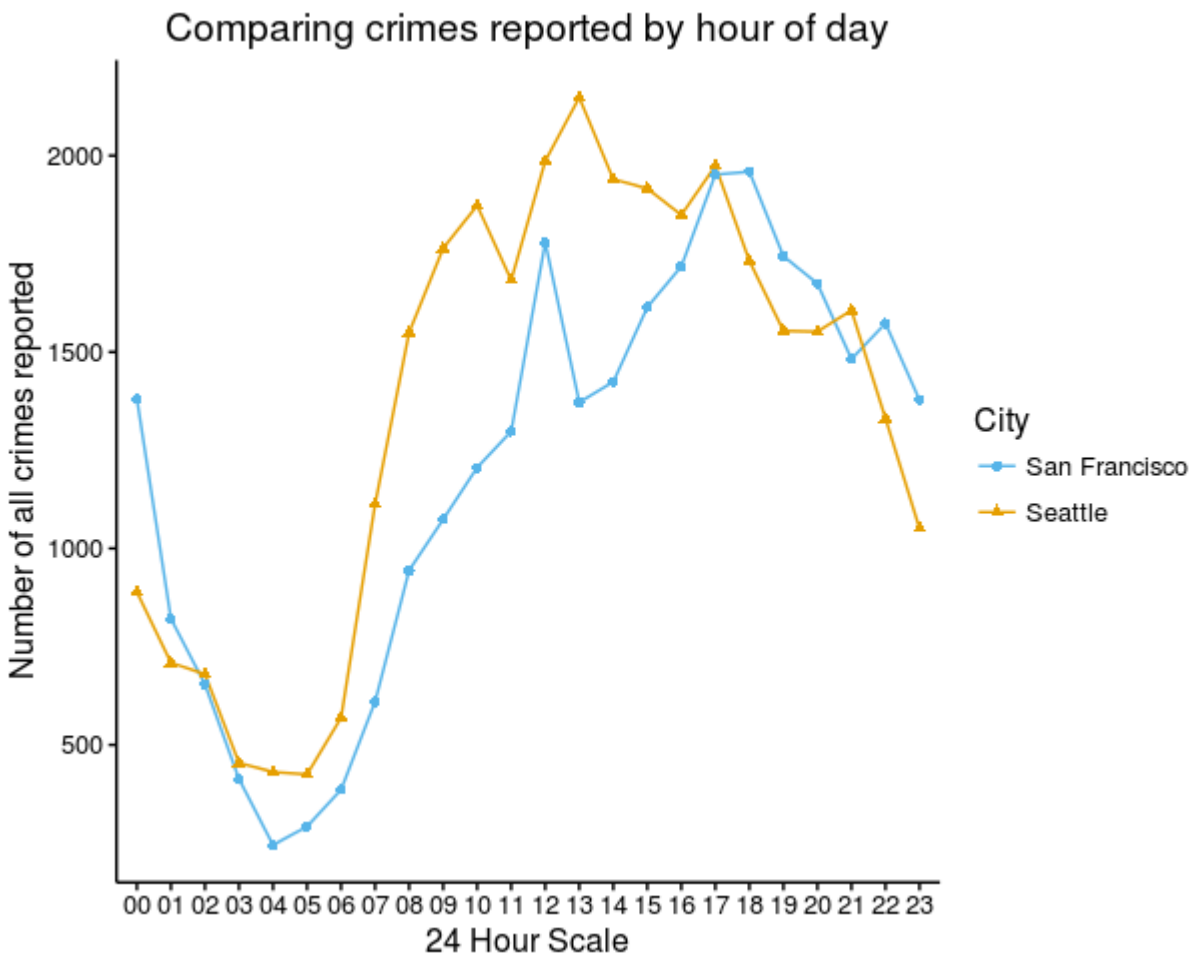


## Thefts peak between 12 noon and 7 pm in Seattle

NOTE I tried to use Jupyter and Rmarkdown to provide an interactive script but it was too time-consuming to try and learn these and still submit the assignment in reasonable time. I have included my full R script at the end of this document.

Since I am not US based, I decided to compare the crimes reported between San Francisco and Seattle, and see if there were any striking similarities or discrepancies between the two cities.

San Francisco central had about 855,000 inhabitants in 2014 compared to Seattle with about 670,000, according to Wikipedia. In other words, Seattle has about 80% or 4/5 the inhabitants of San Francisco. I expected to see roughly this ratio in crimes as well. If not, then one city would be more prone than the other to certain, or perhaps all types of crime. Next, I was interested to compare the crimes reported on an hour by hour basis throughout the entire day and night.

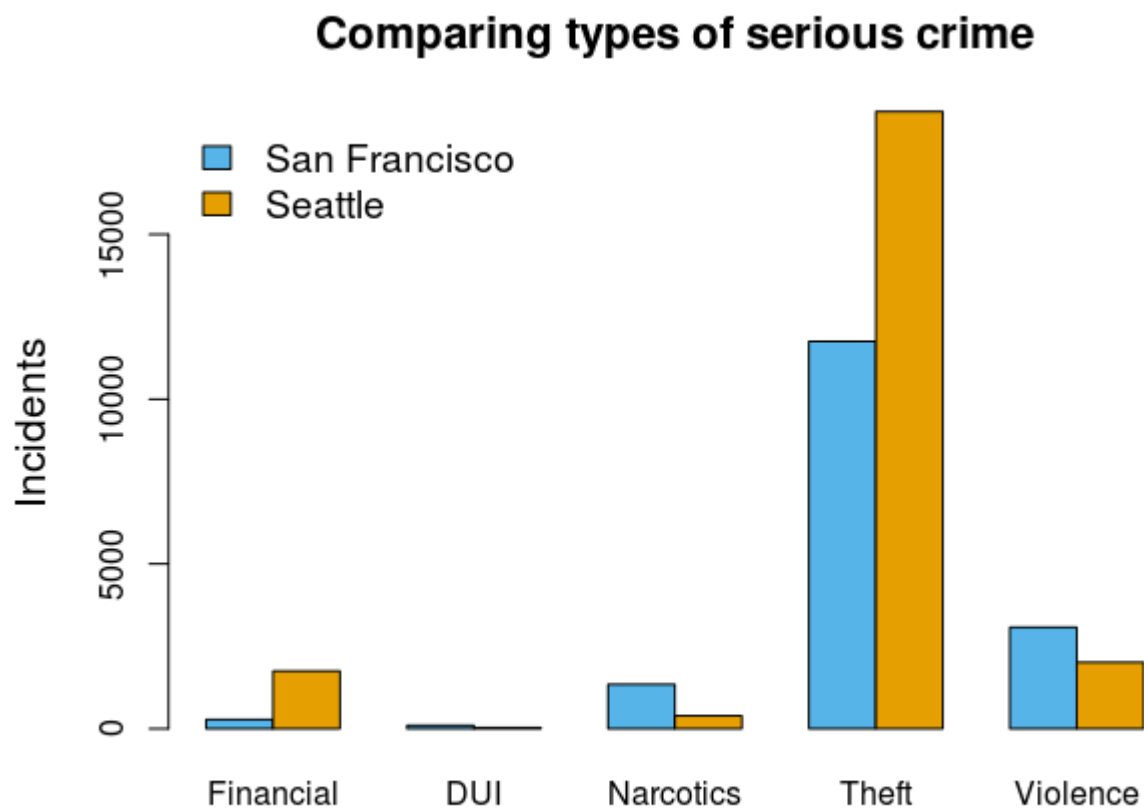


As can be seen, the total number of crimes reported in Seattle exceeded those in San Francisco for most of the day. It's only in the evening up to just past midnight where there were more crimes reported.

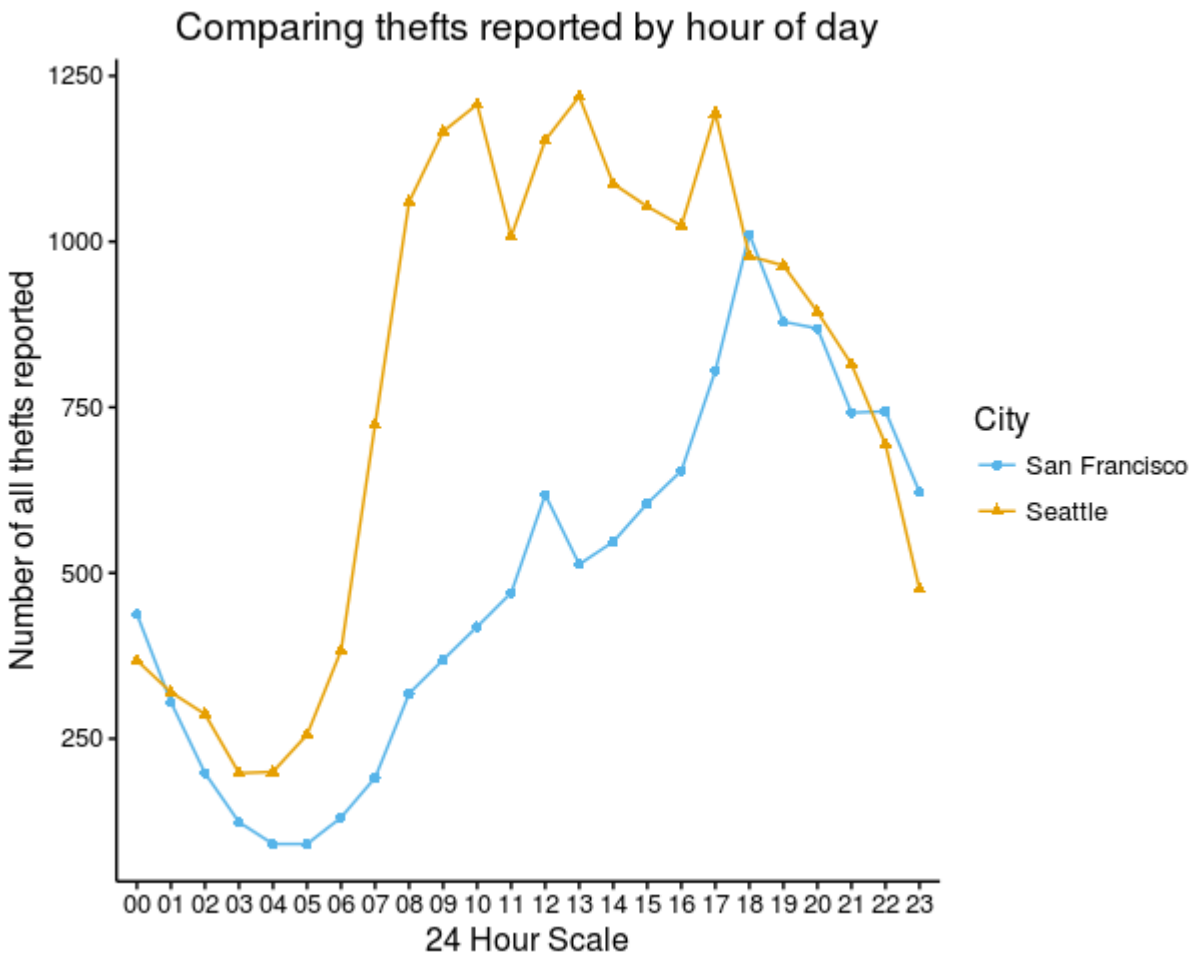
I then decided to examine the types of crime. Generally speaking the US seems to distinguish between infractions (very minor), misdemeanors (can include jail time), and felonies (serious crime, and leading to incarceration in a penitentiary (or worse).

The categorisation differs from state to state, so I decided to create my own categories of the types of crime to compare. I wanted to remove all the infractions, and then categorise the remaining crimes as:

- Financial (including extortion, bribery etc)
- Driving under the influence (alcohol or drugs)
- Narcotics (trafficking, smuggling etc)
- Theft (grouping together burglary etc – in other words where the victim is not present, and purse snatching etc – where the victim is not injured)
- Violence – where the victim is either threatened with violence or where actual violence is used against the victim.



As can be seen, Seattle exceeded the 4 to 5 ratio for financial and crimes related to theft, so I decided to examine when during the day thefts tend to occur when comparing San Francisco and Seattle.



As can be seen, in San Francisco thefts peaked at round 6 to 7pm, i.e. the time when most people leave the office, whereas in Seattle thefts peaked and remained high the entire working day and afterwards, i.e. between 8am and 8pm.

The overall conclusion is thus that one has to be especially vigilant the entire day when in central Seattle.

My code:

```
setwd("~/datasci_course_materials/assignment6/")
```

```
library(plyr)
```

```
library(ggplot2)
```

```
library(reshape2)
```

```
##### Plot 1 - compare all data by hour
```

```
sf_incidents <- read.csv("~/datasci_course_materials/assignment6/sanfrancisco_incidents_summer_2014.csv")
```

```
sea_incidents <- read.csv("~/datasci_course_materials/assignment6/seattle_incidents_summer_2014.csv")
```

```
sf_incidents$hr24 <- sapply(as.character(sf_incidents$Time), FUN=function(x) {strsplit(x, split=':')[[1]][1]})
```

```
table(sf_incidents$hr24) # San Francisco incidents by the hour
```

```
sea_incidents$hr24 <- sapply(as.character(strptime(sea_incidents$Date.Reported, "%m/%d/%Y %l:%M:%S %p")), FUN=function(x) {strsplit(x, split='[-:" "]')[[1]][4]})
```

```
table(sea_incidents$hr24) # Seattle incidents by the hour
```

```
# Data hardcoded / transposed from the two tables above due to my lack of skill with R
```

```
sfsea_df <- data.frame(City=rep(c("Seattle", "San Francisco"),  
each=24), Hour=rep(c("00", "01", "02", "03", "04", "05", "06", "07", "08", "09", "10", "11", "12", "13", "14", "15", "16", "17",  
"18", "19", "20", "21", "22", "23"), 2),  
Crimes=c(889,709,680,454,431,425,569,1113,1550,1764,1873,1683,1986,2148,1940,1916,1849,1974,1732,155  
4,1552,1606,1331,1051,  
1381,821,654,413,245,292,386,610,944,1074,1206,1298,1779,1372,1424,1614,1718,1952,1959,1744,1674,148  
2,1573,1378))
```

```
p <- ggplot(data=sfsea_df, aes(x=Hour, y=Crimes, group=City, color=City, shape=City)) +
```

```
  geom_line() +
```

```
  geom_point() +
```

```
  labs(title="Comparing crimes reported by hour of day", x="24 Hour Scale", y="Number of all crimes  
reported") +
```

```
  theme_classic()
```

```
p + theme(axis.line.x = element_line(color = "black", size =.5), axis.line.y = element_line(color = "black", size  
=.5)) + scale_color_manual(values=c('#56B4E9', '#E69F00'))
```

##### Plot 2 - break out the serious crime for comparison

# Categorise crimes in San Francisco and Seattle

```
cat_violence <- c('ASSAULT','SUICIDE','KIDNAPPING','ROBBERY','HOMICIDE','THREATS','ARSON',
  'ASSLT-AGG-BODYFORCE','ASSLT-AGG-GUN','ASSLT-POLICE-BODYFORCE','ASSLT-AGG-POLICE-
GUN','ASSLT-AGG-POLICE-WEAPON',
  'ASSLT-AGG-WEAPON','ROBBERY-BANK-BODYFORCE','ROBBERY-BANK-GUN','ROBBERY-BANK-
WEAPON','ROBBERY-BUSINESS-BODYFORCE',
  'ROBBERY-BUSINESS-GUN','ROBBERY-BUSINESS-WEAPON','ROBBERY-RESIDENCE-
BODYFORCE','ROBBERY-RESIDENCE-GUN','ROBBERY-RESIDENCE-WEAPON',
  'ROBBERY-STREET-BODYFORCE','ROBBERY-STREET-GUN','ROBBERY-STREET-WEAPON','THREATS-
KILL','THREATS-OTHER','THREATS-WEAPON',
  'HOMICIDE-JUST-GUN','HOMICIDE-JUST-WEAPON','HOMICIDE-PREMEDITATED-GUN')
cat_narc <- c('DRUG/NARCOTIC',
  'NARC-DRUG TRAFFIC LOITERING','NARC-EQUIPMENT/PARAPHENALIA','NARC-FORGERY-
PRESCRIPTION','NARC-FOUND-AMPHETAMINE',
  'NARC-FOUND-COCAINE','NARC-FOUND-HEROIN','NARC-FOUND-MARIJU','NARC-FOUND-
METH','NARC-FOUND-OPIUM','NARC-FOUND-OTHER',
  'NARC-FOUND-SYNTHETIC','NARC-FRAUD-PRESCRIPTION','NARC-POSSESS-AMPHETAMINE','NARC-
POSSESS-COCAINE','NARC-POSSESS-HALLUCINOGEN',
  'NARC-POSSESS-HEROIN','NARC-POSSESS-MARIJU','NARC-POSSESS-METH','NARC-POSSESS-
OTHER','NARC-POSSESS-PILL/TABLET',
  'NARC-PRODUCE-MARIJU','NARC-SELL-AMPHETAMINE','NARC-SELL-COCAINE','NARC-SELL-
HEROIN','NARC-SELL-MARIJU','NARC-SELL-METH',
  'NARC-SELL-SYNTHETIC','NARC-SMUGGLE-OTHER')
cat_financial <-
c('BRIBERY','EMBEZZLEMENT','EXTORTION','FORGERY/COUNTERFEITING','FRAUD','GAMBLING',
  'COUNTERFEIT','EMBEZZLE','FORGERY-CHECK','FORGERY-CREDIT CARD','FORGERY-OTH','FRAUD-
CHECK','FRAUD-COMPUTER',
  'FRAUD-CREDIT CARD','FRAUD-IDENTITY THEFT','FRAUD-OTHER','FRAUD-WIRE-ELECTRONIC')
cat_theft <- c('BURGLARY','LARCENY/THEFT','ROBBERY','STOLEN PROPERTY','VEHICLE THEFT',
  'BURGLARY-FORCE-NONRES','BURGLARY-FORCE-RES','BURGLARY-NOFORCE-NONRES','BURGLARY-
NOFORCE-RES','BURGLARY-SECURE PARKING-NONRES',
  'BURGLARY-SECURE PARKING-RES','THEFT-AUTOACC','THEFT-AUTO PARTS','THEFT-BICYCLE','THEFT-
BOAT','THEFT-BUILDING',
  'THEFT-CARPROWL','THEFT-COINOP','THEFT-LICENSE PLATE','THEFT-MAIL','THEFT OF
SERVICES','THEFT-OTH','THEFT-PKPOCKET',
  'THEFT-PRSNATCH','THEFT-SHOPLIFT','VEH-THEFT-AUTO','VEH-THEFT-MTRCYCLE','VEH-THEFT-
OTHVEH','VEH-THEFT-TRAILER',
  'VEH-THEFT-TRUCK')
cat_dui <- c('DRIVING UNDER THE INFLUENCE','DUI-LIQUOR','DUI-DRUGS')
```

```
# San Francisco categorisation
```

```
sf_incidents$cat <- 'Other'  
sf_incidents$cat[sf_incidents$Category %in% cat_violence] <- 'Violence'  
sf_incidents$cat[sf_incidents$Category %in% cat_narc] <- 'Narcotics'  
sf_incidents$cat[sf_incidents$Category %in% cat_financial] <- 'Financial'  
sf_incidents$cat[sf_incidents$Category %in% cat_theft] <- 'Theft'  
sf_incidents$cat[sf_incidents$Category %in% cat_dui] <- 'DUI'
```

```
# Seattle categorisation
```

```
sea_incidents$cat <- 'Other'  
sea_incidents$cat[sea_incidents$Offense.Type %in% cat_violence] <- 'Violence'  
sea_incidents$cat[sea_incidents$Offense.Type %in% cat_narc] <- 'Narcotics'  
sea_incidents$cat[sea_incidents$Offense.Type %in% cat_financial] <- 'Financial'  
sea_incidents$cat[sea_incidents$Offense.Type %in% cat_theft] <- 'Theft'  
sea_incidents$cat[sea_incidents$Offense.Type %in% cat_dui] <- 'DUI'
```

```
table(sea_incidents$cat)  
table(sf_incidents$cat)
```

```
# Data hardcoded / transposed from the two tables above due to my lack of skill with R
```

```
data <- structure(list(Fin=c(279L, 1748L), DUI=c(100L, 34L), Narc=c(1345L, 391L),  
  Theft=c(11754L, 18728L), Viol=c(3076L, 2011L)),  
  .Names=c("Financial", "DUI", "Narcotics", "Theft", "Violence"),  
  class="data.frame",  
  row.names = c(NA, -2L))  
attach(data)
```

```
colours <- c('#56B4E9', '#E69F00')
```

```
barplot(as.matrix(data), main="Comparing types of serious crime", ylab = "Incidents",  
  cex.lab = 1.3, cex.main = 1.4, beside = TRUE, col=colours)
```

```
legend("topleft", c("San Francisco", "Seattle"), cex=1.2, bty="n", fill=colours)
```

```
##### Plot 3 - compare thefts by time of day #####
```

```
sub_sf_theft <- subset(sf_incidents, cat %in% c('Theft'))
```

```
sub_sf_theft$hr24 <- sapply(as.character(sub_sf_theft$Time), FUN=function(x) {strsplit(x, split=':')[[1]][1]})
```

```
table(sub_sf_theft$hr24)
```

```
sub_sea_theft <- subset(sea_incidents, cat %in% c('Theft'))
```

```
sub_sea_theft$hr24 <- sapply(as.character(strptime(sub_sea_theft$Date.Reported, "%m/%d/%Y %l:%M:%S %p")), FUN=function(x) {strsplit(x, split=':-" "')[[1]][4]})
```

```
table(sub_sea_theft$hr24)
```

```
# Data hardcoded / transposed from the two tables above due to my lack of skill with R
```

```
sfsea_theft_df <- data.frame(City=rep(c("San Francisco", "Seattle"),  
each=24), Hour=rep(c("00", "01", "02", "03", "04", "05", "06", "07", "08", "09", "10", "11", "12", "13", "14", "15", "16", "17",  
"18", "19", "20", "21", "22", "23"), 2),  
Thefts=c(438,305,198,124,91,91,131,191,318,369,419,470,618,513,547,605,654,805,1011,879,869,742,744,62  
2,  
368,320,287,198,200,256,383,724,1060,1166,1207,1008,1153,1219,1087,1053,1024,1194,978,964,894,815,69  
4,476))
```

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

```
p2 <- ggplot(data=sfsea_theft_df, aes(x=Hour, y=Thefts, group=City, color=City, shape=City)) +  
  geom_line() +  
  geom_point() +  
  labs(title="Comparing thefts reported by hour of day", x="24 Hour Scale", y="Number of all thefts reported")  
+  
  theme_classic()
```

```
p2 + theme(axis.line.x = element_line(color = "black", size =.5), axis.line.y = element_line(color = "black", size  
=.5)) + scale_color_manual(values=c('#56B4E9', '#E69F00'))
```