ap3650_nyc_crime_data_visualization

Anita

March 13, 2018

```
library(data.table)
library(vcdExtra)
library(extracat)
library(ggplot2)
library(dplyr)
library(tidyverse)
library(lubridate)

#fread("NYPD_Complaint_Data_Historic.csv",na.strings="",colClasses = c(PARKS_NM="c",HADEVELOPT="c"))->c
#fread("NYPD_Complaint_Data_Historic.csv",na.strings="")->crime_df
crime_df <- read.csv("NYPD_Complaint_Data_Historic.csv", header=TRUE)
```

Data Manipulations

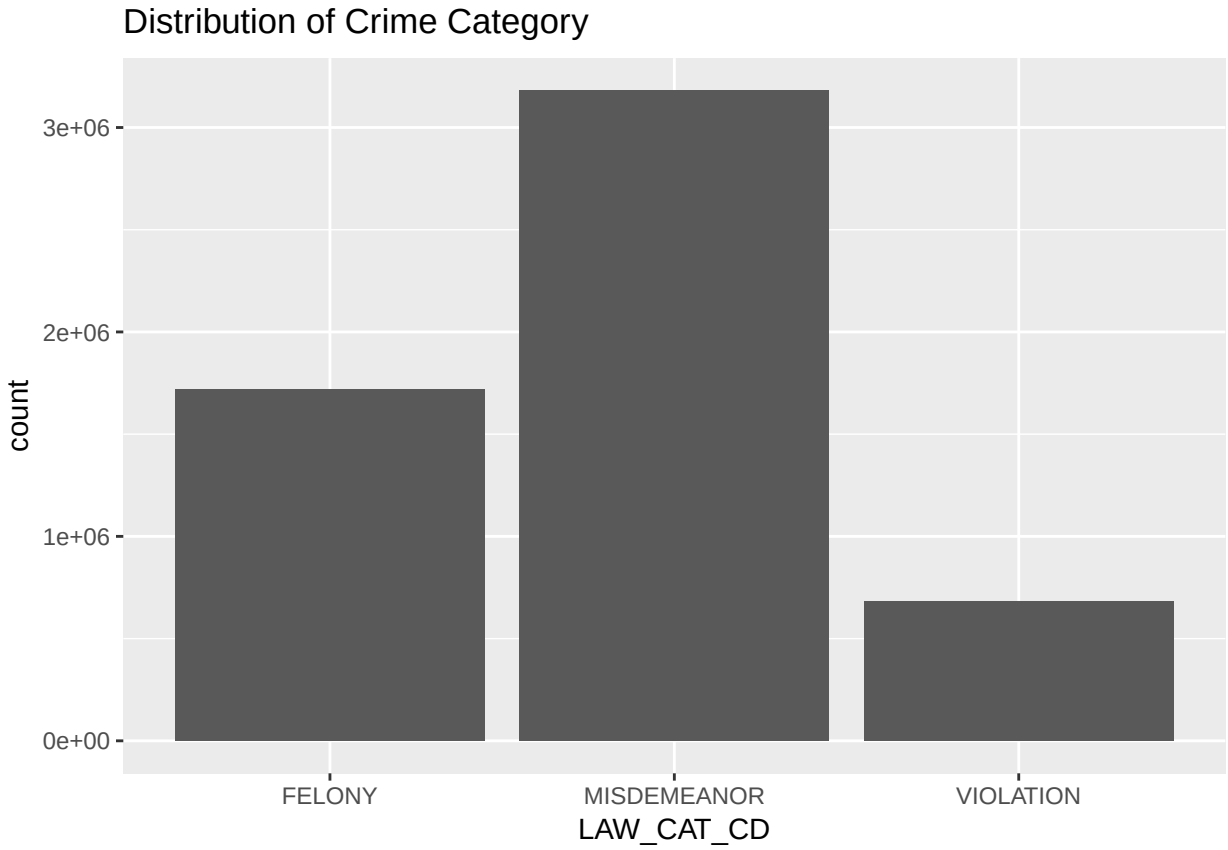
```
#Convert dates and times to correct format
crime_df$CMPLNT_FR_DT <- as.Date(crime_df$CMPLNT_FR_DT, format='%m/%d/%Y')
crime_df$CMPLNT_TO_DT <- as.Date(crime_df$CMPLNT_TO_DT, format='%m/%d/%Y')
crime_df$RPT_DT <- as.Date(crime_df$RPT_DT, format='%m/%d/%Y')

crime_df$CMPLNT_FR_TM <- as.POSIXct(crime_df$CMPLNT_FR_TM, format='%H:%M:%S')
crime_df$CMPLNT_TO_TM <- as.POSIXct(crime_df$CMPLNT_TO_TM, format='%H:%M:%S')
```

Plots

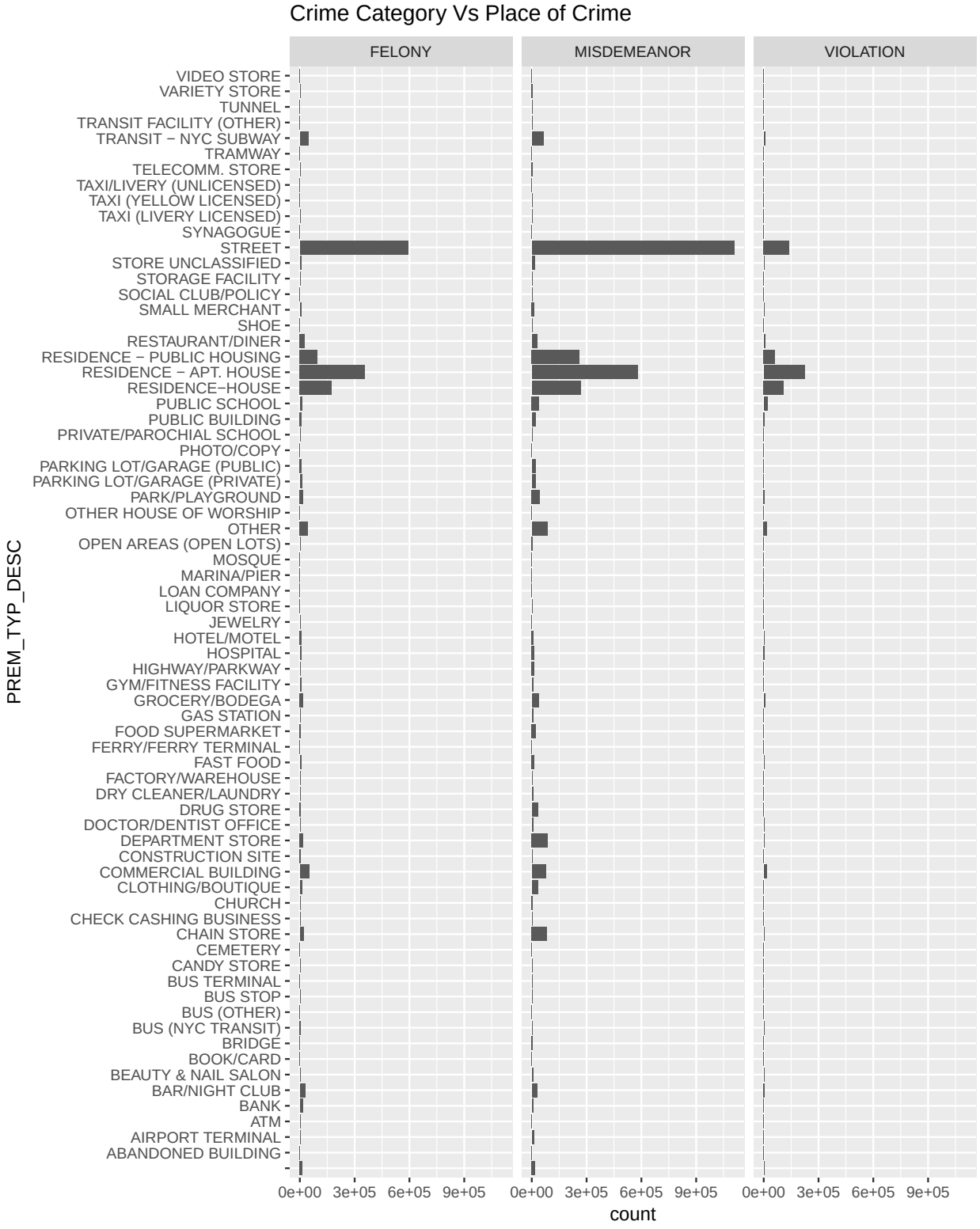
Warm-up Plot :-) Bar Chart

```
  ggplot(crime_df,aes(LAW_CAT_CD)) +
    geom_bar() +
    ggtitle("Distribution of Crime Category")
```



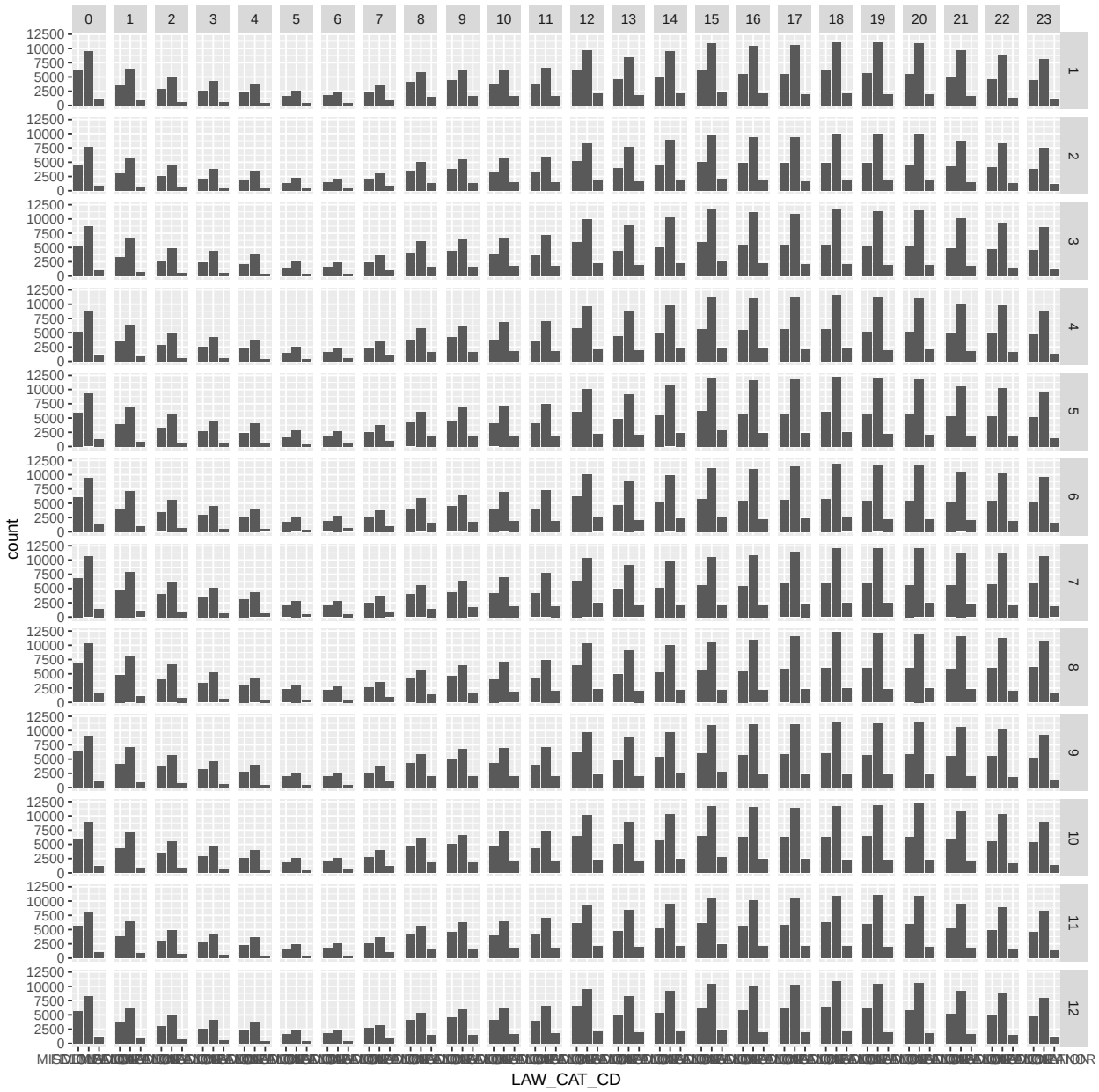
Type of Offense

```
ggplot(crime_df,aes(PREM_TYP_DESC)) +
  geom_bar() +
  facet_wrap(~LAW_CAT_CD) +
  coord_flip() +
  ggtitle("Crime Category Vs Place of Crime")
```


Crime Category Vs Place of Crime
FELONY
MISDEMEANOR
VIOLATION
PREM_TYP_DESC
VIDEO STORE
VARIETY STORE
TUNNEL
TRANSIT FACILITY (OTHER)
TRANSIT − NYC SUBWAY
TRAMWAY
TELECOMM. STORE
TAXI/LIVERY (UNLICENSED)
TAXI (YELLOW LICENSED)
TAXI (LIVERY LICENSED)
SYNAGOGUE
STREET
STORE UNCLASSIFIED
STORAGE FACILITY
SOCIAL CLUB/POLICY
SMALL MERCHANT
SHOE
RESTAURANT/DINER
RESIDENCE − PUBLIC HOUSING
RESIDENCE − APT. HOUSE
RESIDENCE−HOUSE
PUBLIC SCHOOL
PUBLIC BUILDING
PRIVATE/PAROCHIAL SCHOOL
PHOTO/COPY
PARKING LOT/GARAGE (PUBLIC)
PARKING LOT/GARAGE (PRIVATE)
PARK/PLAYGROUND
OTHER HOUSE OF WORSHIP
OTHER
OPEN AREAS (OPEN LOTS)
MOSQUE
MARINA/PIER
LOAN COMPANY
LIQUOR STORE
JEWELRY
HOTEL/MOTEL
HOSPITAL
HIGHWAY/PARKWAY
GYM/FITNESS FACILITY
GROCERY/BODEGA
GAS STATION
FOOD SUPERMARKET
FERRY/FERRY TERMINAL
FAST FOOD
FACTORY/WAREHOUSE
DRY CLEANER/LAUNDRY
DRUG STORE
DOCTOR/DENTIST OFFICE
DEPARTMENT STORE
CONSTRUCTION SITE
COMMERCIAL BUILDING
CLOTHING/BOUTIQUE
CHURCH
CHECK CASHING BUSINESS
CHAIN STORE
CEMETERY
CANDY STORE
BUS TERMINAL
BUS STOP
BUS (OTHER)
BUS (NYC TRANSIT)
BRIDGE
BOOK/CARD
BEAUTY & NAIL SALON
BAR/NIGHT CLUB
BANK
ATM
AIRPORT TERMINAL
ABANDONED BUILDING
0e+00
3e+05
6e+05
9e+05
count

Month and Time and Type of Crime

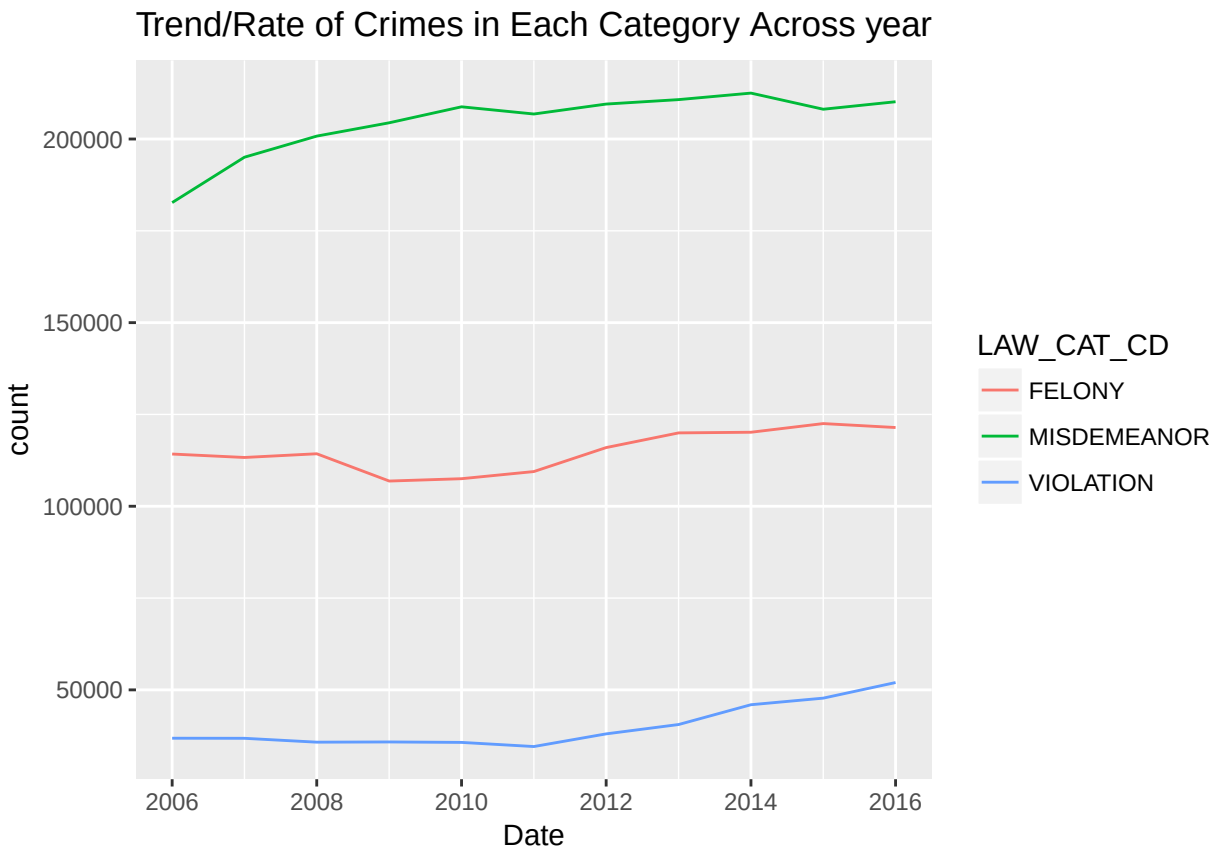
```
crime_df <- crime_df %>% drop_na()
ggplot(crime_df, aes(LAW_CAT_CD)) +
geom_bar() +
#facet_wrap(~month(CMPLNT_FR_DT))
#facet_wrap(~hour(CMPLNT_FR_TM))
facet_grid(month(CMPLNT_FR_DT)~hour(CMPLNT_FR_TM))
```



Time Series - Trend of Crime Rate

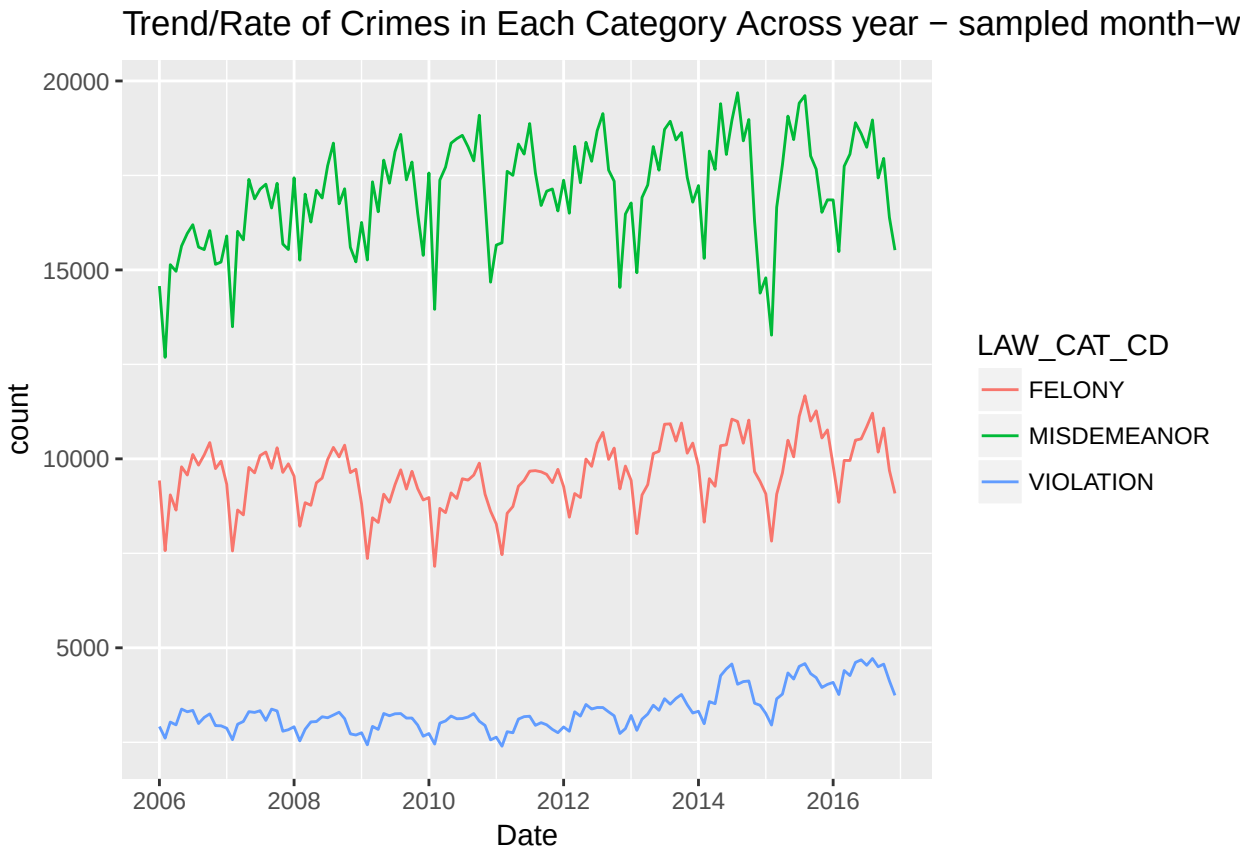
```
crime_time <- crime_df %>%
                filter(year(CMPLNT_FR_DT)>2005) %>%
                group_by(Date=floor_date(CMPLNT_FR_DT, "year"),LAW_CAT_CD) %>%
                summarize(count=n())

ggplot(crime_time, aes(Date,count, color=LAW_CAT_CD))+
      geom_line() +
      ggtitle("Trend/Rate of Crimes in Each Category Across year")
```



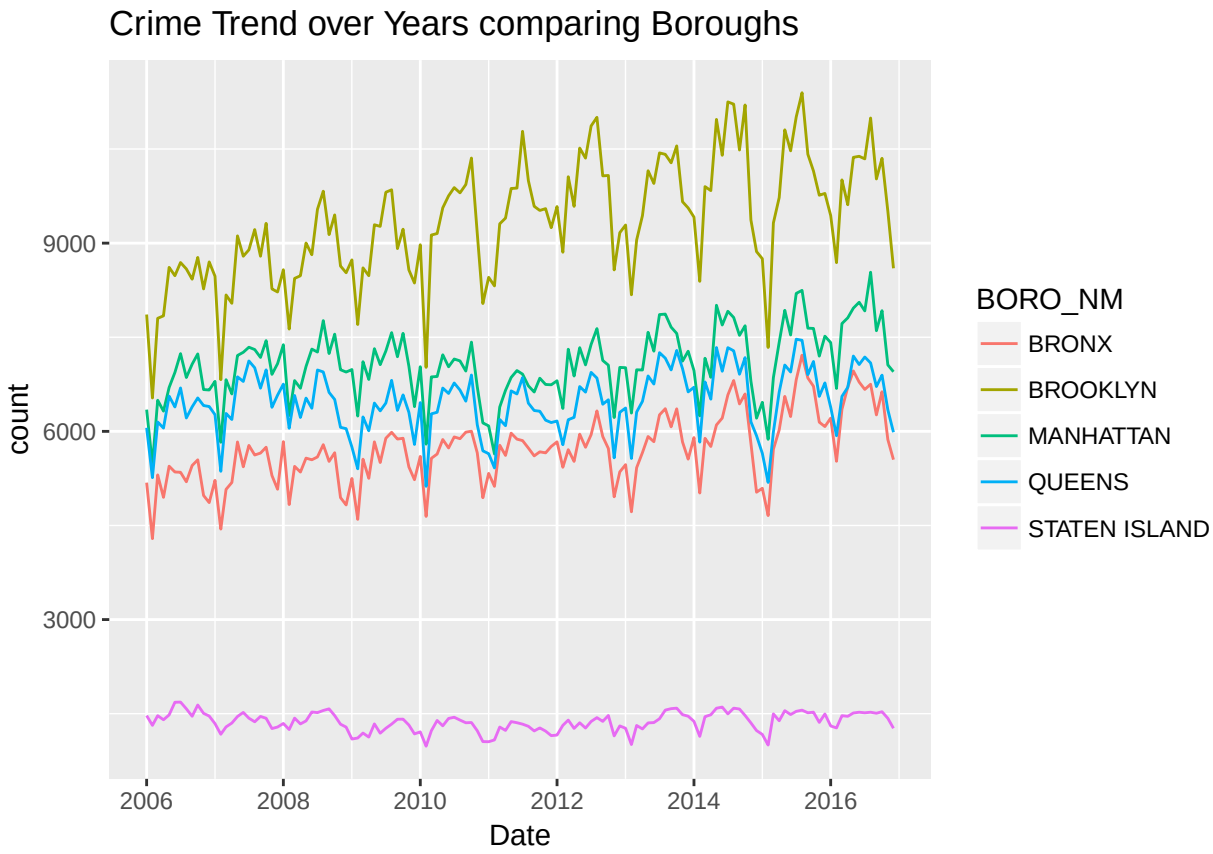

```
crime_time <- crime_df %>%
                filter(year(CMPLNT_FR_DT)>2005) %>%
                group_by(Date=floor_date(CMPLNT_FR_DT, "month"),LAW_CAT_CD) %>%
                summarize(count=n())

ggplot(crime_time, aes(Date,count, color=LAW_CAT_CD))+
      geom_line() +
      ggtitle("Trend/Rate of Crimes in Each Category Across year - sampled month-wise")
```




```
crime_boro <- crime_df %>%
                  filter(year(CMPLNT_FR_DT)> 2005 & BORO_NM != "") %>%
                  group_by(Date=floor_date(CMPLNT_FR_DT,"month"),BORO_NM) %>%
                  summarize(count=n())

ggplot(crime_boro,aes(Date,count,color=BORO_NM)) +
  geom_line() +
  ggtitle("Crime Trend over Years comparing Boroughs")
```

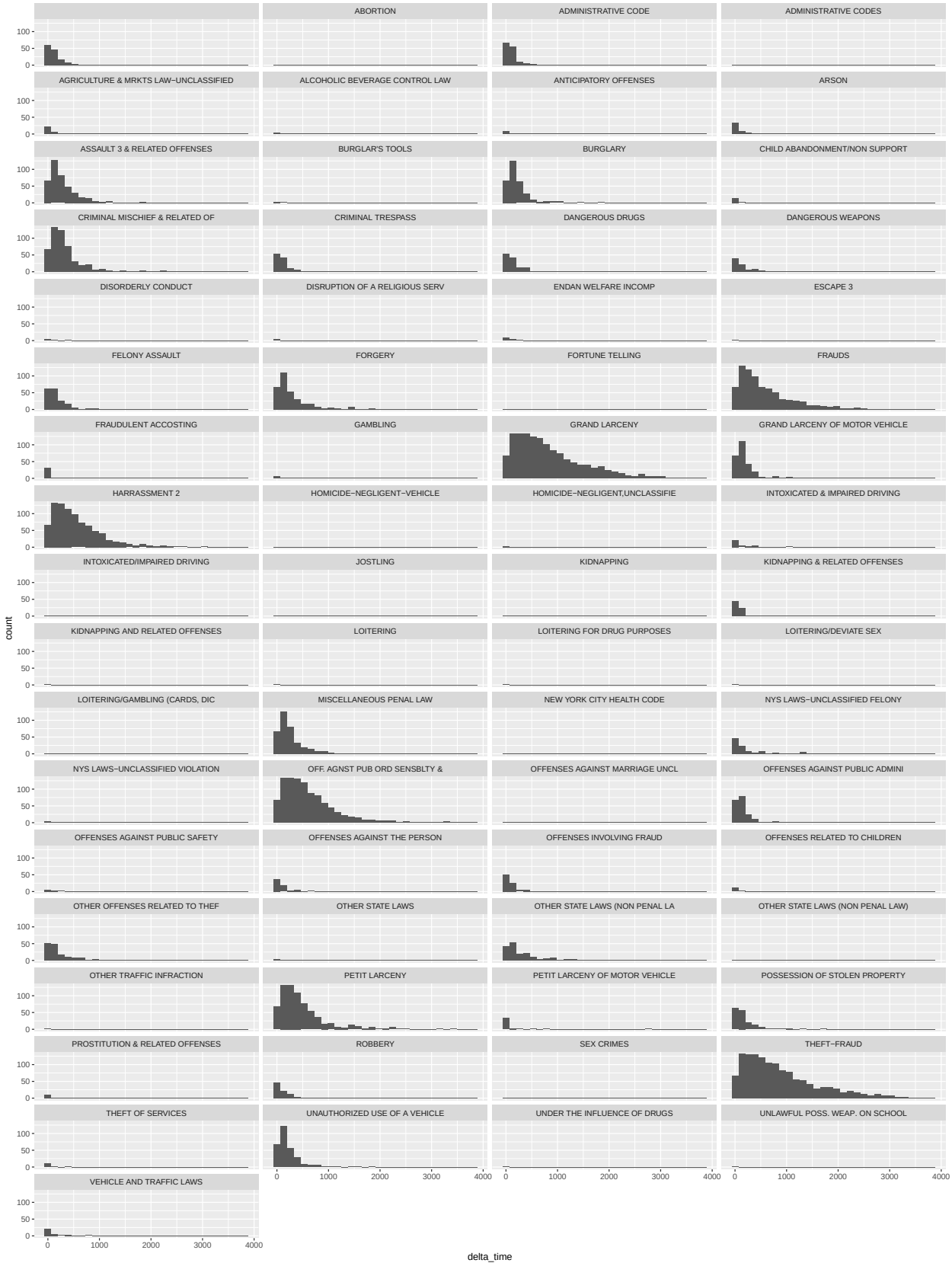


* Shows monthly pattern similar to Jingbo's * Year pattern fluctuates * Some NM_BORO are empty * Gaps between bororughs reduces towards later years

length of Crime Vs Type of Crime

```
crime_time <- crime_df %>%  drop_na() %>%
                  filter(year(CMPLNT_FR_DT)>2005) %>%
                  mutate(delta_time = as.numeric(CMPLNT_TO_DT - CMPLNT_FR_DT)) %>%
                  group_by(OFNS_DESC, delta_time) %>%
                  summarize(count=n())

ggplot(crime_time,aes(delta_time)) +
  geom_histogram(na.rm=TRUE) +
  facet_wrap(~OFNS_DESC, ncol = 4)
```

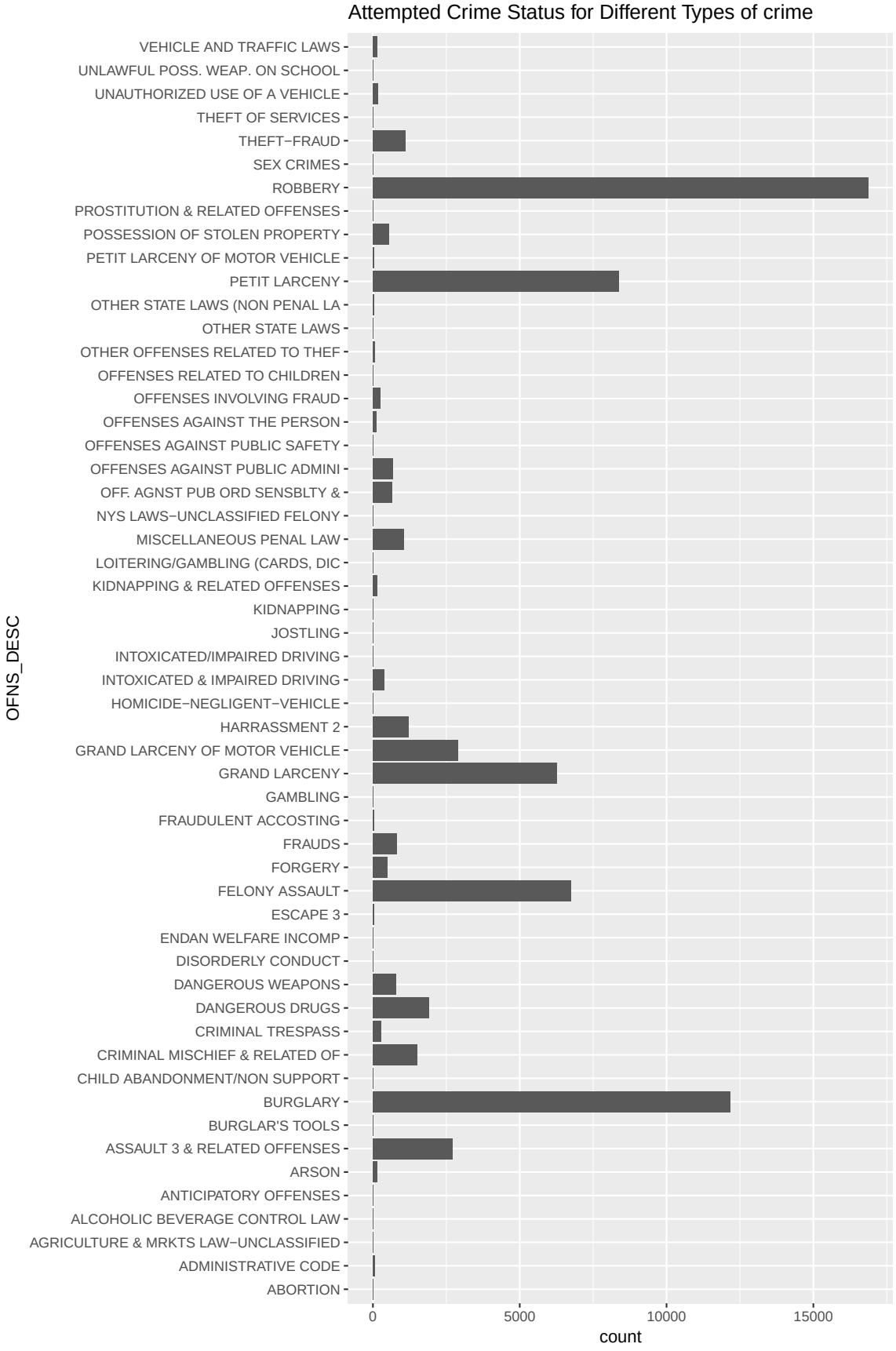


1. There are some cases where there might be typo on "To Date" especially year might be typo
2. Observed larceny(grand and petite have lot of cases)

3. There are blank OFfense category

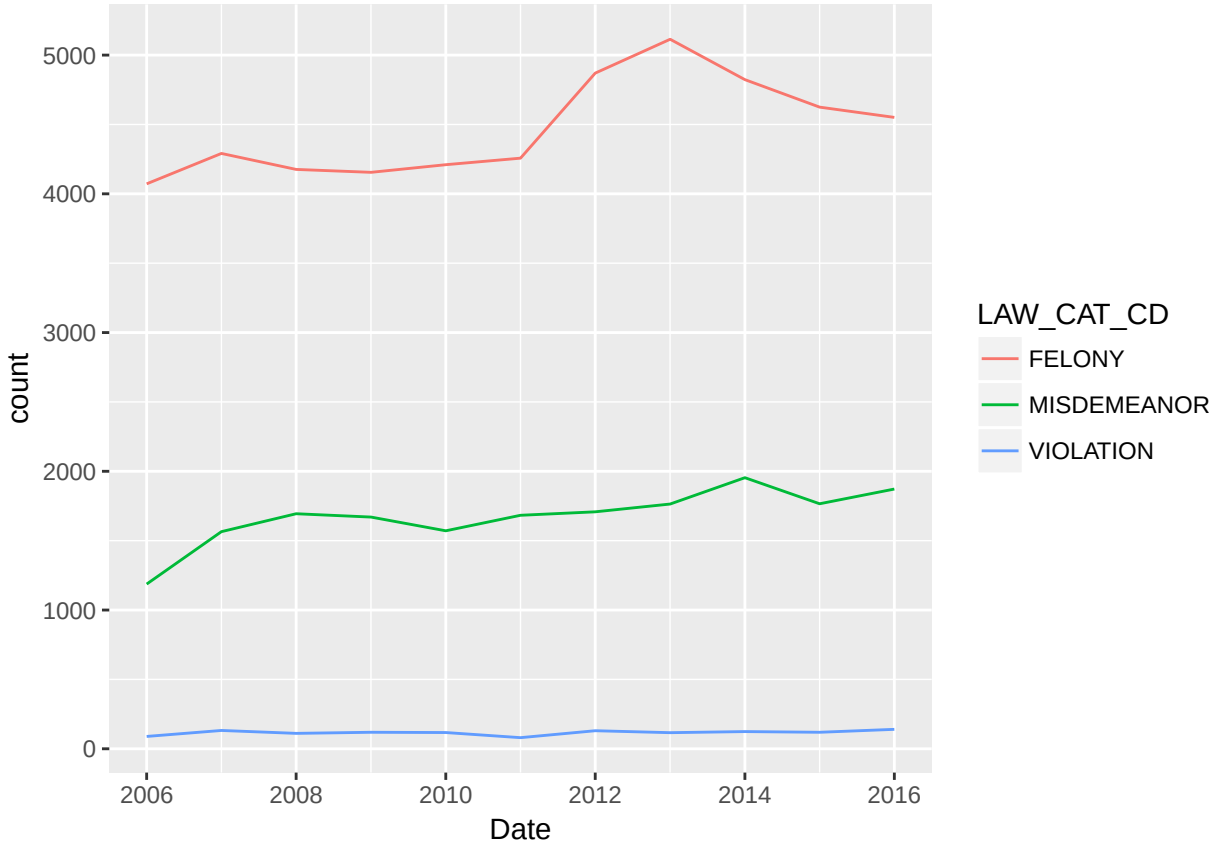
Atempted Crime vs Type of Crime

```
crime_stat <- crime_df %>%
                filter(CRM_ATPT_CPTD_CD == "ATTEMPTED" & OFNS_DESC !="") %>%
                group_by(OFNS_DESC) %>%
                summarize(count=n())
ggplot(crime_stat,aes(OFNS_DESC,count)) +
    geom_col() +
    coord_flip() +
    ggtitle("Attempted Crime Status for Different Types of crime")
```


Attempted Crime Status for Different Types of crime
OFNS_DESC
VEHICLE AND TRAFFIC LAWS
UNLAWFUL POSS. WEAP. ON SCHOOL
UNAUTHORIZED USE OF A VEHICLE
THEFT OF SERVICES
THEFT−FRAUD
SEX CRIMES
ROBBERY
PROSTITUTION & RELATED OFFENSES
POSSESSION OF STOLEN PROPERTY
PETIT LARCENY OF MOTOR VEHICLE
PETIT LARCENY
OTHER STATE LAWS (NON PENAL LA
OTHER STATE LAWS
OTHER OFFENSES RELATED TO THEF
OFFENSES RELATED TO CHILDREN
OFFENSES INVOLVING FRAUD
OFFENSES AGAINST THE PERSON
OFFENSES AGAINST PUBLIC SAFETY
OFFENSES AGAINST PUBLIC ADMINI
OFF. AGNST PUB ORD SENSBLTY &
NYS LAWS−UNCLASSIFIED FELONY
MISCELLANEOUS PENAL LAW
LOITERING/GAMBLING (CARDS, DIC
KIDNAPPING & RELATED OFFENSES
KIDNAPPING
JOSTLING
INTOXICATED/IMPAIRED DRIVING
INTOXICATED & IMPAIRED DRIVING
HOMICIDE−NEGLIGENT−VEHICLE
HARRASSMENT 2
GRAND LARCENY OF MOTOR VEHICLE
GRAND LARCENY
GAMBLING
FRAUDULENT ACCOSTING
FRAUDS
FORGERY
FELONY ASSAULT
ESCAPE 3
ENDAN WELFARE INCOMP
DISORDERLY CONDUCT
DANGEROUS WEAPONS
DANGEROUS DRUGS
CRIMINAL TRESPASS
CRIMINAL MISCHIEF & RELATED OF
CHILD ABANDONMENT/NON SUPPORT
BURGLARY
BURGLAR'S TOOLS
ASSAULT 3 & RELATED OFFENSES
ARSON
ANTICIPATORY OFFENSES
ALCOHOLIC BEVERAGE CONTROL LAW
AGRICULTURE & MRKTS LAW−UNCLASSIFIED
ADMINISTRATIVE CODE
ABORTION
0
5000
10000
15000
count

Attempted Crime Trend

```
crime_stat <- crime_df %>% drop_na() %>%
                filter(CRM_ATPT_CPTD_CD=="ATTEMPTED" & year(CMPLNT_FR_DT)>2005 & LAW_CAT_CD != "") %>%
                group_by(Date=floor_date(CMPLNT_FR_DT,"year"),LAW_CAT_CD) %>%
                summarize(count=n())
ggplot(crime_stat, aes(Date,count,color=LAW_CAT_CD)) +
  geom_line()
```



To find Top 10 Crime Categories, mosaic plots building blocks

```
crime_top <- crime_df %>% drop_na() %>%
                filter(OFNS_DESC!="") %>%
                group_by(OFNS_DESC) %>%
                summarize(count=n())

ggplot(crime_top, aes(reorder(OFNS_DESC,count), count)) +
  geom_col() +
  coord_flip() +
  ggtitle(" Top Crime Categories")
```

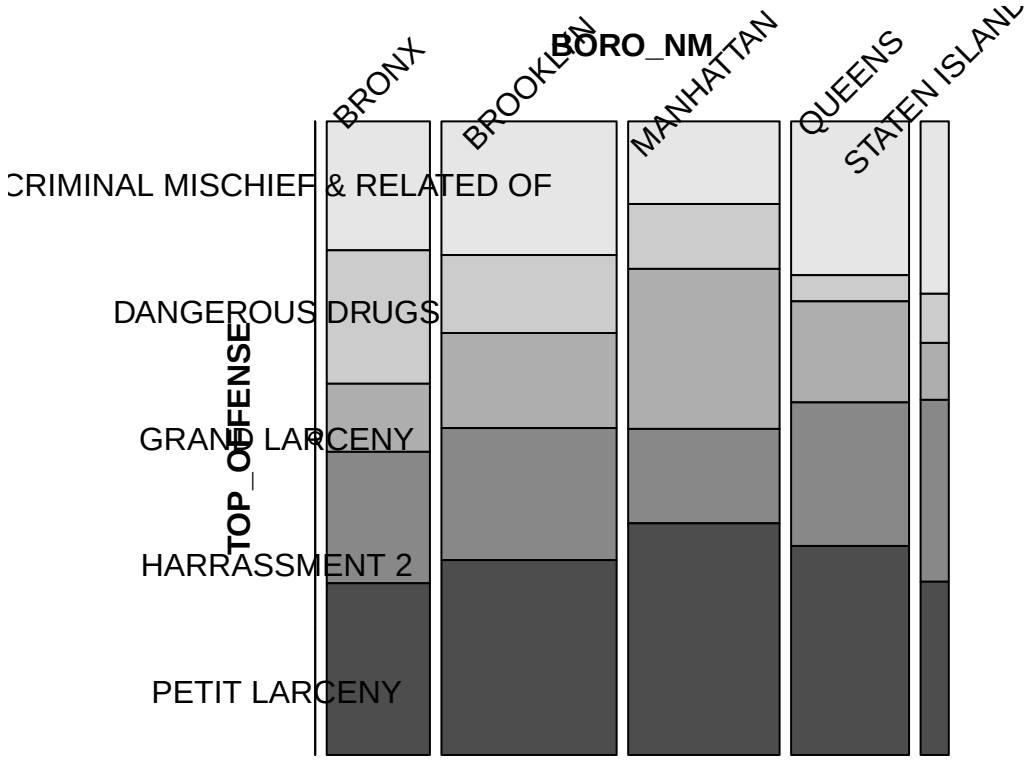

Top Crime Categories
reorder(OFNS_DESC, count)
PETIT LARCENY
HARRASSMENT 2
CRIMINAL MISCHIEF & RELATED OF
ASSAULT 3 & RELATED OFFENSES
GRAND LARCENY
DANGEROUS DRUGS
OFF. AGNST PUB ORD SENSBLTY &
BURGLARY
ROBBERY
FELONY ASSAULT
GRAND LARCENY OF MOTOR VEHICLE
MISCELLANEOUS PENAL LAW
DANGEROUS WEAPONS
OFFENSES AGAINST PUBLIC ADMINI
INTOXICATED & IMPAIRED DRIVING
CRIMINAL TRESPASS
FORGERY
VEHICLE AND TRAFFIC LAWS
THEFT−FRAUD
FRAUDS
POSSESSION OF STOLEN PROPERTY
OFFENSES INVOLVING FRAUD
UNAUTHORIZED USE OF A VEHICLE
OFFENSES AGAINST THE PERSON
ADMINISTRATIVE CODE
OTHER OFFENSES RELATED TO THEF
ARSON
OTHER STATE LAWS (NON PENAL LA
NYS LAWS−UNCLASSIFIED FELONY
BURGLAR'S TOOLS
FRAUDULENT ACCOSTING
THEFT OF SERVICES
KIDNAPPING & RELATED OFFENSES
GAMBLING
OFFENSES RELATED TO CHILDREN
PETIT LARCENY OF MOTOR VEHICLE
DISORDERLY CONDUCT
PROSTITUTION & RELATED OFFENSES
OFFENSES AGAINST PUBLIC SAFETY
ALCOHOLIC BEVERAGE CONTROL LAW
CHILD ABANDONMENT/NON SUPPORT
JOSTLING
OTHER STATE LAWS
AGRICULTURE & MRKTS LAW−UNCLASSIFIED
LOITERING/GAMBLING (CARDS, DIC
ENDAN WELFARE INCOMP
ESCAPE 3
ANTICIPATORY OFFENSES
SEX CRIMES
NEW YORK CITY HEALTH CODE
HOMICIDE−NEGLIGENT−VEHICLE
UNLAWFUL POSS. WEAP. ON SCHOOL
DISRUPTION OF A RELIGIOUS SERV
INTOXICATED/IMPAIRED DRIVING
HOMICIDE−NEGLIGENT,UNCLASSIFIE
NYS LAWS−UNCLASSIFIED VIOLATION
LOITERING FOR DRUG PURPOSES
ADMINISTRATIVE CODES
LOITERING/DEVIATE SEX
LOITERING
KIDNAPPING
FORTUNE TELLING
OTHER STATE LAWS (NON PENAL LAW)
OTHER TRAFFIC INFRACTION
ABORTION
UNDER THE INFLUENCE OF DRUGS
OFFENSES AGAINST MARRIAGE UNCL
KIDNAPPING AND RELATED OFFENSES
0e+00
2e+05
4e+05
6e+05
count

Boro, Juris, Crime Categories

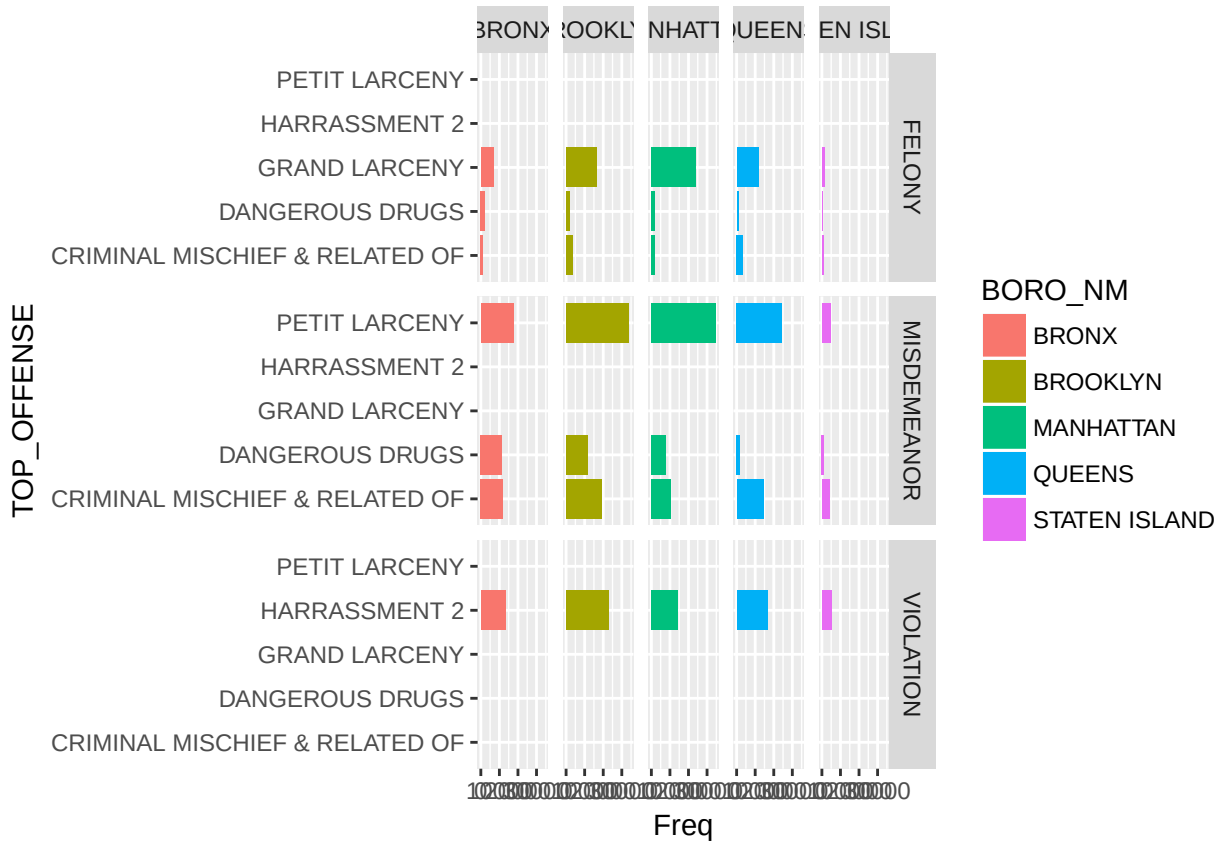
```
                #group_by(BORO_NM,JURIS_DESC,OFNS_DESC) %>%
                #mutate(count=n()) %>%
top_ofns   <- c("PETIT LARCENY", "HARRASSMENT 2", "CRIMINAL MISCHIEF & RELATED OF", "ASSAUALT 3 & REL
crime_sort <- crime_df %>% drop_na() %>%
                filter(BORO_NM != "", JURIS_DESC !="", (OFNS_DESC == top_ofns))%>%
                group_by(BORO_NM,LAW_CAT_CD,OFNS_DESC) %>%
                summarize(Freq=n())

crime_sort$TOP_OFFENSE = crime_sort$OFNS_DESC[,drop=TRUE]

#doubledecker(OFNS_DESC~BORO_NM, data=crime_sort, gp = gpar(fill = c("grey90", "red")))
mosaic(TOP_OFFENSE~BORO_NM, direction=c("v"), labeling=labeling_border(rot_labels=c(45,0,0, 0)), crim
```

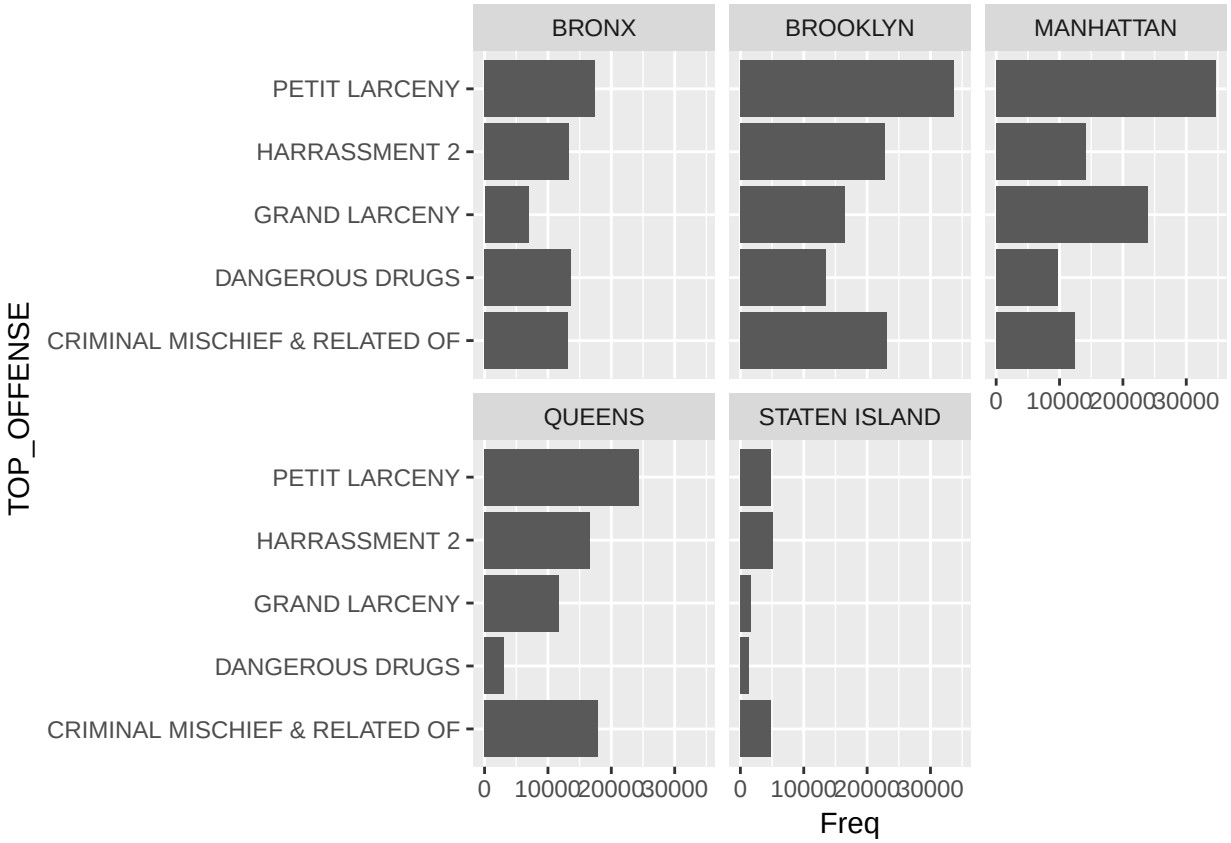



```
#doubledecker(TOP_OFFENSE~BORO_NM, data=crime_sort)
ggplot(crime_sort, aes(TOP_OFFENSE,Freq, fill=BORO_NM)) +
  geom_col() +
  facet_grid(LAW_CAT_CD~ BORO_NM) +
  coord_flip()
```



** The above plot shows something surprising, the categories are not standard, need to research more. For example, dangerous drugs is under Felony as well as Misdemeanor!! **

```
ggplot(crime_sort, aes(TOP_OFFENSE,Freq)) +
        geom_col() +
        facet_wrap(~BORO_NM) +
        coord_flip()
```

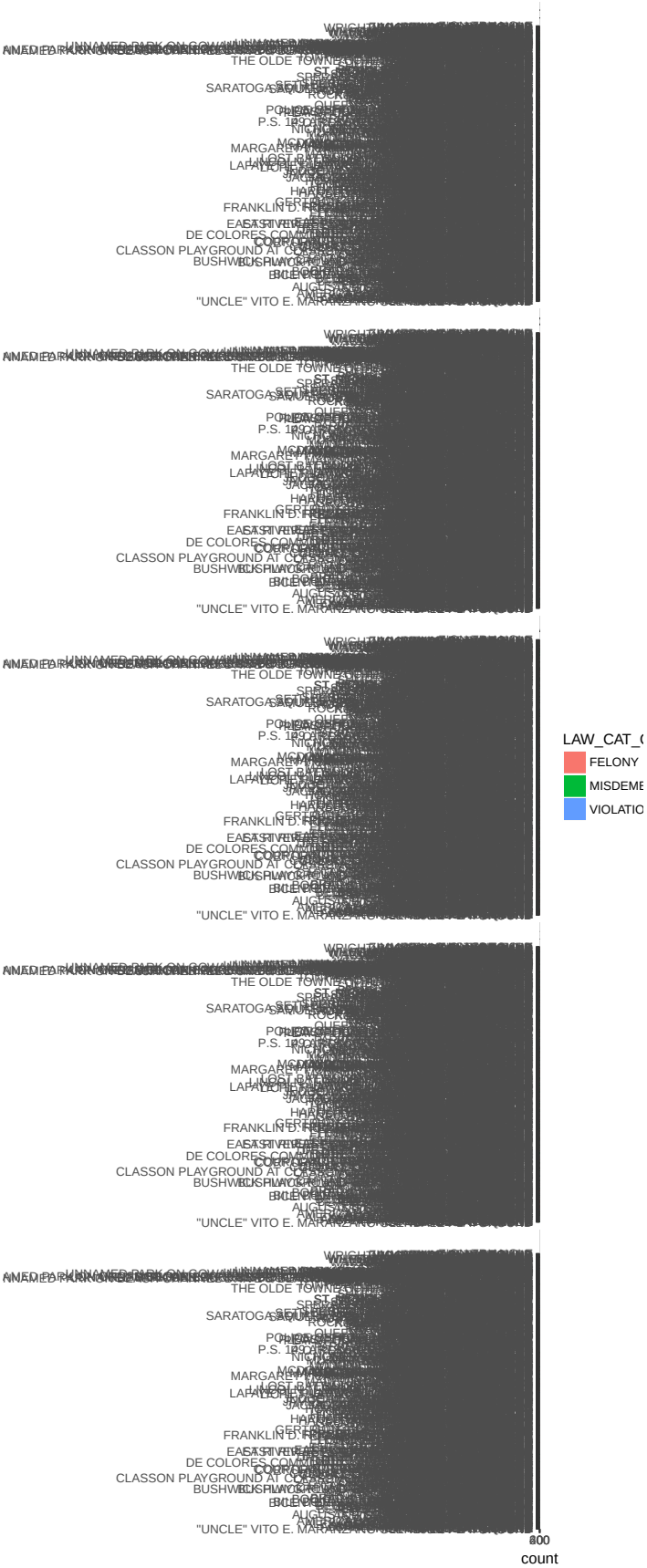


- I tired indivial Crime Types, the colors were too confusing as lot of categories

```
crime_parks <- crime_df %>% drop_na() %>%
                filter(BORO_NM!="",PARKS_NM!="",LAW_CAT_CD!="") %>%
                group_by(BORO_NM,PARKS_NM,LAW_CAT_CD) %>%
                summarize(count=n())

crime_parks <- crime_parks %>%
                arrange(desc(count))

ggplot(crime_parks,aes(PARKS_NM,count, fill=LAW_CAT_CD)) +
  geom_col() +
  facet_wrap(~BORO_NM, ncol=1) +
  coord_flip()
```


LAW_CAT_C
FELONY
MISDEME
VIOLATIO
800
count

trial on ggmap

```
library(ggmap)
```

Note that the `echo = FALSE` parameter was added to the code chunk to prevent printing of the R code that generated the plot.