Statistical Analysis from Crime Data of Seattle, Summer 2014 - Ravi R

Overview and Objectives

This is a data analysis conducted on the Seattle Crime Data of summer 2014. Some of the questions that we plan to address are :

- which incidents are most frequent & how these incidents vary by the time of the day
- which offense type is predominant in the evenings
- during what periods of the day are robberies most common

Analysis and code

Getting the data ready

We use the R programming language for this assignment. We first sort out the data to convert the date to the right format and then add a column for hour (0-23 hrs) so that we can find out the trends in the day. We then include only the fields for Offense type and Hour of the day as we aren't interested in any other fields.

Finally we summarize it by the Offense type to get a frequency of each of the offense type in every hour & then we get the top 10 offenses and plot it. The following code achieves it:

```
df <- read.csv("seattle_summer.csv")
library(ggplot2)
dtimes <- df$Occurred.Date.or.Date.Range.Start
x <- strptime(dtimes, format="%m/%d/%Y %I:%M:%S %p")
hour <- x$hour #hour of the crime
df$hour_of_crime <- hour
library(plyr)
```

```
## Warning: package 'plyr' was built under R version 3.2.2
```

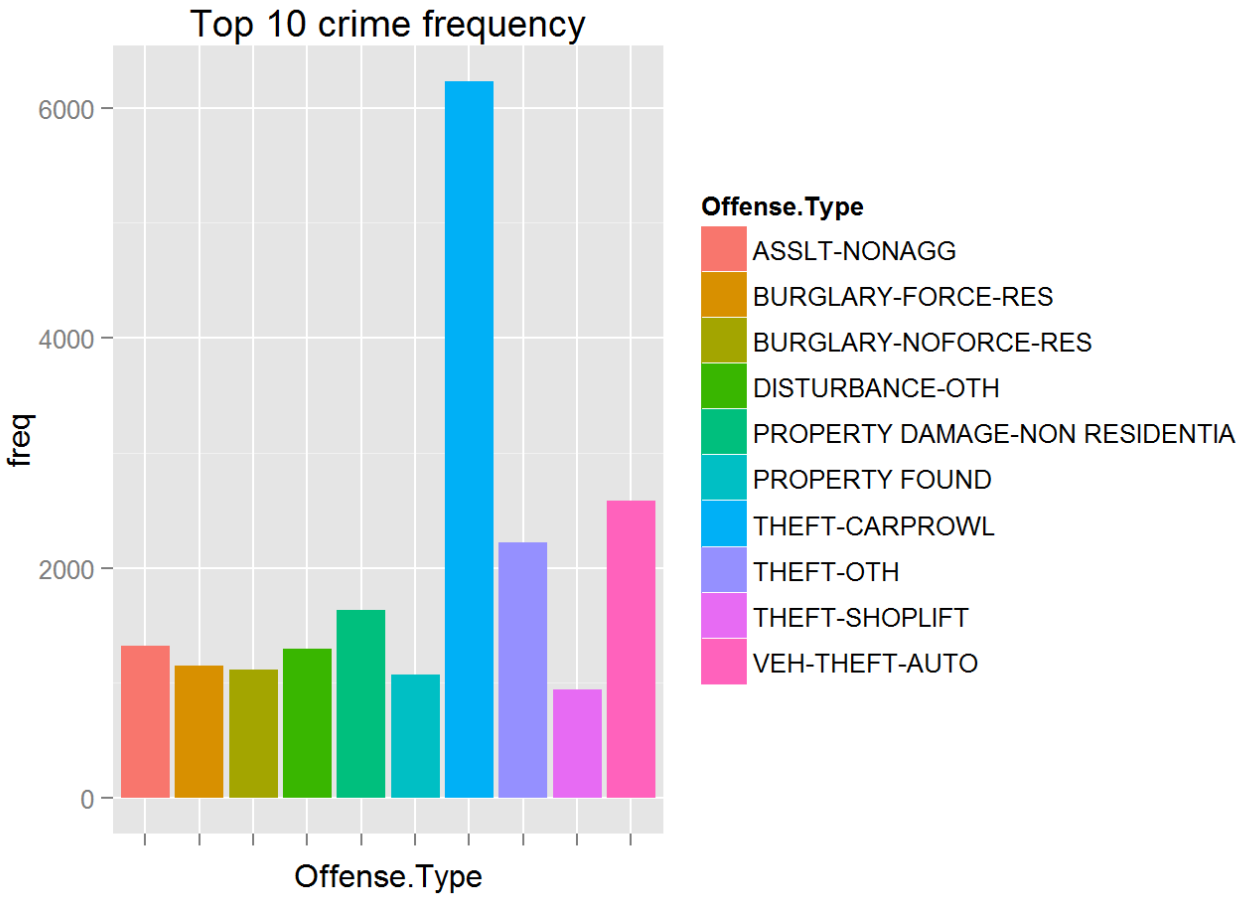
```
df_new <- df[ , which(names(df) %in% c("Offense.Type","hour_of_crime"))]
summary <- ddply(df_new, c("Offense.Type", "hour_of_crime"), summarize, freq=len
gth(hour_of_crime))
summary_ot <- ddply(df_new, c("Offense.Type"), summarize, freq=length(hour_of_cr
ime))
top10 <- head(summary_ot[with(summary_ot, order(-freq)), ], 10)
summary_top10 <- summary[summary$Offense.Type %in% top10$Offense.Type,]

#Just check if the data has been got correctly for one of the categories
#summary[summary$Offense.Type=="THREATS-OTHER",]
#Well we have for all the 24 hours 0-23 !
#Now to plot the

#ggplot(summary, aes(x=Offense.Type, y=freq, color=hour_of_crime))+geom_point()

#ggplot(summary, aes(x=hour_of_crime, y=freq, color=Offense.Type))+geom_point()

ggplot(data=top10, aes(y=freq, x=Offense.Type, fill=Offense.Type),height=1200, w
idth=1200) +
    geom_bar(stat="identity") + #theme(axis.text.x=element_text(angle=90,hjust=1
,vjust=0.5)) +
theme(axis.text.x = element_blank()) +
ggtitle("Top 10 crime frequency")
```



This gives us a good overview picture of what the crime situation is. We see that overall, the top offense is 'Theft-CarProwl'.

Next we analyze the spread of these crimes over time.

Data Analytics and Conclusions

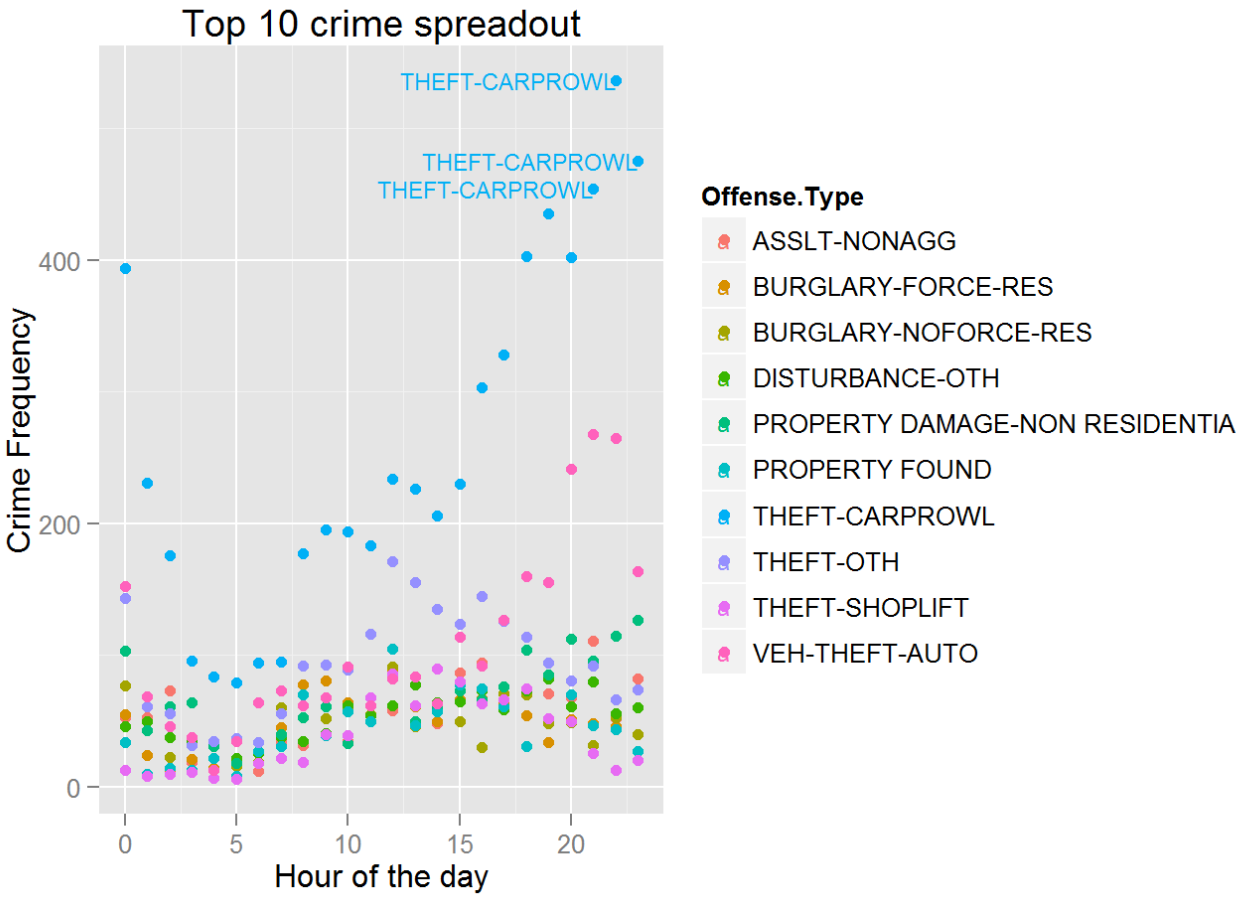
For the purpose of visualiztion, we categorize the hour of crime as ordinal (since it has an importance of order), frequency of offense as numeric and the offense type as nominal data (portrayed as different colors) for the visualization.

For each of the questions that we would like to answer we do the analytics and the conclusion in this section.

With the code in hand we now plot a graph to see the overall trend and in particular find out which incidents are most frequent in the evenings.

```
#png("plot.png", width=12500, height=10000, res=300)


ggplot(summary_top10, aes(x=hour_of_crime, y=freq, color=Offense.Type), height=9
60, width=1200)+ geom_point() + geom_text(aes(label=ifelse(freq>450,as.character
(Offense.Type),'')),hjust=1,just=-100, size=3) + xlab("Hour of the day") + ylab(
"Crime Frequency") + ggtitle("Top 10 crime spreadout")
```



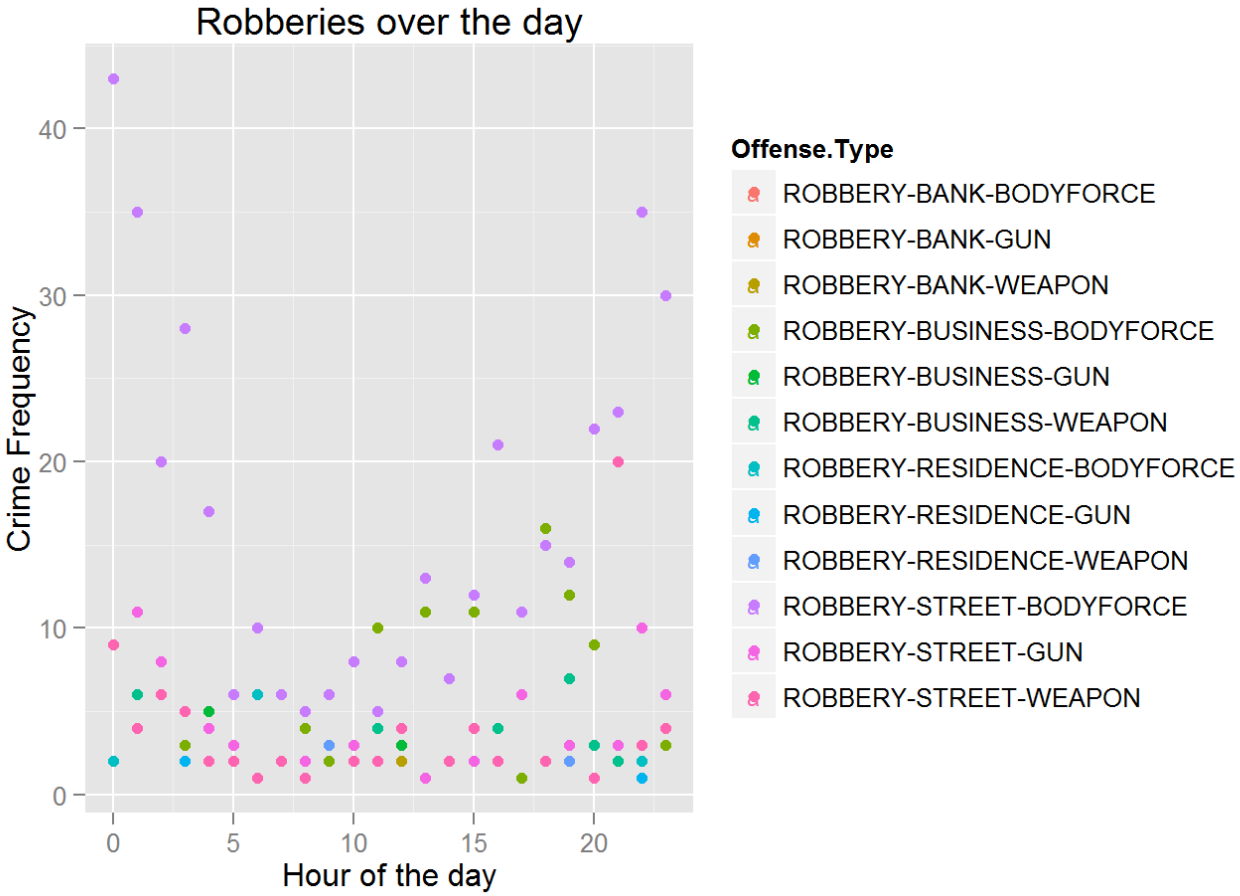

```
#dev.off()
```

Based on the graph, we can conculde that “Theft-Carprowl” as the most frequent offense type in the evening.

Finally let’s look at the trend for robberies. For this we filter the Offense type for words containing the text “ROBBERY” (which includes many categories of robberies) and we plot the graph of robbery over the whole day

```
summary_subset <- summary[grep("ROBBERY", summary$Offense.Type),]

ggplot(summary_subset, aes(x=hour_of_crime, y=freq, color=Offense.Type))+ geom_p
oint() + geom_text(aes(label=ifelse(freq>450,as.character(Offense.Type),'')),hju
st=1,just=-100, size=3) + xlab("Hour of the day") + ylab("Crime Frequency") + gg
title("Robberies over the day")
```



From the graph we can conclude that robberies pick up after about 2000hrs(or 8PM) and continue to be high around till about 5AM.

Key Findings

In the summer of 2014 in Seattle, most offense types were “Theft - Carprowl” and they had a higher frequency in the evenings. Also robberies were predominant at night.