

Huda Ouda, Saif ALKhuraisat, Shaheer Ahmed, Sonali Kannan, Zainab Alhashmi

12/29/2020

Contents

1	Introduction	2
2	Context	2
3	About the Data	2
4	Initializing	3
5	NYC311 Data Extract	3
6	Findings	3
6.1	Top 10 Agencies That Received Complaints	3
6.2	Top 10 Types of Complaints Made	5
6.3	Distribution of Complaints by Borough	6
6.4	Location of Complaints in Each Borough	6
6.5	Yearly Distribution of Complaints	7
6.6	Time-Series Analysis of Calls Made each Month per Year	8
6.7	Complaints Made During Peak Times of January and March	9
6.8	Complaints Location During January and March	10
6.9	Processing Time of Service Requests	11
6.10	Status of Top 5 Complaints	13
6.11	Data for Population of Boroughs (Joining NYC311 and Population Data)	13
6.12	Population of Each Borough	14
6.13	Trend of Total Number of Complaints per Capita	15
6.14	Total Distribution of Heating Complaints	16
6.15	Trend of Heating Complaints per Capita	17
7	Conclusion	18

8	Appendix	19
8.1	Data Dictionary for NYC311 Data	19
8.2	Data Dictionary for Joined Population Data	19

1 Introduction

This report provides insights from the NYC311 dataset which reflects all the calls received from the residents of different boroughs. 311 is a helpline number that provides access to New York City government services through eight platforms: Call Center, Social Media, Mobile App, Text, Video Relay Service and TTY/text telephone. NYC311 is a database for all the complaints received by people who are living in New York. The line “311” is similar to the idea of “911”, however, it is for non-emergency calls. Its mission is to provide the public with quick and easy access to all NYC government services to achieve best customer service. The service started to widespread in 2010 till date.

With this report, we are hoping that these insights could provide some proactive measures the agencies could work on to reduce number of complaints logged by residents.

We have worked with two datasets using RMarkdown, NYC311 and the population changes over the years.

2 Context

New York City, the most populous metropolitan area in the United States, is home to over 8.3 million residents. To serve this massive and diverse population the City operates the nation’s largest and most complex municipal government with more than 350,000 city employees and 120 agencies, offices and organizations offering over 4,000 different services to residents.

Most of the calls from the New Yorkers are related to complaints about public space, buildings, noise, street conditions as well as other non-emergency issues. After receiving the calls, the call center agents at 311 forward the issues to the relevant departments in NY City to take an action and then close the loop of the complaint. This service is not only benefiting the people of New York, but also benefiting the government.

The NYC311 works like a net that holds the city together. Whenever there is a call request, the NYC311 system automatically saves the location in a huge database that feed’s info through NYC government.

Through this huge database, data scientists use NYC311 for analysis, and accurate measurement of data to eventually improve city government.

3 About the Data

The NYC311 dataset includes 51 attributes excluding the unique key variable and over 9 million rows. Each observation/row in the file, represents a complaint logged by a resident. The attributes describe this complaint by including which agency it belongs to, when the creation date is, time, the location and status among other descriptive attributes.

As for the additional dataset we used, the population estimates over years, it includes a total of 165 attributes in it’s original version since it is showing the population trend from 2010 – 2019, yearly basis as well as many other attributes which were not useful for our analysis. After some cleaning and rearranging we brought the dataset down to 3 attributes including “Borough”, “Year”, and “Population” containing 25 entries.

4 Initializing

In this section we initialize the data by loading relevant libraries and storing the data in a variable called 'nyc311'. This will store the observations of the csv file to the object after which we will replace the space in the names of columns by '.' to eliminate the use for quotations every time we try to name a column.

5 NYC311 Data Extract

Here we can see an extract of the data we are going to be working with. It lays out a brief of the attributes that we are using for our analysis and the information contained in them.

Table 1: Table continues below

Agency	Complaint.Type	Status	Borough	Location.Type
NYPD	Vending	Closed	BRONX	Street/Sidewalk
NYPD	Blocked Driveway	Open	BROOKLYN	Street/Sidewalk
NYPD	Noise - Street/Sidewalk	Open	BROOKLYN	Street/Sidewalk
NYPD	Noise - Street/Sidewalk	Assigned	BROOKLYN	Street/Sidewalk
NYPD	Noise - Street/Sidewalk	Closed	MANHATTAN	Street/Sidewalk
NYPD	Noise - Street/Sidewalk	Closed	BROOKLYN	Street/Sidewalk

Created.Date	Closed.Date	Latitude	Longitude	Incident.Zip
2015-04-14 02:14:40	2015-04-14 03:03:22	40.83	-73.82	10465
2015-04-14 02:10:12	NA	40.62	-73.94	11234
2015-04-14 02:03:01	NA	40.62	-74	11204
2015-04-14 02:02:40	NA	40.71	-73.96	11211
2015-04-14 02:00:04	2015-04-14 02:47:33	40.8	-73.96	10025
2015-04-14 01:52:15	2015-04-14 02:11:10	40.69	-73.96	11205

6 Findings

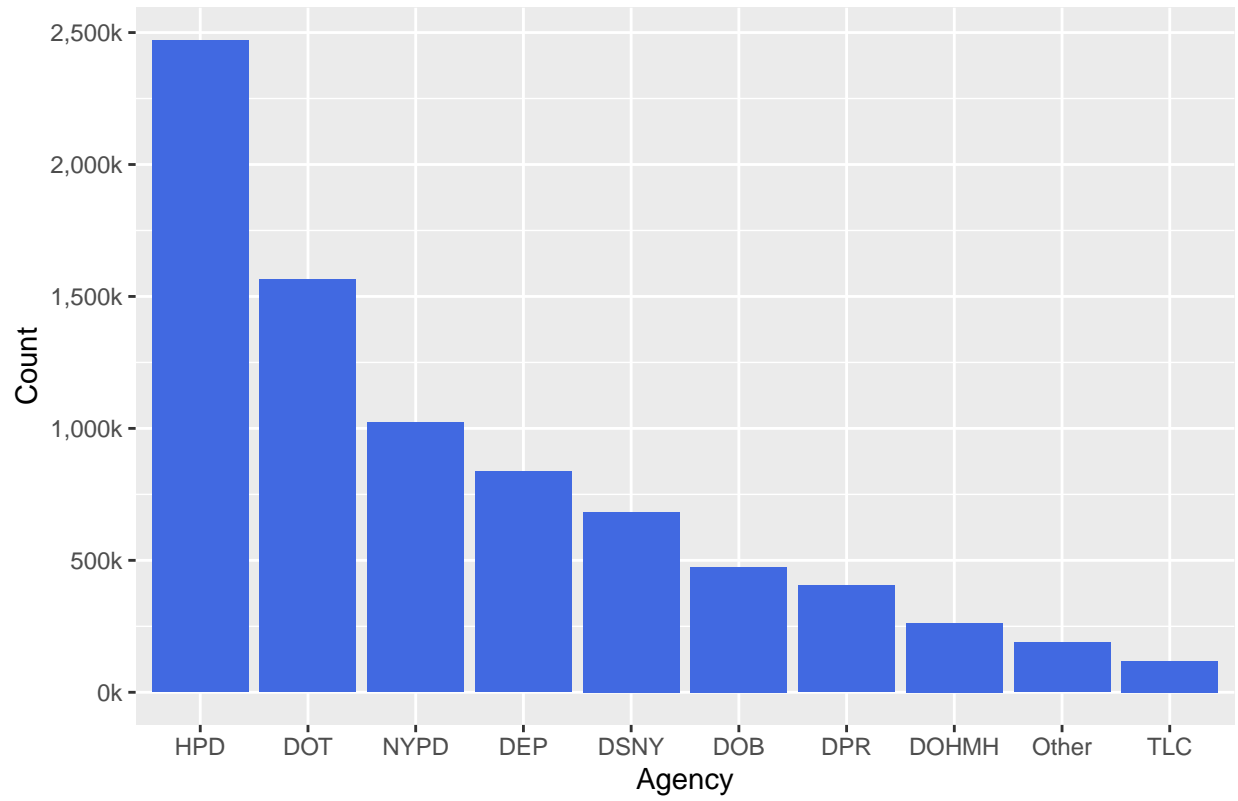
Here we make use of visualization techniques to get a look inside the data and what its trying to tell us. We look for trends and insights that could be of use to anyone who is concerned with them.

But first lets analyze our selected 'sample' to see if it has any null rows or duplicates and fix them

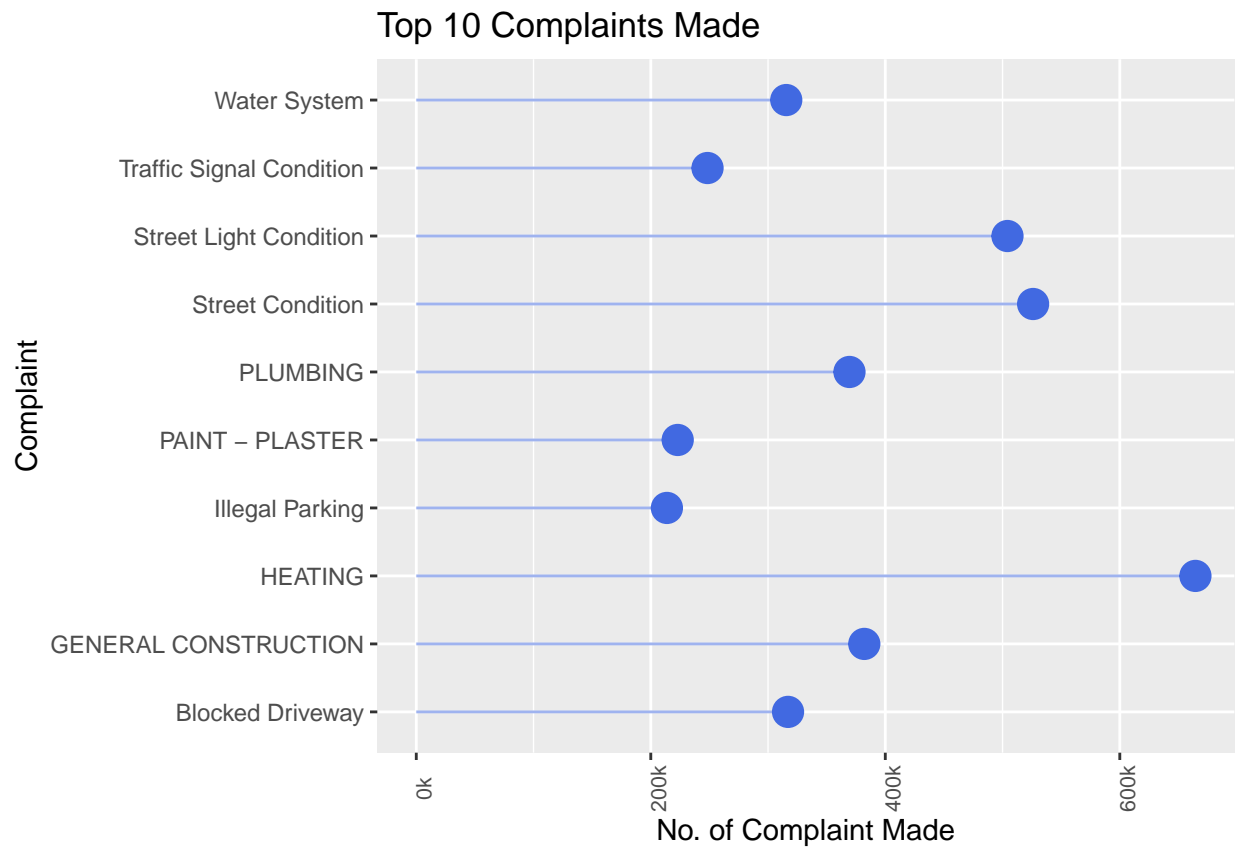
6.1 Top 10 Agencies That Received Complaints

Here we see the top 10 agencies to which complaints have been forwarded to where Housing Preservation & Development received the most of the complaints followed by Department of Transportation and then New York Police Department. This shows most complaints were not of a emergency response nature.

Top 10 Agencies That Received Complaints

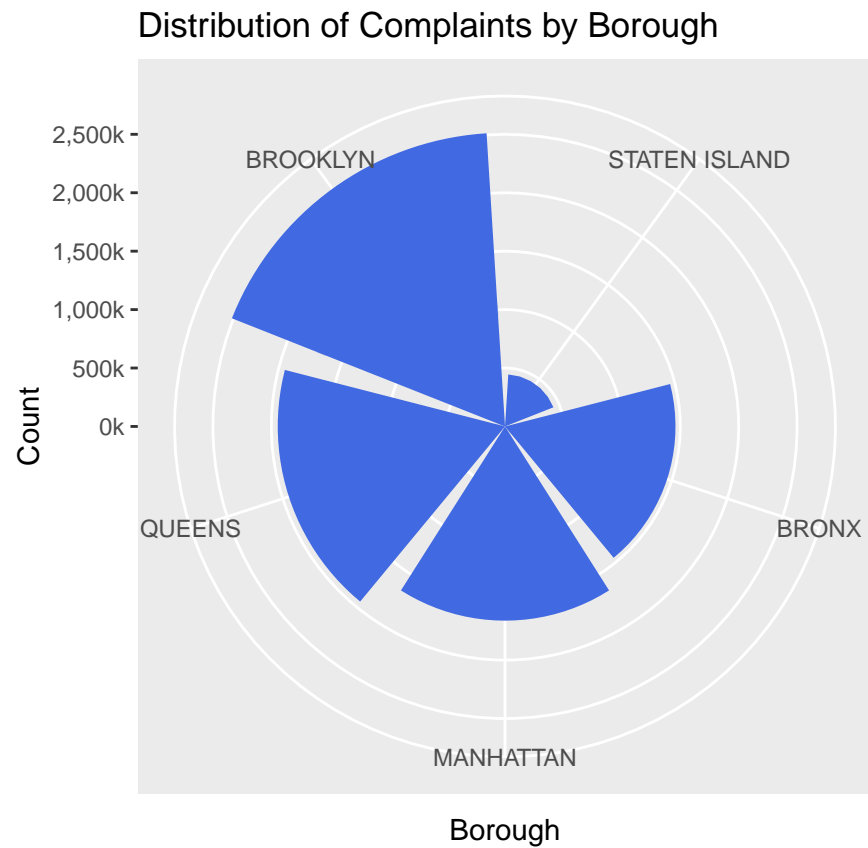


6.2 Top 10 Types of Complaints Made



From the above graph we can identify “HEATING” as the most registered complaint throughout New York State followed by “Street Condition” which will be the focus of our analysis later.

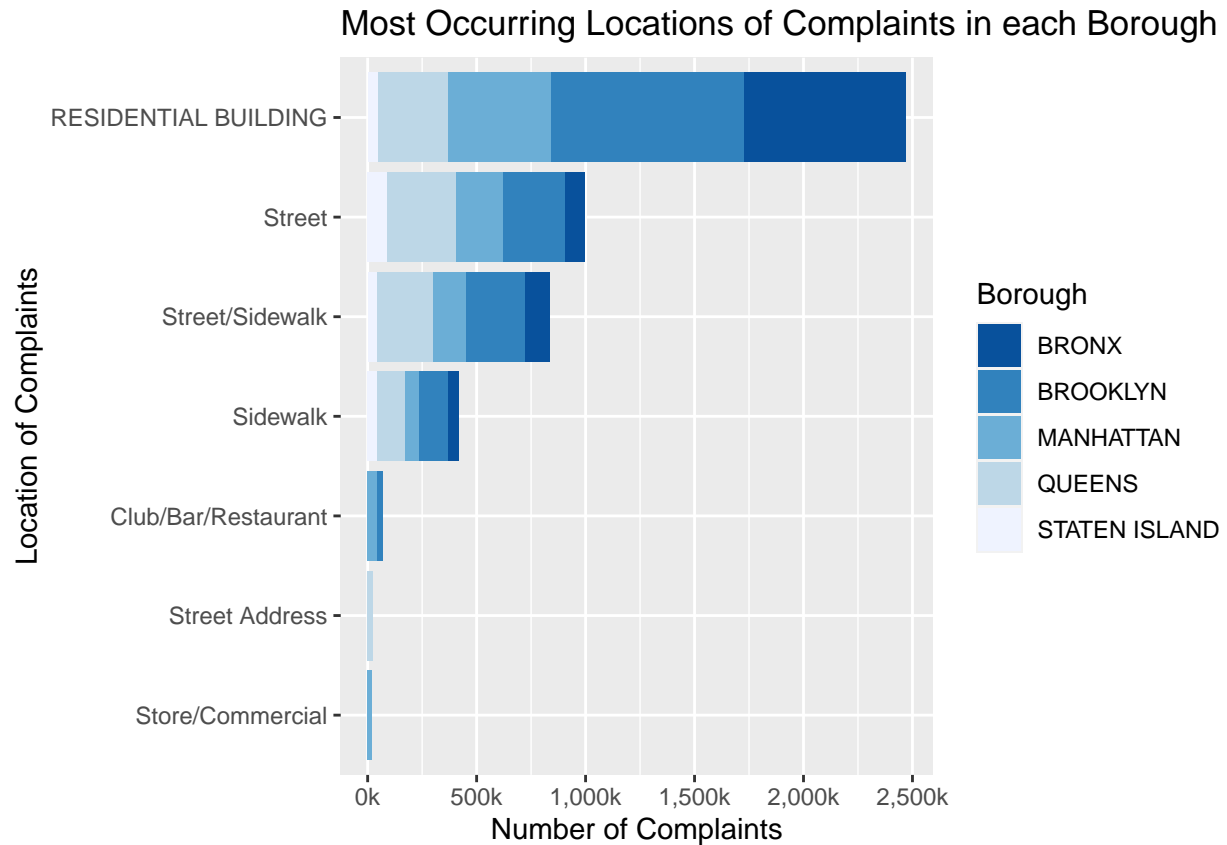
6.3 Distribution of Complaints by Borough



Brooklyn seems to have the highest number of registered complaints and could be classified as a troubled Borough followed by Queens and Manhattan.

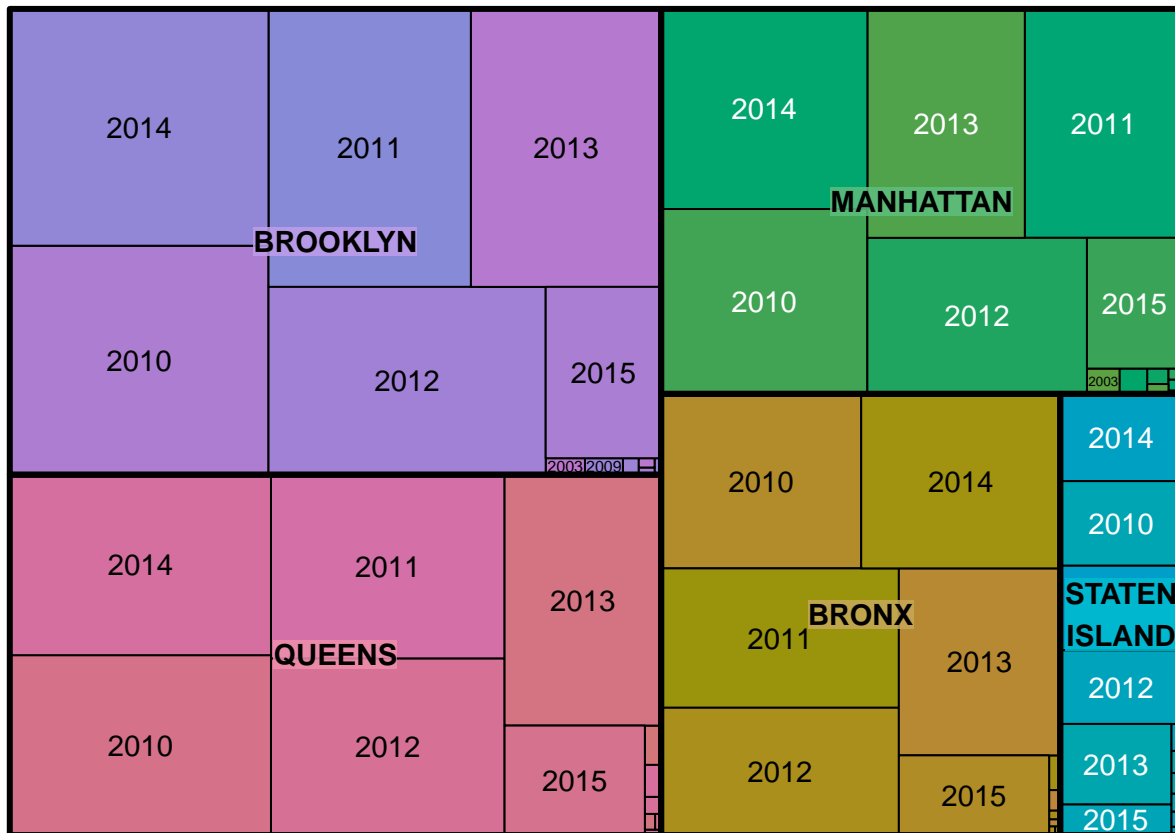
6.4 Location of Complaints in Each Borough

The below bar graph shows that most complaints that come in are about something happening in residential buildings followed by streets and sidewalks.



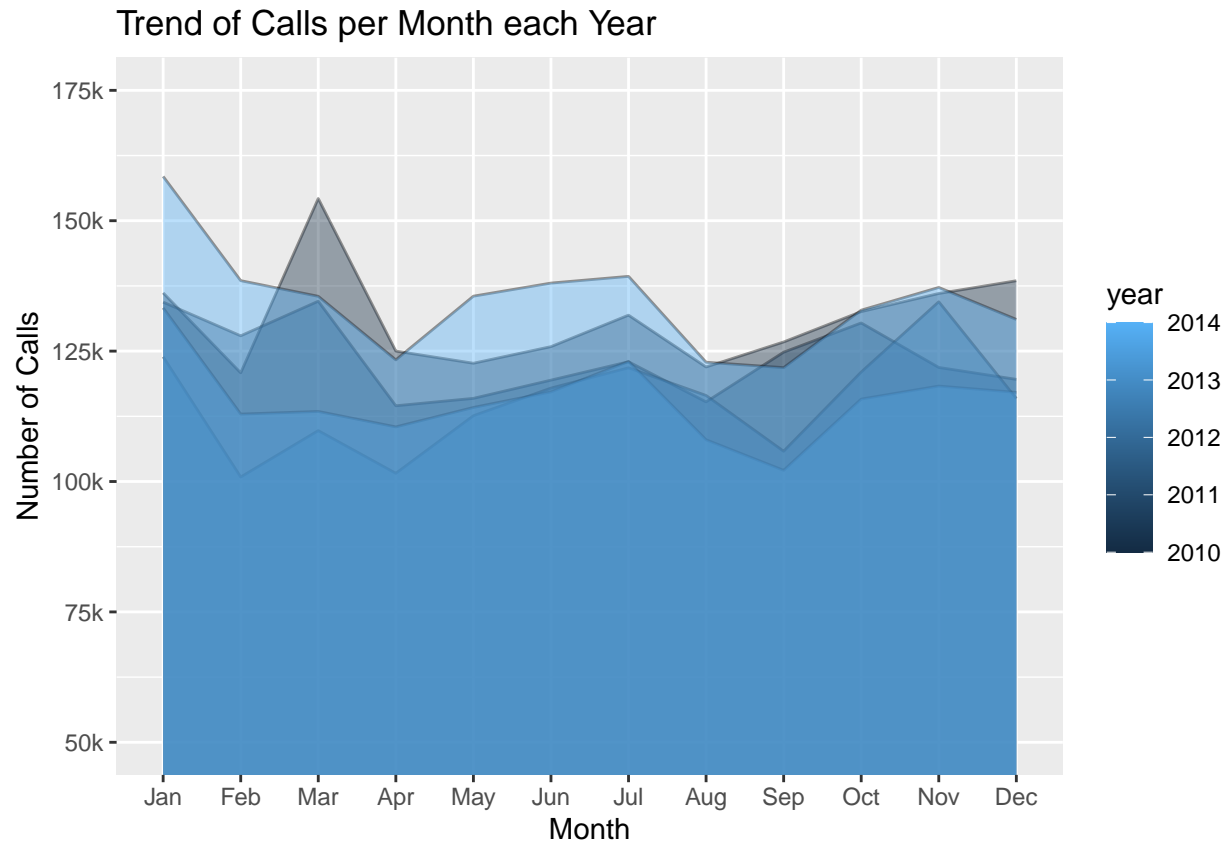
6.5 Yearly Distribution of Complaints

Let's have a look at the distribution of the number of complaints in each Borough every year. This will give us a better understanding of the years to pay attention to for each Borough and give us a head start to look at any time-series analysis we might be interested in. We have to keep in mind the reason for the section for 2015 being so small is because the data is only available till April and so is incomplete.



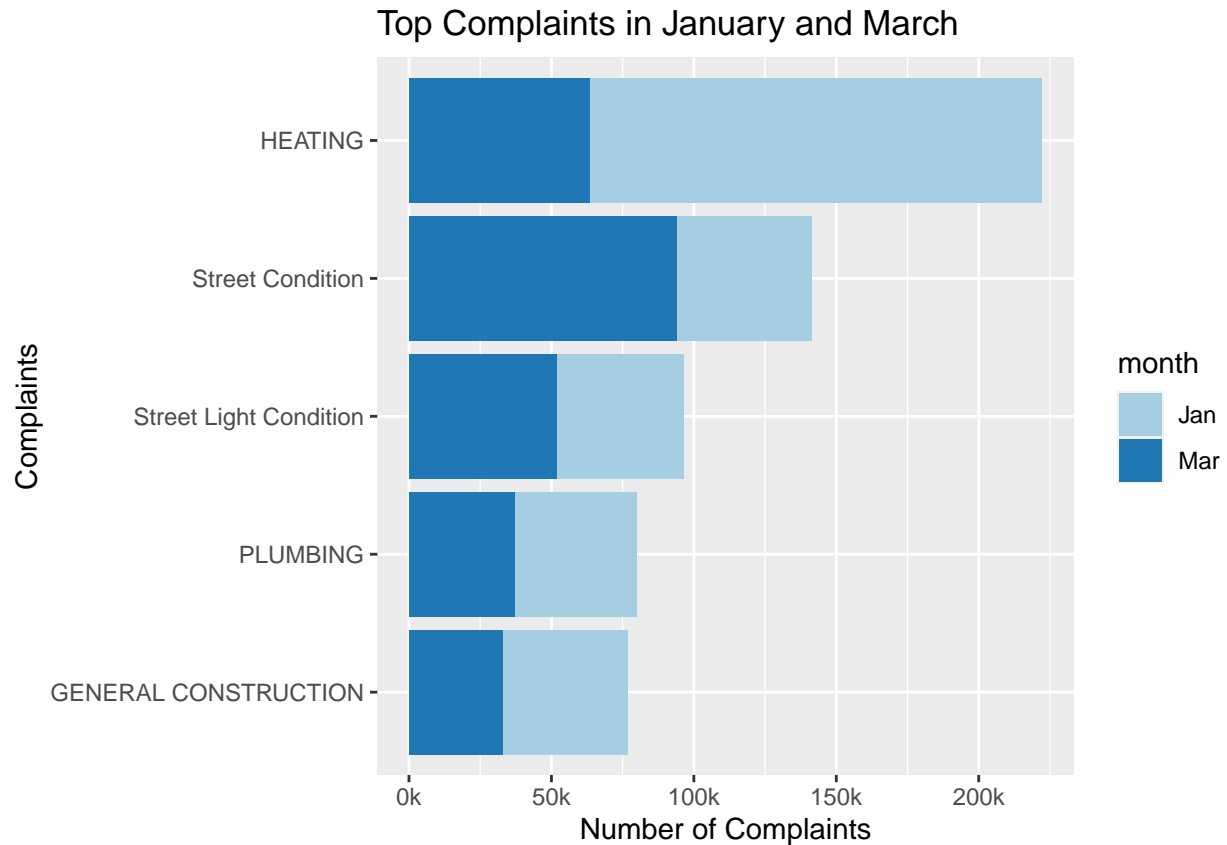
6.6 Time-Series Analysis of Calls Made each Month per Year

Each area graph below shows us the trend of calls for each year. The number of calls peak during January and March consistently for each year. And then we have a steady number of calls made throughout the summer as well. All time-series analysis from this point will exclude all years before 2010 and the year 2015 as it has incomplete data.



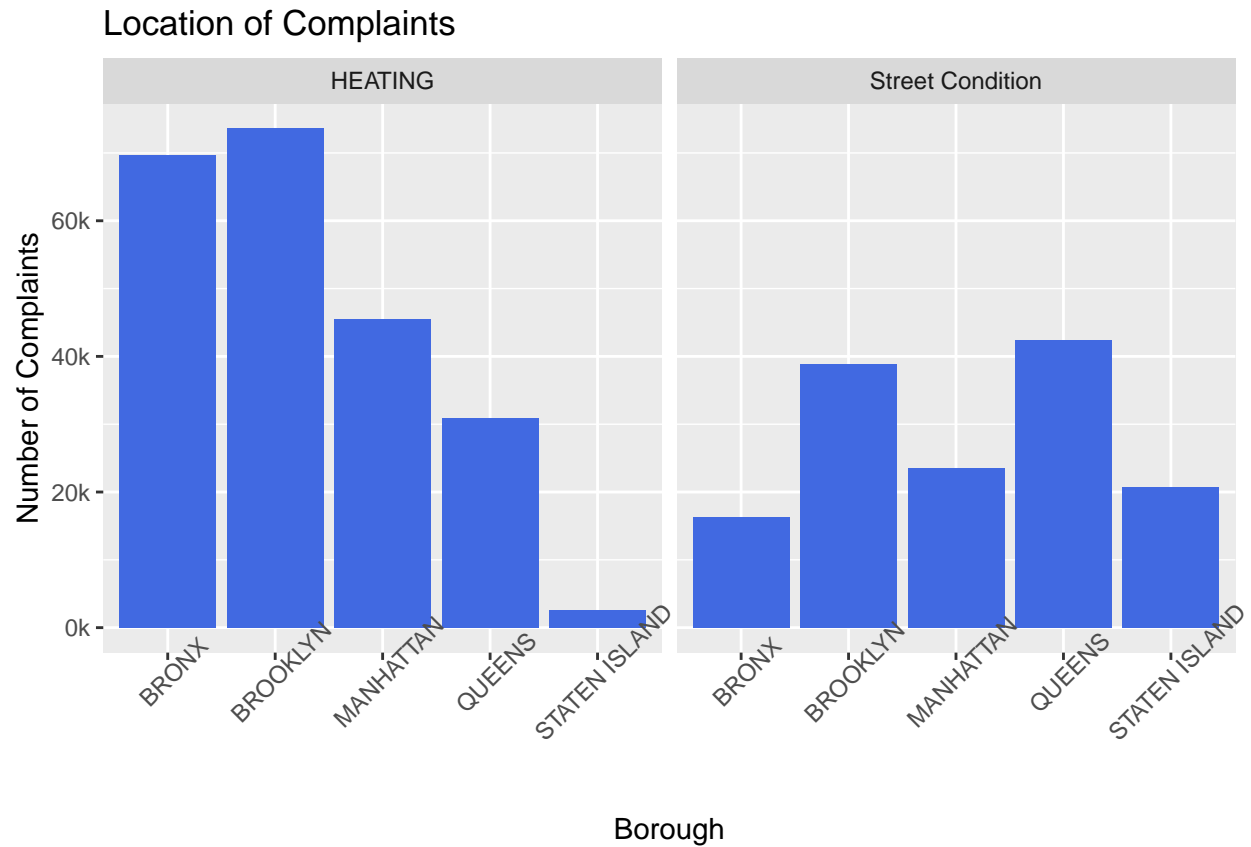
6.7 Complaints Made During Peak Times of January and March

The most number of complaints made in January were concerning heating issues which makes sense as that is one of the peak times of winter. And the peak we get in March has the highest complaints of “Street Condition”. But where do these complaints come from?



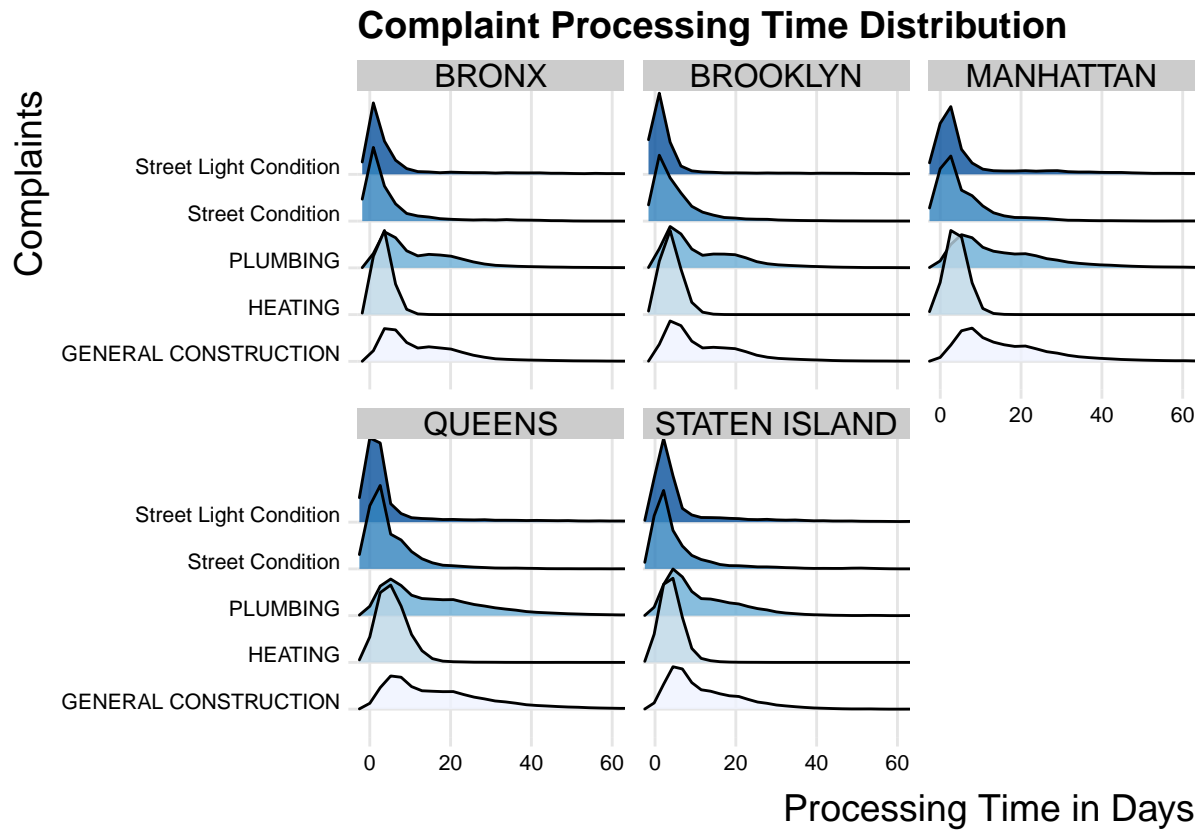
6.8 Complaints Location During January and March

Most heating complaints come from Brooklyn and most street condition complaints from Queens. Looking back at the “Most Occurring Locations of Complaints in each Borough” which states that the top 2 locations of complaints are in residential building and streets, the above plot proves consistent with heating complaints coming from residential building and street condition problems as such as well.



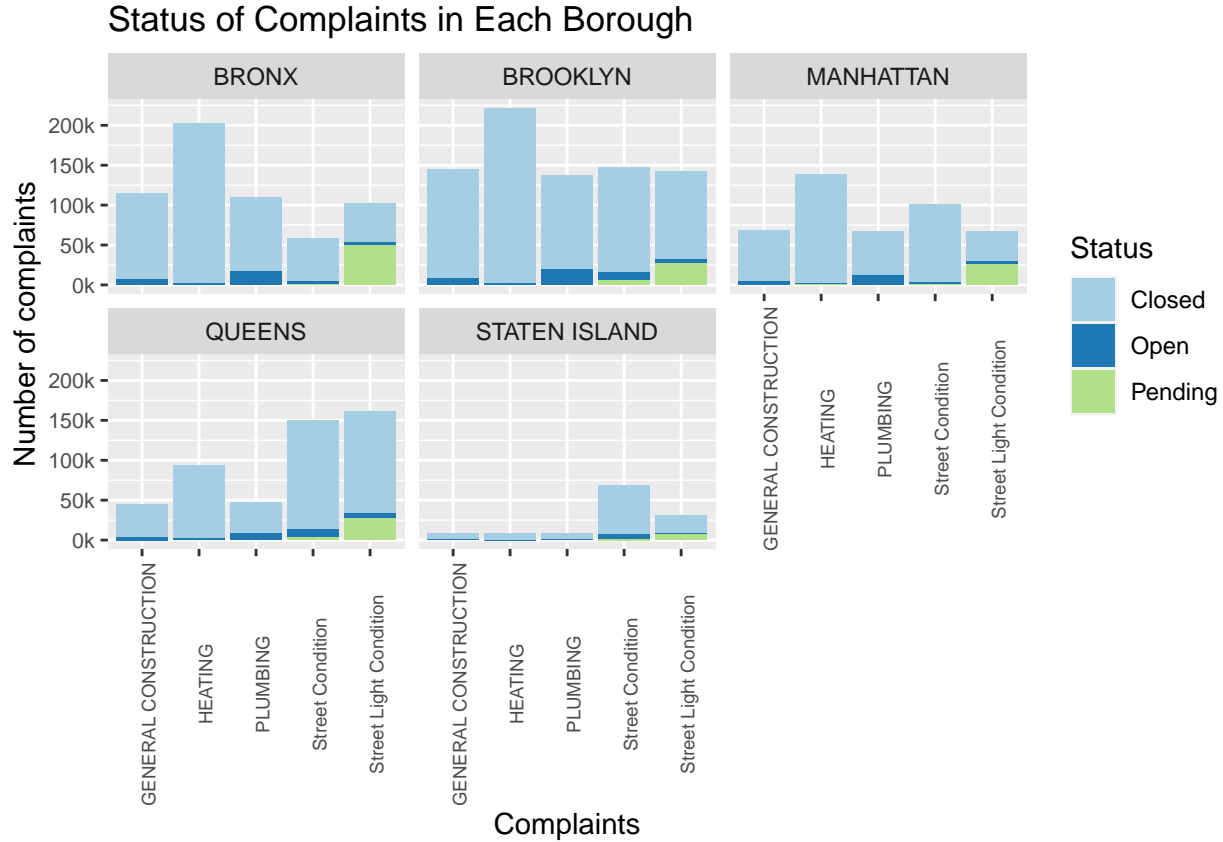
6.9 Processing Time of Service Requests

We had a good look about involved agencies, type of complaints, boroughs and geolocation of complaints yet we should have a holistic view and that why we need to dive into the distribution of processing time of the top 5 complaints in each borough



As we can see, relatively most complaints are resolved within 10 days. Heating and street lighting condition problems seem to be the highest in number of complaints made and are the quickest to be resolved. The longest time taken is general construction.

6.10 Status of Top 5 Complaints



The above graph represents the status of the complaints in each Borough. We have mapped all relevant statuses to “Open” and “Closed” and kept “Pending” as it is since these 3 were identified to be the most relevant of all categories of statuses. Most cases present in this dataset seemed to have been closed but street light condition complaints have quite a few number which are still open and pending.

6.11 Data for Population of Boroughs (Joining NYC311 and Population Data)

We managed to find a dataset containing information on the population of each Borough throughout the years 2010 to 2015. Since the NYC311 data lacks complete information on the complaints made in 2015 it was omitted from the below analysis. The data for population was very untidy with a lot of columns and unnecessary data. Some cleaning process was applied to get the data into a simple format with 25 rows and 3 relevant columns such as Borough name, Year, and population corresponding to each year for every Borough.

The below table represents an extract of the cleaned population dataset, summarized NYC311 complaints dataset as well as the joined dataset formed once the population data was joined respectively. The summarized NYC311 complaints data represents the highest number of complaints made each year, which happens to consistently be “HEATING” in every Borough.

Table showing the population data of each Borough per year

Borough	year	population
BRONX	2010	1387298
BROOKLYN	2010	2509828

Borough	year	population
MANHATTAN	2010	1588767
QUEENS	2010	2234701
STATEN ISLAND	2010	469615
BRONX	2011	1397335

Table showing summarized data of the top complaints every year from NYC311

Borough	year	Complaint.Type	Heating.Complaints	Total.Complaints
BRONX	2010	HEATING	47088	289575
BRONX	2011	HEATING	43155	274692
BRONX	2012	HEATING	41758	250779
BRONX	2013	HEATING	48578	249935
BRONX	2014	HEATING	22291	287532
BROOKLYN	2010	HEATING	57220	491359

Table showing the joined dataset of the above two tables.

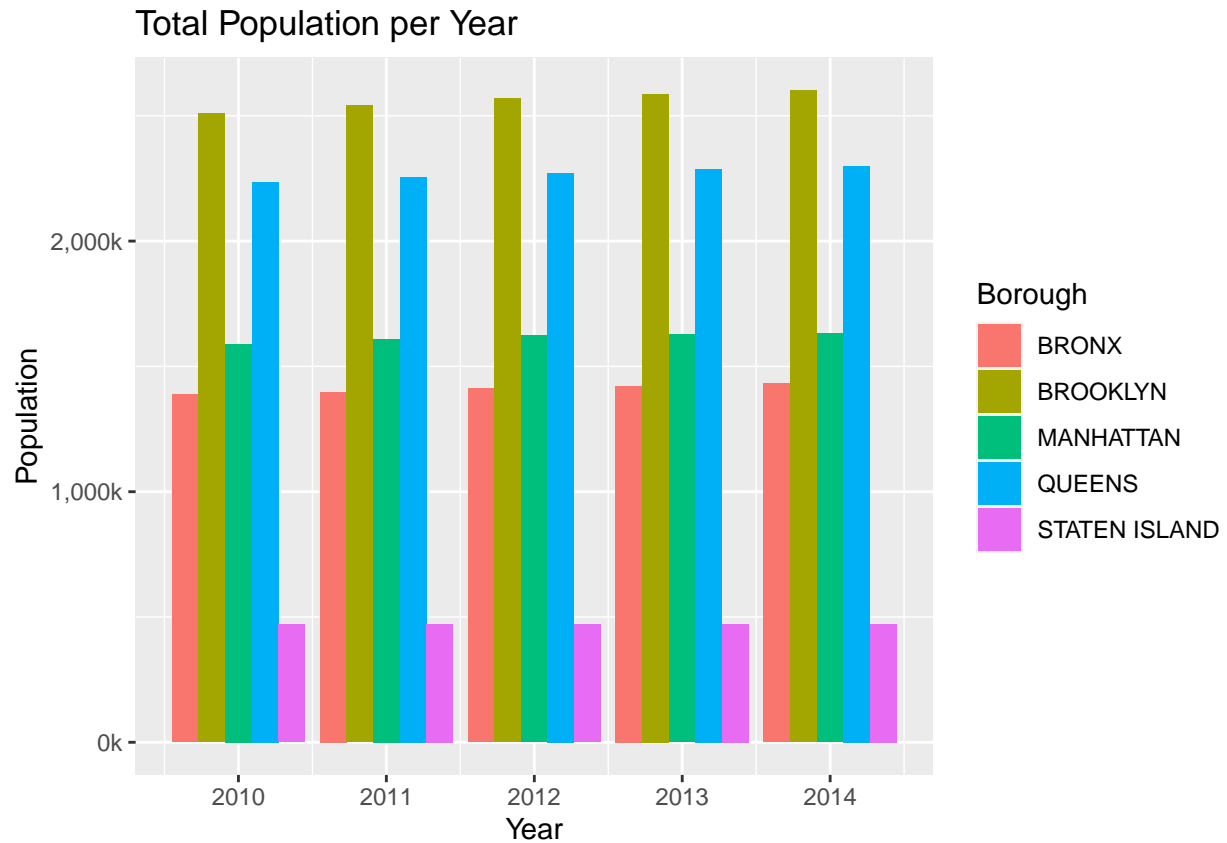
Table 5: Table continues below

Borough	year	Complaint.Type	Heating.Complaints	Total.Complaints
BRONX	2010	HEATING	47088	289575
BRONX	2011	HEATING	43155	274692
BRONX	2012	HEATING	41758	250779
BRONX	2013	HEATING	48578	249935
BRONX	2014	HEATING	22291	287532
BROOKLYN	2010	HEATING	57220	491359

population
1387298
1397335
1411496
1421928
1430942
2509828

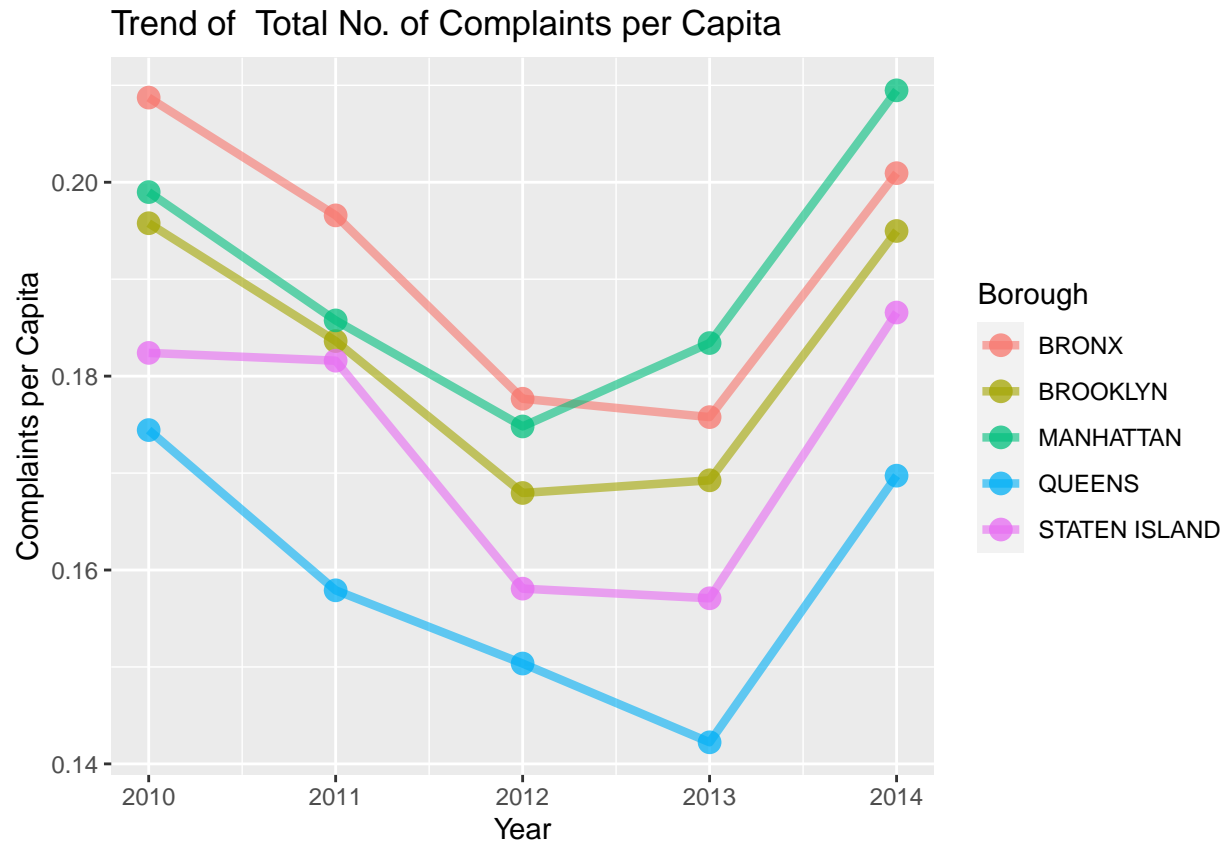
6.12 Population of Each Borough

We already know that Brooklyn comes out highest in terms of the number of complaints registered out of all Boroughs. We can now confirm that is due to Brooklyn being the highest in population out of all. There isn't much drastic change in the change of population each increasing year.



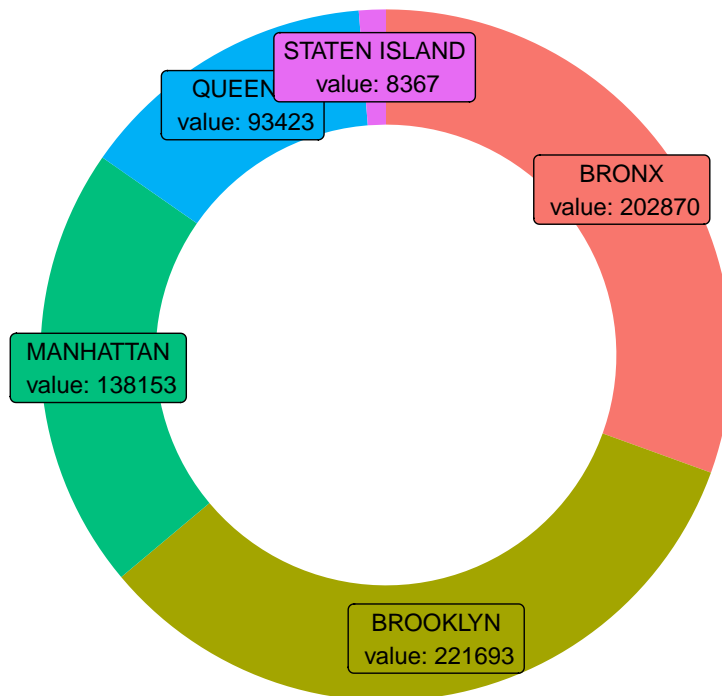
6.13 Trend of Total Number of Complaints per Capita

We can now find out what the situation in each Borough looks when we see the number of complaints made per capita over the years. Dividing the number of complaints by the population allows us to look at the trend in an un-biased manner by removing the population bias since there is a drastic difference in the population of Brooklyn and Staten Island (nearly 2 million people). The trend we see is slightly alarming as complaints seem to have sharply risen after 2013 and as of 2014, Manhattan leads in the number of complaints made per capita followed by Bronx and Brooklyn.



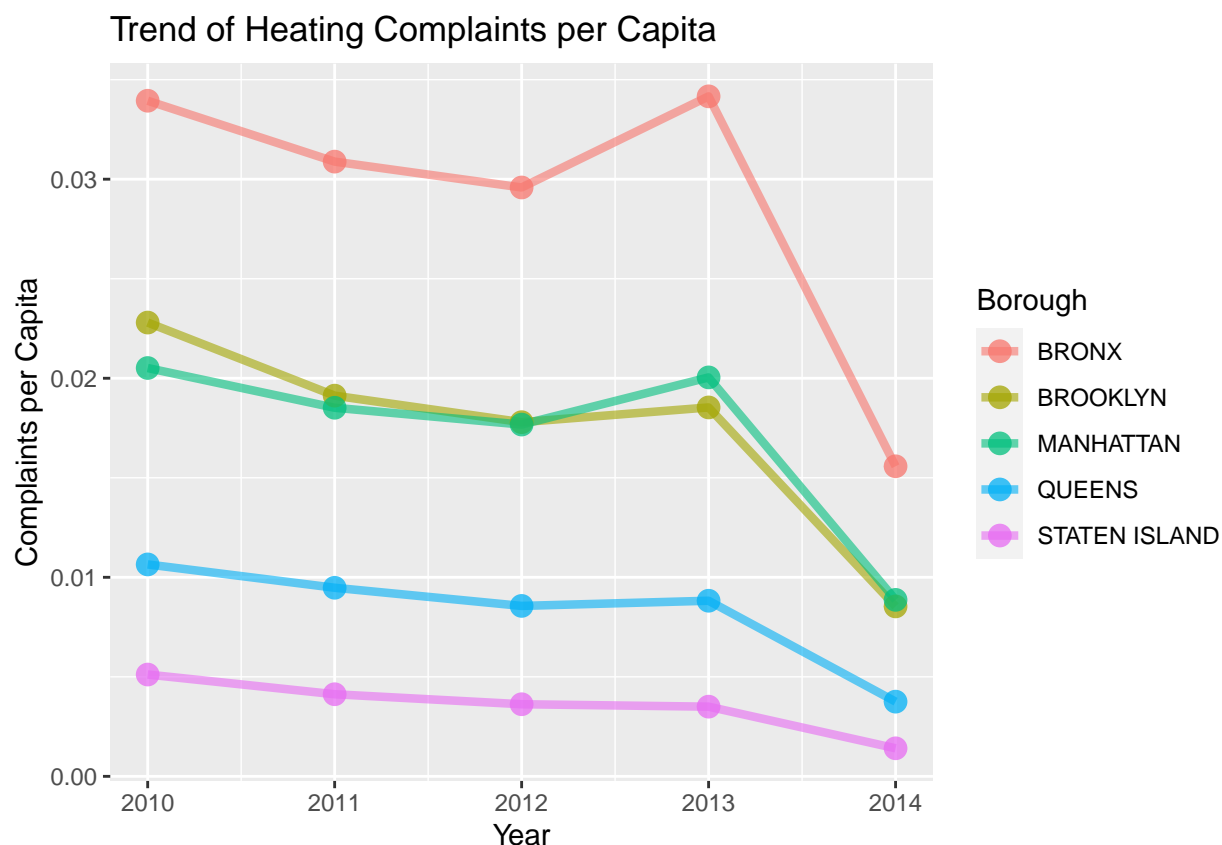
6.14 Total Distribution of Heating Complaints

We know that heating complaints constitutes the highest number of complaints made every year. So which Boroughs report the highest number of these complaints overall? The below graph gives us a look into this insight.



6.15 Trend of Heating Complaints per Capita

Now that we know which Boroughs have the highest number of heating complaints and even the number of complaints, let's see how their trend looks to be over time. The below line graph shows us a trend that we see for the complaints made for each Borough throughout 2010 to 2014 in regards to heating complaints. We can see a clear drop in number of cases being registered especially after 2013 but Bronx maintains the most number of heating complaints made per capita with Brooklyn and Manhattan battling for second and third place.



7 Conclusion

We have used two datasets for our insights in this report, NYC311, and population changes over the years in each borough.

Our exploration of the data revealed many insights on the agencies receiving complaints, the distribution of complaints, and trends over time.

The top three agencies that received complaints are HPD, DOT, and NYPD. Also, livable spaces (residential buildings) registered the highest number of complaints across all boroughs than in commercial areas (store, club, bar, or restaurants).

We found out the majority of the cases reported in this database had been closed by April 2015 which is where our data cuts off at.

Moreover, Heating problems remain prevalent throughout the state of New York but is one of the quickest to be resolved. These complaints reach their peaks in January and March consistently every year.

Per capita analysis removes Brooklyn from the top of the list even though it has a high number of populations and put Manhattan followed by Bronx. This helps in avoiding population bias.

Based on our insights, we recommend the below to ensure excellent services provided to the NYC population:

- The relevant agencies must investigate the gas supply to buildings with central heaters and the quality control of different heater types installed in buildings and sold in the state.
- More workers for these agencies need to be available in Brooklyn as it comes out on top in most analysis of complaints that we have conducted mostly due to having the largest population in New York.

8 Appendix

8.1 Data Dictionary for NYC311 Data

Description	Column.Name
Agency Complaint was Reported to	Agency
Type of Complaint	Complaint.Type
Status of Submitted Complaint	Status
Borough from which Complaint was filled	Borough
Location Type of Complaint	Location.Type
Date Complaint was Created	Created.Date
Date Complaint was Closed	Closed.Date
Latitude of Location of Complaint	Latitude
Longitude of Location of Complaint	Longitude
Zip code of the Location of Complaint	Incident.Zip

8.2 Data Dictionary for Joined Population Data

Description	Column.Name
Name of Borough where complaint was registered	Borough
Year complaint was registered	year
Type of complaint	Complaint.Type
Number of complaints related to 'HEATING'	Heating.Complaints
Total number of complaints registered	Total.Complaints
Total population at the time	population