ap3650_nyc_crime_data_visualization

Anita

March 13, 2018

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


#fread("NYPD_Complaint_Data_Historic.csv",na.strings="",colClasses = c(ParkName="c",HADEVELOPT="c"))->c
#crime_df <- fread("NYPD_Complaint_Data_Historic.csv",na.strings="")
#crime_df_1 <- read.csv("NYPD_Complaint_Data_Historic.csv", header=TRUE)

## Copied from rj2168.rmd for uniform read and variable names

var_names <- c("Id", "DateStart", "TimeStart", "DateEnd", "TimeEnd", "DateReport", "ClassCode", "Offens
               "IntClassCode", "IntOffenseDesc", "AtptCptdStatus", "Level", "Jurisdiction", "Boro", "Pc
               "PremDesc", "ParkName", "HousingDevName", "XCoord", "YCoord", "Lat", "Long", "Lat_Long")

crime_df <- fread("NYPD_Complaint_Data_Historic.csv",na.strings="", col.names = var_names, stringsAsFac
```

```
##
Read 0.0% of 5580035 rows
Read 3.6% of 5580035 rows
Read 7.0% of 5580035 rows
Read 10.6% of 5580035 rows
Read 13.8% of 5580035 rows
Read 17.0% of 5580035 rows
Read 19.9% of 5580035 rows
Read 23.3% of 5580035 rows
Read 26.9% of 5580035 rows
Read 28.7% of 5580035 rows
Read 29.9% of 5580035 rows
Read 33.7% of 5580035 rows
Read 37.5% of 5580035 rows
Read 40.5% of 5580035 rows
Read 43.0% of 5580035 rows
Read 45.9% of 5580035 rows
Read 48.9% of 5580035 rows
Read 51.8% of 5580035 rows
Read 53.9% of 5580035 rows
Read 57.3% of 5580035 rows
Read 61.3% of 5580035 rows
Read 64.7% of 5580035 rows
Read 67.4% of 5580035 rows
Read 71.1% of 5580035 rows
```

```
Read 75.3% of 5580035 rows
Read 79.4% of 5580035 rows
Read 83.7% of 5580035 rows
Read 87.1% of 5580035 rows
Read 90.3% of 5580035 rows
Read 94.1% of 5580035 rows
Read 98.2% of 5580035 rows
Read 5580035 rows and 24 (of 24) columns from 1.362 GB file in 00:00:49
```

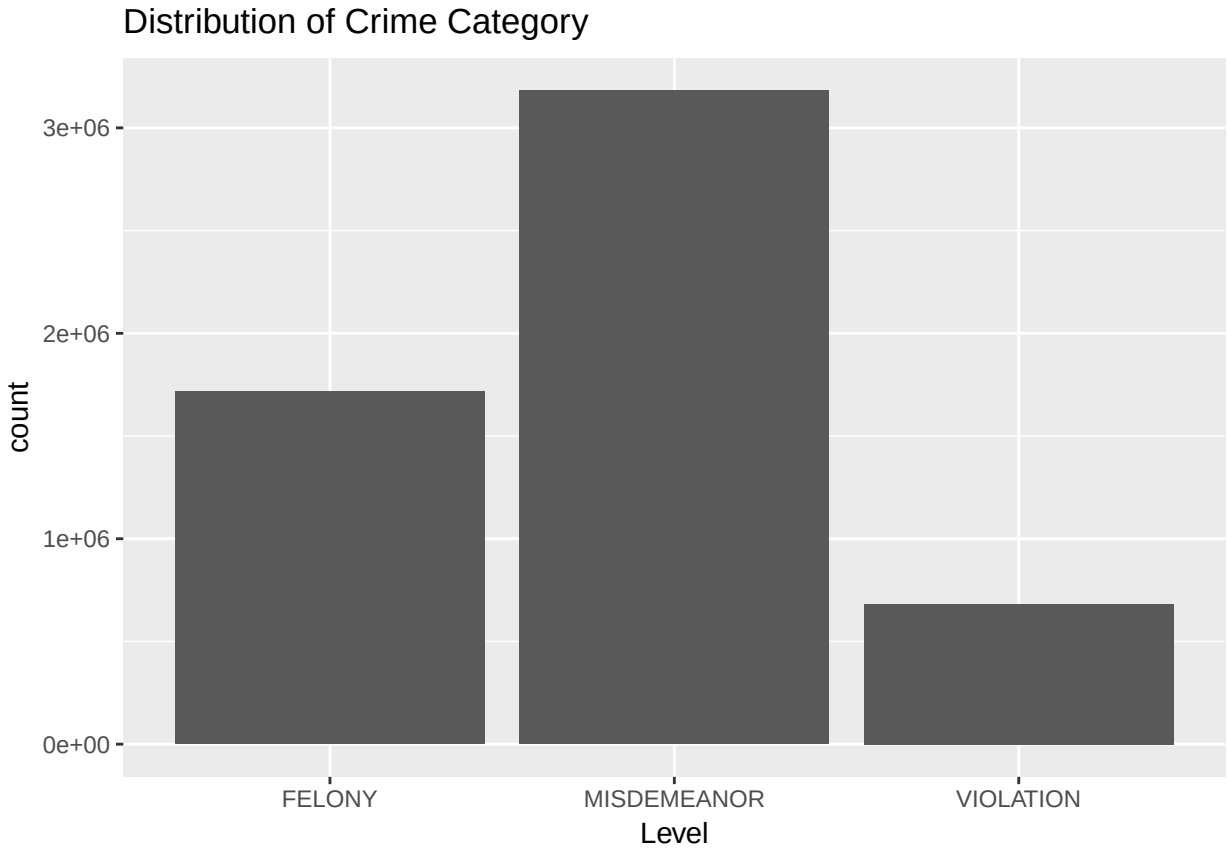
Data Manipulations

```
#Convert dates and times to correct format

#New Variable Names
crime_df$DateStart  <- as.Date(crime_df$DateStart, format='%m/%d/%Y')
crime_df$DateEnd    <- as.Date(crime_df$DateEnd, format='%m/%d/%Y')
crime_df$DateReport <- as.Date(crime_df$DateReport, format='%m/%d/%Y')

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

Plots

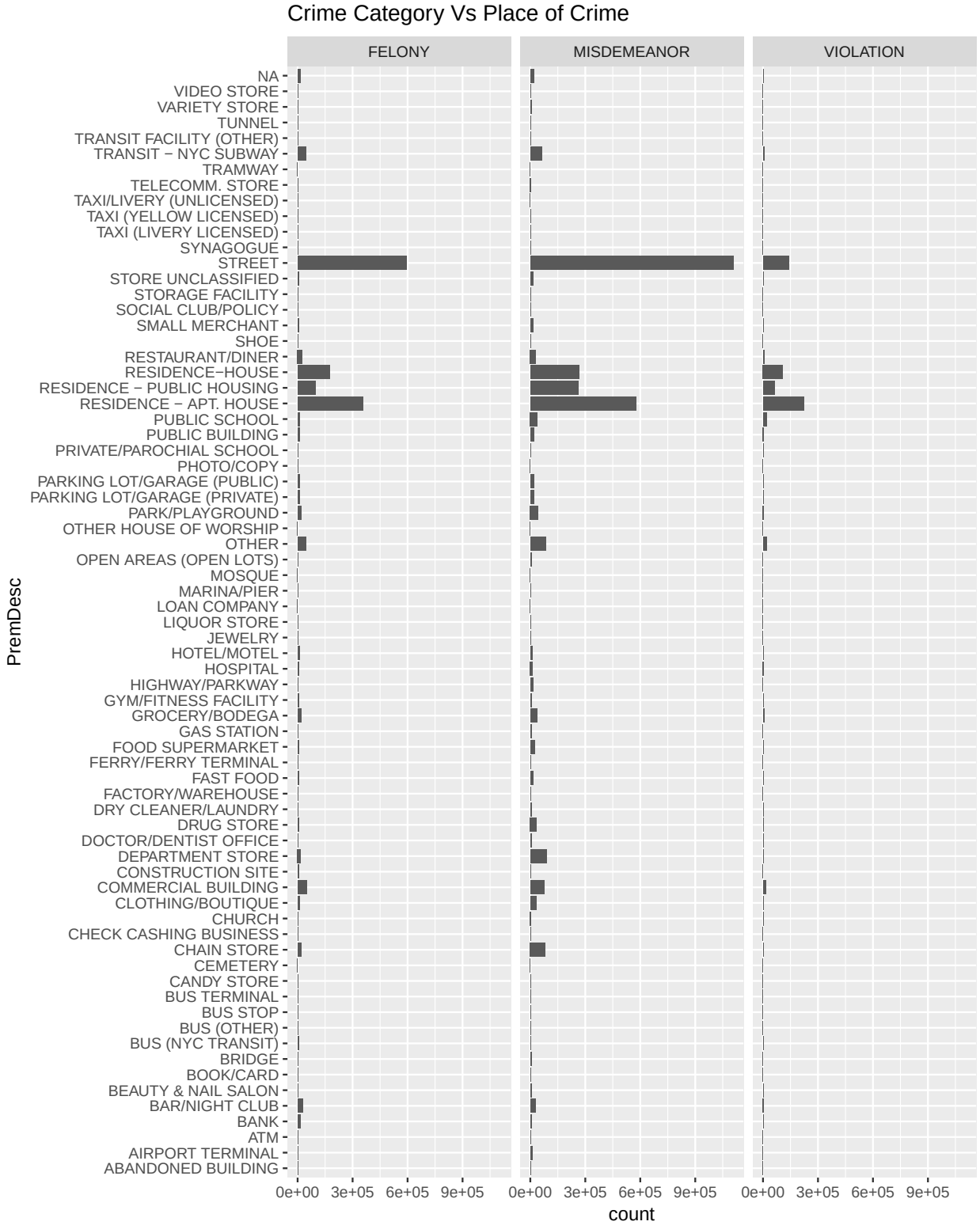
Warm-up Plot :-) Bar Chart

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



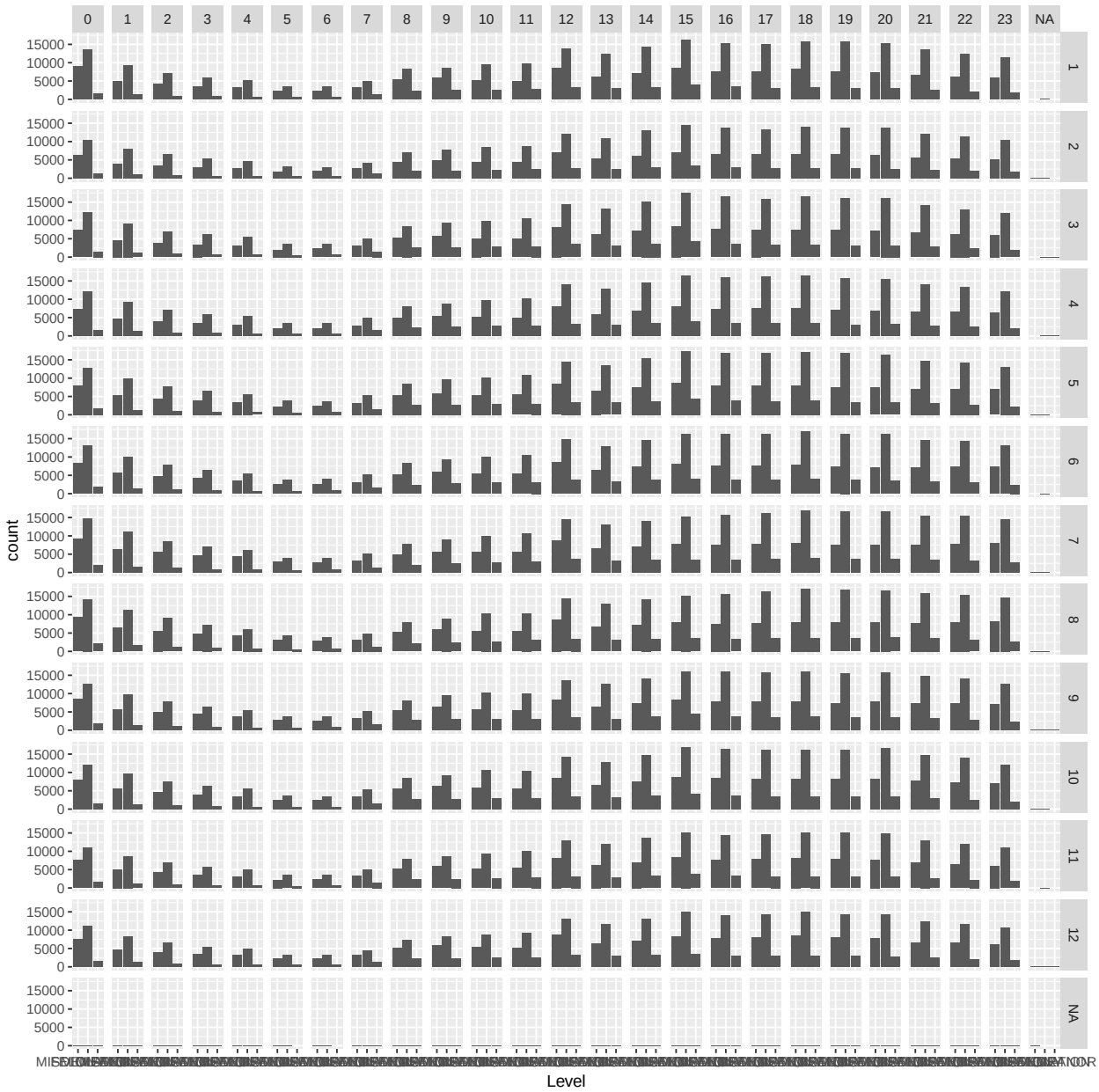
Type of Offense

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


Crime Category Vs Place of Crime
FELONY
MISDEMEANOR
VIOLATION
PremDesc
NA
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−HOUSE
RESIDENCE − PUBLIC HOUSING
RESIDENCE − APT. 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

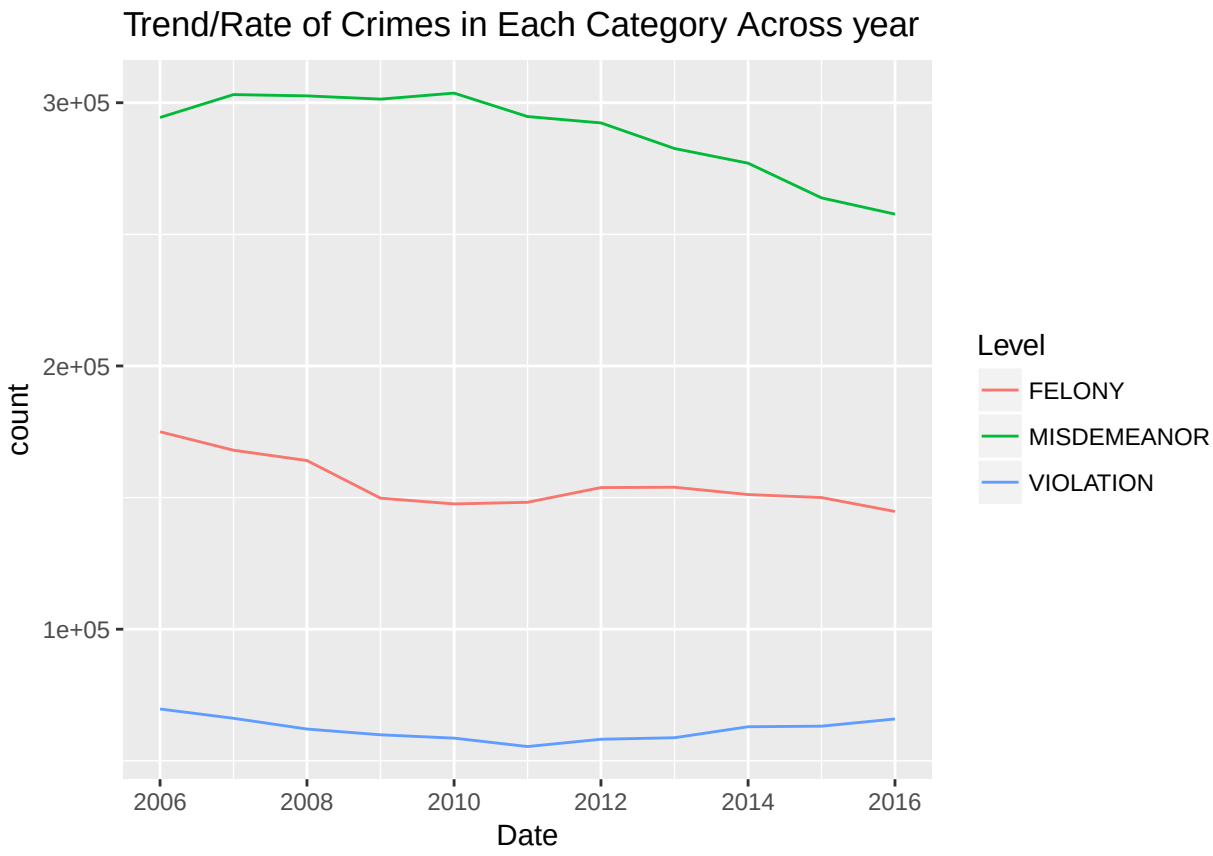
Month and Time and Type of Crime

```
#crime_df <- crime_df %>% drop_na()
ggplot(crime_df, aes(Level)) +
geom_bar() +
#facet_wrap(~month(DateStart))
#facet_wrap(~hour(TimeStart))
facet_grid(month(DateStart)~hour(TimeStart))
```

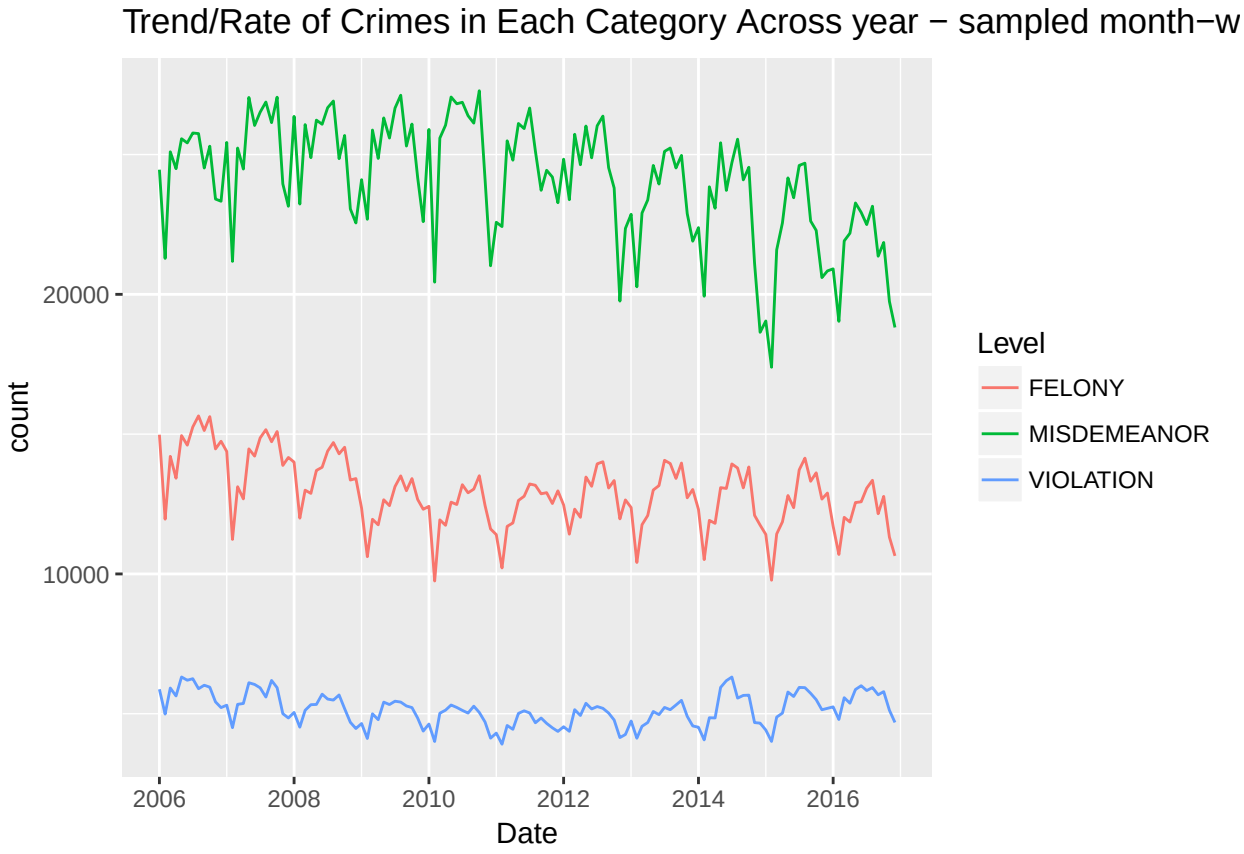


Time Series - Trend of Crime Rate

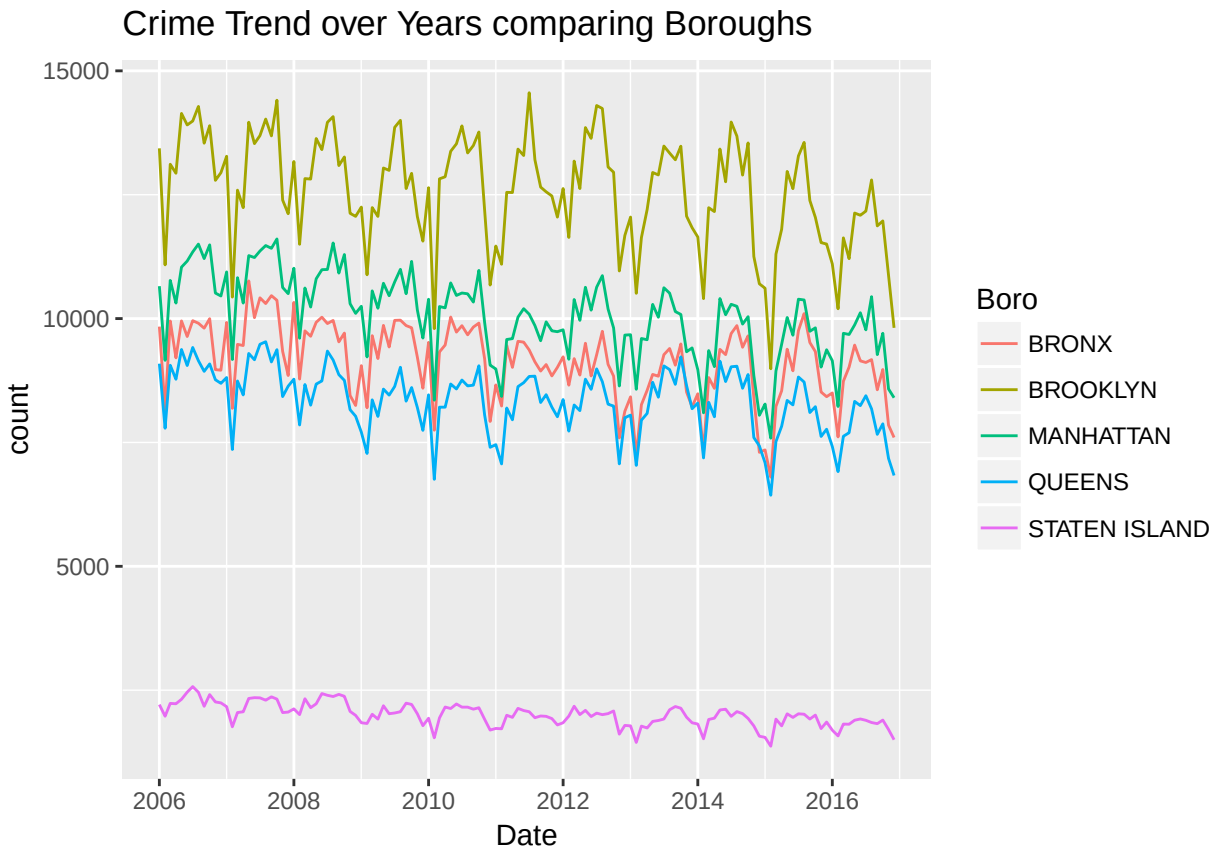
```
crime_time <- crime_df %>%
                filter(year(DateStart)>2005) %>%
                group_by(Date=floor_date(DateStart, "year"),Level) %>%
                summarize(count=n())

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




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

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




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

ggplot(crime_boro,aes(Date,count,color=Boro)) +
  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

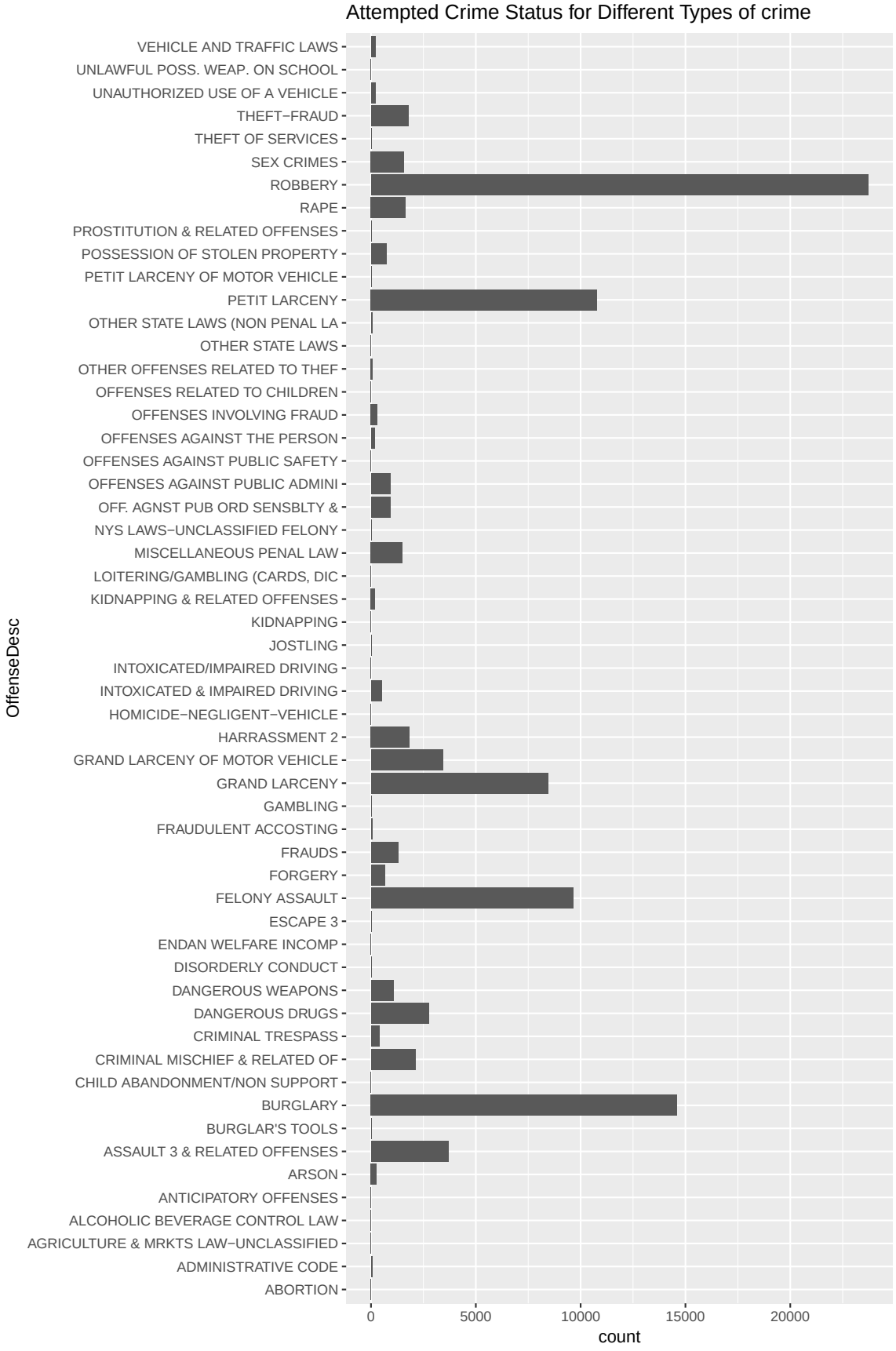
```
#crime_time <- crime_df %>%  drop_na() %>%
crime_time <- crime_df %>%
                  filter(year(DateStart)>2005) %>%
                  mutate(delta_time = as.numeric(DateEnd - DateStart)) %>%
                  group_by(OffenseDesc, delta_time) %>%
                  summarize(count=n())

ggplot(crime_time,aes(delta_time)) +
  geom_histogram(na.rm=TRUE) +
  facet_wrap(~OffenseDesc, 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

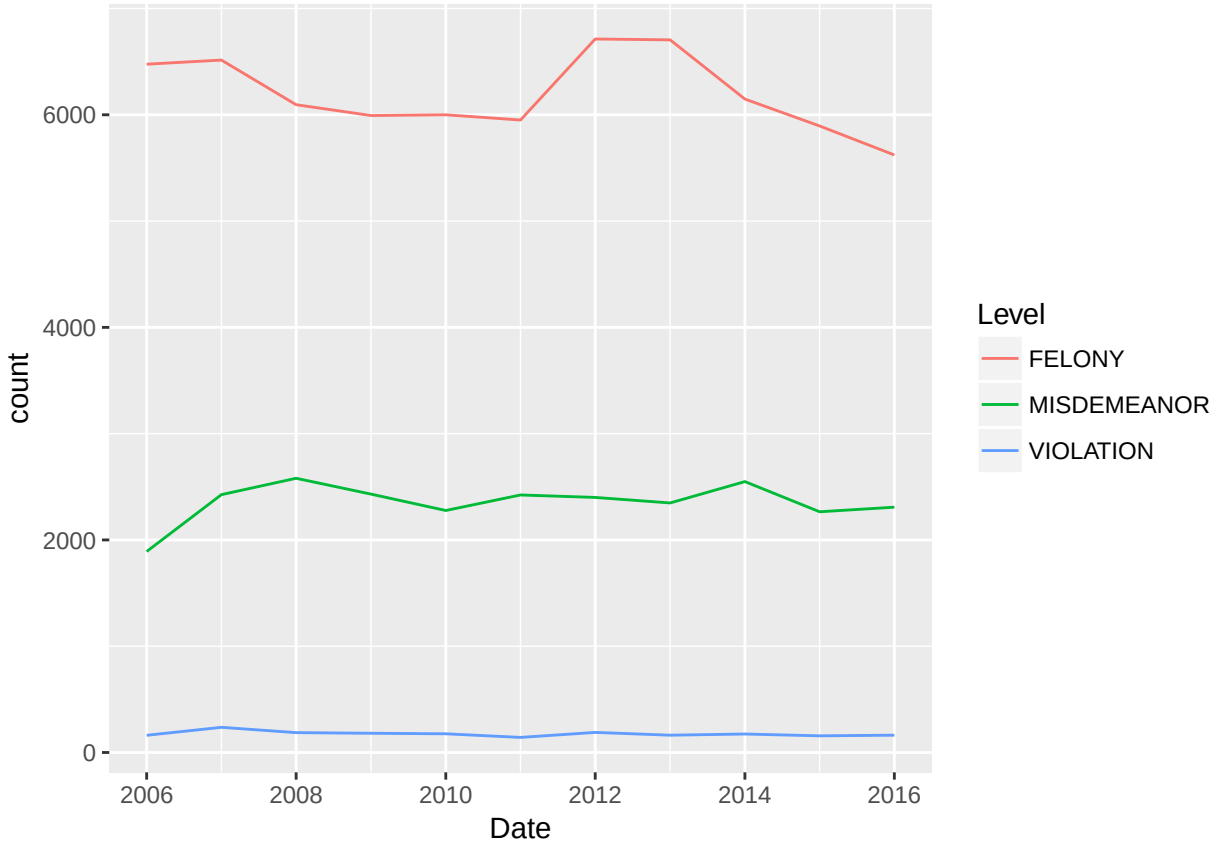
Atempted Crime vs Type of Crime

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


Attempted Crime Status for Different Types of crime
VEHICLE AND TRAFFIC LAWS
UNLAWFUL POSS. WEAP. ON SCHOOL
UNAUTHORIZED USE OF A VEHICLE
THEFT−FRAUD
THEFT OF SERVICES
SEX CRIMES
ROBBERY
RAPE
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
OffenseDesc
0
5000
10000
15000
20000
count

Attempted Crime Trend

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



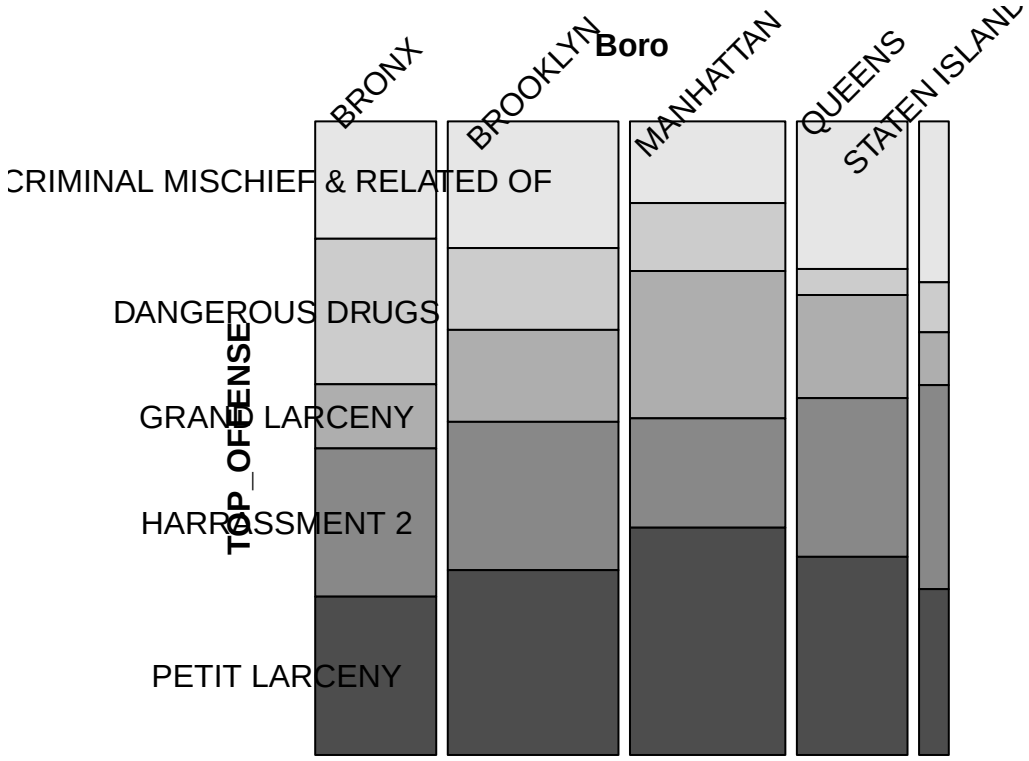
To find Top 10 Crime Categories, mosaic plots building blocks

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

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

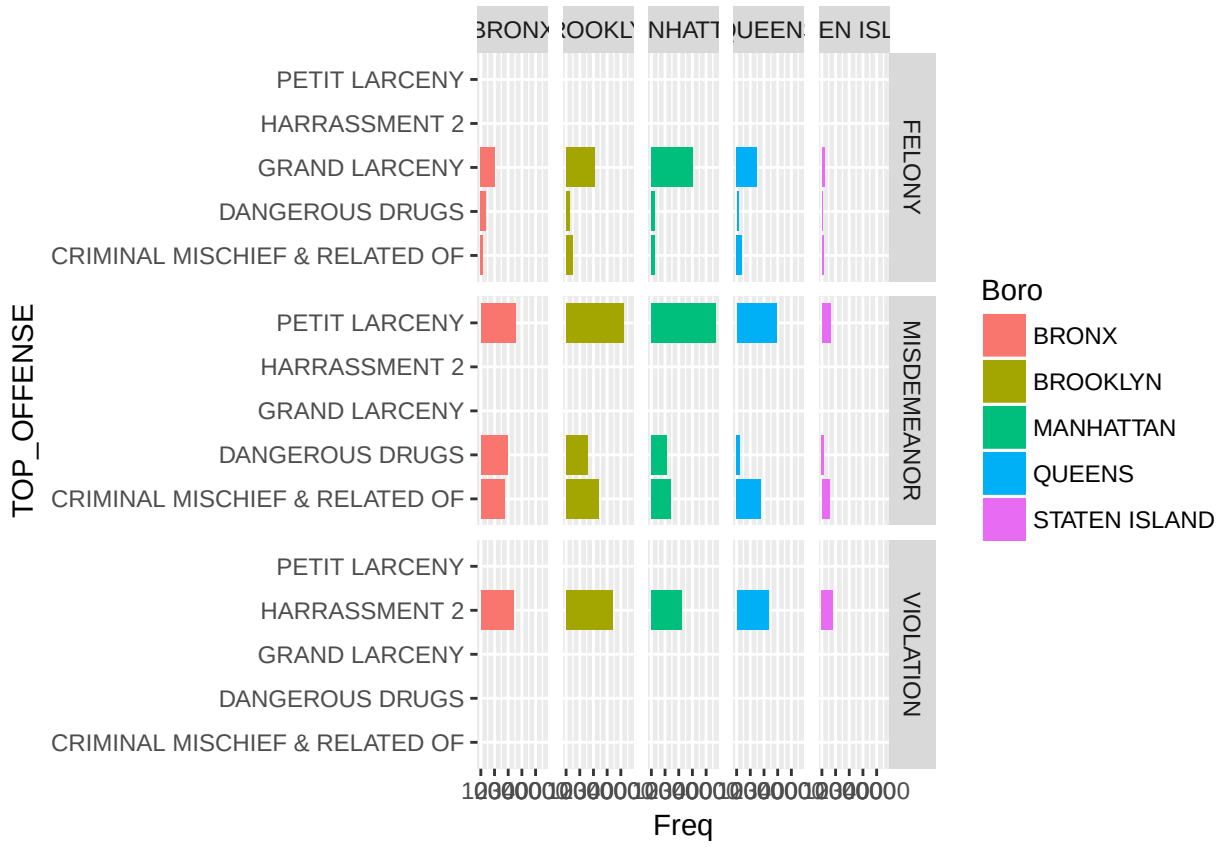

Top Crime Categories
PETIT LARCENY
HARRASSMENT 2
ASSAULT 3 & RELATED OFFENSES
CRIMINAL MISCHIEF & RELATED OF
GRAND LARCENY
DANGEROUS DRUGS
OFF. AGNST PUB ORD SENSBLTY &
ROBBERY
FELONY ASSAULT
BURGLARY
DANGEROUS WEAPONS
MISCELLANEOUS PENAL LAW
OFFENSES AGAINST PUBLIC ADMINI
GRAND LARCENY OF MOTOR VEHICLE
INTOXICATED & IMPAIRED DRIVING
CRIMINAL TRESPASS
VEHICLE AND TRAFFIC LAWS
THEFT−FRAUD
SEX CRIMES
FORGERY
FRAUDS
POSSESSION OF STOLEN PROPERTY
OFFENSES INVOLVING FRAUD
UNAUTHORIZED USE OF A VEHICLE
RAPE
ARSON
OFFENSES AGAINST THE PERSON
ADMINISTRATIVE CODE
OTHER OFFENSES RELATED TO THEF
OTHER STATE LAWS (NON PENAL LA
MURDER & NON−NEGL. MANSLAUGHTER
NYS LAWS−UNCLASSIFIED FELONY
THEFT OF SERVICES
FRAUDULENT ACCOSTING
BURGLAR'S TOOLS
KIDNAPPING & RELATED OFFENSES
GAMBLING
OFFENSES RELATED TO CHILDREN
PROSTITUTION & RELATED OFFENSES
PETIT LARCENY OF MOTOR VEHICLE
DISORDERLY CONDUCT
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
HOMICIDE−NEGLIGENT−VEHICLE
NEW YORK CITY HEALTH CODE
UNLAWFUL POSS. WEAP. ON SCHOOL
DISRUPTION OF A RELIGIOUS SERV
INTOXICATED/IMPAIRED DRIVING
HOMICIDE−NEGLIGENT,UNCLASSIFIE
LOITERING FOR DRUG PURPOSES
LOITERING
NYS LAWS−UNCLASSIFIED VIOLATION
ADMINISTRATIVE CODES
LOITERING/DEVIATE SEX
OTHER TRAFFIC INFRACTION
FORTUNE TELLING
KIDNAPPING
OTHER STATE LAWS (NON PENAL LAW)
ABORTION
UNDER THE INFLUENCE OF DRUGS
OFFENSES AGAINST MARRIAGE UNCL
KIDNAPPING AND RELATED OFFENSES
reorder(OffenseDesc, count)
0
250000
500000
750000
count

Boro, Juris, Crime Categories

```
                #group_by(Boro,Jurisdiction,OffenseDesc) %>%
                #mutate(count=n()) %>%
top_ofns   <- c("PETIT LARCENY", "HARRASSMENT 2", "CRIMINAL MISCHIEF & RELATED OF", "ASSAUALT 3 & REL
crime_sort <- crime_df %>%
                filter(Boro != "", Jurisdiction !="", (OffenseDesc == top_ofns))%>%
                group_by(Boro,Level,OffenseDesc) %>%
                summarize(Freq=n())

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

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

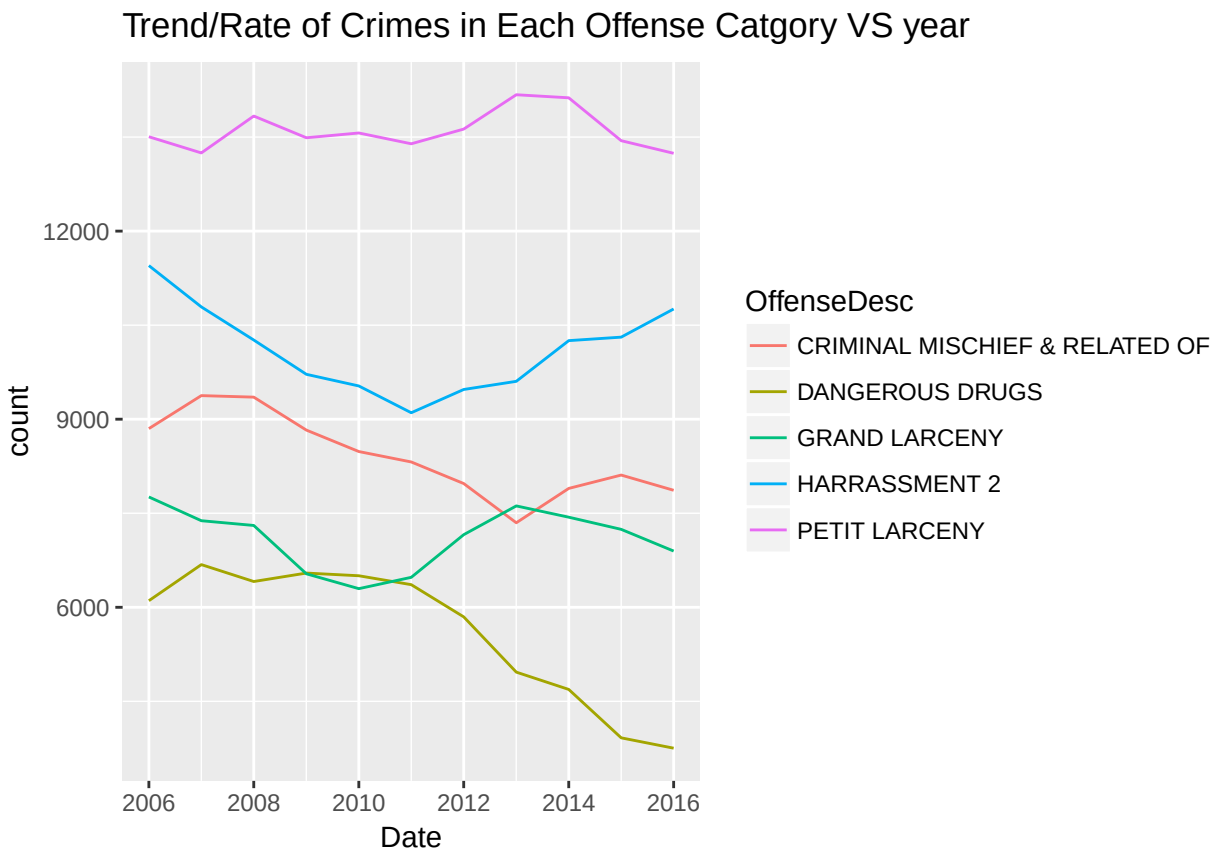



```
#doubledecker(TOP_OFFENSE~Boro, data=crime_sort)
ggplot(crime_sort, aes(TOP_OFFENSE,Freq, fill=Boro)) +
  geom_col() +
  facet_grid(Level~ Boro) +
  coord_flip()
```

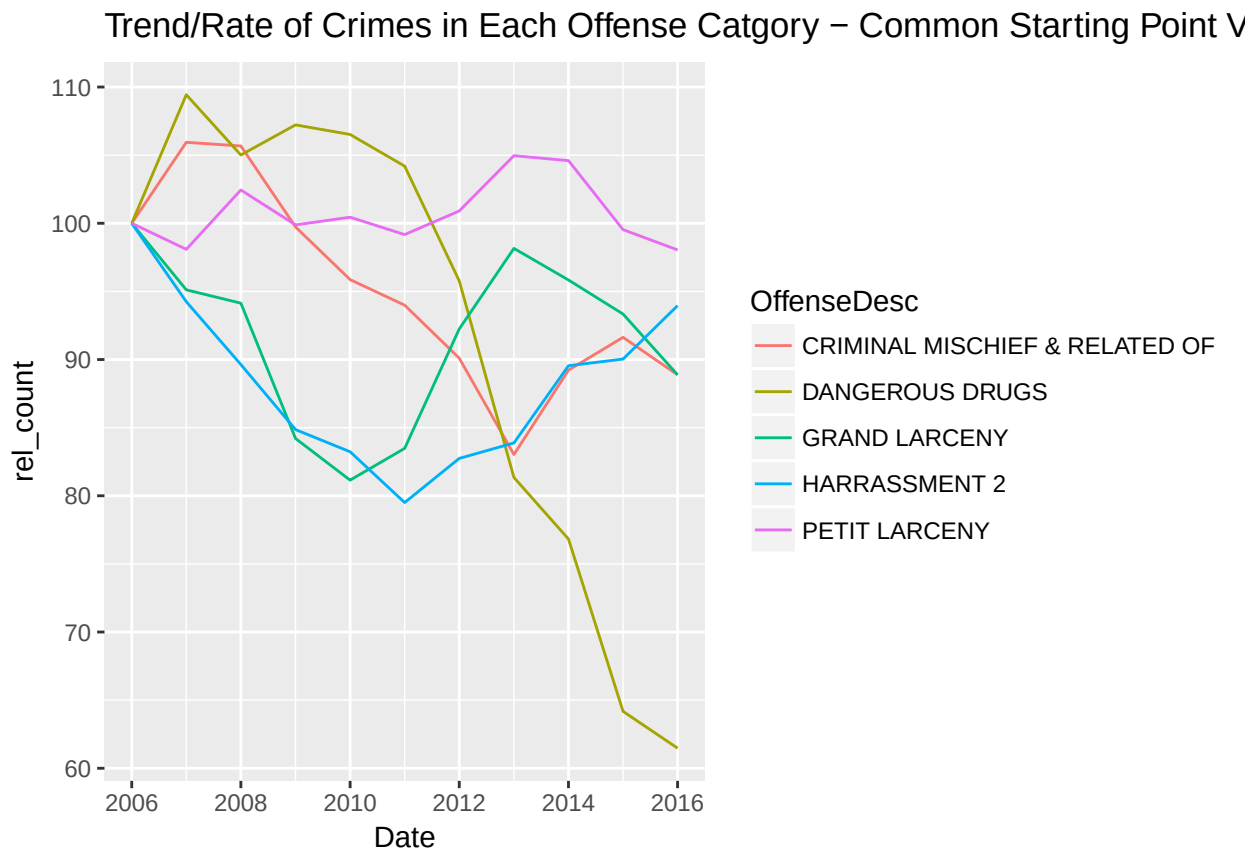


Time Trend of Top OFFENSE Category

```
top_ofns   <- c("PETIT LARCENY", "HARRASSMENT 2", "CRIMINAL MISCHIEF & RELATED OF", "ASSAUALT 3 & RE
crime_time_top_ofns <- crime_df %>%
                        filter(year(DateStart)>2005, (OffenseDesc == top_ofns)) %>%
                        group_by(Date=floor_date(DateStart, "year"),OffenseDesc) %>%
                        summarize(count=n())

crime_time_top_ofns <- crime_time_top_ofns %>%
                        group_by(OffenseDesc) %>%
                        mutate(rel_count = count*100/count[1])

ggplot(crime_time_top_ofns, aes(Date,count, color = OffenseDesc))+
      geom_line() +
      ggtitle("Trend/Rate of Crimes in Each Offense Catgory VS year")
```

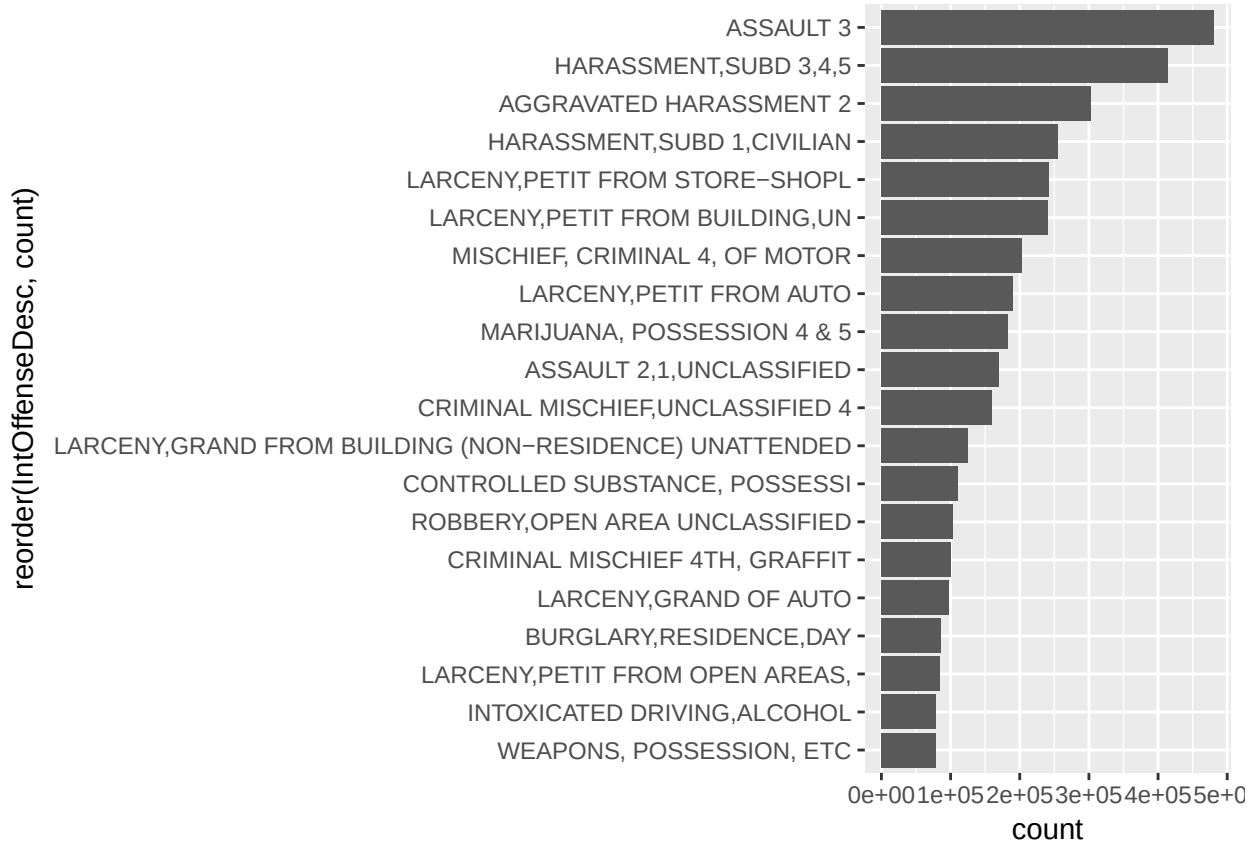



```
ggplot(crime_time_top_ofns, aes(Date,rel_count, color = OffenseDesc))+
      geom_line() +
      ggtitle("Trend/Rate of Crimes in Each Offense Catgory - Common Starting Point VS year")
```

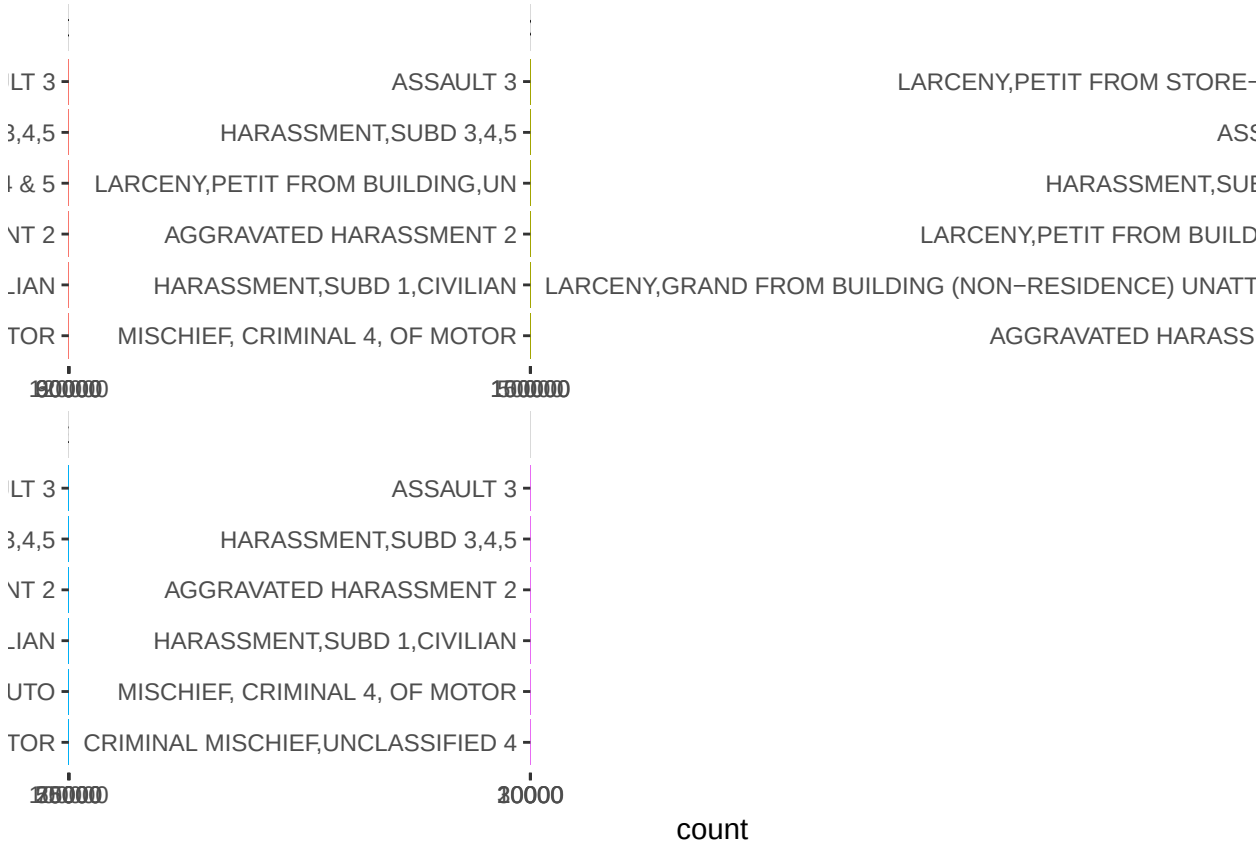


Crime PD top

```
crime_pd_top_all_boro <- crime_df %>%
                  filter(!is.na(Boro)) %>%
                  filter(IntOffenseDesc != "" && !is.na(IntOffenseDesc) && OffenseDesc != "" && Boro
                  group_by(Boro,IntOffenseDesc) %>%
                  summarize(count = n()) %>%
                  top_n(n=6, wt=count) %>%
                  arrange(Boro, desc(count))


crime_pd_top <- crime_df %>%
                  filter(IntOffenseDesc != "" && !is.na(IntOffenseDesc)  && (Boro != "")) %>%
                  group_by(IntOffenseDesc) %>%
                  summarize(count = n()) %>%
                  top_n(n=20, wt=count)


ggplot(crime_pd_top, aes(reorder(IntOffenseDesc,count), count)) +
  geom_col() +
  coord_flip()
```



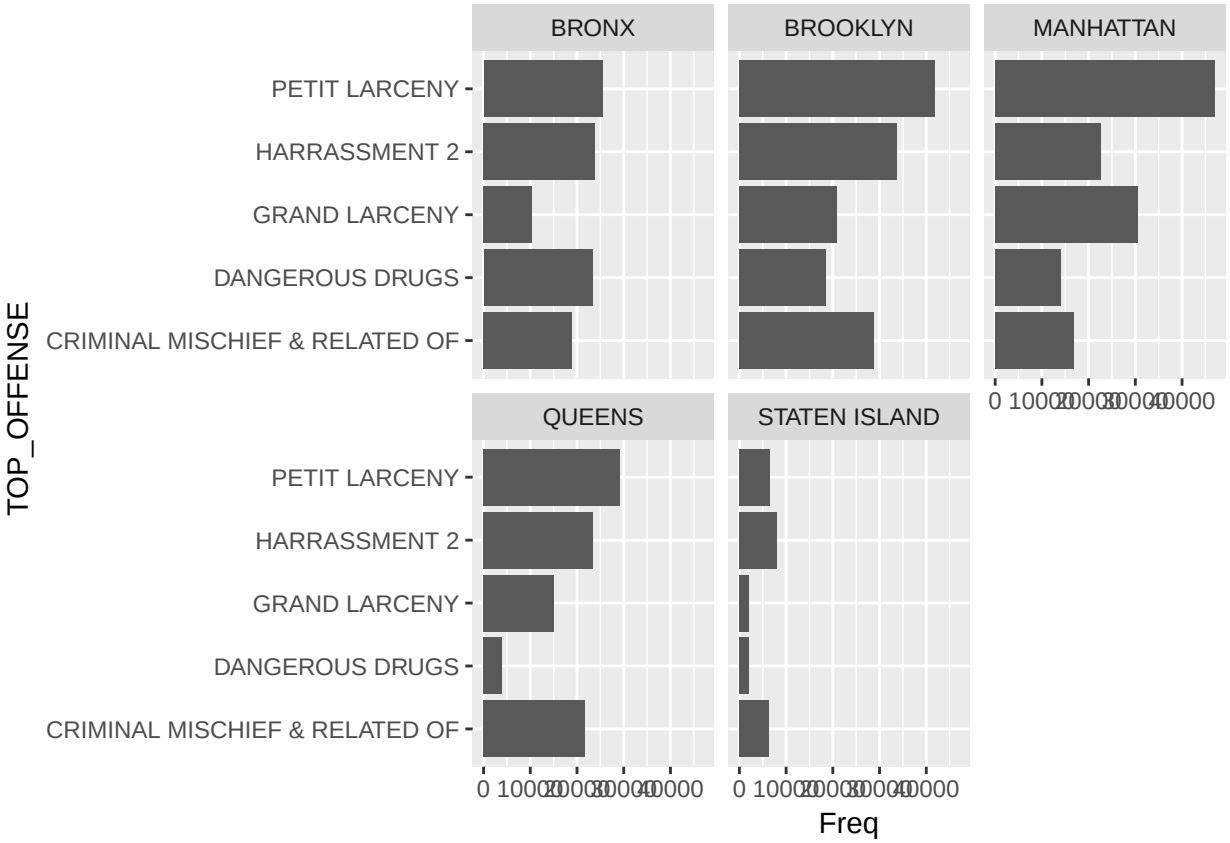

```
ggplot(crime_pd_top_all_boro, aes(reorder(IntOffenseDesc,count),count, fill=Boro)) +
  geom_col() +
  coord_flip() +
  facet_wrap(~Boro, scales="free")
```



total classification of overall crimes (pd_desc) -> 409

** 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!! **

```
## use freq density here
ggplot(crime_sort, aes(TOP_OFFENSE,Freq)) +
         geom_col() +
         facet_wrap(~Boro) +
         coord_flip()
```



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

```
crime_parks <- crime_df %>%
                filter(Boro!="",ParkName!="",Level!="") %>%
                group_by(Boro,ParkName,Level) %>%
                summarize(count=n())

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

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

crime_parks_1 <- crime_df %>%
                filter(Boro!="",ParkName!="",OffenseDesc!="") %>%
                group_by(Boro,ParkName,OffenseDesc) %>%
                summarize(count=n())

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

getPalette = colorRampPalette(brewer.pal(18, "Set1"))

ggplot(crime_pk_1 ,aes(ParkName, count, fill=OffenseDesc)) +
  geom_col() +
```

```
    facet_wrap(~Boro, ncol=1, scales="free_y") +
    scale_fill_manual(values = getPalette(18)) +
    # scale_fill_brewer(palette="Set3") +
    coord_flip()
```


VAN CORTLANDT PARK
ST. MARY'S PARK BRONX
MACOMBS DAM PARK
FRANZ SIGEL PARK
CROTONA PARK
CLAREMONT PARK
BRONX PARK
AQUEDUCT WALK
PROSPECT PARK
MCCARREN PARK
DYKER BEACH PARK
CONEY ISLAND BEACH & BOARDWALK
BROOKLYN BRIDGE PARK
WASHINGTON SQUARE PARK
RIVERSIDE PARK
RANDALL'S ISLAND PARK
MARCUS GARVEY PARK
HARLEM RIVER PARK
CENTRAL PARK
ROCKAWAY BEACH AND BOARDWALK
FOREST PARK
FLUSHING MEADOWS CORONA PARK
TOMPKINSVILLE PARK
TAPPEN PARK
STAPLETON PLAYGROUND
OCEAN BREEZE PARK
LT. LIA PLAYGROUND
FRANKLIN D. ROOSEVELT BOARDWALK AND BEACH
FABER POOL AND PARK
DUGAN PLAYGROUND
DE MATTI PLAYGROUND
CPL. THOMPSON PARK
CONFERENCE HOUSE PARK
CLOVE LAKES PARK
BLOOMINGDALE PARK
ParkName
count
OffenseDesc
ASSAULT 3 & RELATED OFFENSES
BURGLARY
CRIMINAL MISCHIEF & RELATED OF
CRIMINAL TRESPASS
DANGEROUS DRUGS
DANGEROUS WEAPONS
FELONY ASSAULT
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

```r
ggplot(crime_pk_1 ,aes(ParkName, count, fill=OffenseDesc)) +
  geom_col() +
  facet_wrap(~Boro, ncol=1, scales="free") +
  scale_fill_manual(values = getPalette(18)) +
  coord_flip()
```


VAN CORTLANDT PARK
ST. MARY'S PARK BRONX
MACOMBS DAM PARK
FRANZ SIGEL PARK
CROTONA PARK
CLAREMONT PARK
BRONX PARK
AQUEDUCT WALK
PROSPECT PARK
MCCARREN PARK
DYKER BEACH PARK
CONEY ISLAND BEACH & BOARDWALK
BROOKLYN BRIDGE PARK
WASHINGTON SQUARE PARK
RIVERSIDE PARK
RANDALL'S ISLAND PARK
MARCUS GARVEY PARK
HARLEM RIVER PARK
CENTRAL PARK
ROCKAWAY BEACH AND BOARDWALK
FOREST PARK
FLUSHING MEADOWS CORONA PARK
TOMPKINSVILLE PARK
TAPPEN PARK
STAPLETON PLAYGROUND
OCEAN BREEZE PARK
LT. LIA PLAYGROUND
FRANKLIN D. ROOSEVELT BOARDWALK AND BEACH
FABER POOL AND PARK
DUGAN PLAYGROUND
DE MATTI PLAYGROUND
CPL. THOMPSON PARK
CONFERENCE HOUSE PARK
CLOVE LAKES PARK
BLOOMINGDALE PARK
ParkName
count
OffenseDesc
ASSAULT 3 & RELATED OFFENSES
BURGLARY
CRIMINAL MISCHIEF & RELATED OF
CRIMINAL TRESPASS
DANGEROUS DRUGS
DANGEROUS WEAPONS
FELONY ASSAULT
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

trial on ggmap

```
library(ggmap)

#NYC <- get_map(location = "new york city",  color = "bw", zoom = 15, source = "google")
#ggmap(NYC)
#
#ggplot()+geom_point(data = crime_df, aes(x = Longitude, y = Latitude ,colour = factor(Level)))
```

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