

# Crime Visualisation of San Francisco - Analytics Report

*Paulo Cardoso*

*March 06, 2016*

## Overview

This is an investigation that will be an Exploratory Data Analysis focused on Visualisations. The object of this analysis are the incidents of crimes in the summer of 2014 on Seattle and San Francisco. The data set used was provided by Coursera and is available on: ([https://github.com/uwescience/datasci\\_course\\_materials/blob/master/assignment6/sanfrancisco\\_incidents\\_summer\\_2014.csv](https://github.com/uwescience/datasci_course_materials/blob/master/assignment6/sanfrancisco_incidents_summer_2014.csv))

This report has for objective to answer the following questions:

- I) What is the time when most incidents occur?
- II) How frequently incidents occur?
- III) What is the distribution of the 5 most important occurrences by day of the week?
- IV) What is the distribution of incidents by district?

## Set up

Loading the external libraries and the data.

```
# Loading Libraries
```

```
library(ggplot2)
library(scales)
library(ggmap)
```

```
## Warning: package 'ggmap' was built under R version 3.2.3
```

```
# Loading Data
```

```
dSanF <- read.csv(file = "sanfrancisco_incidents_summer_2014.csv", stringsAsFactors = FALSE)
```

```
# Cleaning Data
```

```
dSanF$Category <- factor(dSanF$Category)
dSanF$DayOfWeek <- factor(dSanF$DayOfWeek)
dSanF$date <- as.Date(dSanF$date, "%m/%d/%Y")
dSanF$time <- as.POSIXct(dSanF$time, format="%H:%M")
dSanF$hour <- as.POSIXlt(dSanF$time)$hour
```

```
dSanF$Category <-
```

```
  factor(dSanF$Category, levels(dSanF$Category)[order(summary(dSanF$Category), decreasing = T)])
```

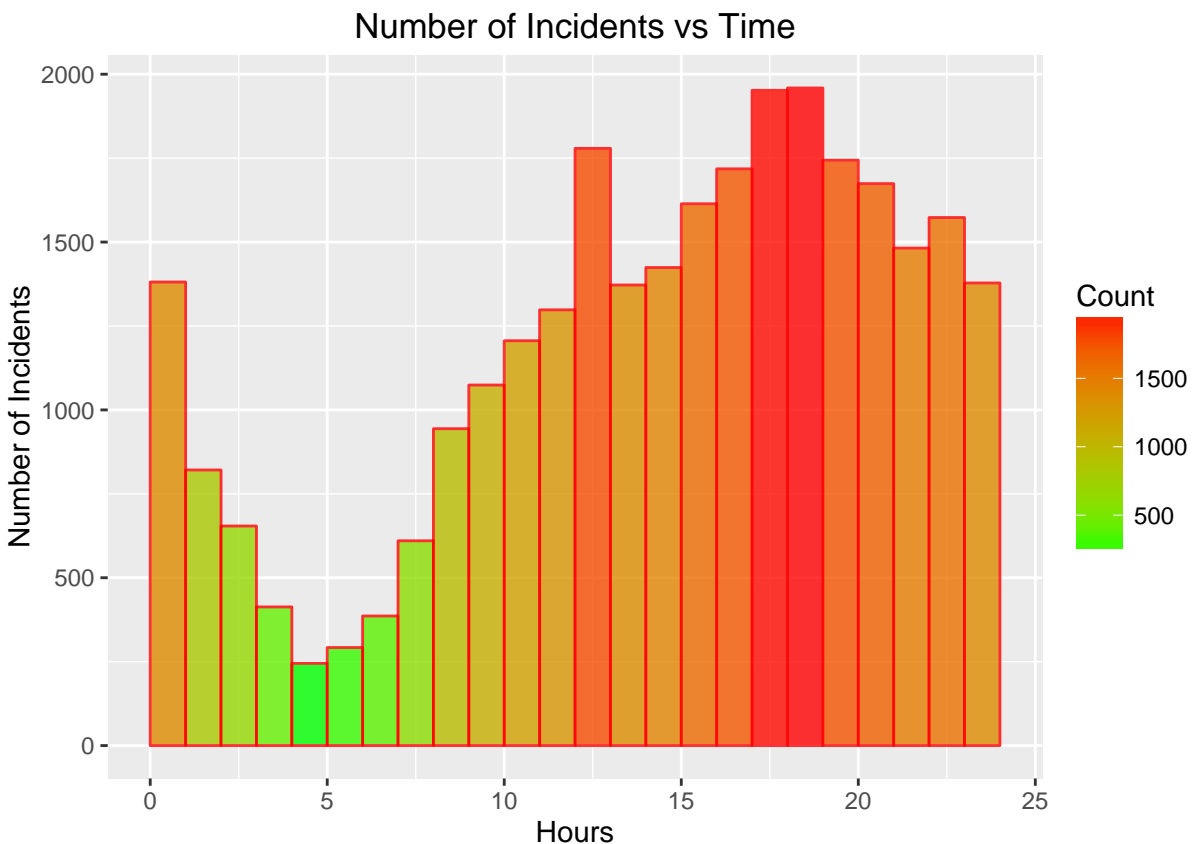
```
dSanF$DayOfWeek <- factor(dSanF$DayOfWeek, c("Sunday", "Monday", "Tuesday", "Wednesday", "Thursday", "Friday", "Saturday"))
```

## Exploratory Data Analyses

I) What is the time when most incidents occur?

R.: As can be observed below, the time of day with the highest incidence of crimes is between 6:00 and 7:59 pm. With almost 2,000 incidents per hour. On the other hand the safest time of day is between 4:00 and 7:59 am, with less than 500 occurrences recorded per hour.

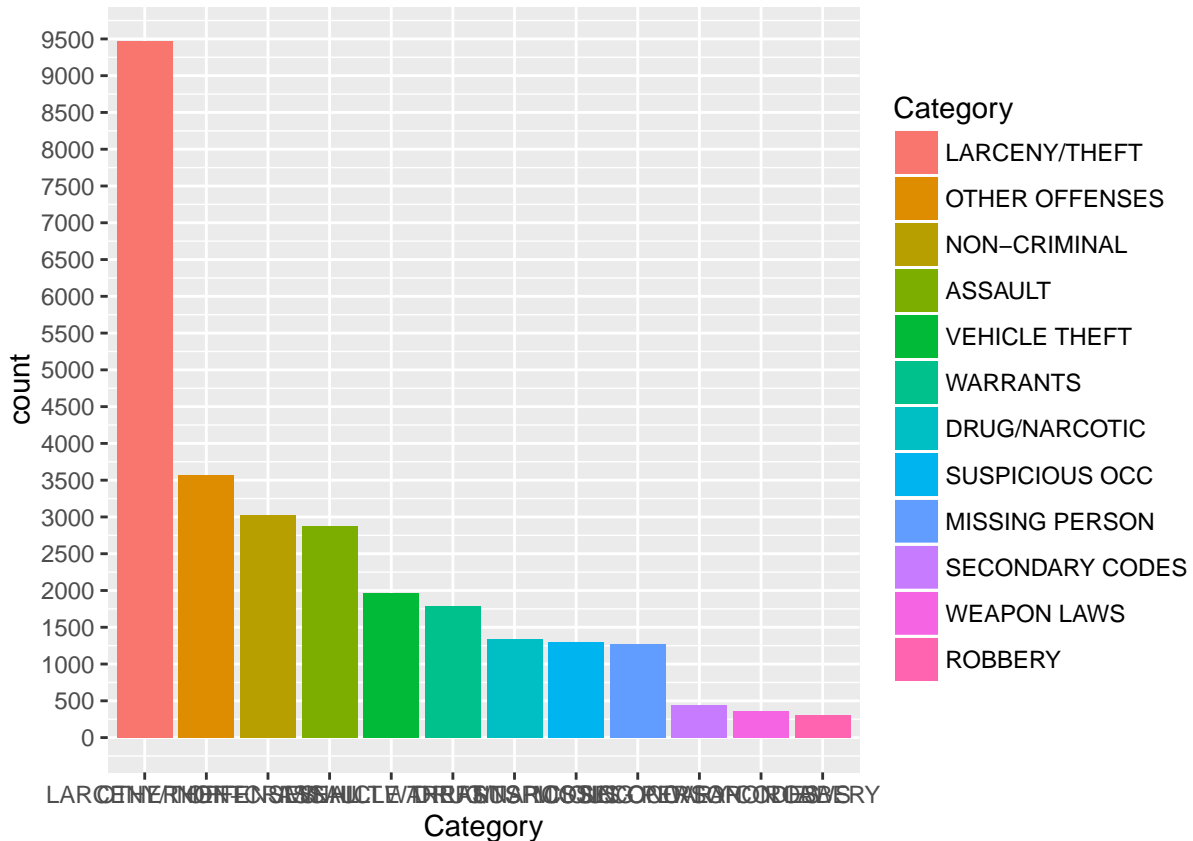
```
ggplot(data=dSanF, aes(dSanF$hour)) +  
  geom_histogram(breaks=seq(0, 24, by = 1), alpha = 0.8, col="red", aes(fill=..count..)) + scale_fill_g
```



II) How frequently incidents occur?

R.: The frequency of the 11 main incidents can be seen below. As we can see the huge amount of larceny by registering a number of occurrences of almost 9500 cases. Amount that is more than 2.5 times the amount of the second type of crime more occurrences.

```
dataP2 <- subset(dSanF, Category %in% names(head(sort(summary(dSanF$Category), decreasing = T), sum((c  
p2 <- ggplot(dataP2, aes(x=Category, fill=Category)) + geom_bar(stat="count") + scale_y_continuous(  
p2
```



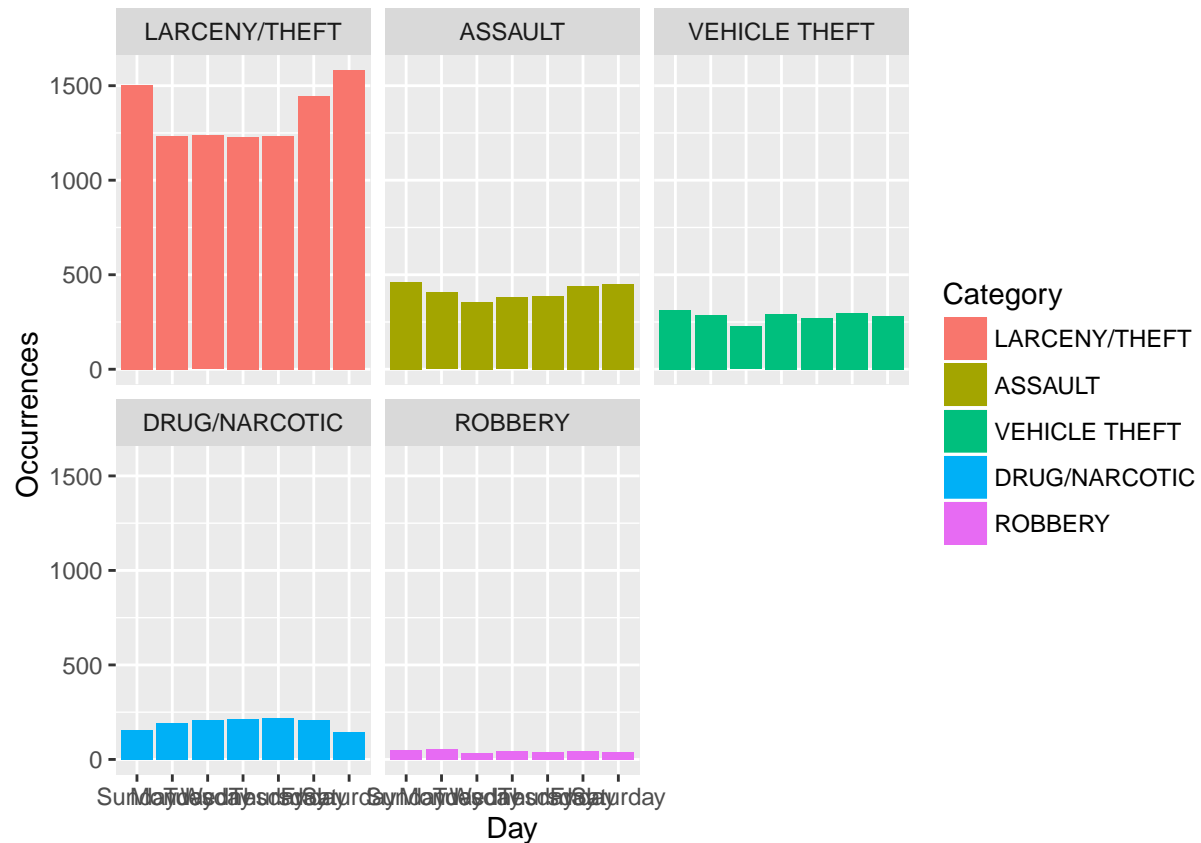
III) What is the distribution of the 5 most important occurrences by day of the week?

The graph below shows the distribution of the 5 most important occurrences per day of the week. Also it is observed that the crimes of aggression, vehicle theft, drug / narcotic and robbery has virtually stable trends during the week. On the other hand Lacerny / theft is stable from Monday to Thursday, rising on Friday and reaching its highest point during the weekend.

```
setCrimes <- subset( dSanF, dSanF$Category %in% levels(dSanF$Category)[c(1, 4, 5, 7, 12)])

dataP3 <- aggregate(setCrimes$Category, by=list(Category=setCrimes$Category, Day=setCrimes$DayOfWeek),
names(dataP3)[3] <- "Occurrences"

ggplot(dataP3, aes(x=Day, y=Occurrences, fill=Category)) + geom_bar(stat="identity") + facet_wrap(~Category)
```



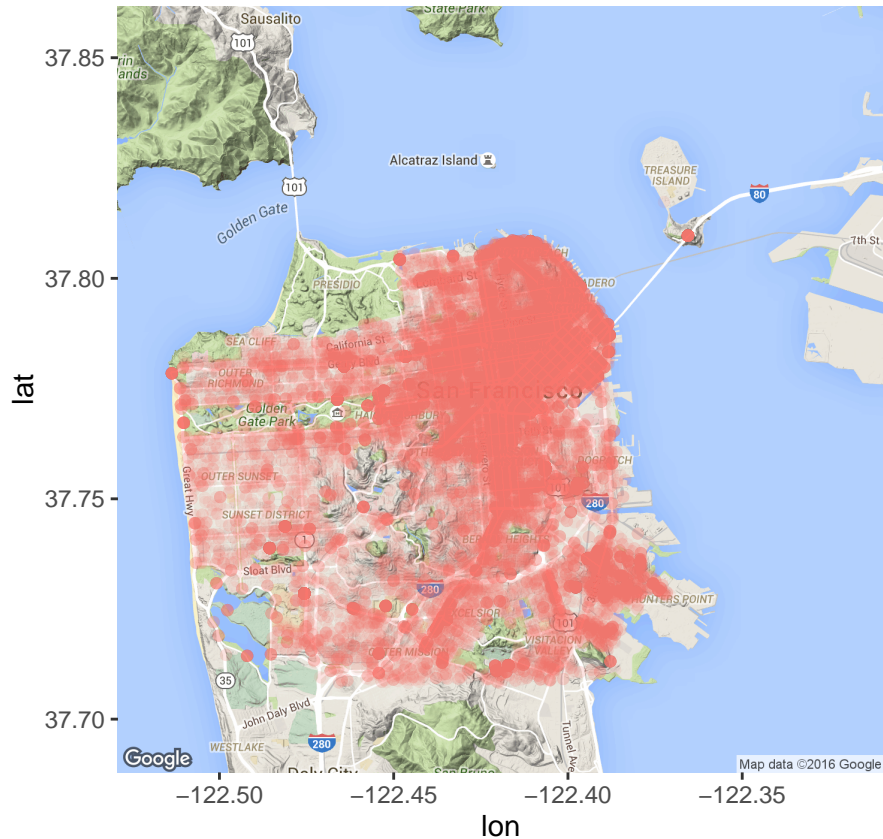
IV) What is the distribtion of incidents by district?

The map below shows the incidence of crime occurrences in areas of the city. And as can be seen the highest incidence of crimes is in the northeast of the city of San Francisco.

```
map <- get_map(location = 'San francisco', zoom = 12)
```

```
## Map from URL : http://maps.googleapis.com/maps/api/staticmap?center=San+francisco&zoom=12&size=640x640
## Information from URL : http://maps.googleapis.com/maps/api/geocode/json?address=San%20francisco&sens
```

```
ggmap(map) + geom_point(aes(x = X, y = Y, color='red'), data = dSanF, alpha = .1) + theme(legend.position = 'bottom')
```



## Conclusion

What can be concluded based on the analysis conducted above is:

- The region northeast of the city is the region with the highest incidence of reported criminal occurrences, so it is the most insecure area of the city.
- The preferred time by outlaw is from 6:00 to 7:59 pm. Time that they occur most amounts of crime, so is the period the more uncertain day. Already the safest day period is from 4:00 to 7:59 am.
- The type of crime most committed in the city of San Francisco is the crime of larceny, with almost 9500 cases in the summer of 2014. The number of larceny cases is so big that it is 2.5 times greater than the second.
- The cases of larceny increase on weekends.