



An Analysis of NYC311 Noise Complaints during the year 2010

By



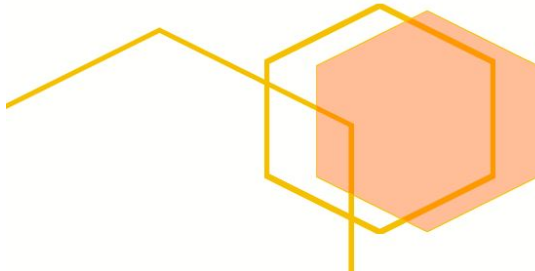
Raees Rizwan



Noorah Mohammed

ISTE 782: VISUAL ANALYTICS

SUBMITTED TO: MICK MCQUAID



Contents

Introduction	3
Motivation	3
Objective	4
Methodology	4
Visualizations	5
Where does noise pollution occur?	5
Most Populated Boroughs	5
Most Active Boroughs	6
Heat Map by Borough	7
What are contributable factors?	8
Commonly Occuring Noise Complaints	8
Loud Music/Party by Borough	9
Association Between Age Group and Loud Music	10
When does noise pollution take place?	11
Distribution of Noise by Month	12
Distribution of Noise by Season	13
Distribution of Noise by Days	14
Distribution of Noise by Hour	15
Conclusion	16
Future Research	16
Appendix	17
Data Discription	17
Session Info	18

Introduction



Non-emergency 311 call systems, used increasingly in U.S. cities, offer numerous advantages. New York City receives 311 calls for non-emergency services from its residents, businesses, and visitors. NYC311 is available 24/7. In addition, the general public can access their website to submit queries in more than 50 languages and can get help in more than 175 languages by calling 311. They provide citizens a quick, convenient way to kvetch about problems in their neighborhoods, and get a response. NYC311's fundamental goal is to provide the public with swift and easy access to all New York City government services and information while offering the best customer service. In addition, they enable city governments to identify patterns and address issues proactively.

Motivation

There is an incidental boon: Huge volumes of accumulated data. A handful of data scientists have begun to peer this data as a treasure trove. Specifically, they have started combing through it to explore connections between variables by visualizing the data in more detail, which can tell us about the policies and culture of the cities in question.

Objective

The primary objective of this report is to perform valuable analysis by unveiling worthwhile visualizations on the 311 data, by anchoring on the complaint type and filtering on boroughs and years. Subsequently, develop insights from the visualizations that could benefit NYC authorities to optimize the allocation of its resources in certain areas of each borough. In simple words, this report aims to use data visualization techniques to understand the following:

- *The most common noise complaints.*
- *The most contributable factor for noise pollution.*
- *The hours, months, and years with the highest complaint frequencies.*
- *How the above features varied between five boroughs.*

Methodology

The initial dataset of NYC311 complaint data contained over 9 million rows coming from various government departments spanning from the year 2003 to 2015. We first enhanced the NYC311 service request data using various data cleaning techniques such as `tydr` and `dplyr`, as it contained many infelicities. Next, we constrained our analysis to a specific year by filtering our dataset by the year 2010. Subsequently, the population by age data set was joined to the NYC311 dataset to discover the connection between age and complaint type. We sought for a way to consolidate complaint types to make the analysis more interpretable. For example, for the “Noise” category, we grouped the complaints through regular expressions on any type that contained the words “Noise”. After this consolidation, we then sought to uncover more insight into seasonality and timing of incident reporting. With that, we also generated additional features from the timestamps to be able to filter by month, day of the year and hour.

Following are some basic questions on noise pollution in New York City in 2010 that we seek to answer as we progress:

1. **WHERE:** Where does noise pollution occur?
2. **WHAT:** What are contributable factors?
3. **WHEN:** When does noise pollution take place?

To answer these questions, several tasks needed to be done. It includes from a basic data cleaning to preparing data for visualization.

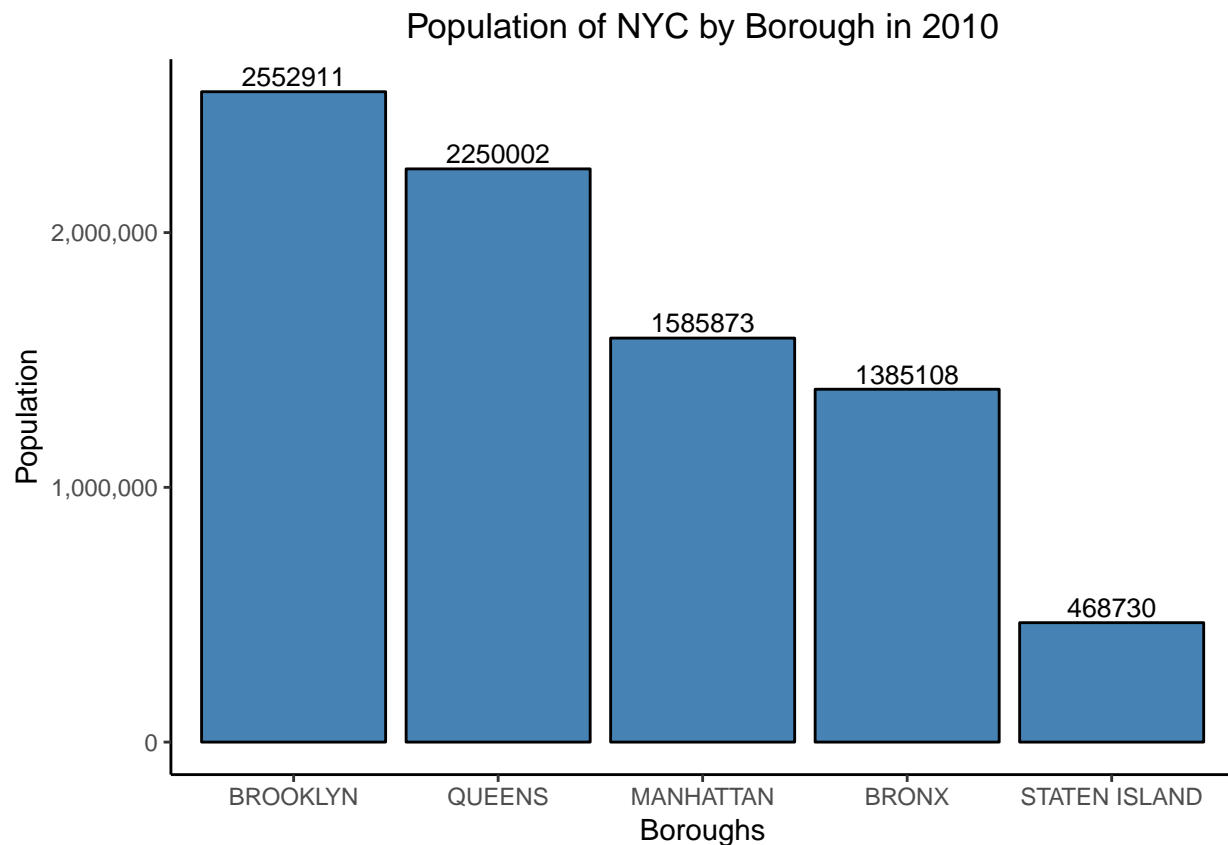
Visualizations

Where does noise pollution occur?

Manhattan, Bronx, Brooklyn, Queens, and Staten Island are 5 boroughs in New York City. To analyze which borough contributes the most towards noise pollution, it is important to look at the number of population in each borough. It is apposite to deem that the higher population in a borough results in a higher number of complaints.

Most Populated Boroughs

To support the above-mentioned notion, a bar chart is plotted below to show the population of NYC by borough.

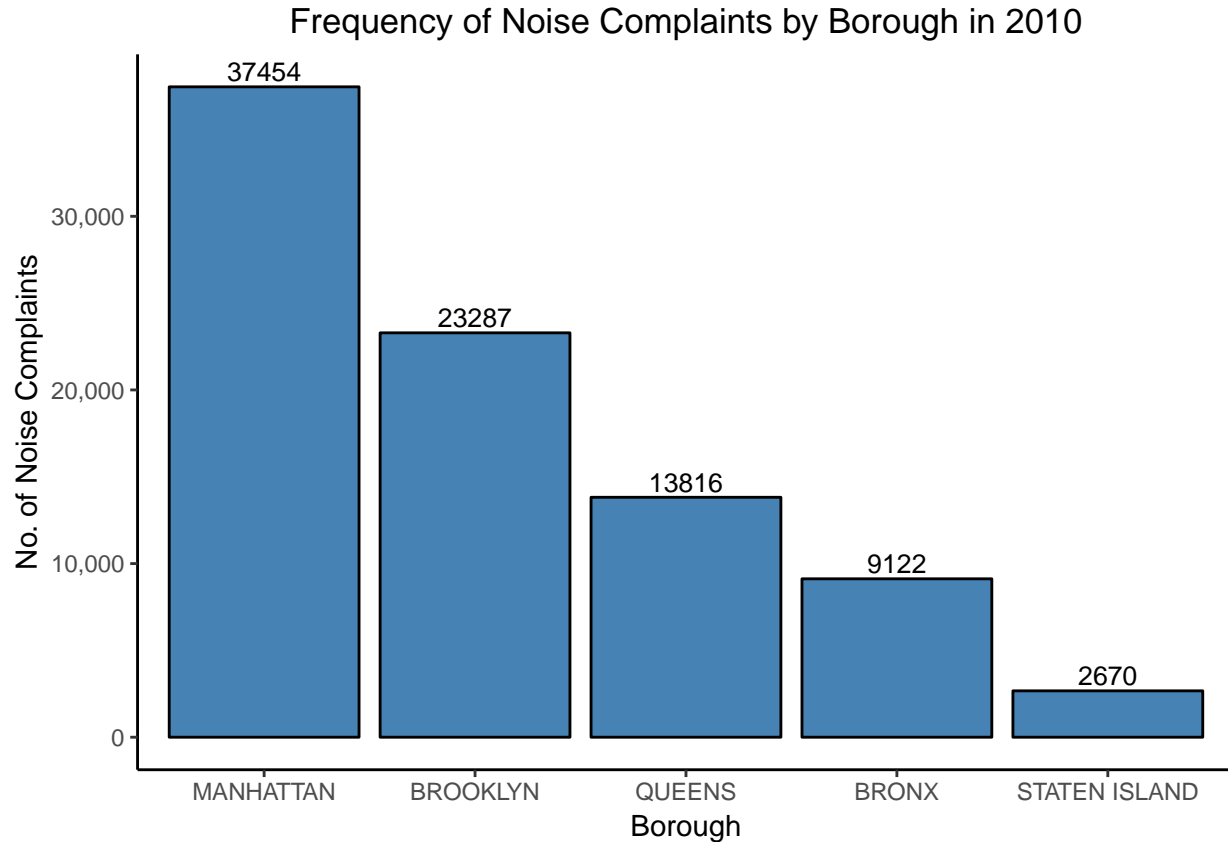


From the above bar chart, we can note the following:

- Brooklyn and Queens are the top 2 boroughs with highest population count respectively.
- Manhattan and Bronx has relatively moderate population count.
- The least population count is perceived in the borough of Staten Island.

Most Active Boroughs

To understand which borough is responsible for generating the most noise pollution in NYC, a bar chart is plotted below illustrating the frequency of noise complaints by each borough throughout the year 2010.



From the above bar chart, we can note the following:

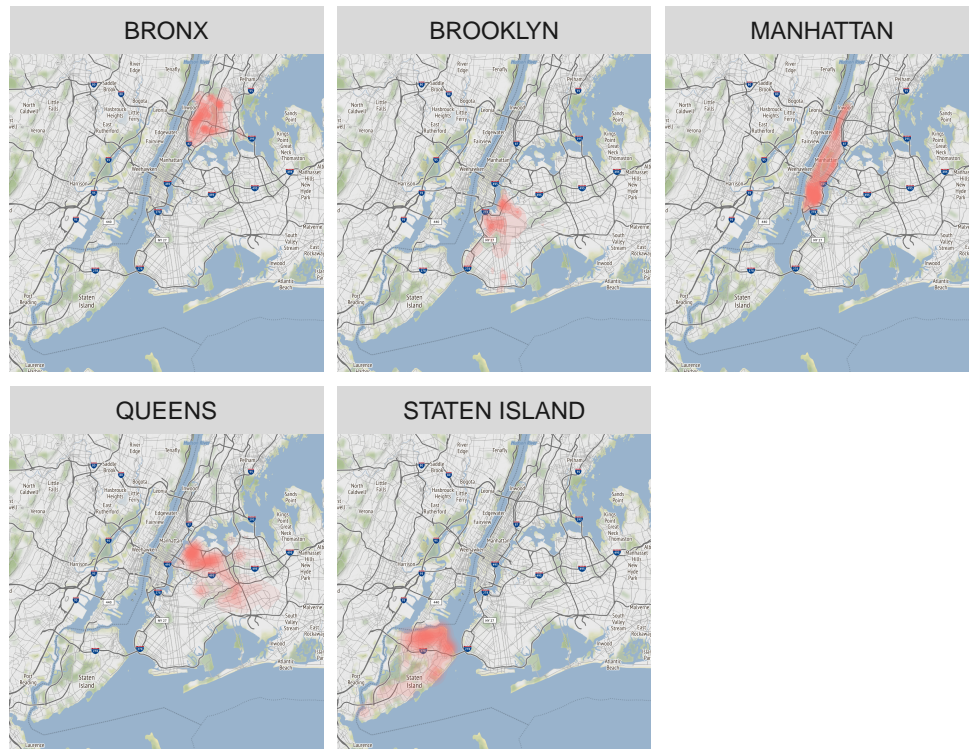
- Most amounts of noise complaints are received from Manhattan.
- Brooklyn and Queens are ranked second and third respectively in contributing noise pollution.
- The least amounts of noise complaints are received from Staten Island.

While Manhattan is considered as the leading borough in contributing noise pollution, some other interesting trends were notable. From the previous visual which depicts the total population by borough in 2010, it can be observed that Brooklyn and Queens have the highest number of the population respectively. However, the noise complaints are much higher in Manhattan relative to other Boroughs. This contradicting result inspired to analyze into more detail. Hence, the population of New York City was further investigated by different age groups in the later stages.

Heat Map by Borough

After understanding the contribution of noise pollution by each borough, it interesting to investigate which part of the boroughs do the noise complaints come from. The following heat maps show the spatial distribution of noise pollution by each borough.

Distribution of Noise Complaints by Borough in 2010



From the above maps, we can note the following:

- For Manhattan, noise pollution is distributed almost the whole of the borough.
- For Brooklyn and Staten Island, noise pollution is distributed in northern areas of those boroughs.
- For Bronx, noise pollution is distributed in the southern-west area.
- For queens, noise pollution is distributed in the western area.

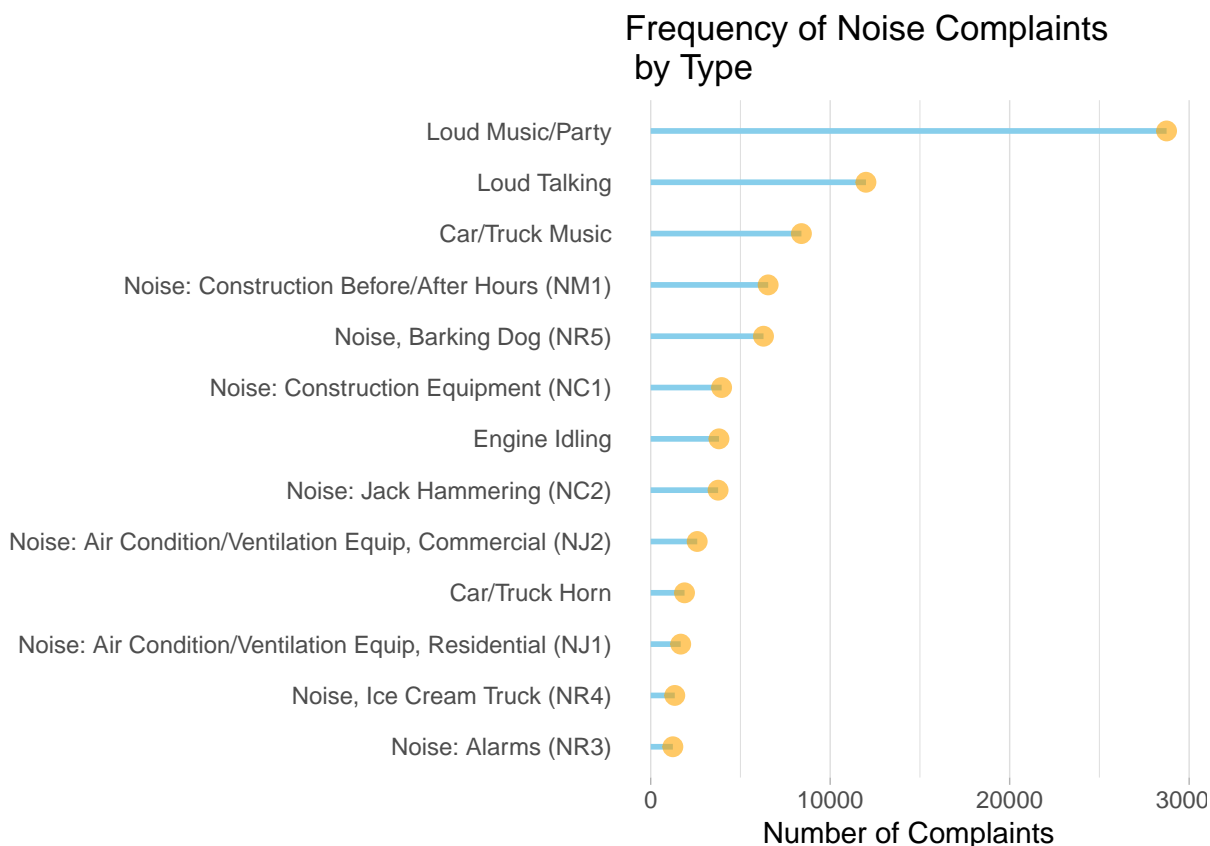
By now, we have answered the first question regarding where noise pollution occurs the most. These analyses might encourage NYC authorities to focus more on above-mentioned locations to tackle noise pollution problem. Subsequently, the report attempts to investigate the most contributable factors to noise pollution.

What are contributable factors?

We have developed an understanding of the locations of the noise being produced. Now, it is worth to find the source of noise pollution in New York City. There are over 35 complaint type descriptors in the dataset that I grouped into 11 categories based on their significance.

Commonly Occuring Noise Complaints

Below plot illustrates the distribution of noise pollution by noise type.



Note: The figure above is clipped from the right after rendering to pdf. The x-axis value at the rightmost is 30000.

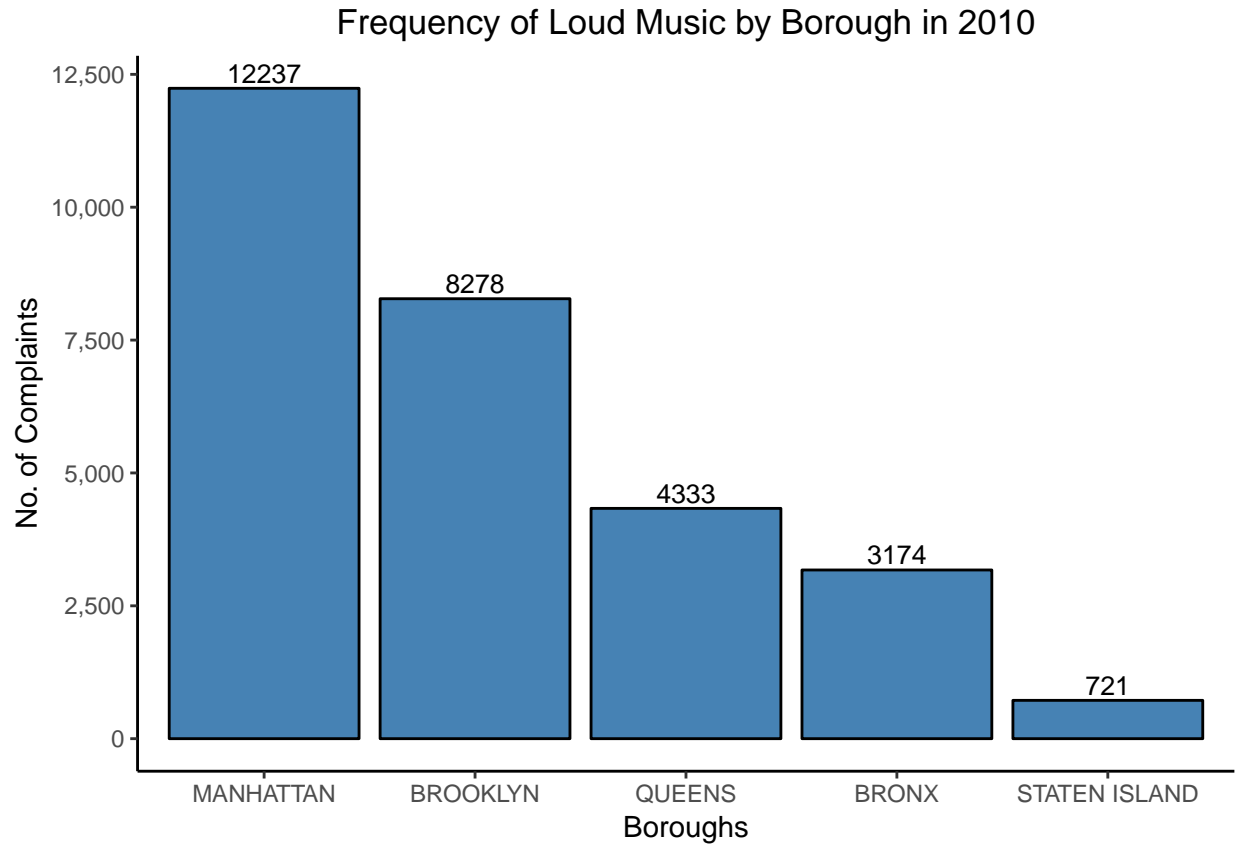
From the above figure, we can note the following:

- Loud Music/Party was by far the most frequent of all the noise related complaints.
- Top three complaint types across all boroughs are Loud Music/Party, Loud Talking and Car/Truck Music respectively.

Since Loud Music/Party shows the most dominance in the noise type category, let's dig deeper into Loud Music/Party category to get some interesting insights.

Loud Music/Party by Borough

From the previous analyses, we learned that **Loud Music/Party** is the most contributable factor to noise pollution in New York City. Investigating loud music by borough may provide us with insights regarding its source. A bar chart is plotted below illustrating the frequency of loud music by each borough throughout the year 2010.



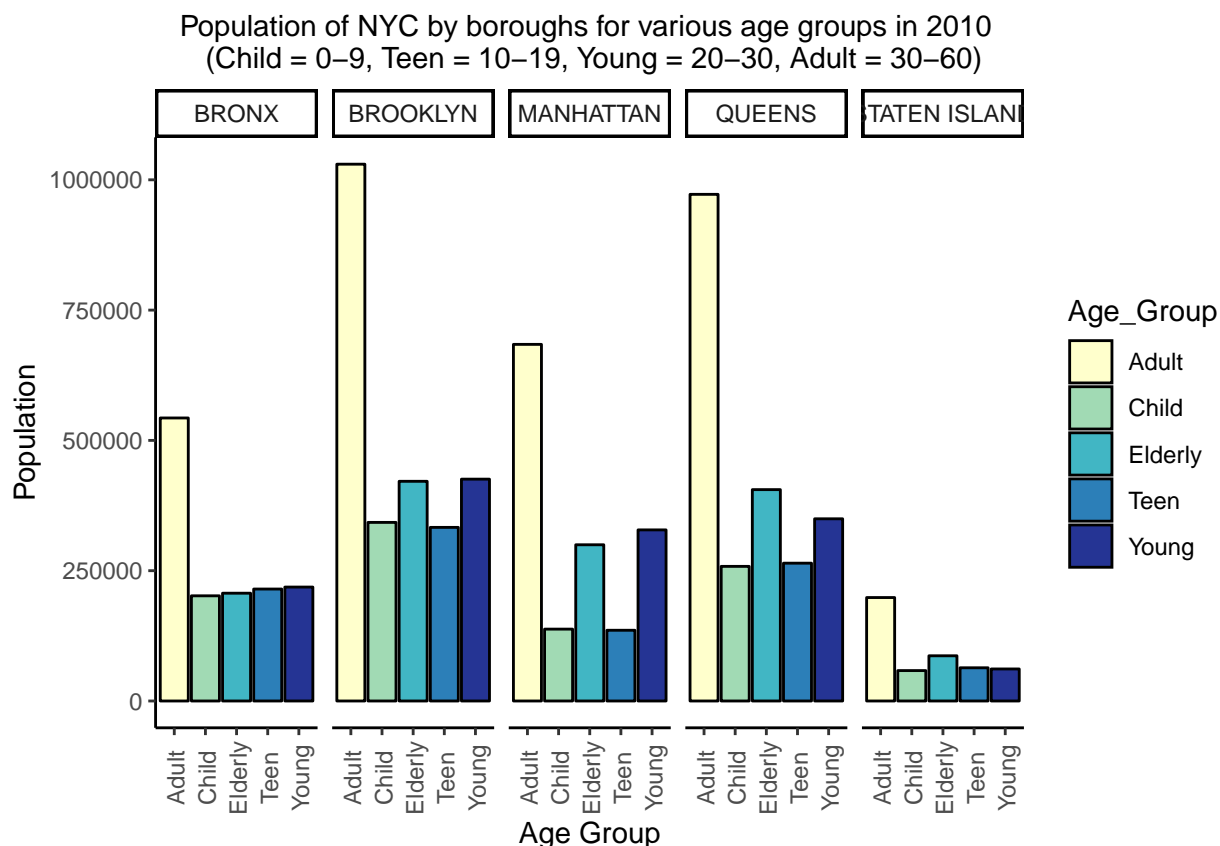
From the above bar chart, we can note the following:

- Manhattan produced the highest number of **Loud Music** complaints in 2010. This is expected as Manhattan receives the most noise complaints relative to the other boroughs.
- Brooklyn and Queens are ranked second and third respectively in contributing to **Loud Music** complaints.
- The least amounts of **Loud Music** complaints are received from Staten Island.

Manhattan seems to be the borough of interest to investigate the source for **Loud Music** complaints. Therefore, in the following section, a new data of population by age group was introduced to examine the relation between **Age Group** and **Loud Music**.

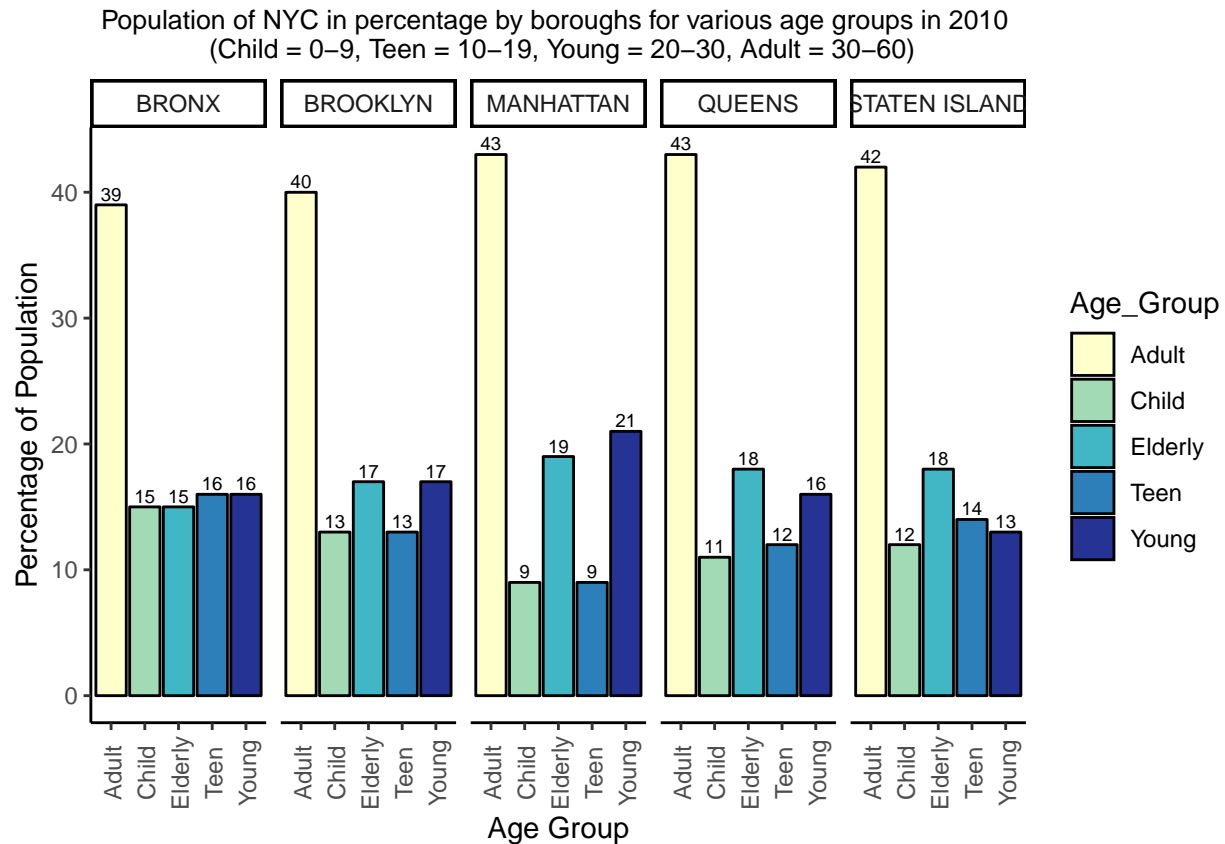
Association Between Age Group and Loud Music

It is understandable that young undergraduates and graduates between the age of 20-30 are linked heavily with loud parties. Therefore, we can expect a positive correlation between the number of young people and loud music. To support this hypothesis, a bar chart is plotted illustrating the population by the borough for various age groups. For our hypothesis to be true, the number of young people must be high in the borough of Manhattan relative to the other boroughs.



The above-resulted bar chart does not provide conclusive evidence to support our initial hypothesis. On the contrary to what we expected, Brooklyn and Queens have a higher number of youngsters than Manhattan. This result is plausible as they have the topmost number of the population respectively. However, we expect taking a step further by analyzing the age group based on the percentage of the population in each borough may give us some decisive evidence to support our assumption.

The below plot illustrates the population of New York City in different boroughs by various age groups.



From the above plot, it is evident that Manhattan has the highest percentage of young people relative to other boroughs. This result gives us marginal evidence to support our initial attempt to show a positive relationship between young people and loud music. However, the close values of other boroughs do not explain the significantly large amount of loud music complaints in Manhattan.

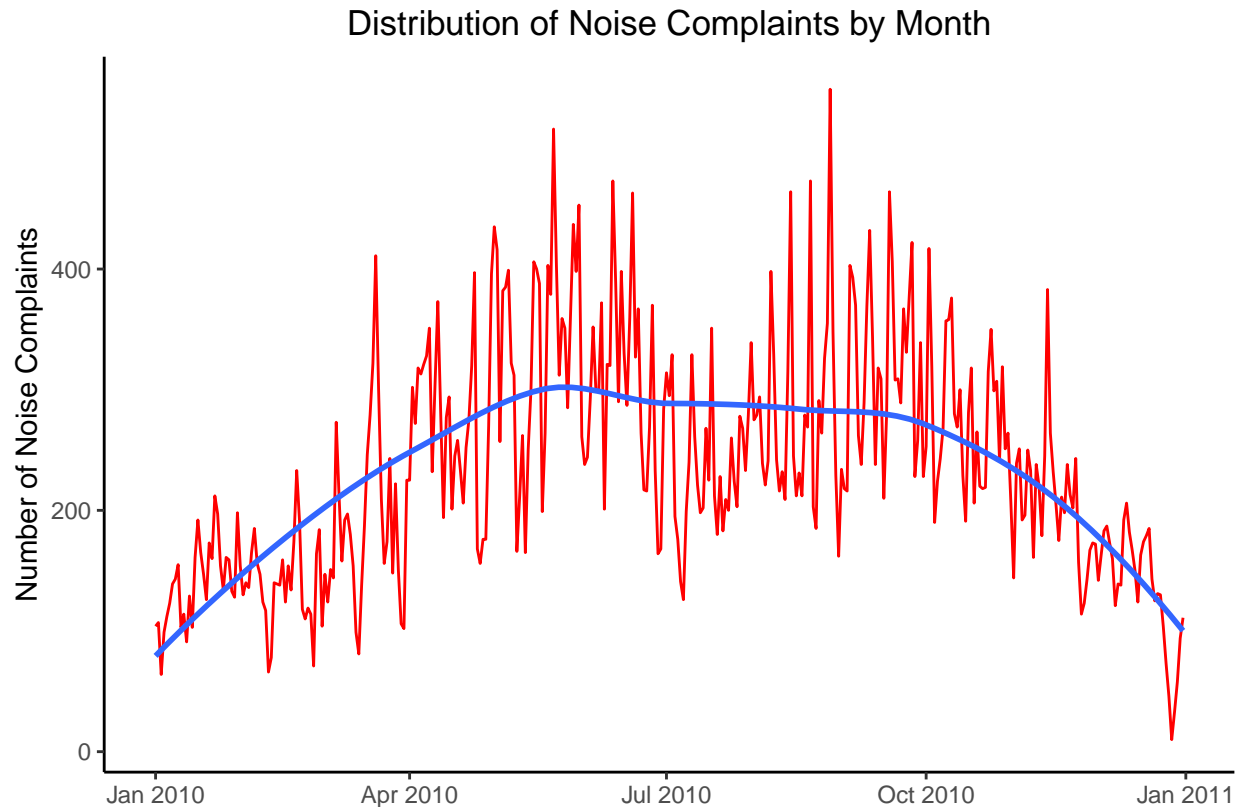
Note: Although the visualizations did not provide any conclusive insights, it was worth for the research experience

When does noise pollution take place?

To answer the question of when does noise pollution take place, we must analyze all the noise complaints by month, day and hour. Therefore, this section expounds the distribution of noise complaints by Month, Season, Day and Meridiem.

Distribution of Noise by Month

The below line chart describes the distribution of all noise complaints by month during the year 2010.



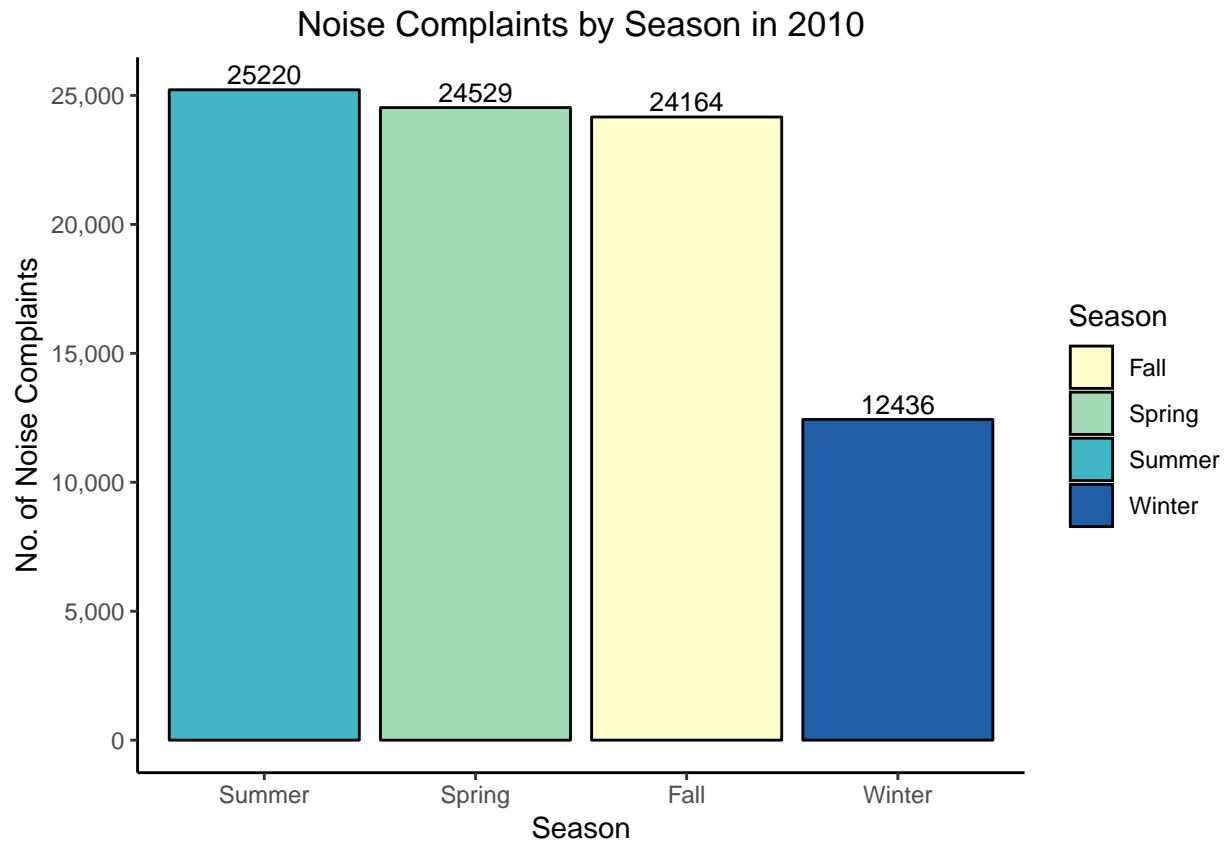
From the above line chart, we can note the following:

- The noise complaints are not evenly distributed across months.
- The highest peaks were recorded during the month of May, June, Aug, and Sept. This suggests that most of the noise complaints were received during the middle months of the year.
- The lowest peaks were recorded during the month of Jan, Feb, Mar, Nov, and Dec. This suggests that the beginning and end months of the year are relatively quieter.

The above visualization consist of a specific pattern that entices to analyze the noise complaints by seasons.

Distribution of Noise by Season

From the above visualization, we can infer that noise complaints may not be evenly distributed across seasons as well. Therefore, following bar chart illustrates the distribution of noise complaints by season can be useful to develop some insights.



From the above bar chart, we can note the following:

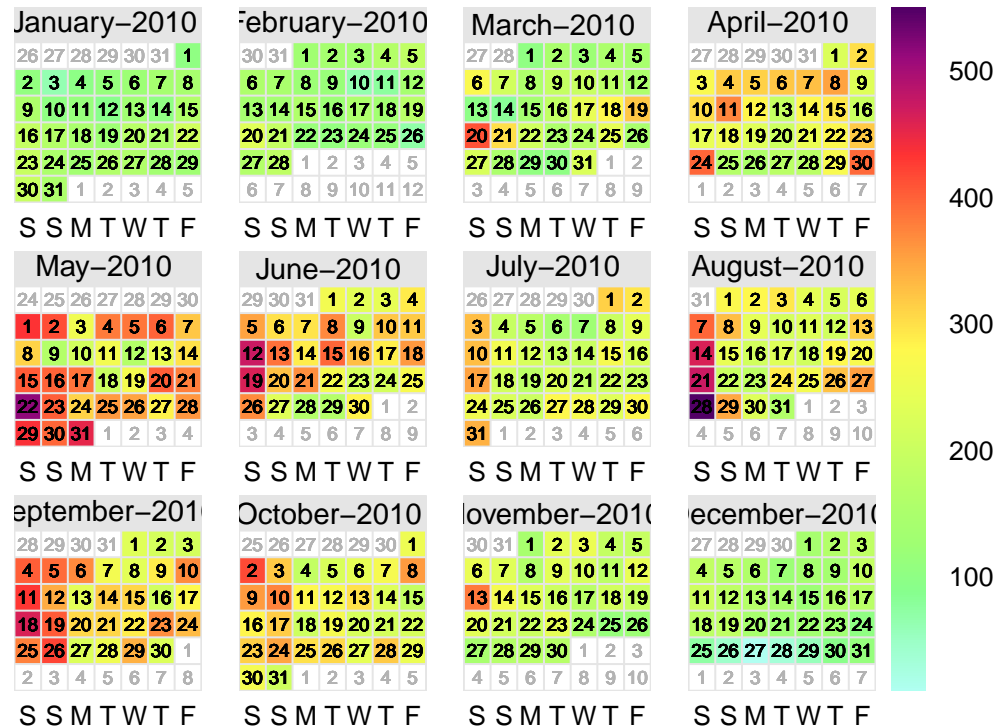
- The highest number of noise complaints were filed during the summer season.
- Spring and Fall seasons also received noise complaints closer to the summer season.
- The lowest number of noise complaints were filed during the winter season.

The significant drop of noise complaints during the winter season indicates that people tend to be quieter in the months with cold temperatures.

Distribution of Noise by Days

The calendar below shows the level of calls by days throughout the year 2010.

Noise Complaints by Day in 2010



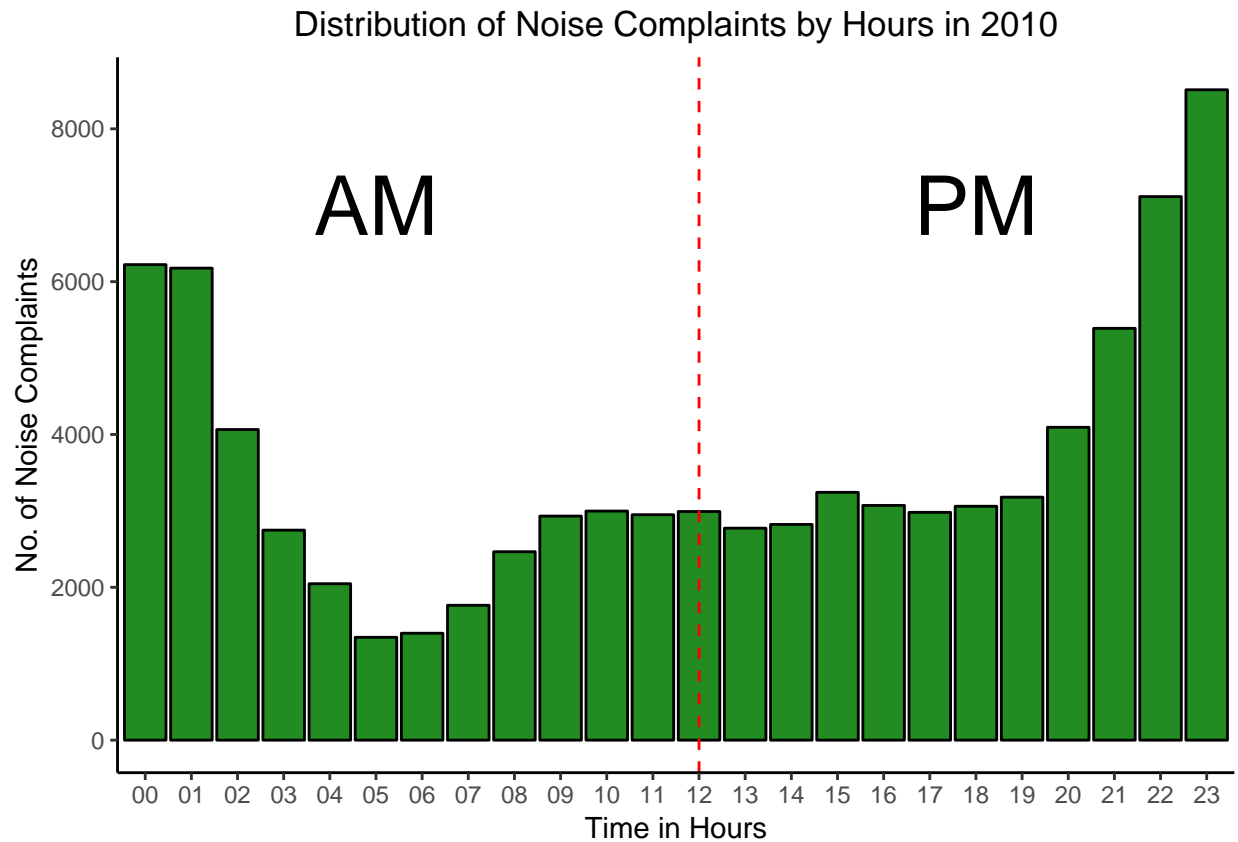
From the above calendar, we can note the following:

- Month of Apr, May, Jun, and Sep generate more noise complaints. This result is consistent with the season analysis performed earlier.
- Saturday and Sunday have the highest noise complaint volume.

Since Saturday and Sunday are weekends in New York City, the above result showing the highest volume of noise complaints is persuasive.

Distribution of Noise by Hour

From the previous visualizations, we have developed an understanding of when the noise pollution occurs the most. In addition to that, the below bar chart is plotted to describe the distribution of noise complaints during each hour of the day.



From the above bar chart, we can note the following:

- Disproportionately more complaints were filed in the early morning and nighttime hours than during the day.
- The total number of noise complaints received after noon is greater than that of before noon.

These results are consistent with the fact that most people party during the weekends from midnight to early morning.

Conclusion

This report endeavored to analyze the noise pollution in New York City based on three basic questions - Where, What, and When.

- WHERE: Maximum amounts of noise complaints are received from Manhattan. Contemplating spatial distribution by borough, noise pollution for Manhattan and Queens is distributed almost the whole of each borough. On the other hand, for Brooklyn and Staten Island, noise pollution is distributed in northern areas of those boroughs. For the Bronx, noise pollution is distributed in the southern area.
- WHAT: The most contributable factor to noise pollution in New York City is Loud Music/Party.
- WHEN: From the calendar heat map, noise pollution primarily occurs during the weekend. It might be because many people invite friends to their houses and have a party playing music, making some serious noise. Loud noise complaints start to peak from May- September when it becomes warmer, schools are off, and there are more activities to do outside. May and September have the highest volume of calls.

Most calls tend to be placed during 10 pm- 1 am through Friday and Saturday, and we would recommend seasonal staffing to receive calls during the summer months, particularly in the evenings and on weekends. The NYPD might also want to come up with a program or adjust hours, since they are the agency usually tasked with investigating calls, particularly from what makes up most of the calls: Loud Noise or Loud Music complaints. NYC311 can alter staffing to compare with hours and long periods of high objection volume.

The results of these visualizations and analyses uncover critical knowledge into the volume and distribution of noise complaints filed with NYC311. Understanding these patterns offers a few different ways by which NYC311 can build and enhance their objective of delivering the general population with quick and simple access to the numerous administrative services available. In whole, it seems 311's activities can make some good efficiency gains from the insights above.

Future Research

The short period of time assigned for research on this task limited the scope of the analyses. Future investigations of these data will benefit from incorporating demographic and socioeconomic data of the complaint areas. If given more time, it would be interesting to relate crime by borough data with the 311 noise data and find some correlation between them.

Appendix

Data Discription

Variable	Data Type	Description
Created Month	Char	Month service request was created
Created Day	Char	Day service request was created
Created Year	Char	Year service request was created
Time	Char	Time service request was created
Meridiem	Char	Meridiem (AM/PM) service request was created
Agency	Char	This is acronym of responding City Government Agency
Complaint Type	Char	This is the first level of a hierarchy identifying the topic of the incident or condition. Complaint Type may have a corresponding Descriptor (below) or may stand alone
Descriptor	Char	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
Incident Zip	Char	5-digit incident location zip code, provided by geo validation
Borough	Char	Divisions of New York City Provided by the submitter and confirmed by geovalidation
Latitude	numeric	Geo based lat value of the incident location
Longitude	numeric	Geo based long value of the incident location
Population	numeric	Number of population
Age	Char	Classes of population age in the range of 5

Session Info

```
sessionInfo()
```

```
## R version 3.6.0 (2019-04-26)
## Platform: x86_64-w64-mingw32/x64 (64-bit)
## Running under: Windows 10 x64 (build 16299)
##
## Matrix products: default
##
## locale:
## [1] LC_COLLATE=English_United States.1252
## [2] LC_CTYPE=English_United States.1252
## [3] LC_MONETARY=English_United States.1252
## [4] LC_NUMERIC=C
## [5] LC_TIME=English_United States.1252
##
## attached base packages:
## [1] stats      graphics  grDevices  utils      datasets  methods   base
##
## other attached packages:
##  [1] ggmap_3.0.0          RColorBrewer_1.1-2 openair_2.6-5
##  [4] scales_1.0.0         tinytex_0.13       kableExtra_1.1.0
##  [7] xtable_1.8-4         data.table_1.12.2  forcats_0.4.0
## [10] stringr_1.4.0        dplyr_0.8.1        purrr_0.3.2
## [13] readr_1.3.1          tidyr_0.8.3        tibble_2.1.1
## [16] ggplot2_3.1.1        tidyverse_1.2.1
##
## loaded via a namespace (and not attached):
##  [1] Rcpp_1.0.1           lubridate_1.7.4     lattice_0.20-38
##  [4] png_0.1-7            assertthat_0.2.1    digest_0.6.19
##  [7] R6_2.4.0             cellranger_1.1.0    plyr_1.8.4
## [10] backports_1.1.4      evaluate_0.14       httr_1.4.0
## [13] pillar_1.4.1         RgoogleMaps_1.4.3   rlang_0.3.4
## [16] curl_3.3             lazyeval_0.2.2      readxl_1.3.1
## [19] rstudioapi_0.10      hexbin_1.27.3       Matrix_1.2-17
## [22] rmarkdown_1.13       labeling_0.3        splines_3.6.0
## [25] webshot_0.5.1        munsell_0.5.0       broom_0.5.2
## [28] compiler_3.6.0       modelr_0.1.4        xfun_0.7
## [31] pkgconfig_2.0.2      mgcv_1.8-28         htmltools_0.3.6
## [34] tidyselect_0.2.5     viridisLite_0.3.0   crayon_1.3.4
## [37] withr_2.1.2          bitops_1.0-6        MASS_7.3-51.4
## [40] grid_3.6.0           nlme_3.1-139        jsonlite_1.6
```

## [43]	gtable_0.3.0	magrittr_1.5	cli_1.1.0
## [46]	stringi_1.4.3	reshape2_1.4.3	mapproj_1.2.6
## [49]	latticeExtra_0.6-28	xml2_1.2.0	generics_0.0.2
## [52]	rjson_0.2.20	tools_3.6.0	glue_1.3.1
## [55]	jpeg_0.1-8	maps_3.3.0	hms_0.4.2
## [58]	yaml_2.2.0	colorspace_1.4-1	cluster_2.0.8
## [61]	rvest_0.3.4	knitr_1.23	haven_2.1.0