Hmk4-Q3

Mingfeng Li

November 14, 2018

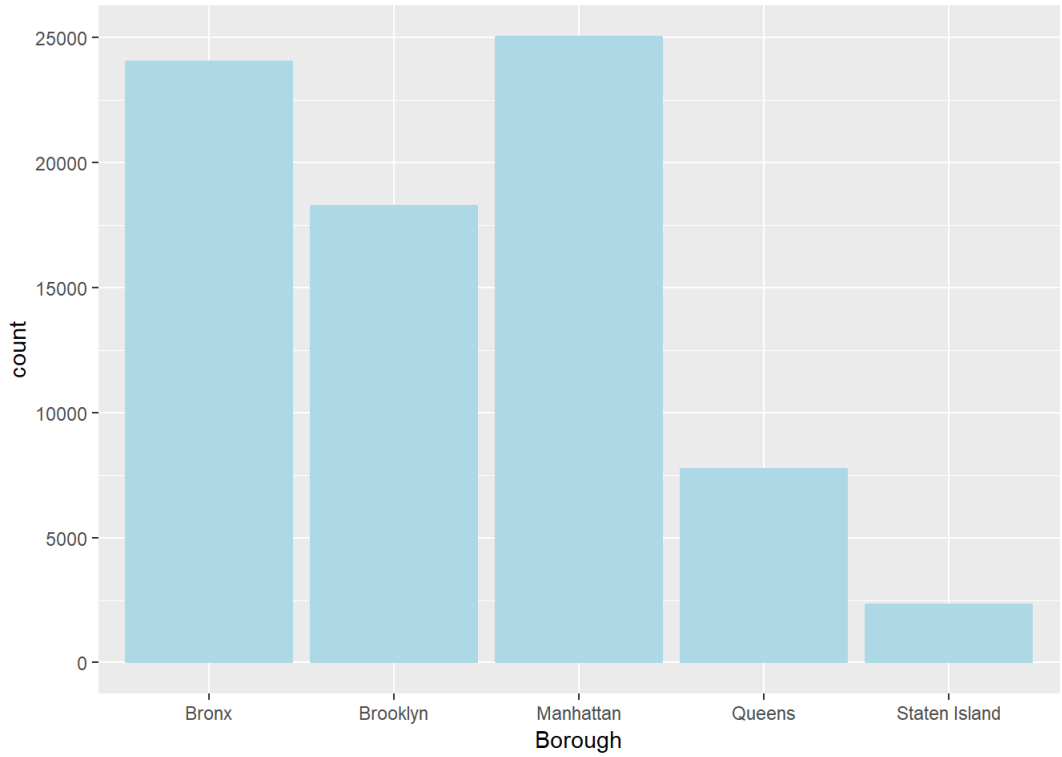
Question 3.

```
data<-read.csv("C:/Users/Mingfeng/OneDrive/Desktop/Columbia/Fall 2018/STAT 5702/Homework 4/Rodent_Inspection
_77K_Sample_Clean.csv",header=TRUE,na.strings = c("n/a","NA"))
library(tidyverse)
library(ggplot2)
```

The dataset of our project comes from NYC OpenData. The dataset is on Rodent Inspections in NYC from 2009 to 2015. For this homework, I picked the variables of BORO_CODE and BOROUGH and visualized in one histogram and two maps.

First of all, The amount of inspections by borough in NYC is checked. It can be noticed that most inspections happened in Manhattan and then followed by Bronx,Brooklyn,Queens and State Island. Borough is too large to generate enough information, hence zipcode was mapped instead in the next map.

```
ggplot(data, aes(data$BOROUGH))+
  geom_bar(fill='lightblue',color='lightblue')+
  xlab("Borough")
```

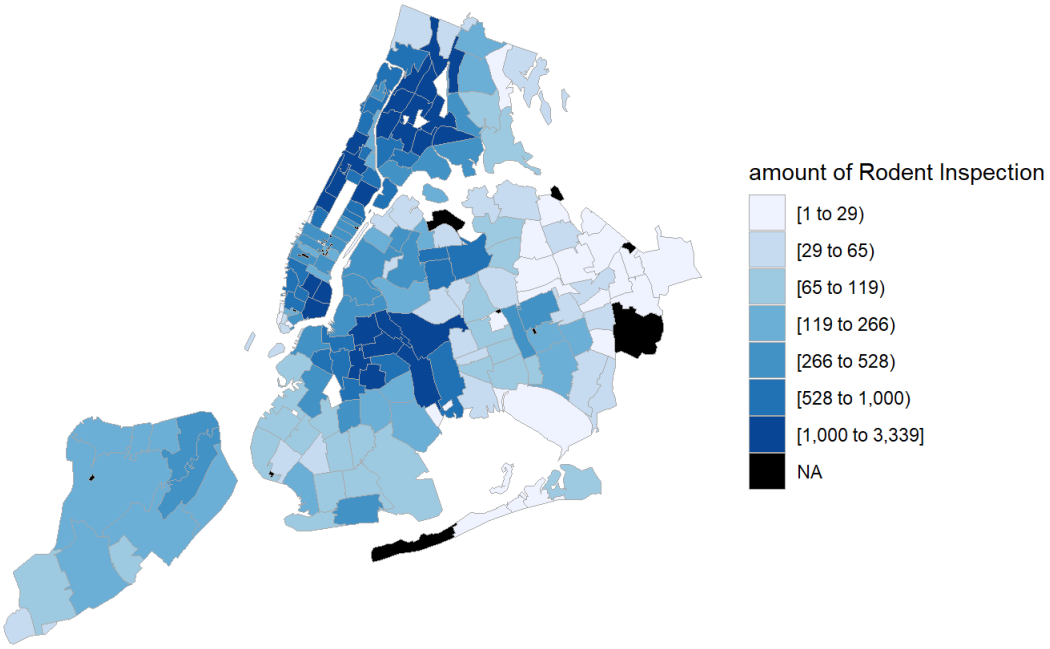


Shown in the map, the inspections amounts is higher in the lower east side of Bronx, upper side of Brooklyn, lower east side of Manhattan and the region around the Century Park. These regions are all highly residental areas hence intuitively the result is what we expected.

This brings us another question. If all the inspections have fixed the problem, and are there difference among the soving rates in each borough?

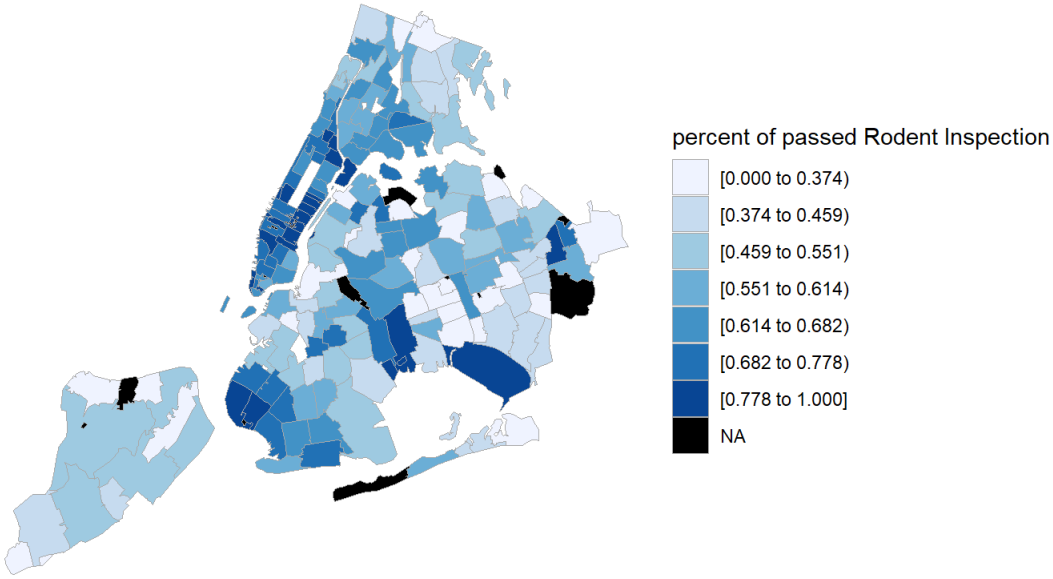
```
library("choroplethr")
library(devtools)
install_github('arilamstein/choroplethrZip@v1.5.0')
library(choroplethrZip)
data_dist<-data%>%group_by(ZIP_CODE)%>%
  summarize(value=sum(JOB_TICKET_OR_WORK_ORDER_ID!="NA"))
colnames(data_dist)<-c("region","value")
data_dist$region<-as.character(data_dist$region)
nyc_fips = c(36005, 36047, 36061, 36081, 36085)
zip_choropleth(data_dist,
               county_zoom = nyc_fips,
               title = "the amount of Rodent Insepection Records in New York City",
               legend="amount of Rodent Inspection")
```

the amount of Rodent Insepection Records in New York City




```
data_result<-data%>%group_by(ZIP_CODE)%>%
  summarize(value=sum(RESULT=="Passed Inspection")/sum(JOB_TICKET_OR_WORK_ORDER_ID!="NA"))
colnames(data_result)<-c("region","value")
data_result$region<-as.character(data_result$region)
nyc_fips = c(36005, 36047, 36061, 36081, 36085)
zip_choropleth(data_result,
               county_zoom = nyc_fips,
               title = "the Percent of passed inspection in New York City by zipcode",
               legend="percent of passed Rodent Inspection")
```

the Percent of passed inspection in New York City by zipcode



The map below shows the percent of passed inspection by zipcode. It can be observed that the two map highlight different regions. Manhattan still have a high percent of resolved cases while, the rate is low in Bronx despite the great amount of rodent inspections.

Data Source: https://data.cityofnewyork.us/Health/Rodent-Inspection/p937-wjvj