311 Service Calls of New York City

Yisong Tao

10/14/2016

311 is a public service phone number of many communities in North America, which provides access to many non-emergency municipal services[1]. New York City now supports both phone call and online interfaces to 311 service, and provide open access to 311 service requests data from 2010 to present. There are several well-written articles online utilizing this dataset to come up with insights into data-driven urban management[2] or specific issues in NYC[3].

The questions I want to answer by analyze and visualization of New York 311 data are: what are the top complaints? Are they the same across 5 boroughs? Try to characterize the top complaints.

```
setwd("~/Desktop/Project1")
library(dplyr)
library(ggplot2)
library(ggthemes)
```

The NYC 311 service requests data can be accessed from https://nycopendata.socrata.com/Social-Services/ 311-Service-Requests-from-2010-to-Present/erm2-nwe9/data. The dataset for this project was downloaded on 10/01/2016, which was a 9.15 Gb csv file containing over 137 million requests from 2010-01-01 to 2016-09-30 02:11:54.

```
fileurl <- "https://nycopendata.socrata.com/api/views/erm2-nwe9/rows.csv?accessType=DOWNLOAD"
file <- "data.csv"
download.file(fileurl, file, method = "curl" )
data <- data.table::fread(file, sep = ",", header = TRUE,
                          stringsAsFactors = FALSE,
                          na.strings = c("N/A", "", "NA", "Unspecified"))
nrow(data)            #dataset contains 13720953 observations
names(data)           #dataset contanins 53 variables
saveRDS(data, "all_data.RDS")
data_311 <- readRDS("all_data.RDS")
```

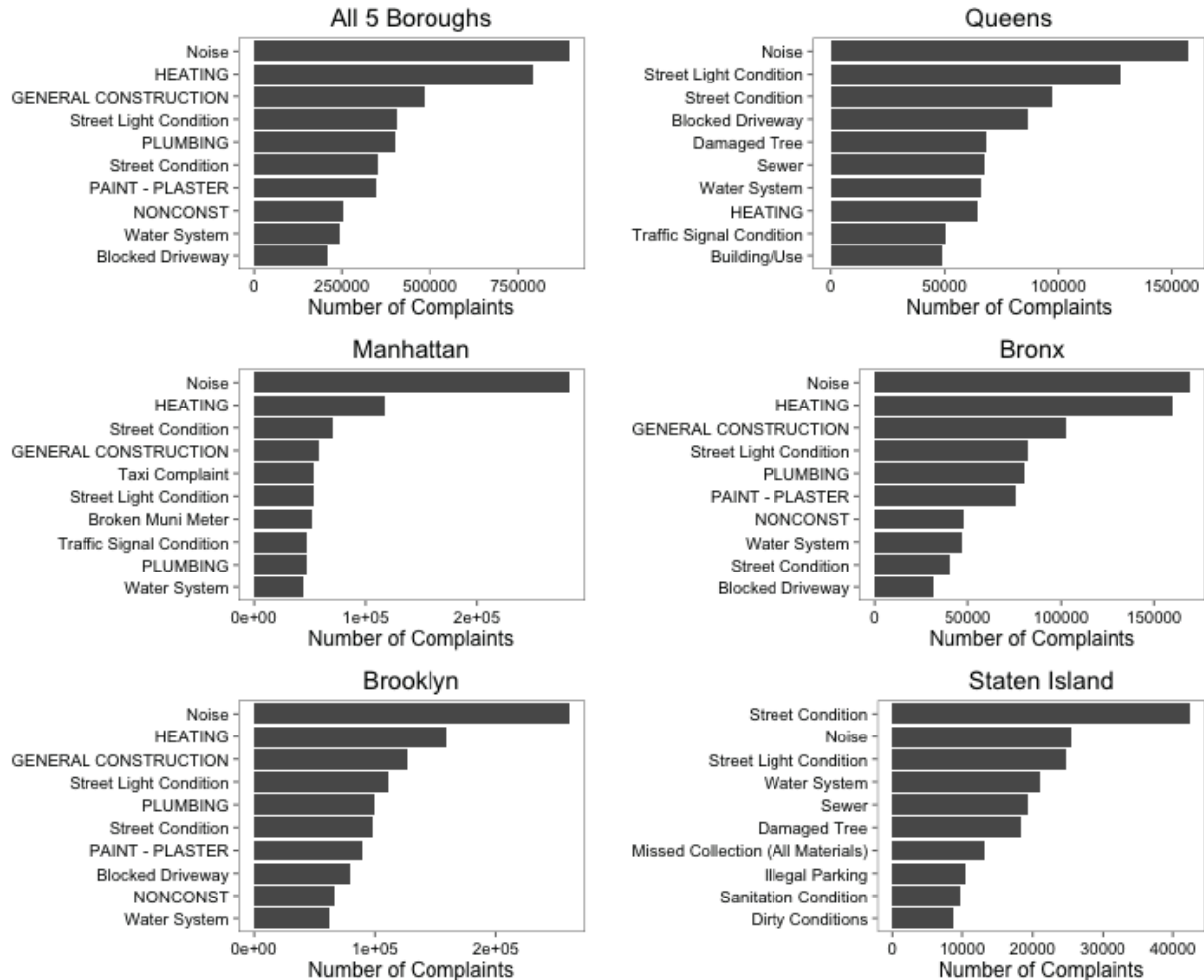## 1. What are the most complained issues in New York?

```
#data_311 <- readRDS("all_data.RDS")
#åall_complaints <- data_311 %>%select(2, 6, 25) #Select "Complaint Type" and "Borough" columns
#all_complaints$`Created Date` <- as.POSIXct(strptime(all_complaints$`Created Date`, "%m/%d/%Y %I:%M:%S
#all_complaints$Year <- format(all_complaints$`Created Date`, "%Y")
#all_complaints <- filter(all_complaints, Year != "2016")
#all_complaints <- filter(all_complaints, Year != "2015")
#all_complaints <- filter(all_complaints, Year != "2014")
#saveRDS(all_complaints, "all_complaints_1013.RDS")
all_complaints <- readRDS("all_complaints_1013.RDS")
all_complaints$`Complaint Type`[grepl("^Noise.*", all_complaints$`Complaint Type`)] <- "Noise"
all_complaints <- all_complaints %>%
    group_by(Borough, `Complaint Type`) %>% summarise(Count = n())
all_complaints_NY <- all_complaints %>%
```

```r
    group_by(`Complaint Type`) %>%
    summarise(Count = sum(Count)) %>%
    arrange(desc(Count))
top10_complaints_NY <- top_n(all_complaints_NY, 10, Count)
figure1 <- function(df){
    return(ggplot(df) + geom_bar(aes(x=reorder(`Complaint Type`,Count) , y = Count),
            stat = "identity") + theme_few() +
    xlab("") + ylab("Number of Complaints") + coord_flip() )
}
p_top10_complaints_NY  <- figure1(top10_complaints_NY) + ggtitle("All 5 Boroughs")
top10_complaints_borough <- top_n(group_by(all_complaints, Borough), 10, Count)
top10_complaints_borough <- arrange(top10_complaints_borough, desc(Borough))
top10_complaints_Man <- filter(top10_complaints_borough, Borough == "MANHATTAN")
p_top10_complaints_Man  <- figure1(top10_complaints_Man) + ggtitle("Manhattan")
top10_complaints_Qns <- filter(top10_complaints_borough, Borough == "QUEENS")
p_top10_complaints_Qns  <- figure1(top10_complaints_Qns) + ggtitle("Queens")
top10_complaints_Bn <- filter(top10_complaints_borough, Borough == "BROOKLYN")
p_top10_complaints_Bn  <- figure1(top10_complaints_Bn) + ggtitle("Brooklyn")
top10_complaints_Brx <- filter(top10_complaints_borough, Borough == "BRONX")
p_top10_complaints_Brx  <- figure1(top10_complaints_Brx) + ggtitle("Bronx")
top10_complaints_SI <- filter(top10_complaints_borough, Borough == "STATEN ISLAND")
p_top10_complaints_SI  <- figure1(top10_complaints_SI) + ggtitle("Staten Island")
#### The following function allos multiple graphs in one plot, the following multiplot
#### function is from http://www.cookbook-r.com/Graphs/Multiple_graphs_on_one_page_(ggplot2)/
multiplot <- function(..., plotlist=NULL, file, cols=1, layout=NULL) {
  library(grid)
  plots <- c(list(...), plotlist)
  numPlots = length(plots)
  if (is.null(layout)) {
    layout <- matrix(seq(1, cols * ceiling(numPlots/cols)),
                  ncol = cols, nrow = ceiling(numPlots/cols))
  }
  if (numPlots==1) {
    print(plots[[1]])
  } else {
    grid.newpage()
    pushViewport(viewport(layout = grid.layout(nrow(layout), ncol(layout))))
    for (i in 1:numPlots) {
      matchidx <- as.data.frame(which(layout == i, arr.ind = TRUE))
      print(plots[[i]], vp = viewport(layout.pos.row = matchidx$row,
                                    layout.pos.col = matchidx$col))
    }
  }
}
####
multiplot(p_top10_complaints_NY, p_top10_complaints_Man, p_top10_complaints_Bn,
        p_top10_complaints_Qns, p_top10_complaints_Brx, p_top10_complaints_SI,
        cols=2)
```

The above barplot shows the top 10 complained issues from New York City and each of the 5 boroughs. Noise is the most complained issue in New York City(11.84% of all complaints) and each borough except for Staten Island, where noise is the second most complained after street condition. Heating, street light condition and street condition are also among the most complained issue.

2. Regarding noise complaints, which neighborhood/zip codes are making most of the noises in New York City?

```r
noise_complaints <- select(data_311, 6, 7, 9, 25, 51, 52)
noise_complaints$Descriptor <- as.factor(noise_complaints$Descriptor)
noise_complaints <- filter(noise_complaints, grepl("Noise", `Complaint Type`))
noise_complaints$Type <- NA
noise_complaints$Type[grepl("Dog|Animals",
                            noise_complaints$Descriptor)] <- "Dog and other animals"
noise_complaints$Type[grepl("Air Condition|air condition",
                            noise_complaints$Descriptor)] <- "Air Conditioner"
noise_complaints$Type[grepl("Banging",
                            noise_complaints$Descriptor)] <- "Banging/Pounding"
noise_complaints$Type[grepl("Truck|Vehicle|Boat|Private Carting|Engine|Flying|Hovering|Honking", noise_
noise_complaints$Type[grepl("Construction|Jack Hammering|Manufacturing", noise_complaints$Descriptor)] <
```

3

```
noise_complaints$Type[grepl("Loud Music|Television|Talking",
                            noise_complaints$Descriptor)] <- "Music/TV/Talking"
noise_complaints$Type[grepl("Alarm",
                            noise_complaints$Descriptor)] <- "Alarm"
noise_complaints$Type[is.na(noise_complaints$Type)] <- "Other"
noise_complaints<-select(noise_complaints, 2, 3, 4, 5, 6, 7)
noise_complaints$`Incident Zip` <- substr(noise_complaints$`Incident Zip`, 1, 5)
nyc_list <- readRDS("nyc_list.RDS") ## load a list of NYC zip codes
noise_complaints <- filter(noise_complaints, `Incident Zip` %in% nyc_list)
saveRDS(noise_complaints, "noise_complaints.RDS")
```

```
noise_complaints <- readRDS("noise_complaints.RDS")
noise_sum<- as.data.frame(table(noise_complaints$Descriptor))
noise_sum <- noise_sum[noise_sum$Freq > 0,]
noise_sum <- noise_sum[order(-noise_sum$Freq),]
```

There are 36 different descriptors for all 1923816 noise complaints. Top 10 complained noise types are:

```
##                                                Var1    Freq
## 927                          Loud Music/Party 1019651
## 153                          Banging/Pounding  292170
## 928                               Loud Talking  174473
## 1033 Noise: Construction Before/After Hours (NM1)  106863
## 252                            Car/Truck Music   64556
## 1024                    Noise, Barking Dog (NR5)   46242
## 1034       Noise: Construction Equipment (NC1)   37788
## 563                               Engine Idling   28469
## 929                            Loud Television   20416
## 1035              Noise: Jack Hammering (NC2)   20275
```

I grouped the 36 types into 8 types to simplify the plots. These 8 types of noises are *Construction*, *Dog and other animals*, *Vehicle*, *Music/TV/Talking*, *Alarm*, *Air Conditioner*, *Banging/Pounding* and *Other*. The following analysis also uses the population data from 2010 census to normalize the results. Due to the low population in certain zip codes, some of the per-capita values had to removed as out-liers.

```
## load NY state population data by zip code from 2010 census
nyc_pop <- read.csv("aff_download/DEC_10_SF1_P1_with_ann.csv", skip = 1,
                    header = T, stringsAsFactors = F)

nyc_pop$Zip <- as.factor(nyc_pop$Id2)
noise_complaints <- readRDS("noise_complaints.RDS")
library(leaflet)
library(tmap)
nyczipgeo <- readRDS("nyczipgeo.RDS") ## load NYC zip code shape map
noise_sum_zipcode <- as.data.frame(table(noise_complaints$`Incident Zip`))
noise_sum_zipcode$Zip <- noise_sum_zipcode$Var1
noise_sum_zipcode <- left_join(noise_sum_zipcode, nyc_pop)
noise_sum_zipcode <- noise_sum_zipcode %>% select(Zip, Freq, Total) %>%
    mutate( Count = Freq/Total)
noise_sum_zipcode$Count[noise_sum_zipcode$Count == Inf] <- NA
noise_sum_zipcode$Count[which.max(noise_sum_zipcode$Count)] <- NA
nycmap <- append_data(nyczipgeo, noise_sum_zipcode, key.shp = "ZCTA5CE10", key.data = "Zip")
```

4

```r
nyc_map<- tm_shape(nycmap) +
    tm_fill("Count", title = "All Noise", palette = "YlOrRd") +
    tm_borders(alpha = 0.5) +
    tm_style_natural(legend.frame = F, legend.bg.color = NA)


noise_sum_zipcode_type <- noise_complaints %>%
    group_by(`Incident Zip`, Type) %>%
    summarise(Count = n())
noise_sum_zipcode_type <- reshape2::dcast(noise_sum_zipcode_type, `Incident Zip` ~ Type)
noise_sum_zipcode_type <- left_join(noise_sum_zipcode_type, nyc_pop,
                                    by =c("Incident Zip" = "Zip"))
noise_sum_zipcode_type$`Air Conditioner` <- noise_sum_zipcode_type$`Air Conditioner`/ noise_sum_zipcode,
noise_sum_zipcode_type$`Air Conditioner`[noise_sum_zipcode_type$`Air Conditioner`== Inf] <- NA
noise_sum_zipcode_type$`Air Conditioner`[which.max(noise_sum_zipcode_type$`Air Conditioner`)] <- NA
noise_sum_zipcode_type$Alarm <- noise_sum_zipcode_type$Alarm/ noise_sum_zipcode_type$Total
noise_sum_zipcode_type$Alarm[noise_sum_zipcode_type$Alarm== Inf] <- NA
noise_sum_zipcode_type$Vehicle <- noise_sum_zipcode_type$Vehicle/ noise_sum_zipcode_type$Total
noise_sum_zipcode_type$Vehicle[noise_sum_zipcode_type$Vehicle== Inf] <- NA
noise_sum_zipcode_type$Construction <- noise_sum_zipcode_type$Construction/ noise_sum_zipcode_type$Total
noise_sum_zipcode_type$Construction[noise_sum_zipcode_type$Construction== Inf] <- NA
noise_sum_zipcode_type$Construction[which.max(noise_sum_zipcode_type$Construction)] <- NA
noise_sum_zipcode_type$`Banging/Pounding` <- noise_sum_zipcode_type$`Banging/Pounding`/ noise_sum_zipcod
noise_sum_zipcode_type$`Banging/Pounding`[noise_sum_zipcode_type$`Banging/Pounding`== Inf] <- NA
noise_sum_zipcode_type$`Dog and other animals` <- noise_sum_zipcode_type$`Dog and other animals`/ noise_
noise_sum_zipcode_type$`Dog and other animals`[noise_sum_zipcode_type$`Dog and other animals`== Inf] <-
noise_sum_zipcode_type$`Music/TV/Talking` <- noise_sum_zipcode_type$`Music/TV/Talking`/ noise_sum_zipcod
noise_sum_zipcode_type$`Music/TV/Talking`[noise_sum_zipcode_type$`Music/TV/Talking`== Inf] <- NA
noise_sum_zipcode_type$Other <- noise_sum_zipcode_type$Other/ noise_sum_zipcode_type$Total
noise_sum_zipcode_type$Other[noise_sum_zipcode_type$Other== Inf] <- NA


nyc_noise_map <- append_data(nyczipgeo, noise_sum_zipcode_type,
                             key.shp = "ZCTA5CE10", key.data = "Incident Zip")
nyc_noise_map_construction <- tm_shape(nyc_noise_map) +
    tm_fill("Construction", title = "Construction", palette = "YlOrRd") +
    tm_borders(alpha = 0.5) +
    tm_style_natural(legend.frame = F, legend.bg.color = NA)
nyc_noise_map_dogs <- tm_shape(nyc_noise_map) +
    tm_fill("Dog and other animals",
            title = "Dog and other Animals", palette = "YlOrRd") +
    tm_borders(alpha = 0.5) +
    tm_style_natural(legend.frame = F, legend.bg.color = NA)
nyc_noise_map_vehicle <- tm_shape(nyc_noise_map) +
    tm_fill("Vehicle", title = "Vehicle", palette = "YlOrRd") +
    tm_borders(alpha = 0.5) +
    tm_style_natural(legend.frame = F, legend.bg.color = NA)
nyc_noise_map_music <- tm_shape(nyc_noise_map) +
    tm_fill("Music/TV/Talking", title = "Music/TV/Talking",
            palette = "YlOrRd") +
    tm_borders(alpha = 0.5) +
    tm_style_natural(legend.frame = F, legend.bg.color = NA)
nyc_noise_map_alarm <- tm_shape(nyc_noise_map) +
    tm_fill("Alarm", title = "Alarm", palette = "YlOrRd") +
    tm_borders(alpha = 0.5) +
```

```
    tm_style_natural(legend.frame = F, legend.bg.color = NA)
nyc_noise_map_ac <- tm_shape(nyc_noise_map) +
    tm_fill("Air Conditioner", title = "Air Conditioner",
            palette = "YlOrRd") + tm_borders(alpha = 0.5) +
    tm_style_natural(legend.frame = F, legend.bg.color = NA)
nyc_noise_map_other <- tm_shape(nyc_noise_map) +
    tm_fill("Other", title = "Other", palette = "YlOrRd") +
    tm_borders(alpha = 0.5) +
    tm_style_natural(legend.frame = F, legend.bg.color = NA)

nyc_noise_map_banging <- tm_shape(nyc_noise_map) +
    tm_fill("Banging/Pounding",
            title = "Banging/Pounding", palette = "YlOrRd") +
    tm_borders(alpha = 0.5) +
    tm_style_natural(legend.frame = F, legend.bg.color = NA)
nyc_map
```
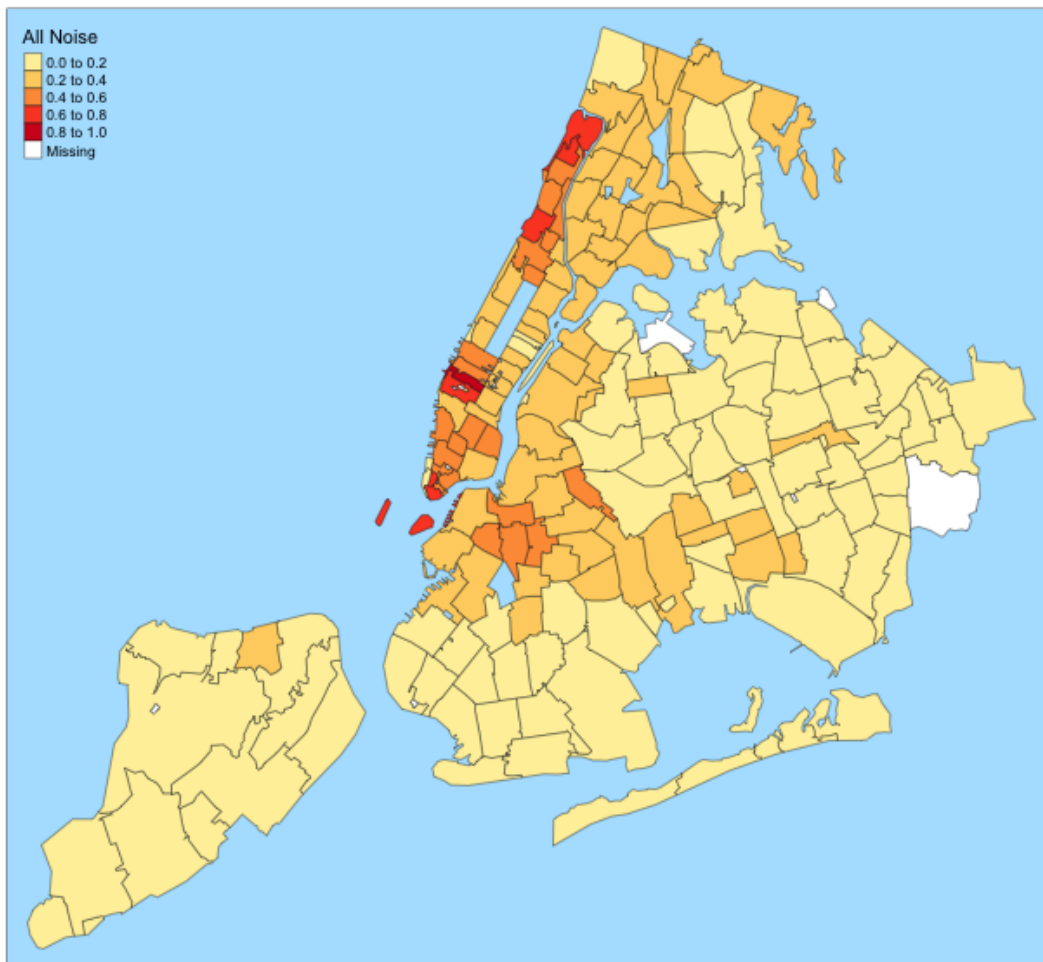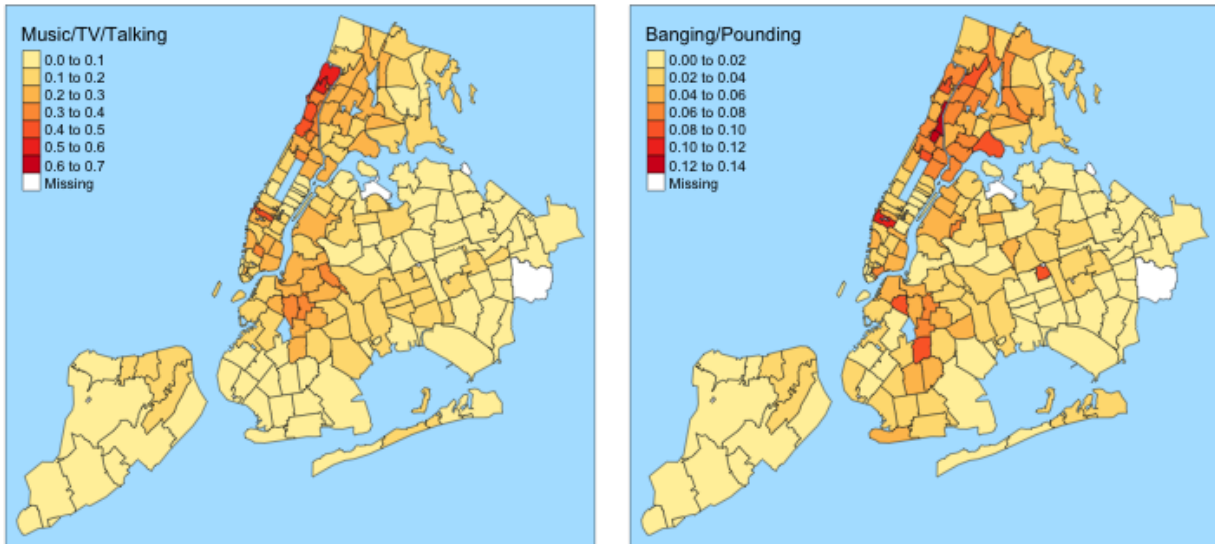

The neighborhoods complaining about noises most are northern Manhattan, mid Manhattan and lower Manhattan, an area in Brooklyn(zip codes 11205, 11216, 11217 and 11238) and several neighborhoods in Bronx.

```
multiplot(nyc_noise_map_music, nyc_noise_map_banging, cols=2)
```


"Loud Music/Party" is the most complained type of noise in NYC, accounting for over 50% of all noise complaints. It's not surprising that its spatial distribution is very similar to that of over all noise. "Banging/Pounding" is the second most complained type. It is most severe in the neighborhood of mid Manhattan(Zip code 10001) and Upper Manhattan (zip codes 10030 an 10039).

```
multiplot(nyc_noise_map_construction, nyc_noise_map_vehicle, cols=2)
```
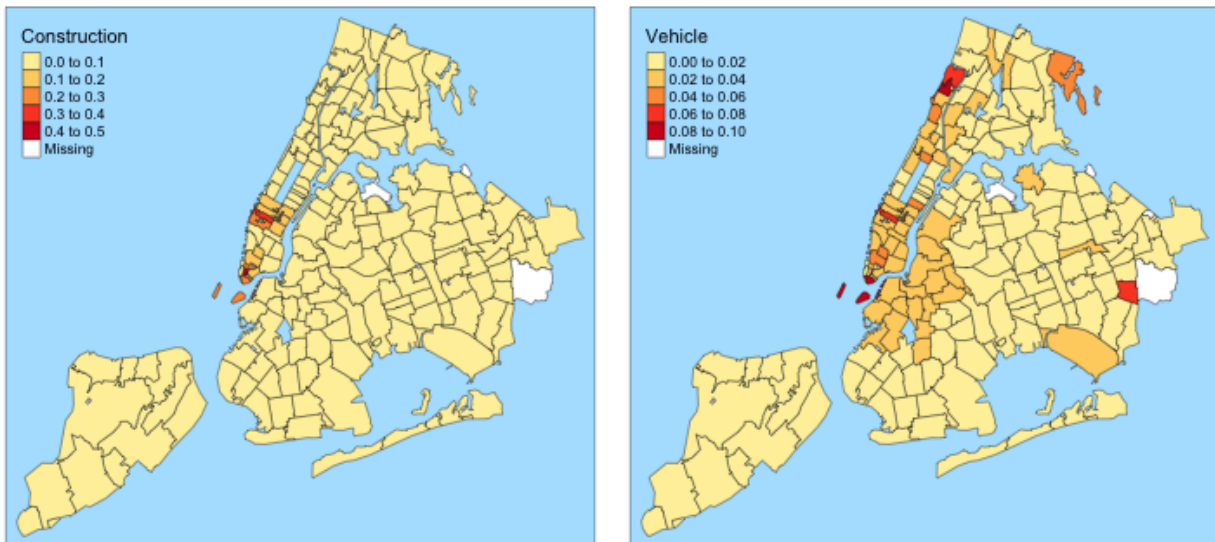

Complaints about construction noises are most common in mid and low Manhattan, while vehcile noises are most complained around mid, low and northern Manhattan.

```
multiplot(nyc_noise_map_ac, nyc_noise_map_other, cols=2)
```
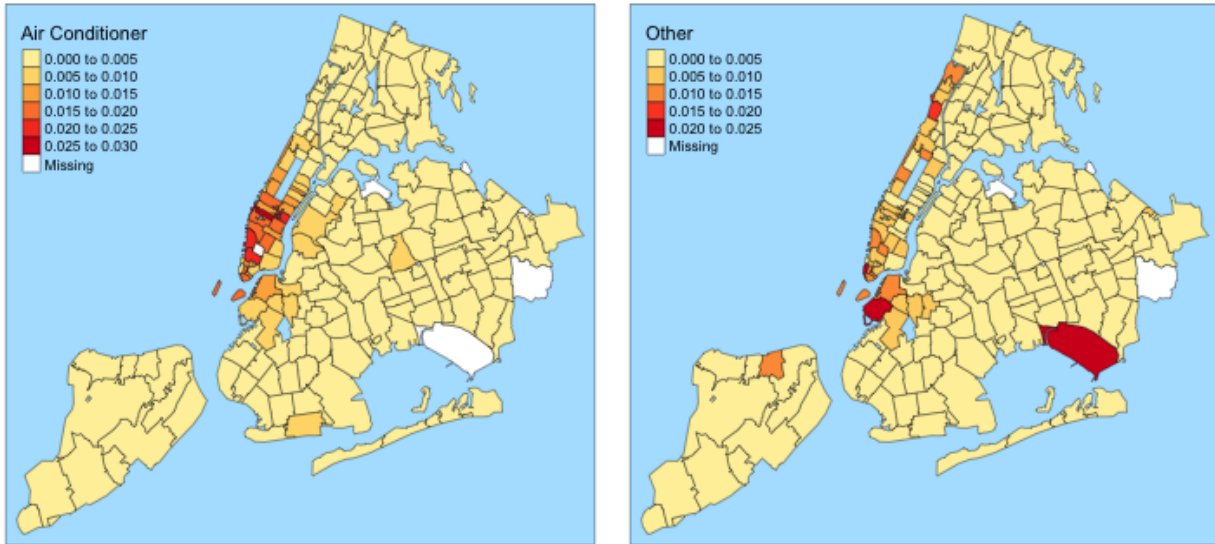
Air condition and ventilation noises are most complained in mid and low Manhattan. The category "Other" is mostly complained near JFK air port, Red Hook neighborhood of Brooklyn and Washington Heights of Manhattan.

```
multiplot(nyc_noise_map_dogs, nyc_noise_map_alarm, cols=2)
```
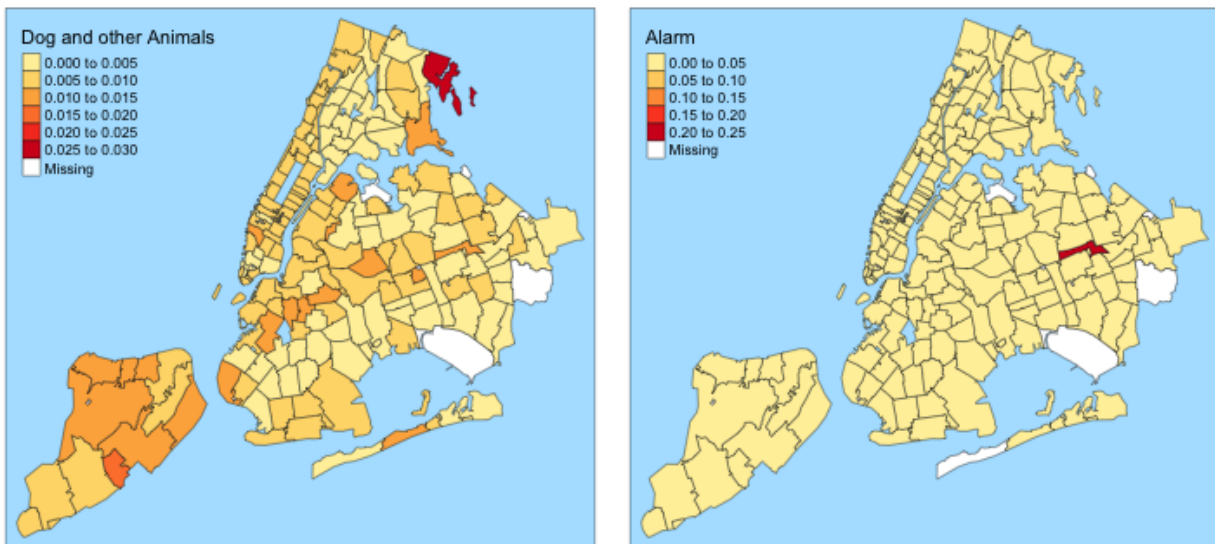

Complaints about barking dogs and other animals are relatively rare in NYC and it spreads across all boroughs. It is mostly common in the neighborhood(zip code 10464) near Westchester County. Complaints about "Alarm" are very peculiar as the neighborhood around Kew Gardens (zip code 11366) contributed near 20% of the complaints, alomst 10x more than any other zip codes in NYC.

3. How about other common complaints such as "Heating" and "Street Condition"?

```
#other_complaints <- all_complaints%>%select(2, 6, 7, 9, 25, 51, 52)
#other_complaints$`Created Date` <- as.POSIXct(strptime(data_311$`Created Date`, "%m/%d/%Y %I:%M:%S %p")
```

```
#other_complaints$Month <- format(other_complaints$`Created Date`, "%b")
#other_complaints$`Incident Zip` <- substring(other_complaints$`Incident Zip`, 1, 5)
#saveRDS(other_complaints, "other_complaints.RDS")
#other_complaints <- readRDS("other_complaints.RDS")
#Heating <- filter(other_complaints, grepl("HEATING", other_complaints$`Complaint Type`))
#Heating$Month <- as.factor(Heating$Month)
#Heating <- select(Heating, -1)
#saveRDS(Heating, "Heating.RDS")
Heating <- readRDS("Heating.RDS")
Heating_plot <- ggplot(data = Heating) +
    theme_few() + scale_fill_few() + ylab("Number of Complaints") +
    geom_bar(aes(x= Month, fill = Borough), stat = "count") + theme(legend.position = "top") +
    scale_x_discrete(limits = c ("Jan", "Feb", "Mar", "Apr", "May","Jun",
                                 "Jul", "Aug", "Sep", "Oct", "Nov", "Dec")) +
    ggtitle("Heating")

#StreetCondition <- filter(other_complaints, grepl("Street Condition", other_complaints$`Complaint Type
#StreetCondition$Month <- as.factor(StreetCondition$Month)
#StreetCondition <- select(StreetCondition, -1)
#saveRDS(StreetCondition, "StreetCondition.RDS")
StreetCondition <- readRDS("StreetCondition.RDS")
StreetCondition_plot <- ggplot(data = StreetCondition) +
    theme_few() + scale_fill_few() + ylab("Number of Complaints") +
    geom_bar(aes(x= Month, fill = Borough), stat = "count") +
    scale_x_discrete(limits = c ("Jan", "Feb", "Mar", "Apr", "May","Jun",
                                 "Jul", "Aug", "Sep", "Oct", "Nov", "Dec")) +
    theme(legend.position = "none") +
    ggtitle("Street Condition")

multiplot(Heating_plot, StreetCondition_plot, cols = 2)
```
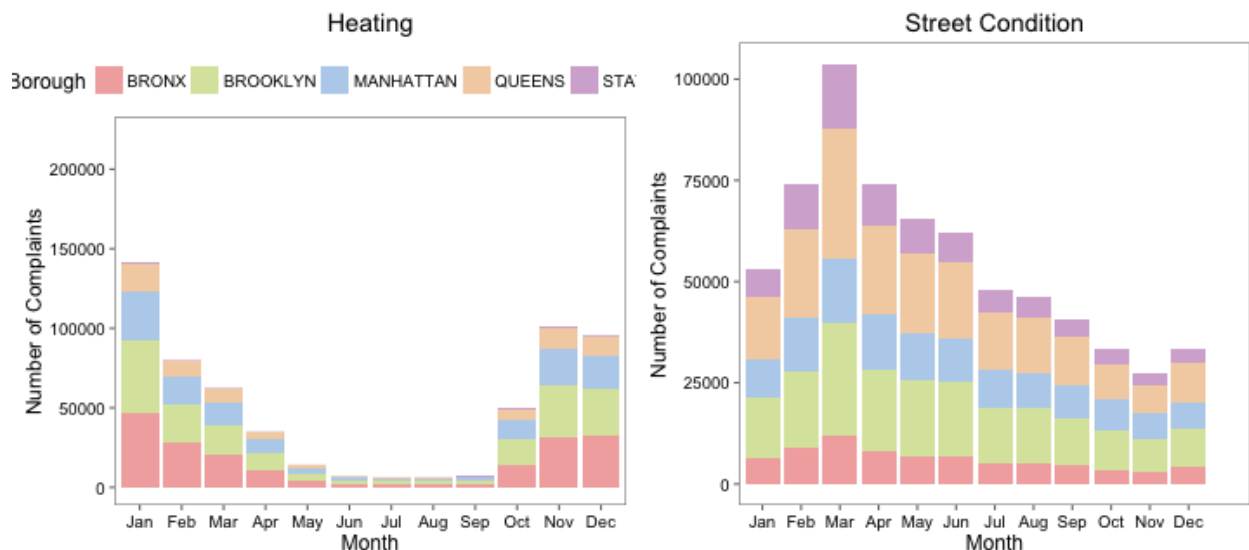

Complaints about heating is most often in cold weather form October to April. This is as expected. The situation is similar with complaints about street conditions, which is most complained from Februry to Jun, after all the long and cold NY winter time.

```r
nyc_list <- readRDS("nyc_list.RDS") ## load a list of NYC zip codes
Heating_sum <- Heating %>% filter(`Incident Zip` %in% nyc_list) %>%
    group_by(`Incident Zip`) %>% summarise(Count = n()) %>% mutate (Zip = `Incident Zip`)

Heating_sum <- left_join(Heating_sum, nyc_pop)
Heating_sum <- Heating_sum %>% select(Zip, Count, Total) %>% mutate(CountperCapita = Count/Total)
heating_map <- append_data(nyczipgeo, Heating_sum,
                           key.shp = "ZCTA5CE10", key.data = "Zip")
heating_map_plot <- tm_shape(heating_map) +
    tm_fill("CountperCapita", title = "Heating Complaints", palette = "YlOrRd") +
    tm_borders(alpha = 0.5) +
    tm_style_natural(legend.frame = F, legend.bg.color = NA)
StreetCondition_sum <- StreetCondition %>%
    filter(`Incident Zip` %in% nyc_list) %>%
    group_by(`Incident Zip`) %>% summarise(Count = n()) %>% mutate (Zip = `Incident Zip`)
StreetCondition_sum <- left_join(StreetCondition_sum, nyc_pop)
StreetCondition_sum <- StreetCondition_sum %>% select(Zip, Count, Total) %>% mutate(CountperCapita = Cou
StreetCondition_map <- append_data(nyczipgeo, StreetCondition_sum,
                                   key.shp = "ZCTA5CE10", key.data = "Zip")
StreetCondition_map_plot <- tm_shape(StreetCondition_map) +
    tm_fill("CountperCapita", title = "Street Condition Complaints",
            palette = "YlOrRd") + tm_borders(alpha = 0.5) +
    tm_style_natural(legend.frame = F, legend.bg.color = NA)

multiplot(heating_map_plot, StreetCondition_map_plot, cols = 2)
```
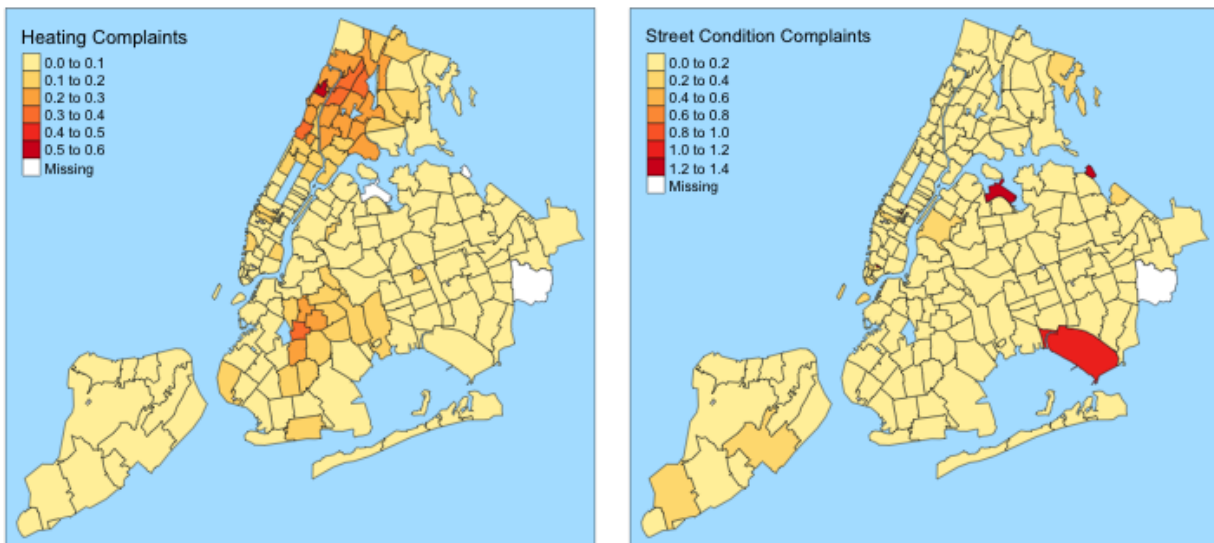

It appears some neighborhoods in Upper Manhattan suffers most heating problems, some neighborhoods in Brooklyn and Bronx also have some heating issues. As for road conditions, both JFK and LaGuardia air port appear to have reported most street condition issues.

Summary:

After exploratory analysis of NYC 311 dataset, we conclude that New Yorkers' most complained issue is noise, most of which are from Manhattan and some of its adjacent neighborhoods in Bronx and Brooklyn. Heating and street condition are also often complained, these two complaints have their own correlation to

the cold weather. Their time and spatial pattern could be useful for the city to distribute the resources to help resolving these issues.

References:

1. https://en.wikipedia.org/wiki/3-1-1
2. What a Hundred Million Calls to 311 Reveal About New York - Wired
3. Mapping New York's Noisiest Neighborhoods - The New Yorker