homework ii

Nirbhay Pherwani, Dhiren Chandnani

2019-12-09

Introduction

The aim of this exploratory data analysis is to understand the NYC 311 data and observe any noticeable trends and make good predictions Each day, NYC311 receives thousands of requests related to several hundred types of non-emergency services, including noise complaints, plumbing issues, and illegally parked cars. These requests are received by NYC311 and forwarded to the relevant agencies, such as the Police, Buildings or Transportation. The agency responds to the request, addresses it and the request is then closed.

Initialization

Here we load the tidyverse packages and the `data.table` package and load the nyc311 data set. Then we fix the column names of the nyc311 data so that they have no spaces.

```
library(tidyverse)
```

```
## -- Attaching packages ----------------------- tidyverse 1.2.1 --

## v ggplot2 3.2.1     v purrr   0.3.2
## v tibble  2.1.3     v dplyr   0.8.3
## v tidyr   0.8.3     v stringr 1.4.0
## v readr   1.3.1     v forcats 0.4.0

## -- Conflicts -------------------------- tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
```

```
library(data.table)
```

```
##
## Attaching package: 'data.table'

## The following objects are masked from 'package:dplyr':
##
##     between, first, last

## The following object is masked from 'package:purrr':
##
##     transpose
```

```
nyc311<-fread("311_Service_Requests_from_2010_to_Present.csv")
names(nyc311)<-names(nyc311) %>%
  stringr::str_replace_all("\\s", ".")
```

Removing columns

Getting rid of non-required columns

```
nyc311 = select(nyc311, -10:-15,-17:-19,-23,-27:-49)
```

Description

Here we describe the data, showing both a sample and a data dictionary.

The head of the table

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

```
library(xtable)
options(xtable.comment=FALSE)
options(xtable.booktabs=TRUE)
narrow<-nyc311 %>%
  select(Agency,
Complaint.Type,
Descriptor,
Incident.Zip,
Status,
Borough)
xtable(head(narrow))
```

	Agency	Complaint.Type	Descriptor	Incident.Zip	Status	Borough
1	NYPD	Vending	In Prohibited Area	10465	Closed	BRONX
2	NYPD	Blocked Driveway	No Access	11234	Open	BROOKLYN
3	NYPD	Noise - Street/Sidewalk	Loud Music/Party	11204	Open	BROOKLYN
4	NYPD	Noise - Street/Sidewalk	Loud Talking	11211	Assigned	BROOKLYN
5	NYPD	Noise - Street/Sidewalk	Loud Talking	10025	Closed	MANHATTAN
6	NYPD	Noise - Street/Sidewalk	Loud Talking	11205	Closed	BROOKLYN

Data Dictionary

Unique Key - A key to uniquely identify each service request (SR)

Created Date - Date of creation of service request (Format: MM/DD/YY HH:MM:SS AM/PM)

Closed Date - Date on which the SR was closed ((Format: MM/DD/YY HH:MM:SS AM/PM))

Agency - The responding City Government Agency acronym

Agency Name - Full name of the responding City Government Agency

Complaint Type - Information about the topic of the incident or condition.

Descriptor - Dependent on the complaint type. Contains more information on the incident or condition.

Status - Status of SR submitted (Suggested values: Assigned, Cancelled, Closed, etc.)

Due Date - Date when SR is supposed to be updated (Format: MM/DD/YY HH:MM:SS AM/PM)

Resolution Action Updated Date - Date when responding agency last updated the SR. (Format: MM/DD/YY HH:MM:SS AM/PM)

Location Type - Describes the type of location used in the address information

Incident Zip - Incident location zip code, provided by geo validation.

Incident Address - House number of incident address provided by submitter.

Street Name - Street name of incident address provided by the submitter

Cross Street 1 - First Cross street based on the geo validated incident location

Cross Street 2 - Second Cross Street based on the geo validated incident location

Intersection Street 1 - First intersecting street based on geo validated incident location

Intersection Street 2 - Second intersecting street based on geo validated incident location

Address Type - Type of incident location information available (Values: Address, Block face, Intersection, LatLong, Placename).

City - City of the incident location provided by geo validation.

Landmark - If the incident location is identified as a Landmark the name of the landmark will display here

Facility Type - If available, this field describes the type of city facility associated to the SR

Community Board - Provided by geovalidation.

Borough - Provided by the submitter and confirmed by geo validation.

X Coordinate (State Plane) - Geo validated, X coordinate of the incident location.

Y Coordinate (State Plane) - Geo validated, Y coordinate of the incident location.

Latitude - Geo based Lat of the incident location

Longitude - Geo based Long of the incident location

Location - Combination of the geo based lat & long of the incident location

Park Facility Name - If the incident location is a Parks Dept facility, the Name of the facility will appear here

Park Borough - The borough of incident if it is a Parks Dept facility

Vehicle Type - If the incident is a taxi, this field describes the type of TLC vehicle.

Taxi Company Borough - If the incident is identified as a taxi, this field will display the borough of the taxi company.

Taxi Pick Up Location - If the incident is identified as a taxi, this field displays the taxi pick up location

Bridge Highway Name - If the incident is identified as a Bridge/Highway, the name will be displayed here.

Bridge Highway Direction - If the incident is identified as a Bridge/Highway, the direction where the issue took place would be displayed here.

Road Ramp - If the incident location was Bridge/Highway this column differentiates if the issue was on the Road or the Ramp.

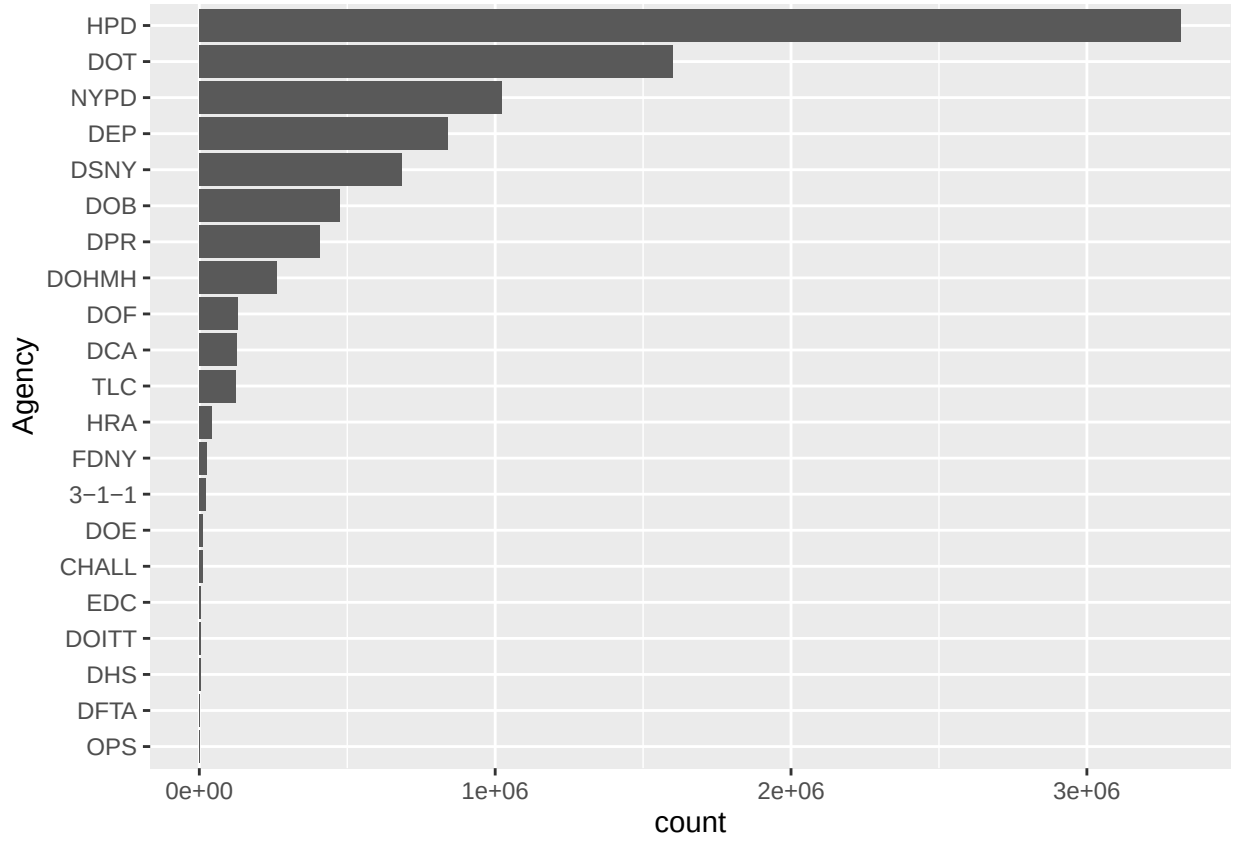
Bridge Highway Segment - Additional information on the section of the Bridge/Highway were the incident took place.

Exploration

Here we explore the columns in the data set.

The following plot shows the distribution of the service requests in various agencies. It can be seen that HPD (Department of Housing Preservation and Development) receives the most service calls followed by the DOT (Department of Transportation).

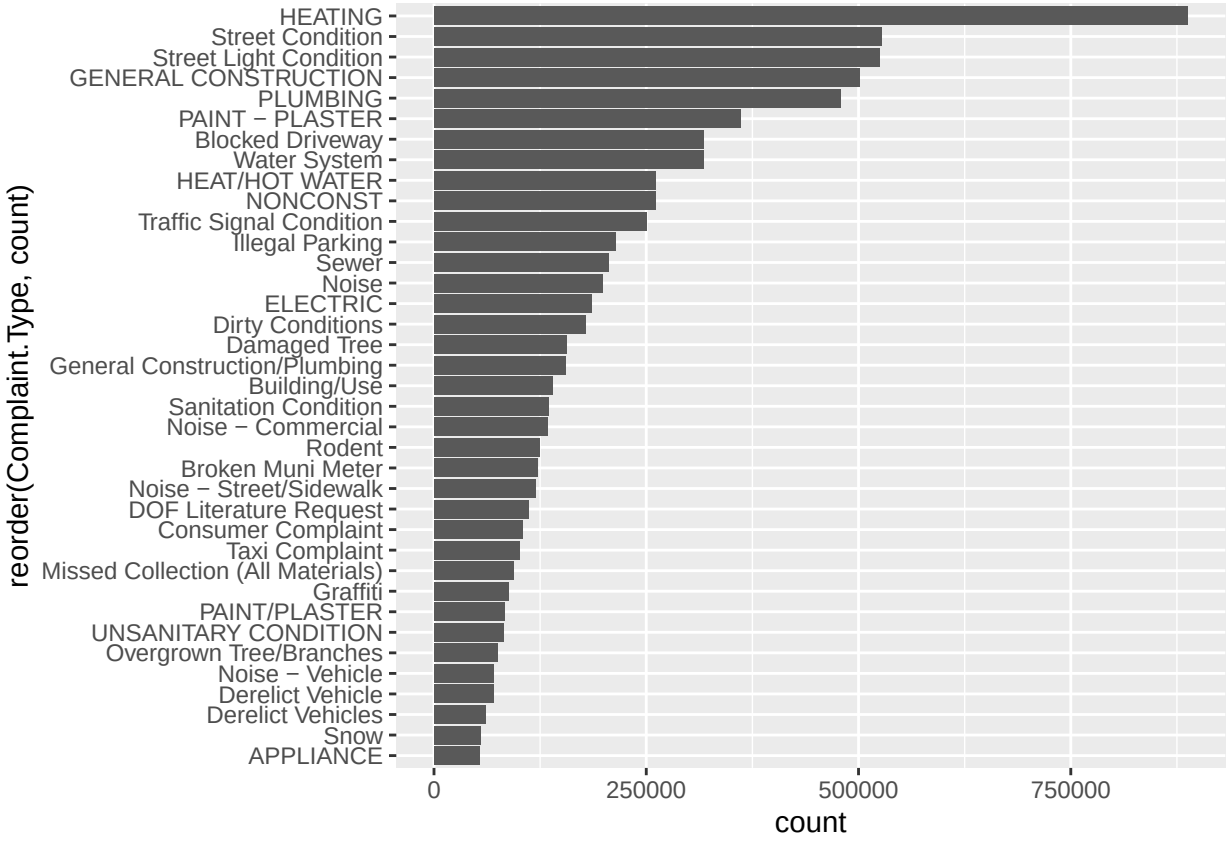
```
bigAgency <- narrow %>%
  group_by(Agency) %>%
  summarize(count=n()) %>%
  filter(count>1000)
bigAgency$Agency<-factor(bigAgency$Agency,
  levels=bigAgency$Agency[order(bigAgency$count)])
p<-ggplot(bigAgency,aes(x=Agency,y=count)) +
   geom_bar(stat="identity") +
   coord_flip()
p
```



Most common complaints

The below graph gives us a crucial information about the common types of complaint. It shows that the highest count of complaints recieved is regarding the heating problem.

```
complaints <- narrow %>%
  group_by(Complaint.Type) %>%
  summarize(count=n()) %>%
  filter(count>50000)
complaints$Complaint.Type<-factor(complaints$Complaint.Type,
  levels=complaints$Complaint.Type[order(complaints$Complaint.Type)])
p<-ggplot(complaints,aes(x=reorder(Complaint.Type, count),y=count)) +
   geom_bar(stat="identity") +
   coord_flip()
p
```

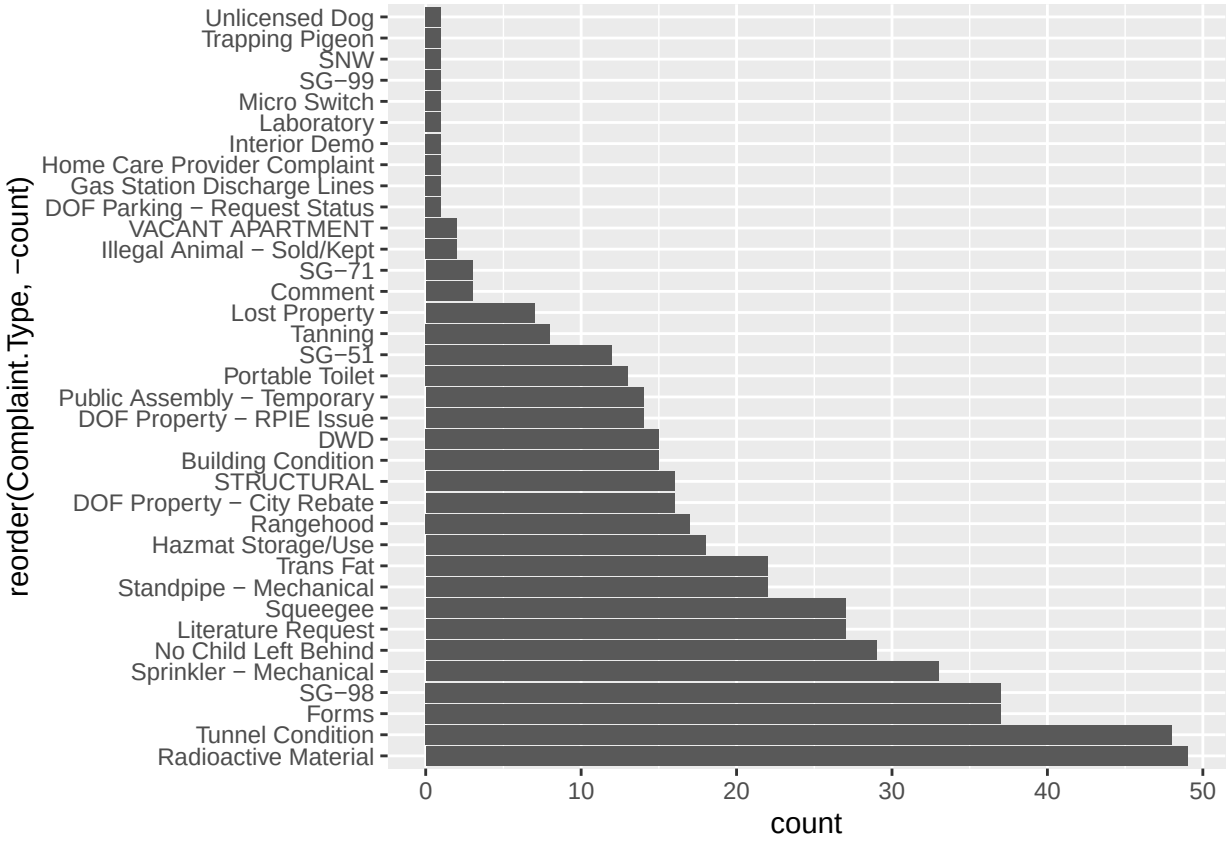


Least common complaints

This plot shows us the least common complaint types. The most humorous of them all is Trapping pigeon with just 1 complaint.

```
complaints <- narrow %>%
  group_by(Complaint.Type) %>%
  summarize(count=n()) %>%
  filter(count<50)
complaints$Complaint.Type<-factor(complaints$Complaint.Type,
  levels=complaints$Complaint.Type[order(complaints$Complaint.Type)])
p<-ggplot(complaints,aes(x=reorder(Complaint.Type, -count),y=count)) +
   geom_bar(stat="identity") +
```

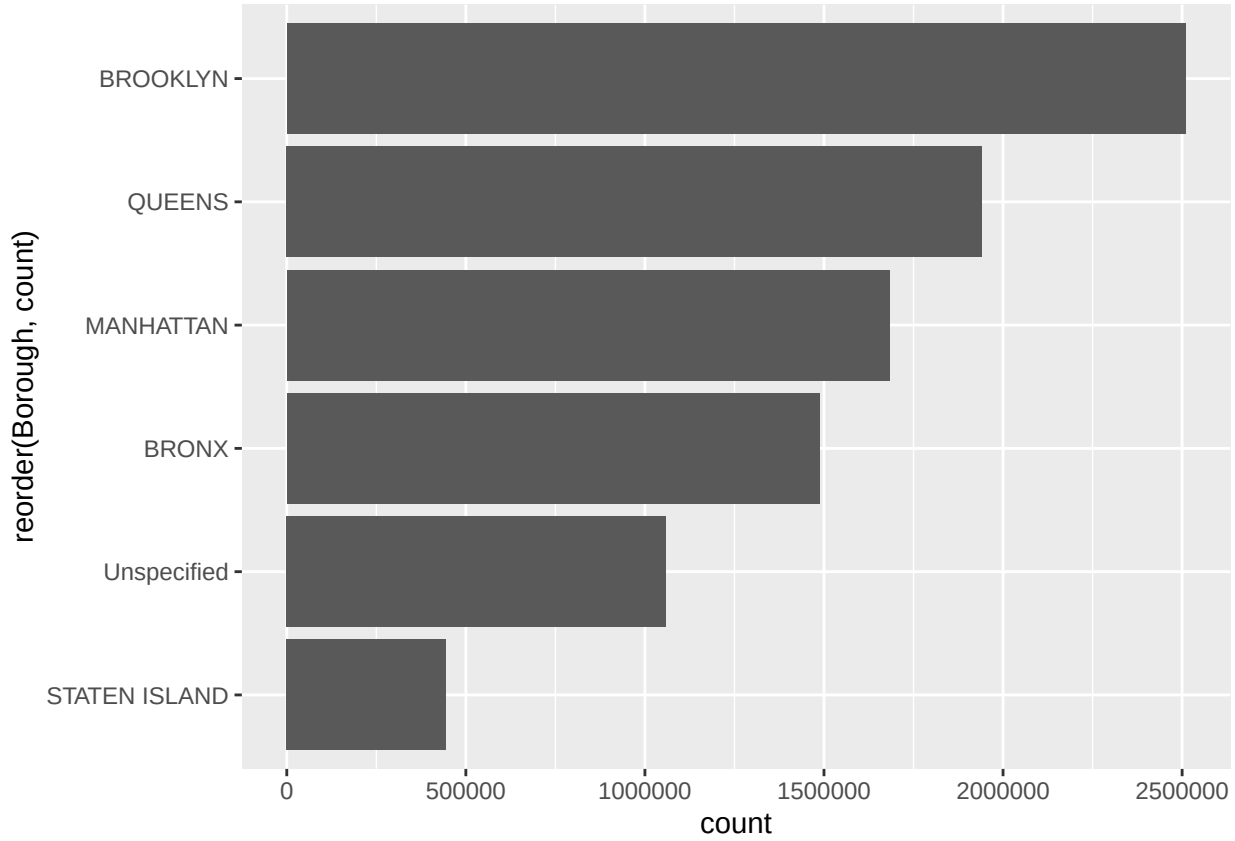
```
   coord_flip()
p
```



Complaints by Borough

The below graph talks about the Borough with most complaints. Brooklyn has the highest number of complaints followed by Queens.

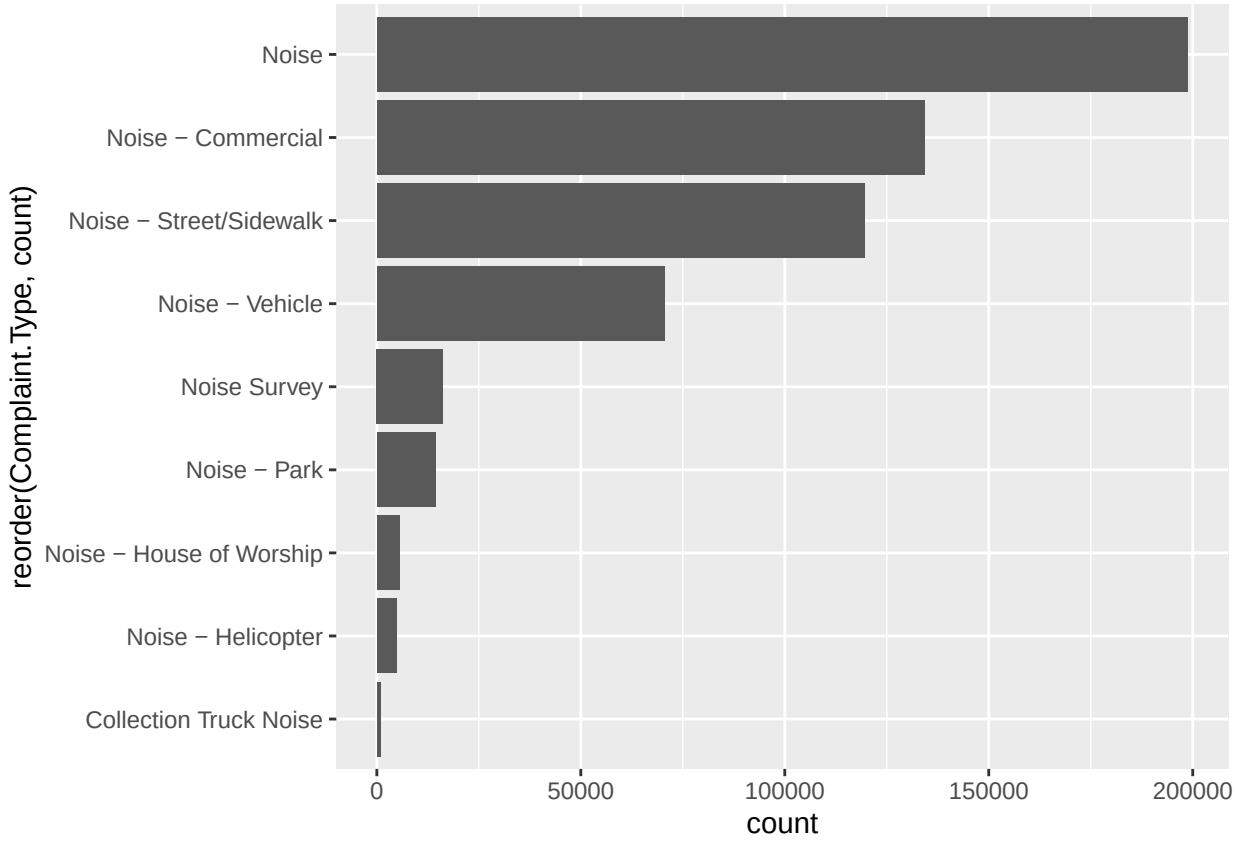
```
complaints <- nyc311 %>%
  group_by(Borough) %>%
  summarize(count=n())
p<-ggplot(complaints,aes(x=reorder(Borough, count),y=count)) +
   geom_bar(stat="identity") +
   coord_flip()
p
```



Noise complaints

The top three complaint types across all boroughs, Commercial, Street/Sidewalk, and Vehicle are also the top three complaint types within each borough.

```
complaints <- nyc311 %>%
  group_by(Complaint.Type) %>%
  summarize(count=n()) %>%
  filter(str_detect(Complaint.Type, "Noise"))
p<-ggplot(complaints,aes(x=reorder(Complaint.Type, count),y=count)) +
   geom_bar(stat="identity") +
  coord_flip()
p
```



Trend of 311 calls

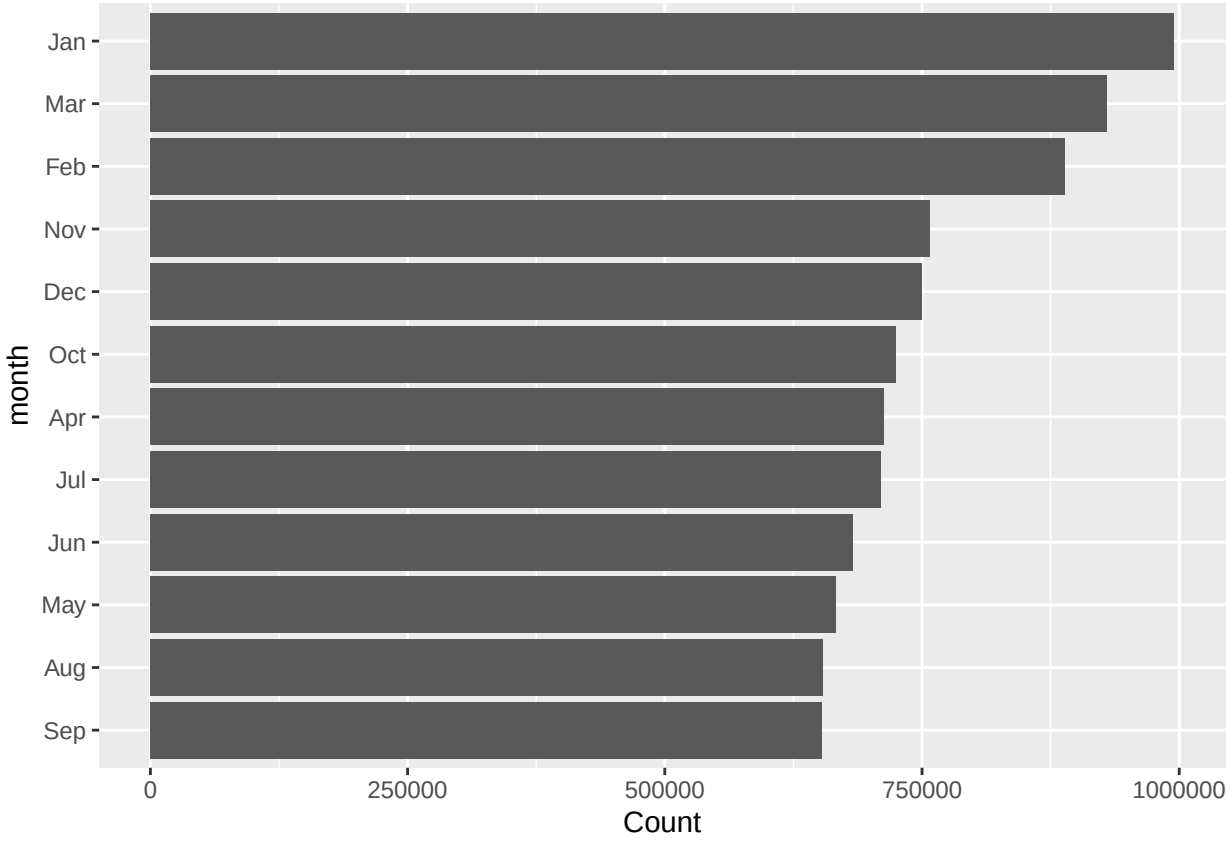
This is one of the most important graphs of this analysis. It shows the trend of 311 calls across various months. As you can see, January and March are the top 2 busiest months of service request.

```
library(lubridate)
```

```
##
## Attaching package: 'lubridate'

## The following objects are masked from 'package:data.table':
##
##     hour, isoweek, mday, minute, month, quarter, second, wday,
##     week, yday, year

## The following object is masked from 'package:base':
##
##     date
```

```
nyc311 %>%
  mutate(month = month.abb[month(mdy_hms(Created.Date))]) %>%
  filter(!is.na(month)) %>%
  group_by(month) %>%
```

```
  summarise(Count = n()) %>%
  arrange(desc(Count)) %>%
  ungroup() %>%
  mutate(month = reorder(month,Count)) %>%

    ggplot(aes(x = month,y = Count)) +
    geom_bar(stat='identity') +
    coord_flip()
```



Next we include a crosstabulation.

Crosstab between Borough and Complaint Type

```
xtabA<-dplyr::filter(narrow,
  Complaint.Type=='HEATING' |
  Complaint.Type=='GENERAL CONSTRUCTION' |
  Complaint.Type=='PLUMBING'
)
xtabB<-select(xtabA,Borough,"Complaint.Type")
library(gmodels)
CrossTable(xtabB$Borough,xtabB$'Complaint.Type')
```

```
##
##
##    Cell Contents
## |-------------------------|
```

```
## |                       N |
## | Chi-square contribution |
## |           N / Row Total |
## |           N / Col Total |
## |         N / Table Total |
## |-------------------------|
##
##
## Total Observations in Table:  1868064
##
##
##                | xtabB$Complaint.Type
## xtabB$Borough  | GENERAL CONSTRUCTION |              HEATING |             PLUMBING |            Row Total |
## ---------------|----------------------|----------------------|----------------------|----------------------|
##         BRONX  |               107626 |               195246 |               103964 |               406836 |
##                |               23.326 |               19.145 |                1.030 |                      |
##                |                0.265 |                0.480 |                0.256 |                0.218 |
##                |                0.215 |                0.220 |                0.217 |                      |
##                |                0.058 |                0.105 |                0.056 |                      |
## ---------------|----------------------|----------------------|----------------------|----------------------|
##      BROOKLYN  |               132552 |               190268 |               128383 |               451203 |
##                |             1076.405 |             2717.190 |             1398.387 |                      |
##                |                0.294 |                0.422 |                0.285 |                0.242 |
##                |                0.264 |                0.214 |                0.268 |                      |
##                |                0.071 |                0.102 |                0.069 |                      |
## ---------------|----------------------|----------------------|----------------------|----------------------|
##     MANHATTAN  |                61453 |               137458 |                63103 |               262014 |
##                |             1123.330 |             1347.582 |              245.877 |                      |
##                |                0.235 |                0.525 |                0.241 |                0.140 |
##                |                0.123 |                0.155 |                0.132 |                      |
##                |                0.033 |                0.074 |                0.034 |                      |
## ---------------|----------------------|----------------------|----------------------|----------------------|
##        QUEENS  |                41277 |                75776 |                43604 |               160657 |
##                |               79.707 |                4.192 |              142.183 |                      |
##                |                0.257 |                0.472 |                0.271 |                0.086 |
##                |                0.082 |                0.085 |                0.091 |                      |
##                |                0.022 |                0.041 |                0.023 |                      |
## ---------------|----------------------|----------------------|----------------------|----------------------|
## STATEN ISLAND  |                 8329 |                 6011 |                 7525 |                21865 |
##                |             1030.062 |             1845.525 |              657.654 |                      |
##                |                0.381 |                0.275 |                0.344 |                0.012 |
##                |                0.017 |                0.007 |                0.016 |                      |
##                |                0.004 |                0.003 |                0.004 |                      |
## ---------------|----------------------|----------------------|----------------------|----------------------|
##   Unspecified  |               150277 |               282916 |               132296 |               565489 |
##                |               15.587 |              750.862 |             1106.709 |                      |
##                |                0.266 |                0.500 |                0.234 |                0.303 |
##                |                0.300 |                0.319 |                0.276 |                      |
##                |                0.080 |                0.151 |                0.071 |                      |
## ---------------|----------------------|----------------------|----------------------|----------------------|
##  Column Total  |               501514 |               887675 |               478875 |              1868064 |
##                |                0.268 |                0.475 |                0.256 |                      |
## ---------------|----------------------|----------------------|----------------------|----------------------|
##
##
```

Of all the service requests originating from Brooklyn, the most common complaint type was of Heating. 282916 of all SRs with complaint type 'Heating' are Unspecified.

Crosstab between Borough and Status

```
xtabA<-dplyr::filter(narrow,
  Status=='Closed' |
  Status=='Open' |
```

```
  Status=='Assigned' |
  Status=='Pending'
)
xtabB<-select(xtabA,Borough,"Status")
library(gmodels)
CrossTable(xtabB$Borough,xtabB$'Status')
```

```
##
##
##    Cell Contents
## |-------------------------|
## |                       N |
## | Chi-square contribution |
## |           N / Row Total |
## |           N / Col Total |
## |         N / Table Total |
## |-------------------------|
##
##
## Total Observations in Table:  9092480
##
##
##                | xtabB$Status
## xtabB$Borough |  Assigned |    Closed |      Open |   Pending | Row Total |
## --------------|-----------|-----------|-----------|-----------|-----------|
##         BRONX |     23157 |   1190472 |    203951 |     68633 |   1486213 |
##               |     9.926 |  5310.591 | 22196.472 | 13804.840 |           |
##               |     0.016 |     0.801 |     0.137 |     0.046 |     0.163 |
##               |     0.167 |     0.153 |     0.227 |     0.255 |           |
##               |     0.003 |     0.131 |     0.022 |     0.008 |           |
## --------------|-----------|-----------|-----------|-----------|-----------|
##      BROOKLYN |     39975 |   2119906 |    274308 |     71078 |   2505267 |
##               |    79.159 |   301.204 |  2891.649 |   127.496 |           |
##               |     0.016 |     0.846 |     0.109 |     0.028 |     0.276 |
##               |     0.288 |     0.272 |     0.305 |     0.264 |           |
##               |     0.004 |     0.233 |     0.030 |     0.008 |           |
## --------------|-----------|-----------|-----------|-----------|-----------|
##     MANHATTAN |     25851 |   1429165 |    182086 |     43515 |   1680617 |
##               |     1.583 |    69.368 |  1545.257 |   780.039 |           |
##               |     0.015 |     0.850 |     0.108 |     0.026 |     0.185 |
##               |     0.186 |     0.184 |     0.203 |     0.162 |           |
##               |     0.003 |     0.157 |     0.020 |     0.005 |           |
## --------------|-----------|-----------|-----------|-----------|-----------|
##        QUEENS |     41130 |   1673100 |    161352 |     62725 |   1938307 |
##               |  4507.695 |   106.198 |  4754.895 |   499.559 |           |
##               |     0.021 |     0.863 |     0.083 |     0.032 |     0.213 |
##               |     0.296 |     0.215 |     0.180 |     0.233 |           |
##               |     0.005 |     0.184 |     0.018 |     0.007 |           |
## --------------|-----------|-----------|-----------|-----------|-----------|
## STATEN ISLAND |      8196 |    367007 |     47382 |     21006 |    443591 |
##               |   300.336 |   434.806 |   287.426 |  4724.810 |           |
##               |     0.018 |     0.827 |     0.107 |     0.047 |     0.049 |
##               |     0.059 |     0.047 |     0.053 |     0.078 |           |
##               |     0.001 |     0.040 |     0.005 |     0.002 |           |
## --------------|-----------|-----------|-----------|-----------|-----------|
##   Unspecified |       460 |   1006480 |     29376 |      2169 |   1038485 |
##               | 14942.660 | 15445.449 | 52273.311 | 26552.907 |           |
##               |     0.000 |     0.969 |     0.028 |     0.002 |     0.114 |
##               |     0.003 |     0.129 |     0.033 |     0.008 |           |
##               |     0.000 |     0.111 |     0.003 |     0.000 |           |
## --------------|-----------|-----------|-----------|-----------|-----------|
##  Column Total |    138769 |   7786130 |    898455 |    269126 |   9092480 |
##               |     0.015 |     0.856 |     0.099 |     0.030 |           |
## --------------|-----------|-----------|-----------|-----------|-----------|
##
##
```

23157 of all SRs with a status 'Assigned' are from the Bronx borough. 71078 of all SRs with a status 'Pending' are from the Brooklyn borough. Of all the 1038485 Unspecified location service requests, 1006480 have been closed.

Crosstab between Noise complaints and Status

```
xtabA<-dplyr::filter(narrow,
  Complaint.Type=='Noise - Commercial' |
  Complaint.Type=='Noise - Vehicle'
)
xtabB<-select(xtabA,Complaint.Type,"Status")
library(gmodels)
CrossTable(xtabB$Complaint.Type,xtabB$'Status')
```

```
##
##
##    Cell Contents
## |-------------------------|
## |                       N |
## | Chi-square contribution |
## |           N / Row Total |
## |           N / Col Total |
## |         N / Table Total |
## |-------------------------|
##
##
## Total Observations in Table:  204662
##
##
##                      | xtabB$Status
## xtabB$Complaint.Type |  Assigned |    Closed |      Open | Row Total |
## ---------------------|-----------|-----------|-----------|-----------|
##   Noise - Commercial |        10 |    134160 |        52 |    134222 |
##                      |     1.806 |     0.001 |     2.762 |           |
##                      |     0.000 |     1.000 |     0.000 |     0.656 |
##                      |     1.000 |     0.656 |     0.825 |           |
##                      |     0.000 |     0.656 |     0.000 |           |
## ---------------------|-----------|-----------|-----------|-----------|
##      Noise - Vehicle |         0 |     70429 |        11 |     70440 |
##                      |     3.442 |     0.003 |     5.264 |           |
##                      |     0.000 |     1.000 |     0.000 |     0.344 |
##                      |     0.000 |     0.344 |     0.175 |           |
##                      |     0.000 |     0.344 |     0.000 |           |
## ---------------------|-----------|-----------|-----------|-----------|
##         Column Total |        10 |    204589 |        63 |    204662 |
##                      |     0.000 |     1.000 |     0.000 |           |
## ---------------------|-----------|-----------|-----------|-----------|
##
##
```

This crosstab is about the status of noise complaints. In previous plots, we saw that commercial and vehicular noise had the most complaints. This crosstab helps us look at the status of these complaints. Most of them look closed. Only 52 noise commercial complaints are open.

Conclusion

After looking at the various plots, we can say that the Borough with the most complaints in Brooklyn and the most common type of complaint was regarding Heating. We also looked at the various noise complaints and observed that commercial noise had the most service request. Looking at the trend, January month had the most Service Request calls so the agencies need to be more active during that month.