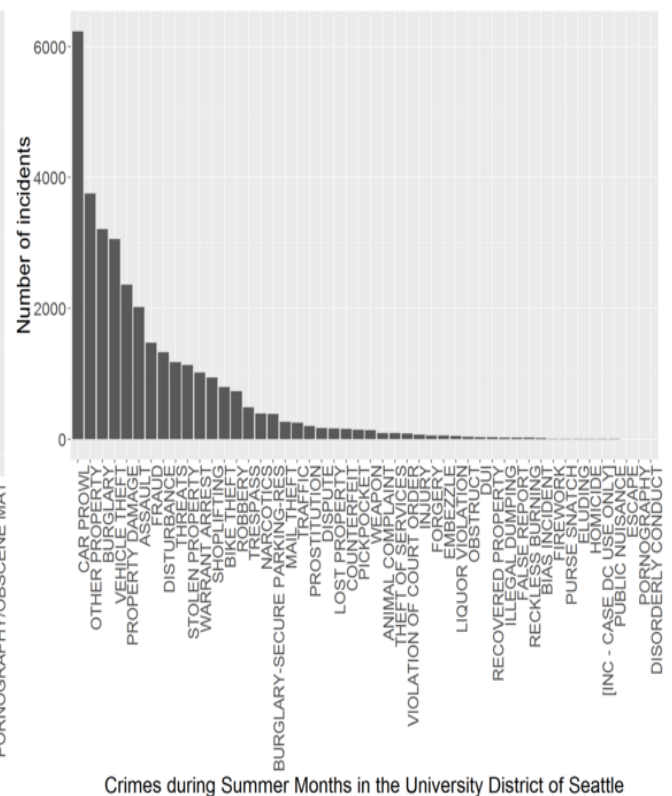
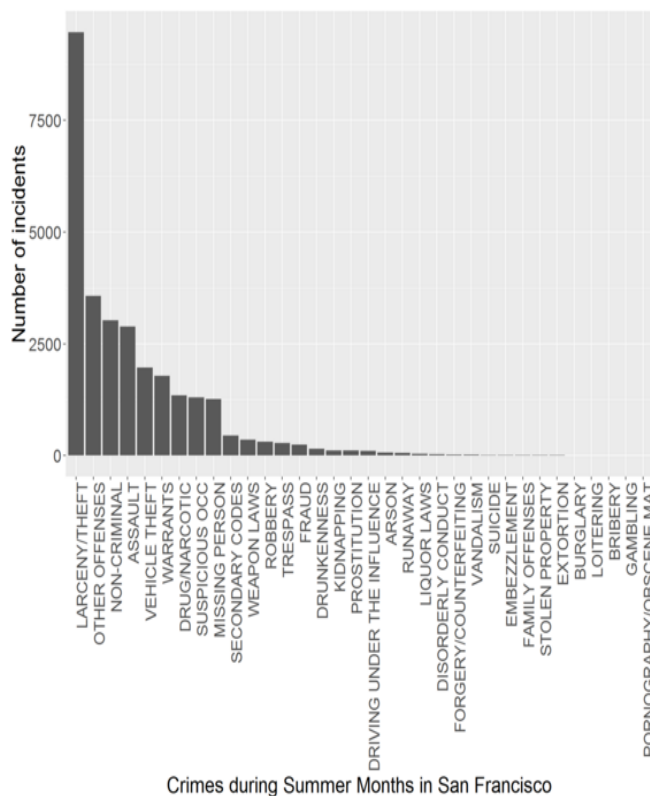


# Crime Analytics: Visualization of Crime Incident Reports for Summer 2014 in San Francisco and Seattle

- In this assignment, some *expoloratory analysis* is done on the **criminal incident data** from **Seattle** and **San Francisco** to visualize patterns and contrast and compare patterns across the two cities.
- Data** used: The real crime dataset from **Summer (June-Aug) 2014** for both of two US cities *Seattle* and *San Francisco* has been used for the analysis. These reduced datasets are available on the github repository:  
[https://github.com/uwescience/datasci\\_course\\_materials/tree/master/assignment6](https://github.com/uwescience/datasci_course_materials/tree/master/assignment6).
- All of the analysis / visualization is done using R and with the library *graphic grammar plot* and should be easily reproducible. Almost all of the relevant code is embedded in the html, rest can be found from the *Rmd* file uploaded in the same github repository.
- First we observe from the following visualization that the crime of category **LARCENY/THEFT** is the most common in San Francisco, whereas **CAR PROWL** is most common crime in the University District of Seattle, during the Summer Months.

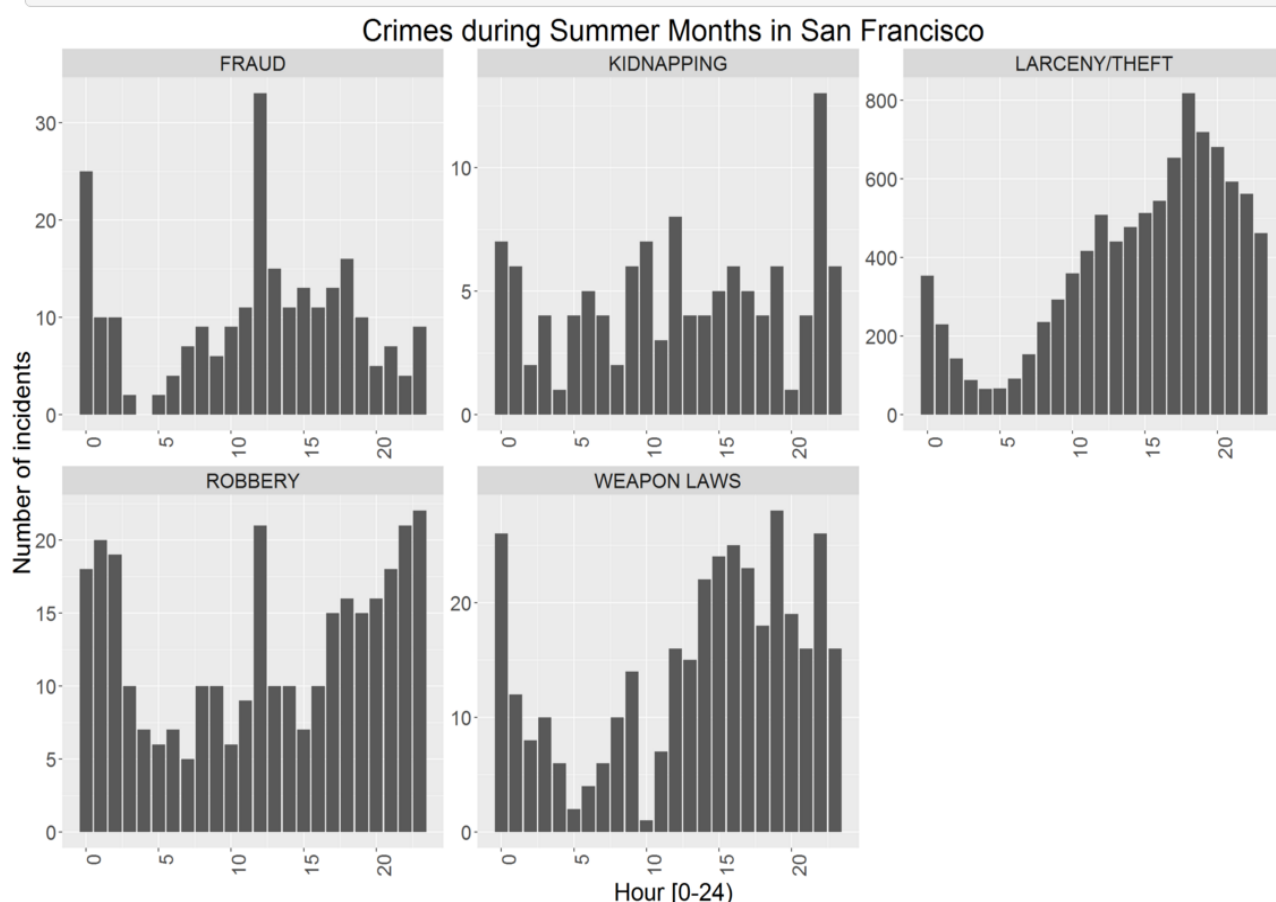
```
library(ggplot2)
library(dplyr)
p1 <- ggplot(df, aes(x=reorder(Category, Category, function(x)-length(x)))) + geom_bar() +
  xlab('Crimes during Summer Months in San Francisco') +
  ylab('Number of incidents') +
  theme(axis.text.x = element_text(angle = 90, hjust = 1), legend.position="none", text = element_text(size = 20))
p2 <- ggplot(df2, aes(x=reorder(Summarized.Offense.Description, Summarized.Offense.Description, function(x)-length(x)))) +
  geom_bar() +
  xlab('Crimes during Summer Months in the University District of Seattle') +
  ylab('Number of incidents') +
  theme(axis.text.x = element_text(angle = 90, hjust = 1), legend.position="none", text = element_text(size = 20))
multiplot(p1, p2, cols=2)
```



## Analysis to find how the number of the crime incidents change over time in a day (below are the **key findings** in bold text)

- We see from the following visualization that the violent crimes such as **LARCENY/THEFT, ROBBERY, WEAPON LAWS** increase at night (and in the early morning), whereas Crimes such as **FRAUD, KIDNAPPING** are quite common in the day time too, during **Summer Months in San Francisco** (as can be seen from the following figure). Notice that the Hour (in the X-axis) is represented in 24-hr format, so that 0-5 represents the hours from 12 AM - 5 AM in the morning, whereas 15-20 represents the hours from 3 PM - 8 PM at night. The Y-axis represents the total number of incidents happened, so that we have a *time series* of numbers of incidents for a few crime incidents we focus on.

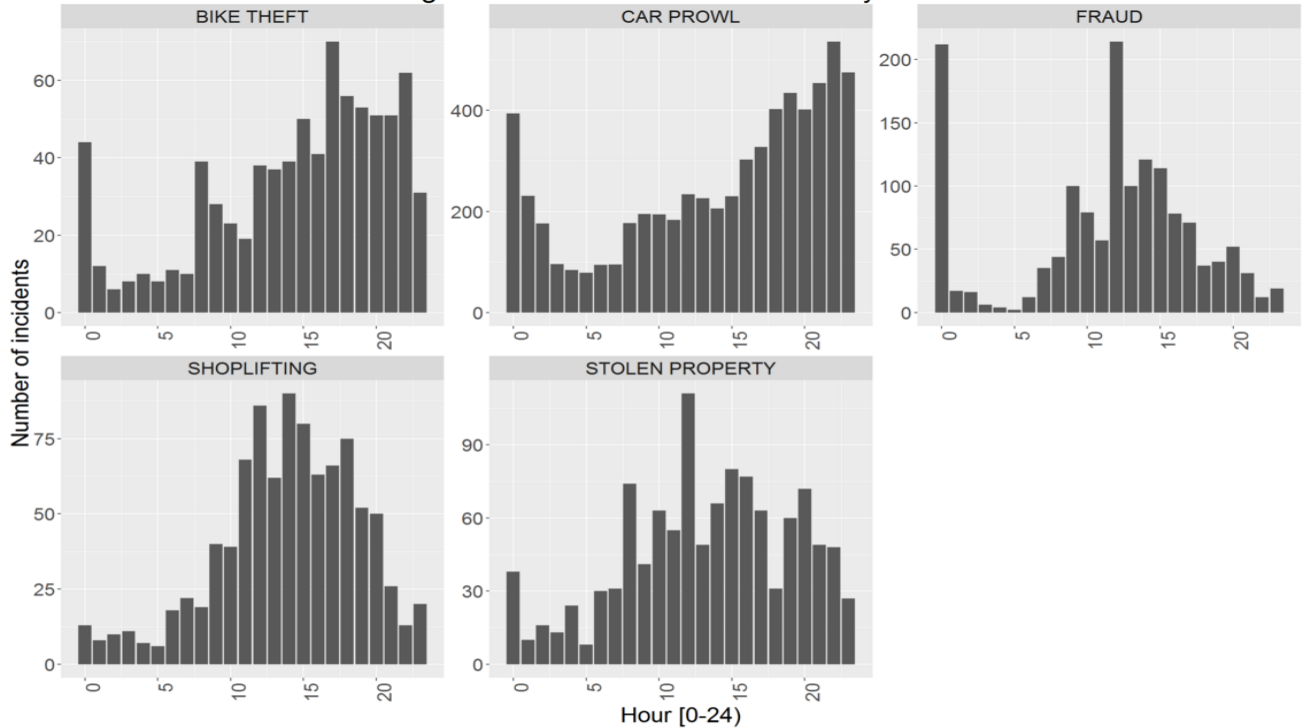
```
df1 <- df %>% filter(Category %in% c('LARCENY/THEFT', 'ROBBERY', 'WEAPON LAWS', 'FRAUD', 'KIDNAPPING'))
ggplot(df1, aes(Hr)) + geom_bar() +
  #scale_x_continuous(breaks=seq(0, 24, 2)) +
  facet_wrap(~Category, scale='free') +
  xlab('Hour [0-24)') +
  ylab('Number of incidents') +
  ggtitle('Crimes during Summer Months in San Francisco') +
  theme(axis.text.x = element_text(angle = 90, hjust = 1), legend.position="none",
        text = element_text(size = 20))
```



- Similarly, we can observe from the following visualization that the crimes such as **CAR PROWL, BIKE THEFT** increase at night, whereas the crimes such as **FRAUD, SHOPLIFTING, STOLEN PROPERTY** are quite common in the day time too, during Summer Months in the University District of Seattle.

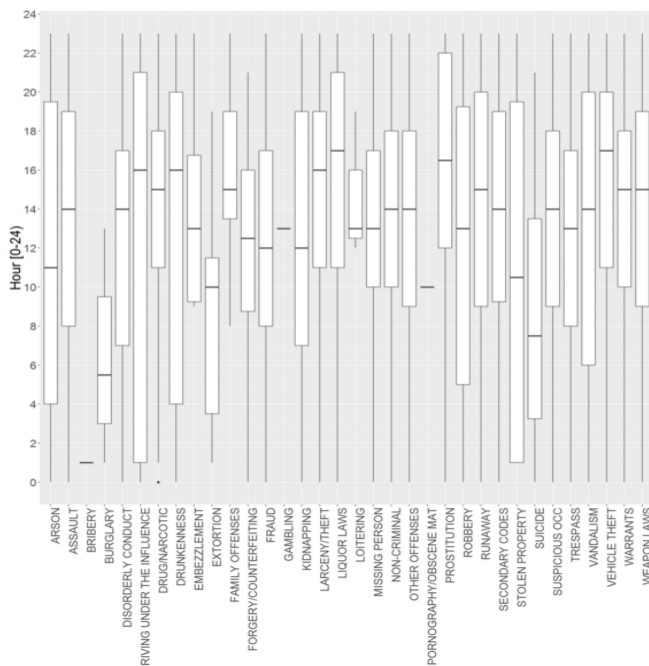
```
df1 <- df2 %>% filter(Summarized.Offense.Description %in% c('CAR PROWL', 'BIKE THEFT', 'SHOPLIFTING', 'FRAUD', 'STOLEN PROPE
RTY'))
ggplot(df1, aes(Hr)) + geom_bar() +
  #scale_x_continuous(breaks=seq(0, 24, 5)) +
  facet_wrap(~Summarized.Offense.Description, scale='free') +
  xlab('Hour [0-24)') +
  ylab('Number of incidents') +
  ggtitle('Crimes during Summer Months in the University District of Seattle') +
  theme(axis.text.x = element_text(angle = 90, hjust = 1), legend.position="none",
        text = element_text(size = 20))
```

## Crimes during Summer Months in the University District of Seattle

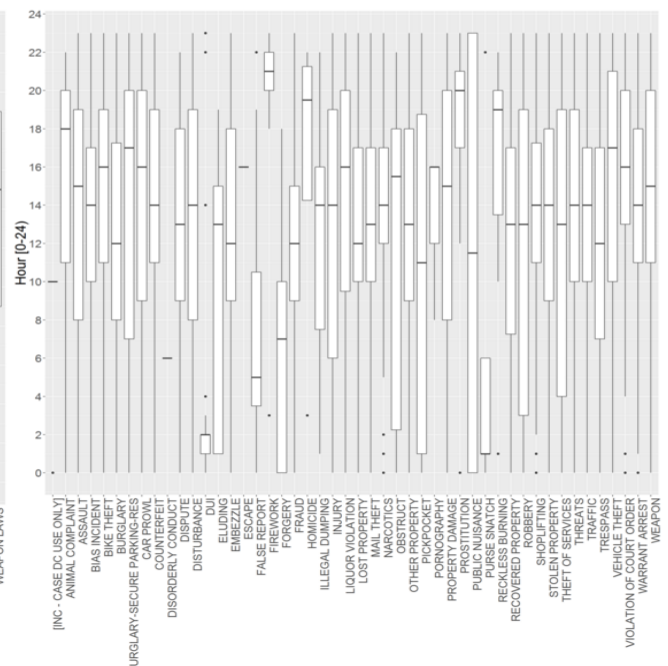


- As can be seen from the below figures, the median hour when LARCENY/THEFT incidents happened was 4 PM in San Francisco whereas the median hour when BIKE THEFT incidents happened was 5 PM in the University District of Seattle, during the Summer Months. The median hour when BURGLARY incidents happened was around 6 AM in San Francisco and was 2 PM in the University District of Seattle, during the Summer Months.

```
p1 <- ggplot(df, aes(Category, Hr)) + geom_boxplot() +
  xlab('Crimes during Summer Months in San Francisco') +
  ylab('Hour [0-24]') +
  scale_y_continuous(breaks=seq(0, 24, 2)) +
  theme(axis.text.x = element_text(angle = 90, hjust = 1), legend.position="none", text = element_text(size = 20))
p2 <- ggplot(df2, aes(Summarized.Offense.Description, Hr)) + geom_boxplot() +
  xlab('Crimes during Summer Months in the University District of Seattle') +
  ylab('Hour [0-24]') +
  scale_y_continuous(breaks=seq(0, 24, 2)) +
  theme(axis.text.x = element_text(angle = 90, hjust = 1), legend.position="none", text = element_text(size = 20))
multiplot(p1, p2, cols=2)
```



Crimes during Summer Months in San Francisco



Crimes during Summer Months in the University District of Seattle

- Also, as can be seen from the below figures, the overall hourly crime incidents pattern for both the cities looks kind of similar. Both of the cities have the least number of crimes incidents happened in between 4 AM - 6 AM in the morning during summer months. The maximum number of crime incidents happened in San Francisco in between 4 PM - 6 PM in the evening whereas the most crime-prone hours for the University District of Seattle was 6 PM - 8 PM at night.

```
df1 <- df %>% group_by(Hr) %>% summarise(numIncidents=n())
p1 <- ggplot(df1, aes(x=Hr, y=numIncidents)) + geom_point() + stat_smooth() +
  scale_x_continuous(breaks=seq(0, 24, 2)) +
  xlab('Hour [0-24)') +
  ylab('Number of hourly incidents') +
  ggtitle('Crimes during Summer Months in San Francisco') +
  theme(axis.text.x = element_text(angle = 90, hjust = 1), legend.position="none", text = element_text(size = 20))
df2 <- df %>% group_by(Hr) %>% summarise(numIncidents=n())
p2 <- ggplot(df2, aes(x=Hr, y=numIncidents)) + geom_point() + stat_smooth() +
  scale_x_continuous(breaks=seq(0, 24, 2)) +
  xlab('Hour [0-24)') +
  ylab('Number of hourly incidents') +
  ggtitle('Crimes during Summer in the University District of Seattle') +
  theme(axis.text.x = element_text(angle = 90, hjust = 1), legend.position="none", text = element_text(size = 20))
multiplot(p1, p2, cols=2)
```

