**Percentage of Drivers Fully Stopped at a Stop Sign**

**1. Introduction:**

Crimes are increasing day by day and so an analysis of criminals stopped by the police is done. The weapons they possessed, location where they were caught, their sex, age was analyzed. Also, whether the police was wearing uniform or not, the reason of force used and whether arrest was made. The data was collected and then graphs were drawn using R studio and R programming language. The data consisted of 45788 observations. The variables were sex, date of birth, city, height, weight, race, etc. From that huge data set some meaningful observations were made.

**2. Body:**

**2.1 Data:**

The data is there at following link: <https://www.dropbox.com/s/1z44m3pnaosbsc0/PoliceCaughts.csv?dl=0>

**2.2 Methods:**

R programming language was used.

1. First the PoliceCaughts.csv file was read.

PoliceCaughts=read.csv("PoliceCaughts.csv")

1. For time and city analysis: The following code is for city Brooklyn:

PoliceCaughtsMornBrooklyn=subset(PoliceCaughts, PoliceCaughts$timestop>=600&PoliceCaughts$timestop<1200&PoliceCaughts$city==2)

PoliceCaughtsNightBrooklyn=subset(PoliceCaughts, PoliceCaughts$timestop>=0&PoliceCaughts$timestop<600&PoliceCaughts$city==2)

PoliceCaughtsEveningBrooklyn=subset(PoliceCaughts, PoliceCaughts$timestop>=1800&PoliceCaughts$timestop<=2359&PoliceCaughts$city==2)

PoliceCaughtsAfterNoonBrooklyn=subset(PoliceCaughts, PoliceCaughts$timestop>=1200&PoliceCaughts$timestop<1800&PoliceCaughts$city==2)

par(mfrow=c(2,2))

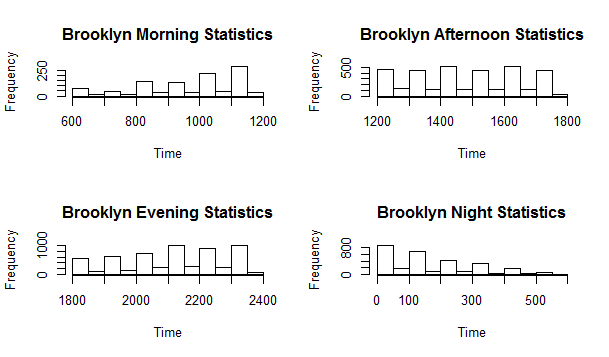
hist(PoliceCaughtsMornBrooklyn$timestop, xlab = "Time", ylab="Frequency", main="Brooklyn Morning Statistics")

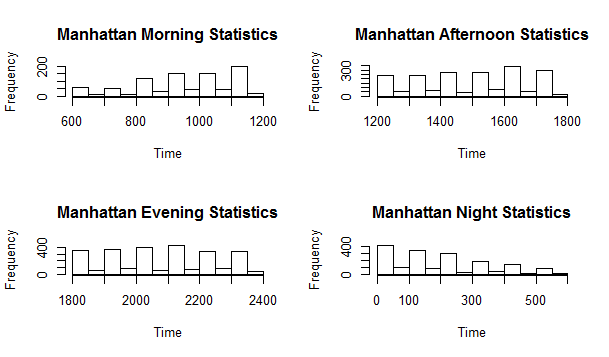
hist(PoliceCaughtsAfterNoonBrooklyn$timestop, xlab = "Time", ylab="Frequency", main="Brooklyn Afternoon Statistics")

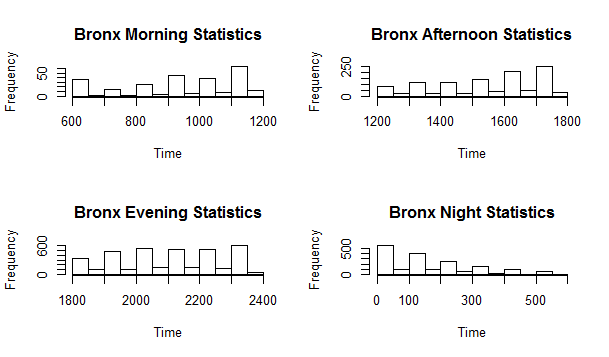
hist(PoliceCaughtsEveningBrooklyn$timestop, xlab = "Time", ylab="Frequency", main="Brooklyn Evening Statistics")

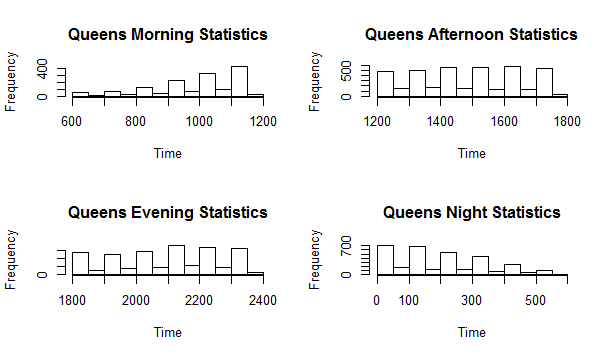
hist(PoliceCaughtsNightBrooklyn$timestop, xlab = "Time", ylab="Frequency", main="Brooklyn Night Statistics")

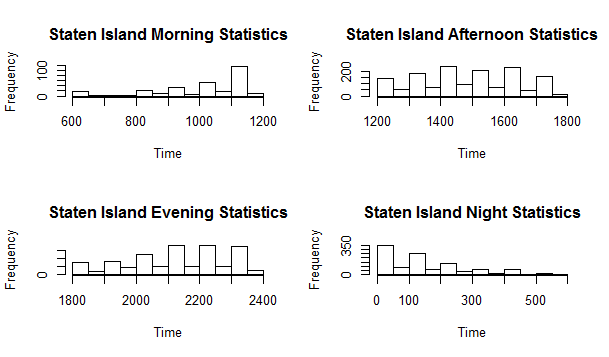
Similarly, code was written for all the other cities.











1. For gender analysis:

male=c(nrow(PoliceCaughtsMaleManhattan),nrow(PoliceCaughtsMaleBrooklyn),nrow(PoliceCaughtsMaleBronx),nrow(PoliceCaughtsMaleQueens),nrow(PoliceCaughtsMaleStaten))

female=c(nrow(PoliceCaughtsFemaleManhattan),nrow(PoliceCaughtsFemaleBrooklyn),nrow(PoliceCaughtsFemaleBronx),nrow(PoliceCaughtsFemaleQueens),nrow(PoliceCaughtsFemaleStaten))

plot(region, male, xlab="region", ylab="",type="l", col="black", main="Gender Statistics for each region")

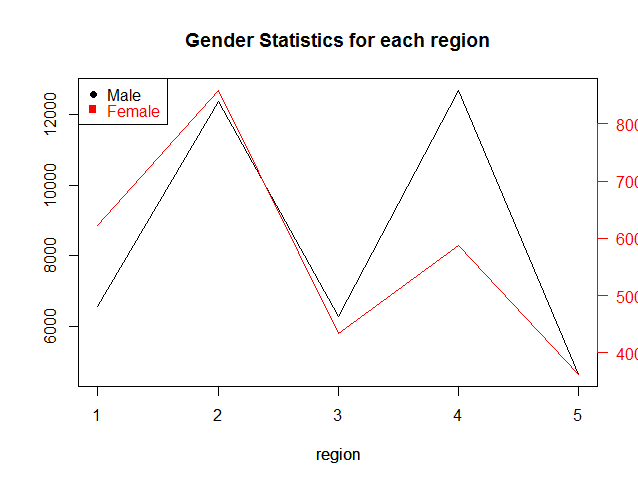
par(new=TRUE)

plot(region, female, axes="F", xlab="region", ylab="", type="l", col="red")

axis(4, ylim=c(4000,13000), col="red",col.axis="red",las=1)

mtext("Cell Density",side=4,col="red",line=4)

legend("topleft",legend=c("Male","Female"), text.col=c("black","red"),pch=c(16,15),col=c("black","red"))



1: Manahttan, 2: Brooklyn, 3: Bronx, 4. Queens, 5: Staten Island are the regions respectively. This code plots two line graphs in a single box so that we can analyze the male and female ratio.

1. For age analysis:

young=subset(PoliceCaughts,PoliceCaughts$age<30)

middleAged=subset(PoliceCaughts,PoliceCaughts$age>=30&PoliceCaughts$age<60)

old=subset(PoliceCaughts,PoliceCaughts$age>=60&PoliceCaughts$age<100)

superold=subset(PoliceCaughts,PoliceCaughts$age>=100)

par(mfrow=c(2,2))

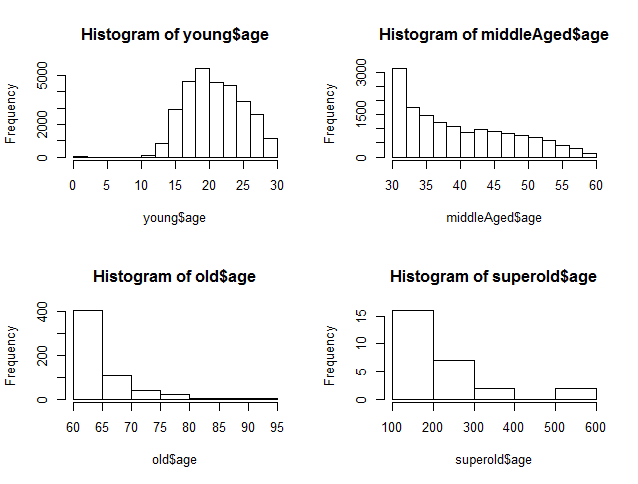
hist(young$age)

hist(middleAged$age)

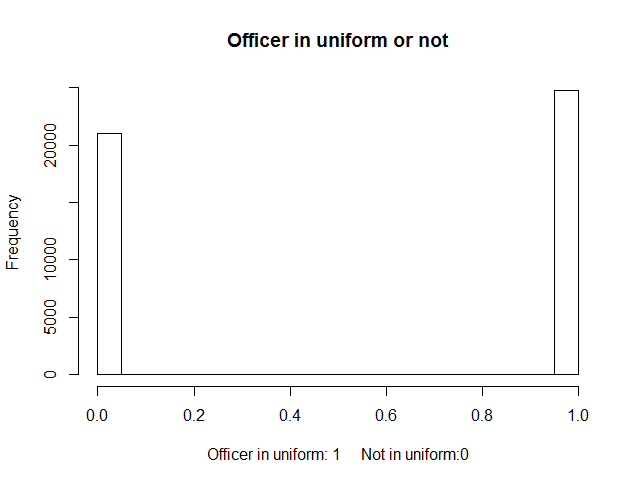
hist(old$age)

hist(superold$age)

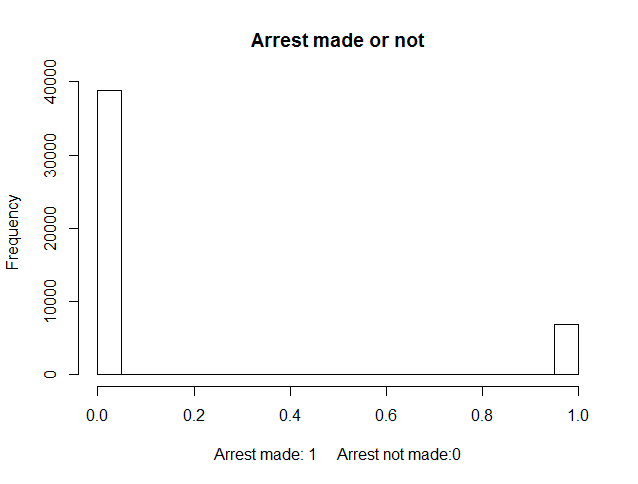
This will plot all the histograms for various ages.



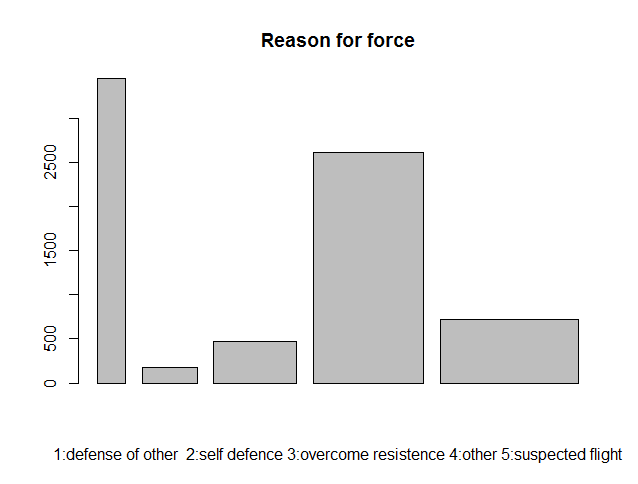
1. Officer with uniform worn or not:



1. Arrest made or not:



1. Reason for force used:



1. Number of weapons suspected:

[,1] [,2] [,3] [,4] [,5] [,6]

[1,] "pistol" "riflshot" "asltweap" "knifcuti" "machgun" "othrweap"

[2,] "197" "2" "0" "1068" "0" "363"

The data frame showing the number of each weapon.

* 1. **Analysis:**

1. Time and city analysis:

For Brooklyn city, around 1000 people were caught during evening which was more than other part of the day and more than any other city at any particular time. Manhattan had the maximum of 400 people caught during evening which was more than any other part of the day and was least maximum than any other city. For all the cities the maximum people caught was always in the evening except Staten island and queens.

1. Gender analysis:

From the gender analysis, it can be seen that more male were caught than female. The total number of male caught is more for Queens and Brooklyn. More male and female were caught in Brooklyn than any other city. So, it can be inferred that maximum number of people were caught in Brooklyn with Queens being the second maximum and following it is Manhattan.

1. Age analysis:

The age of the maximum number of people caught was between 15 to 30. Very less people with age above 100 were caught.

1. Officer in uniform or not.

It can be observed that there are many officers who were not in uniform and so unarmed and that the number is almost equal to the number of officers in uniform.

1. Arrest made or not:

It is observed that the number of arrests not made is too high than the numbers of arrests made.

1. Reason of force used is analyzed through the graph and the suspected criminals with harmful weapons like knifcuti and pistol was more.

**3. Results/ Conclusion/ Discussion:**

The number of officers without uniform should be minimized as it is risky and harmful for them to be unarmed. The number of arrests should be increased. Also, Brooklyn rules should be made stricter so that the number of crimes and criminals reduce. Also, the young generation should be taught not to make crimes so, that their number decreases. Also, the officer should have enough arms so that he can save himself from weapons like knifcuti and pistol.