# Exploratory Data Analysis, Assignment 1

Gayathri Prerepa (gp2254)

February 19, 2022

# This is the exploratory data analysis of NYC crime complaints.

# - The first step in any Exploratory Analysis is to ask questions.

# - Then find the patterns and trends with the help of graphical and statistical methods in the data to make informed decisions and to answer the question.

# - This includes finding how many records there are total, looking at the anomalies of the data, and looking for what information could be interesting for further analysis.

# We’ll address the following questions as we go ahead in this EDA:

# 1. Summary of the dataset

# 2. How much crime does each borough have?

# 3. Which borough has the highest number of crimes?

# 4. Visualisation of crimes each month and categorised by borough

# 5. Complaints per borough as a pie chart

# 6. What are the different crime categories (level of offense)

# 7. Which departments are solving crimes by month?

# 8. How many of each of the crimes have taken place?

# 9. How many victims are present in each age group?

# 10. Which age group is most likely to be a victim?

# 11. How many crimes are happening between 12am - 6am

# 12. How many Murders take place in this time frame?

# 13. How many Murders happening between 12 am to 6am per age group

# 14. Distribution of the co-ordinates of the above crimes

# 15. How many street crimes occur by borough?

# 16. What are the races of victims of street crimes?

# EXPLORARTORY DATA ANALYSIS OF NYC CRIME COMPLAINTS:

# 1. Summary of the dataset:

# 1.1 Loading all the required libraries

```
library(ggplot2)
library(dplyr)
library(tidyverse)
library(lubridate)
library(plotly)
library(gridExtra)
```

# 1.2 Read the data into dataframe

```
data <- read.csv("NYC_complaints.csv")
head(data)
```

| | CMPLNT… | ADDR_PC… | BOR… | CMPLNT_F… | CMPLNT_F… | CMPLNT_T… | CMPLNT_T… |
|---|---|---|---|---|---|---|---|
| | <int> | <int> | <chr> | <chr> | <chr> | <chr> | <chr> |
| 1 | 903695881 | 69 | | 12/17/2021 | 22:13:00 | | |
| 2 | 400462399 | 113 | | 12/17/2021 | 6:21:00 | | |
| 3 | 587910690 | 113 | | 12/13/2021 | 20:05:00 | | |
| 4 | 186105368 | 52 | BRONX | 12/7/2021 | 22:49:00 | | |
| 5 | 185325394 | 113 | | 12/6/2021 | 17:25:00 | | |

| | CMPLNT… | ADDR_PC… | BOR… | CMPLNT_F… | CMPLNT_F… | CMPLNT_T… | CMPLNT_T… |
|---|---|---|---|---|---|---|---|
| | <int> | <int> | <chr> | <chr> | <chr> | <chr> | <chr> |
| 6 | 791525475 | 44 | | 12/5/2021 | 22:16:00 | | |

6 rows | 1-8 of 37 columns

# The data includes diverse datatypes - this analysis includes datetimes, multi-layered factors, and even latitude/longitude point data

# 1.3 Printing dimension of the dataset

```
glimpse(data)
```

```
## Rows: 449,506
## Columns: 36
## $ CMPLNT_NUM                <int> 903695881, 400462399, 587910690, 186105368, 1~
## $ ADDR_PCT_CD               <int> 69, 113, 113, 52, 113, 44, 47, 46, 75, 73, 10~
## $ BORO_NM                   <chr> "", "", "", "BRONX", "", "", "BRONX", "", "",~
## $ CMPLNT_FR_DT              <chr> "12/17/2021", "12/17/2021", "12/13/2021", "12~
## $ CMPLNT_FR_TM              <chr> "22:13:00", "6:21:00", "20:05:00", "22:49:00"~
## $ CMPLNT_TO_DT              <chr> "", "", "", "", "", "", "", "", "", "", "", "~
## $ CMPLNT_TO_TM              <chr> "", "", "", "", "", "", "", "", "", "", "", "~
## $ CRM_ATPT_CPTD_CD          <chr> "COMPLETED", "COMPLETED", "COMPLETED", "COMPL~
## $ HADEVELOPT                <chr> "", "", "", "", "", "", "", "", "", "", "", "~
## $ HOUSING_PSA               <int> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, N~
## $ JURISDICTION_CODE         <int> NA, NA, NA, 0, NA, NA, 0, NA, NA, NA, NA, NA,~
## $ JURIS_DESC                <chr> "N.Y. POLICE DEPT", "N.Y. POLICE DEPT", "N.Y.~
## $ KY_CD                     <int> 101, 101, 101, 118, 101, 101, 118, 101, 101, ~
## $ LAW_CAT_CD                <chr> "FELONY", "FELONY", "FELONY", "FELONY", "FELO~
## $ LOC_OF_OCCUR_DESC         <chr> "OUTSIDE", "OUTSIDE", "OUTSIDE", "", "INSIDE"~
## $ OFNS_DESC                 <chr> "MURDER & NON-NEGL. MANSLAUGHTER", "MURDER & ~
## $ PARKS_NM                  <chr> "", "", "", "", "", "", "", "", "", "", "", "~
## $ PATROL_BORO               <chr> "", "", "", "PATROL BORO BRONX", "", "", "PAT~
## $ PD_CD                     <int> NA, NA, NA, 792, NA, NA, 792, NA, NA, NA, NA,~
## $ PD_DESC                   <chr> "", "", "", "WEAPONS POSSESSION 1 & 2", "", "~
## $ PREM_TYP_DESC             <chr> "", "", "", "STREET", "", "", "STREET", "", "~
## $ RPT_DT                    <chr> "12/17/2021", "12/17/2021", "12/13/2021", "12~
## $ STATION_NAME              <chr> "", "", "", "", "", "", "", "", "", "", "", "~
## $ SUSP_AGE_GROUP            <chr> "25-44", "", "", "", "25-44", "", "", "25-44"~
## $ SUSP_RACE                 <chr> "BLACK", "", "", "", "BLACK", "", "", "BLACK"~
## $ SUSP_SEX                  <chr> "M", "", "", "", "M", "", "", "M", "", "M", "~
## $ TRANSIT_DISTRICT          <int> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, N~
## $ VIC_AGE_GROUP             <chr> "25-44", "25-44", "25-44", "UNKNOWN", "25-44"~
## $ VIC_RACE                  <chr> "BLACK", "BLACK", "BLACK", "UNKNOWN", "BLACK"~
## $ VIC_SEX                   <chr> "M", "F", "M", "E", "M", "M", "E", "F", "M", ~
## $ X_COORD_CD                <int> 1011203, 1043252, 1042087, 1017088, 1046176, ~
## $ Y_COORD_CD                <int> 174515, 187998, 190443, 260895, 193100, 24603~
## $ Latitude                  <dbl> 40.64565, 40.68250, 40.68922, 40.88272, 40.69~
## $ Longitude                 <dbl> -73.90288, -73.78727, -73.79145, -73.88125, -~
## $ Lat_Lon                   <chr> "(40.64564719600002, -73.90287588699994)", "(~
## $ New.Georeferenced.Column <chr> "POINT (-73.90287588699994 40.6456471960002)~
```

# 1.4 Dropping columns that are not in immediate use

```
df <- data[ -c(2,6,7,9,10,11,13,15,17,18,19,20,23,27,31,32,36) ]
head(df)
```

| CMPLNT… | BOR… | CMPLNT_F… | CMPLNT_F… | CRM_ATPT_CPT… | JURIS_DESC | LAW |
|---|---|---|---|---|---|---|
| <int> | <chr> | <chr> | <chr> | <chr> | <chr> | <chr |

| | CMPLNT…<br><int> | BOR…<br><chr> | CMPLNT_F…<br><chr> | CMPLNT_F…<br><chr> | CRM_ATPT_CPT…<br><chr> | JURIS_DESC<br><chr> | LAW<br><chr |
|---|---|---|---|---|---|---|---|
| 1 | 903695881 | | 12/17/2021 | 22:13:00 | COMPLETED | N.Y. POLICE DEPT | FEL |
| 2 | 400462399 | | 12/17/2021 | 6:21:00 | COMPLETED | N.Y. POLICE DEPT | FEL |
| 3 | 587910690 | | 12/13/2021 | 20:05:00 | COMPLETED | N.Y. POLICE DEPT | FEL |
| 4 | 186105368 | BRONX | 12/7/2021 | 22:49:00 | COMPLETED | N.Y. POLICE DEPT | FEL |
| 5 | 185325394 | | 12/6/2021 | 17:25:00 | COMPLETED | N.Y. POLICE DEPT | FEL |
| 6 | 791525475 | | 12/5/2021 | 22:16:00 | COMPLETED | N.Y. POLICE DEPT | FEL |

6 rows | 1-8 of 20 columns

# 1.5 Renaming the columns for better understanding and analysis

```
names(df) <- c("ID","Borough","Date","Time","Crime Status","Jurisdiction","Level of offense", "O
ffense", "Premise" , "Report Date", " Suspect age", "Suspect race","Suspect sex", "Victim age",
" Victim race", " Victim sex","Latitude", "Longtitude", "Cordinates")
head(df)
```

| | ID<br><int> | Boro…<br><chr> | Date<br><chr> | Time<br><chr> | Crime Status<br><chr> | Jurisdiction<br><chr> | Level of offense<br><chr> |
|---|---|---|---|---|---|---|---|
| 1 | 903695881 | | 12/17/2021 | 22:13:00 | COMPLETED | N.Y. POLICE DEPT | FELONY |
| 2 | 400462399 | | 12/17/2021 | 6:21:00 | COMPLETED | N.Y. POLICE DEPT | FELONY |
| 3 | 587910690 | | 12/13/2021 | 20:05:00 | COMPLETED | N.Y. POLICE DEPT | FELONY |
| 4 | 186105368 | BRONX | 12/7/2021 | 22:49:00 | COMPLETED | N.Y. POLICE DEPT | FELONY |
| 5 | 185325394 | | 12/6/2021 | 17:25:00 | COMPLETED | N.Y. POLICE DEPT | FELONY |
| 6 | 791525475 | | 12/5/2021 | 22:16:00 | COMPLETED | N.Y. POLICE DEPT | FELONY |

6 rows | 1-8 of 20 columns

# 1.6 Printing dimensions and details of the new dataset

```
glimpse(df)
```

```
## Rows: 449,506
## Columns: 19
## $ ID                  <int> 903695881, 400462399, 587910690, 186105368, 1853253~
## $ Borough             <chr> "", "", "", "BRONX", "", "", "BRONX", "", "", "", "~
## $ Date                <chr> "12/17/2021", "12/17/2021", "12/13/2021", "12/7/202~
## $ Time                <chr> "22:13:00", "6:21:00", "20:05:00", "22:49:00", "17:~
## $ `Crime Status`      <chr> "COMPLETED", "COMPLETED", "COMPLETED", "COMPLETED",~
## $ Jurisdiction        <chr> "N.Y. POLICE DEPT", "N.Y. POLICE DEPT", "N.Y. POLIC~
## $ `Level of offense`  <chr> "FELONY", "FELONY", "FELONY", "FELONY", "FELONY", "~
## $ Offense             <chr> "MURDER & NON-NEGL. MANSLAUGHTER", "MURDER & NON-NE~
## $ Premise             <chr> "", "", "", "STREET", "", "", "STREET", "", "", "",~
## $ `Report Date`       <chr> "12/17/2021", "12/17/2021", "12/13/2021", "12/7/202~
## $ ` Suspect age`      <chr> "25-44", "", "", "", "25-44", "", "", "25-44", "", ~
## $ `Suspect race`      <chr> "BLACK", "", "", "", "BLACK", "", "", "BLACK", "", ~
## $ `Suspect sex`       <chr> "M", "", "", "", "M", "", "", "M", "", "M", "", "",~
## $ `Victim age`        <chr> "25-44", "25-44", "25-44", "UNKNOWN", "25-44", "18-~
## $ ` Victim race`      <chr> "BLACK", "BLACK", "BLACK", "UNKNOWN", "BLACK", "BLA~
## $ ` Victim sex`       <chr> "M", "F", "M", "E", "M", "M", "E", "F", "M", "M", "~
## $ Latitude            <dbl> 40.64565, 40.68250, 40.68922, 40.88272, 40.69648, 4~
## $ Longtitude          <dbl> -73.90288, -73.78727, -73.79145, -73.88125, -73.776~
## $ Cordinates          <chr> "(40.64564719600002, -73.90287588699994)", "(40.682~
```

# 1.7 Formatting the date to get date month and year separately

```
df$Date <- as.Date(as.character(df$Date), format = "%m/%d/%y")
df$date2 <- df$Date
df <- separate(df, col = date2, into = c("year","month","day"), sep ="-")
```

# One of the key aims of an EDA is to identify problems within the data and fix (“clean”) them where possible. This can make later analyses more accurate, and provide insight into ways that the data collection process could improve.

# 1.8 Replacing empty cells with value “NOT REPORTED”

```
df <- df%>%mutate_if(is.character, list(~na_if(.,"")))
df[is.na(df)]<- "NOT REPORTED"
head(df)
```

| | ID | Borough | Date | Time | Crime Status | Jurisdiction | Level of offe |
|---|---|---|---|---|---|---|---|
| | <int> | <chr> | <date> | <chr> | <chr> | <chr> | <chr> |
| 1 | 903695881 | NOT REPORTED | 2020-12-17 | 22:13:00 | COMPLETED | N.Y. POLICE DEPT | FELONY |
| 2 | 400462399 | NOT REPORTED | 2020-12-17 | 6:21:00 | COMPLETED | N.Y. POLICE DEPT | FELONY |
| 3 | 587910690 | NOT REPORTED | 2020-12-13 | 20:05:00 | COMPLETED | N.Y. POLICE DEPT | FELONY |
| 4 | 186105368 | BRONX | 2020-12-07 | 22:49:00 | COMPLETED | N.Y. POLICE DEPT | FELONY |
| 5 | 185325394 | NOT REPORTED | 2020-12-06 | 17:25:00 | COMPLETED | N.Y. POLICE DEPT | FELONY |
| 6 | 791525475 | NOT REPORTED | 2020-12-05 | 22:16:00 | COMPLETED | N.Y. POLICE DEPT | FELONY |

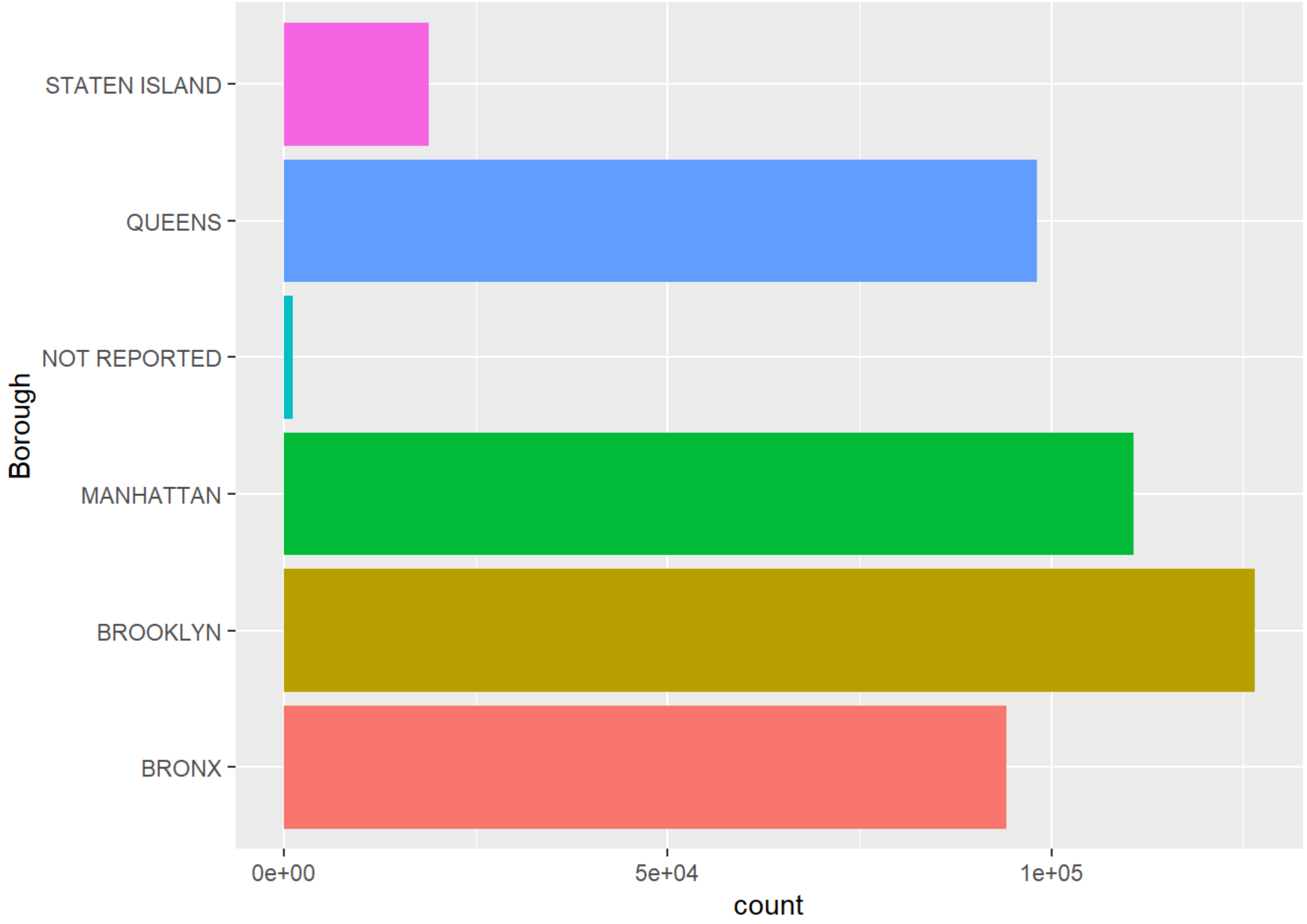
6 rows | 1-8 of 23 columns

# Doing this as no information is known to assign a default value to it. By doing this we can identify inconsistencies in the data and how it affects the analysis.

# 2. How much crime does each borough have?

# 2.1 Count of crimes in each borough

```
cf <- ggplot(df, aes(x = Borough, fill=as.factor(Borough))) + geom_bar(width=0.9, stat="count")
 + theme(legend.position="none") + coord_flip()
print(cf)
```



3. Which borough has the highest number of crimes?

- From the above visualisation, we can say that Brooklyn has the highest crimes among the boroughs and number of “not reported” crimes are less.

4. Visualisation of crimes each month, day and categorised by borough

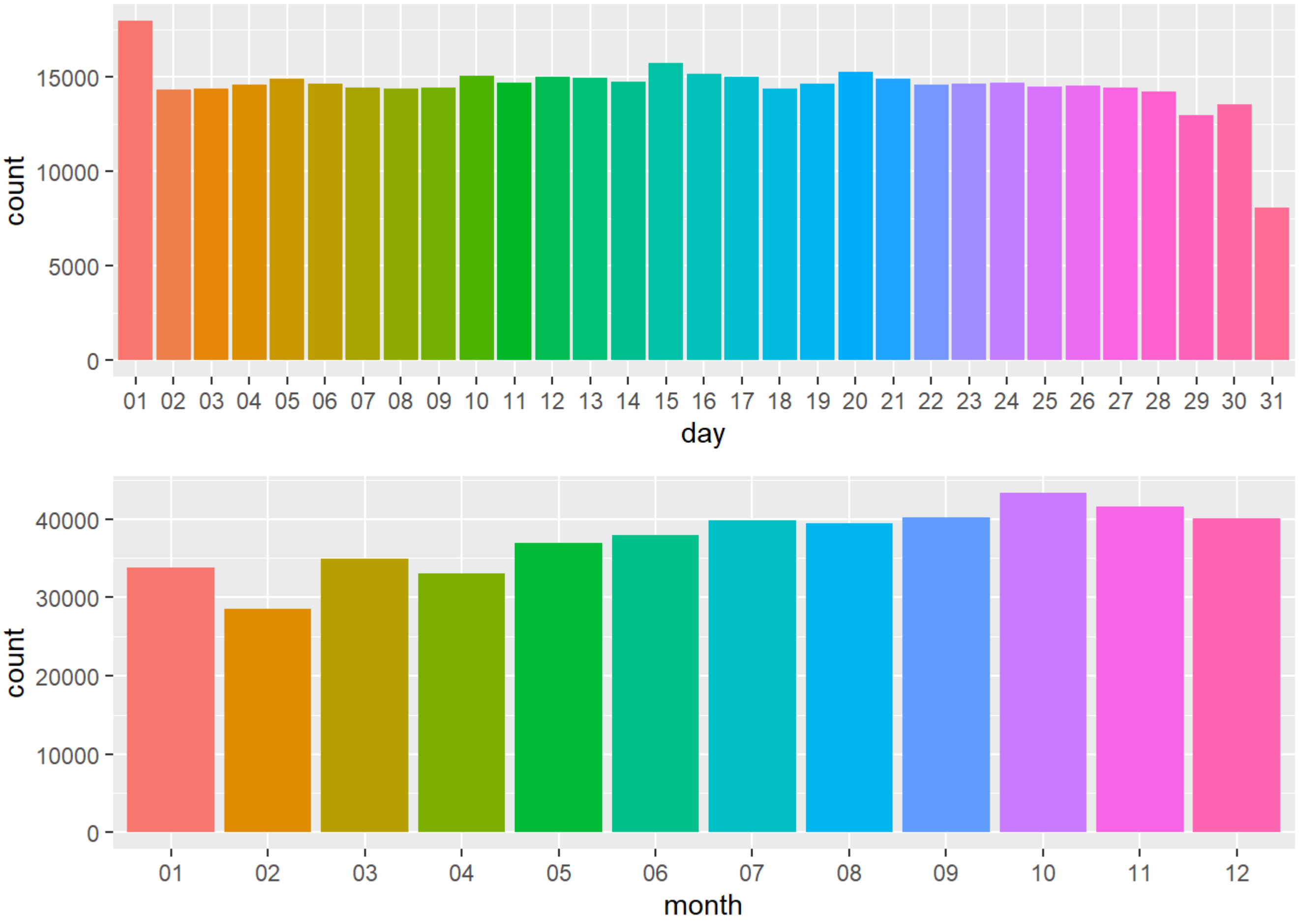
Let’s analyze the datetime fields formatted previously. Namely, crime record frequency by: Month of Year.

# 4.1 Count of crimes per month

```
mf <- ggplot(df, aes(x = month, fill=as.factor(month))) + geom_bar(width=0.9, stat="count") + theme(legend.position= "none")

dff <- ggplot(df, aes(x = day, fill=as.factor(day))) + geom_bar(width=0.9, stat="count") + theme(legend.position= "none")

grid.arrange(dff, mf)
```



The month of october, followed by november and december had the highest crimes with more than 40000 cases, and February the least with below 30000 complaints recorded.

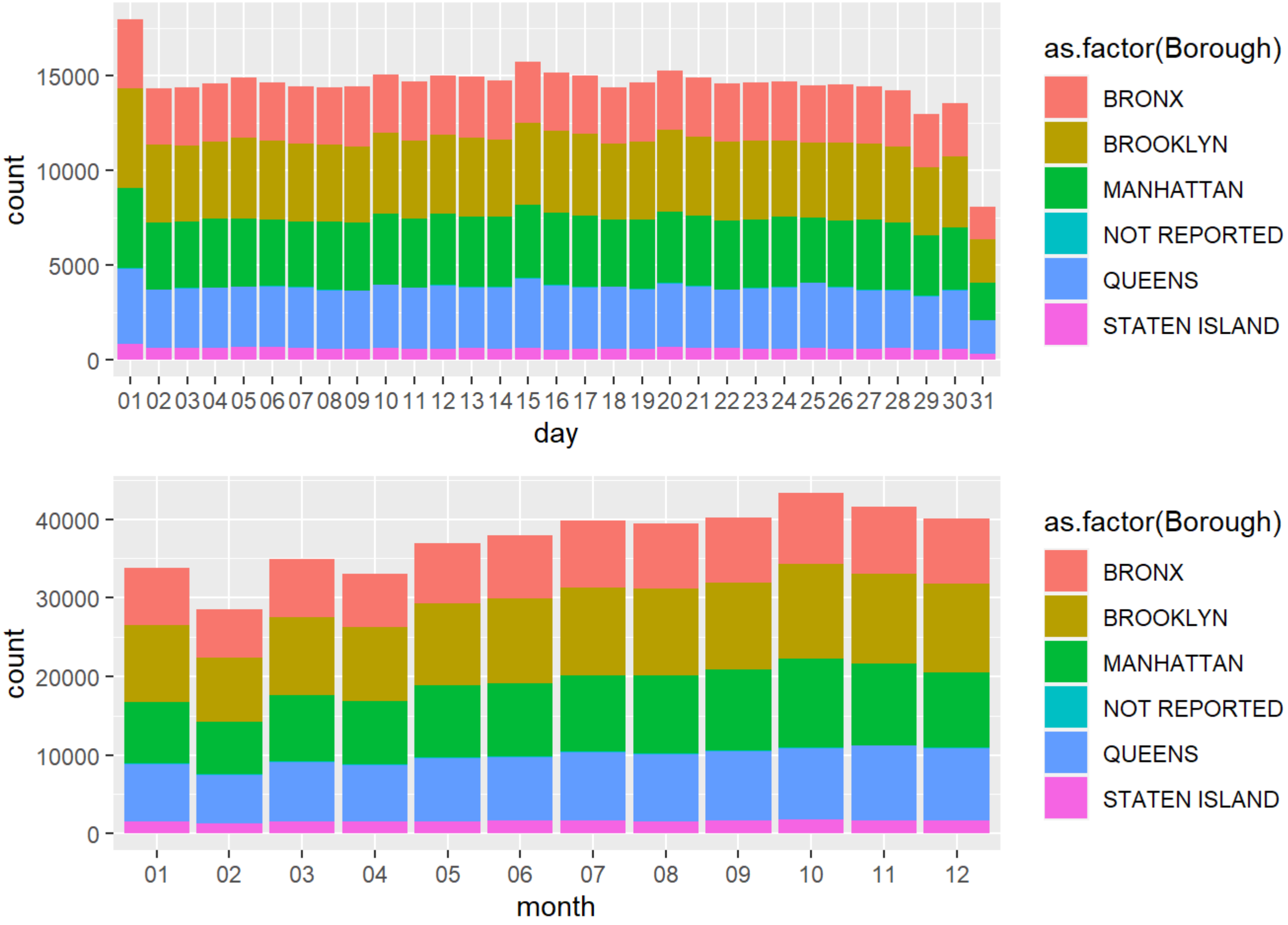
The first of every month has the most number of complaints with approximately 17500 complaints recorded and the last day of every month had the least compaints with nearly 7500 complaints, which is 10000 less than on first day.

## 4.2 Count of crimes per month and day categorised by borough

```
mb <- ggplot(df, aes(x = month, fill=as.factor(Borough))) + geom_bar(width=0.9, stat="count") +
 theme(legend.position= "right")

db <- ggplot(df, aes(x = day, fill=as.factor(Borough))) + geom_bar(width=0.9, stat="count") + th
eme(legend.position= "right")

grid.arrange(db, mb)
```

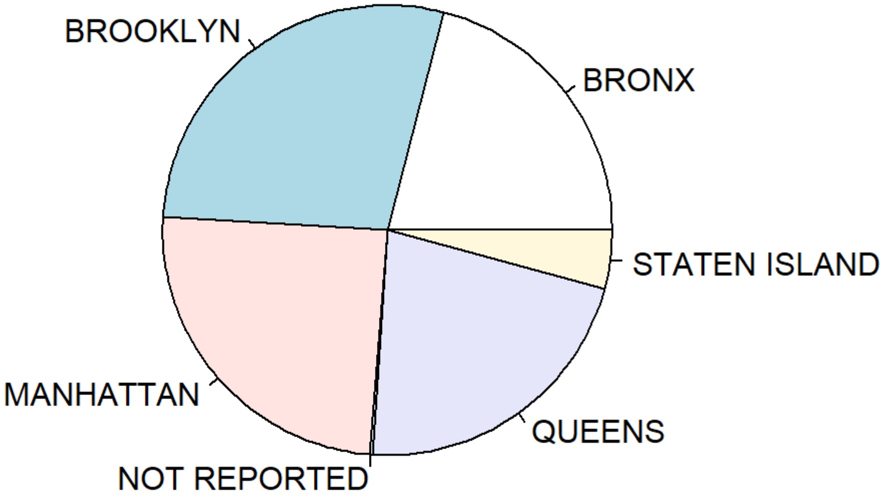


# - By observation, Brooklyn has the highest and Staten island has the least portion of crimes in most of the months and days.

# - All boroughs seem to have constant number of complaints per month throughout the year.

# 5. Complaints per borough as a pie chart

```
cnt <-pie(table(df$`Borough`))
```



#

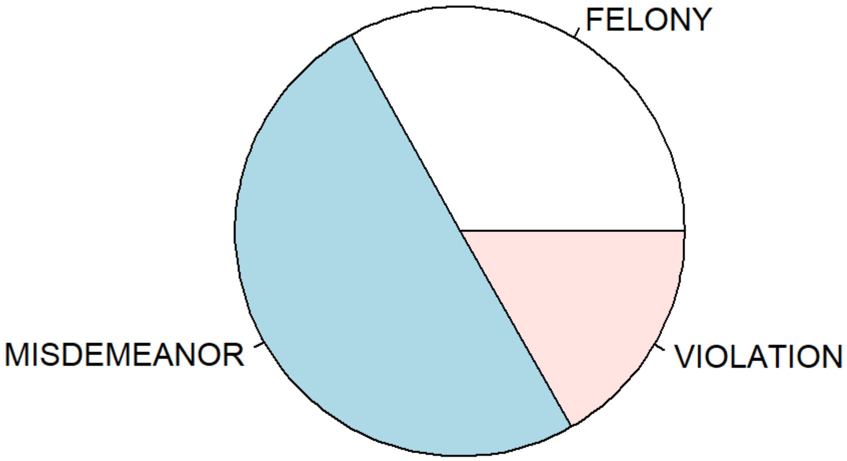
Portion of “Not reported” boroughs is negligible compared to the whole data available.

# 6. What are the different crime categories (level of offense)

# All offenses are split into 3 categories in the order: FELONY, MISDEMEANOR, and VIOLATION.

# In terms of crime severity, we can infer by names and the frequencies of these categories: FELONY is the most severe, then MISDEMEANOR and VIOLATION.

```
lc <- pie(table(df$`Level of offense`))
```



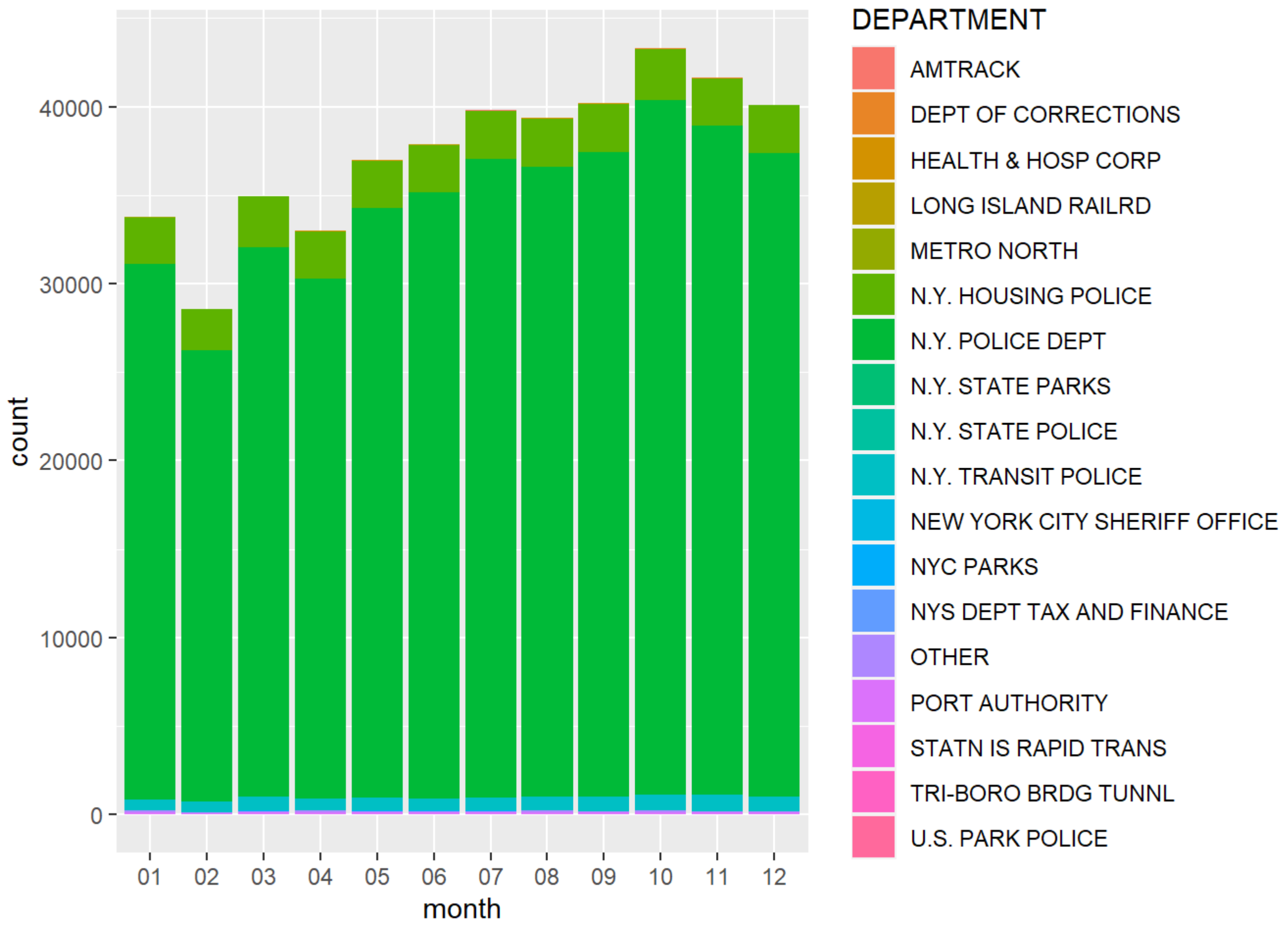
# - Misdemeanor is about 50% of offenses, felony and violation are 33% and 27% repectively

# 7. Which departments are solving crimes by month?

```
pd<-ggplot(data=df,aes(x=`month`,fill=`Jurisdiction`))+geom_histogram(stat="count")+ scale_fill_
discrete(name="DEPARTMENT")
```

```
## Warning: Ignoring unknown parameters: binwidth, bins, pad
```

```
print(pd)
```



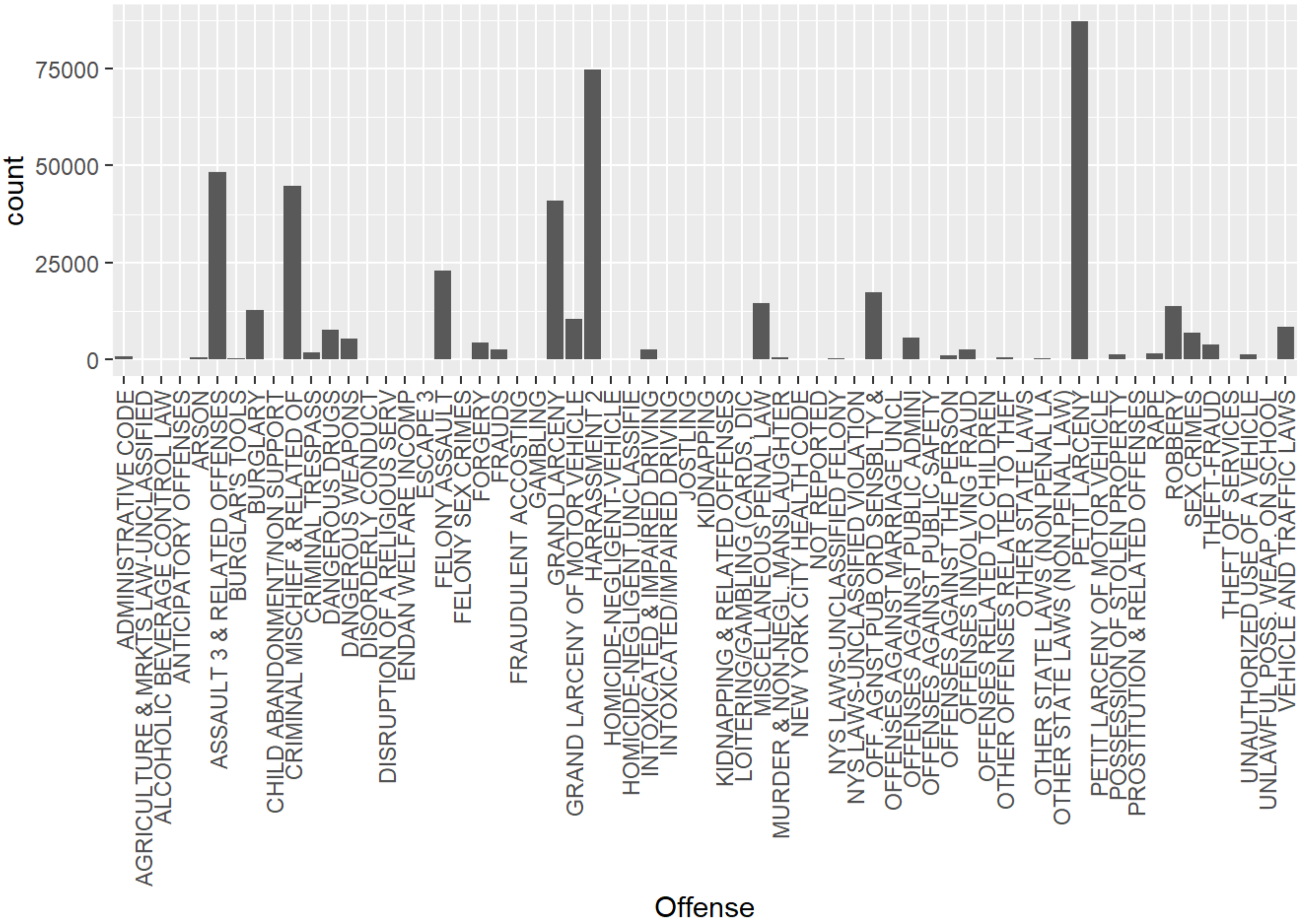
- Most of the crime complaints are taken up by the N.Y. Police Department.

# 8. How many of each of the crimes have taken place?

```
oc <- ggplot(df, aes(x=`Offense`))+geom_histogram(stat ="count")+theme(axis.text.x=element_text
(angle=90,hjust=1,vjust=0.5))
```

```
## Warning: Ignoring unknown parameters: binwidth, bins, pad
```

```
print(oc)
```



# - Petit Larceny followed by Harrasement 2 and assault are the highest complaints recorded overall

# 9. How many victims are present in each age group?

```
ag <- table(df$`Victim age`)
print(ag)
```

```
##
##      -1      -3      -4     -48     -51     -61     -62    -921    -935    -943
##       2       1       1       2       1       1       1       1       1       1
##    -960     <18   18-24   25-44   45-64     65+     936     945     963     970
##       1   14408   39869  166937   90170   21532       1       1       1       1
## UNKNOWN
##  116573
```

# - We can see that around 14 age group values do not align with reality, but this is a very small inconsistency comapred to the whole data.

# - But what is concering is the 116573 unknown values(almost 1/3 of the data), this will affect the accuracy.

# 10. Which age group is most likely to be a victim?

```
ag[ag==max(ag)]
```

```
##  25-44
## 166937
```

# - Of the known data, ages between 25 and 44 are most common victims of crimes with a count of 166937

# 10.1 Converting time column from character to time stamp

```
df$Time <- as.POSIXct(df$Time, format = "%H:%M:%S")
df$Time <- format(df$Time, "%H:%M:%S")
```

# 11. How many crimes are happening between 12am - 6am

```
df1 <- df %>% filter(Time < "06:00:00" & Time > "00:00:00")
head(df1)
```

| | ID | Borough | Date | Time | Crime Status | Jurisdiction | Level of offe |
|---|---|---|---|---|---|---|---|
| | <int> | <chr> | <date> | <chr> | <chr> | <chr> | <chr> |
| 1 | 276296223 | BRONX | 2020-12-01 | 00:01:00 | COMPLETED | N.Y. POLICE DEPT | FELONY |
| 2 | 694082264 | NOT REPORTED | 2020-10-02 | 01:25:00 | COMPLETED | N.Y. POLICE DEPT | FELONY |
| 3 | 474078722 | NOT REPORTED | 2020-09-16 | 05:15:00 | COMPLETED | N.Y. POLICE DEPT | FELONY |
| 4 | 843162354 | NOT REPORTED | 2020-09-15 | 01:13:00 | COMPLETED | N.Y. POLICE DEPT | FELONY |
| 5 | 715480292 | NOT REPORTED | 2020-08-26 | 01:34:00 | COMPLETED | N.Y. POLICE DEPT | FELONY |
| 6 | 644667510 | NOT REPORTED | 2020-08-14 | 04:11:00 | COMPLETED | N.Y. POLICE DEPT | FELONY |

6 rows | 1-8 of 23 columns

# 11.1 Printing dimension and calculating percentage

```
print(nrow(df1))
```

```
## [1] 64501
```

```
print(nrow(df1)*100/nrow(df))
```

```
## [1] 14.34931
```

# - There are totally 64501 crimes haepping between 12 am - 6 am, ie, past midnight and

# before early hours. This is 14.3% of all the crimes

# 12. How many Murders take place in this time frame?

```
df2 <- df1 %>% filter(df1$`Offense` == "MURDER & NON-NEGL. MANSLAUGHTER")
head(df2)
```

| | ID<br><int> | Borough<br><chr> | Date<br><date> | Time<br><chr> | Crime Status<br><chr> | Jurisdiction<br><chr> | Level of offe<br><chr> |
|---|---|---|---|---|---|---|---|
| 1 | 694082264 | NOT REPORTED | 2020-10-02 | 01:25:00 | COMPLETED | N.Y. POLICE DEPT | FELONY |
| 2 | 474078722 | NOT REPORTED | 2020-09-16 | 05:15:00 | COMPLETED | N.Y. POLICE DEPT | FELONY |
| 3 | 843162354 | NOT REPORTED | 2020-09-15 | 01:13:00 | COMPLETED | N.Y. POLICE DEPT | FELONY |
| 4 | 715480292 | NOT REPORTED | 2020-08-26 | 01:34:00 | COMPLETED | N.Y. POLICE DEPT | FELONY |
| 5 | 644667510 | NOT REPORTED | 2020-08-14 | 04:11:00 | COMPLETED | N.Y. POLICE DEPT | FELONY |
| 6 | 754095161 | NOT REPORTED | 2020-08-14 | 00:58:00 | COMPLETED | N.Y. POLICE DEPT | FELONY |

6 rows | 1-8 of 23 columns

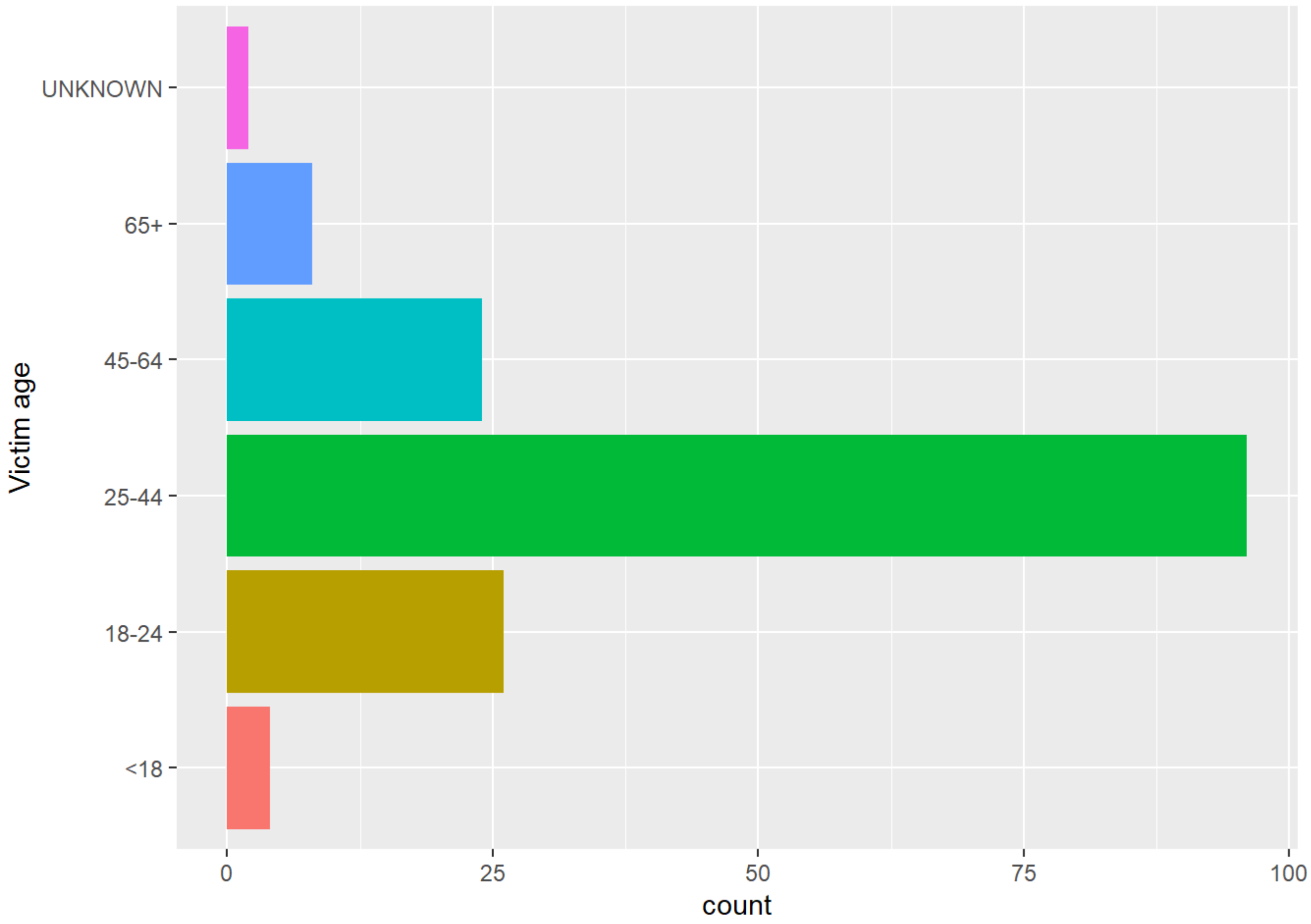
```
print(nrow(df2))
```

```
## [1] 160
```

# - 160 Of the 64501 crimes happening between 12 am 6 am are murders.

# 13. How many Murders happening between 12 am to 6 am per age group

```
mg<-ggplot(df2, aes(x =`Victim age`, fill=as.factor(`Victim age`))) + geom_bar(width=0.9, stat=
"count") + theme(legend.position="none") + coord_flip()
print(mg)
```

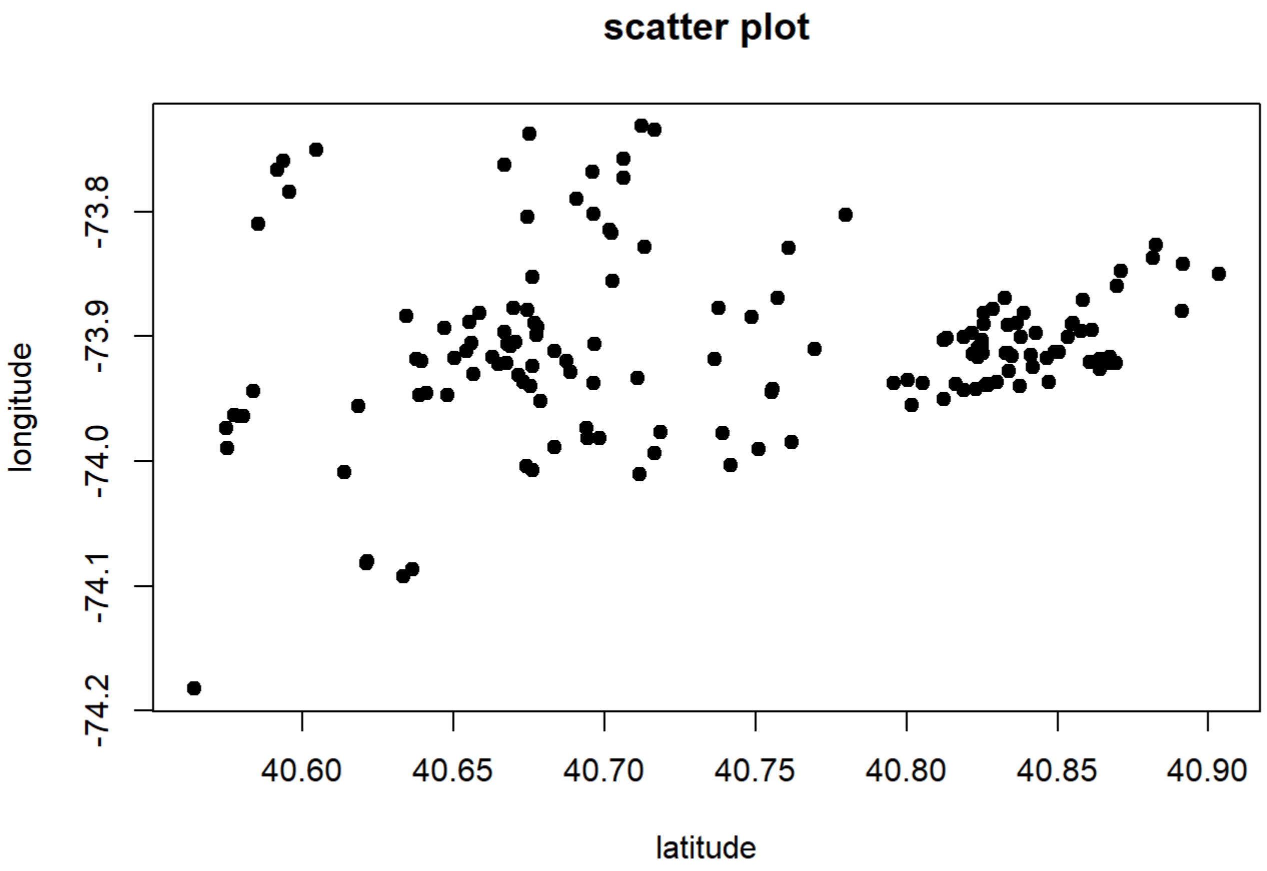


- Age groups between 25-44 are most vulnerable to muders after midnight and before daylight, and the least affected are below 18 and over 65. This could lead to some interesting and insightful analysis.

## 14. Distribution of the co-ordinates of the above crimes

### 14.1 Scatterplot:

```
x<- df2$Latitude
y<- df2$Longtitude
sp<-plot( x,y, main = "scatter plot", xlab = "latitude", ylab = "longitude", pch = 19)
```



# 15. How many street crimes occur by borough?

```
df3 <- df %>% filter(df$`Premise` == "STREET")
head(df3)
```

| | ID | Borough | Date | Time | Crime Status | Jurisdiction | Level of offense |
|---|---|---|---|---|---|---|---|
| | <int> | <chr> | <date> | <chr> | <chr> | <chr> | <chr> |
| 1 | 186105368 | BRONX | 2020-12-07 | 22:49:00 | COMPLETED | N.Y. POLICE DEPT | FELONY |
| 2 | 276296223 | BRONX | 2020-12-01 | 00:01:00 | COMPLETED | N.Y. POLICE DEPT | FELONY |
| 3 | 504183189 | MANHATTAN | 2020-09-12 | 16:55:00 | COMPLETED | N.Y. POLICE DEPT | FELONY |
| 4 | 180721837 | BRONX | 2020-07-06 | 16:15:00 | COMPLETED | N.Y. POLICE DEPT | FELONY |
| 5 | 300711837 | BRONX | 2020-02-13 | 15:15:00 | COMPLETED | N.Y. POLICE DEPT | MISDEMEANOR |
| 6 | 187909876 | MANHATTAN | 2020-01-27 | 03:05:00 | COMPLETED | N.Y. POLICE DEPT | FELONY |

6 rows | 1-8 of 23 columns

# 15.1 Calculating percentage

```
print(nrow(df3))
```

```
## [1] 127641
```

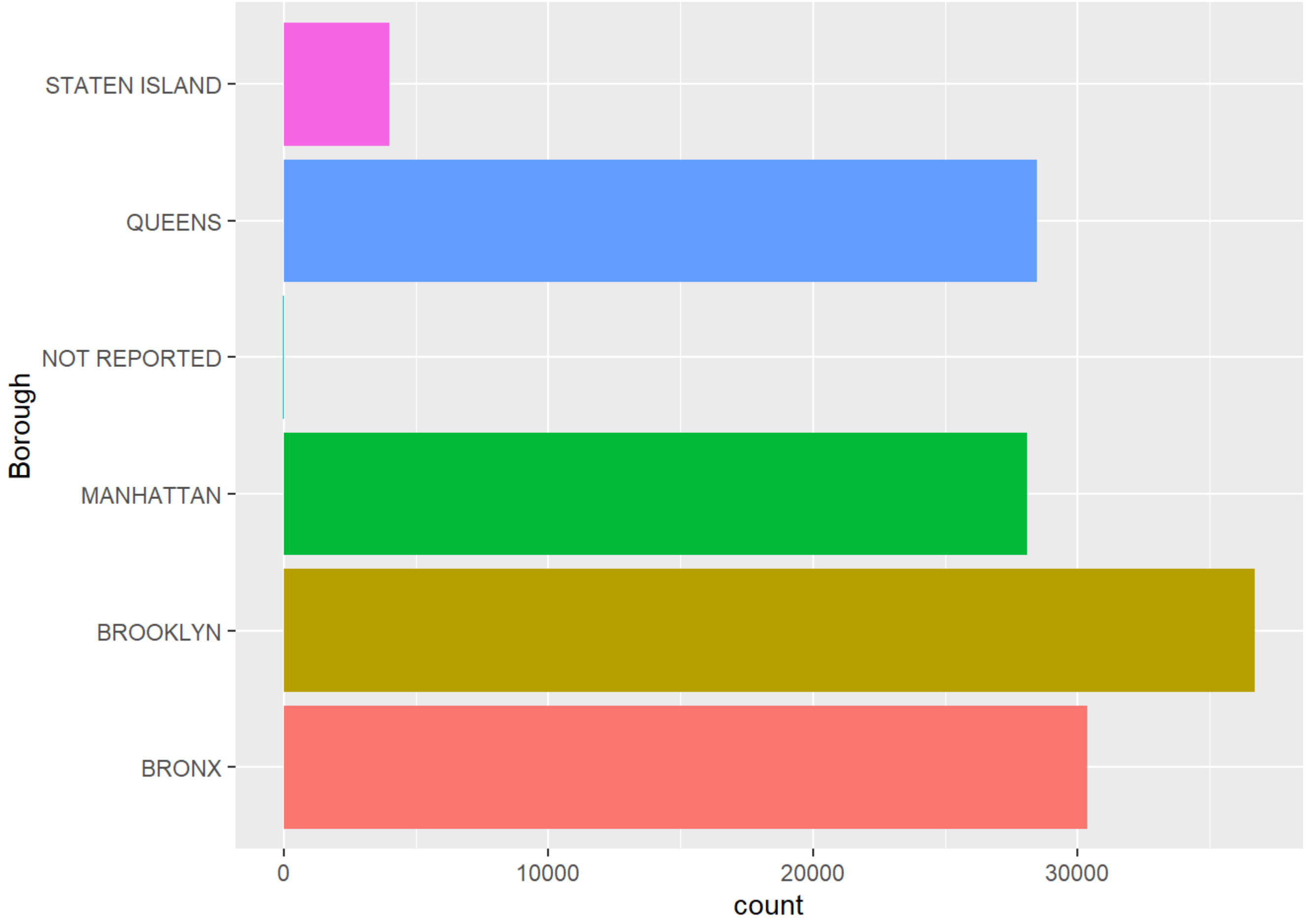
```
print(nrow(df3)*100/nrow(df))
```

```
## [1] 28.39584
```

# - There are 127641 street crimes, which accounts to 28.4% of total crimes recorded

# 15.2 Histogram of street crimes by borough

```
sb<-ggplot(df3, aes(x = `Borough`, fill=as.factor(`Borough`))) + geom_bar(width=0.9, stat="coun
t") + theme(legend.position="none") + coord_flip()
print(sb)
```



# - Brooklyn has the most street crimes followed by Bronx.

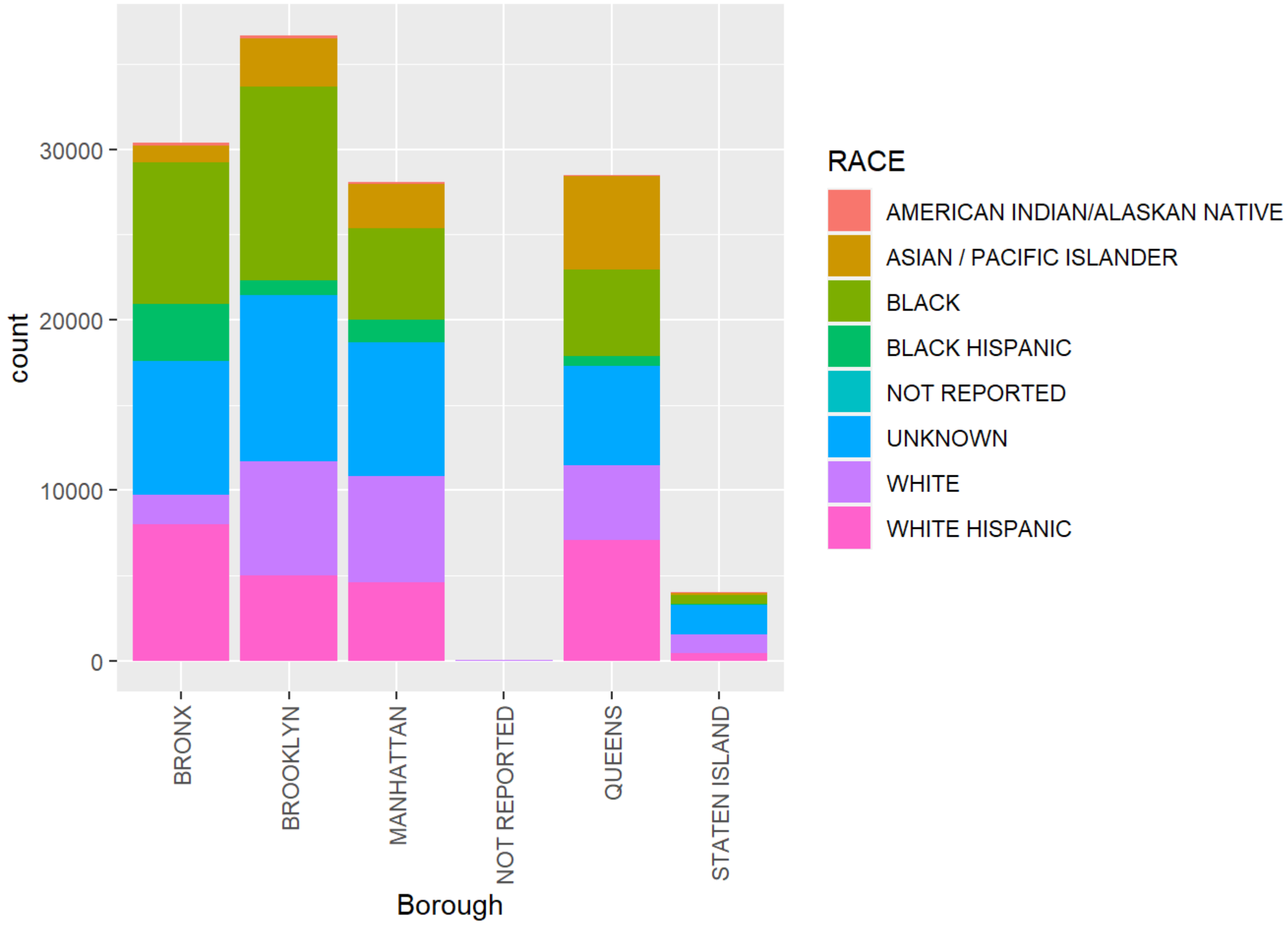
# - Staten island has the least, with less than 5000 street crimes recorded.

# 16. What are the races of victims of street crimes?

```
sr<- ggplot(data=df3,aes(x=`Borough`,fill=` Victim race`))+geom_histogram(stat="count")+ scale_f
ill_discrete(name="RACE")+theme(axis.text.x=element_text(angle=90,hjust=1,vjust=0.5))
```

```
## Warning: Ignoring unknown parameters: binwidth, bins, pad
```

```
print(sr)
```



- In Bronx and Brooklyn, the race most affected by street crimes are Black

# CONCLUSIONS:

## Which borough has the highest and lowest number of crimes?

- Brooklyn and Staten Island

## Which departments are solving crimes by month?

- N.Y. Police department

Which age group is most likely to be a victim?

- 25-44

How many crimes are happening between 12am - 6am

- 64501, 14.3% of all recorded crimes

How many Murders take place in this time frame?

- 160 murders

What is the race of most/least victims of street crimes?

- Most victims are Black and least are American indians/Alaskan natives.