Coursera Crime Visualization

Javad Hashtroudian

October 10, 2017

Load needed libraries

load libraries and set working directory and clean workspace

```
knitr::opts_chunk$set(echo = FALSE)
library(ggplot2)
library(dplyr)
```

```
##
## Attaching package: 'dplyr'
```

```
## The following objects are masked from 'package:stats':
##
##     filter, lag
```

```
## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

```
library(tidyr)

setwd("c:/assignment6")
rm(list = ls())
```

load both csvs

and do a little EDA (expolrotory data analysis)

```
##   IncidntNum      Category                      Descript DayOfWeek
## 1  140734311         ARSON            ARSON OF A VEHICLE    Sunday
## 2  140736317  NON-CRIMINAL                LOST PROPERTY    Sunday
## 3  146177923 LARCENY/THEFT GRAND THEFT FROM LOCKED AUTO    Sunday
## 4  146177531 LARCENY/THEFT GRAND THEFT FROM LOCKED AUTO    Sunday
## 5  140734220  NON-CRIMINAL               FOUND PROPERTY    Sunday
## 6  140734349 DRUG/NARCOTIC      POSSESSION OF MARIJUANA    Sunday
##         Date  Time PdDistrict     Resolution                   Address
## 1 08/31/2014 23:50    BAYVIEW           NONE LOOMIS ST / INDUSTRIAL ST
## 2 08/31/2014 23:45    MISSION           NONE    400 Block of CASTRO ST
## 3 08/31/2014 23:30   SOUTHERN           NONE  1000 Block of MISSION ST
## 4 08/31/2014 23:30   RICHMOND           NONE       FULTON ST / 26TH AV
## 5 08/31/2014 23:23   RICHMOND           NONE  800 Block of LA PLAYA ST
## 6 08/31/2014 23:13   SOUTHERN ARREST, BOOKED        11TH ST / MINNA ST
##           X        Y                             Location         PdId
## 1 -122.4056 37.73832 (37.7383221869053, -122.405646994567) 1.407343e+13
## 2 -122.4350 37.76177 (37.7617677182954, -122.435012093789) 1.407363e+13
## 3 -122.4098 37.78004 (37.7800356268394, -122.409795194505) 1.461779e+13
## 4 -122.4853 37.77252 (37.7725176473142, -122.485262988324) 1.461775e+13
## 5 -122.5099 37.77231 (37.7723131976814, -122.509895418239) 1.407342e+13
## 6 -122.4166 37.77391  (37.773907074489, -122.416578493475) 1.407343e+13
```

```
##   RMS.CDW.ID General.Offense.Number Offense.Code Offense.Code.Extension
## 1     483839             2015218538         2202                      0
## 2     481252             2015213067         2610                      0
## 3     481375             2015210301         2316                      0
## 4     481690             2015209327         2599                      0
## 5     478198             2015207880         2399                      3
## 6     480485             2015904103         2308                      0
##          Offense.Type Summary.Offense.Code Summarized.Offense.Description
## 1   BURGLARY-FORCE-RES                2200                        BURGLARY
## 2 FRAUD-IDENTITY THEFT                2600                           FRAUD
## 3           THEFT-MAIL                2300                       MAIL THEFT
## 4          COUNTERFEIT                2500                      COUNTERFEIT
## 5            THEFT-OTH                2300                   OTHER PROPERTY
## 6       THEFT-BUILDING                2300                   OTHER PROPERTY
##           Date.Reported Occurred.Date.or.Date.Range.Start
## 1 06/28/2015 10:31:00 AM            06/28/2014 10:31:00 AM
## 2 06/24/2015 11:09:00 AM            06/01/2014 12:00:00 AM
## 3 06/22/2015 09:22:00 AM            08/31/2014 09:00:00 AM
## 4 06/21/2015 03:52:00 PM            06/20/2014 01:38:00 PM
## 5 06/20/2015 11:59:00 AM            06/01/2014 11:59:00 AM
## 6 06/19/2015 02:55:00 PM            06/19/2014 02:45:00 PM
##   Occurred.Date.Range.End       Hundred.Block.Location District.Sector
## 1  06/28/2015 10:31:00 AM        6XX BLOCK OF NW 74 ST               J
## 2  06/24/2015 11:09:00 AM        23XX BLOCK OF 43 AV E               C
## 3                                81XX BLOCK OF 11 AV SW              F
## 4                                  6XX BLOCK OF PINE ST              M
## 5  11/01/2014 12:00:00 PM 77XX BLOCK OF SUNNYSIDE AV N               J
## 6  07/10/2014 02:45:00 PM 35XX BLOCK OF S FERDINAND ST               R
##   Zone.Beat Census.Tract.2000 Longitude Latitude
## 1        J2          2900.301 -122.3647 47.68252
## 2        C2          6300.100 -122.2771 47.63990
## 3        F3         11300.501 -122.3493 47.52923
## 4        M2          8200.100 -122.3348 47.61237
## 5        J3          2700.202 -122.3294 47.68596
## 6        R3         10300.401 -122.2875 47.55785
##                          Location Month Year
## 1  (47.68252427, -122.364671996)     6 2014
## 2 (47.639900761, -122.277080248)     6 2014
## 3 (47.529232299, -122.349312181)     8 2014
## 4 (47.612368448, -122.334817763)     6 2014
## 5 (47.685959879, -122.329378505)     6 2014
## 6 (47.557854802, -122.287477902)     6 2014
```

```
## 'data.frame':    28993 obs. of  13 variables:
##  $ IncidntNum: int  140734311 140736317 146177923 146177531 140734220 140734349 140734349 140734349 140738147
##  $ Category  : chr  "ARSON" "NON-CRIMINAL" "LARCENY/THEFT" "LARCENY/THEFT" ...
##  $ Descript  : chr  "ARSON OF A VEHICLE" "LOST PROPERTY" "GRAND THEFT FROM LOCKED AUTO" "GRAND THEFT FROM LOC
##  $ DayOfWeek : chr  "Sunday" "Sunday" "Sunday" "Sunday" ...
##  $ Date      : chr  "08/31/2014" "08/31/2014" "08/31/2014" "08/31/2014" ...
##  $ Time      : chr  "23:50" "23:45" "23:30" "23:30" ...
##  $ PdDistrict: chr  "BAYVIEW" "MISSION" "SOUTHERN" "RICHMOND" ...
##  $ Resolution: chr  "NONE" "NONE" "NONE" "NONE" ...
##  $ Address   : chr  "LOOMIS ST / INDUSTRIAL ST" "400 Block of CASTRO ST" "1000 Block of MISSION ST" "FULTON S
##  $ X         : num  -122 -122 -122 -122 -123 ...
##  $ Y         : num  37.7 37.8 37.8 37.8 37.8 ...
##  $ Location  : chr  "(37.7383221869053, -122.405646994567)" "(37.7617677182954, -122.435012093789)" "(37.7800
##  $ PdId      : num  1.41e+13 1.41e+13 1.46e+13 1.46e+13 1.41e+13 ...
```

```
## 'data.frame':    32779 obs. of  19 variables:
##  $ RMS.CDW.ID                        : int  483839 481252 481375 481690 478198 480485 470170 465137 461710 456
##  $ General.Offense.Number            : num  2.02e+09 2.02e+09 2.02e+09 2.02e+09 2.02e+09 ...
##  $ Offense.Code                      : chr  "2202" "2610" "2316" "2599" ...
##  $ Offense.Code.Extension            : int  0 0 0 0 3 0 0 0 0 1 ...
##  $ Offense.Type                      : chr  "BURGLARY-FORCE-RES" "FRAUD-IDENTITY THEFT" "THEFT-MAIL" "COUNTERF
##  $ Summary.Offense.Code              : chr  "2200" "2600" "2300" "2500" ...
##  $ Summarized.Offense.Description    : chr  "BURGLARY" "FRAUD" "MAIL THEFT" "COUNTERFEIT" ...
##  $ Date.Reported                     : chr  "06/28/2015 10:31:00 AM" "06/24/2015 11:09:00 AM" "06/22/2015 09:2
##  $ Occurred.Date.or.Date.Range.Start: chr  "06/28/2014 10:31:00 AM" "06/01/2014 12:00:00 AM" "08/31/2014 09:0
##  $ Occurred.Date.Range.End           : chr  "06/28/2015 10:31:00 AM" "06/24/2015 11:09:00 AM" "" "" ...
##  $ Hundred.Block.Location            : chr  "6XX BLOCK OF NW 74 ST" "23XX BLOCK OF 43 AV E" "81XX BLOCK OF 11
##  $ District.Sector                   : chr  "J" "C" "F" "M" ...
##  $ Zone.Beat                         : chr  "J2" "C2" "F3" "M2" ...
##  $ Census.Tract.2000                 : num  2900 6300 11301 8200 2700 ...
##  $ Longitude                         : num  -122 -122 -122 -122 -122 ...
##  $ Latitude                          : num  47.7 47.6 47.5 47.6 47.7 ...
##  $ Location                          : chr  "(47.68252427, -122.364671996)" "(47.639900761, -122.277080248)" "
##  $ Month                             : int  6 6 8 6 6 6 6 8 8 7 ...
##  $ Year                              : int  2014 2014 2014 2014 2014 2014 2014 2014 2014 2014 ...
```

```
##    IncidntNum          Category           Descript
##  Min.   : 10284385   Length:28993       Length:28993
##  1st Qu.:140545607   Class :character   Class :character
##  Median :140632022   Mode  :character   Mode  :character
##  Mean   :142017280
##  3rd Qu.:140719664
##  Max.   :990367398
##   DayOfWeek             Date               Time
##  Length:28993       Length:28993       Length:28993
##  Class :character   Class :character   Class :character
##  Mode  :character   Mode  :character   Mode  :character
##
##
##
##   PdDistrict         Resolution          Address                X
##  Length:28993       Length:28993       Length:28993       Min.   :-122.5
##  Class :character   Class :character   Class :character   1st Qu.:-122.4
##  Mode  :character   Mode  :character   Mode  :character   Median :-122.4
##                                                           Mean   :-122.4
##                                                           3rd Qu.:-122.4
##                                                           Max.   :-122.4
##        Y           Location             PdId
##  Min.   :37.71   Length:28993       Min.   :1.028e+12
##  1st Qu.:37.76   Class :character   1st Qu.:1.405e+13
##  Median :37.78   Mode  :character   Median :1.406e+13
##  Mean   :37.77                      Mean   :1.420e+13
##  3rd Qu.:37.79                      3rd Qu.:1.407e+13
##  Max.   :37.82                      Max.   :9.904e+13
```

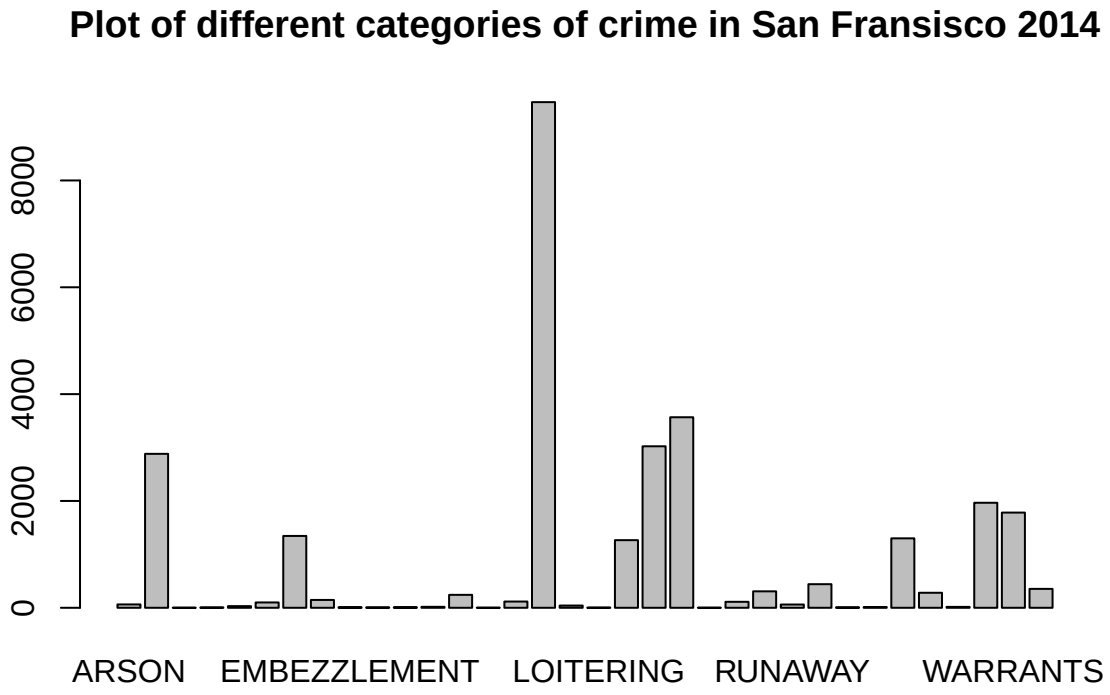
```
##    RMS.CDW.ID      General.Offense.Number Offense.Code
##  Min.   :  12470   Min.   :2.015e+07      Length:32779
##  1st Qu.:  47246   1st Qu.:2.014e+09      Class :character
##  Median : 999859   Median :2.014e+09      Mode  :character
##  Mean   : 578650   Mean   :2.020e+09
##  3rd Qu.:1030117   3rd Qu.:2.014e+09
##  Max.   :1125252   Max.   :2.014e+11
```

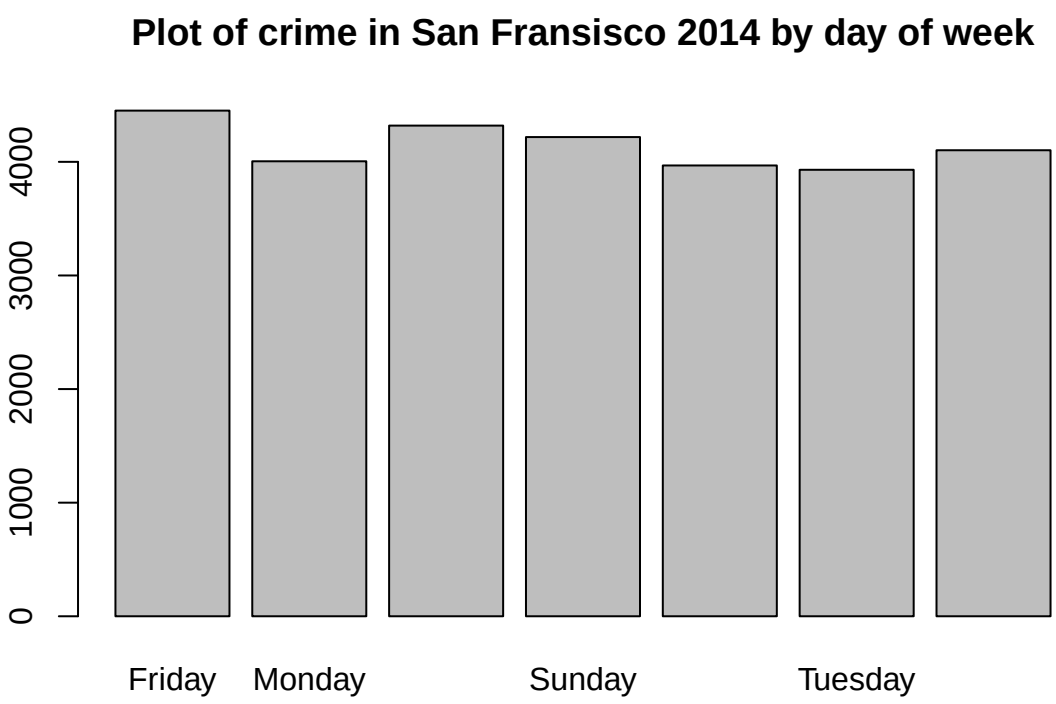
```
##
##  Offense.Code.Extension Offense.Type       Summary.Offense.Code
##  Min.   : 0.0           Length:32779       Length:32779
##  1st Qu.: 0.0           Class :character   Class :character
##  Median : 0.0           Mode  :character   Mode  :character
##  Mean   : 4.2
##  3rd Qu.: 1.0
##  Max.   :98.0
##
##  Summarized.Offense.Description Date.Reported
##  Length:32779                   Length:32779
##  Class :character               Class :character
##  Mode  :character               Mode  :character
##
##
##
##
##  Occurred.Date.or.Date.Range.Start Occurred.Date.Range.End
##  Length:32779                      Length:32779
##  Class :character                  Class :character
##  Mode  :character                  Mode  :character
##
##
##
##
##  Hundred.Block.Location District.Sector     Zone.Beat
##  Length:32779           Length:32779       Length:32779
##  Class :character       Class :character   Class :character
##  Mode  :character       Mode  :character   Mode  :character
##
##
##
##
##  Census.Tract.2000   Longitude         Latitude       Location
##  Min.   :  100.1   Min.   :-122.4   Min.   : 0.00   Length:32779
##  1st Qu.: 4600.1   1st Qu.:-122.3   1st Qu.:47.58   Class :character
##  Median : 7500.4   Median :-122.3   Median :47.61   Mode  :character
##  Mean   : 6737.7   Mean   :-114.7   Mean   :44.65
##  3rd Qu.: 9200.2   3rd Qu.:-122.3   3rd Qu.:47.66
##  Max.   :26500.1   Max.   :   0.0   Max.   :47.75
##  NA's   :134
##      Month            Year
##  Min.   :6.000   Min.   :2014
##  1st Qu.:6.000   1st Qu.:2014
##  Median :7.000   Median :2014
##  Mean   :6.981   Mean   :2014
##  3rd Qu.:8.000   3rd Qu.:2014
##  Max.   :8.000   Max.   :2014
##
```

Note the two cities report their crime statistics in differen ways

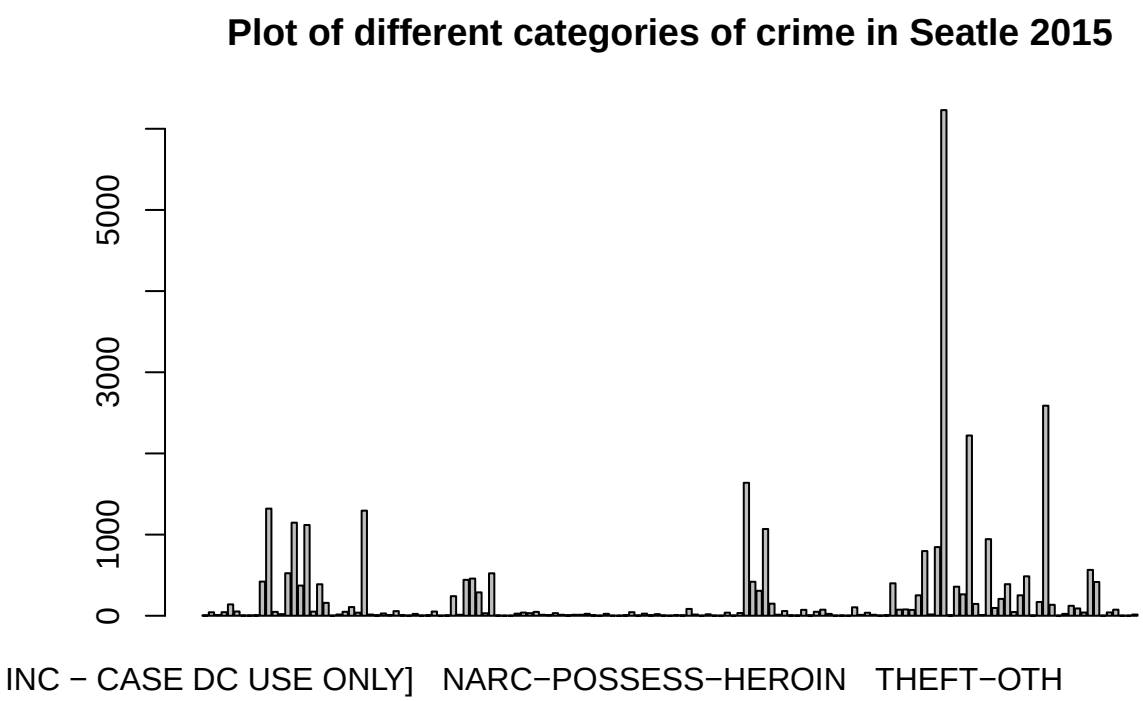
Therefore I decided to examine only one city But I will chose which city

Next I decided to do some very preliminary plots on the cities to decide what to examine with more details

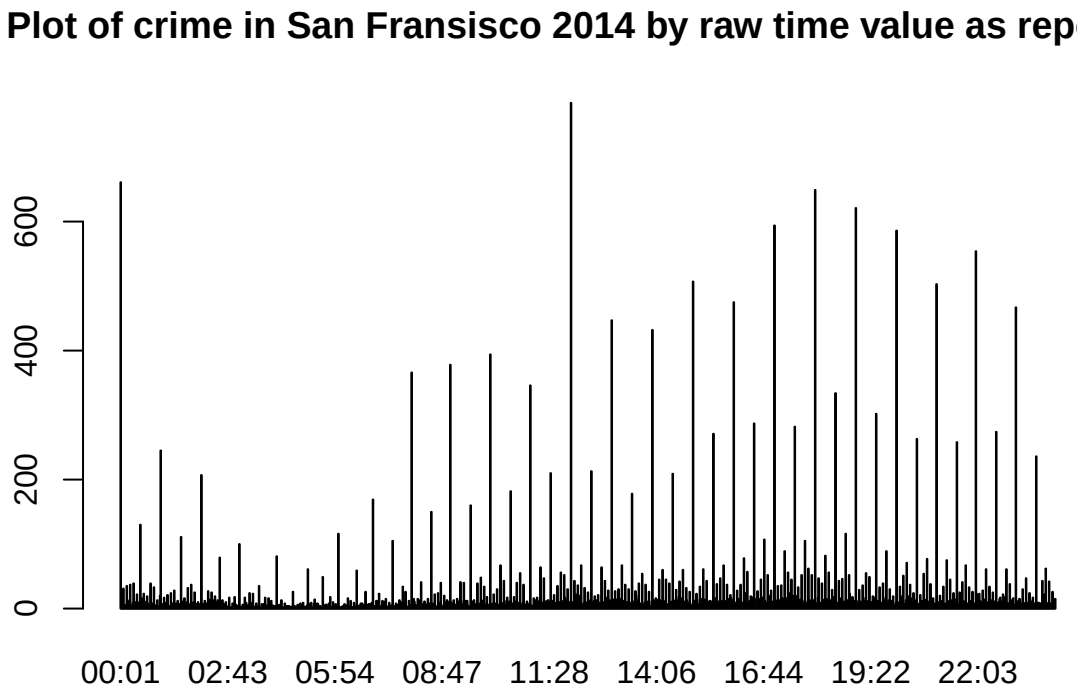



Plot of crime in San Fransisco 2014 by day of week
0
1000
2000
3000
4000
Friday
Monday
Sunday
Tuesday

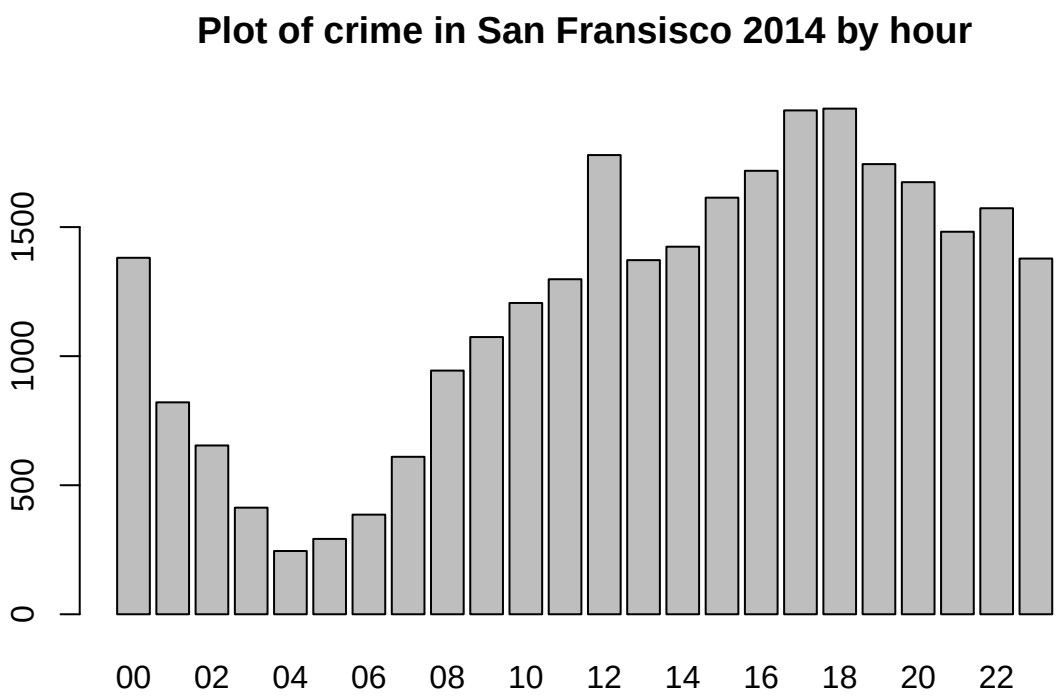
Finally a quick look at one column from Seatle before I make my mind up



Based on the above preliminary data analytics I decided to examine the San Fransisco data in more detail.



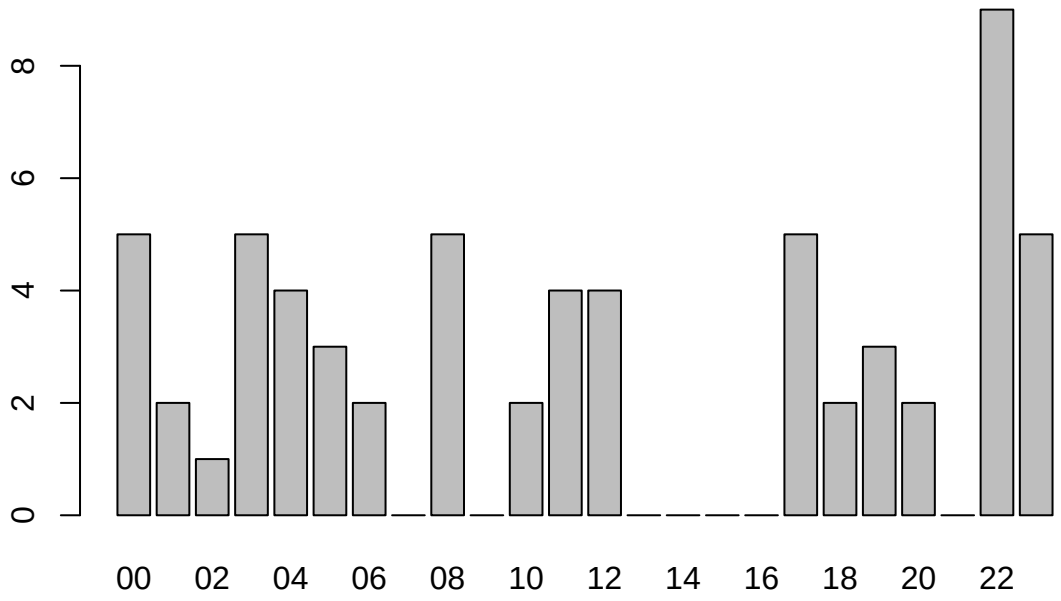
It appears that crime goes down after 10 pm to a minimum around 4am However the above graph may well be misleading as there are two many categories



So next I will look at just the hours

Definitely it appears to confirm previous idea that that crime goes down after 10 pm to a minimum around 4am Also shows a peack at midday

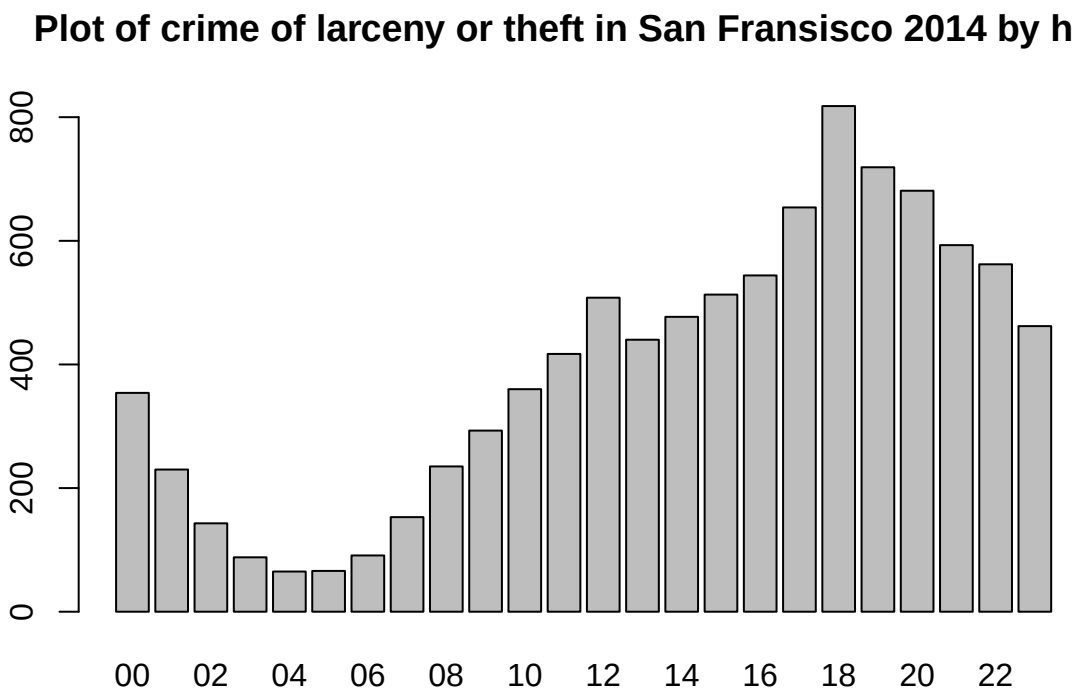
Now I'm wondering if different crimes vary by time



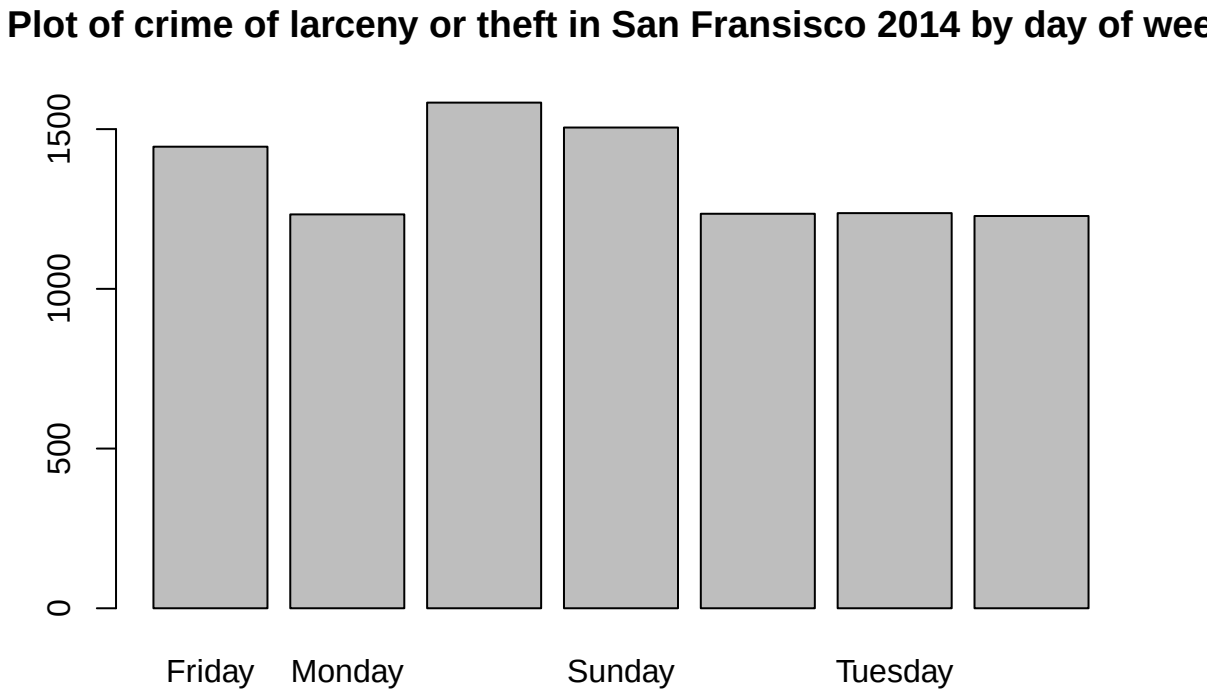
There is not enough arson data to het a good satistic for now. So next I looked at crimes with many instances

```
##                      ARSON                    ASSAULT
##                         63                       2882
##                    BRIBERY                   BURGLARY
##                          1                          6
##          DISORDERLY CONDUCT DRIVING UNDER THE INFLUENCE
##                         31                        100
##               DRUG/NARCOTIC                DRUNKENNESS
##                       1345                        147
##                EMBEZZLEMENT                  EXTORTION
##                         10                          7
##             FAMILY OFFENSES      FORGERY/COUNTERFEITING
##                         10                         18
##                      FRAUD                   GAMBLING
##                        242                          1
##                  KIDNAPPING              LARCENY/THEFT
##                        117                       9466
##                 LIQUOR LAWS                 LOITERING
##                         42                          3
##              MISSING PERSON               NON-CRIMINAL
##                       1266                       3023
##              OTHER OFFENSES     PORNOGRAPHY/OBSCENE MAT
##                       3567                          1
##                PROSTITUTION                   ROBBERY
##                        112                        308
##                     RUNAWAY            SECONDARY CODES
##                         61                        442
##             STOLEN PROPERTY                   SUICIDE
```

```
##                         8                        14
##          SUSPICIOUS OCC                  TRESPASS
##                      1300                       281
##               VANDALISM                VEHICLE THEFT
##                        17                      1966
##                WARRANTS                 WEAPON LAWS
##                      1782                       354
```

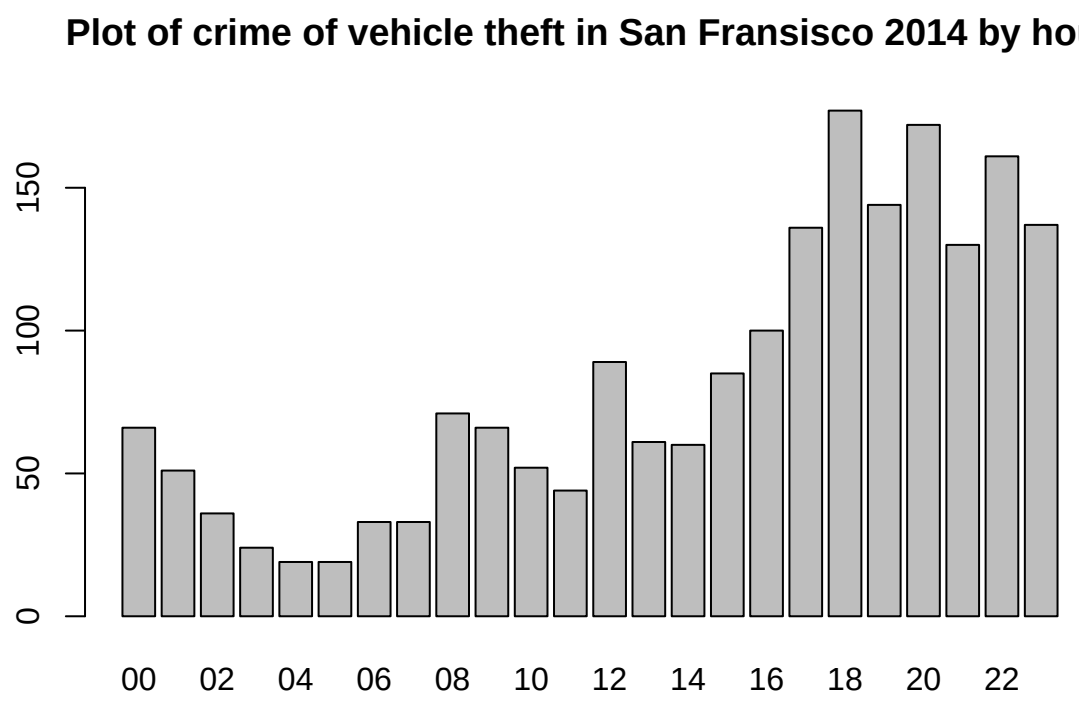


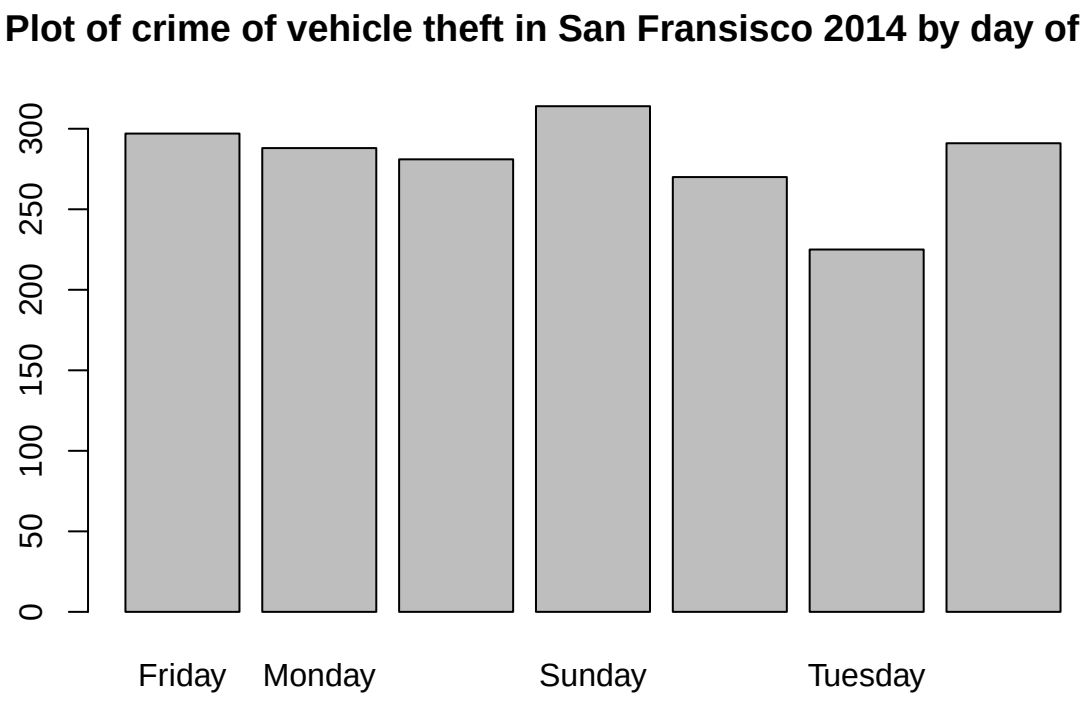
So I'll look at LARCENY/THEFT



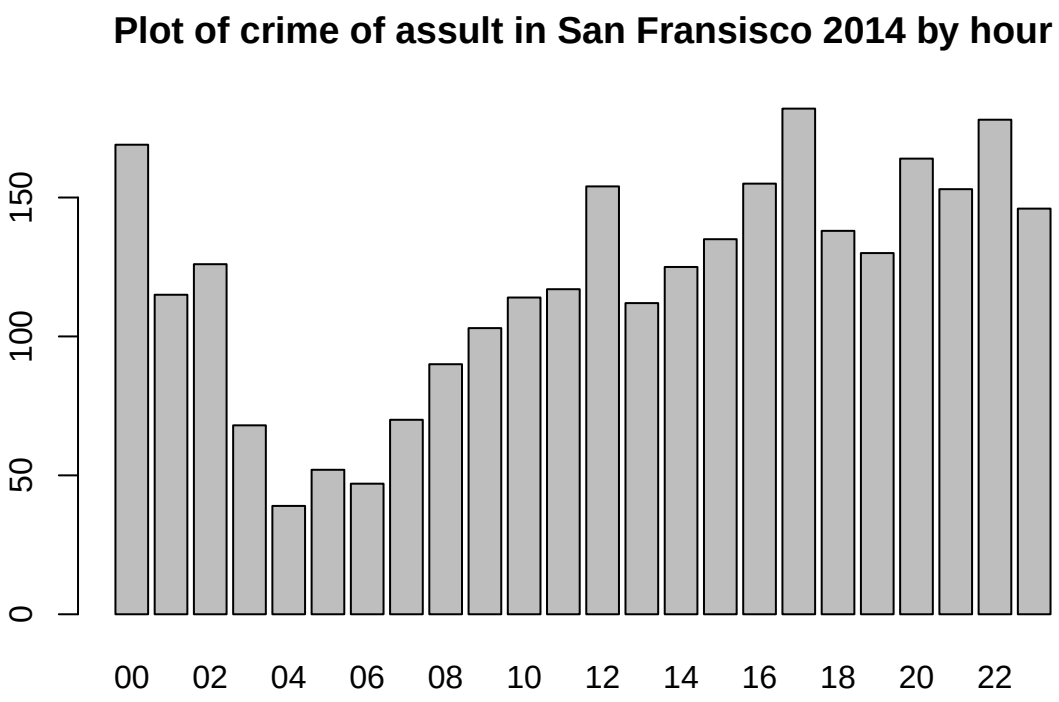
Note this case the specifc (LARCENY/THEFT) crime looks very similar to the totality of crime

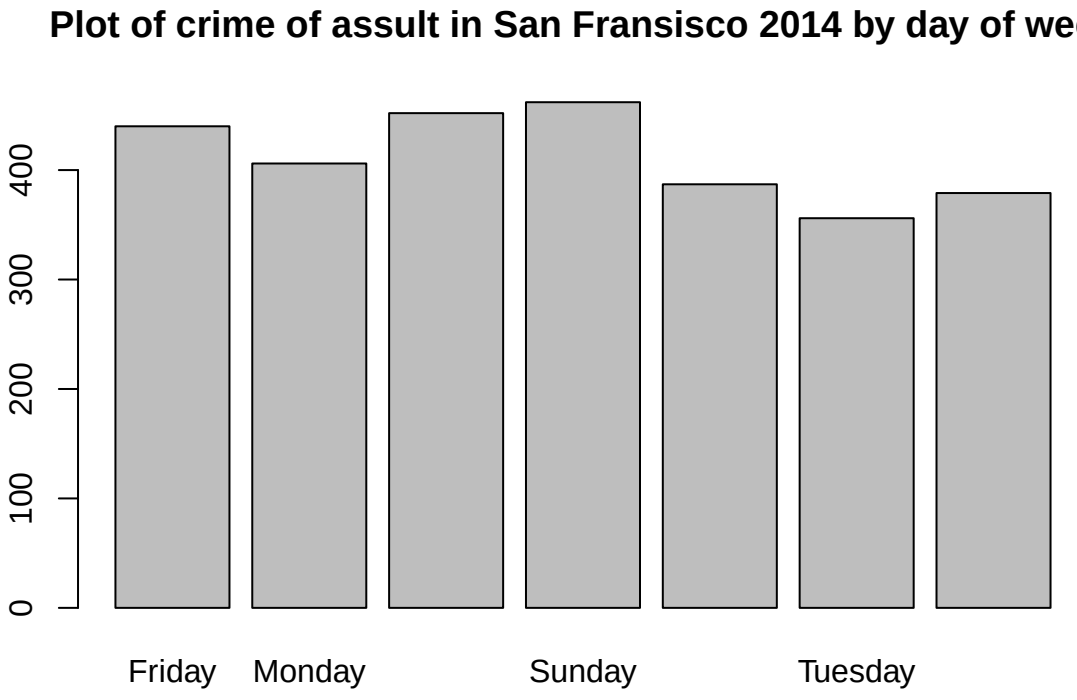
So now I'm curious about same plots for VEHICLE THEFT





Next do the same for ASSAULT





It is interesting that assults continue to be high until 2am My guess is that this relates to people drinking alcohol There are

not enough DRIVING UNDER THE INFLUENC to do any corrolation so I look up typical times bars close in San Fransisco At https://www.tripadvisor.com/ShowTopic-g60713-i30-k1347482-Time_that_bars_close-San_Francisco_California.html I discover that 2am is when they stop serving liquor Also: Most hotels' bars close at 01:00. Most bars with entertainment close at 01:30. Most gin mills close at 01:50. Last call for alcohol is usually twenty minutes before closing.

So my visualization made me predict a corralation between time of assults and bar closing times