NYC Open Data Analysis

DOB Permit Issuance

Created By:

Yogesh J. Sawant

yogeshja@buffalo.edu

NYC Open Data Analysis:-**DOB Permit Issuance**

Description:

Department of Building(DOB) is an organization that issues permits for constructions and demolition activities in New York. The construction industry must submit an application to DOB with details of the construction job they would like to complete. The primary types of application, aka job type, are: New Building, Demolition, and Alterations Type 1, 2, and 3. Each job type can have multiple work types, such as general construction, boiler, elevator, and plumbing. Each work type will receive a separate permit. (See the DOB Job Application Filings dataset for information about each job application.) Each row/record in this dataset represents the life cycle of one permit for one work type. The dataset is updated daily with new records, and each existing record will be updated as the permit application moves through the approval process to reflect the latest status of the application.

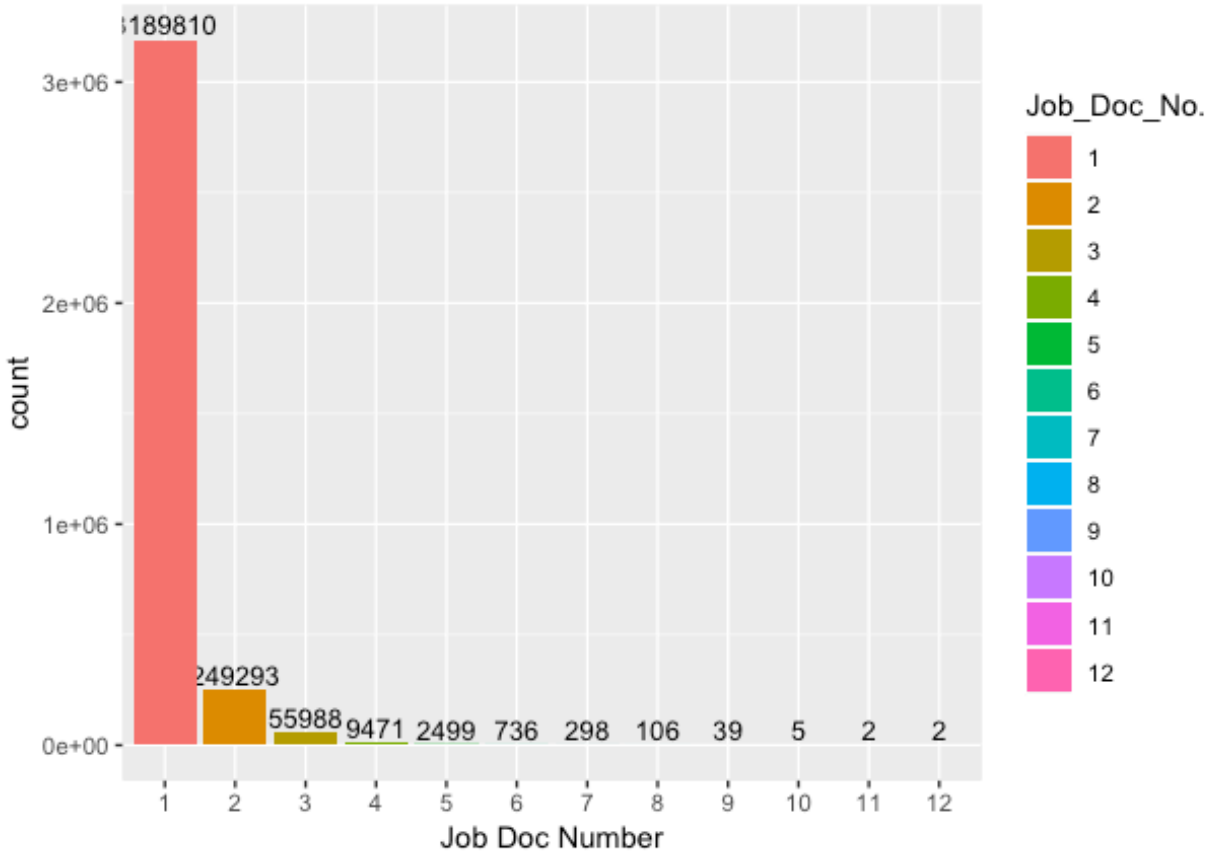
This Project Analyses the above mentioned data and the sequence and presents the investigations using visualization and statistics developed in R. There are numerous number of investigations which can be drawn from the data given by DOB.

- **Job Doc # v/s Count:**

  Code in R:

```
train_data <- read.csv("./DOB_Permit_Issuance.csv")

library(ggplot2)
library(dplyr)

        ###Job Doc #

jdn<-unique(train_data$Job.doc...)
t<-table(train_data$Job.doc...)
jdndf<-as.data.frame(t)
Job_Doc_No.<-jdndf$Var1
p<-ggplot(jdndf, aes(x=jdndf$Var1, y=jdndf$Freq,fill=Job_Doc_No.)) +
          geom_text(aes(label=jdndf$Freq), vjust=-0.3, size=3.5)+
          labs(x = "Job Doc Number",y ="count")+
          geom_bar(stat="identity")
print(p)
```

Output Plot:



The above Bar Plot displays the Job Doc # and count. According to the graph we can say the count of job Doc 1 was very high as compared to the other job Doc # ,it goes upto 3.1 Million filings all over new York.

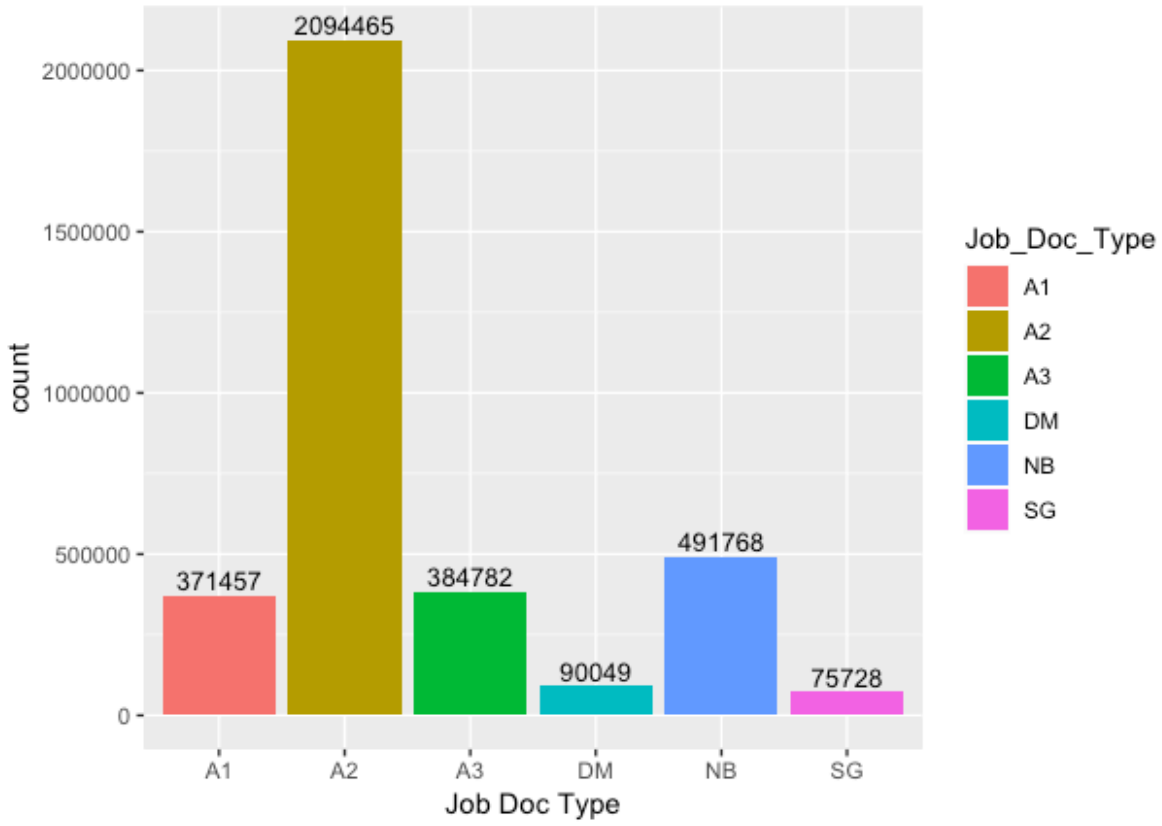
- **Job Doc Type v/s Count:**

  Code in R:

```
train_data <- read.csv("./DOB_Permit_Issuance.csv")

library(ggplot2)
library(dplyr)

        ###Job Doctype

        jdt<-unique(train_data$Job.Type)
        t1<-table(train_data$Job.Type)
```

```
        jdtdf<-as.data.frame(t1)
        Job_Doc_Type<-jdtdf$Var1
        p<-ggplot(jdtdf, aes(x=jdtdf$Var1, y=jdtdf$Freq,fill=Job_Doc_Type)) +
          geom_text(aes(label=jdtdf$Freq), vjust=-0.3, size=3.5)+
          labs(x = "Job Doc Type",y ="count")+
          geom_bar(stat="identity")
print(p)
```

Output Plot:



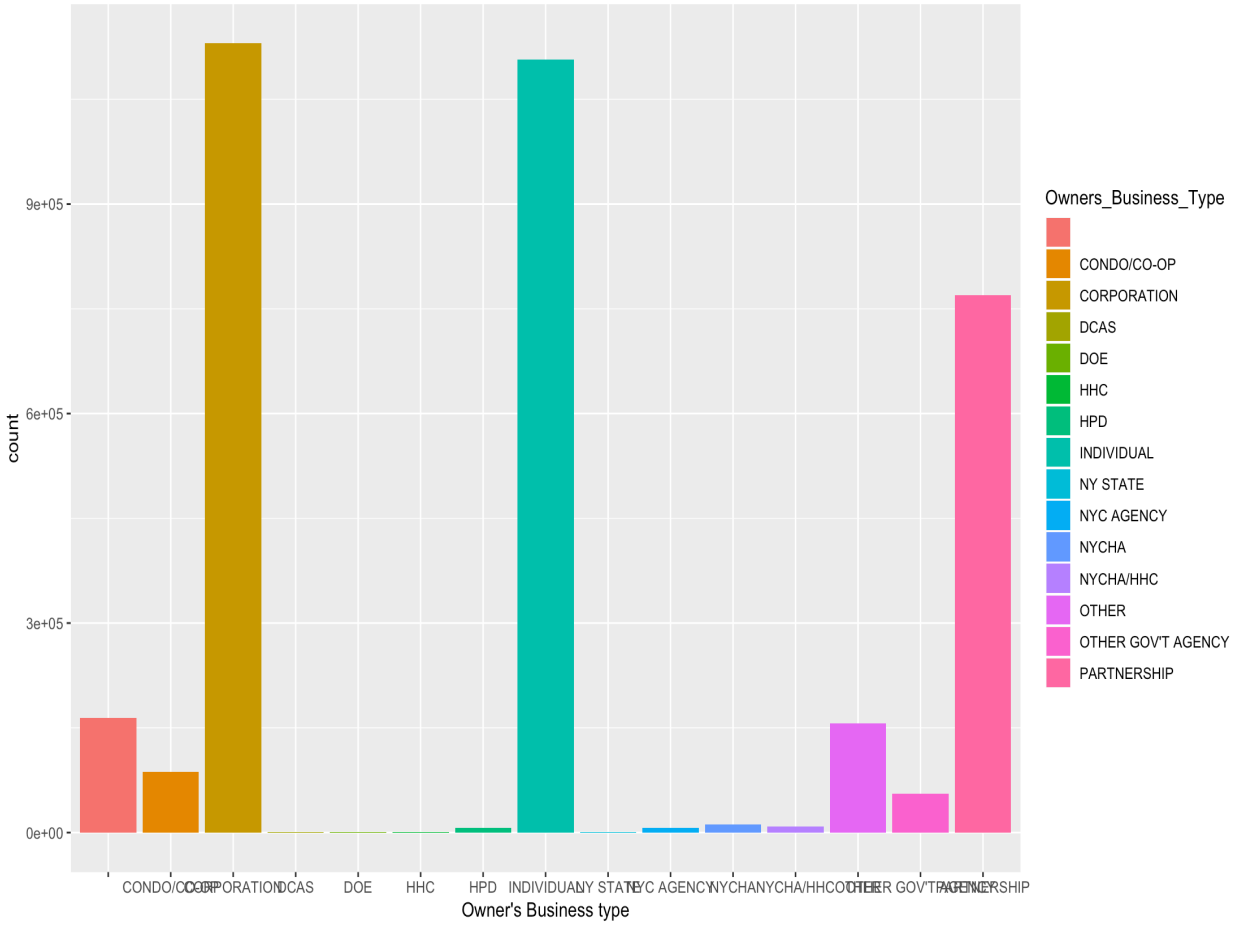
It is clearly evident from the above Bar Plot that the Alterations of the Type A2 the count is about 2.09 Million, so we can say that Type 2 alteration are very much prominent in NYC region, followed by A3 and A1 and so on.

- **Owner Business Type v/s Count:**

Code in R:

```
train_data <- read.csv("./DOB_Permit_Issuance.csv")
library(ggplot2)
library(dplyr)

        df<-data.frame(train_data$Owner.s.Business.Type)
        Owners_Business_Type<-df$train_data.Owner.s.Business.Type
        p<-ggplot(df,
        aes(x=df$train_data.Owner.s.Business.Type,fill=Owners_Business_Type)) +
          labs(x = "Owner's Business type",y ="count")+
          geom_histogram(stat="count")

print(p)
```

Output Plot:



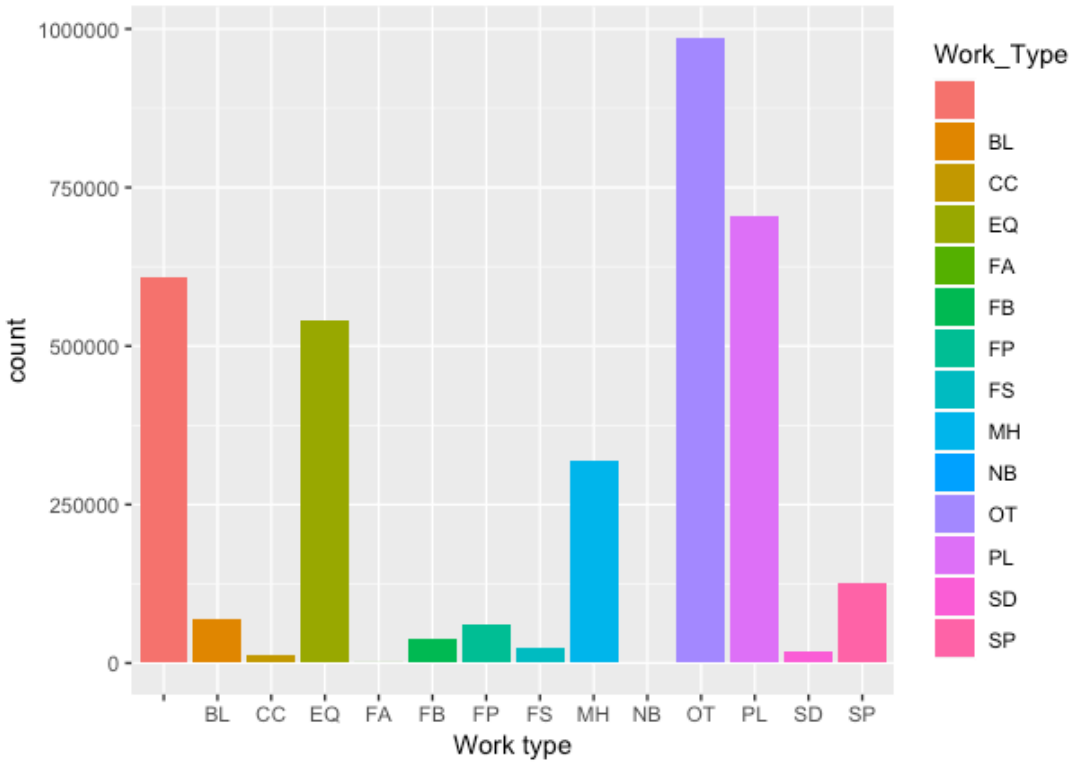
The above Histogram Plot helps us to understand what all types of Owner's businesses are involved in NYC for Development purposes .This graphs signifies that the Corporation Business type, Individual Business type and other government agency make large fillings for the development and construction at NYC.

- **Work Type v/s Count:**

Code in R:

```
train_data <- read.csv("./DOB_Permit_Issuance.csv")
library(ggplot2)
library(dplyr)

        x<-unique(train_data$Work.Type)
        df1<-data.frame(train_data$Work.Type)
        Work_Type<-df1$train_data.Work.Type
        p<-ggplot(df, aes(x=df1$train_data.Work.Type,fill=Work_Type)) +
          labs(x = "Work type",y ="count")+
          geom_histogram(stat="count")

print(p)
```

Output Plot:



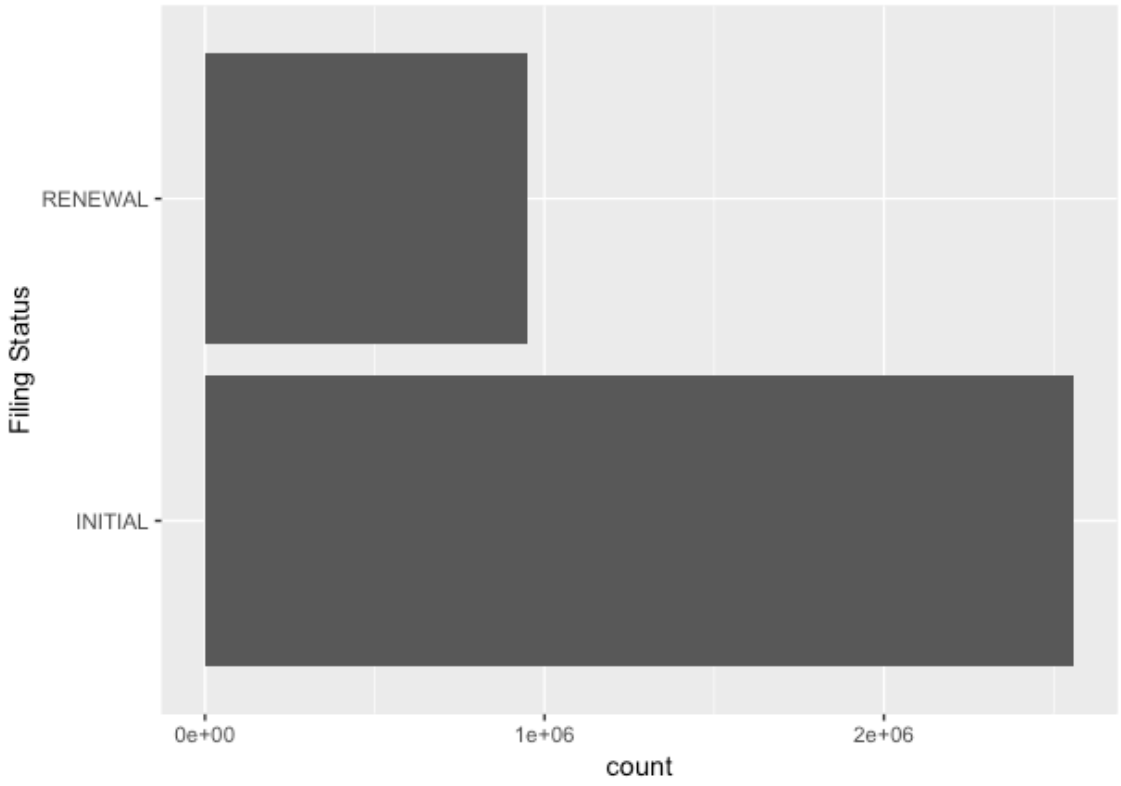
This Histogram Plot tells us about the type of works that are taking place in new york city and most of the place in new york city is used by Boiler and other construction work types.

- **Filing Status:**

Code in R:

```
train_data <- read.csv("./DOB_Permit_Issuance.csv")
library(ggplot2)
library(dplyr)
        fs<-unique(train_data$Filing.Status)
        dfs<-data.frame(train_data$Filing.Status)
        p<-ggplot(dfs, aes(x=dfs$train_data.Filing.Status)) +
          labs(x = "Filing Status",y ="count")+
          geom_histogram(stat="count")+
          coord_flip()

print(p)
```

Output Plot:



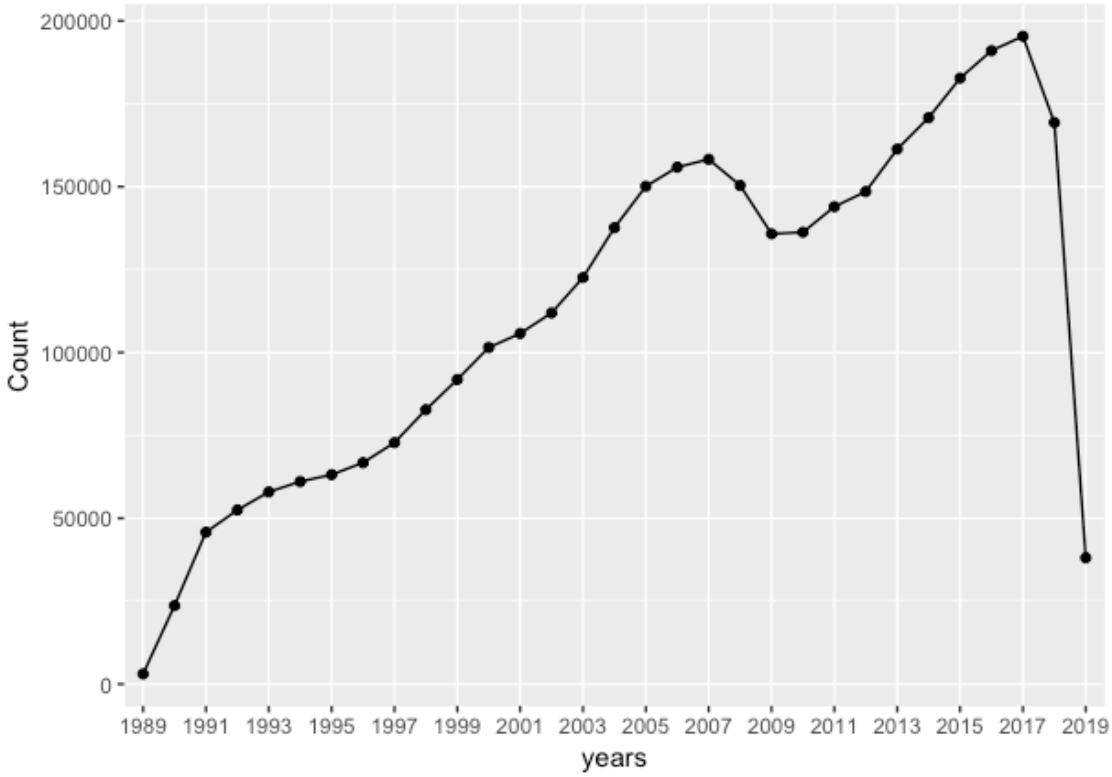
Plot makes it evident that the most of the filing is still in initial status. Which tells that new ventures are filing for places in NYC.

- **Issuance Count over the years:**

Code in R:

```
train_data <- read.csv("./DOB_Permit_Issuance.csv")
library(ggplot2)
library(dplyr)

x<-data.frame(train_data$Issuance.Date)
z<-data.frame(substr(x$train_data.Issuance.Date,7,10))
a<-data.frame(table(z))
a<-a[-1,]
p<-ggplot(a, aes(x=a$z, y=a$Freq)) +
geom_point(aes(group=31))+
geom_line(aes(group=1))+
scale_x_discrete(breaks =seq(1989,2019,2))+
#scale_x_discrete(labels=a$z)+
labs(x = "years",y ="Count")
print(p)
```

Output Line Plot:



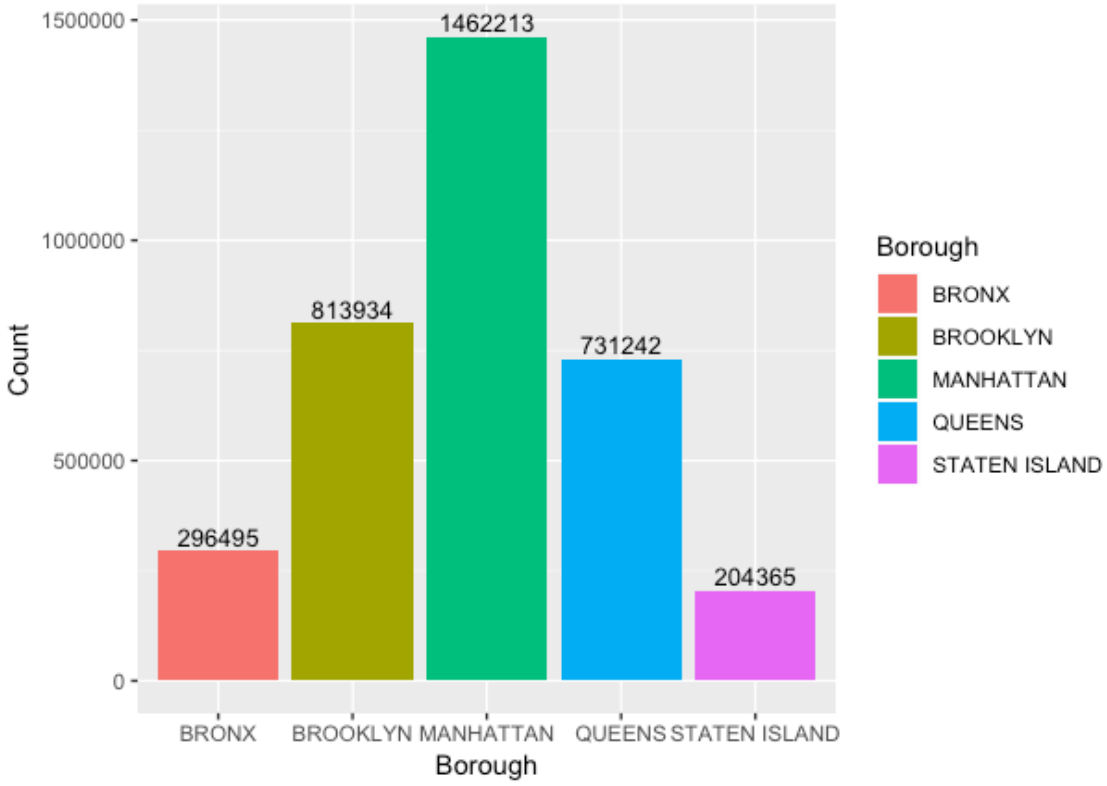
The Issuance count over the years had a gradual increase from 1989 and was highest in 2017,later dropped drastically in 2019. There can be numerous reasons for this like government changes, rules and regulation changes or lack of documents proof etc.

Locations and coverage in different borough in NYC:

1. Code in R:

```
train_data <- read.csv("./DOB_Permit_Issuance.csv")
library(ggplot2)
library(dplyr)
        x1<-unique(train_data$BOROUGH)
        t<-table(train_data$BOROUGH)
        df<-as.data.frame(t)
        Borough<-df$Var1
        p<-ggplot(df, aes(x=df$Var1, y=df$Freq,fill=Borough)) +
         geom_text(aes(label=df$Freq), vjust=-0.3, size=3.5)+
         labs(x = "Borough",y ="Count")+
         geom_bar(stat="identity")
print(p)
```

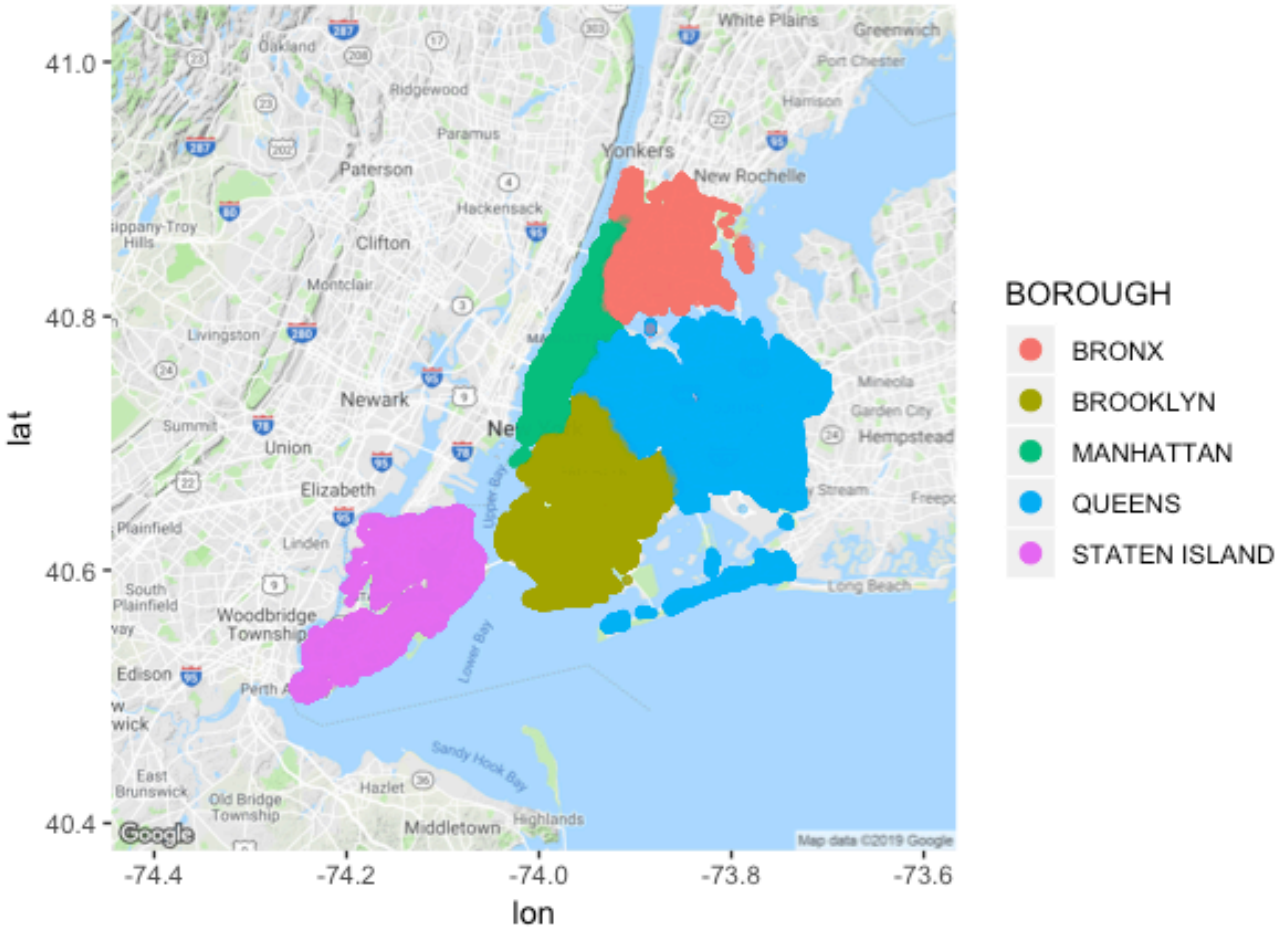
Output Plot:



Code in R:

```
train_data <- read.csv("/Users/demo/Documents/TOPOS/DOB_Permit_Issuance.csv")

library(ggplot2)
library(dplyr)

        LONGITUDE<-train_data$LONGITUDE
        LATITUDE<-train_data$LATITUDE
        data<-
        data.frame(train_data$BOROUGH,train_data$LONGITUDE,train_data$LATITU
        DE)

        if(!requireNamespace("devtools")) install.packages("devtools")
        devtools::install_github("dkahle/ggmap", ref = "tidyup")
        ggmap(get_googlemap())
        register_google(key = 'AIzaSyBaWvzN-f15VOLu5TmuIIyBZDBAi1Xi3Mc')
        geocode("New York City")
        register_google(key = 'AIzaSyBaWvzN-f15VOLu5TmuIIyBZDBAi1Xi3Mc')

        nycmap <- get_map(location="New York City",zoom = 10)
        nyc1 <- ggmap(nycmap)
        nyc1 <- nyc1 + geom_point(data=data, aes(data$train_data.LONGITUDE,
        data$train_data.LATITUDE, color = data$train_data.BOROUGH), alpha = .2,
        size = 1)
        nyc1 <- nyc1 + guides(color = guide_legend(title="BOROUGH",override.aes =
        list(alpha = 1, size = 3)))
print(nyc1)
```

Output Map:



In the above Plot 1, It is tells us about the count of filings borough wise in NYC, and it evident that Manhattan has large number of filings each year .

The Plot 2 is plotted by making use of the Google API(Geocoding API) and ggmap.
This graph helps us to understand the coverage of requests of permit all over new York.

The Queens are is has larger spread as compared to other borough.