ap3650_nyc_crime_data_visualization

Anita

March 13, 2018

```r
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

```r
#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')
```
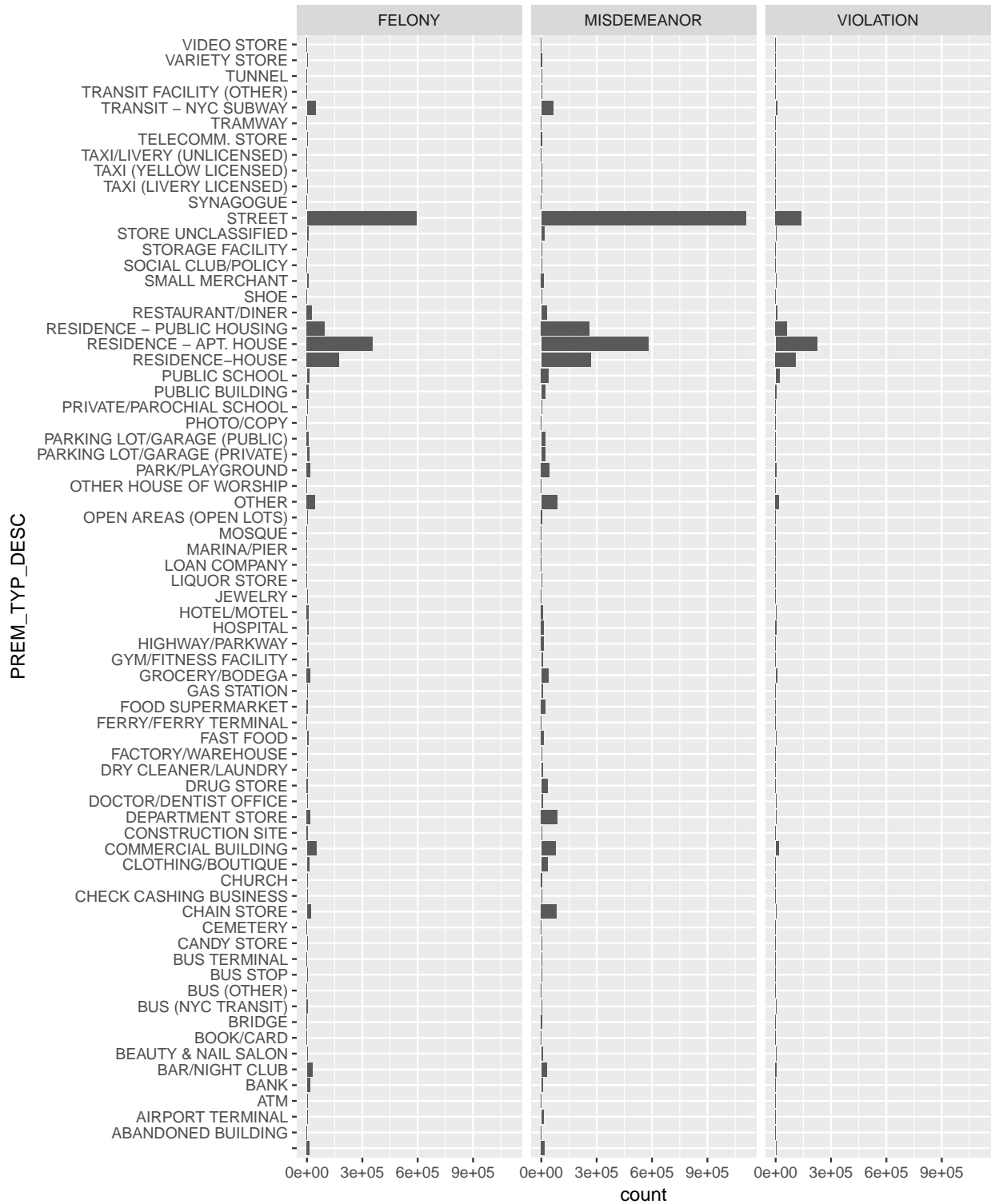
Plots

Warm-up Plot :-) Bar Chart

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

## Distribution of Crime Category
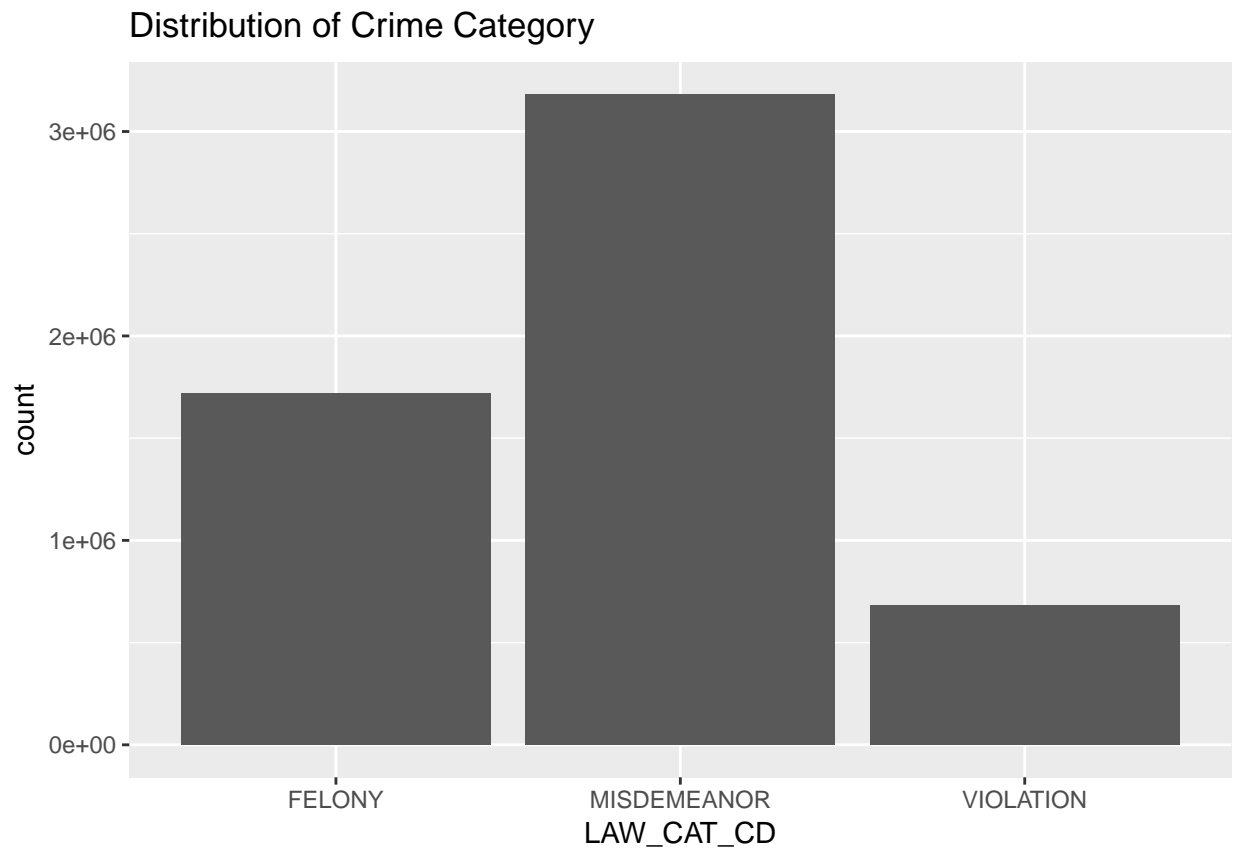
Type of Offense

```r
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

Month and Time and Type of Crime

```r
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))
```
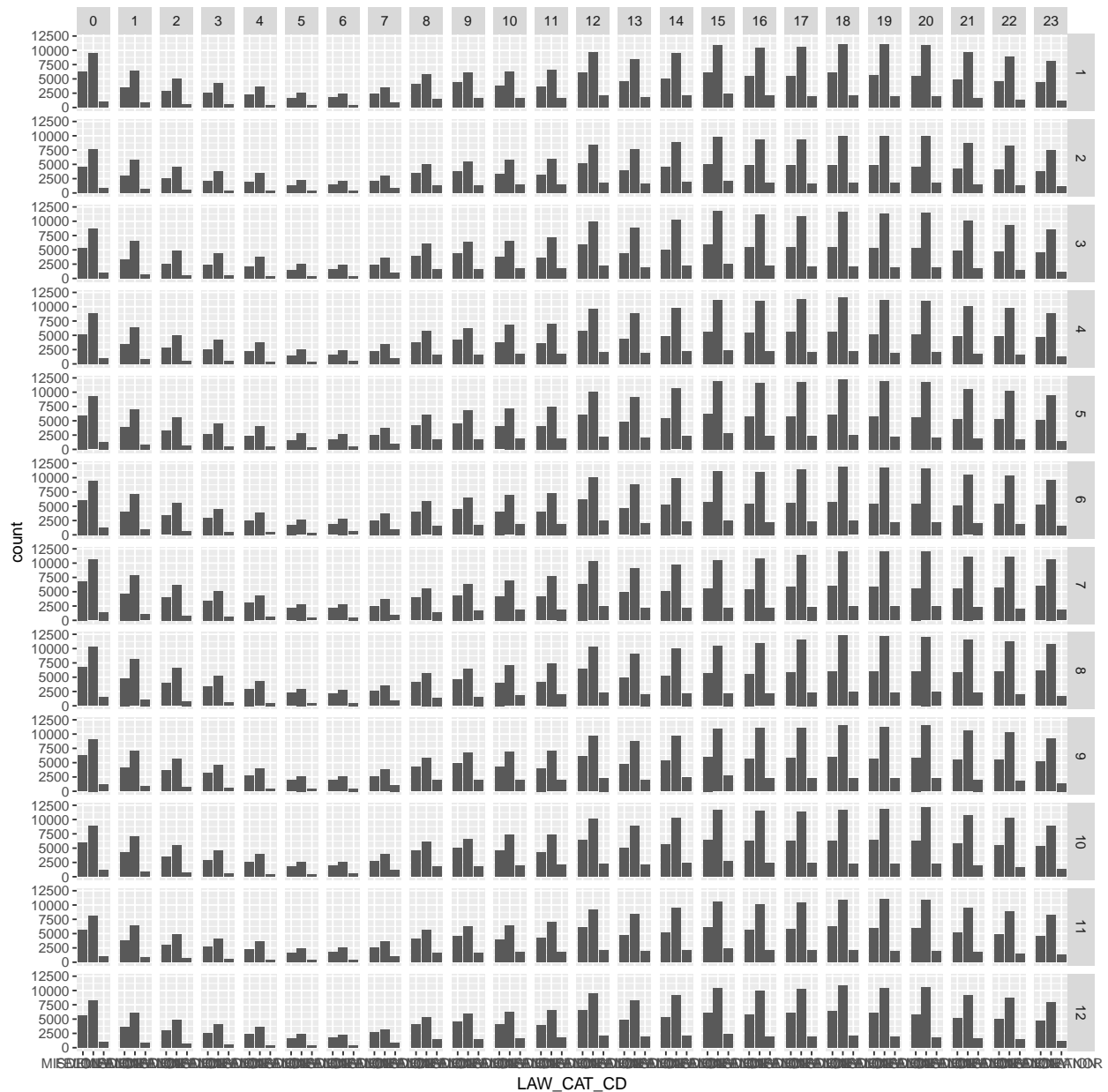

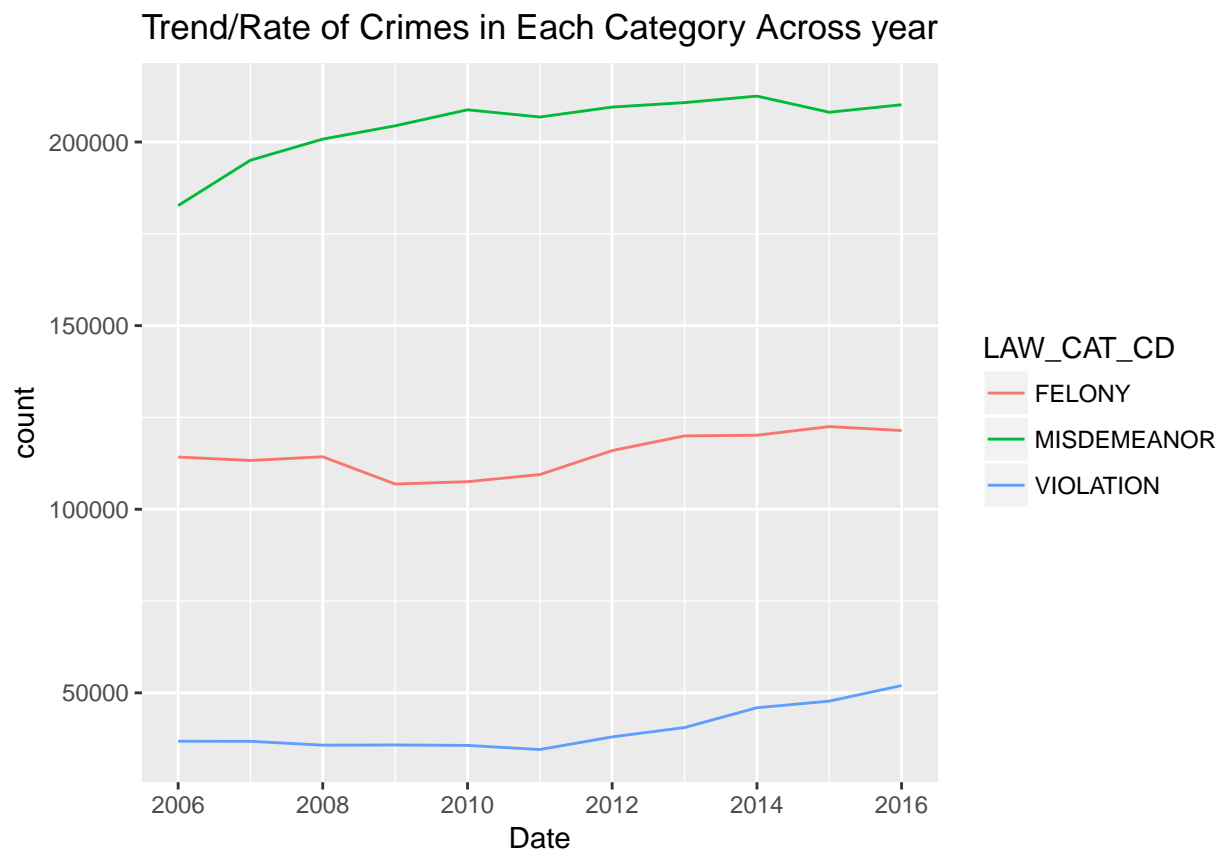
Time Series - Trend of Crime Rate

```
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")
```


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")
```

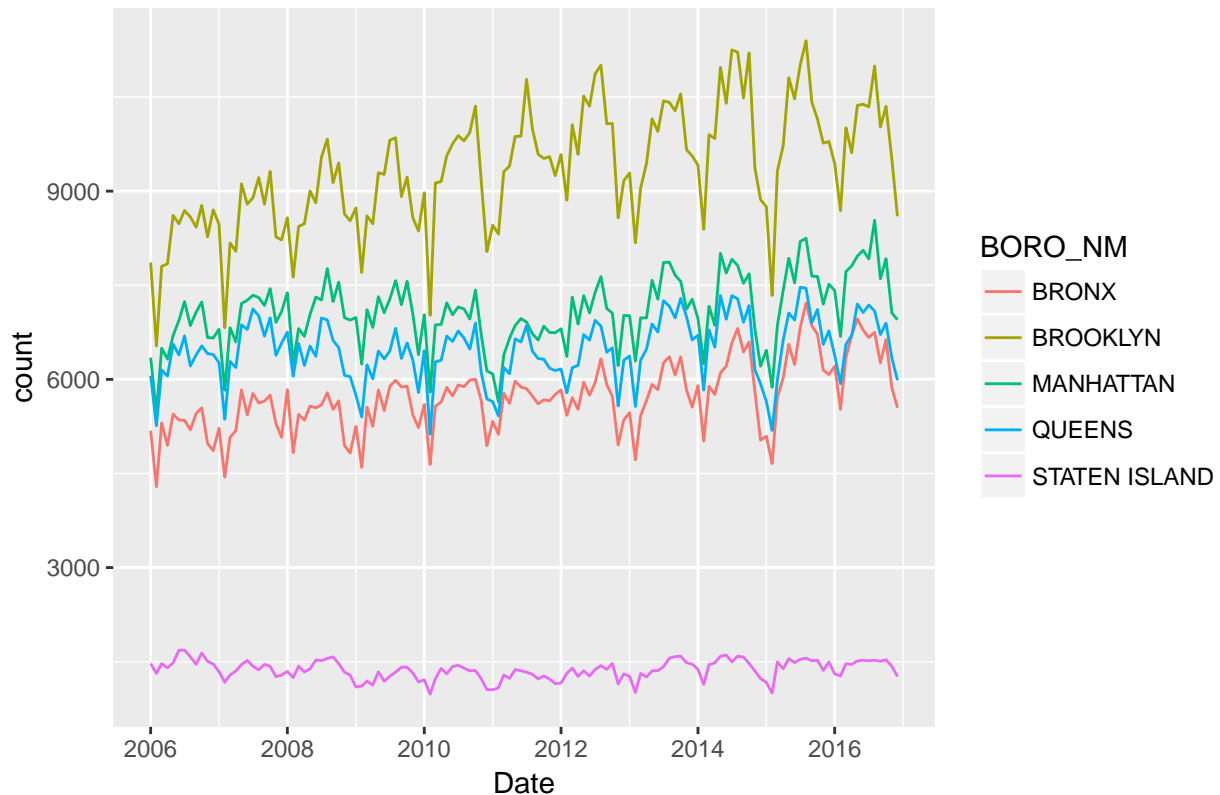## Trend/Rate of Crimes in Each Category Across year – sampled month–w


```r
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")
```

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

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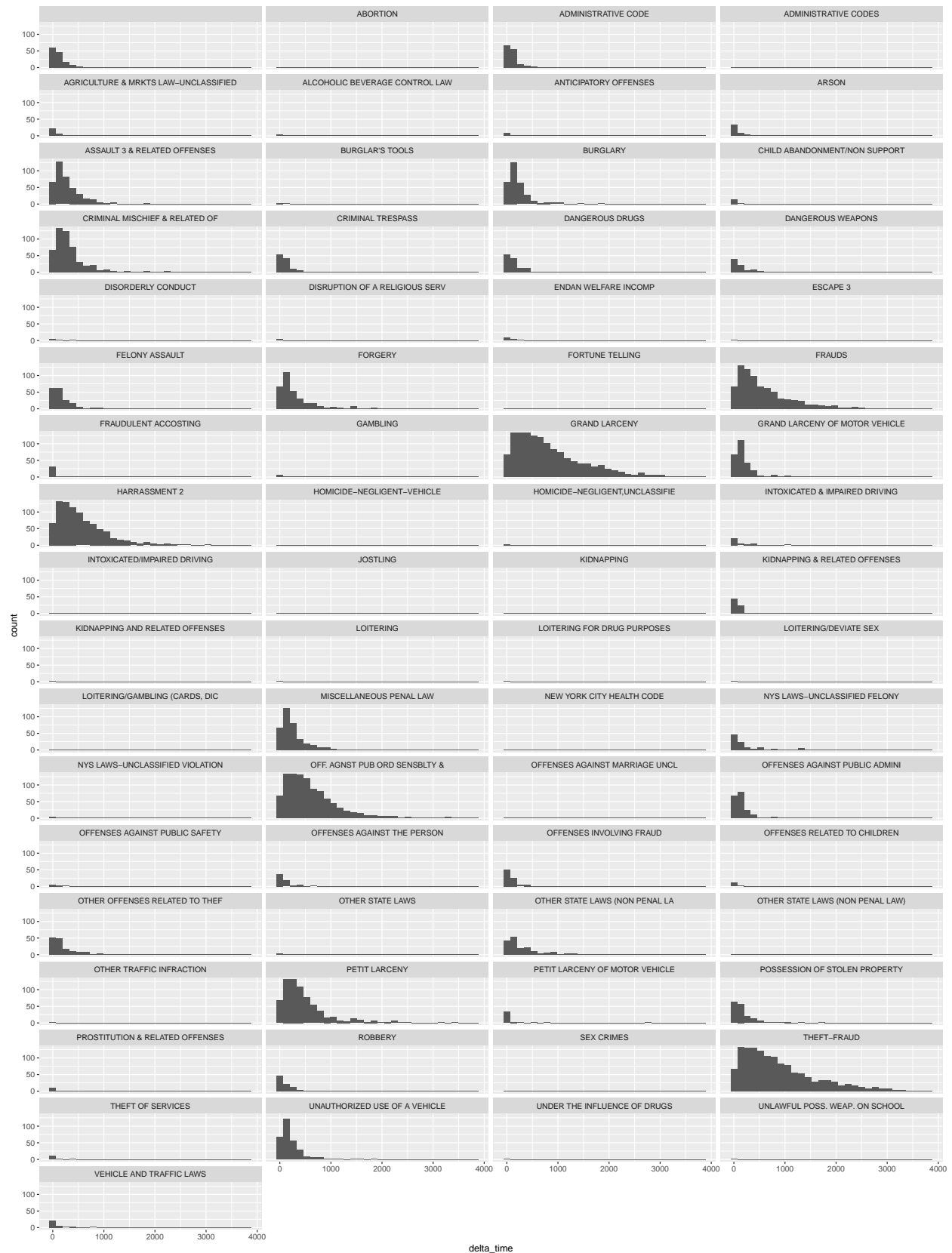
1. There are some cases where there might be typo on "To Date" especially year might be typo
2. Observed larceny(grand and petite have lot of cases)

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
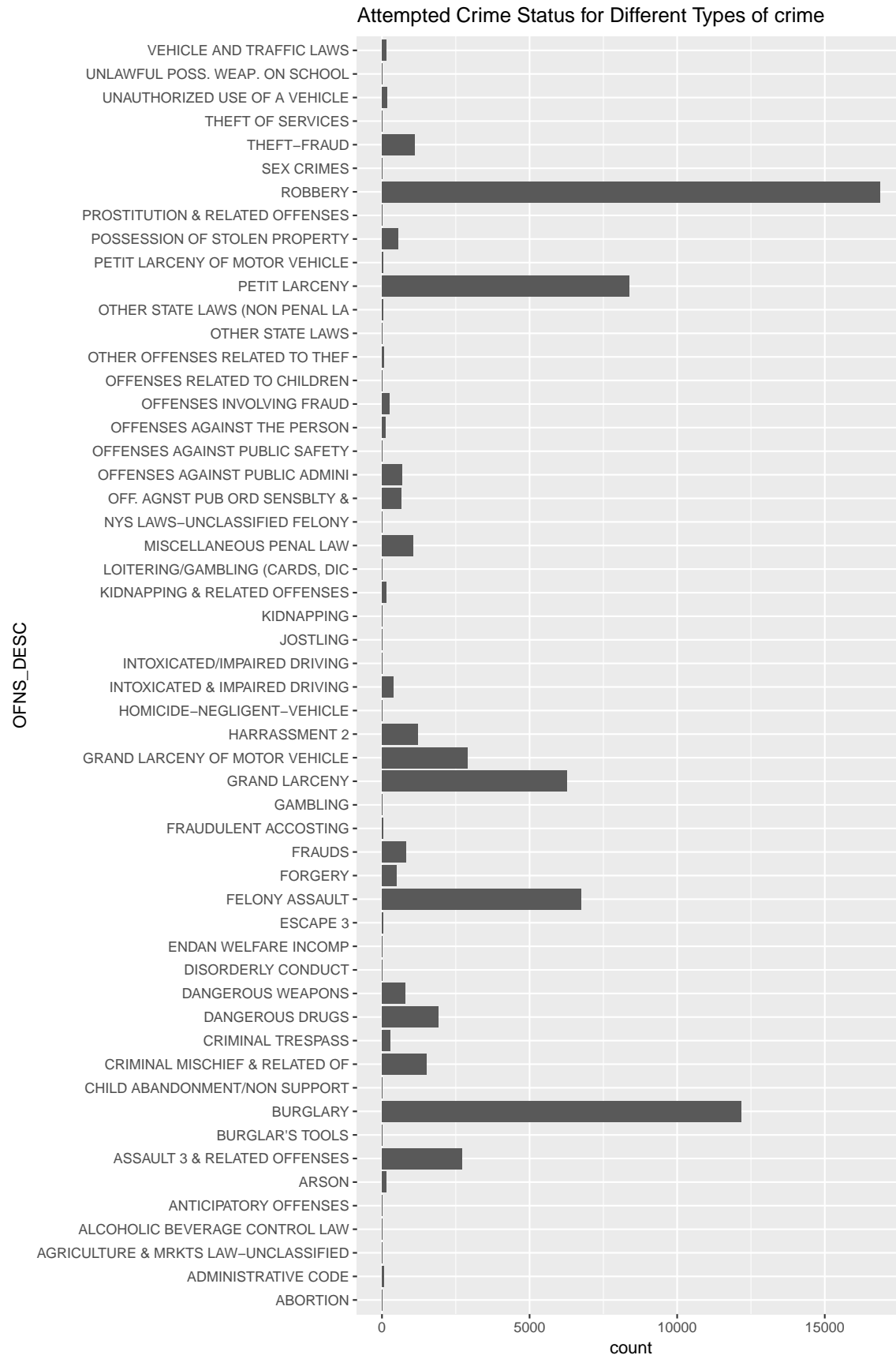
Attempted Crime Trend

```
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()
```
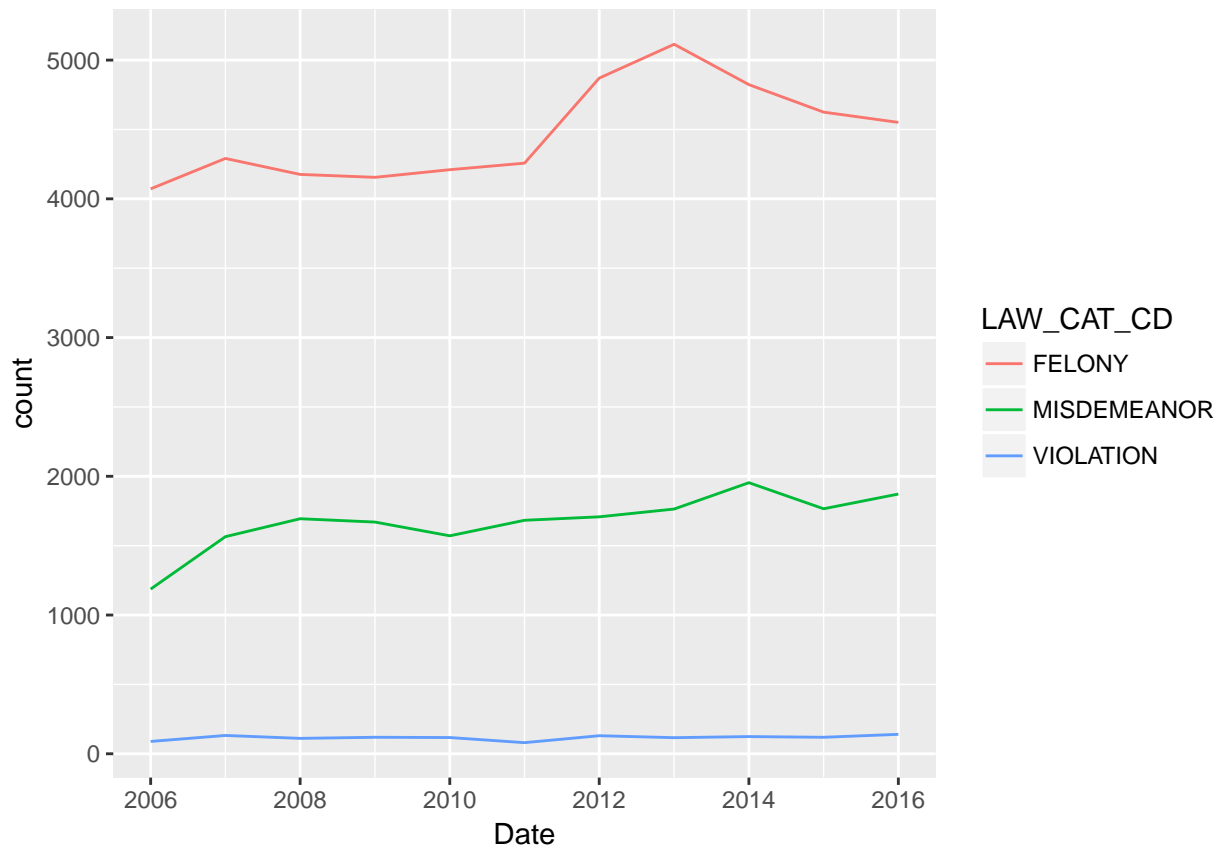

To find Top 10 Crime Categories, mosaic plots building blocks

```
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
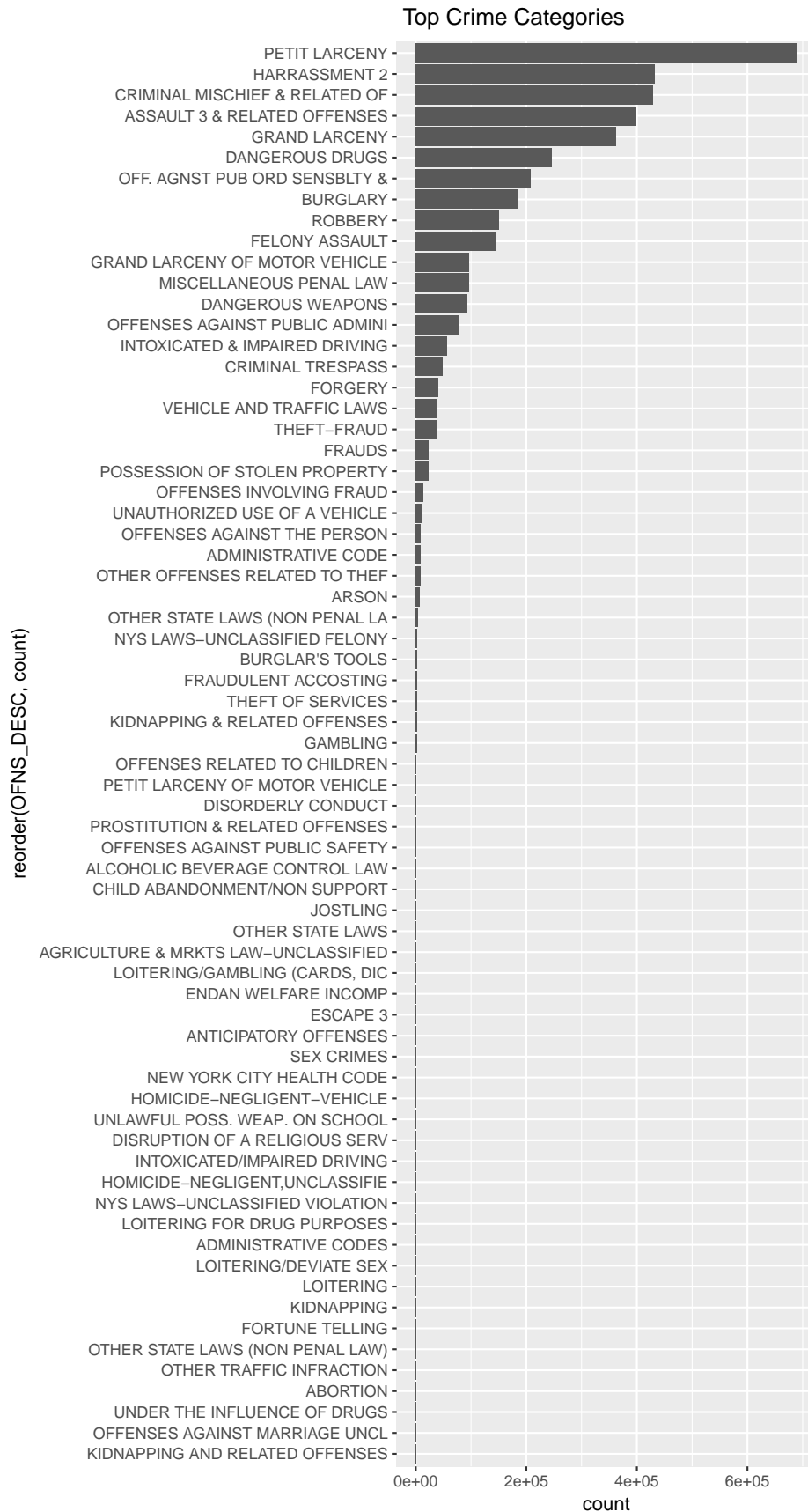
Boro, Juris, Crime Categories

```
                  #group_by(BORO_NM,JURIS_DESC,OFNS_DESC) %>%
                  #mutate(count=n()) %>%
top_ofns   <- c("PETIT LARCENY", "HARRASSMENT 2", "CRIMINAL MISCHIEF & RELATED OF", "ASSAUALT 3 & RELA
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)), crime
```
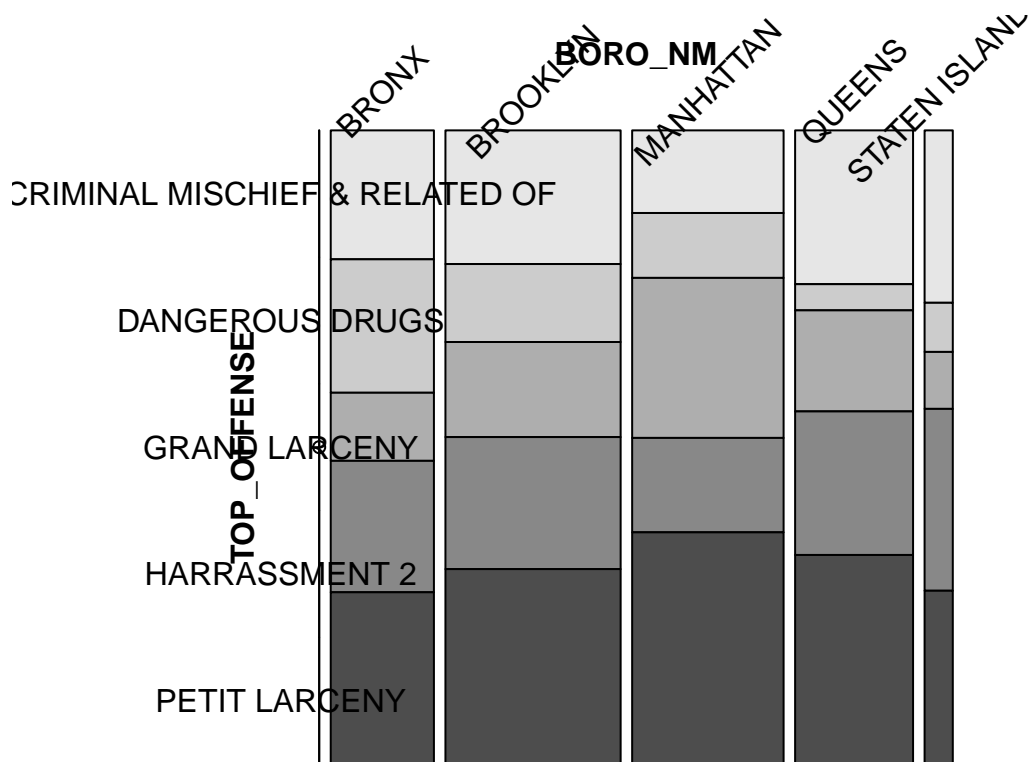


```
#doubledecker(TOP_OFFENSE~BORO_NM, data=crime_sort)
ggplot(crime_sort, aes(TOP_OFFENSE,Freq, fill=BORO_NM)) +
  geom_col() +
  facet_grid(LAW_CAT_CD~ BORO_NM) +
  coord_flip()
```
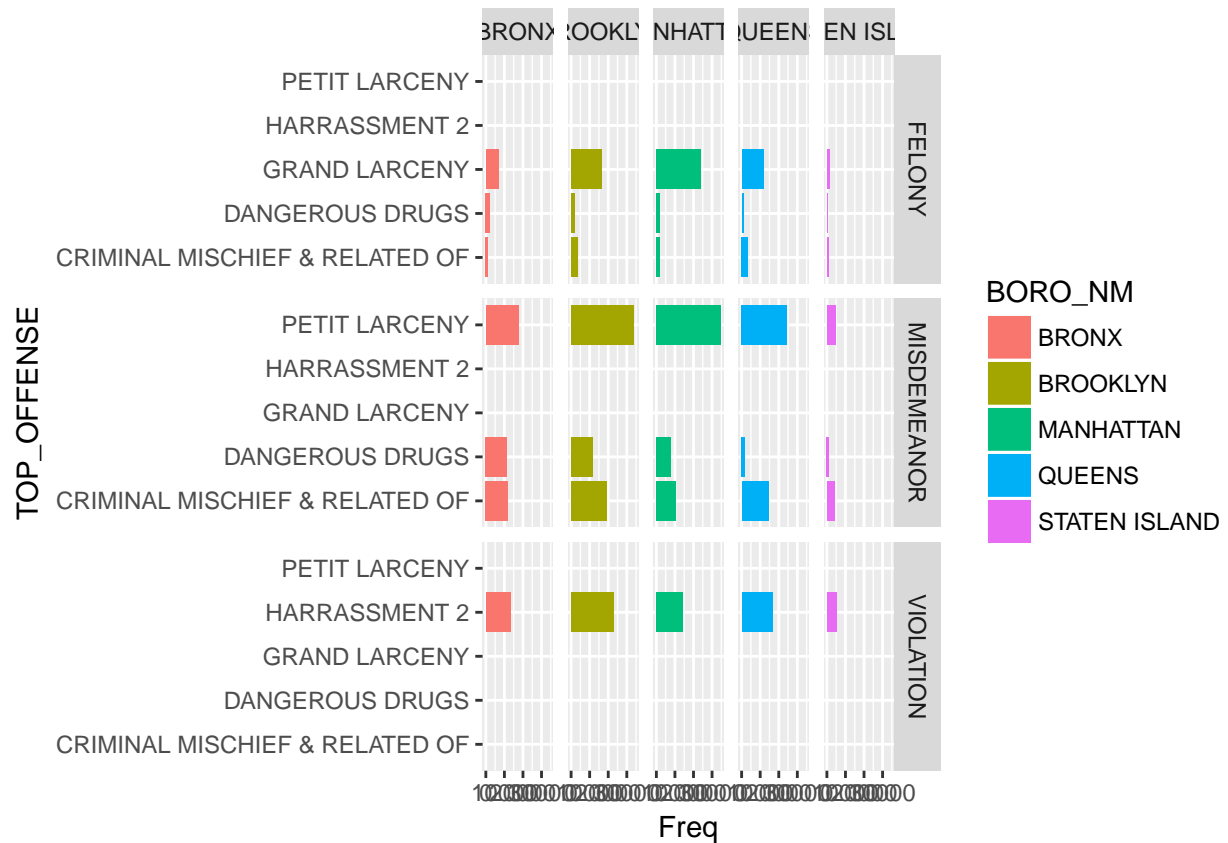
** The above plot shows something surprising, the categories are not standard, need to research more. For example, dangerous drugs is under Felony as well as Misdemeanor!! **

```
ggplot(crime_sort, aes(TOP_OFFENSE,Freq)) +
        geom_col() +
        facet_wrap(~BORO_NM) +
        coord_flip()
```
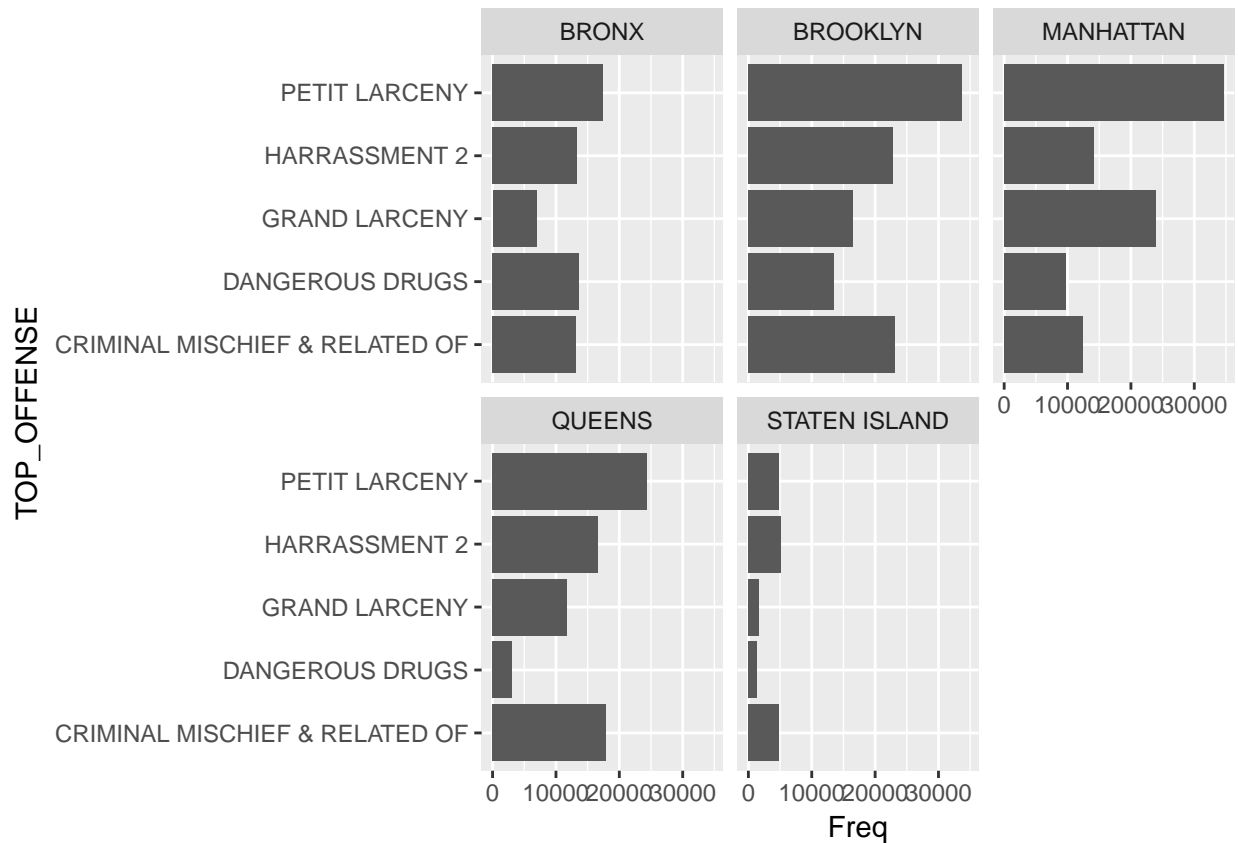
- I tired indivial Crime Types, the colors were too confusing as lot of categories

```r
crime_parks <- tbl_df(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))

crime_pk <- crime_parks %>%
                group_by(BORO_NM) %>%
                top_n(n=10, wt=count)

crime_parks_1 <- crime_df %>% drop_na() %>%
                filter(BORO_NM!="",PARKS_NM!="",OFNS_DESC!="") %>%
                group_by(BORO_NM,PARKS_NM,OFNS_DESC) %>%
                summarize(count=n())

crime_pk_1 <- crime_parks_1 %>%
                group_by(BORO_NM) %>%
                top_n(n=10, wt=count)

ggplot(crime_pk_1 ,aes(PARKS_NM, count, fill=OFNS_DESC)) +
    geom_col() +
    facet_wrap(~BORO_NM, ncol=1, scales="free_y") +
```

```
    coord_flip()
```

PARKS_NM

VAN CORTLANDT PARK
TREMONT PARK
ST. MARY'S PARK BRONX
MACOMBS DAM PARK
FRANZ SIGEL PARK
EWEN PARK
CROTONA PARK
CLAREMONT PARK

PROSPECT PARK
MCCARREN PARK
CONEY ISLAND BEACH & BOARDWALK
BROOKLYN BRIDGE PARK

RIVERSIDE PARK
RANDALL'S ISLAND PARK
MARCUS GARVEY PARK
JACKIE ROBINSON PARK MANHATTAN
HARLEM RIVER PARK
CENTRAL PARK

ROCKAWAY BEACH AND BOARDWALK
FOREST PARK
FLUSHING MEADOWS CORONA PARK

TOMPKINSVILLE PARK
TAPPEN PARK
STAPLETON PLAYGROUND
LT. LIA PLAYGROUND
FRANKLIN D. ROOSEVELT BOARDWALK AND BEACH
FABER POOL AND PARK
DE MATTI PLAYGROUND
CPL. THOMPSON PARK
CONFERENCE HOUSE PARK
CLOVE LAKES PARK
BLOOMINGDALE PARK

count

OFNS_DESC
- ASSAULT 3 & RELATED OFFENSES
- BURGLARY
- CRIMINAL MISCHIEF & RELATED OF
- CRIMINAL TRESPASS
- DANGEROUS DRUGS
- GRAND LARCENY
- HARRASSMENT 2
- INTOXICATED & IMPAIRED DRIVING
- MISCELLANEOUS PENAL LAW
- OFF. AGNST PUB ORD SENSBLTY &
- OFFENSES AGAINST PUBLIC ADMINI
- OFFENSES INVOLVING FRAUD
- PETIT LARCENY
- ROBBERY

17

```
ggplot(crime_pk_1 ,aes(PARKS_NM, count, fill=OFNS_DESC)) +
  geom_col() +
  facet_wrap(~BORO_NM, ncol=1, scales="free") +
  coord_flip()
```

PARKS_NM

VAN CORTLANDT PARK
TREMONT PARK
ST. MARY'S PARK BRONX
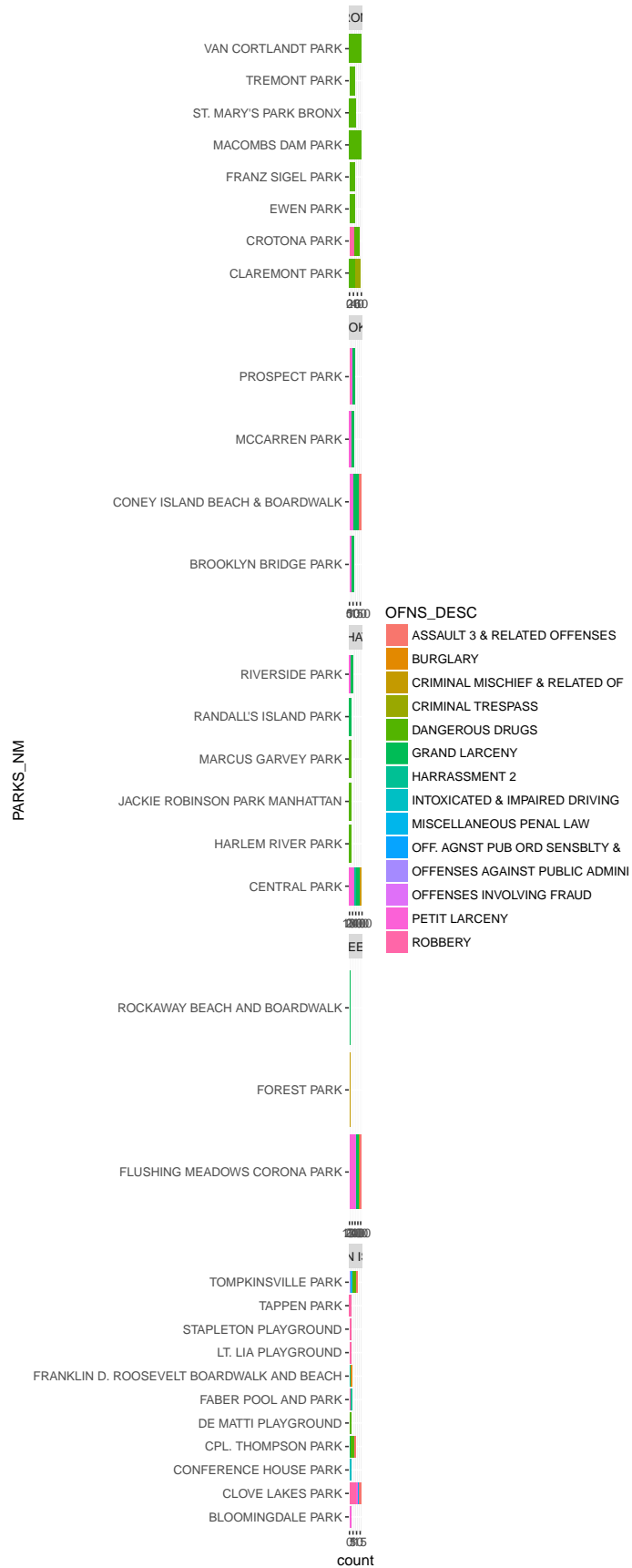MACOMBS DAM PARK
FRANZ SIGEL PARK
EWEN PARK
CROTONA PARK
CLAREMONT PARK

PROSPECT PARK
MCCARREN PARK
CONEY ISLAND BEACH & BOARDWALK
BROOKLYN BRIDGE PARK

RIVERSIDE PARK
RANDALL'S ISLAND PARK
MARCUS GARVEY PARK
JACKIE ROBINSON PARK MANHATTAN
HARLEM RIVER PARK
CENTRAL PARK

ROCKAWAY BEACH AND BOARDWALK
FOREST PARK
FLUSHING MEADOWS CORONA PARK

TOMPKINSVILLE PARK
TAPPEN PARK
STAPLETON PLAYGROUND
LT. LIA PLAYGROUND
FRANKLIN D. ROOSEVELT BOARDWALK AND BEACH
FABER POOL AND PARK
DE MATTI PLAYGROUND
CPL. THOMPSON PARK
CONFERENCE HOUSE PARK
CLOVE LAKES PARK
BLOOMINGDALE PARK

count

OFNS_DESC
- ASSAULT 3 & RELATED OFFENSES
- BURGLARY
- CRIMINAL MISCHIEF & RELATED OF
- CRIMINAL TRESPASS
- DANGEROUS DRUGS
- GRAND LARCENY
- HARRASSMENT 2
- INTOXICATED & IMPAIRED DRIVING
- MISCELLANEOUS PENAL LAW
- OFF. AGNST PUB ORD SENSBLTY &
- OFFENSES AGAINST PUBLIC ADMINI
- OFFENSES INVOLVING FRAUD
- PETIT LARCENY
- ROBBERY

19

trial on ggmap

```r
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.