homework iii

Nirbhay Pherwani, Dhiren Chandnani

2019-09-24

Introduction

We have worked on four scenarios to get some insights on the different fields and how they are related. Fields we worked on include Complaint.Type, Location.Type, Status, Borough, Agency. We used google maps, cross tabs and bar plots to answer raised questions.

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 ---------------------------------------------------------------------------
```

```
## v ggplot2 3.2.1     v purrr   0.3.2
## v tibble  2.1.3     v dplyr   0.8.3
## v tidyr   1.0.0     v stringr 1.4.0
## v readr   1.3.1     v forcats 0.4.0
```

```
## -- 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", ".")
mini311<-nyc311[sample(nrow(nyc311),10000),]
write.csv(mini311,"mini311.csv")
```

Reading the Sampled Data

```
sample<-fread("mini311.csv")
```

Data Exploration

I. Noise Complaints

In this section we are trying to display all the noise complaints in New York and explore them more categorically.

Selecting all noise complaints out of the sampled data

```
complaintLocations <- sample %>%
  select(Complaint.Type,
    Longitude,
    Latitude
  )
noiseComplaintLocations <- complaintLocations %>%
  filter(str_detect(Complaint.Type,"Noise"))
```

Plotting the noise complaints on a map

```
library(ggmap)
```

```
## Google's Terms of Service: https://cloud.google.com/maps-platform/terms/.
```

```
## Please cite ggmap if you use it! See citation("ggmap") for details.
```

```
library(curl)
```

```
##
## Attaching package: 'curl'
```

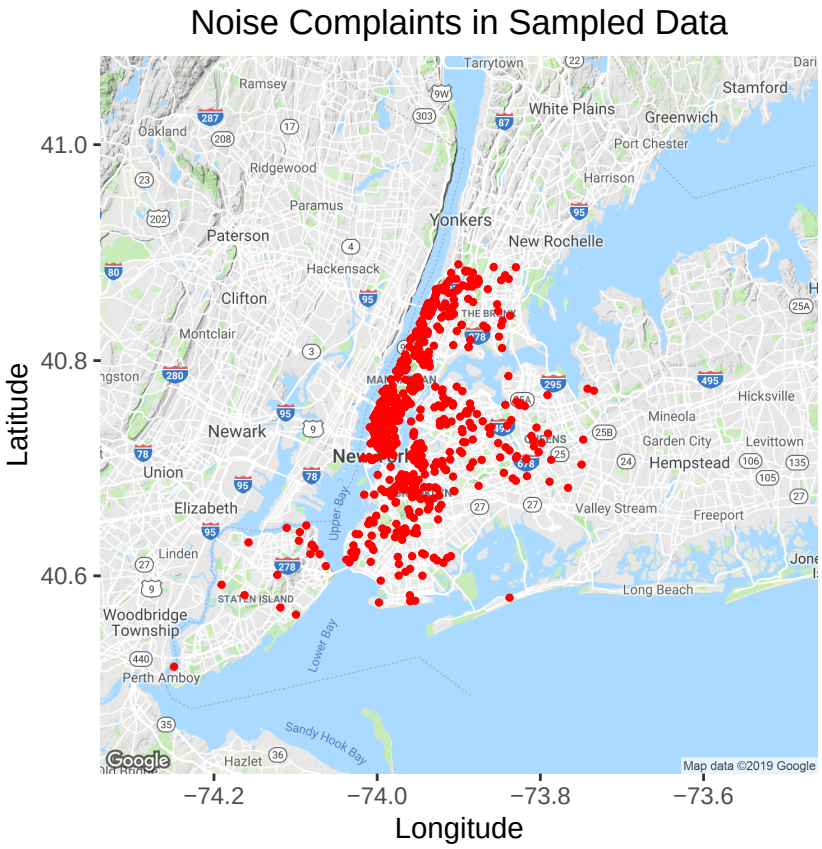
```
## The following object is masked from 'package:readr':
##
##     parse_date
```

```
key <- "AIzaSyClTqcMNpFm9_rFaaXH6ptzDpmTmAEwml4"
register_google(key=key)
nyc_map <- get_map(location=c(lon=-73.9,lat=40.75),
           maptype="terrain",zoom=10)
```

```
## Source : https://maps.googleapis.com/maps/api/staticmap?center=40.75,-73.9&zoom=10&size=640x640&scal
```

```
map <- ggmap(nyc_map) +
  geom_point(data=noiseComplaintLocations,aes(x=Longitude,y=Latitude),
         size=0.8,color="red") +
  ggtitle("Noise Complaints in Sampled Data") +
  theme(plot.title=element_text(hjust=0.5)) +
  xlab("Longitude") + ylab("Latitude")
map
```

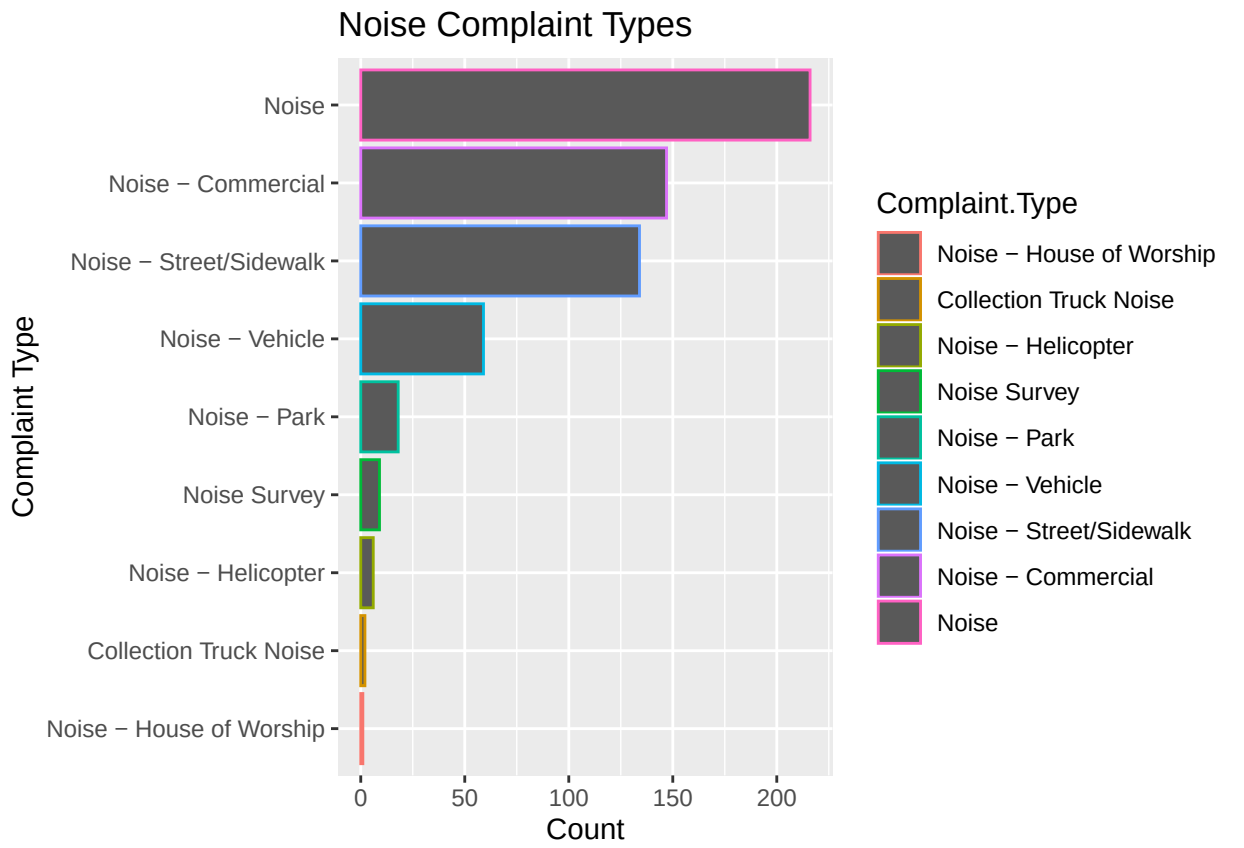
```
## Warning: Removed 7 rows containing missing values (geom_point).
```



What are the different noise complaints and which has the highest count after general noise complaints?

```
noiseComplaintData<-noiseComplaintLocations %>%
  group_by(Complaint.Type) %>%
  summarise(Count = n()) %>%
  arrange(desc(Count)) %>%
  ungroup() %>%
  mutate(Complaint.Type = reorder(Complaint.Type,Count))

ggplot(data=noiseComplaintData, aes(x = Complaint.Type,y = Count, color = Complaint.Type)) +
    xlab("Complaint Type") + ylab("Count") +
    geom_bar(stat='identity') +
```

```
    ggtitle("Noise Complaint Types") +
    coord_flip()
```

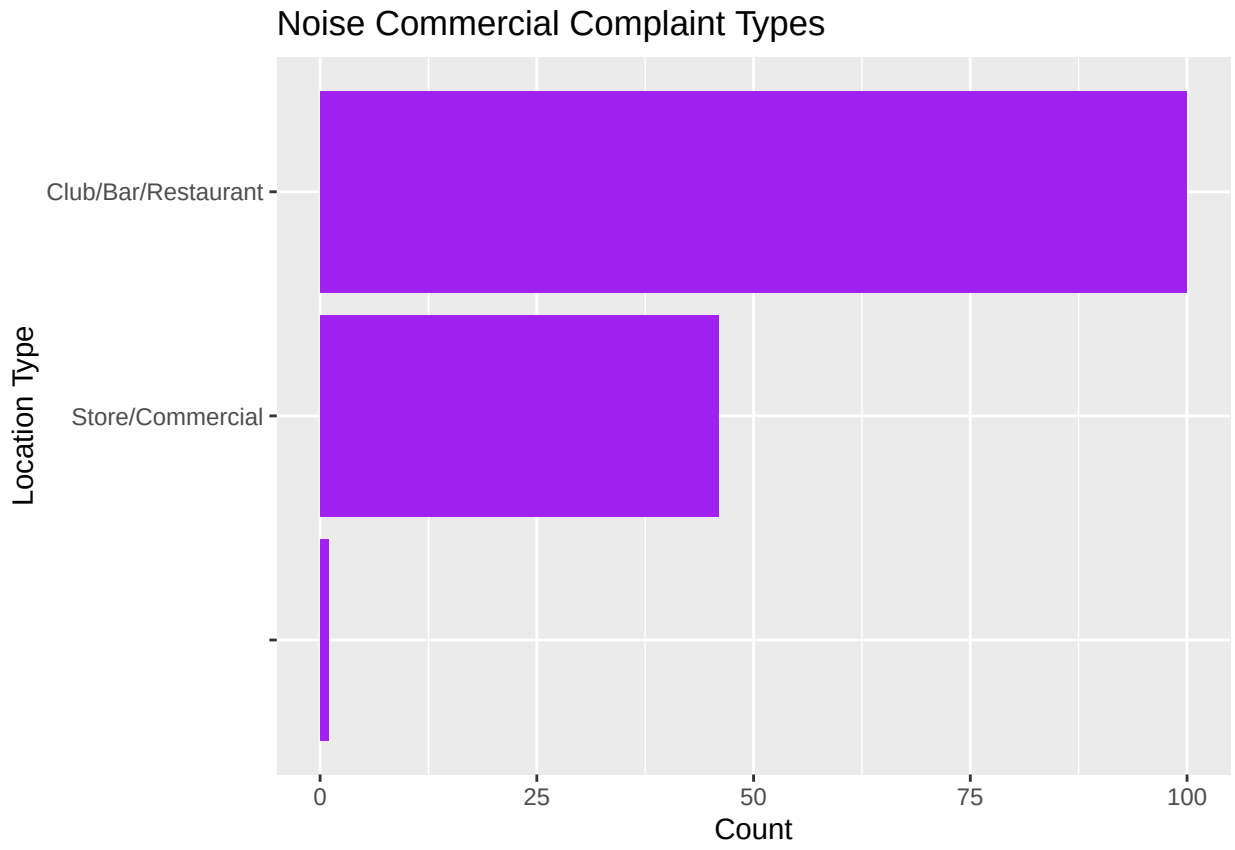


As we can see, Noise - Commercial is the category with second highest number of noise complaints. Next, we will see the distribution for commercial noise complaints.

Commercial-Noise complaints

```
require(dplyr)
complaintLocationsData <- sample %>%
  dplyr::select(Location.Type,
         Complaint.Type,
         Longitude,
         Latitude)

commercialNoiseData <- complaintLocationsData %>%
  filter(complaintLocationsData$Complaint.Type == "Noise - Commercial")

commercialNoiseCounts <- commercialNoiseData %>%
  group_by(Location.Type) %>%
  summarise(Count = n()) %>%
  arrange(desc(Count)) %>%
  ungroup() %>%
  mutate(Location.Type = reorder(Location.Type,Count))
```

```
ggplot(data=commercialNoiseCounts, aes(x = Location.Type,y = Count)) +
    xlab("Location Type") + ylab("Count") +
    geom_bar(stat='identity', fill="purple") +
    ggtitle("Noise Commercial Complaint Types") +
    coord_flip()
```



As we can see, Clubs/Bars/Restaurants contribute to the most number of commercial noise complaints. Lets see what it looks like on a map.

Selecting Commercial-Noise complaints by Clubs/Bars/Restaurants (Location Type) out of the sampled data

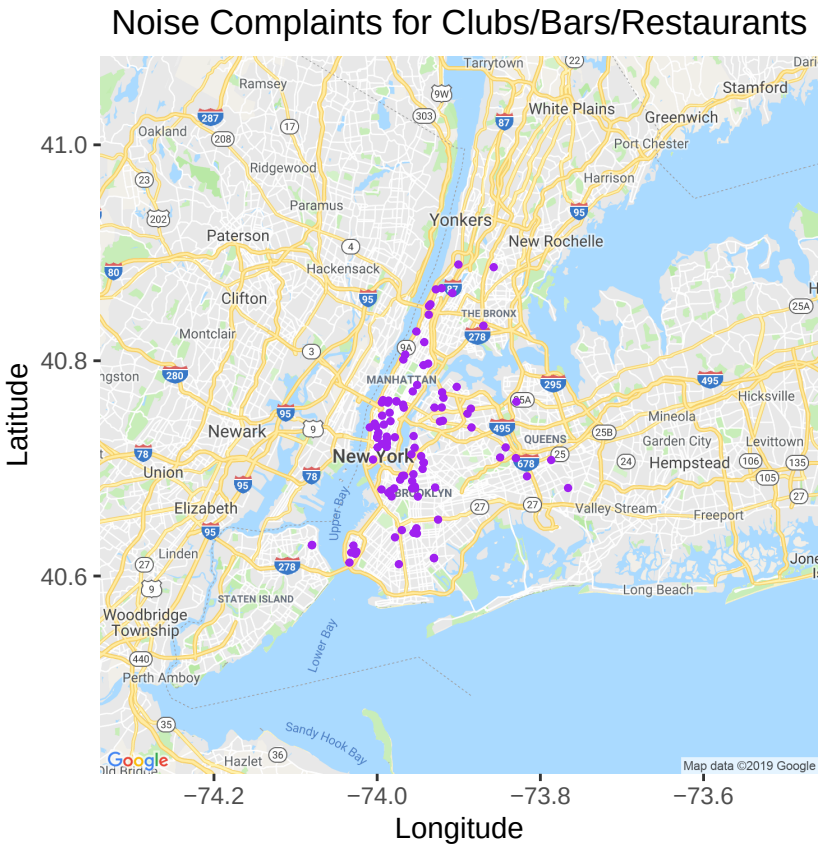
```
cBRLocs <- commercialNoiseData %>%
  filter(Location.Type=="Club/Bar/Restaurant")

cBRLocs<- cBRLocs %>%
  select(Location.Type,
    Longitude,
    Latitude
)
```

Plotting the Clubs/Bars/Restaurants noise complaints on a map

```
library(ggmap)
library(curl)
key <- "AIzaSyClTqcMNpFm9_rFaaXH6ptzDpmTmAEwml4"
register_google(key=key)
nyc_map <- get_map(location=c(lon=-73.9,lat=40.75),
           maptype="roadmap",zoom=10)
```

```
## Source : https://maps.googleapis.com/maps/api/staticmap?center=40.75,-73.9&zoom=10&size=640x640&scal
```

```
map <- ggmap(nyc_map) +
  geom_point(data=cBRLocs,aes(x=Longitude,y=Latitude),
         size=0.8,color="purple") +
  ggtitle("Noise Complaints for Clubs/Bars/Restaurants") +
  theme(plot.title=element_text(hjust=0.5)) +
  xlab("Longitude") + ylab("Latitude")
map
```



II. Heating Complaints

In this section, we will be exploring the heating complaints in New York, categorized by Borough.

Selecting all the heating complaints by Borough

```
heatingLocations <- sample %>%
  select(Borough, Complaint.Type,
    Longitude,
    Latitude,
    Status
  )
heatingLocationsFiltered <- heatingLocations %>%
  filter(heatingLocations$Complaint.Type == "HEATING")

boroughHeatingLocations<- heatingLocationsFiltered %>%
  select(Borough,
    Longitude,
    Latitude
)
```

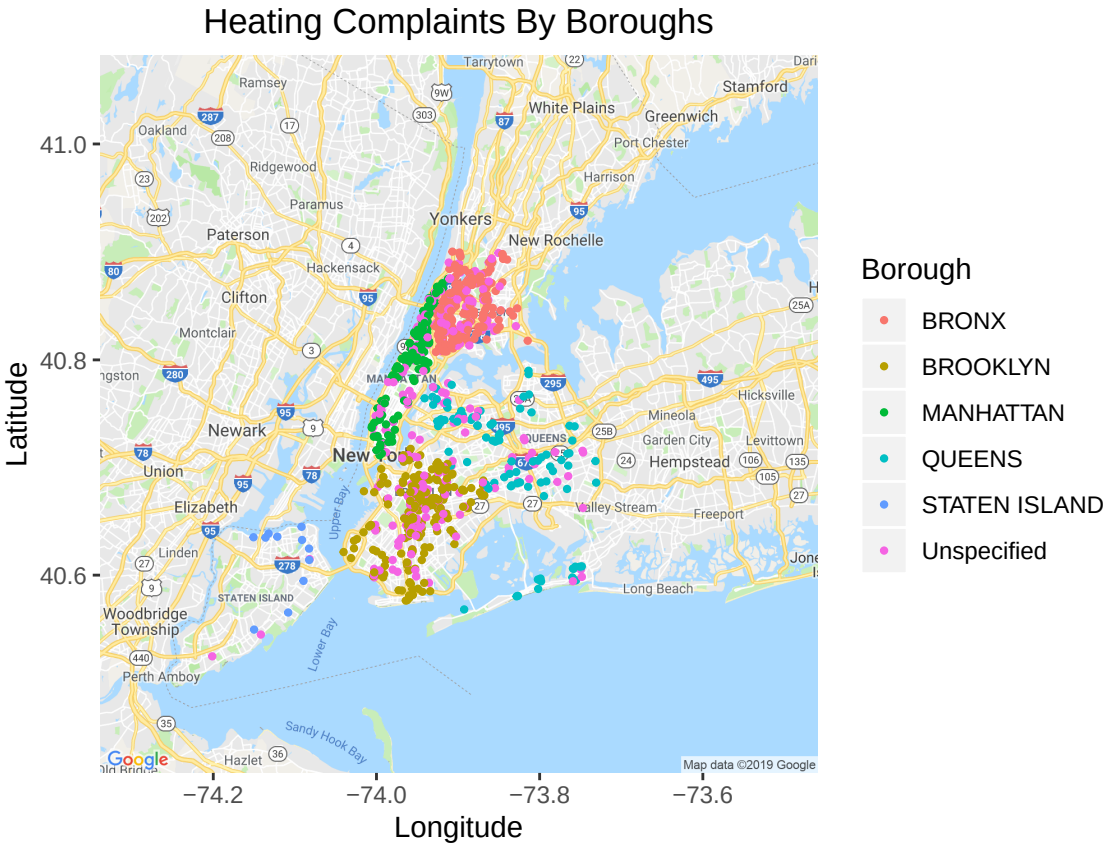
Lets see the how heatings complaints look on the map when plotted borough-wise

```
library(ggmap)
library(curl)
key <- "AIzaSyClTqcMNpFm9_rFaaXH6ptzDpmTmAEwml4"
register_google(key=key)
nyc_map <- get_map(location=c(lon=-73.9,lat=40.75),
           maptype="roadmap",zoom=10)
```

```
## Source : https://maps.googleapis.com/maps/api/staticmap?center=40.75,-73.9&zoom=10&size=640x640&scal
```

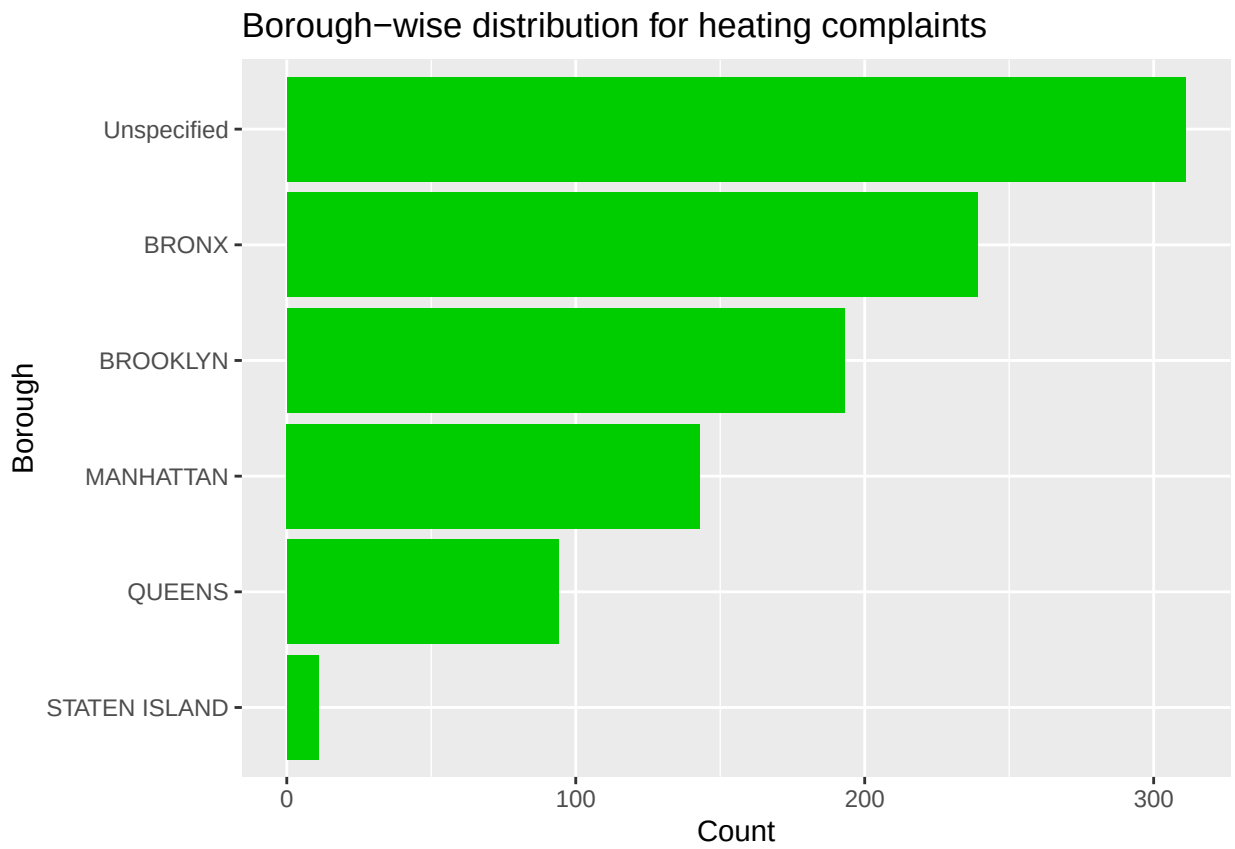
```
map <- ggmap(nyc_map) +
  geom_point(data=boroughHeatingLocations,aes(x=Longitude,y=Latitude, color=Borough),
         size=0.8) +
  ggtitle("Heating Complaints By Boroughs") +
  theme(plot.title=element_text(hjust=0.5)) +
  xlab("Longitude") + ylab("Latitude")
map
```

```
## Warning: Removed 17 rows containing missing values (geom_point).
```



Which known borough has the highest number of Heating complaints ?

```
heatingBoroughCounts<-sample %>%
  filter(Complaint.Type=="HEATING") %>%
  group_by(Borough) %>%
  summarise(Count = n()) %>%
  arrange(desc(Count)) %>%
  ungroup() %>%
  mutate(Borough = reorder(Borough,Count))

ggplot(data=heatingBoroughCounts, aes(x = Borough,y = Count)) +
  xlab("Borough") + ylab("Count") +
    geom_bar(stat='identity', fill="green3") +
    ggtitle("Borough-wise distribution for heating complaints") +
    coord_flip()
```



Looking at the bar chart, Brooklyn seems to be the known location having the most number of heating complaints.

Let's see how the status wise distribution looks for these boroughs

```
xtabA<-dplyr::filter(heatingLocationsFiltered)
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:  991
##
##
##                | xtabB$Status
## xtabB$Borough |    Closed |      Open | Row Total |
## --------------|-----------|-----------|-----------|
##         BRONX |       237 |         2 |       239 |
##               |     0.003 |     0.276 |           |
```

```
##               |     0.992 |     0.008 |     0.241 |
##               |     0.242 |     0.167 |           |
##               |     0.239 |     0.002 |           |
## --------------|-----------|-----------|-----------|
##      BROOKLYN |       191 |         2 |       193 |
##               |     0.001 |     0.049 |           |
##               |     0.990 |     0.010 |     0.195 |
##               |     0.195 |     0.167 |           |
##               |     0.193 |     0.002 |           |
## --------------|-----------|-----------|-----------|
##     MANHATTAN |       140 |         3 |       143 |
##               |     0.011 |     0.929 |           |
##               |     0.979 |     0.021 |     0.144 |
##               |     0.143 |     0.250 |           |
##               |     0.141 |     0.003 |           |
## --------------|-----------|-----------|-----------|
##        QUEENS |        93 |         1 |        94 |
##               |     0.000 |     0.017 |           |
##               |     0.989 |     0.011 |     0.095 |
##               |     0.095 |     0.083 |           |
##               |     0.094 |     0.001 |           |
## --------------|-----------|-----------|-----------|
## STATEN ISLAND |        11 |         0 |        11 |
##               |     0.002 |     0.133 |           |
##               |     1.000 |     0.000 |     0.011 |
##               |     0.011 |     0.000 |           |
##               |     0.011 |     0.000 |           |
## --------------|-----------|-----------|-----------|
##   Unspecified |       307 |         4 |       311 |
##               |     0.000 |     0.015 |           |
##               |     0.987 |     0.013 |     0.314 |
##               |     0.314 |     0.333 |           |
##               |     0.310 |     0.004 |           |
## --------------|-----------|-----------|-----------|
##  Column Total |       979 |        12 |       991 |
##               |     0.988 |     0.012 |           |
## --------------|-----------|-----------|-----------|
##
##
```

Again, we can see that Brooklyn has the max number of heating complaints, so lets see the status wise map plot of Brooklyn's heating complaints.

Status wise brooklyn heating complaints

```
heatingStatusLocations <- heatingLocationsFiltered %>%
  select(Borough, Status,
    Longitude,
    Latitude
  )
heatingStatusLocationsFiltered <- heatingStatusLocations %>%
  filter(heatingStatusLocations$Borough == "BROOKLYN")

brooklynHeatingStatus<- heatingStatusLocationsFiltered %>%
  select(Status,
    Longitude,
    Latitude
)
```

Map plots for the complaints

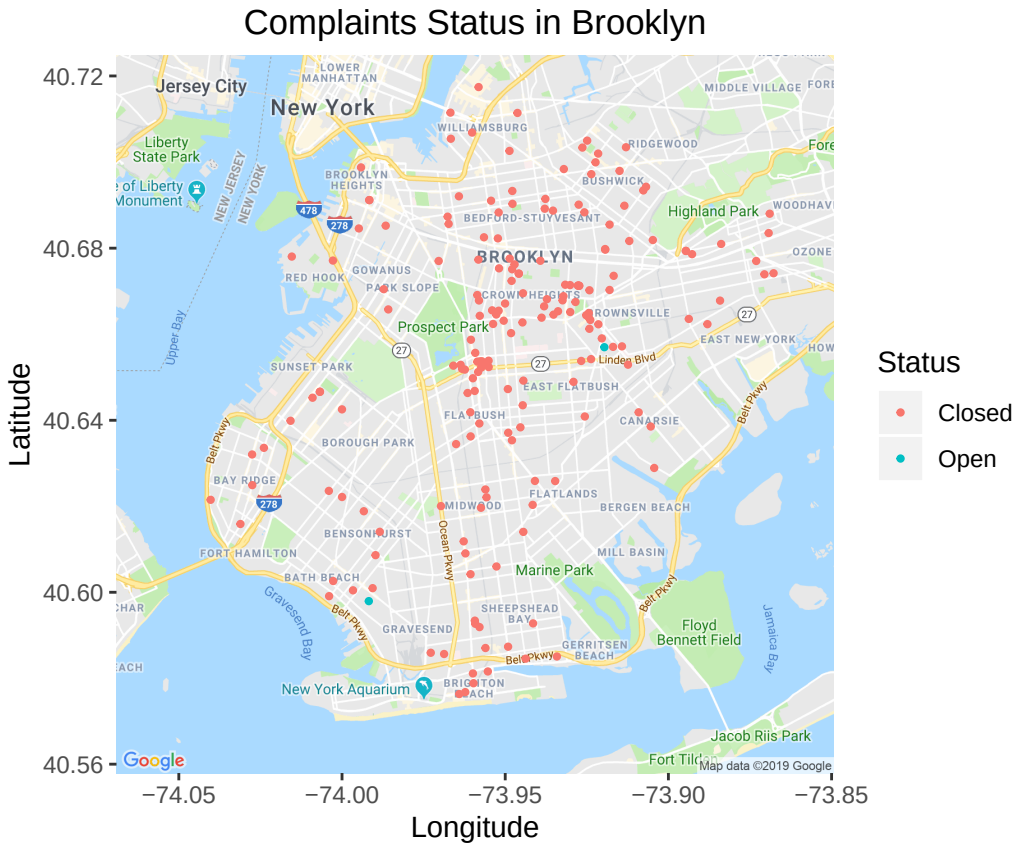
```
library(ggmap)
library(curl)
key <- "AIzaSyClTqcMNpFm9_rFaaXH6ptzDpmTmAEwml4"
register_google(key=key)
nyc_map <- get_map(location="flatbush",
           maptype="roadmap",zoom=12)
```

```
## Source : https://maps.googleapis.com/maps/api/staticmap?center=flatbush&zoom=12&size=640x640&scale=2
```

```
## Source : https://maps.googleapis.com/maps/api/geocode/json?address=flatbush&key=xxx
```

```
map <- ggmap(nyc_map) +
  geom_point(data=brooklynHeatingStatus,aes(x=Longitude,y=Latitude, color=Status),
         size=0.8) +
  ggtitle("Complaints Status in Brooklyn") +
  theme(plot.title=element_text(hjust=0.5)) +
  xlab("Longitude") + ylab("Latitude")
map
```

```
## Warning: Removed 4 rows containing missing values (geom_point).
```



Most of the heating complaints seem to be closed in Brooklyn.

III. A look at the Agency field

Which agency has the highest number of complaints reported?

```
xtabA<-dplyr::filter(sample)
xtabB<-select(xtabA,Agency,"Status")
library(gmodels)
CrossTable(xtabB$Agency,xtabB$'Status')
```

```
##
##
##    Cell Contents
## |-------------------------|
## |                       N |
## | Chi-square contribution |
## |           N / Row Total |
## |           N / Col Total |
## |         N / Table Total |
## |-------------------------|
##
##
## Total Observations in Table:  10000
##
##
##              | xtabB$Status
## xtabB$Agency |   Assigned |     Closed | Email Sent |       Open |    Pending |    Started |  Row Total |
## -------------|------------|------------|------------|------------|------------|------------|------------|
##        3-1-1 |          0 |         18 |          0 |          5 |          0 |          0 |         23 |
##              |      0.373 |      0.135 |      0.053 |      3.354 |      0.678 |      0.014 |            |
##              |      0.000 |      0.783 |      0.000 |      0.217 |      0.000 |      0.000 |      0.002 |
##              |      0.000 |      0.002 |      0.000 |      0.005 |      0.000 |      0.000 |            |
##              |      0.000 |      0.002 |      0.000 |      0.001 |      0.000 |      0.000 |            |
## -------------|------------|------------|------------|------------|------------|------------|------------|
##        CHALL |          0 |          0 |         15 |          0 |          0 |          0 |         15 |
##              |      0.243 |     12.803 |   6491.774 |      1.468 |      0.443 |      0.009 |            |
##              |      0.000 |      0.000 |      1.000 |      0.000 |      0.000 |      0.000 |      0.002 |
##              |      0.000 |      0.000 |      0.652 |      0.000 |      0.000 |      0.000 |            |
##              |      0.000 |      0.000 |      0.002 |      0.000 |      0.000 |      0.000 |            |
## -------------|------------|------------|------------|------------|------------|------------|------------|
##          DCA |          0 |        129 |          0 |          0 |          0 |          0 |        129 |
##              |      2.090 |      3.244 |      0.297 |     12.629 |      3.805 |      0.077 |            |
##              |      0.000 |      1.000 |      0.000 |      0.000 |      0.000 |      0.000 |      0.013 |
##              |      0.000 |      0.015 |      0.000 |      0.000 |      0.000 |      0.000 |            |
##              |      0.000 |      0.013 |      0.000 |      0.000 |      0.000 |      0.000 |            |
## -------------|------------|------------|------------|------------|------------|------------|------------|
##          DEP |          2 |        735 |          0 |        160 |          0 |          6 |        903 |
##              |     10.902 |      1.655 |      2.077 |     57.984 |     26.639 |     54.987 |            |
##              |      0.002 |      0.814 |      0.000 |      0.177 |      0.000 |      0.007 |      0.090 |
##              |      0.012 |      0.086 |      0.000 |      0.163 |      0.000 |      1.000 |            |
##              |      0.000 |      0.073 |      0.000 |      0.016 |      0.000 |      0.001 |            |
## -------------|------------|------------|------------|------------|------------|------------|------------|
##          DHS |          0 |          3 |          0 |          0 |          0 |          0 |          3 |
##              |      0.049 |      0.075 |      0.007 |      0.294 |      0.088 |      0.002 |            |
##              |      0.000 |      1.000 |      0.000 |      0.000 |      0.000 |      0.000 |      0.000 |
##              |      0.000 |      0.000 |      0.000 |      0.000 |      0.000 |      0.000 |            |
##              |      0.000 |      0.000 |      0.000 |      0.000 |      0.000 |      0.000 |            |
## -------------|------------|------------|------------|------------|------------|------------|------------|
##          DOB |         18 |        450 |          0 |         82 |          0 |          0 |        550 |
##              |      9.274 |      0.804 |      1.265 |     14.722 |     16.225 |      0.330 |            |
##              |      0.033 |      0.818 |      0.000 |      0.149 |      0.000 |      0.000 |      0.055 |
##              |      0.111 |      0.053 |      0.000 |      0.084 |      0.000 |      0.000 |            |
##              |      0.002 |      0.045 |      0.000 |      0.008 |      0.000 |      0.000 |            |
## -------------|------------|------------|------------|------------|------------|------------|------------|
##          DOE |          0 |          8 |          0 |          0 |          0 |          0 |          8 |
##              |      0.130 |      0.201 |      0.018 |      0.783 |      0.236 |      0.005 |            |
```

```
##              |      0.000 |      1.000 |      0.000 |      0.000 |      0.000 |      0.000 |      0.001 |
##              |      0.000 |      0.001 |      0.000 |      0.000 |      0.000 |      0.000 |            |
##              |      0.000 |      0.001 |      0.000 |      0.000 |      0.000 |      0.000 |            |
## -------------|------------|------------|------------|------------|------------|------------|------------|
##          DOF |          0 |        142 |          0 |          0 |          0 |          0 |        142 |
##              |      2.300 |      3.571 |      0.327 |     13.902 |      4.189 |      0.085 |            |
##              |      0.000 |      1.000 |      0.000 |      0.000 |      0.000 |      0.000 |      0.014 |
##              |      0.000 |      0.017 |      0.000 |      0.000 |      0.000 |      0.000 |            |
##              |      0.000 |      0.014 |      0.000 |      0.000 |      0.000 |      0.000 |            |
## -------------|------------|------------|------------|------------|------------|------------|------------|
##        DOHMH |         50 |        144 |          1 |         11 |         72 |          0 |        278 |
##              |    459.615 |     36.666 |      0.203 |      9.662 |    496.319 |      0.167 |            |
##              |      0.180 |      0.518 |      0.004 |      0.040 |      0.259 |      0.000 |      0.028 |
##              |      0.309 |      0.017 |      0.043 |      0.011 |      0.244 |      0.000 |            |
##              |      0.005 |      0.014 |      0.000 |      0.001 |      0.007 |      0.000 |            |
## -------------|------------|------------|------------|------------|------------|------------|------------|
##        DOITT |          0 |          6 |          0 |          0 |          0 |          0 |          6 |
##              |      0.097 |      0.151 |      0.014 |      0.587 |      0.177 |      0.004 |            |
##              |      0.000 |      1.000 |      0.000 |      0.000 |      0.000 |      0.000 |      0.001 |
##              |      0.000 |      0.001 |      0.000 |      0.000 |      0.000 |      0.000 |            |
##              |      0.000 |      0.001 |      0.000 |      0.000 |      0.000 |      0.000 |            |
## -------------|------------|------------|------------|------------|------------|------------|------------|
##          DOT |         21 |       1582 |          1 |         38 |        149 |          0 |       1791 |
##              |      2.214 |      1.864 |      2.362 |    107.574 |    175.033 |      1.075 |            |
##              |      0.012 |      0.883 |      0.001 |      0.021 |      0.083 |      0.000 |      0.179 |
##              |      0.130 |      0.185 |      0.043 |      0.039 |      0.505 |      0.000 |            |
##              |      0.002 |      0.158 |      0.000 |      0.004 |      0.015 |      0.000 |            |
## -------------|------------|------------|------------|------------|------------|------------|------------|
##          DPR |         63 |        338 |          0 |         61 |          0 |          0 |        462 |
##              |    411.787 |      8.043 |      1.063 |      5.499 |     13.629 |      0.277 |            |
##              |      0.136 |      0.732 |      0.000 |      0.132 |      0.000 |      0.000 |      0.046 |
##              |      0.389 |      0.040 |      0.000 |      0.062 |      0.000 |      0.000 |            |
##              |      0.006 |      0.034 |      0.000 |      0.006 |      0.000 |      0.000 |            |
## -------------|------------|------------|------------|------------|------------|------------|------------|
##         DSNY |          3 |        595 |          1 |         33 |         74 |          0 |        706 |
##              |      6.224 |      0.095 |      0.240 |     18.873 |    135.755 |      0.424 |            |
##              |      0.004 |      0.843 |      0.001 |      0.047 |      0.105 |      0.000 |      0.071 |
##              |      0.019 |      0.070 |      0.043 |      0.034 |      0.251 |      0.000 |            |
##              |      0.000 |      0.059 |      0.000 |      0.003 |      0.007 |      0.000 |            |
## -------------|------------|------------|------------|------------|------------|------------|------------|
##          EDC |          0 |          6 |          0 |          0 |          0 |          0 |          6 |
##              |      0.097 |      0.151 |      0.014 |      0.587 |      0.177 |      0.004 |            |
##              |      0.000 |      1.000 |      0.000 |      0.000 |      0.000 |      0.000 |      0.001 |
##              |      0.000 |      0.001 |      0.000 |      0.000 |      0.000 |      0.000 |            |
##              |      0.000 |      0.001 |      0.000 |      0.000 |      0.000 |      0.000 |            |
## -------------|------------|------------|------------|------------|------------|------------|------------|
##         FDNY |          0 |         36 |          1 |          0 |          0 |          0 |         37 |
##              |      0.599 |      0.619 |      9.836 |      3.622 |      1.091 |      0.022 |            |
##              |      0.000 |      0.973 |      0.027 |      0.000 |      0.000 |      0.000 |      0.004 |
##              |      0.000 |      0.004 |      0.043 |      0.000 |      0.000 |      0.000 |            |
##              |      0.000 |      0.004 |      0.000 |      0.000 |      0.000 |      0.000 |            |
## -------------|------------|------------|------------|------------|------------|------------|------------|
##          HPD |          0 |       3077 |          0 |        578 |          0 |          0 |       3655 |
##              |     59.211 |      0.580 |      8.406 |    135.478 |    107.823 |      2.193 |            |
##              |      0.000 |      0.842 |      0.000 |      0.158 |      0.000 |      0.000 |      0.365 |
##              |      0.000 |      0.361 |      0.000 |      0.590 |      0.000 |      0.000 |            |
##              |      0.000 |      0.308 |      0.000 |      0.058 |      0.000 |      0.000 |            |
## -------------|------------|------------|------------|------------|------------|------------|------------|
##          HRA |          0 |         29 |          3 |          0 |          0 |          0 |         32 |
##              |      0.518 |      0.104 |    116.356 |      3.133 |      0.944 |      0.019 |            |
##              |      0.000 |      0.906 |      0.094 |      0.000 |      0.000 |      0.000 |      0.003 |
##              |      0.000 |      0.003 |      0.130 |      0.000 |      0.000 |      0.000 |            |
##              |      0.000 |      0.003 |      0.000 |      0.000 |      0.000 |      0.000 |            |
## -------------|------------|------------|------------|------------|------------|------------|------------|
##         NYPD |          0 |       1111 |          0 |          0 |          0 |          0 |       1111 |
##              |     17.998 |     27.937 |      2.555 |    108.767 |     32.775 |      0.667 |            |
##              |      0.000 |      1.000 |      0.000 |      0.000 |      0.000 |      0.000 |      0.111 |
##              |      0.000 |      0.130 |      0.000 |      0.000 |      0.000 |      0.000 |            |
```

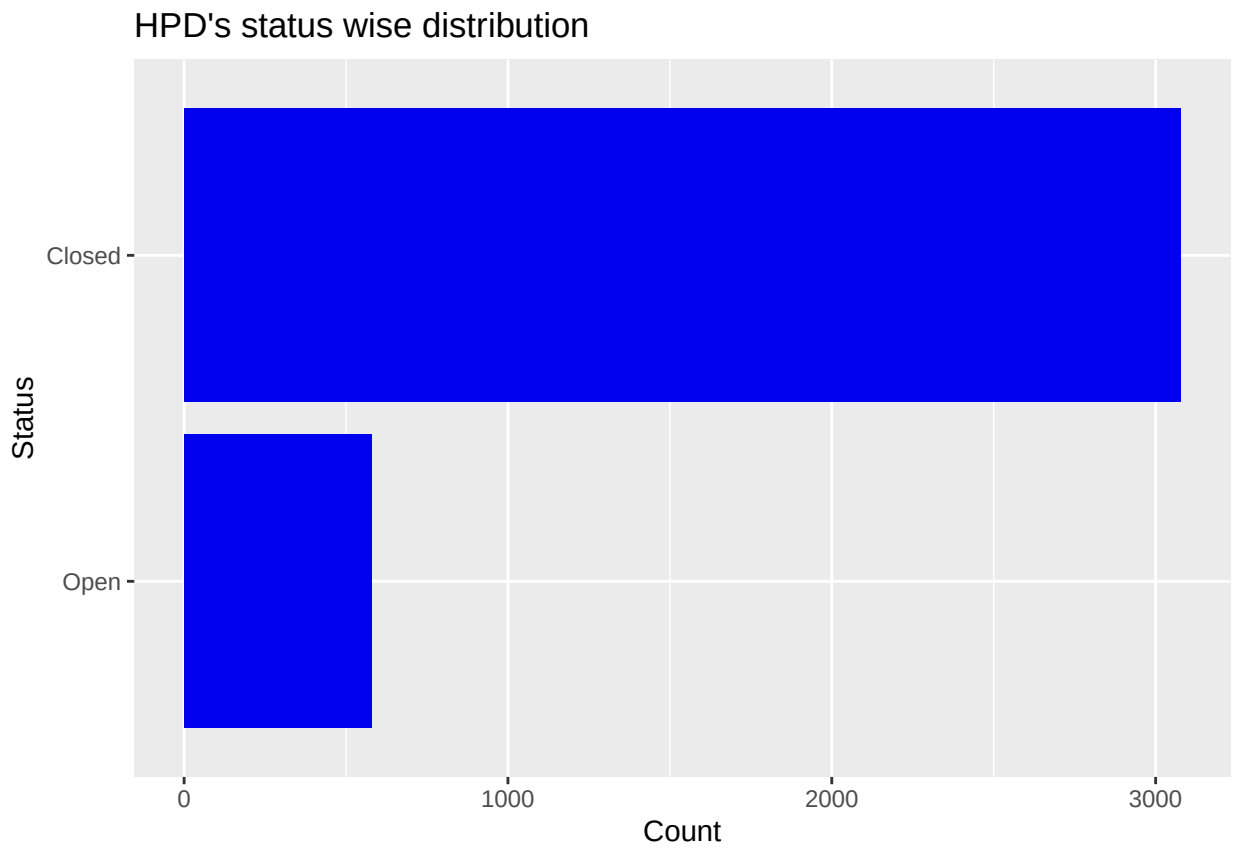
```
##              |      0.000 |      0.111 |      0.000 |      0.000 |      0.000 |      0.000 |            |
## -------------|------------|------------|------------|------------|------------|------------|------------|
##          OEM |          0 |          1 |          0 |          0 |          0 |          0 |          1 |
##              |      0.016 |      0.025 |      0.002 |      0.098 |      0.029 |      0.001 |            |
##              |      0.000 |      1.000 |      0.000 |      0.000 |      0.000 |      0.000 |      0.000 |
##              |      0.000 |      0.000 |      0.000 |      0.000 |      0.000 |      0.000 |            |
##              |      0.000 |      0.000 |      0.000 |      0.000 |      0.000 |      0.000 |            |
## -------------|------------|------------|------------|------------|------------|------------|------------|
##          OPS |          0 |          0 |          1 |          0 |          0 |          0 |          1 |
##              |      0.016 |      0.854 |    432.785 |      0.098 |      0.029 |      0.001 |            |
##              |      0.000 |      0.000 |      1.000 |      0.000 |      0.000 |      0.000 |      0.000 |
##              |      0.000 |      0.000 |      0.043 |      0.000 |      0.000 |      0.000 |            |
##              |      0.000 |      0.000 |      0.000 |      0.000 |      0.000 |      0.000 |            |
## -------------|------------|------------|------------|------------|------------|------------|------------|
##          TLC |          5 |        125 |          0 |         11 |          0 |          0 |        141 |
##              |      3.229 |      0.180 |      0.324 |      0.570 |      4.160 |      0.085 |            |
##              |      0.035 |      0.887 |      0.000 |      0.078 |      0.000 |      0.000 |      0.014 |
##              |      0.031 |      0.015 |      0.000 |      0.011 |      0.000 |      0.000 |            |
##              |      0.001 |      0.013 |      0.000 |      0.001 |      0.000 |      0.000 |            |
## -------------|------------|------------|------------|------------|------------|------------|------------|
## Column Total |        162 |       8535 |         23 |        979 |        295 |          6 |      10000 |
##              |      0.016 |      0.854 |      0.002 |      0.098 |      0.029 |      0.001 |            |
## -------------|------------|------------|------------|------------|------------|------------|------------|
##
##
```

HPD clearly stands out for the maximum number of complaints being reported.

Let's see how HPD is performing on the reported complaints.

Bar chart for HPD's status-wise distribution

```
hpdStatusCounts<-sample %>%
  filter(Agency=="HPD",
  Status=='Closed' |
  Status=='Open') %>%
  group_by(Status) %>%
  summarise(Count = n()) %>%
  arrange(desc(Count)) %>%
  ungroup() %>%
  mutate(Status = reorder(Status,Count))
ggplot(data=hpdStatusCounts, aes(x = Status,y = Count)) +
  ggtitle("HPD's status wise distribution") +
  xlab("Status") + ylab("Count") +
    geom_bar(stat='identity', fill="blue2") +
    coord_flip()
```



A major chunk of the complaints are closed, which shows they are doing a good job!

Let's go ahead and plot complaints reported to HPD on a map.

Selecting agency HPD's status wise location data

```
hpdStatusLocations <- sample %>%
  select(Agency, Status,
    Longitude,
    Latitude
  )
hpdStatusLocationsFiltered <- hpdStatusLocations %>%
  filter(hpdStatusLocations$Agency == "HPD",
  Status=='Closed' |
  Status=='Open')

hpdStatusLocationsFiltered<- hpdStatusLocationsFiltered %>%
  select(Status,
    Longitude,
    Latitude
)
```

Plotting agency HPD's status wise location data on a map

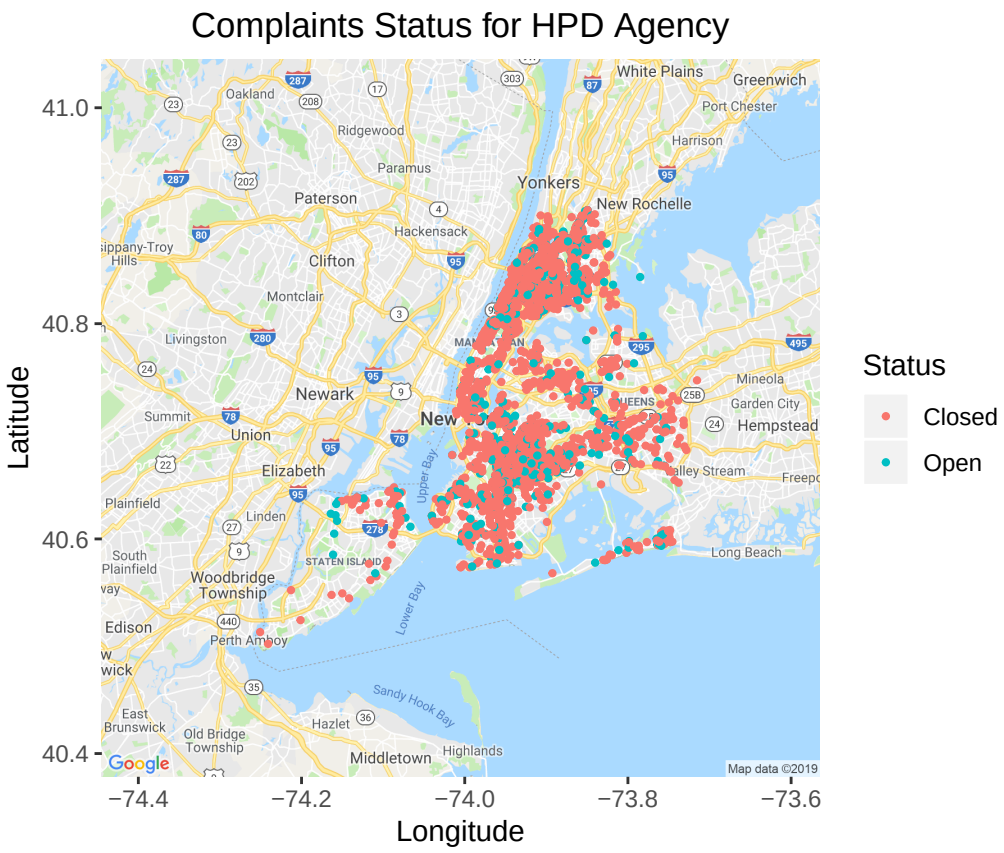
```
library(ggmap)
library(curl)
key <- "AIzaSyClTqcMNpFm9_rFaaXH6ptzDpmTmAEwml4"
register_google(key=key)
nyc_map <- get_map(location="New York City",
           maptype="roadmap",zoom=10)
```

```
## Source : https://maps.googleapis.com/maps/api/staticmap?center=New%20York%20City&zoom=10&size=640x64
```

```
## Source : https://maps.googleapis.com/maps/api/geocode/json?address=New+York+City&key=xxx
```

```
map <- ggmap(nyc_map) +
  geom_point(data=hpdStatusLocationsFiltered,aes(x=Longitude,y=Latitude, color=Status),
         size=0.8) +
  ggtitle("Complaints Status for HPD Agency") +
  theme(plot.title=element_text(hjust=0.5)) +
  xlab("Longitude") + ylab("Latitude")
map
```

```
## Warning: Removed 29 rows containing missing values (geom_point).
```



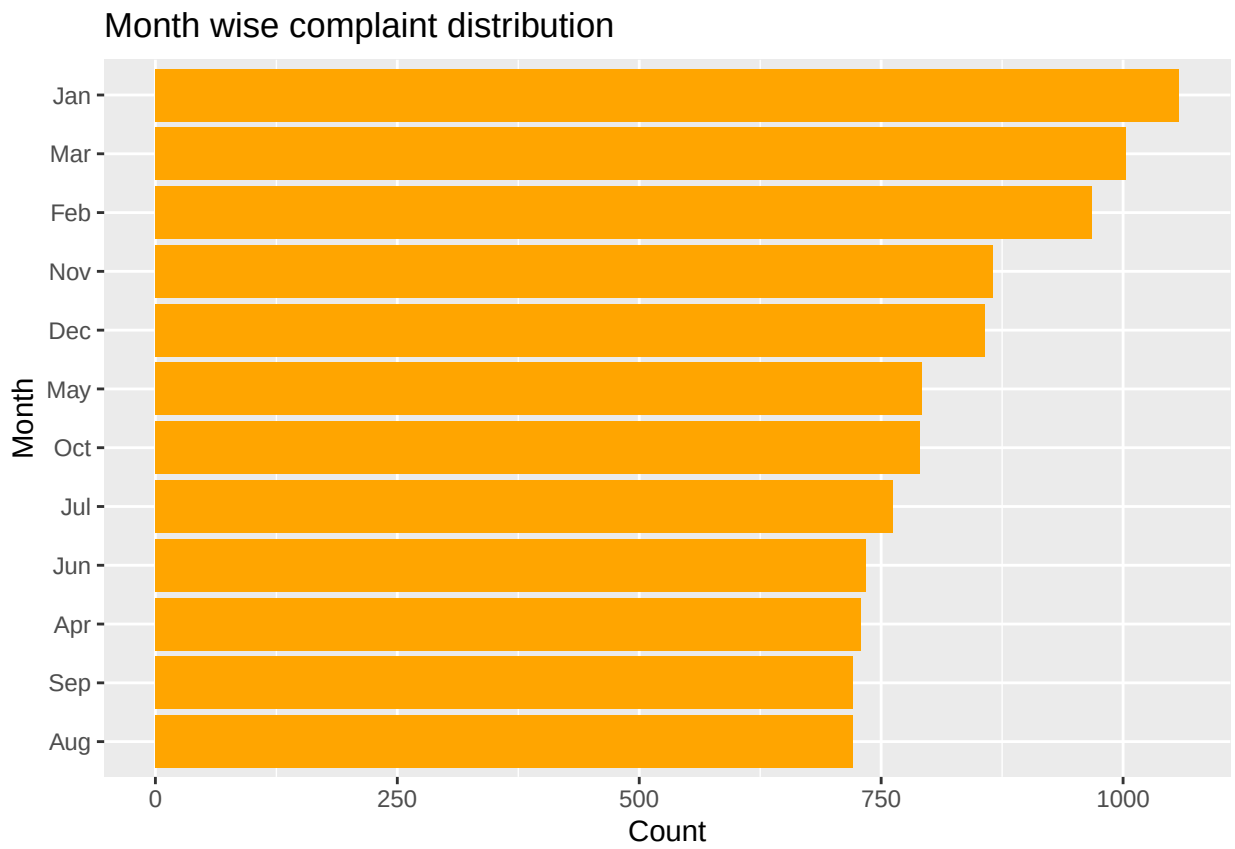
IV. NYC 311 complaints trend

Which month has the maximum number of complaints being reported?

```
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
```

```
trend<-sample %>%
  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(data=trend, aes(x = month,y = Count)) +
  xlab("Month") + ylab("Count") +
    ggtitle("Month wise complaint distribution") +
    geom_bar(stat='identity', fill="orange") +
    coord_flip()
```



We can clearly see that January has the highest number of complaints getting reported.

Let's see January's complaints plotted on a map borough wise.

Map plot for complaints by borough in January

```
janBoroughLocations <- sample %>%
  select(Created.Date, Borough,
    Longitude,
    Latitude
  ) %>%
  mutate(month = month.abb[month(mdy_hms(Created.Date))]) %>%
  filter(!is.na(month), month=='Jan')

janBoroughLocations<- janBoroughLocations %>%
  select(Borough,
    Longitude,
    Latitude
)
```

```
library(ggmap)
library(curl)
key <- "AIzaSyClTqcMNpFm9_rFaaXH6ptzDpmTmAEwml4"
register_google(key=key)
```

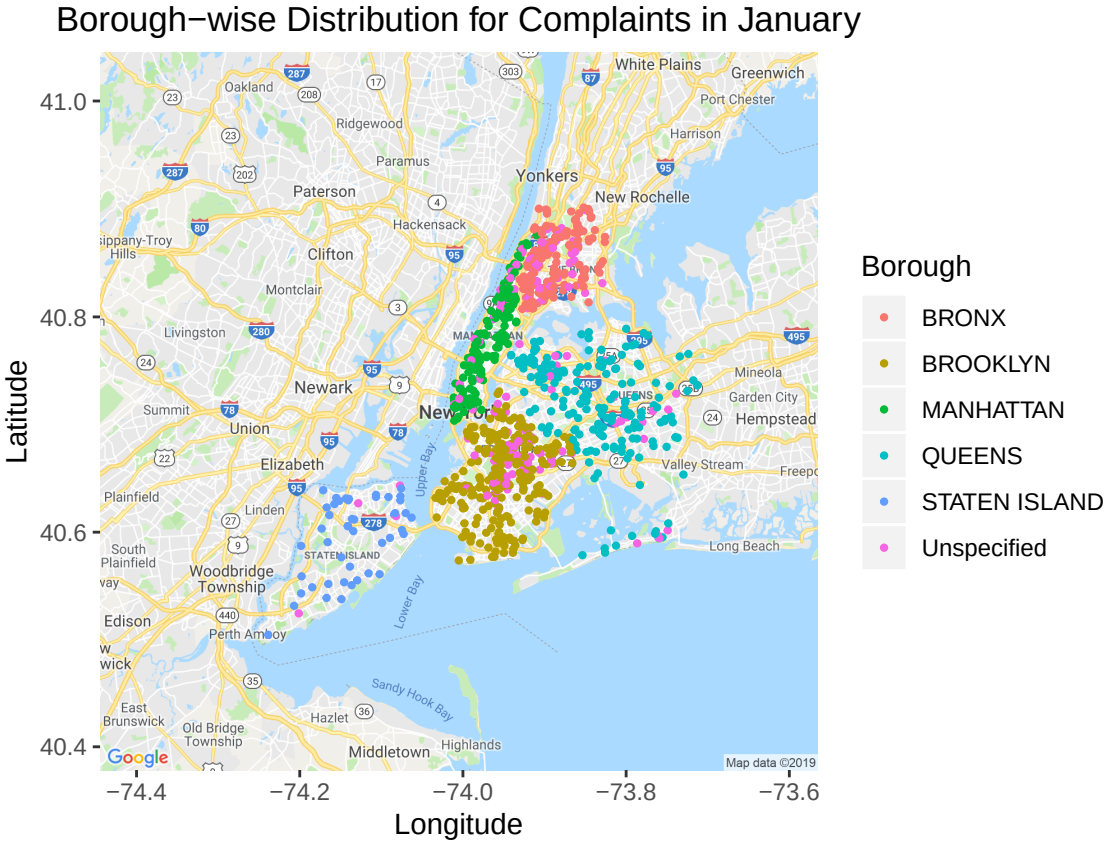
```
nyc_map <- get_map(location="New York City",
           maptype="roadmap",zoom=10)
```

```
## Source : https://maps.googleapis.com/maps/api/staticmap?center=New%20York%20City&zoom=10&size=640x64
```

```
## Source : https://maps.googleapis.com/maps/api/geocode/json?address=New+York+City&key=xxx
```

```
map <- ggmap(nyc_map) +
  geom_point(data=janBoroughLocations,aes(x=Longitude,y=Latitude, color=Borough),
         size=0.8) +
  ggtitle("Borough-wise Distribution for Complaints in January") +
  theme(plot.title=element_text(hjust=0.5)) +
  xlab("Longitude") + ylab("Latitude")
map
```

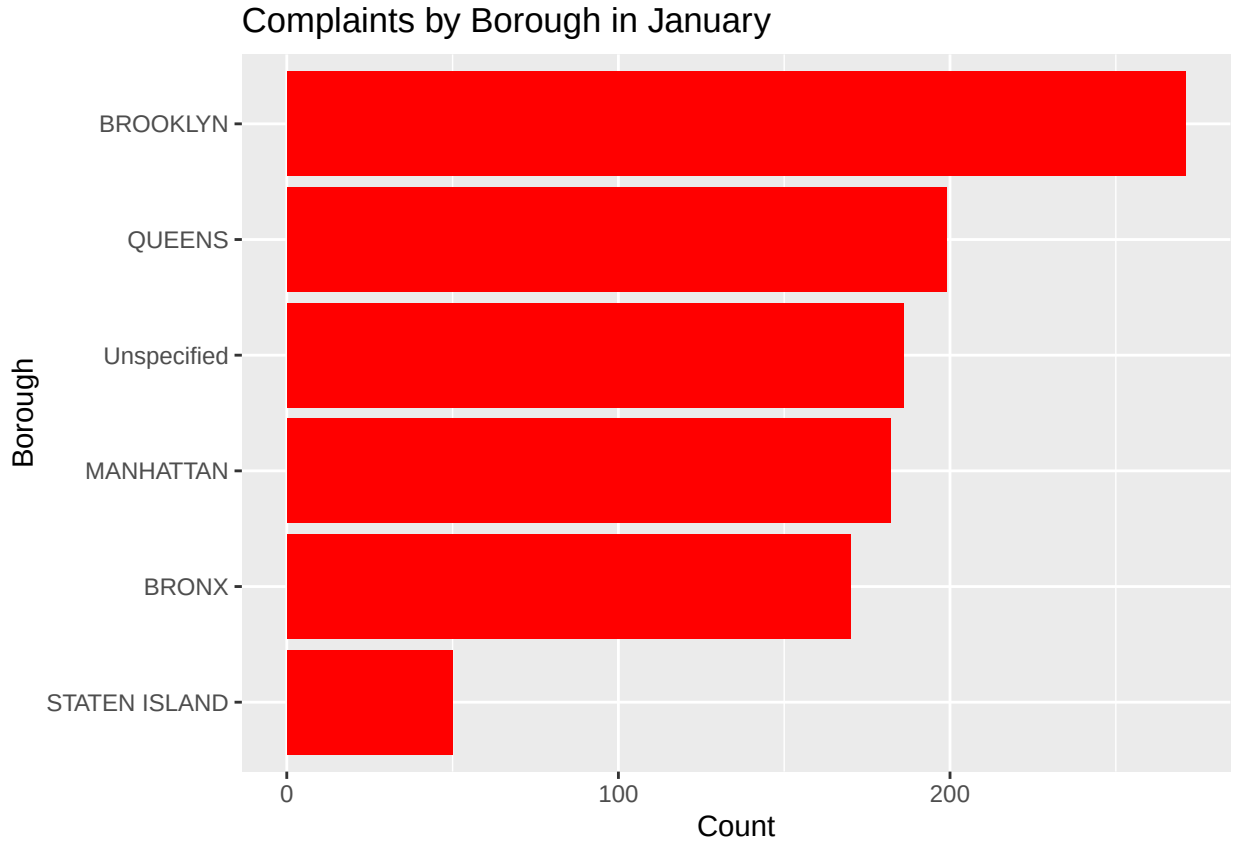
```
## Warning: Removed 85 rows containing missing values (geom_point).
```



Let's see which borough is having the highest complaints being reported in January.

Bar chart for complaints by borough in January

```
janBoroughCounts<- sample %>%
  mutate(month = month.abb[month(mdy_hms(Created.Date))]) %>%
  filter(!is.na(month), month=='Jan') %>%
  group_by(Borough) %>%
  summarise(Count = n()) %>%
  arrange(desc(Count)) %>%
  ungroup() %>%
  mutate(Borough = reorder(Borough,Count))
ggplot(data=janBoroughCounts, aes(x = Borough,y = Count)) +
  ggtitle("Complaints by Borough in January") +
  xlab("Borough") + ylab("Count") +
    geom_bar(stat='identity', fill="red") +
    coord_flip()
```



So we can see that Brooklyn has the maximum number of complaints reported in January.

Lets plot Brooklyn's complaints (January) on a map

Selecting complaints - Brooklyn/January

```
janBrooklynLocations <- sample %>%
  select(Created.Date, Borough,
    Longitude,
    Latitude
```

```
  ) %>%
  mutate(month = month.abb[month(mdy_hms(Created.Date))]) %>%
  filter(!is.na(month), month=='Jan', Borough=="BROOKLYN")

janBrooklynLocations<- janBrooklynLocations %>%
  select(Borough,
    Longitude,
    Latitude
)
```

Plotting complaints - Brooklyn/January

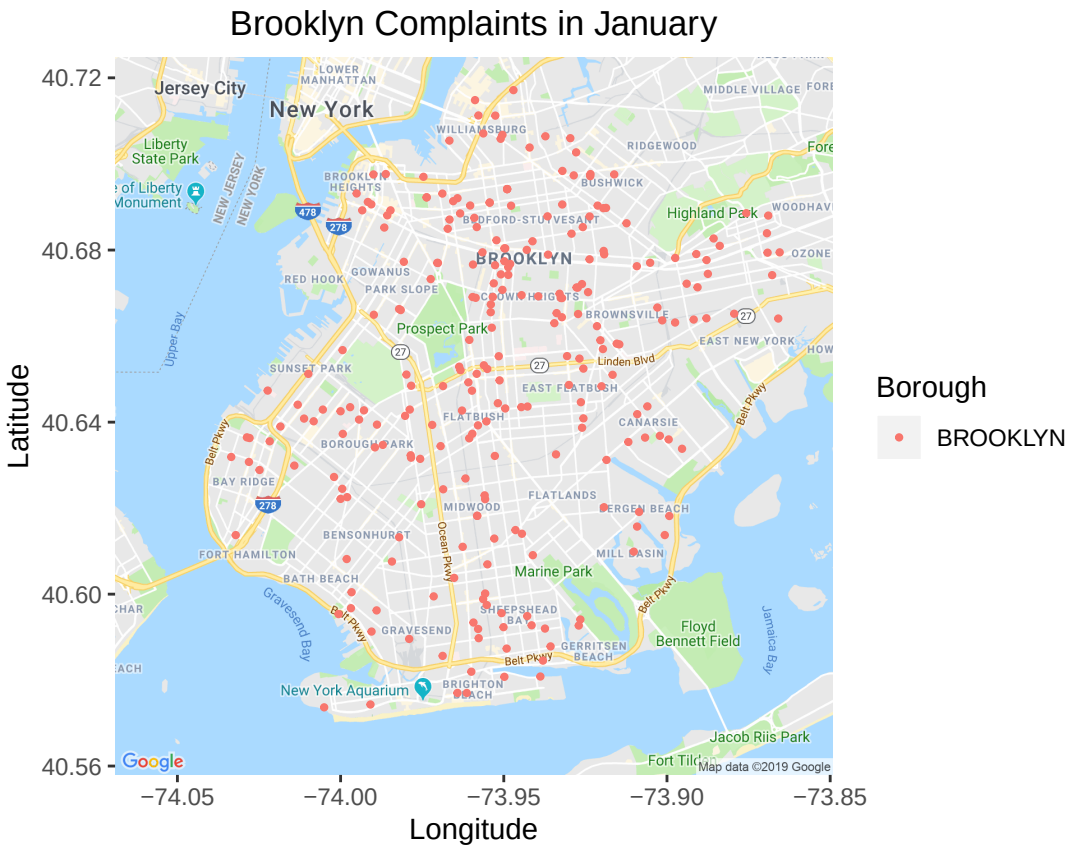
```
library(ggmap)
library(curl)
key <- "AIzaSyClTqcMNpFm9_rFaaXH6ptzDpmTmAEwml4"
register_google(key=key)
nyc_map <- get_map(location="Flatbush",
           maptype="roadmap",zoom=12)
```

```
## Source : https://maps.googleapis.com/maps/api/staticmap?center=Flatbush&zoom=12&size=640x640&scale=2
```

```
## Source : https://maps.googleapis.com/maps/api/geocode/json?address=Flatbush&key=xxx
```

```
map <- ggmap(nyc_map) +
  geom_point(data=janBrooklynLocations,aes(x=Longitude,y=Latitude, color=Borough),
         size=0.8) +
  ggtitle("Brooklyn Complaints in January") +
  theme(plot.title=element_text(hjust=0.5)) +
  xlab("Longitude") + ylab("Latitude")
map
```

```
## Warning: Removed 15 rows containing missing values (geom_point).
```


Brooklyn Complaints in January
Latitude
40.72
40.68
40.64
40.60
40.56
−74.05
−74.00
−73.95
−73.90
−73.85
Longitude
Borough
BROOKLYN
Jersey City
New York
BROOKLYN
Prospect Park
Marine Park
Floyd Bennett Field
Jacob Riis Park
New York Aquarium
Map data ©2019 Google