Data Science At Scale : Visualization of Incident Reports

I analyzed criminal incident data from Seattle and San Francisco in Summer 2014. My goal is to convey the results of my analysis using a series of visualizations in a highly effective manner.

There is significant motivation to perform this analysis, mainly because Seattle and San Francisco are two major American cities frequented by tourists particularly in the summer months. A clear presentation of the results will be highly informative to a tourist.

I used R Statistical Software to analyse the crime data.

Quick look at the data

Load and Inspect data sets

```
> sf.crime.data <-
read.csv('sanfrancisco_incidents_summer_2014.csv')

> seattle.crime.data <-
read.csv('seattle_incidents_summer_2014.csv')

> head(sf.crime.data)

> head(seattle.crime.data)

> table(sf.crime.data$Category)

> table(seattle.crime.data$Offense.Type)
```

Top Ten crimes in San Franscisco

```
>sf.crime.frequency.by.category <- table(sf.crime.data$Category)

> sf.sorted <- sort(sf.crime.frequency.by.category, decreasing =
TRUE)
```

```
  LARCENY/THEFT          OTHER OFFENSES
           9466                    3567
   NON-CRIMINAL                 ASSAULT
           3023                    2882
   VEHICLE THEFT                WARRANTS
           1966                    1782
   DRUG/NARCOTIC            SUSPICIOUS OCC
           1345                    1300
  MISSING PERSON           SECONDARY CODES
           1266                     442
```

Top Ten crimes in Seattle

```
> seattle.crime.frequency.by.category <-
            table(seattle.crime.data$Offense.Type)

> seattle.sorted <- sort(seattle.crime.frequency.by.category,
decreasing = TRUE)
```

```
      THEFT-CARPROWL                VEH-THEFT-AUTO
                6230                          2588
THEFT-OTH PROPERTY DAMAGE-NON RESIDENTIA
                2221                          1638
      ASSLT-NONAGG                DISTURBANCE-OTH
                1320                          1295
  BURGLARY-FORCE-RES          BURGLARY-NOFORCE-RES
                1147                          1118
    PROPERTY FOUND                THEFT-SHOPLIFT
                1069                           944
```

A tourist's perspective of crime data – Map Guide to Crime

A tourist might be curious to know the areas of a city with the most crime in order to be more vigilant while in those areas or she might choose to avoid those areas. A city map with the geo-location of criminal activity will be insightful.

A tourist visiting San Francisco might be concerned about the following criminal activities : MISSING PERSON, ROBBERY, DRUG/NARCOTIC, ASSAULT and LARCENY/THEFT.

A tourist visiting Seattle might be concerned about the following criminal activities: HARASSMENT,ROBBERY-STREET-BODYFORCE,THEFT-PKPOCKET, ASSLT-AGG-BODYFORCE.

Data Science At Scale : Visualization of Incident Reports

Crimes a tourist in SF might be concerned about

```
> sf.tourist.crimes <- subset(sf.crime.data, Category == 'MISSING PERSON' | Category == 'ROBBERY' | Category == 'DRUG/NARCOTIC' | Category == 'ASSAULT' | Category == 'LARCENY/THEFT')

> head(sf.tourist.crimes)
  IncidntNum     Category   Descript DayOfWeek       Date  Time PdDistrict     Resolution              Address        X

3  146177923 LARCENY/THEFT              GRAND THEFT FROM LOCKED AUTO    Sunday 08/31/2014 23:30 SOUTHERN NONE 1000 Block of MISSION ST -122.4098

4  146177531 LARCENY/THEFT              GRAND THEFT FROM LOCKED AUTO    Sunday 08/31/2014 23:30 RICHMOND NONE      FULTON ST / 26TH AV -122.4853

6  140734349 DRUG/NARCOTIC                    POSSESSION OF MARIJUANA    Sunday 08/31/2014 23:13   SOUTHERN ARREST, BOOKED       11TH ST / MINNA ST -122.4166

7  140734349 DRUG/NARCOTIC POSSESSION OF CONTROLLED SUBSTANCE FOR SALE    Sunday 08/31/2014 23:13   SOUTHERN ARREST, BOOKED       11TH ST / MINNA ST -122.4166

.....
```

Crimes a tourist in Seattle might be concerned about

```
> seattle.tourist.crimes <- subset(seattle.crime.data, Offense.Type == 'HARASSMENT' | Offense.Type == 'ROBBERY-STREET-BODYFORCE' | Offense.Type == 'THEFT-PKPOCKET' | Offense.Type == 'ASSLT-AGG-BODYFORCE' )

> head(seattle.tourist.crimes)
  RMS.CDW.ID General.Offense.Number Offense.Code Offense.Code.Extension   Offense.Type Summary.Offense.Code Summarized.Offense.Description          Date.Reported

20     287595          2015112158       5309                  0 HARASSMENT              5300                    THREATS 04/06/2015 12:30:00 PM

62    1104194          2014411364       5309                  0 HARASSMENT              5300                    THREATS 12/12/2014 01:46:00 PM

63    1121051          2014411158       2301                  0 THEFT-PKPOCKET          2300                 PICKPOCKET 12/12/2014 10:28:00 AM
```

Visualize SF crime map

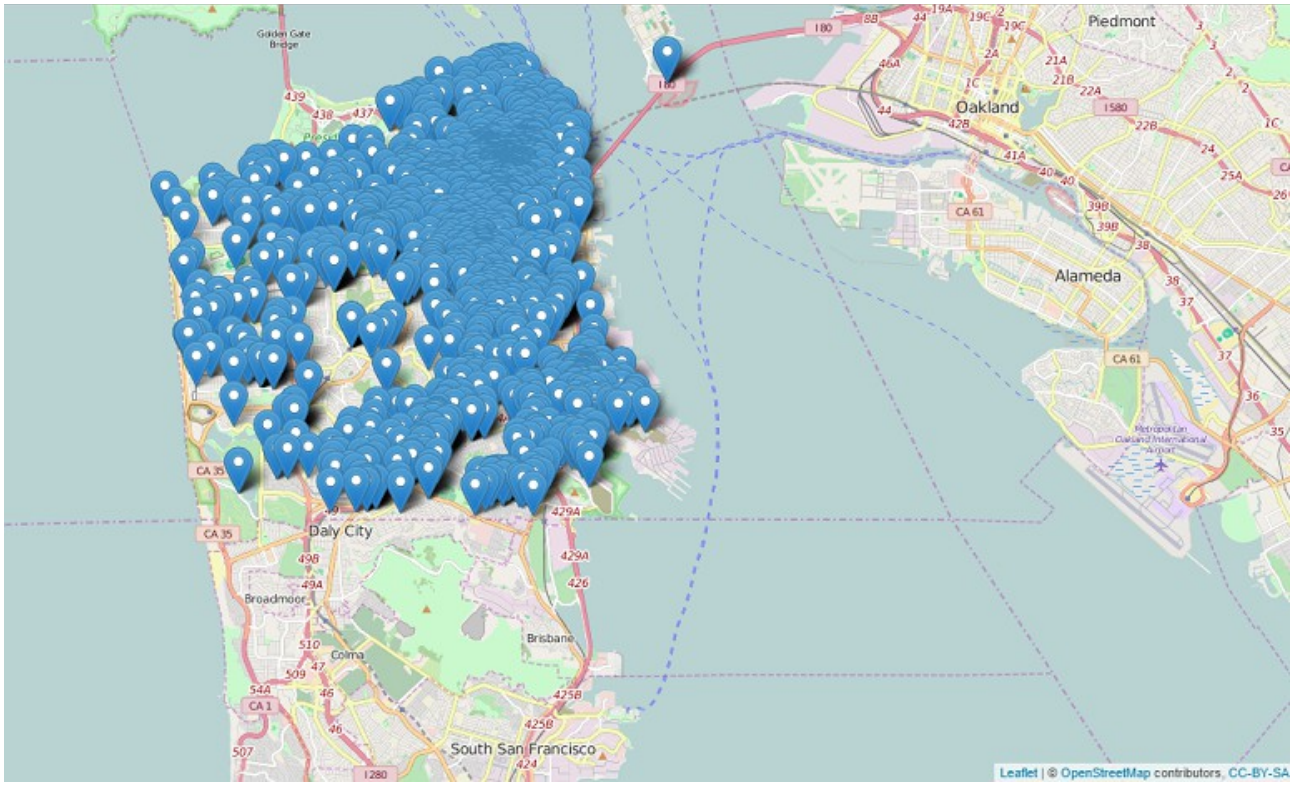
```
library(leaflet)
library(magrittr)
sf.tourist.map <- leaflet()%>%
+   addTiles() %>%
+   setView(-122.42, 37.78, zoom = 15) %>%
+   addMarkers(data = sf.tourist.crimes, lng = ~ X, lat = ~ Y, popup = sf.tourist.crimes$Category)

> sf.tourist.map
```

Visualize Seattle crime map

```
library(leaflet)
library(magrittr)
seattle.tourist.map <- leaflet()%>%
+   addTiles() %>%
+   setView(-122.42, 37.78, zoom = 15) %>%
+   addMarkers(data = seattle.tourist.crimes, lng = ~ X, lat = ~ Y, popup = sattle.tourist.crimes$Offense.Type)

> sf.tourist.map
```

Data Science At Scale : Visualization of Incident Reports

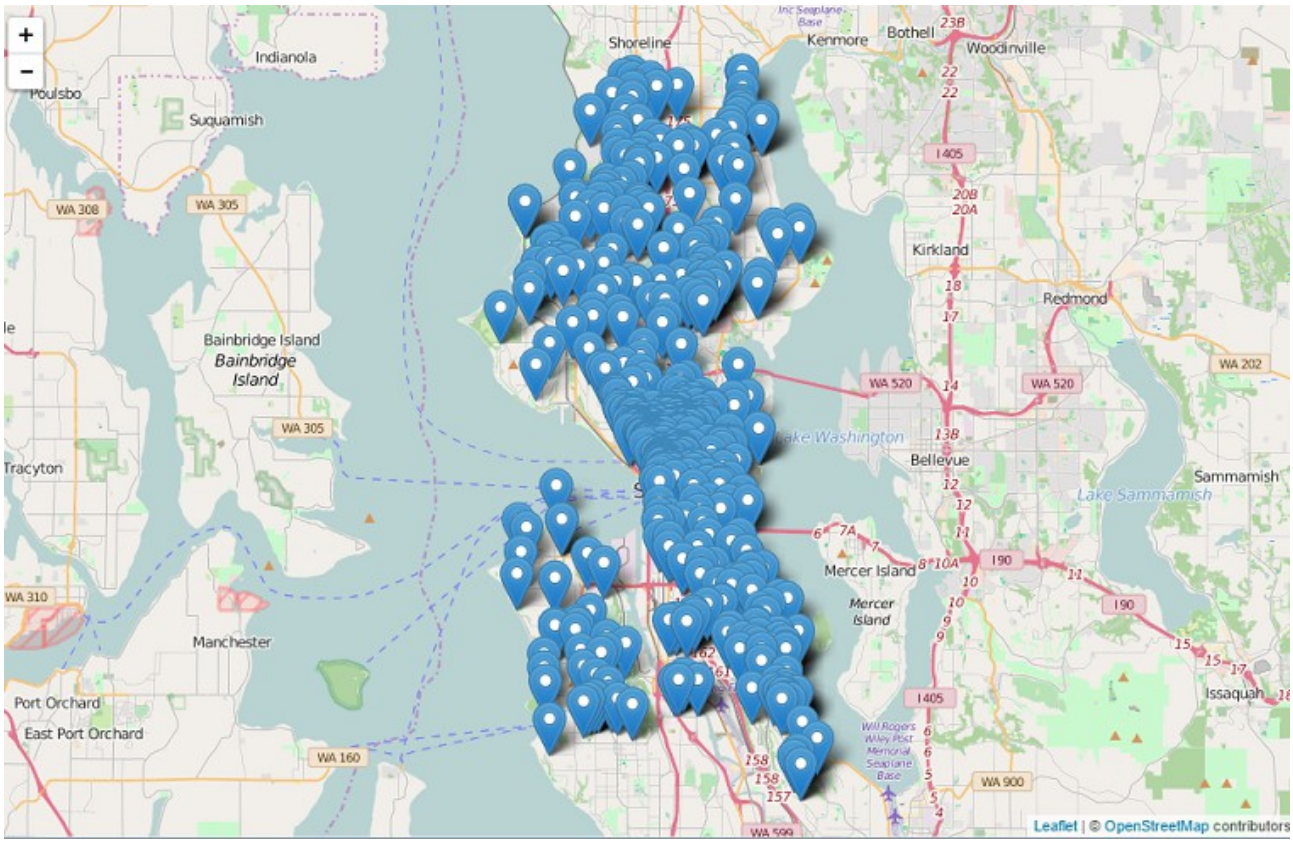
Map of San Francisco Showing some crimes a tourist might be concerned about



A look at the map suggests a tourist is more likely to be safer in Southern San Francisco.

Data Science At Scale : Visualization of Incident Reports

Map of Seattle Showing some crimes a tourist might be concerned about



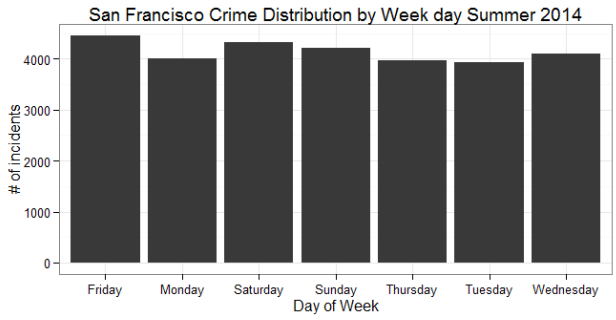
A look at the map suggests a tourist might be safer in Redmond area.

Data Science At Scale : Visualization of Incident Reports

From the perspective of law enforcement, it might be useful to visualize the crime data as a set of histograms.

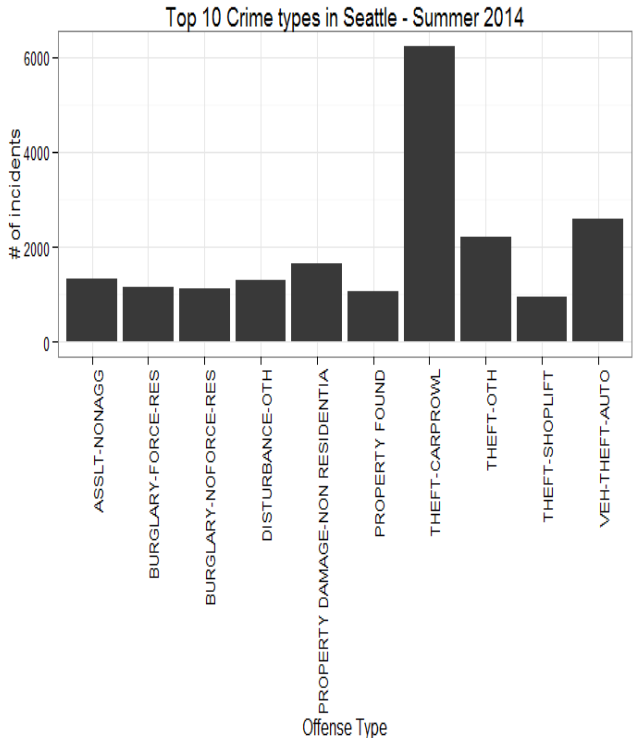
```
sf.crime.data.dayOfWeek <-
table(sf.crime.data$DayOfWeek)

sf.crime.data.dayOfWeek <-
as.data.frame(sf.crime.data.dayOfWeek)

sf.crime.data.dayOfWeek$day <-
row.names(sf.crime.data.dayOfWeek)

sf.crime.data.dayOfWeek$day <-
factor(sf.crime.data.dayOfWeek$day)

sf.crime.data.dayOfWeek.sorted <-
sf.crime.data.dayOfWeek[ with(sf.crime.data.da
yOfWeek, order(-Freq)),]

DayOfWeek_sf_ggplot <-
ggplot( sf.crime.data.dayOfWeek, aes(x=Var1,
y=Freq)) + theme_bw() +
geom_bar(stat="identity") + xlab("Day of
Week") + ylab("# of incidents") + ggtitle("San
Francisco Crime Distribution by Week day
Summer 2014")

DayOfWeek_sf_ggplot
```

Most criminal activities in SF occurred on Fridays



SF Police Department might need to have more officers on patrol on Fridays in Summer 2015 assuming the trend holds.

```
seattle.crime.category <-
table(seattle.crime.data$Offense.Type)
seattle.crime.category<-
as.data.frame(seattle.crime.category)
seattle.crime.category$crime<-
row.names(seattle.crime.category)
seattle.crime.category$crime<-
factor(seattle.crime.category$crime)

#sort crime by frequency
seattle.crime.category.sorted <-
seattle.crime.category[ with(seattle.crime.catego
ry, order(-Freq)),]
top_10_crime_seattle <-
head(seattle.crime.category.sorted, 10)
top_10_crime_seattle

top_10_crime_seattle_ggplot <-
ggplot( top_10_crime_seattle, aes(x=Var1,
y=Freq)) + theme_bw() +
geom_bar(stat="identity") + xlab("Offense
Type") + ylab("# of incidents") + ggtitle("Top
10 Crime types in Seattle - Summer 2014") +
theme(axis.text.x = element_text(angle = 90,
hjust = 1))
top_10_crime_seattle_ggplot
```



Car theft was the most common criminal offense in Seattle in Summer 2014.

Data Science At Scale : Visualization of Incident Reports