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

Finding

Robberies increased at night during summer months in Seattle in 2014 & most offense types in the evening were “Theft - Carprowl”

Overview and Objectives

This is a small data analysis conducted on the Seattle Crime Data of summer 2014. Some of the areas that we take a glimpse of and get some answers for questions are :

- which incidents are most frequent in the evening & how incidents vary by the time of the day
- during what periods of the day are robberies most common

Analysis and code

Getting the data ready

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. 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))
#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()
```

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 or number of crimes per offense type as numeric and the offense type as nominal data (portrayed as different colors) for the visualization.

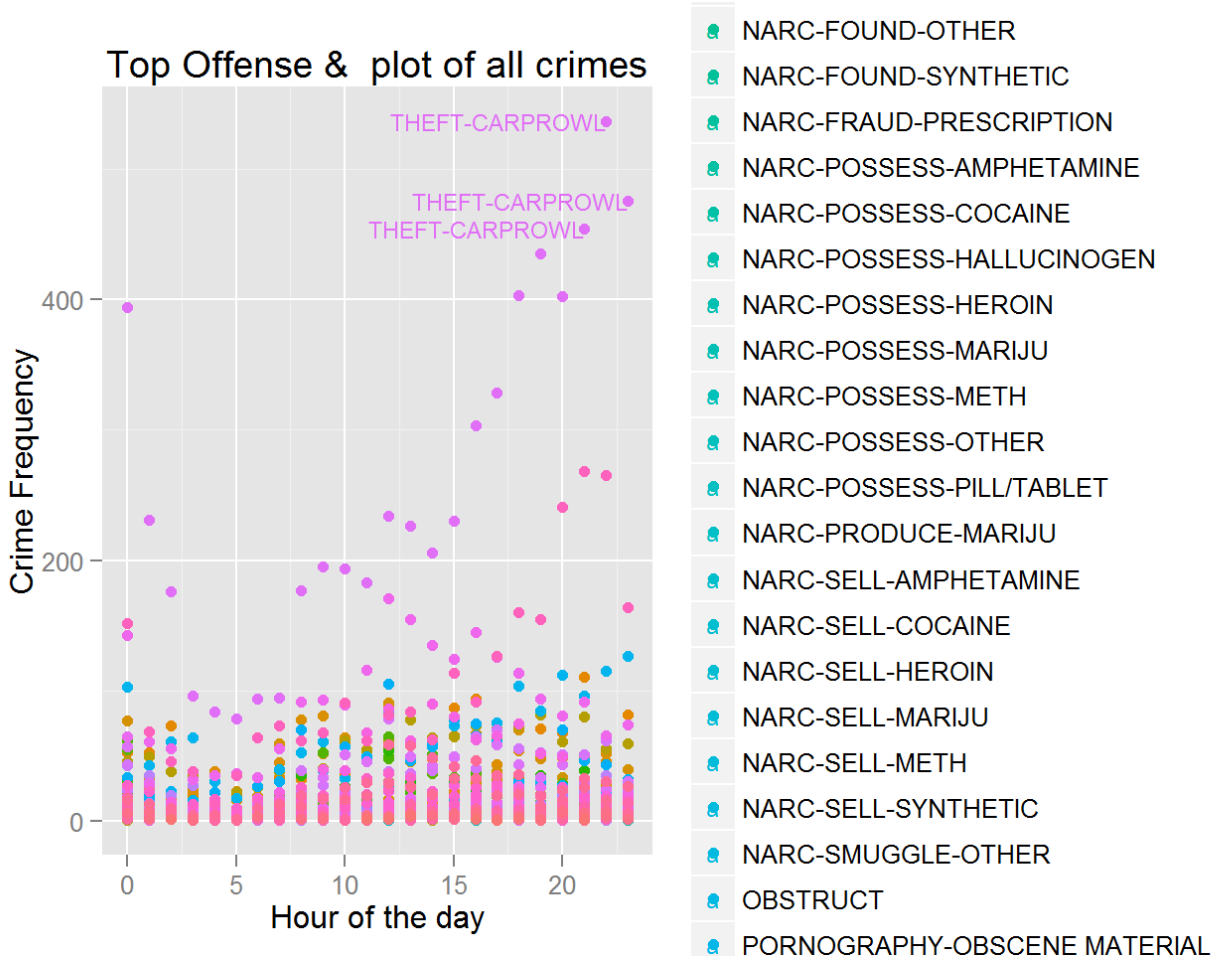
For each of the questions that we would like to answer we do the analytics and the conclusion in this section.

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.

Since the data has lot of points and identifying by color is not that easy we label the points that we are interested in. Based on the trends We keep a threshold of 450 for the # of crimes (we'll label only those Offense types having over 450 for a given hour). Few of the legend items are not visible in this but the information which we mainly want is - a more detailed graph (which cannot fit in this document & which has a higher resolution and fits almost all crime types is available as plot.png in the same repository https://github.com/rravishankar/uow_assignment6/blob/master/plot.png (https://github.com/rravishankar/uow_assignment6/blob/master/plot.png) - this file can be viewed separately maybe on a larger monitor to get the big picture)

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


ggplot(summary, aes(x=hour_of_crime, y=freq, color=Offense.Type), height=960, wi
dth=1200)+ geom_point() + geom_text(aes(label=ifelse(freq>450,as.character(Offen
se.Type),'')),hjust=1,just=-100, size=3) + xlab("Hour of the day") + ylab("Crime
 Frequency") + ggtitle("Top Offense &  plot of all crimes")
```



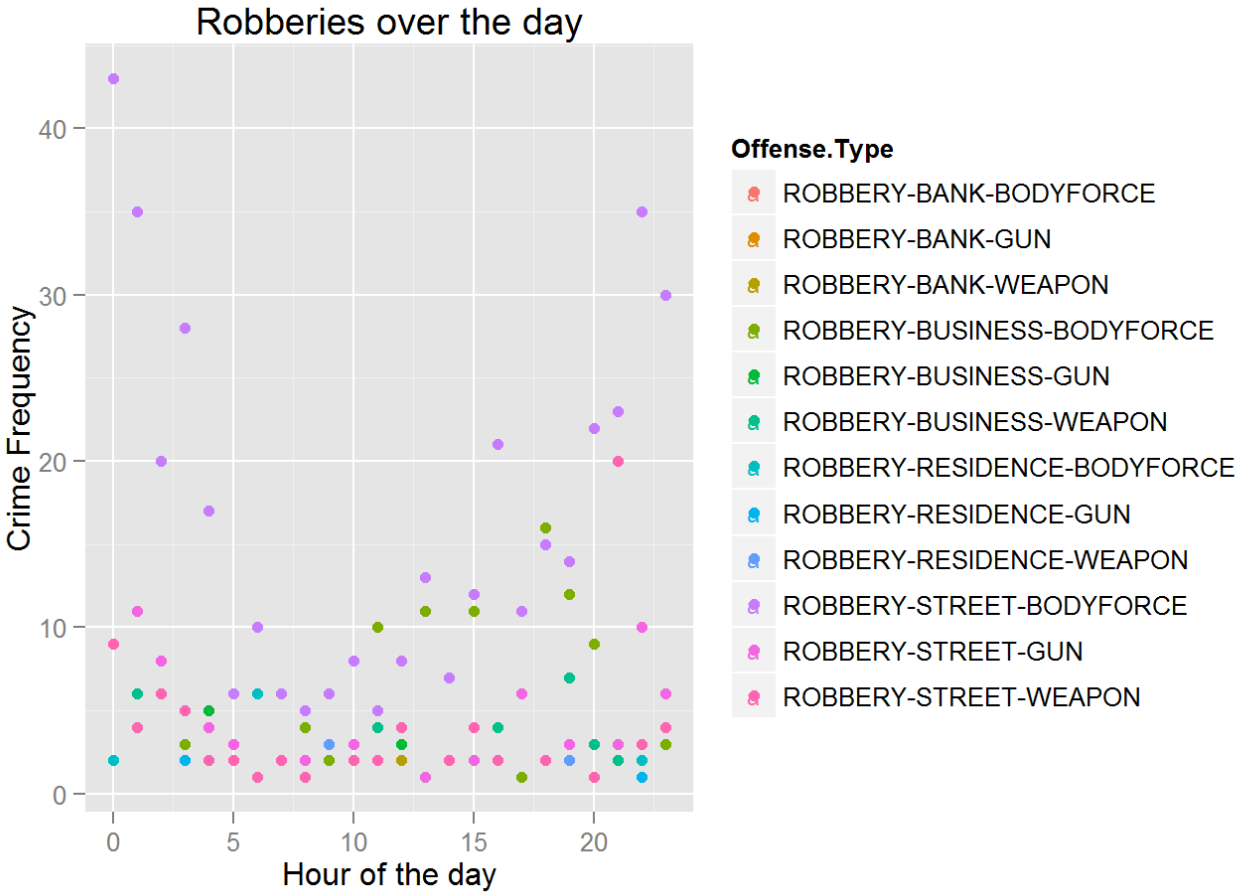

```
#dev.off()
```

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.

Data Visualization Conclusion

The graph visualzations make it easy to observe the trends in the data. It is clear from the visualizations that in Seattle Summer 2014, Theft CarProwl was the most frequently committed offense and robberies were predominant at night.