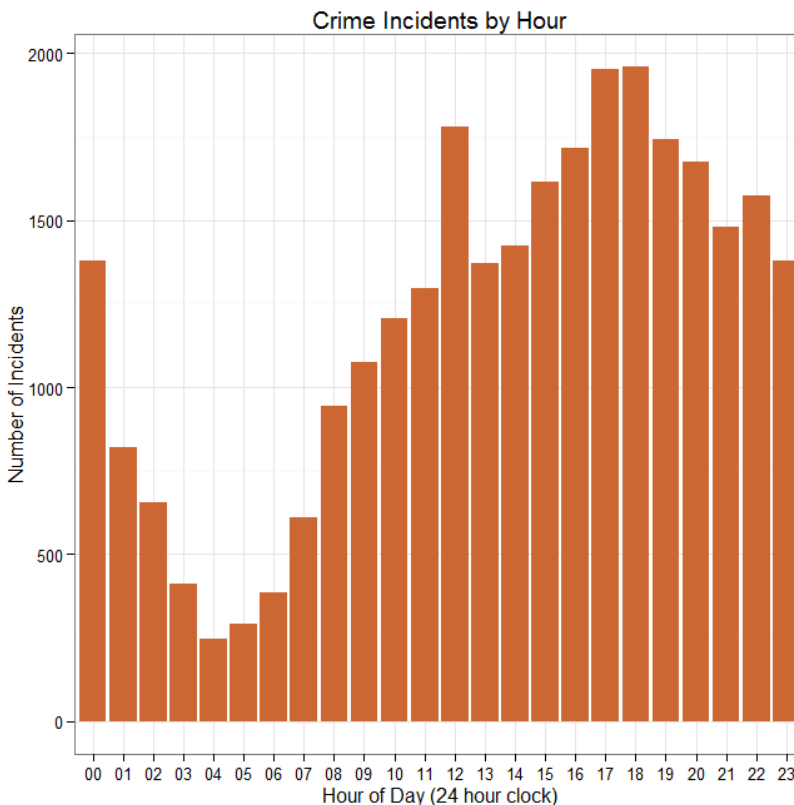


## Exploring the Relationship Between Temporal data and Crime Incidents in San Francisco

Most individuals would correctly assume that the frequency of crime incidents is correlated to temporal data. However, as we'll see in the following notes the incidence of crime appear to be more related to when people just get off of work more than anything else.



Contrary to popular belief most crime incidents do not occur in the early morning hours. In fact the 03:00AM – 06:00AM time block is when a crime incident is least likely to occur.

Un-intuitively we see the most amount of criminal incidents corresponding to when people typically get off work. There's noticeable spike at lunch time and the 05:00PM – 06:00PM time block.

```
# Read Crime Incident Data into a Dataframe
> setwd("C:\\Workspace\\datasci course materials\\assignment6");
> sfoData <- read.csv("sanfrancisco_incidents_summer_2014.csv");

> str(sfoData)
'data.frame':   28993 obs. of  13 variables:
 $ IncidentNum: int  140734311 140736317 146177923 146177531 140734220 140734349 140734349 140734349 ...
 $ Category   : Factor w/ 34 levels "ARSON","ASSAULT",...: 1 20 16 16 20 7 7 6 21 30 ...
 $ Descript   : Factor w/ 368 levels "ABANDONMENT OF CHILD",...: 15 179 143 143 132 247 239 93 107 347 ...
 $ DayOfWeek  : Factor w/ 7 levels "Friday","Monday",...: 4 4 4 4 4 4 4 4 4 4 ...
 $ Date       : Factor w/ 92 levels "06/01/2014","06/02/2014",...: 92 92 92 92 92 92 92 92 92 ...
 $ Time       : Factor w/ 1379 levels "00:01","00:02",...: 1370 1365 1351 1351 1344 1334 1334 1334 1321 1321 ...
 $ PdDistrict: Factor w/ 10 levels "BAYVIEW","CENTRAL",...: 1 4 8 7 7 8 8 8 3 2 ...
 $ Resolution: Factor w/ 16 levels "ARREST, BOOKED",...: 12 12 12 12 12 1 1 1 12 2 ...
 $ Address    : Factor w/ 8055 levels "0 Block of 10TH ST",...: 6843 4022 1098 6111 5096 1263 1263 1263 1575...
 $ X          : num  -122 -122 -122 -122 -123 ...
 $ Y          : num  37.7 37.8 37.8 37.8 37.8 ...
 $ Location   : Factor w/ 8732 levels "(37.7080829769301, -122.419241455854)",...: 1970 3730 5834 4802 7598 ...
```

```
$ PdId      : num  1.41e+13 1.41e+13 1.46e+13 1.46e+13 1.41e+13 ...

# Extract the hour of day from the Time Factor
> sfoData$HourOfDay <- substr(sfoData$Time, 0, 2);

# Plot Crime Incidents by Hour
> ggplot(as.data.frame(table(sfoData$HourOfDay))) +
  geom_bar(aes(Var1, Freq), stat = "identity", fill = "#CC6633") + theme_bw() +
  labs(x = "Hour of Day (24 hour clock)", y = "Number of Incidents", title = "Crime Incidents by Hour");
```