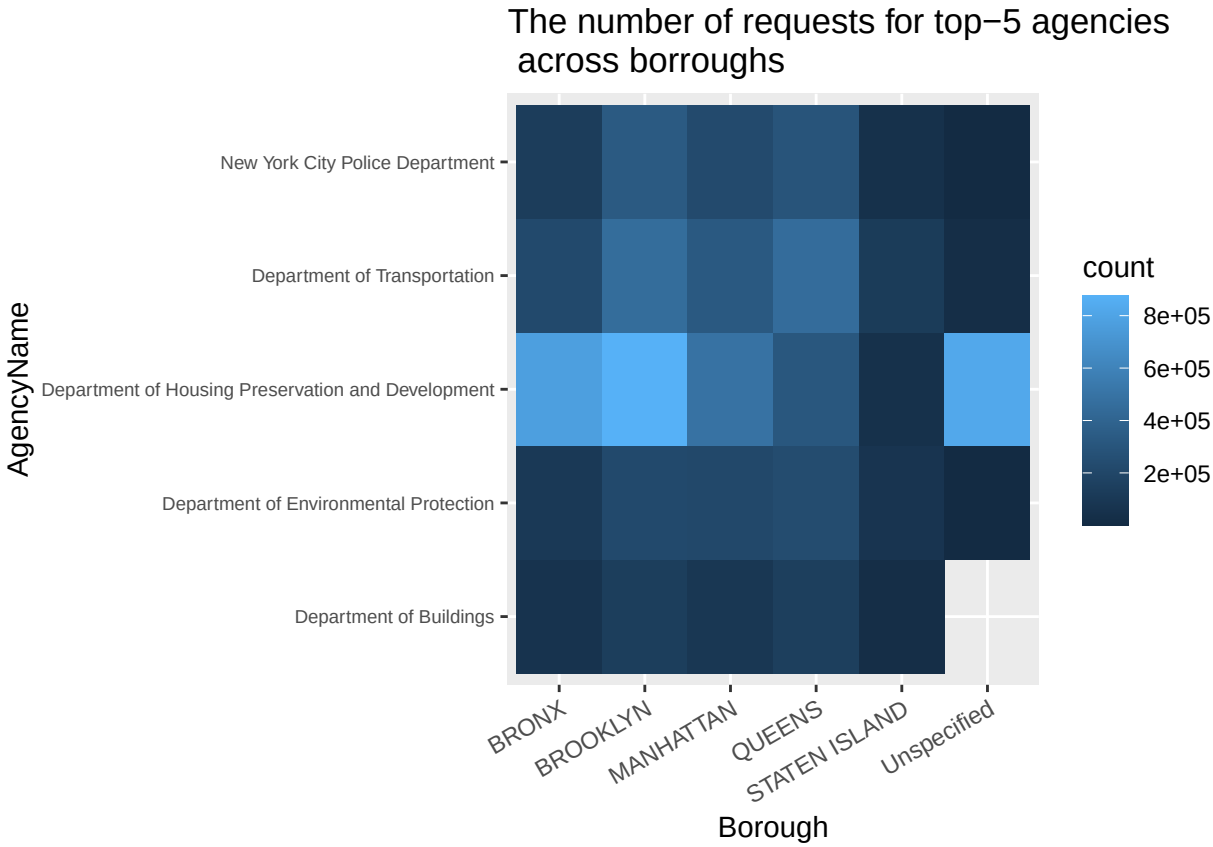


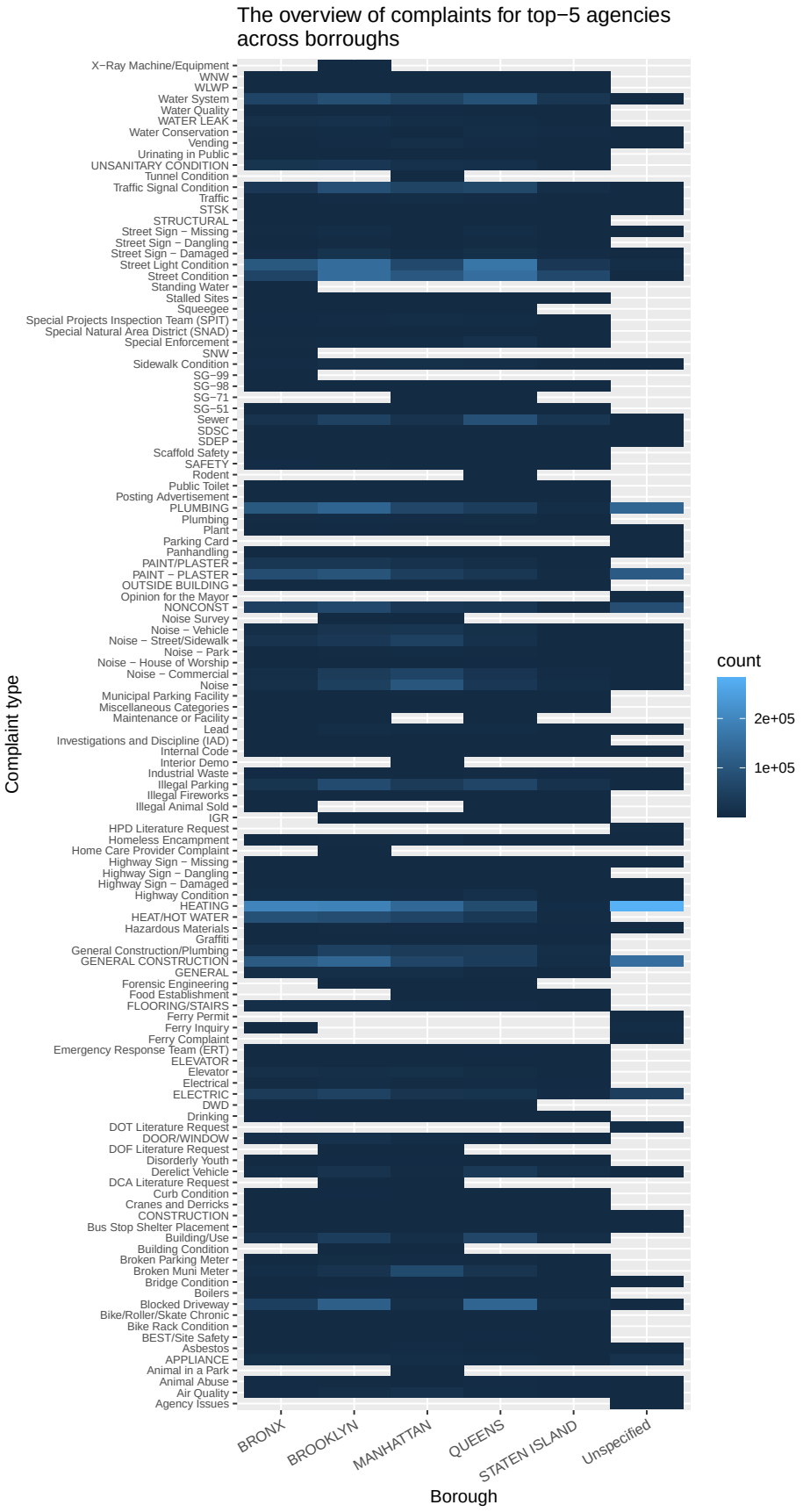
Heat maps

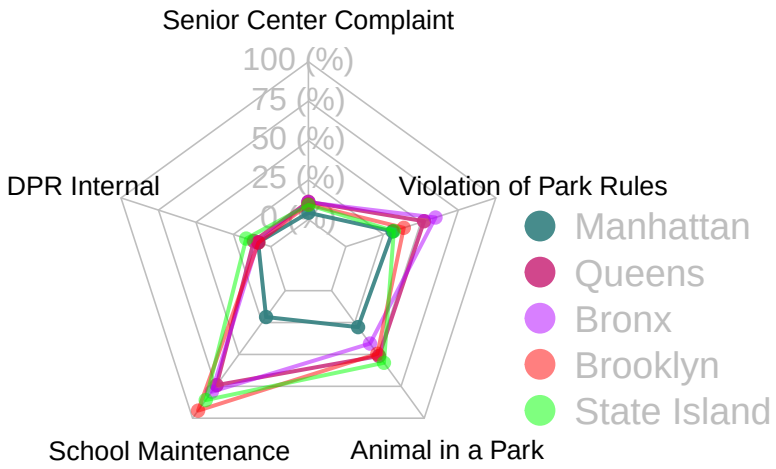
The following plot presents the heat map for The number of request for each agencies, we can see that each agency specializes in some specific tasks, but more than one agency can present work with the same type of occurrence.



For the next plots, we can observe how is the distribution of request among the agencies for each borrough.


The number of requests for top−5 agencies across borroughs
New York City Police Department
Department of Transportation
Department of Housing Preservation and Development
Department of Environmental Protection
Department of Buildings
AgencyName
BRONX
BROOKLYN
MANHATTAN
QUEENS
STATEN ISLAND
Unspecified
Borough
count
8e+05
6e+05
4e+05
2e+05


The overview of complaints for top−5 agencies across borroughs
X−Ray Machine/Equipment
WNW
WLWP
Water System
Water Quality
WATER LEAK
Water Conservation
Vending
Urinating in Public
UNSANITARY CONDITION
Tunnel Condition
Traffic Signal Condition
Traffic
STSK
STRUCTURAL
Street Sign − Missing
Street Sign − Dangling
Street Sign − Damaged
Street Light Condition
Street Condition
Standing Water
Stalled Sites
Squeegee
Special Projects Inspection Team (SPIT)
Special Natural Area District (SNAD)
Special Enforcement
SNW
Sidewalk Condition
SG−99
SG−98
SG−71
SG−51
Sewer
SDSC
SDEP
Scaffold Safety
SAFETY
Rodent
Public Toilet
Posting Advertisement
PLUMBING
Plumbing
Plant
Parking Card
Panhandling
PAINT/PLASTER
PAINT − PLASTER
OUTSIDE BUILDING
Opinion for the Mayor
NONCONST
Noise Survey
Noise − Vehicle
Noise − Street/Sidewalk
Noise − Park
Noise − House of Worship
Noise − Commercial
Noise
Municipal Parking Facility
Miscellaneous Categories
Maintenance or Facility
Lead
Investigations and Discipline (IAD)
Internal Code
Interior Demo
Industrial Waste
Illegal Parking
Illegal Fireworks
Illegal Animal Sold
IGR
HPD Literature Request
Homeless Encampment
Home Care Provider Complaint
Highway Sign − Missing
Highway Sign − Dangling
Highway Sign − Damaged
Highway Condition
HEATING
HEAT/HOT WATER
Hazardous Materials
Graffiti
General Construction/Plumbing
GENERAL CONSTRUCTION
GENERAL
Forensic Engineering
Food Establishment
FLOORING/STAIRS
Ferry Permit
Ferry Inquiry
Ferry Complaint
Emergency Response Team (ERT)
ELEVATOR
Elevator
Electrical
ELECTRIC
DWD
Drinking
DOT Literature Request
DOOR/WINDOW
DOF Literature Request
Disorderly Youth
Derelict Vehicle
DCA Literature Request
Curb Condition
Cranes and Derricks
CONSTRUCTION
Bus Stop Shelter Placement
Building/Use
Building Condition
Broken Parking Meter
Broken Muni Meter
Bridge Condition
Boilers
Blocked Driveway
Bike/Roller/Skate Chronic
Bike Rack Condition
BEST/Site Safety
Asbestos
APPLIANCE
Animal in a Park
Animal Abuse
Air Quality
Agency Issues
Complaint type
BRONX
BROOKLYN
MANHATTAN
QUEENS
STATEN ISLAND
Unspecified
Borough
count
2e+05
1e+05



In the spider graph we can analyse the top complains that comes from schools in different boroughs. As we observe, School Maintence is the most common complain over all the borough schools, but Manhattan is an exception. We can see that Manhattan presents a different distribution of complains compared to other boroughs, we can see that the most common complais for these schools are Violation of Park Rules.

Table: Top-5 agency by the number of complaints received in 2003-2009

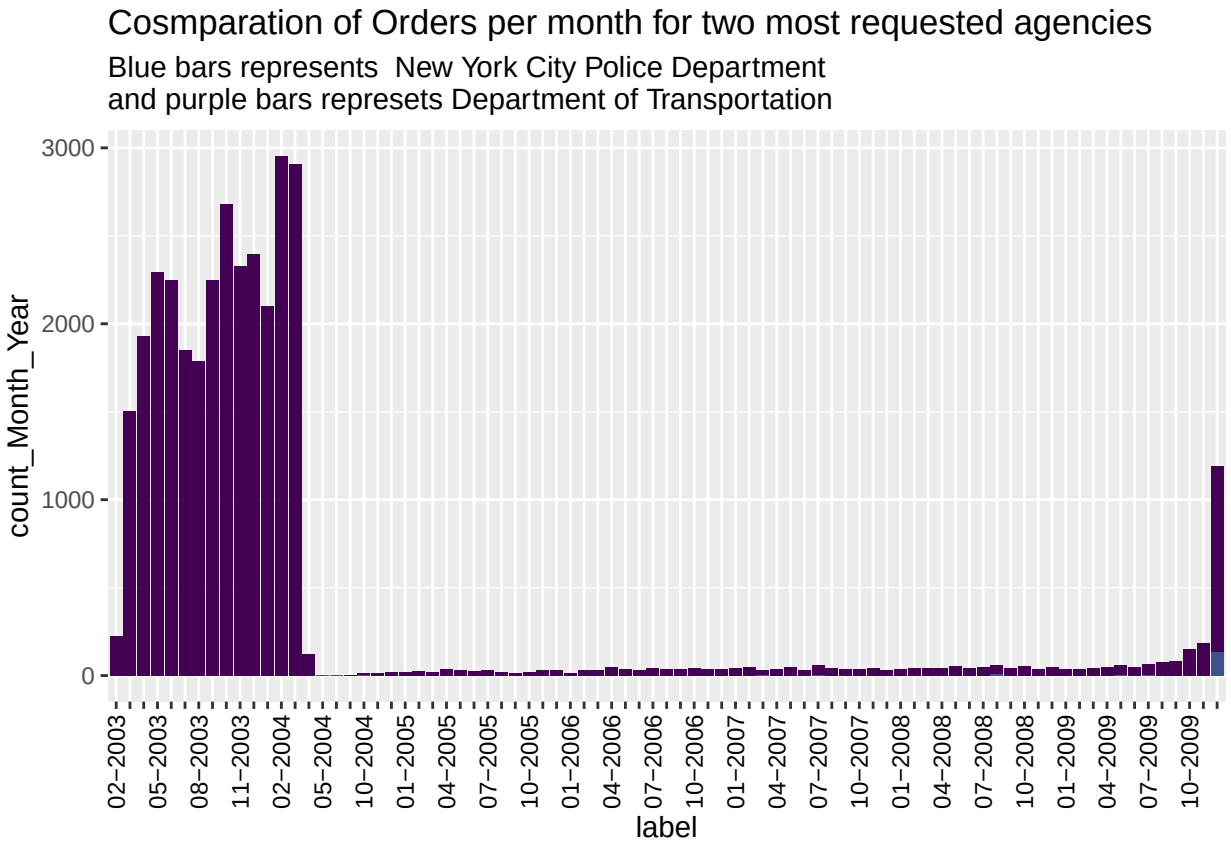
	AgencyName	count
1	Department of Transportation	33244
2	Department of Consumer Affairs	13094
3	Department of Health and Mental Hygiene	10178
4	Department of Parks and Recreation	7369
5	Taxi and Limousine Commission	5018

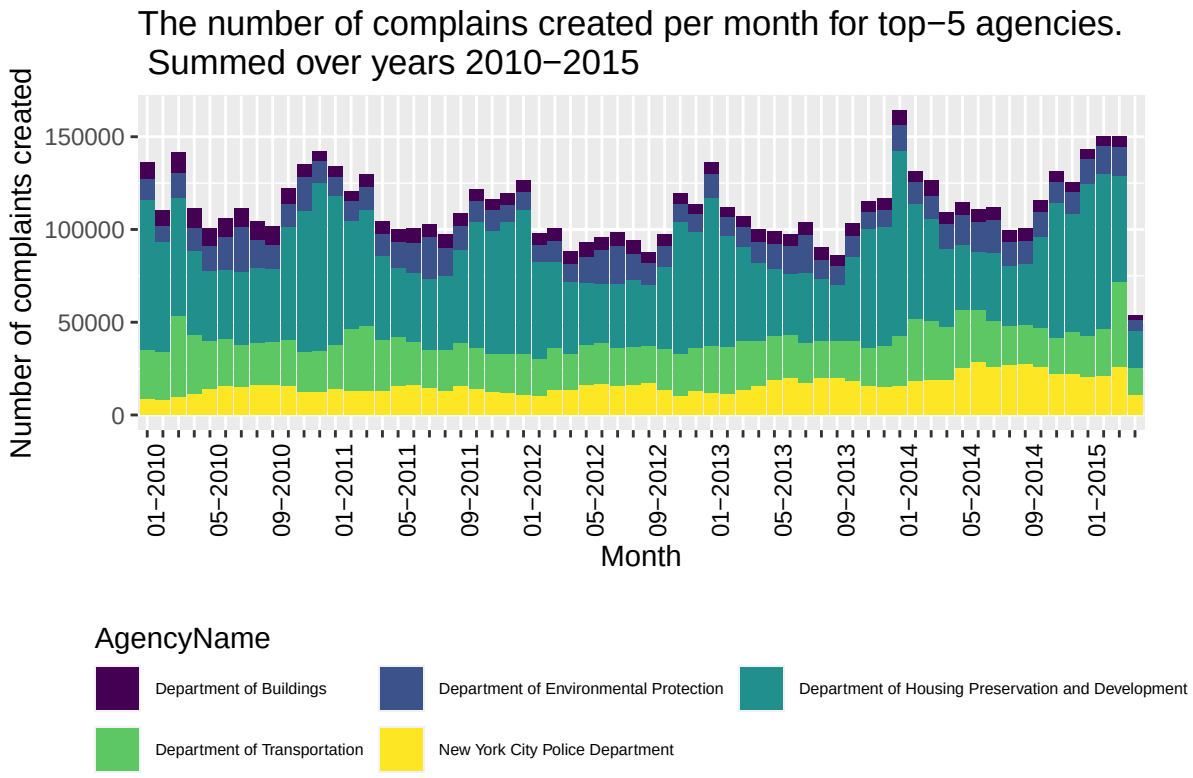
Table: Top-5 agency by the number of complaints received in 2010-2015

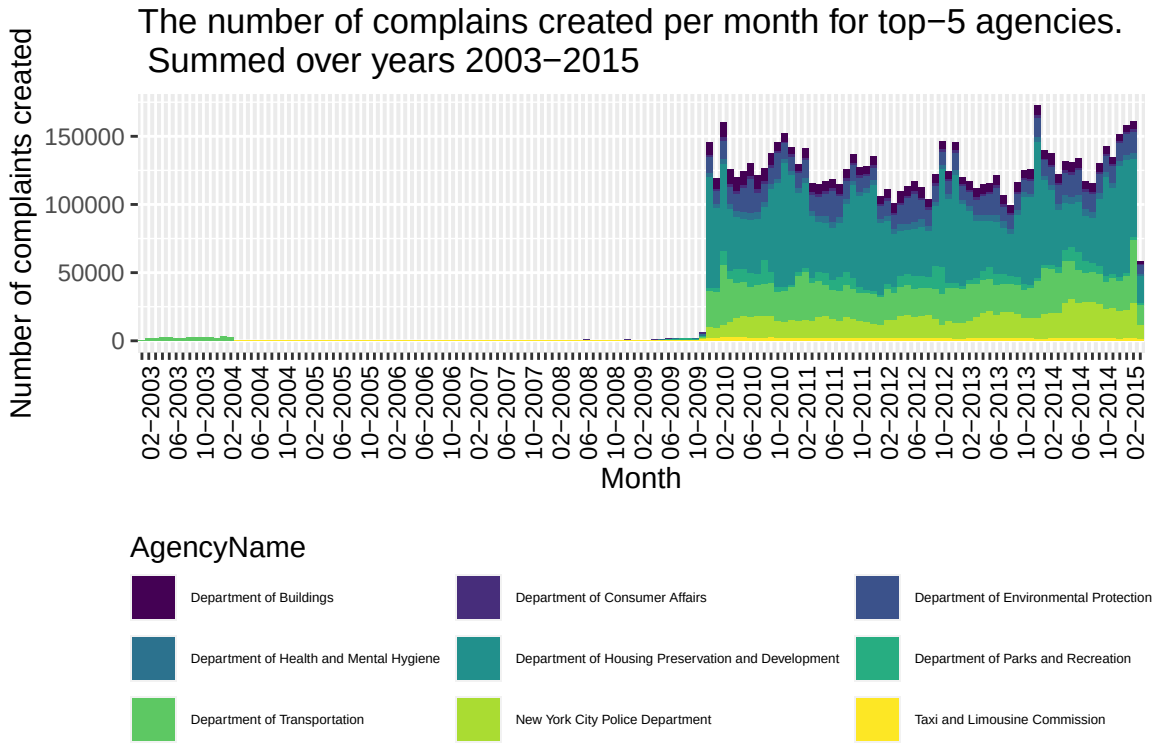
	AgencyName	count
1	Department of Housing Preservation and Development	3318655
2	Department of Transportation	1568151
3	New York City Police Department	1022984
4	Department of Environmental Protection	838774
5	Department of Buildings	452859

Yearly dynamics

We figured out that the dataset can be clearly divided into two disjoint time periods varying by their top-5 most popular services.

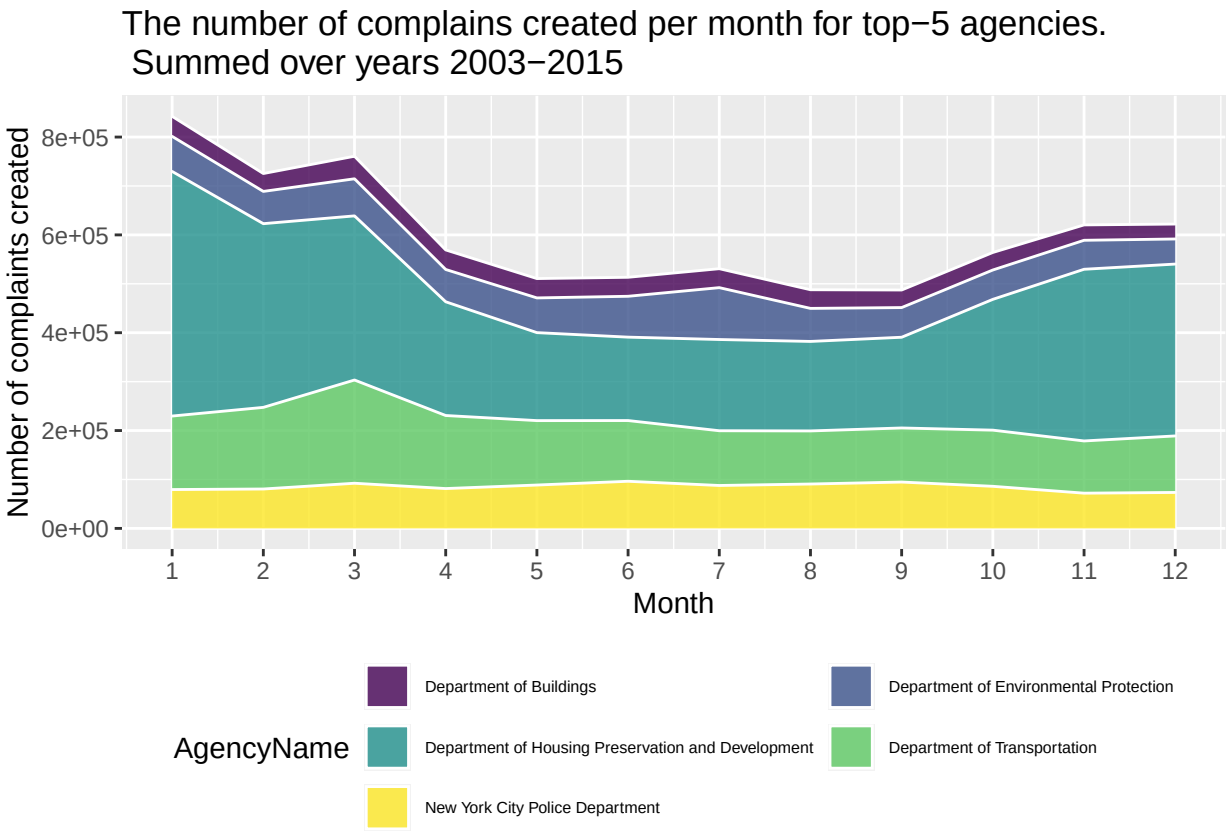



The number of complains created per month for top−5 agencies.
Summed over years 2010−2015
Number of complaints created
150000
100000
50000
0
01−2010
05−2010
09−2010
01−2011
05−2011
09−2011
01−2012
05−2012
09−2012
01−2013
05−2013
09−2013
01−2014
05−2014
09−2014
01−2015
Month
AgencyName
Department of Buildings
Department of Environmental Protection
Department of Housing Preservation and Development
Department of Transportation
New York City Police Department



Seasonal dynamics

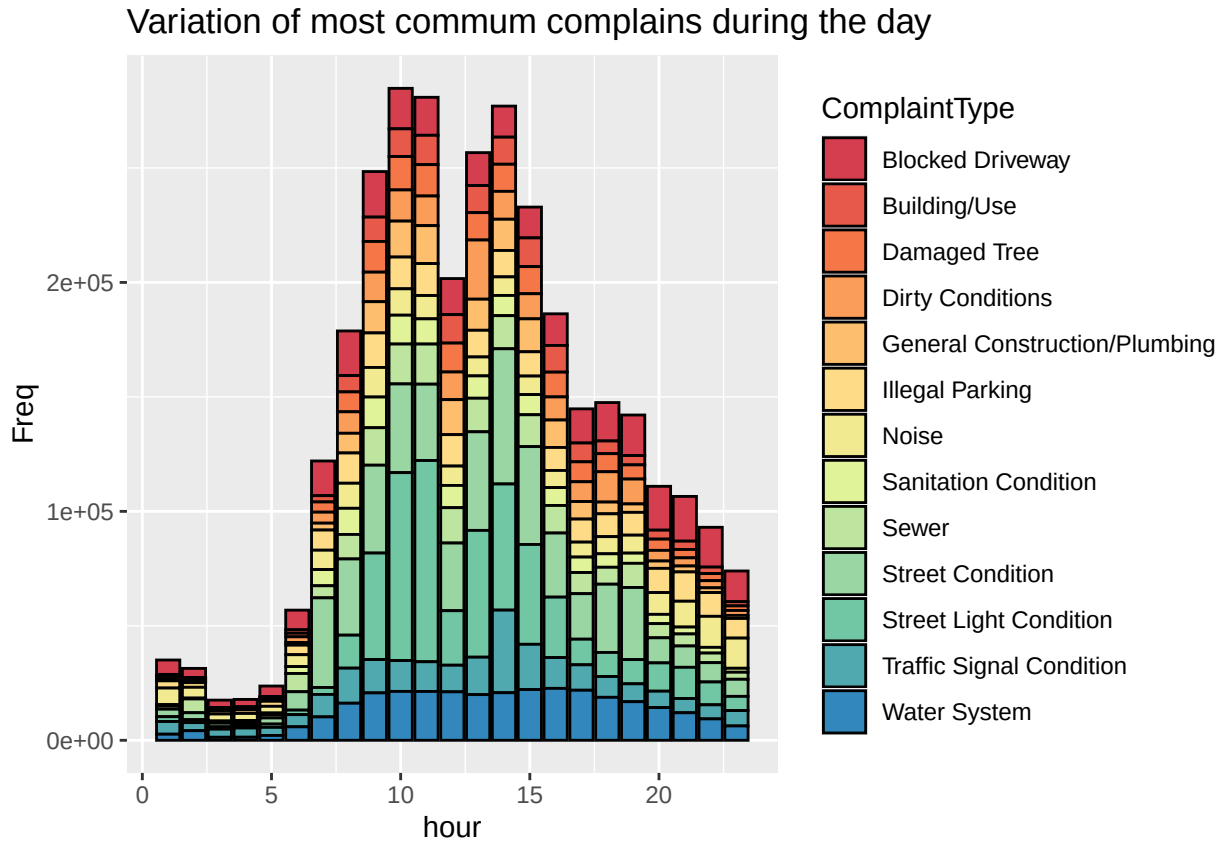
The number of complaints varies for different seasons, and the seasonal dynamics depends on the agency. Departments of Buildings, Environmental protection and Transportation have peaks of requests on March, which are gradually decrease over the summer and starting to rise with the beginning of cold season (October-November). On the contrast, police department (and to some extent Department of Housing Preservation and Development) has almost constant amount of cases, that do not have seasonal variability.

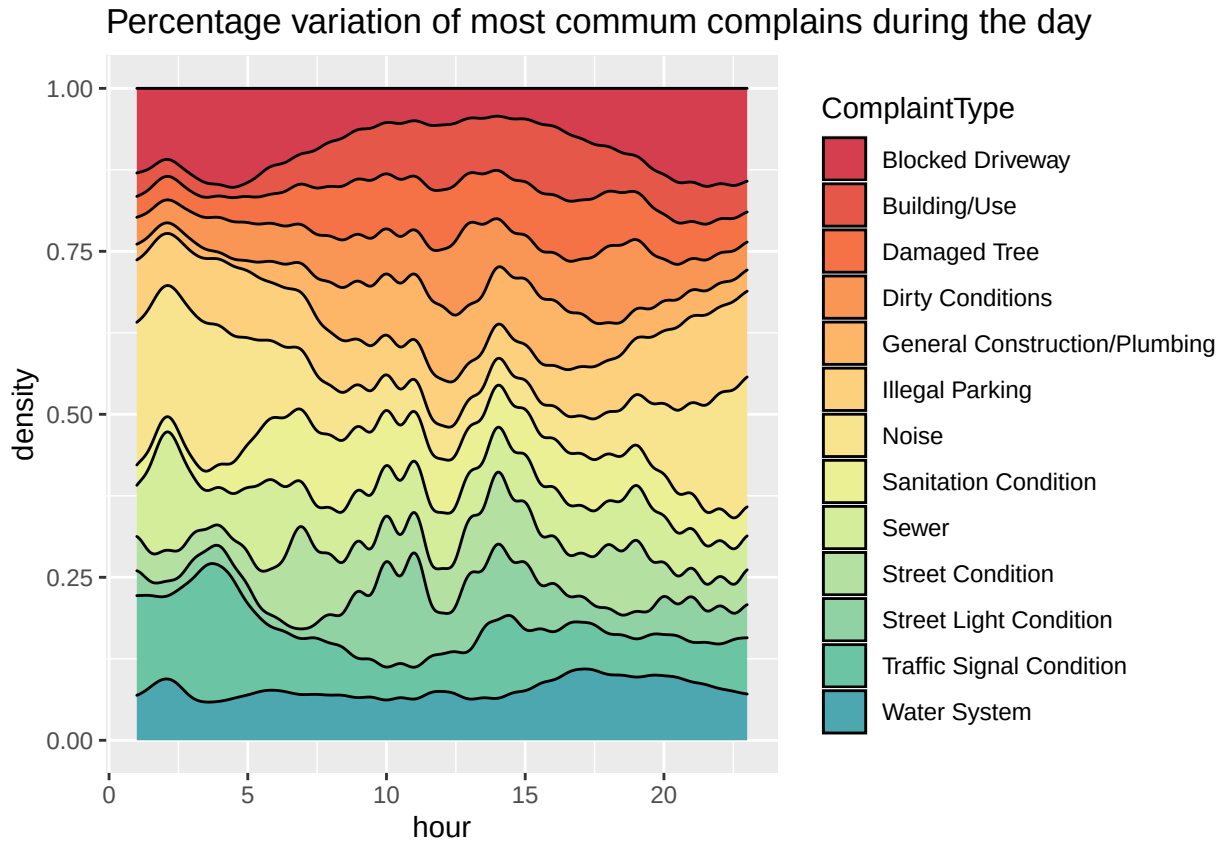


Daily dynamics

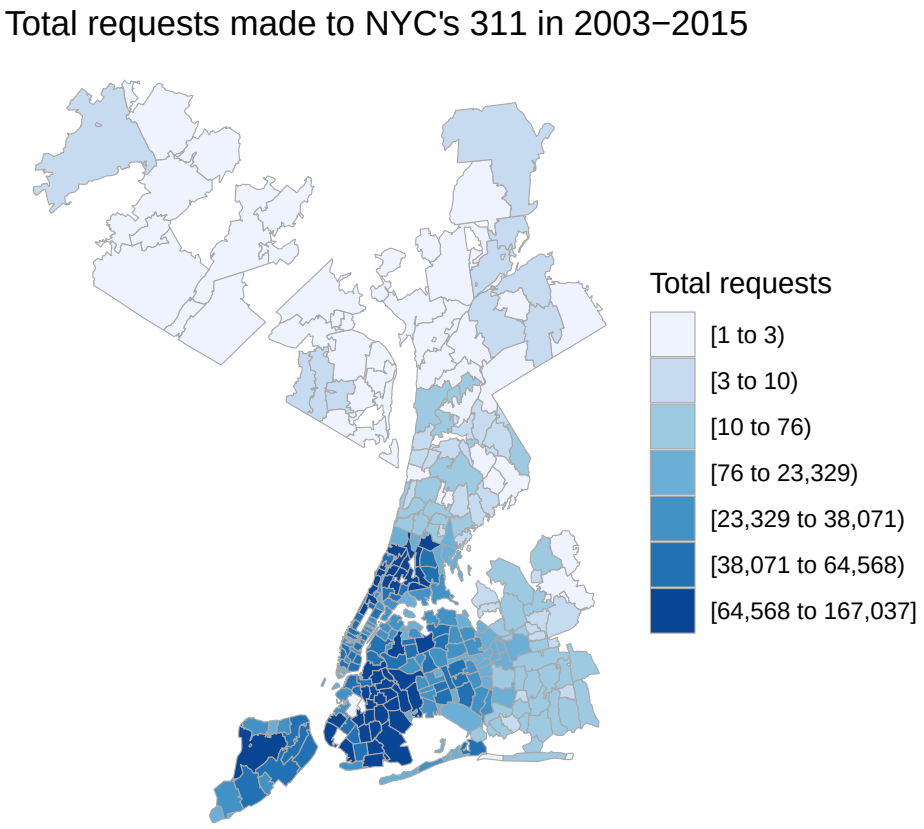
We also analysed the frequency of complaint types depending on hour and found some interesting patterns. Firstly, for the first plot, we can easily observe that most of complains happens in between 10am and 6pm, while between 0am and 5am is the range with the lower volume of complains .

Looking for the second plot, as expected, most driving-related issues have their peaks during the day and are minimal during the night. They also have peaks around 10 am and 3 pm. we observe this kind of behavior for Street Condition, Street Light Condition, Traffic signal condition, General construction, Dirty condition and damage trees complaint types. On the contrary, Noise and (unexpectedly) blocked driveway have their peaks in late evening and early morning.


Variation of most commum complains during the day
Freq
2e+05
1e+05
0e+00
0
5
10
15
20
hour
ComplaintType
Blocked Driveway
Building/Use
Damaged Tree
Dirty Conditions
General Construction/Plumbing
Illegal Parking
Noise
Sanitation Condition
Sewer
Street Condition
Street Light Condition
Traffic Signal Condition
Water System

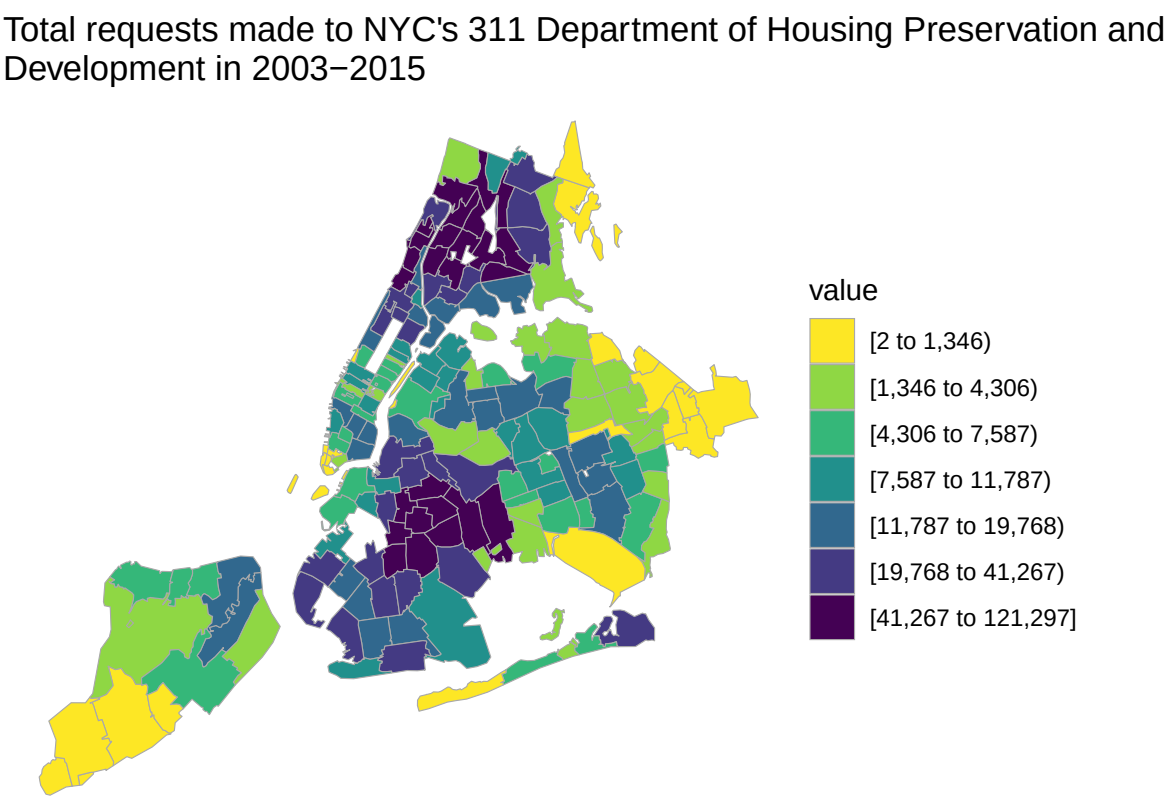

Percentage variation of most commum complains during the day
density
1.00
0.75
0.50
0.25
0.00
0
5
10
15
20
hour
ComplaintType
Blocked Driveway
Building/Use
Damaged Tree
Dirty Conditions
General Construction/Plumbing
Illegal Parking
Noise
Sanitation Condition
Sewer
Street Condition
Street Light Condition
Traffic Signal Condition
Water System

Map analysis

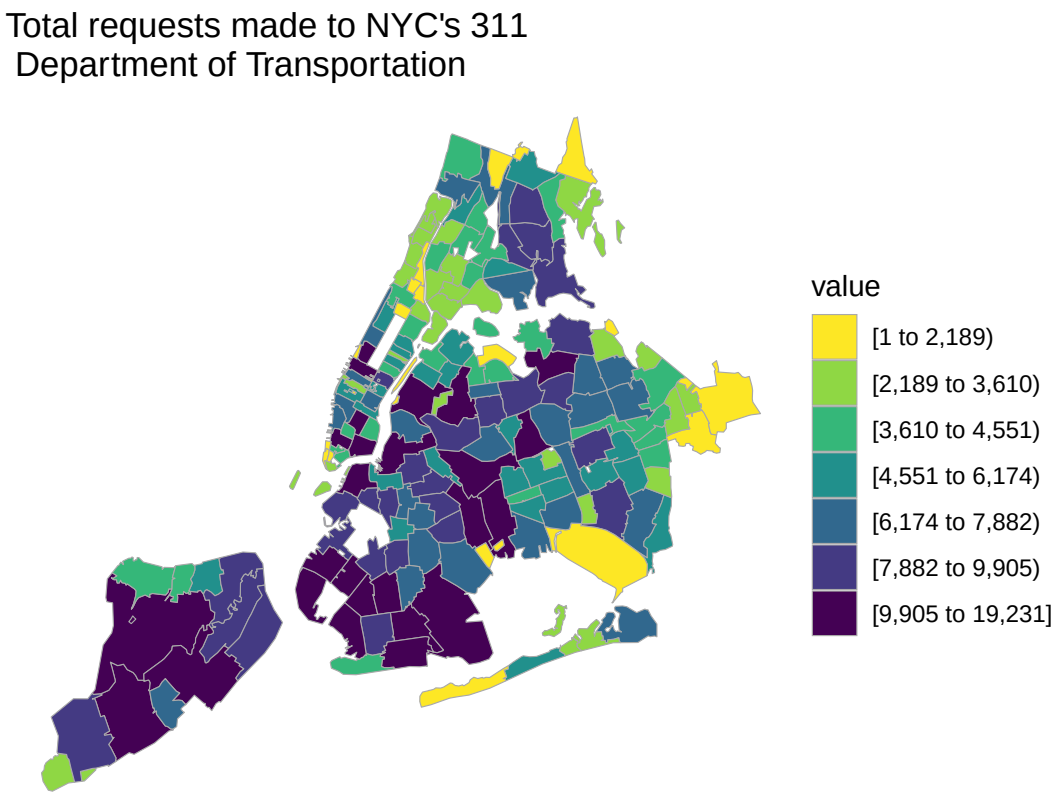


On the map of total requests made to NYC311 in 2003-2015 there are three clusters with the largest number of complaints: a cluster in Staten Island (including Freshkills, CUNY College of Staten Island and New Springville neighborhoods), a cluster in Manhattan/Bronx (including Yorkville, Harlem and full Bronx neighborhoods), and a large cluster located in Brooklyn. There are also several isolated areas in Queens near LaGuardia Airport with high number of complaints.

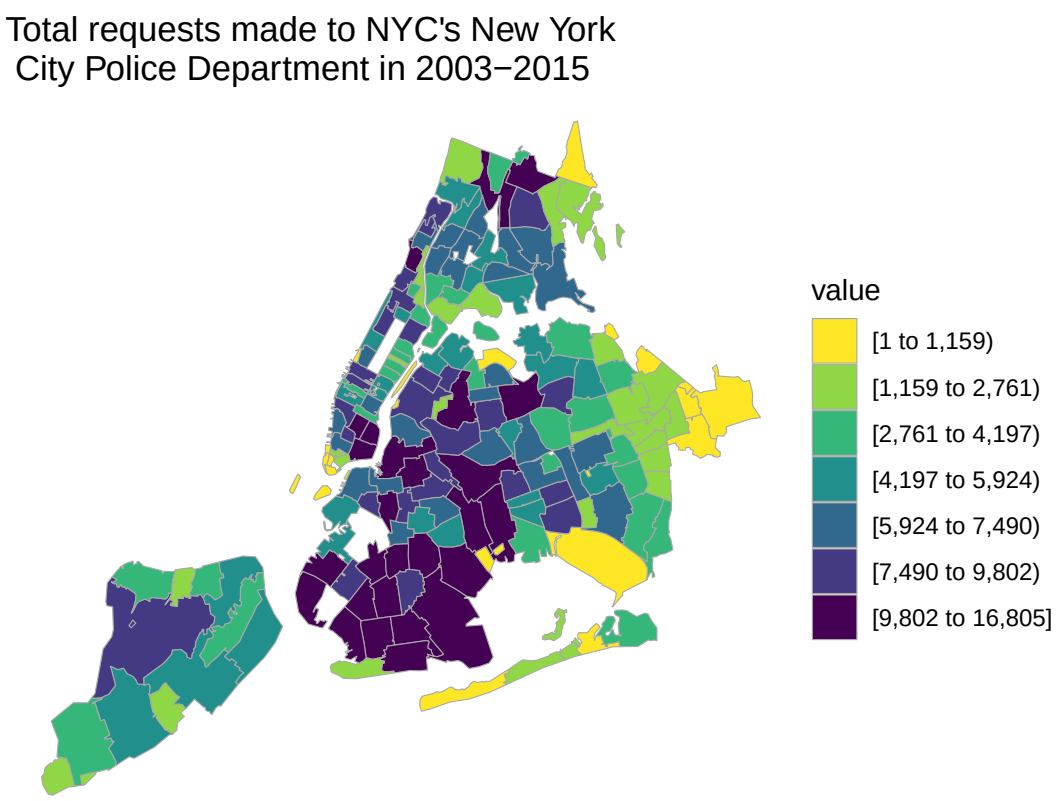
Next we analyze the variability in the number of complaints requested from top-5 agencies. DHPD requests are concentrated in Brooklyn and Bronx - two residential borroughs. The most distant areas (including for some reason LaGuardia airport area) from the NYC center had the less number of complaints. There was not a lot of complaints in the Staten island for DPHD.

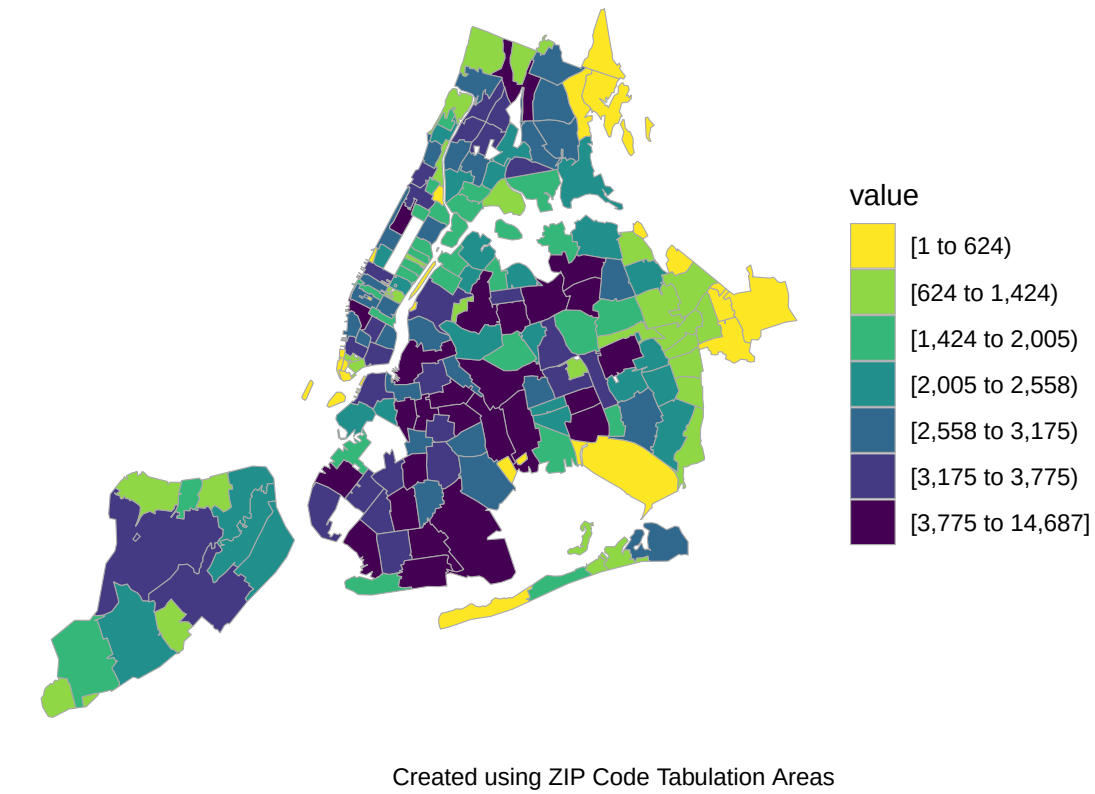


“‘ Brooklyn and Queens had the largest number of requests to DOT agency, as well as the Staten Island near the CUNY College. Manhaettan has one of the least number of complaints to DOT as well as distant areas.



The large cluster in the number of NYPD service (as well DOB services, as the maps are almost identical) requests is located at Brooklyn, it basically spans across the whole borough. Some areas in Queens and Manhattan also have a high value of NYPD requests.


Total requests made to NYC's New York City Police Department in 2003−2015
value
[1 to 1,159)
[1,159 to 2,761)
[2,761 to 4,197)
[4,197 to 5,924)
[5,924 to 7,490)
[7,490 to 9,802)
[9,802 to 16,805]
Created using ZIP Code Tabulation Areas



Conclusion

We analyzed several columns of NYC311 dataset and found out several interesting patterns. First, we made some statistical analyses, observing that the top agencies seems to not have larger completion time of their requests (compared to other not-so-popular agencies), which implies they have the resources to cope with thousands of daily complaints efficiently. The number of complains are rising with the beginning of the cold season (October) and drops during the summer month for most of the popular agencies except the police department. Also, the agencies receive the different number of complaints: there is a group of agencies that are responsible for the majority of complains. Another observation is that the distribution of the complains during the day differs from the night. Distribution during the day is more uniform across the complains types, while at night, some of the complains significantly increases in proportion, like Noise, Sewer, Blocked Driveway and Traffic Signal Condition.

There is a clear time variability in the number of complaints on the level of Months, Hours, and (supposedly Years). Furthermore, this variability depends heavily on the type of receiving agency and the type of complaint. Finally, we made an analyses of the complain requests over the map of New York, observing the variability across neighborhoods We identified geographic variability in requests with several clusters of areas with the largest number of requests: 1) highly populated residential areas (such as Brooklyn, Bronx); 2) business and cultural centers (e.g. Manhattan); 3) Populated community objects (LaGuardia airport, CUNY College of Staten Island). Next, we plan to link our insights to the population density in those areas and compare the rate of requests by 1000 inhabitants.