

VISUAL ANALYTICS

A Report On

Analyzing Heating Complaint Type of NYC311 DatsSet

Submitted By:

Adit Manish Doshi
Neha Manish Vanjara
Yash Narendra Kharade

ad8343@rit.edu
nv6518@rit.edu
yk6014@rit.edu

INTRODUCTION

The 311 Data Set is updated daily and contains information about more than 24 million service requests. 311 is a phone number, used in the US, that allows callers to raise a non-emergency complaint similar to the number 911, which is strictly for the emergency calls.

My teammates and I came to the US this Spring and one of the difficulties we faced was adjusting to this climate. Once our heating system broke and it took hours to fix it. This was when the importance of the heating system struck us. According to the Department of Housing Preservation and Development (HPD), “heat season” has been here for quite some time: Between October 1 and May 31, New York City landlords are required to ensure that apartments are reasonably heated. Landlords must also provide hot water, “365 days per year at a constant minimum temperature of 120 degrees Fahrenheit.”

Of course, that’s not always the case: Every year, there were tens of thousands of heat complaints, including a **whopping 21,894** in one week in 2018 (and there’s regularly a backlog of complaints that HPD has not yet addressed).

Hence in our analysis, we decided to focus our report around Heating Complaint Type.

ORIGIN OF NYC311 DATASET

This data is New York City’s primary source of government information and non-emergency services. NYC311 data originated from reports by New Yorkers to the authorities of a strange smell that was feared that it might cause chemical warfare especially from Aunt Jemima wing of al Qaeda.

The data was enhanced more on January 29, 2009, when another maple syrup event commenced in northern Manhattan. Launched in March 2003, 311 now fields on average more than 50,000 calls a day, offering information about more than 3,600 topics: school closings, recycling rules, homeless shelters, park events, pothole repairs.

AIM

- Which agency receives the most complaints.
- Which month received the most amount of complaints.
- Which was the highest number of complaint types.
- Which Boroughs have the highest number of heating complaints.
- The change in number of calls with the rise of population.

CONTEXT

The Data Scientist can use this dataset to analyze and make predictions about the most common type of complaint received by different areas. This information can further help people. It can help people to decide a better neighborhood for themselves. In case someone prefers a quiet neighborhood, they can avoid places with a high number of noise complaints leading to better and safer housing choice.

Agencies and the Government can use the data to improve their infrastructure and services. If a particular complaint is taking a longer time to resolve, the government can try to figure out if the agencies are responsible for the delay or if it is the type of complaint as some types of complaints will take longer time to resolve over others. It is obvious that Street condition complaint will take longer time to resolve than a block highway complaint.

THE DATA SET

- **NYC311 DATA**

The 311 DataSet has 52 columns and 912493 rows. Each row represents a complaint call. The DataSet also has many missing values and could possibly have duplicate rows.

	UniqueKey	Agency	Borough	ComplaintType
1	30387854	NYPD	BRONX	Vending
2	30388338	NYPD	BROOKLYN	Blocked Driveway
3	30395236	NYPD	BROOKLYN	Noise - Street/Sidewalk
4	30394595	NYPD	BROOKLYN	Noise - Street/Sidewalk
5	30390517	NYPD	MANHATTAN	Noise - Street/Sidewalk

Fig 1

- **NYC POPULATION DATA**

The NYC Population data set was downloaded from <https://data.ny.gov/Government-Finance/Annual-Population-Estimates-for-New-York-State-and/krt9-ym2k>. The DataSet has 52 columns and 912493 rows. Each row represents a population of the different states of New York.

	FIPSCode	Borough	Year	ProgramType	Population
1	36000	New York State	2021	Postcensal Population Estimate	19835913
2	36001	Albany County	2021	Postcensal Population Estimate	313743
3	36003	Allegany County	2021	Postcensal Population Estimate	46106
4	36005	BRONX	2021	Postcensal Population Estimate	1424948
5	36007	Broome County	2021	Postcensal Population Estimate	197240

Fig 2

TIDYING THE DATA SET

Firstly, we remove the missing values from the dataset. After removing the missing values, we're left with 52 columns and 8293313 rows. We then remove the duplicate rows. Before doing that we need to delete the Unique Key column as even if the same complaint is raised more than once, it's still going to have different Unique Id. Hence, Inorder to get distinct rows, it is necessary to get rid

of the Unique Key column first. Now we're left with 51 columns and 7455520 rows.

JOINING THE DATASETS

The population data set had 5 columns and 3654 rows. The second column was named Geography which we changed into Borough so that it matches our Borough column of nyc311 dataset. In the Borough column we replaced the below data values:

New York County	MANHATTAN
Bronx County	BRONX
Kings County	BROOKLYN
Richmond County	STATEN ISLAND
Queens County	QUEENS

Fig 3

We now filtered the program type column so that we just have the Postcensal Population Estimate and Intercensal Population Estimate.

From the new data set we select three columns namely Population, Borough and Year and from the nyc311 data set we select Borough, ComplaintType, IncidentZip, CreatedDate, ClosedDate and Year. For both the dataset we filter out the above mentioned five Boroughs from Fig 3. In addition, we considered the data from 2005 till 2015.

We then join both the data sets using Inner Join.

FINDINGS

We're analyzing the nyc311 data set to see which complaint type has the highest no. of complaints.

Below are the results from our analysis.

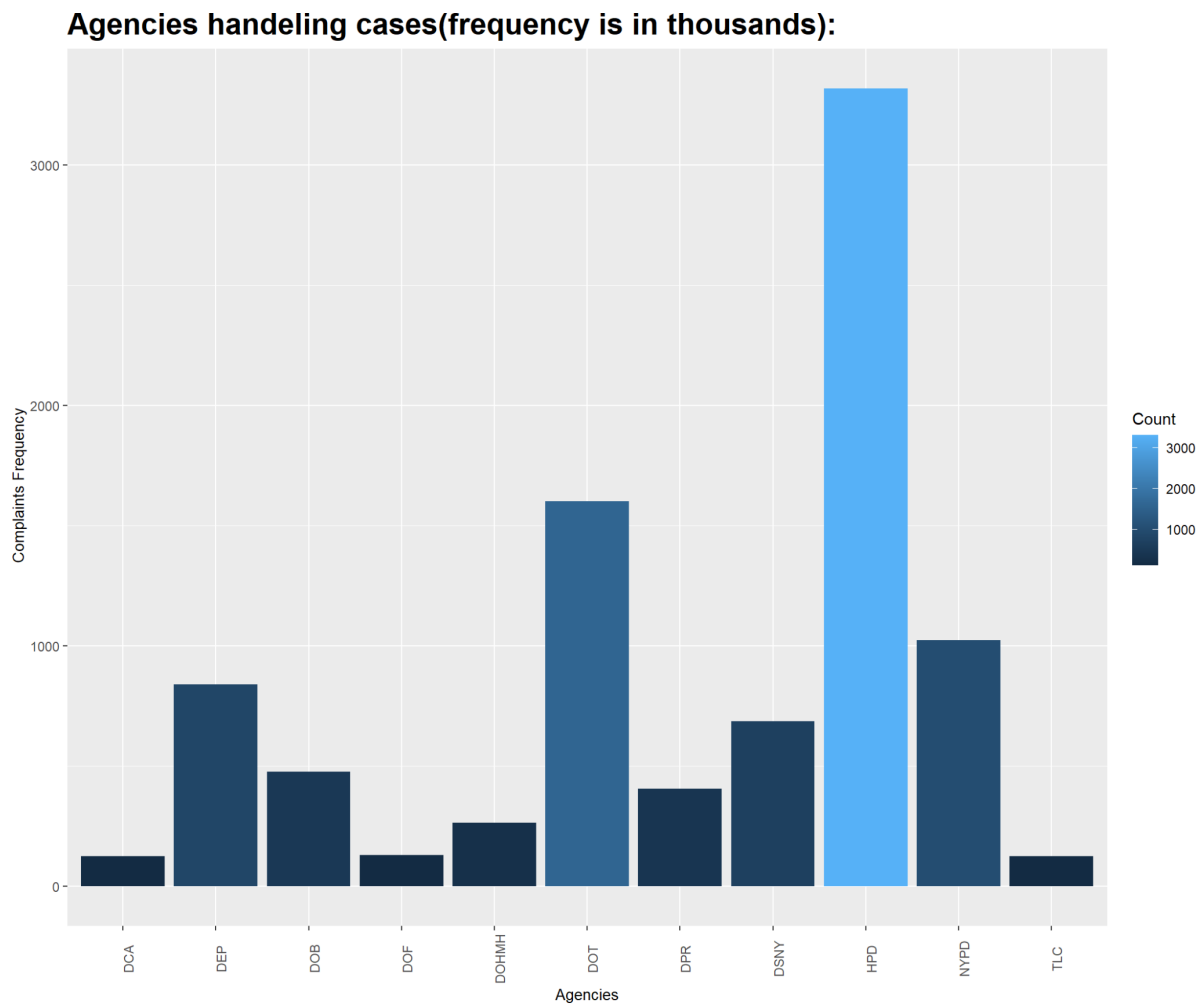


Fig 4

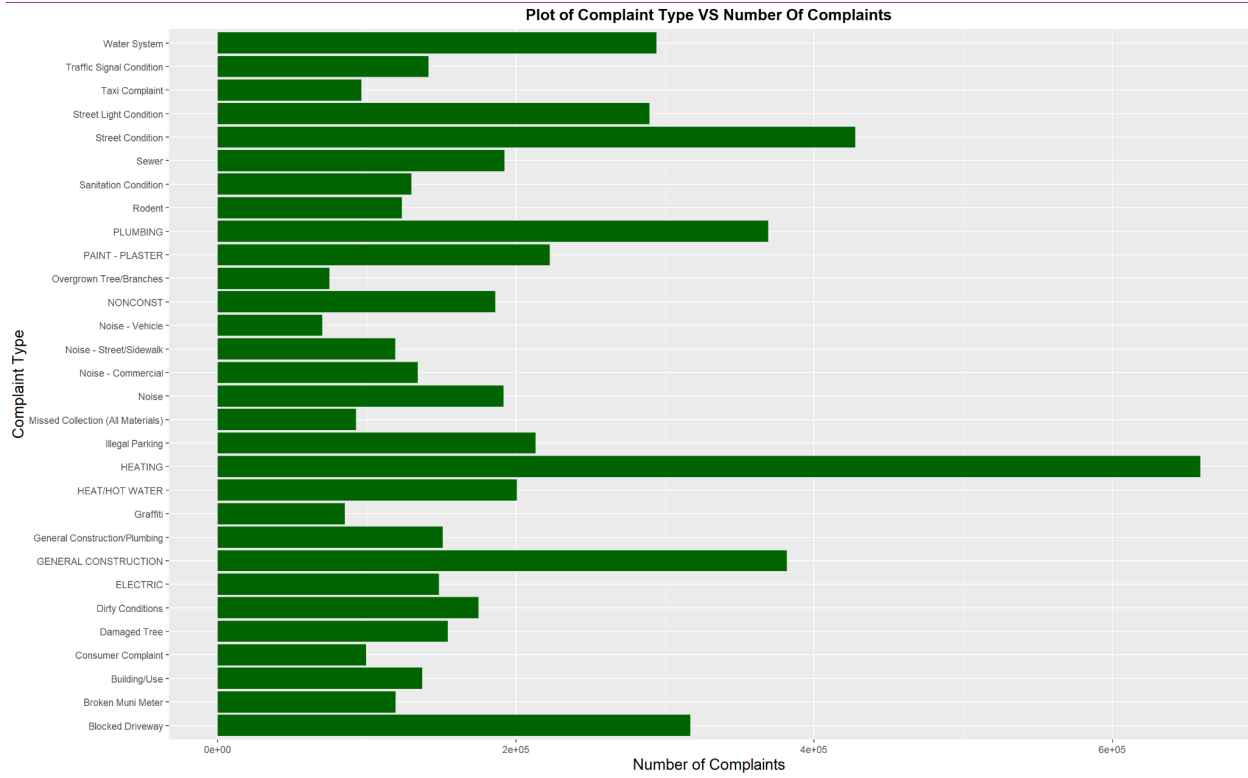


Fig 5

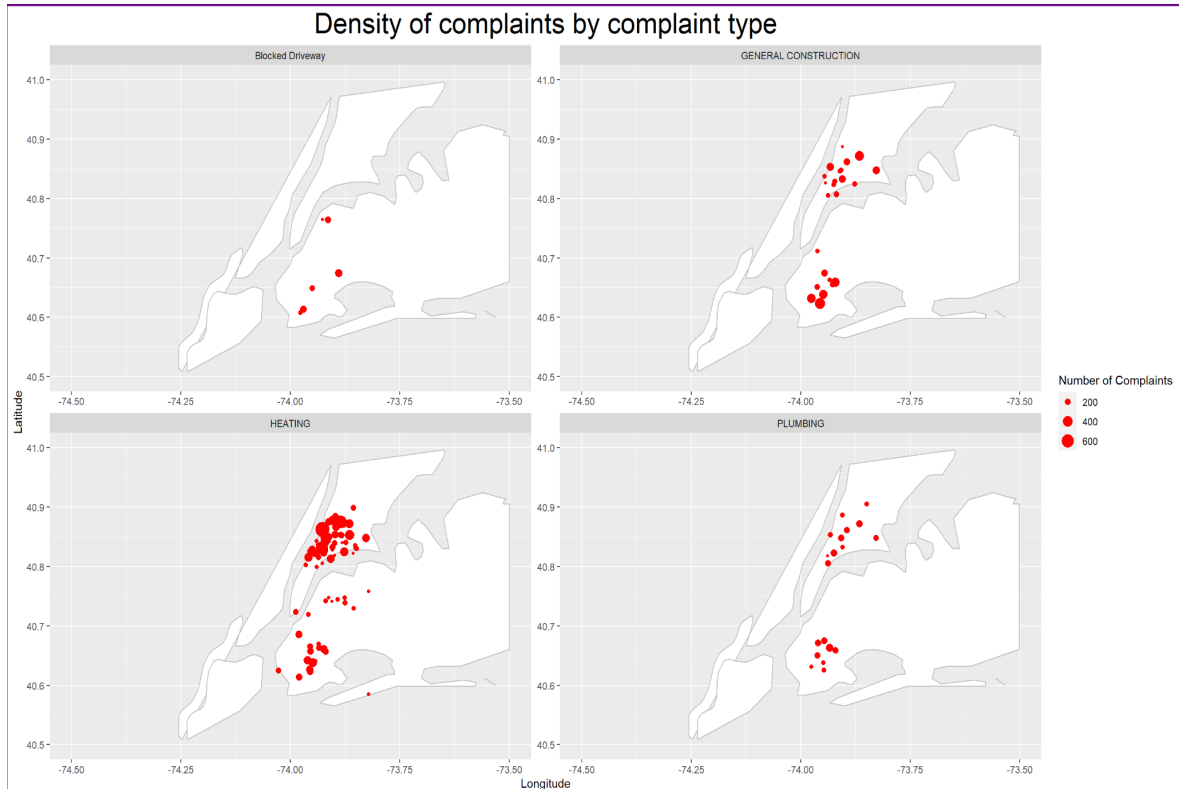


Fig 6

From Fig 4 and Fig 5, we see that the Heating complaint type has the highest number of complaints followed by Street Condition and General Construction. We also observe that Graffiti complaints are the lowest.

Visualizing which Borough has highest Heating complaints

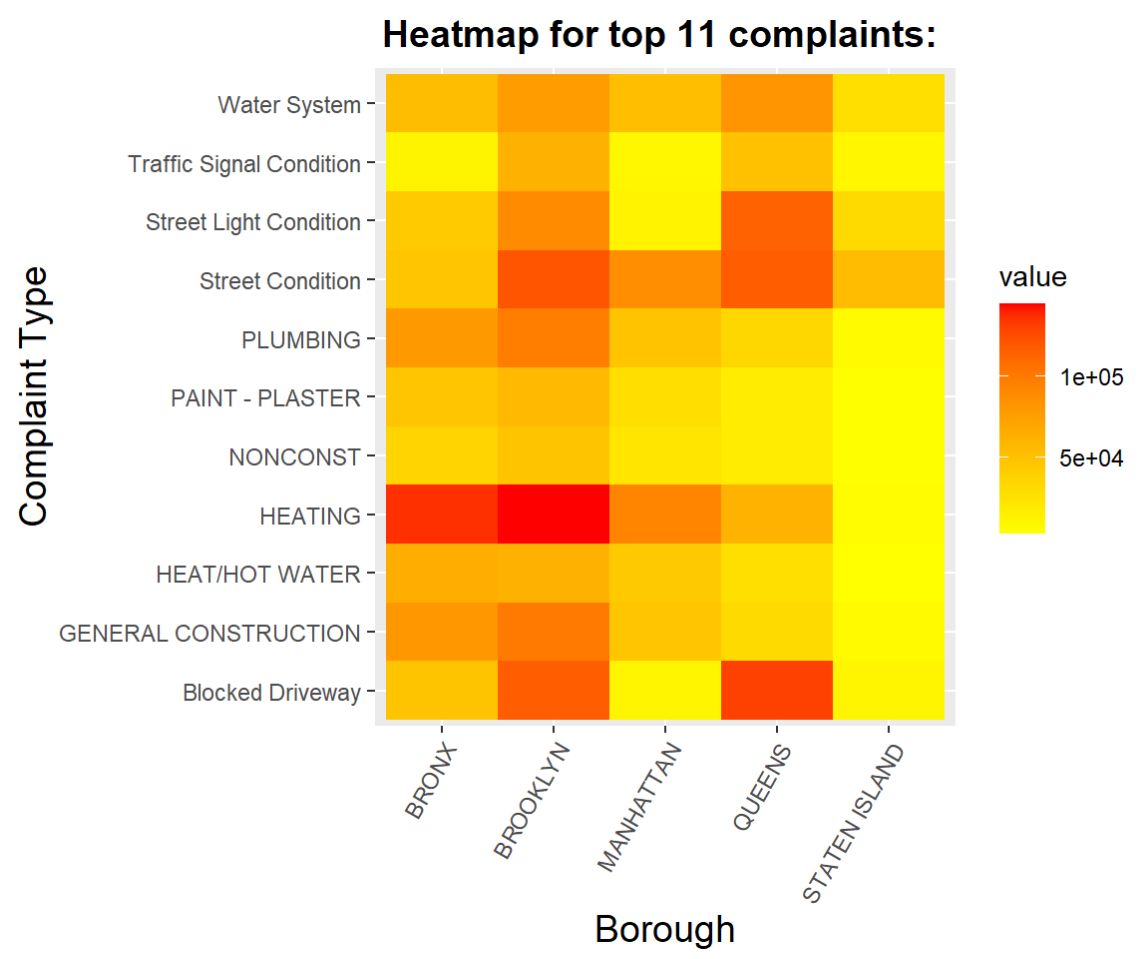


Fig 7

From Fig 6, it shows that Heating, Blocked Driveway, Street Condition have the highest number of complaints. We also notice that the Borough Bronx and Brooklyn have the highest number of heating complaints followed by Manhattan. It also depicts that Staten Island has the lowest number of complaints. The reason for Staten Island has less number of complaints compared to others, could be due to its less population.

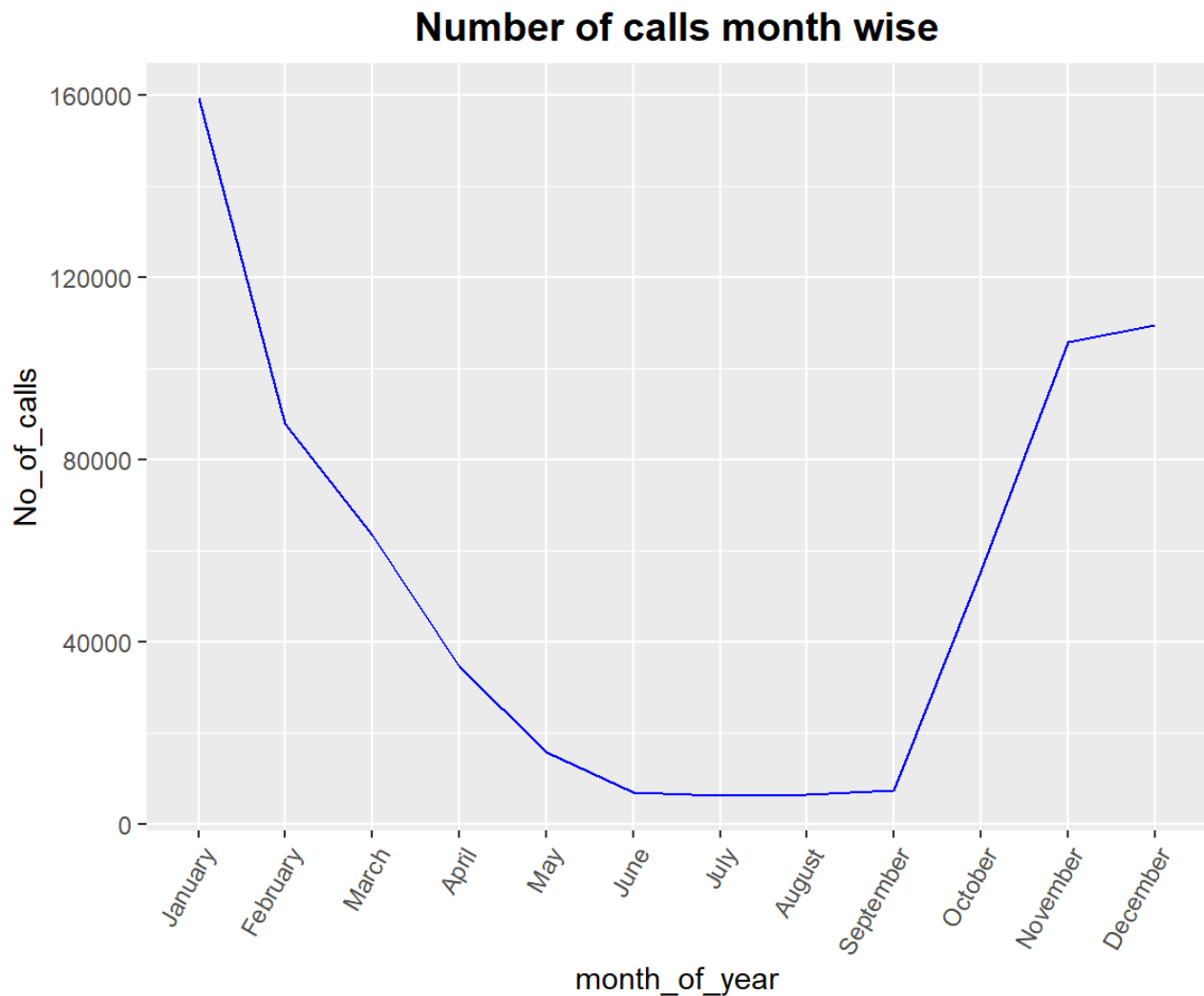


Fig 8

Now we are analyzing the amount of calls that were made during different months of the year. From the above figure, we can see that the highest number of calls were made from the month of October till the month of March. This is

quite obvious as these are the cold months of the country. Usage of heater is a necessity at this time of the year. January received the highest number of calls amongst the others as snowfall is at peak and it's inconvenient without a heating system. We also observe the downward trend from June to September i.e Summer Time and a damaged heater would not be that much of a problem.

Month wise heating calls:

Month	Heating calls
January	159323
February	87836
March	63466
April	34709
May	15875
June	7046
July	6309
August	6522
September	7430
October	55516
November	105661
December	109383

Fig 9

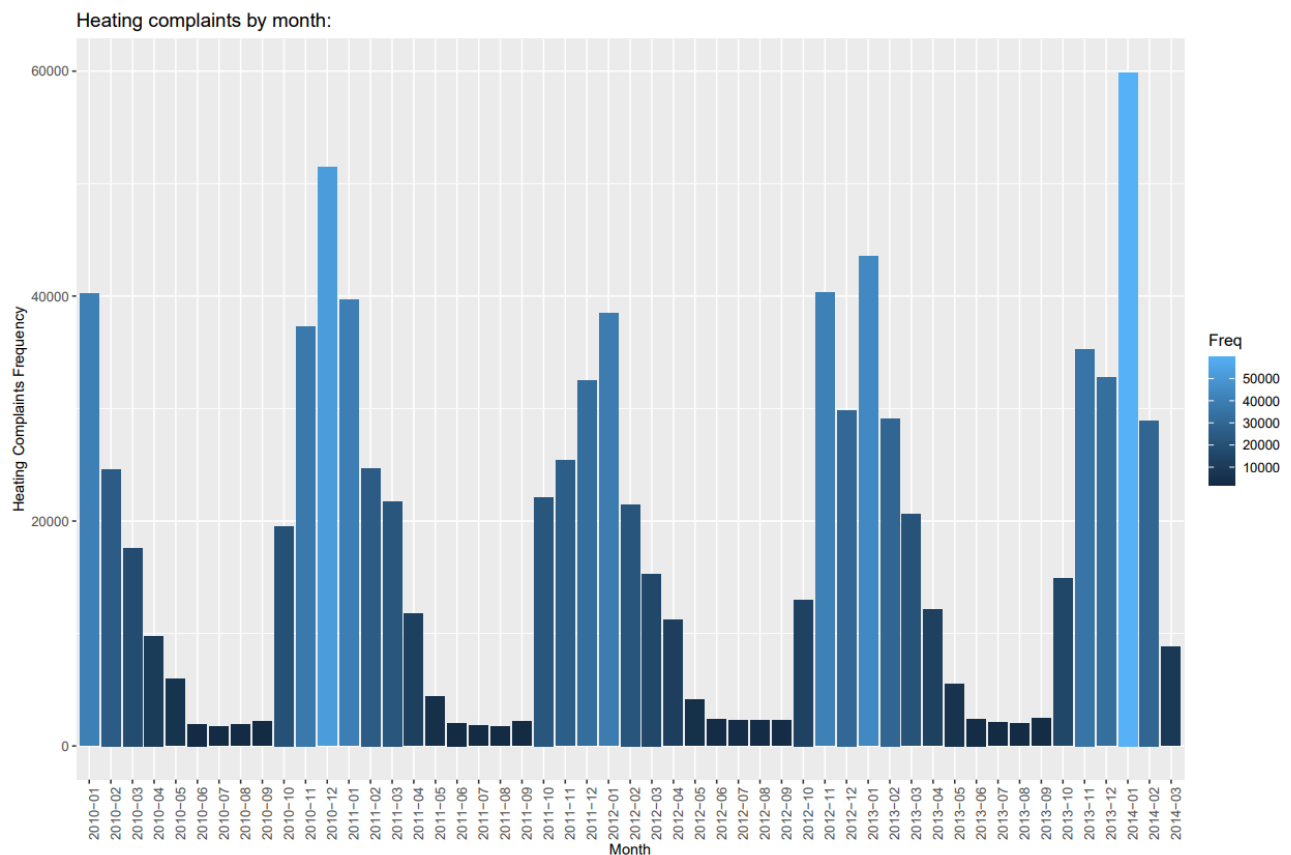


Fig 10

Here we see that compared to other months January had the highest number of calls. During the year 2014, the number of calls made was more than the average number of calls.

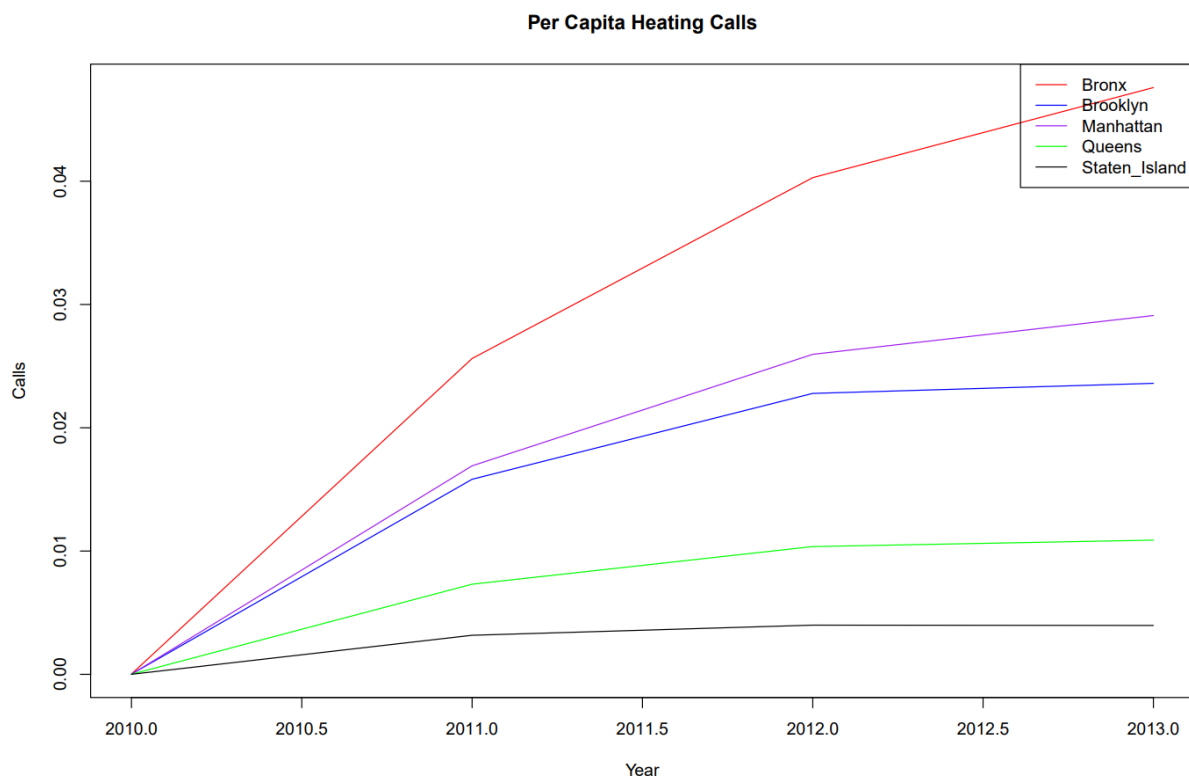


Fig 11

This Figure shows 311 per capita calls for Heating related complaints. It is interesting to see that while all the borough's per capita calls are increasing, the Bronx region is growing at a higher rate and Staten Island has the lowest growth rate. The visualization shows us the evolution of heating complaints per person in each borough over the years 2010 to 2013. The per capita calls for all boroughs increased at the maximum rate during the first year and then slowed down as the years passed.

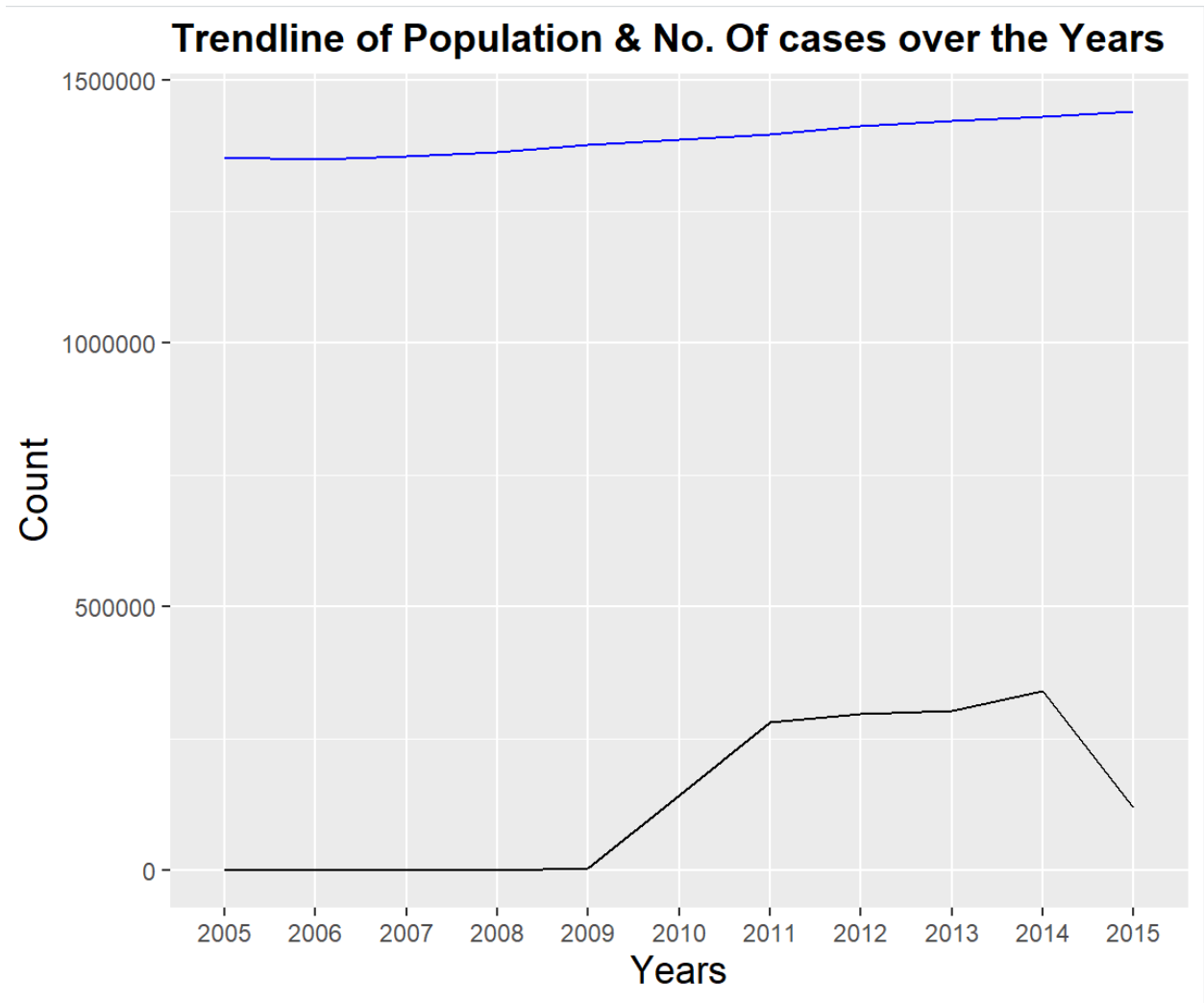


Fig 13

So from the above figure we can see the relationship between population and number of cases is directly proportional. As the population continues to increase, the number of cases also goes up.

CONCLUSION

We observe that the Department of Housing Preservation and Development (HPD) received the highest number of complaints. Heating Complaint Type falls under HPD agency. From the above report we see that the complaint type heating has the highest number of complaints. From the heatmap analysis of top 11 complaints, we notice that the boroughs Bronx and Brooklyn had more number of heating complaints over the other boroughs. This analysis is contradicting as we hoped Manhattan to have more number of complaints due to its population. This could be because Brooklyn and the Bronx have more low income per individual as compared to people of Manhattan because of which they might have poor heating systems in comparison to buildings in Manhattan.

While analyzing the number of calls monthly, we came to know that most of the calls were made between November and February. It is very obvious because it's winter time. We noticed that the highest number of calls were mainly made in the month of January. The month of January received the most number of calls in the year 2014 and that is because January 2014 was four degrees colder than average and with nearly 20 inches of snow, it was the coldest in five years.

From the new data set, we observe that population and number of complaints are directly proportional. Number of complaints was very less between 2005 and 2009, but there was a drastic increase since the year 2010. This trend could be due to the availability of easy access to technology(mobile) during recent years. Now people can register a complaint by a text message, call or even through social media. The agency works continuously to make government services more accessible to non-English speakers, with 311 Online available in more than 50 languages.

FUTURE SCOPE

In this report we've just focused on one complaint type which is Heating. In future, we can analyze other complaint types over different Boroughs to get a broader visualization of the dataset.

There were many columns that were excluded from our analysis but can be used in future analysis to extract more useful data.

APPENDIX

Data Dictionary for NYC311 DataSet

Unique Key	Unique Identifier of service request call in the open data set.
Created Date	Date the call was made and SR was created.
Closed Date	Date SR was closed by the responding agency.
Agency	Acronym of responding city government agencies.
Complaint Type	This is the first level of a hierarchy identifying the topic of the incident or condition.
Resolution.Action.Updated.Date	Date when the responding agency last updated the SR.
Status	Status of SR
Borough	Bronx, Brooklyn, Manhattan, Queens, Staten Island Borough Provided by the submitter and confirmed by geo validation.
Facility Type	If available, this field describes the type of city facility associated to the SR
Incident Type	10479 zip code depending on the area

Descriptor	This is associated to the Complaint Type, and provides further detail on the incident or condition. Descriptor values are dependent on the Complaint Type, and are not always required in SR.
X coordinate	Geo validated, X coordinate of the incident location.
Y coordinate	Geo validated, Y coordinate of the incident location.

Data Dictionary for Population DataSet

FIPS code (Federal Information Processing Standards)	Used for interchange between government agencies and other technical communities in order to ensure uniform practice and organization.
Geography	All the counties/boroughs in New York state.
Year	Year in which the population estimate is taken(ranges from 1970 to 2021).
Program Type	There are three program types: • Postcensal Population Estimate: Population estimate every year from 2010 to 2021. • Census Base Population: This happens every 10 years (1970,1980,1990,2000,2010,2020). • Intercensal Population Estimate: Population estimate every year from 1970 to 2009.
Population	Number of people living in the area.

Session Info

```
sessionInfo()
```

```
R version 4.2.1 (2022-06-23)
```

```
Platform: aarch64-apple-darwin20 (64-bit)
```

```
Running under: macOS Monterey 12.5
```

```
Matrix products: default
```

```
LAPACK:
```

```
/Library/Frameworks/R.framework/Versions/4.2-arm64/Resources/  
lib/libRlapack.dylib
```

```
locale:
```

```
[1]
```

```
en_US.UTF-8/en_US.UTF-8/en_US.UTF-8/C/en_US.UTF-8/en_US.UTF-8
```

```
attached base packages:
```

```
[1] stats      graphics  grDevices  utils      datasets  methods  
base
```

```
other attached packages:
```

```
[1] kableExtra_1.3.4      rworldxtra_1.01      rworldmap_1.3-6  
sp_1.5-0  
[5] knitr_1.39            wordcloud_2.6        RColorBrewer_1.1-3  
reshape_0.8.9  
[9] lubridate_1.8.0       viridis_0.6.2        viridisLite_0.4.0  
ggmap_3.0.0  
[13] devtools_2.4.3        usethis_2.1.6        data.table_1.14.2  
janitor_2.1.0  
[17] forcats_0.5.1         stringr_1.4.0        dplyr_1.0.9  
purrr_0.3.4  
[21] readr_2.1.2           tidyr_1.2.0          tibble_3.1.7  
ggplot2_3.3.6  
[25] tidyverse_1.3.1
```

```
loaded via a namespace (and not attached):
```

[1] bitops_1.0-7	fs_1.5.2	webshot_0.5.3
httr_1.4.3		
[5] bslib_0.3.1	tools_4.2.1	backports_1.4.1
utf8_1.2.2		
[9] R6_2.5.1	DBI_1.1.3	colorspace_2.0-3
withr_2.5.0		
[13] tidyselect_1.1.2		gridExtra_2.3
prettyunits_1.1.1	processx_3.7.0	
[17] curl_4.3.2	compiler_4.2.1	cli_3.3.0
rvest_1.0.2		
[21] xml2_1.3.3	sass_0.4.1	labeling_0.4.2
scales_1.2.0		
[25] callr_3.7.0	systemfonts_1.0.4	digest_0.6.29
foreign_0.8-82		
[29] svglite_2.1.0	rmarkdown_2.14	jpeg_0.1-9
pkgconfig_2.0.3		
[33] htmltools_0.5.2	sessioninfo_1.2.2	highr_0.9
dbplyr_2.2.1		
[37] fastmap_1.1.0	maps_3.4.0	rlang_1.0.3
readxl_1.4.0		
[41] rstudioapi_0.13	jquerylib_0.1.4	farver_2.1.1
generics_0.1.3		
[45] jsonlite_1.8.0	magrittr_2.0.3	dotCall64_1.0-1
Rcpp_1.0.8.3		
[49] munsell_0.5.0	fansi_1.0.3	lifecycle_1.0.1
stringi_1.7.6		
[53] yaml_2.3.5	snakecase_0.11.0	pkgbuild_1.3.1
plyr_1.8.7		
[57] maptools_1.1-4	grid_4.2.1	crayon_1.5.1
lattice_0.20-45		
[61] haven_2.5.0	hms_1.1.1	ps_1.7.1
pillar_1.7.0		
[65] rjson_0.2.21	pkgload_1.3.0	reprex_2.0.1
glue_1.6.2		
[69] evaluate_0.15	remotes_2.4.2	modelr_0.1.8
spam_2.9-0		
[73] png_0.1-7	vctrs_0.4.1	tzdb_0.3.0
RgoogleMaps_1.4.5.3		

[77]	cellranger_1.1.0	gtable_0.3.0	assertthat_0.2.1
	cachem_1.0.6		
[81]	xfun_0.31	broom_1.0.0	memoise_2.0.1
	fields_14.0		
[85]	ellipsis_0.3.2		