

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

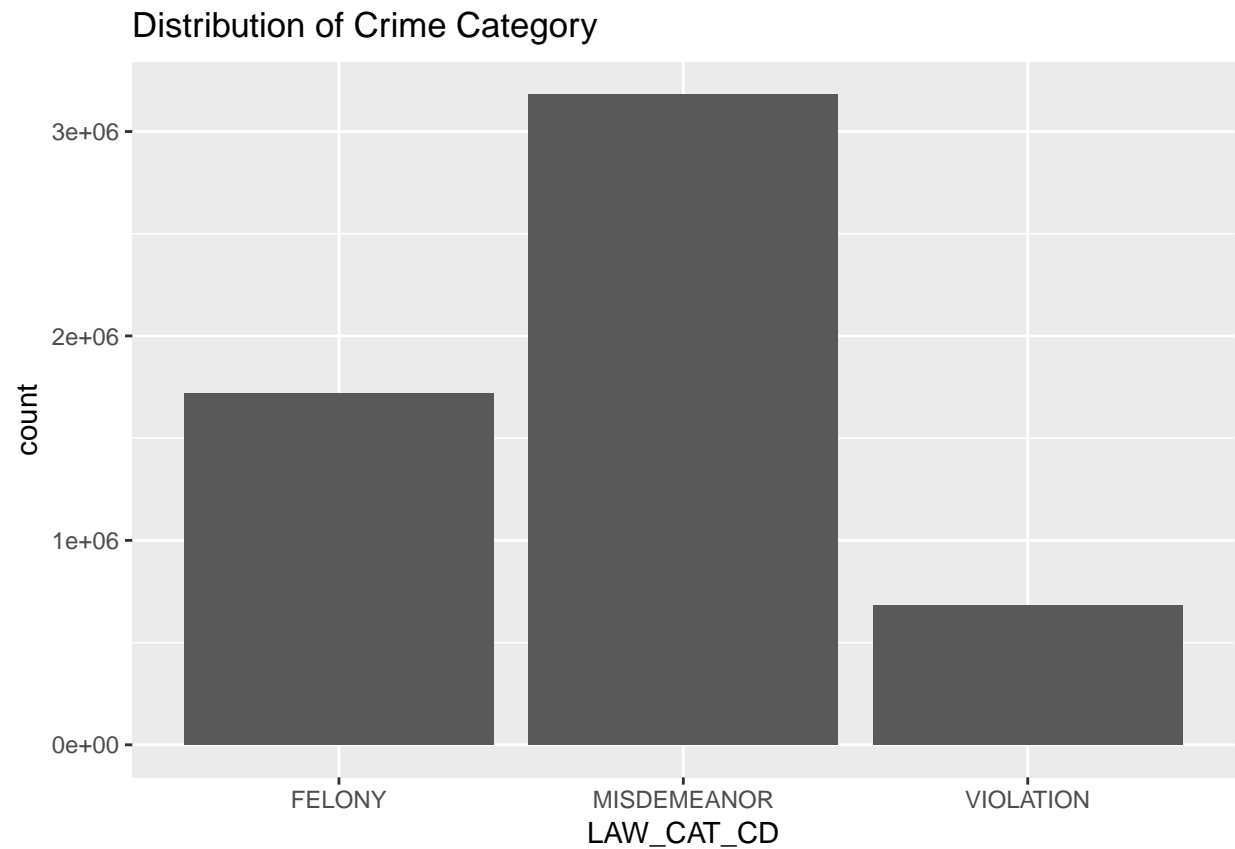
```
#Convert dates and times to correct format
crime_df$CMPLNT_FR_DT <- as.Date(crime_df$CMPLNT_FR_DT, format='%m/%d/%Y')
crime_df$CMPLNT_TO_DT <- as.Date(crime_df$CMPLNT_TO_DT, format='%m/%d/%Y')
crime_df$RPT_DT <- as.Date(crime_df$RPT_DT, format='%m/%d/%Y')

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

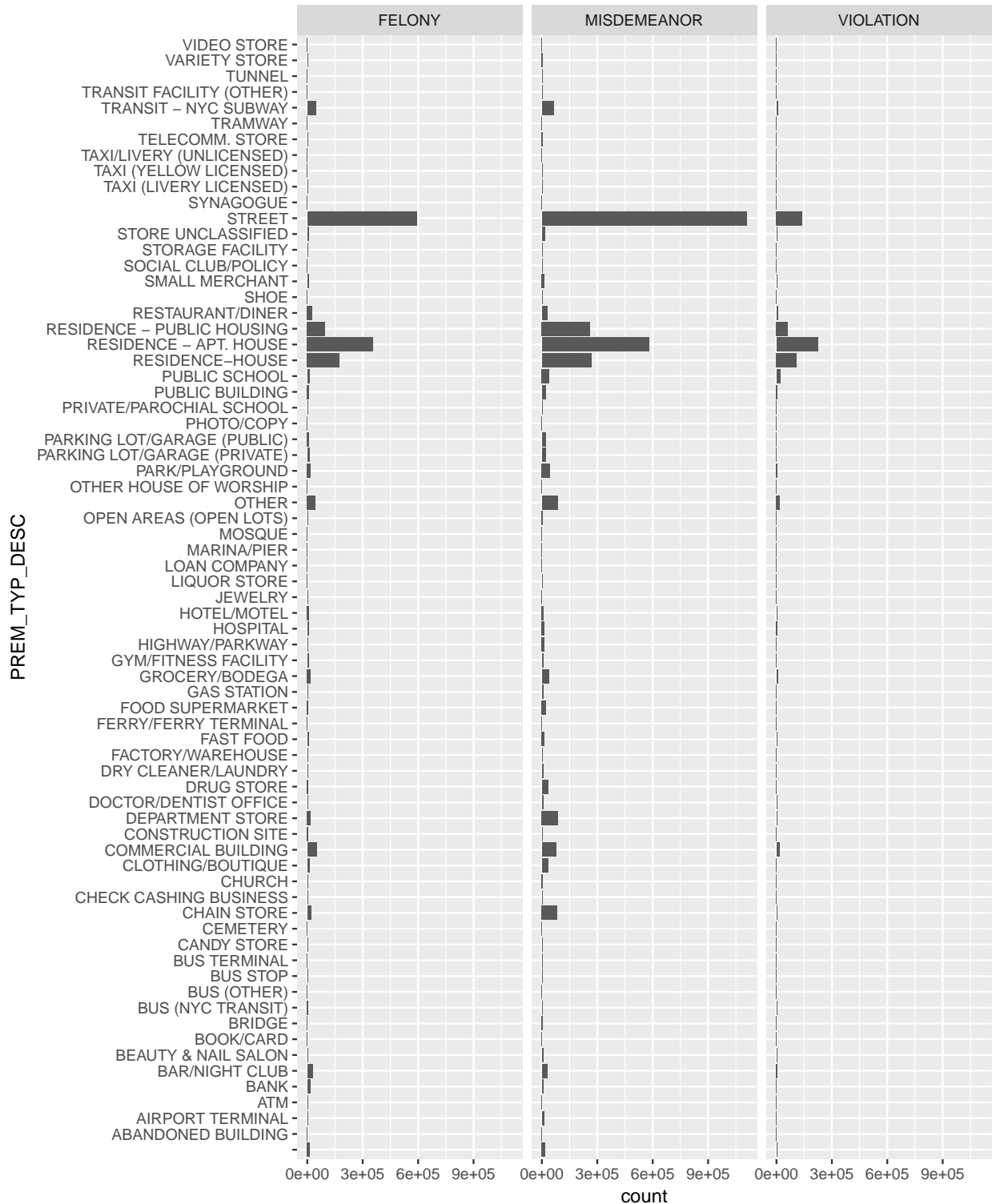
```
ggplot(crime_df,aes(LAW_CAT_CD)) +
  geom_bar() +
  ggtitle("Distribution of Crime Category")
```



Type of Offense

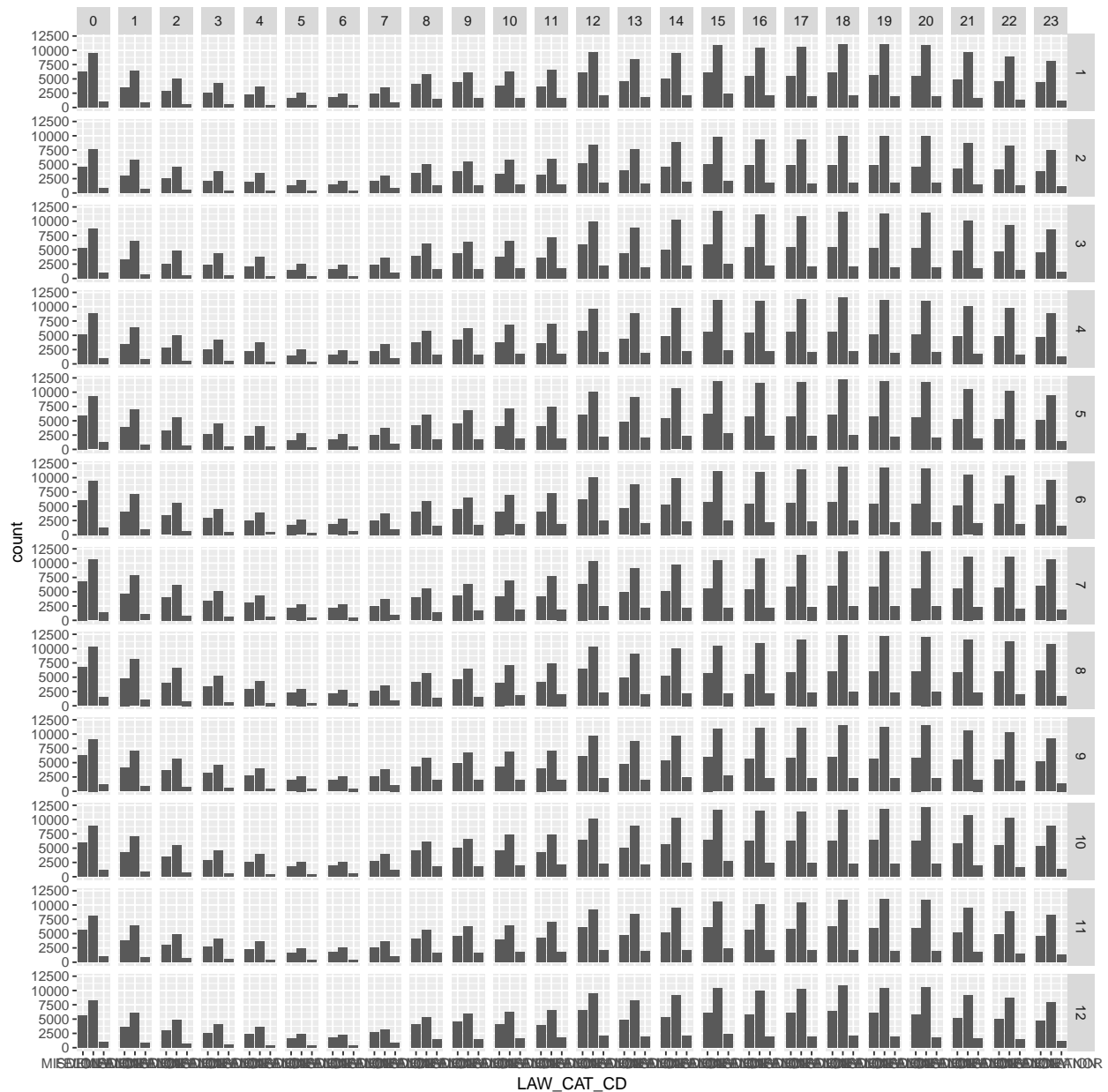
```
ggplot(crime_df, aes(PREM_TYP_DESC)) +  
  geom_bar() +  
  facet_wrap(~LAW_CAT_CD) +  
  coord_flip() +  
  ggtitle("Crime Category Vs Place of Crime")
```

Crime Category Vs Place of Crime



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))
```



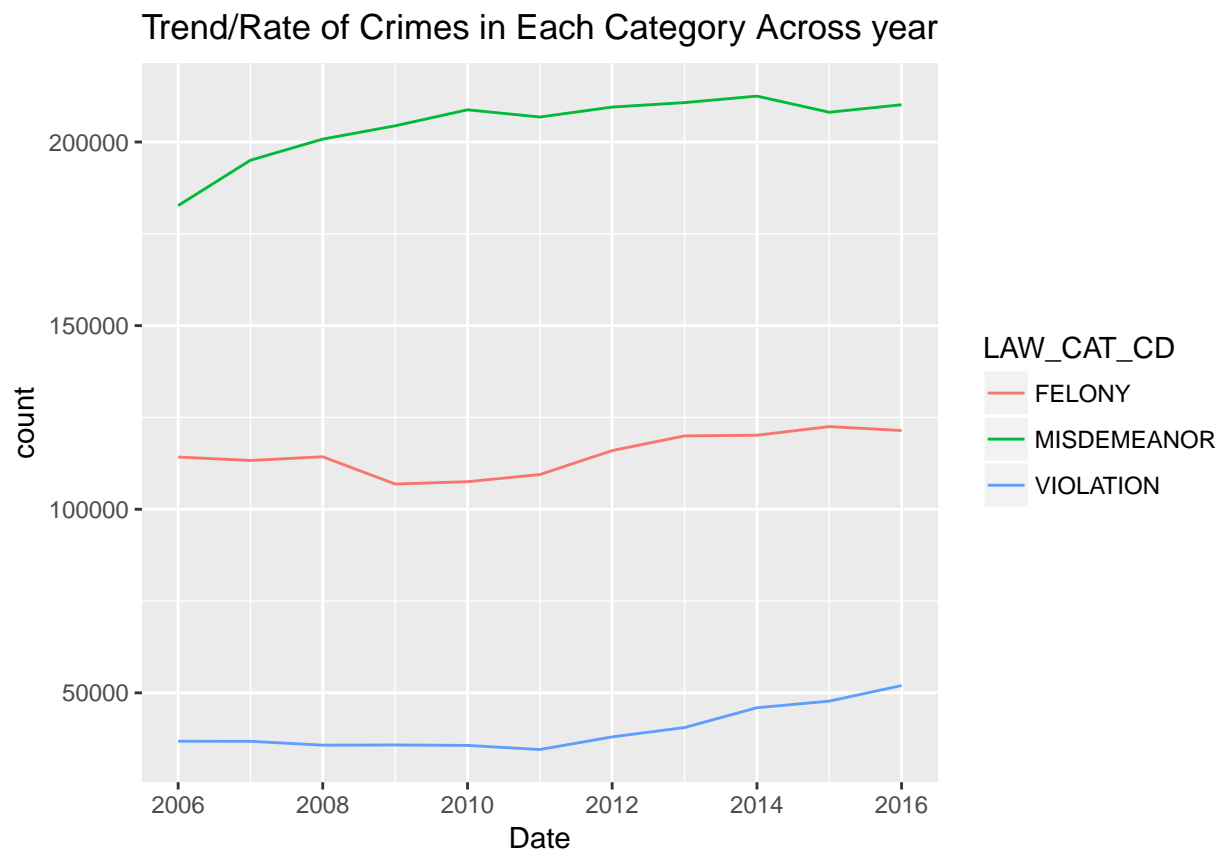
Time Series - Trend of Crime Rate

```

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")

```

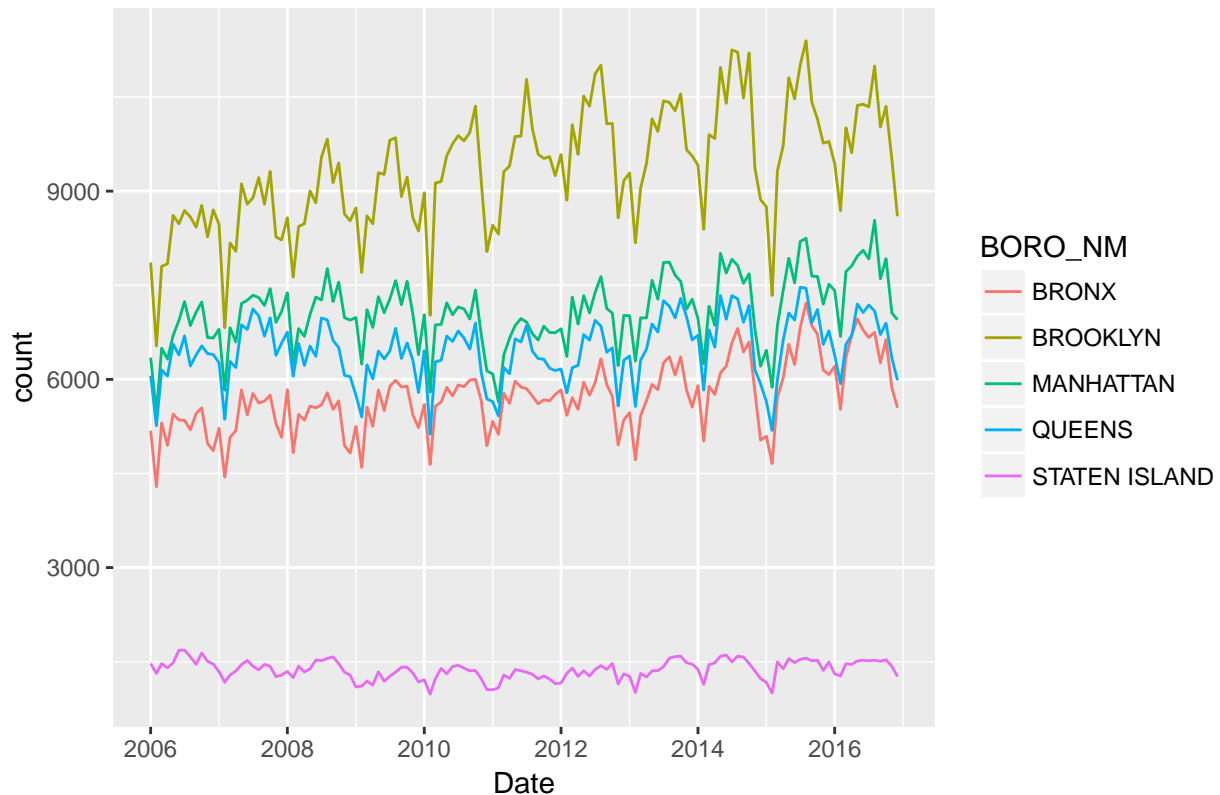
Trend/Rate of Crimes in Each Category Across year – sampled month–w



```
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

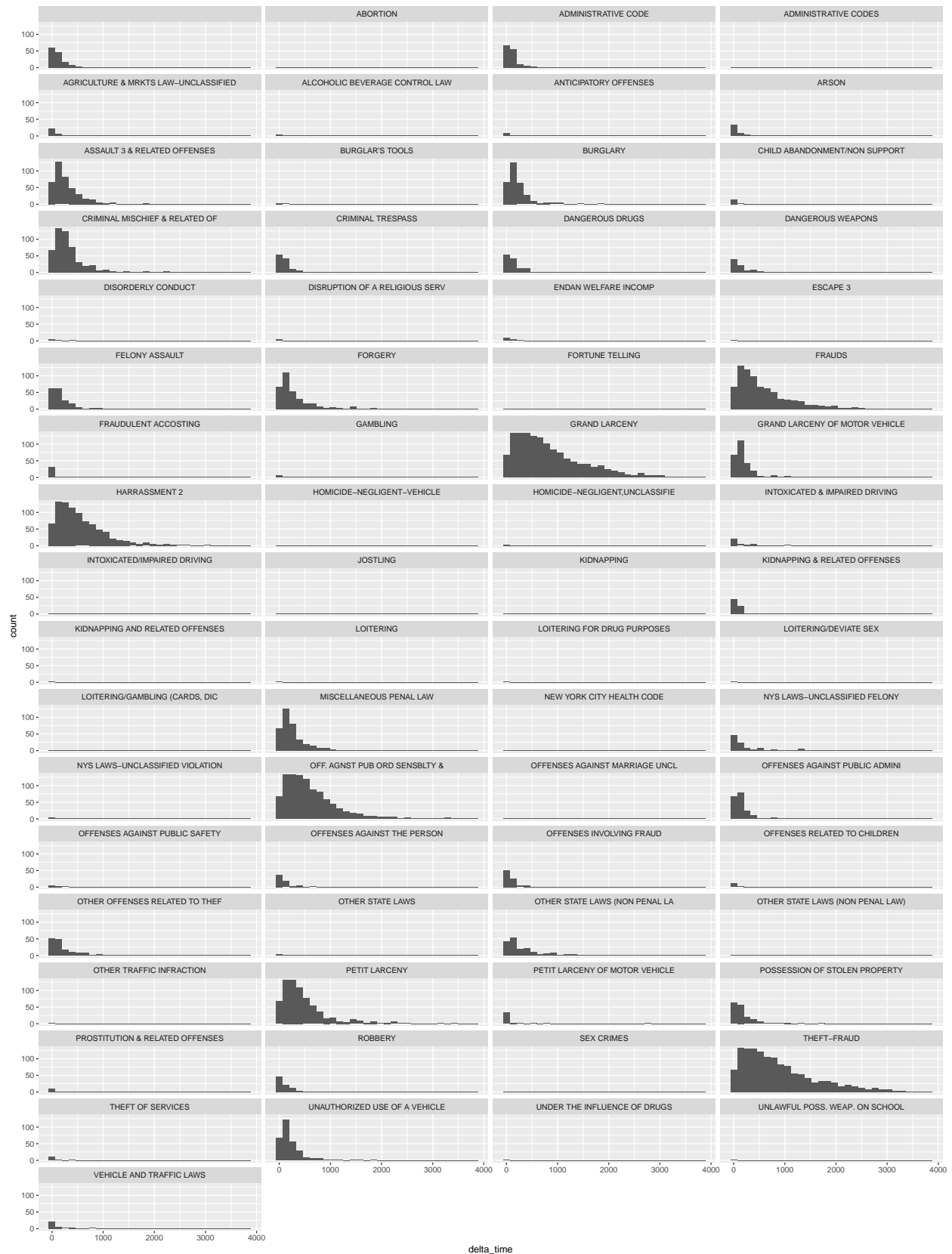


* Shows monthly pattern similar to Jingbo's * Year pattern fluctuates * Some NM_BORO are empty * Gaps between bororughs reduces towards later years

length of Crime Vs Type of Crime

```
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)
```

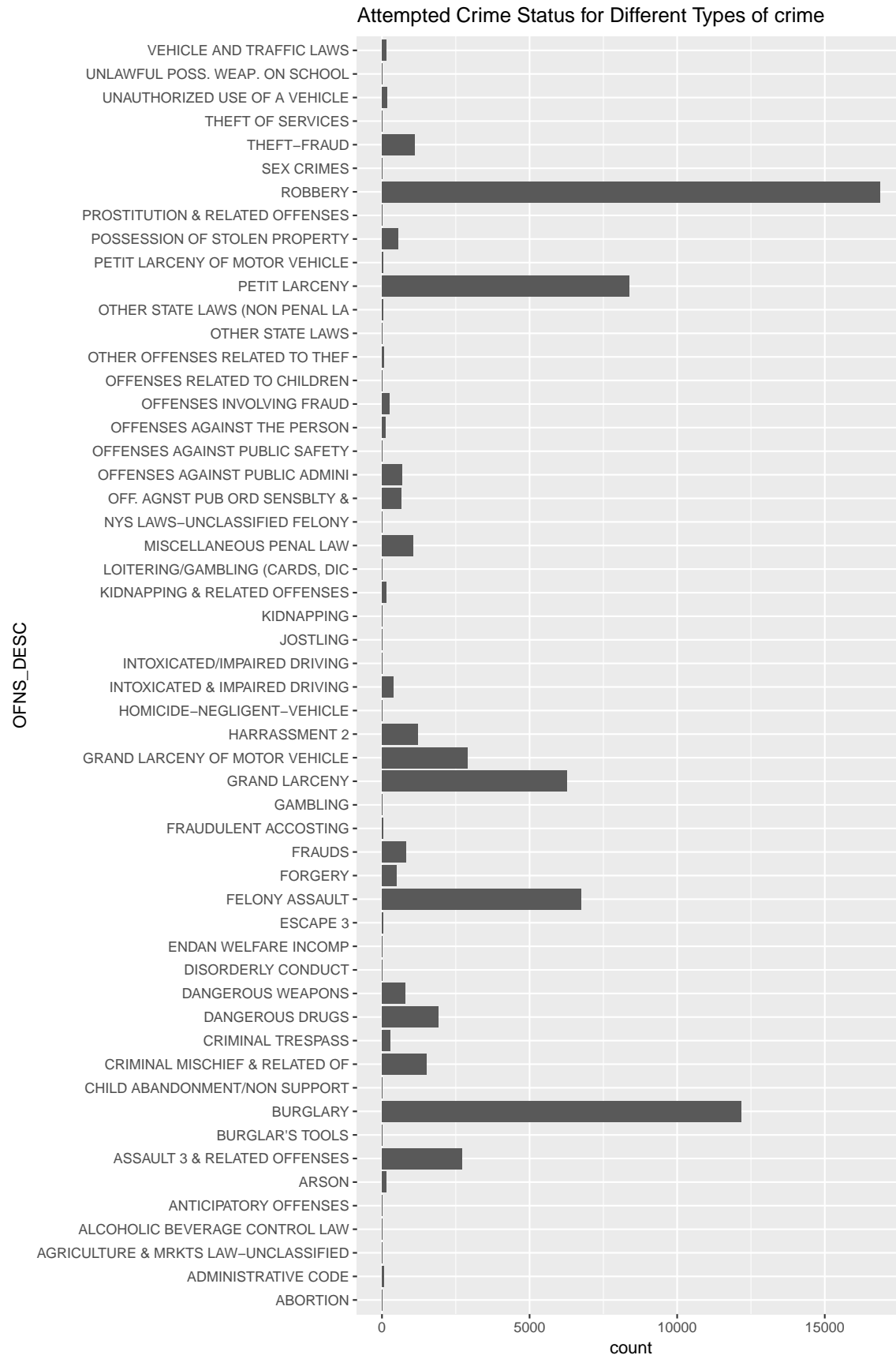


1. There are some cases where there might be typo on "To Date" especially year might be typo
2. Observed larceny(grand and petite have lot of cases)

3. There are blank Offense category

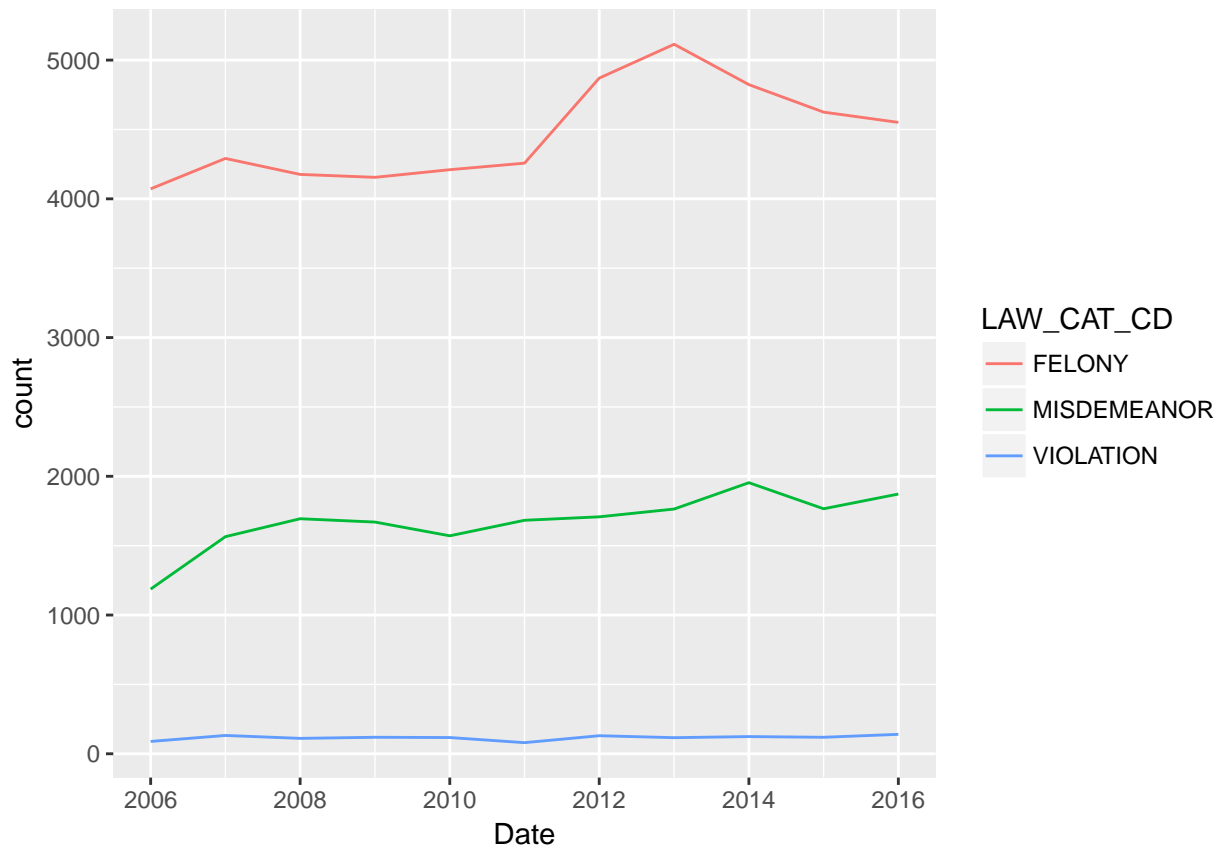
Attempted Crime vs Type of Crime

```
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 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()
```



- I tired indivial Crime Types, the colors were too confusing as lot of categories Note that the `echo = FALSE` parameter was added to the code chunk to prevent printing of the R code that generated the plot.