Crímenes en Baltimore

Autor: Jesé Romero Arbelo

Enero 2018

Contents

Introducción

Según www.usatourist.com:

La ciudad de Baltimore, en el estado de Maryland, está ubicada sobre las costas de la amplia Bahía de Chesapeake, a casi 200 millas (más de 300 km) tierra adentro desde el Océano Atlántico. Está situada a aproximadamente 35 millas (60 km) al noreste de Washington D.C. y a 75 millas (125 km) al sudoeste de Filadelfia, Pensilvania. Es uno de los grandes puertos marinos de la costa este de los Estados Unidos, con enormes instalaciones para embarcaciones de carga y para el transporte de automóviles. Es una ciudad vieja, llena de personalidad y con muchas importantes atracciones históricas.

Descripción del dataset

Esta práctica se centrara en el estudio de los crímenes en la ciudad de Baltimore. Baltimore es una ciudad de Estados Unidos de aproximadamente 611.648 habitantes (2017). En 2016 este era el ratio de crimenes violentos por cada 100.000 habitantes:

- Homicidio: 57.8
- Secuestro: 46.1
- Robos: 694.2
- Asalto con agravante: 740.1

El dataset ha sido extraido de Kaggle y muestra los siguientes campos:

- **CrimeDate** ~ Fecha del suceso
- **CrimeTime** ~ Hora del suceso
- **CrimeCode** ~ Código del crímen
- **Location** ~ Dirección del suceso
- **Description** ~ Tipo de crímen
- **Inside/Outside** ~ Si ha sido en interiores o exteriores
- **Weapon** ~ Arma del crímen
- **Post** ~ Campo desconocido

- **District** ~ Zona de la ciudad
- **Neighborhood** ~ Barrio
- **Longitude** ~ Longitud del crímen
- **Latitude** ~ Latitud del crímen
- **Location 1** ~ Unión de la latitud y la longitud
- **Premise** ~ Móvil del crímen
- **Total Incidents** ~ Cantidad de incidentes

Este dataset es interesante ya que permite comprender el alcance que tiene el delito dentro de una ciudad como es Baltimore y tal vez viendo el alcance criminalistico nos podemos hacer una idea de otras ciudades con tasas de crimen parecidas. Con ello intentamos conocer qué zonas es en las que más crímenes se comete, a qué hora, con que arma...

Existen muchas preguntas que pueden ser resueltas con este dataset y a lo largo del informe podremos hacernos una idea de lo importante que es la seguridad en una ciudad.

Limpieza de los datos

El conjunto de datos contiene distintos elementos en blanco que serán tratados como NA para una posterior visualización y tratamiento, comenzaré asignando y gestionando cada columna importante del dataset.

```
crimes = read.csv("crimes.csv", stringsAsFactors = FALSE)
crimes = as.data.frame(crimes)
summary(crimes)
```

```
##   CrimeDate          CrimeTime          CrimeCode
##  Length:276529      Length:276529      Length:276529
##  Class :character   Class :character   Class :character
##  Mode  :character   Mode  :character   Mode  :character
##
##
##
##
##    Location         Description        Inside.Outside
##  Length:276529      Length:276529      Length:276529
##  Class :character   Class :character   Class :character
##  Mode  :character   Mode  :character   Mode  :character
##
##
##
##
##     Weapon               Post         District         Neighborhood
##  Length:276529      Min.   :  2.0   Length:276529      Length:276529
##  Class :character   1st Qu.:243.0   Class :character   Class :character
##  Mode  :character   Median :511.0   Mode  :character   Mode  :character
##                     Mean   :506.3
##                     3rd Qu.:731.0
##                     Max.   :945.0
##                     NA's   :224
##    Longitude         Latitude      Location.1          Premise
##  Min.   :-76.71   Min.   :39.20   Length:276529      Length:276529
##  1st Qu.:-76.65   1st Qu.:39.29   Class :character   Class :character
##  Median :-76.61   Median :39.30   Mode  :character   Mode  :character
```

```
##  Mean   :-76.62   Mean   :39.31
##  3rd Qu.:-76.59   3rd Qu.:39.33
##  Max.   :-76.53   Max.   :39.37
##  NA's   :2204     NA's   :2204
##  Total.Incidents
##  Min.   :1
##  1st Qu.:1
##  Median :1
##  Mean   :1
##  3rd Qu.:1
##  Max.   :1
##
```

```
str(crimes)
```

```
## 'data.frame':    276529 obs. of  15 variables:
##  $ CrimeDate      : chr  "09/02/2017" "09/02/2017" "09/02/2017" "09/02/2017" ...
##  $ CrimeTime      : chr  "23:30:00" "23:00:00" "22:53:00" "22:50:00" ...
##  $ CrimeCode      : chr  "3JK" "7A" "9S" "4C" ...
##  $ Location       : chr  "4200 AUDREY AVE" "800 NEWINGTON AVE" "600 RADNOR AV" "1800 RAMSAY ST" ...
##  $ Description    : chr  "ROBBERY - RESIDENCE" "AUTO THEFT" "SHOOTING" "AGG. ASSAULT" ...
##  $ Inside.Outside : chr  "I" "O" "Outside" "I" ...
##  $ Weapon         : chr  "KNIFE" "" "FIREARM" "OTHER" ...
##  $ Post           : int  913 133 524 934 113 922 232 123 641 332 ...
##  $ District       : chr  "SOUTHERN" "CENTRAL" "NORTHERN" "SOUTHERN" ...
##  $ Neighborhood   : chr  "Brooklyn" "Reservoir Hill" "Winston-Govans" "Carrollton Ridge" ...
##  $ Longitude      : num  -76.6 -76.6 -76.6 -76.6 -76.6 ...
##  $ Latitude       : num  39.2 39.3 39.3 39.3 39.3 ...
##  $ Location.1     : chr  "(39.2295100000, -76.6054100000)" "(39.3136000000, -76.6321700000)" "(39.34
##  $ Premise        : chr  "ROW/TOWNHO" "STREET" "Street" "ROW/TOWNHO" ...
##  $ Total.Incidents: int  1 1 1 1 1 1 1 1 1 1 ...
```

A priori podemos observar que:

- Existen 2204 casos en los que se desconoce la latitud y la longitud.
- La cantidad de incidentes por caso es siempre 1 (Por lo que en el futuro probablemente eliminaremos esa columna).
- Existen disparidades de datos que significan lo mismo en la columna "Inside.Outside" ("O" = "Outside").

Columna Inside.Outside

Esta columna contiene "I" o "O" dependiendo si ha sido "Inside" o "Outside" respectivamente. A su vez, también contiene valores blank que serán transformados a NAs.

```
crimes$Inside.Outside[crimes$Inside.Outside == ""] = NA

crimes$Inside.Outside[crimes$Inside.Outside == "O"] = "Outside"
crimes$Inside.Outside[crimes$Inside.Outside == "I"] = "Inside"

crimes$Inside.Outside = as.factor(crimes$Inside.Outside)

summary(crimes$Inside.Outside)
```

```
##  Inside Outside    NA's
##  132631  133619   10279
```

Una vez limpia la columna la factorizamos obteniendo dos factores:

- Inside
- Outside

Columna Location

Esta columna contiene valores en blanco que serán transformado a NAs.

```
crimes$Location[crimes$Location == ""] = NA
sum(is.na(crimes$Location))
```

```
## [1] 2207
```

Existen 2207 casos en los que no se conoce la localización.

Columna Weapon type

En esta columna también nos encontramos casos en blanco que pasaremos a WITHOUT WEAPON

```
crimes$Weapon[crimes$Weapon == ""] = "WITHOUT WEAPON"

crimes$Weapon = as.factor(crimes$Weapon)
summary(crimes$Weapon)
```

```
##        FIREARM          HANDS          KNIFE          OTHER WITHOUT WEAPON
##          22312          48995           9650          14620         180952
```

Aquí vemos que la mayoría de casos no se usó arma.

Columna District

Esta columna contiene casos en blanco que pasaremos a NAs.

```
crimes$District[crimes$District == ""] = NA
```

Eliminacion de columnas

Las columnas que no proporcionan información alguna deben ser eliminadas, estas son:

- CrimeCode
- Post
- Location.1
- Total incidents

```
crimes$CrimeCode = NULL
crimes$Post = NULL
crimes$Location.1 = NULL
crimes$Total.Incidents = NULL
summary(crimes)
```

```
##   CrimeDate          CrimeTime           Location
##  Length:276529      Length:276529      Length:276529
##  Class :character   Class :character   Class :character
##  Mode  :character   Mode  :character   Mode  :character
##
##
##
##
##  Description        Inside.Outside              Weapon
##  Length:276529      Inside :132631   FIREARM       : 22312
##  Class :character   Outside:133619   HANDS         : 48995
##  Mode  :character   NA's   : 10279   KNIFE         :  9650
##                                      OTHER         : 14620
##                                      WITHOUT WEAPON:180952
##
##
##    District         Neighborhood         Longitude         Latitude
##  Length:276529      Length:276529      Min.   :-76.71   Min.   :39.20
##  Class :character   Class :character   1st Qu.:-76.65   1st Qu.:39.29
##  Mode  :character   Mode  :character   Median :-76.61   Median :39.30
##                                        Mean   :-76.62   Mean   :39.31
##                                        3rd Qu.:-76.59   3rd Qu.:39.33
##                                        Max.   :-76.53   Max.   :39.37
##                                        NA's   :2204     NA's   :2204
##    Premise
##  Length:276529
##  Class :character
##  Mode  :character
##
##
##
##
```

Columna Neighborhood

Sustituiré los blancos por NA.

```
crimes$Neighborhood[crimes$Neighborhood == ""] = NA
```

Transformación de datos

En este paso transformaré aquellas columnas y crearé algunas otras a partir de estas.

Adición de fechas

A partir de la fecha y la hora del crimen voy a extraer lo siguiente:

- Hora
- Día
- Mes
- Año
- Día de la semana

```
# Hacemos tipo fecha la columna CrimeData
crimes$CrimeDate = as.Date(crimes$CrimeDate, format = '%m/%d/%Y')

# Obtenemos a partir del tiempo y la fecha: Hora, Día, Mes y Año del crimen
crimes$HourOfCrime = as.numeric(substr(crimes$CrimeTime, 0, 2))
crimes$DayOfCrime = as.numeric (format(crimes$CrimeDate, '%d'))
crimes$MonthOfCrime = as.numeric(format(crimes$CrimeDate, '%m'))
crimes$YearOfCrime = as.numeric(format(crimes$CrimeDate, '%Y'))

# Obtenemos el día de la semana en la que se realizó el crimen
crimes$Weekday = wday(crimes$CrimeDate, label=TRUE)

summary(crimes)
```

```
##    CrimeDate           CrimeTime           Location
##  Min.   :2012-01-01   Length:276529      Length:276529
##  1st Qu.:2013-06-04   Class :character   Class :character
##  Median :2014-11-05   Mode  :character   Mode  :character
##  Mean   :2014-11-07
##  3rd Qu.:2016-04-27
##  Max.   :2017-09-02
##
##  Description        Inside.Outside              Weapon
##  Length:276529      Inside :132631   FIREARM       : 22312
##  Class :character   Outside:133619   HANDS         : 48995
##  Mode  :character   NA's   : 10279   KNIFE         :  9650
##                                      OTHER         : 14620
##                                      WITHOUT WEAPON:180952
##
##
##    District         Neighborhood         Longitude         Latitude
##  Length:276529      Length:276529      Min.   :-76.71   Min.   :39.20
##  Class :character   Class :character   1st Qu.:-76.65   1st Qu.:39.29
##  Mode  :character   Mode  :character   Median :-76.61   Median :39.30
##                                        Mean   :-76.62   Mean   :39.31
##                                        3rd Qu.:-76.59   3rd Qu.:39.33
##                                        Max.   :-76.53   Max.   :39.37
##                                        NA's   :2204     NA's   :2204
##    Premise           HourOfCrime      DayOfCrime     MonthOfCrime
##  Length:276529      Min.   : 0.00   Min.   : 1.00   Min.   : 1.00
##  Class :character   1st Qu.: 9.00   1st Qu.: 8.00   1st Qu.: 4.00
##  Mode  :character   Median :15.00   Median :16.00   Median : 6.00
##                     Mean   :13.29   Mean   :15.81   Mean   : 6.45
##                     3rd Qu.:19.00   3rd Qu.:23.00   3rd Qu.: 9.00
##                     Max.   :24.00   Max.   :31.00   Max.   :12.00
##
##   YearOfCrime   Weekday
##  Min.   :2012   dom:37202
```

```
##  1st Qu.:2013   lun:40239
##  Median :2014   mar:39741
##  Mean   :2014   mié:39752
##  3rd Qu.:2016   jue:39387
##  Max.   :2017   vie:41596
##                 sáb:38612
```

Factorización

Este paso factorizaré aquellas columnas que considere necesarias.

```
crimes$Description = as.factor(crimes$Description)
crimes$District = as.factor(crimes$District)
crimes$Neighborhood = as.factor(crimes$Neighborhood)
crimes$Premise = as.factor(crimes$Premise)
summary(crimes$Description)
```

```
##         AGG. ASSAULT                ARSON    ASSAULT BY THREAT
##                27513                 1464                 3503
##           AUTO THEFT              BURGLARY       COMMON ASSAULT
##                26838                42538                45518
##             HOMICIDE              LARCENY    LARCENY FROM AUTO
##                 1559                60528                36295
##                 RAPE ROBBERY - CARJACKING ROBBERY - COMMERCIAL
##                 1637                 1528                 4141
##  ROBBERY - RESIDENCE     ROBBERY - STREET             SHOOTING
##                 2866                17691                 2910
```

```
summary(crimes$District)
```

```
##      CENTRAL      EASTERN NORTHEASTERN     NORTHERN NORTHWESTERN
##        31631        23635        43006        31665        28061
## SOUTHEASTERN     SOUTHERN SOUTHWESTERN      WESTERN         NA's
##        38291        31850        26092        22218           80
```

```
summary(crimes$Neighborhood)
```

```
##                 Downtown                Frankford
##                     9048                     6642
##            Belair-Edison                 Brooklyn
##                     5977                     4516
##              Cherry Hill      Sandtown-Winchester
##                     4086                     4026
##                   Canton             Inner Harbor
##                     3787                     3414
##                   Upton Patterson Park Neighborho
##                     3385                     3368
##             Fells Point Coldstream Homestead Mont
##                     3316                     3307
##           Hamilton Hills                 Mondawmin
##                     3303                     3251
```

```
## Washington Village/Pigtow            McElderry Park
##                      3223                      3004
##      Central Park Heights              Mount Vernon
##                      2960                      2751
##             Broadway East          Carrollton Ridge
##                      2741                      2653
##                   Hampden     East Baltimore Midway
##                      2524                      2460
##            Reservoir Hill                    Oliver
##                      2426                      2376
##     Ellwood Park/Monument        Mid-Town Belvedere
##                      2364                      2362
##              Harlem Park           Charles Village
##                      2172                      2158
##       Baltimore Highlands            Better Waverly
##                      2104                      2028
##             Morrell Park Penrose/Fayette Street Ou
##                      2021                      1955
##               Curtis Bay             Charles North
##                      1932                      1930
##              Waltherson                  Lakeland
##                      1912                      1892
##      Reisterstown Station            Glenham-Belhar
##                      1880                      1873
##             Highlandtown    Ednor Gardens-Lakeside
##                      1835                      1782
##              Howard Park           Washington Hill
##                      1773                      1743
##   Pulaski Industrial Area                 Irvington
##                      1736                      1713
##               Penn North                      Glen
##                      1690                      1676
##              Bolton Hill           Dunbar-Broadway
##                      1674                      1651
##                 Millhill                   Barclay
##                      1645                      1605
##                Riverside                 Poppleton
##                      1605                      1558
##             Madison Park              Federal Hill
##                      1546                      1519
##               Loch Raven                 Greektown
##                      1515                      1470
##          Franklin Square Carroll - Camden Industri
##                      1460                      1454
##                Jonestown                  Woodmere
##                      1429                      1423
##                    Berea        North Harford Road
##                      1406                      1393
##           Johnston Square            Hollins Market
##                      1363                      1357
##                  Oldtown New Southwest/Mount Clare
##                      1348                      1326
##         Upper Fells Point             New Northwood
##                      1313                      1252
```

```
##               Park Circle                Gay Street
##                      1213                      1211
##          Madison-Eastend             Druid Heights
##                      1185                      1183
##                      CARE                 Allendale
##                      1181                      1173
##                   Waverly                  Rosemont
##                      1161                      1156
##                 Arlington    Canton Industrial Area
##                      1151                      1137
##                Lauraville           Greenmount West
##                      1131                      1085
##              Shipley Hill           Hanlon-Longwood
##                      1079                      1077
##         Midtown-Edmondson                  Walbrook
##                      1076                      1073
## Coppin Heights/Ash-Co-Eas             Downtown West
##                      1072                      1071
##                 Remington               Greenspring
##                      1071                      1061
##             Penn-Fallsway      Bridgeview/Greenlawn
##                      1054                      1034
##               Middle East                 Glen Oaks
##                      1027                      1016
##         Edmondson Village                   Cylburn
##                       996                       995
##               Violetville           South Baltimore
##                       969                       951
##                  Pen Lucy                   Cedmont
##                       937                       935
##                   (Other)                    NA's
##                     77941                      2740
```

```
summary(crimes$Premise)
```

```
##     STREET ROW/TOWNHO PARKING LO  APT/CONDO OTHER - IN
##     100919      60502      12043      12002      11459      10757
##     SCHOOL CONVENIENC RETAIL/SMA OTHER - OU       YARD OTHER/RESI
##       7608       4314       3725       3423       3376       3010
##     Street GROCERY/CO      ALLEY HOSP/NURS.  GARAGE ON GAS STATIO
##       2883       2843       2701       1961       1838       1673
## SINGLE HOU DRUG STORE   DRIVEWAY RESTAURANT        BAR  SPECIALTY
##       1624       1564       1551       1516       1372       1342
##    UNKNOWN OFFICE BUI HOTEL/MOTE PORCH/DECK       PARK DEPARTMENT
##       1287       1069       1032        979        884        877
## SHED/GARAG   BUS/AUTO VACANT BUI  RELIGIOUS LIQUOR STO SHOPPING M
##        840        798        768        628        574        571
##  CARRY OUT PUBLIC HOU   Dwelling CLOTHING/S CONSTRUCTI YARD/BUSIN
##        489        459        406        361        358        352
## BARBER/BEA  BUS. PARK  FAST FOOD BUS/RAILRO    STADIUM CAR LOT-NE
##        345        344        321        290        290        288
## CAR  REPAI PUBLIC BUI LIGHT RAIL MARKET STA RECREATION BANK/FINAN
##        281        279        277        245        242        218
## AUTO PARTS TAVERN/NIG WHOLESALE/  WAREHOUSE LAUNDRY/CL     SUBWAY
```

```
##        215        212        212        211        194        184
## POLICE DEP      Alley PLAYGROUND RACE TRACK HARDWARE/B Public Are
##        181        176        174        155        153        145
##  APARTMENT COURT HOUS Parking Lo Common Bus DOCTORS OF INNER HARB
##        137        135        133        112        105        102
##    LIBRARY CONVENTION MINI STORA APT. LOCKE CHAIN FOOD CLUB HOUSE
##        102         97         88         78         75         68
## BUS.  STOR ATM MACHIN Public Hou PIZZA/OTHE FIRE DEPAR HOUSE UNDE
##         52         49         48         44         43         43
##     BAKERY VACANT LOT  PAWN SHOP SALESMAN/C   Hospital   CEMETERY
##         41         41         39         39         36         31
## SCHOOL PLA JEWELRY ST NIGHT DEPO    THEATRE POOL/BOWLI        CAB
##         30         28         28         28         27         26
## BRIDGE-PIE    MTA LOT  BOAT/SHIP    (Other)
##         25         25         22        182
```

Extracción de datos limpios a un CSV

Este paso es exclusivo para extraer los datos a un CSV

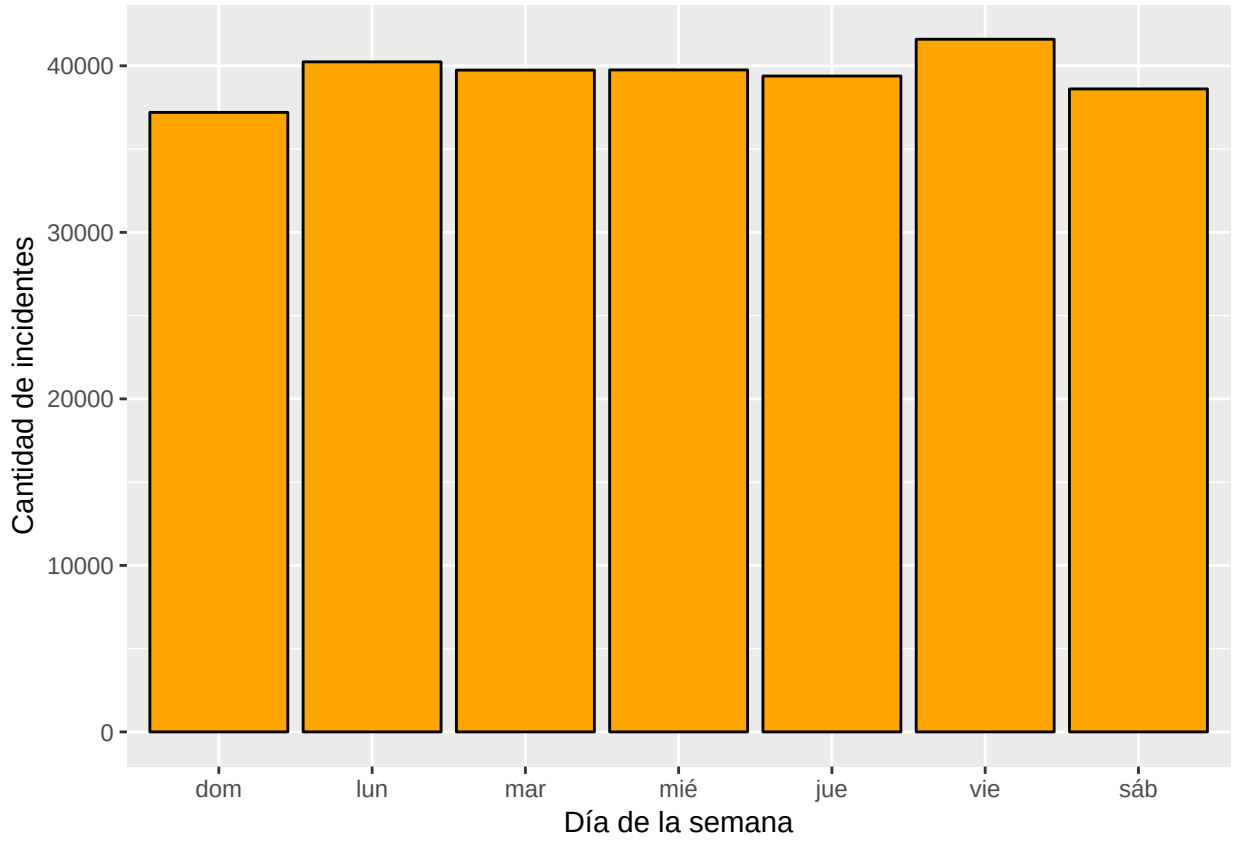
```
write.csv(crimes, "clean_crimes.csv")
```

Análisis de datos

Vamos a realizar un pequeño análisis visual de los datos para tratar de comprender con mayor profundidad la información que disponemos.

Comenzaremos visualizando los crimenes que se cometen a lo largo de una semana.

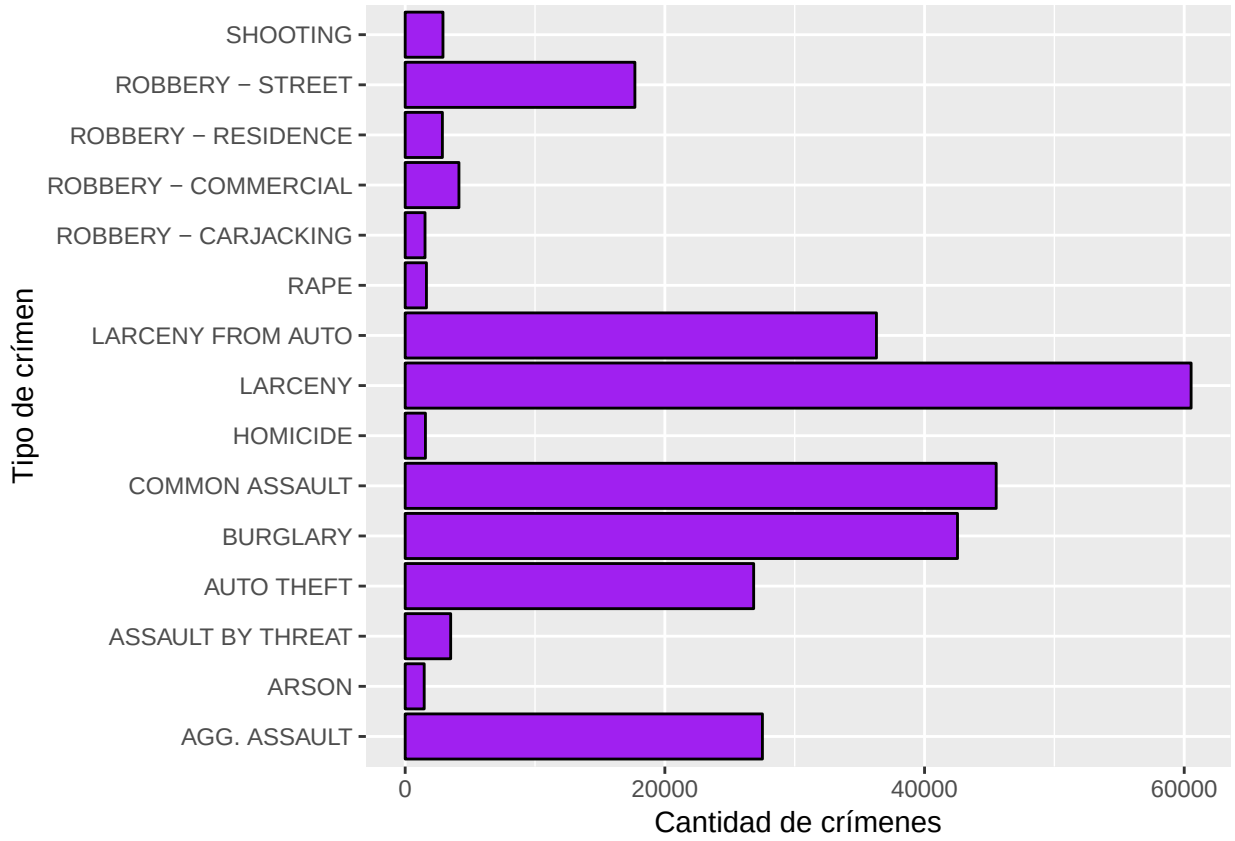
```
ggplot(data=crimes , aes(Weekday)) +
  geom_bar(fill='Orange', color='Black') +
  labs(x = "Día de la semana", y = "Cantidad de incidentes")
```



No existen grandes diferencias entre los días de la semana por lo que se concluye que no es un factor diferenciativo entre distintos crímenes.

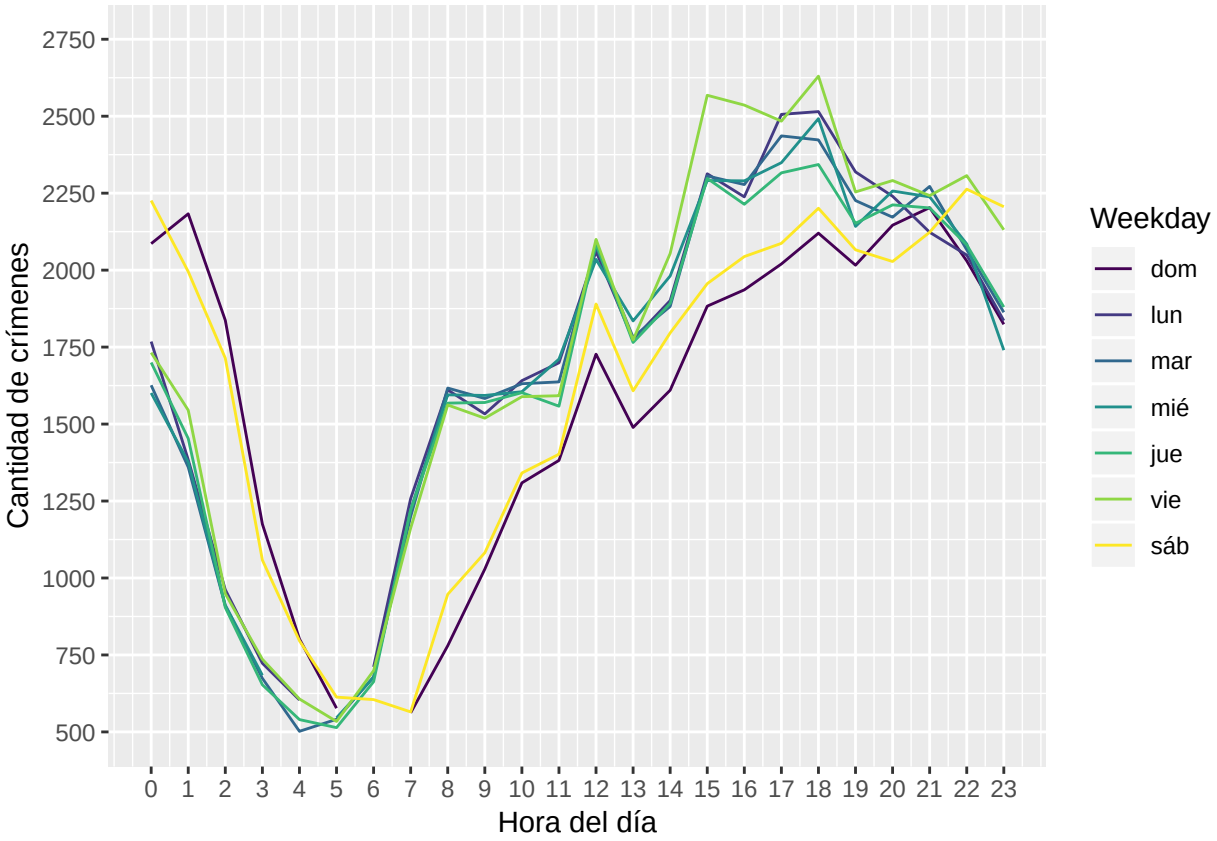
Lo siguiente será observar el tipo de crimen cometido a ver qué casos son los mayoritarios.

```
ggplot(data=crimes , aes(Description)) +
  geom_bar(fill='Purple', color='Black') +
  coord_flip() +
  labs(x = "Tipo de crímen", y = "Cantidad de crímenes")
```



Ahora vamos a ver si existe alguna relación entre la hora del crimen y el día de la semana.

```
ggplot(crimes) + aes(x = HourOfCrime, colour = Weekday) + geom_line(stat = "count") +
  scale_x_continuous(limit = c(0, 23), breaks = seq.int(0, 23, 1)) +
  scale_y_continuous(limit = c(500, 2750), breaks = seq.int(500, 2750, 250)) +
  labs(x = "Hora del día", y = "Cantidad de crímenes")
```

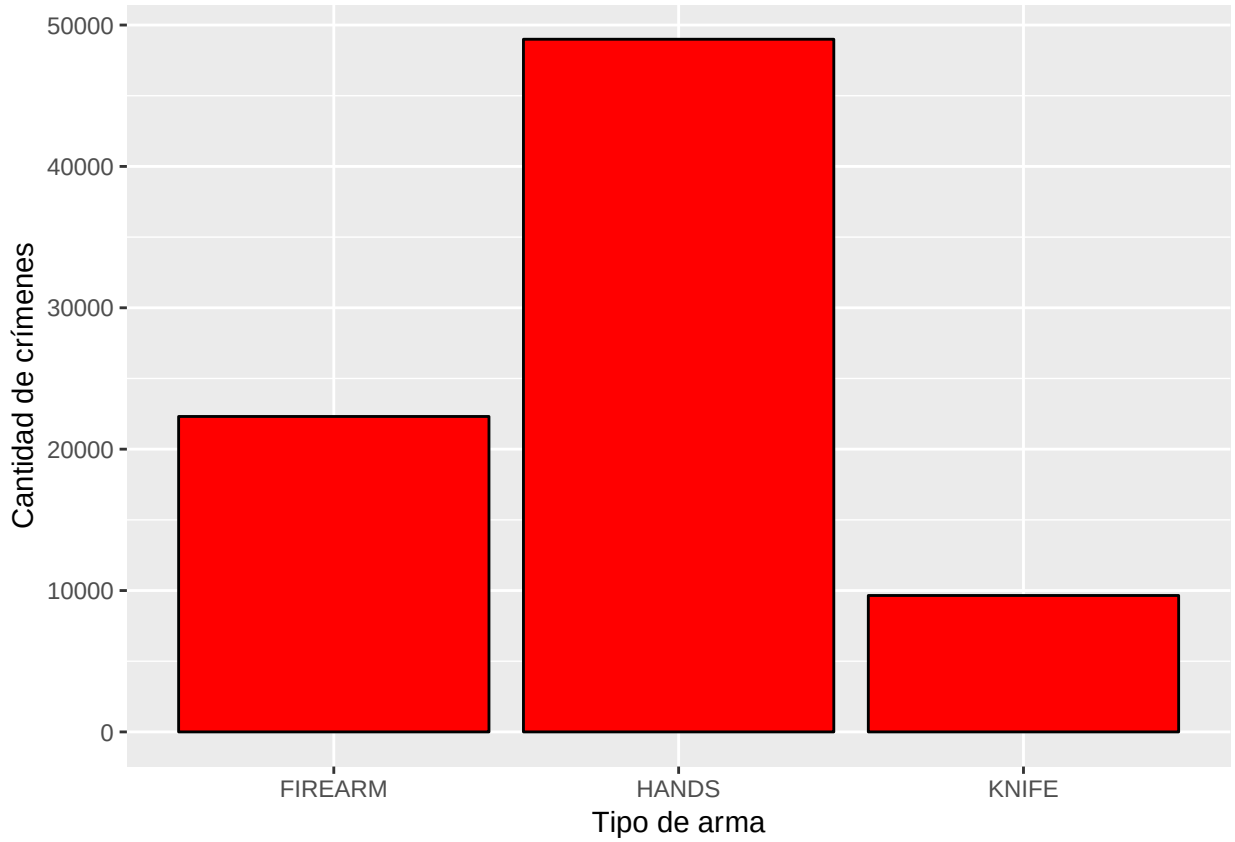


Aquí vemos que la mayoría de los crímenes se realizan durante el día y no de madrugada.

De todos los crímenes que se cometen con armas, veamos si visualizandolo podemos resolver alguna cuestión.

```
with_weapon = crimes %>% filter(Weapon != "WITHOUT WEAPON" & Weapon != "OTHER")

ggplot(data = with_weapon , aes(Weapon)) +
  geom_bar(fill='Red', color='Black') +
  labs(x = "Tipo de arma", y = "Cantidad de crímenes")
```

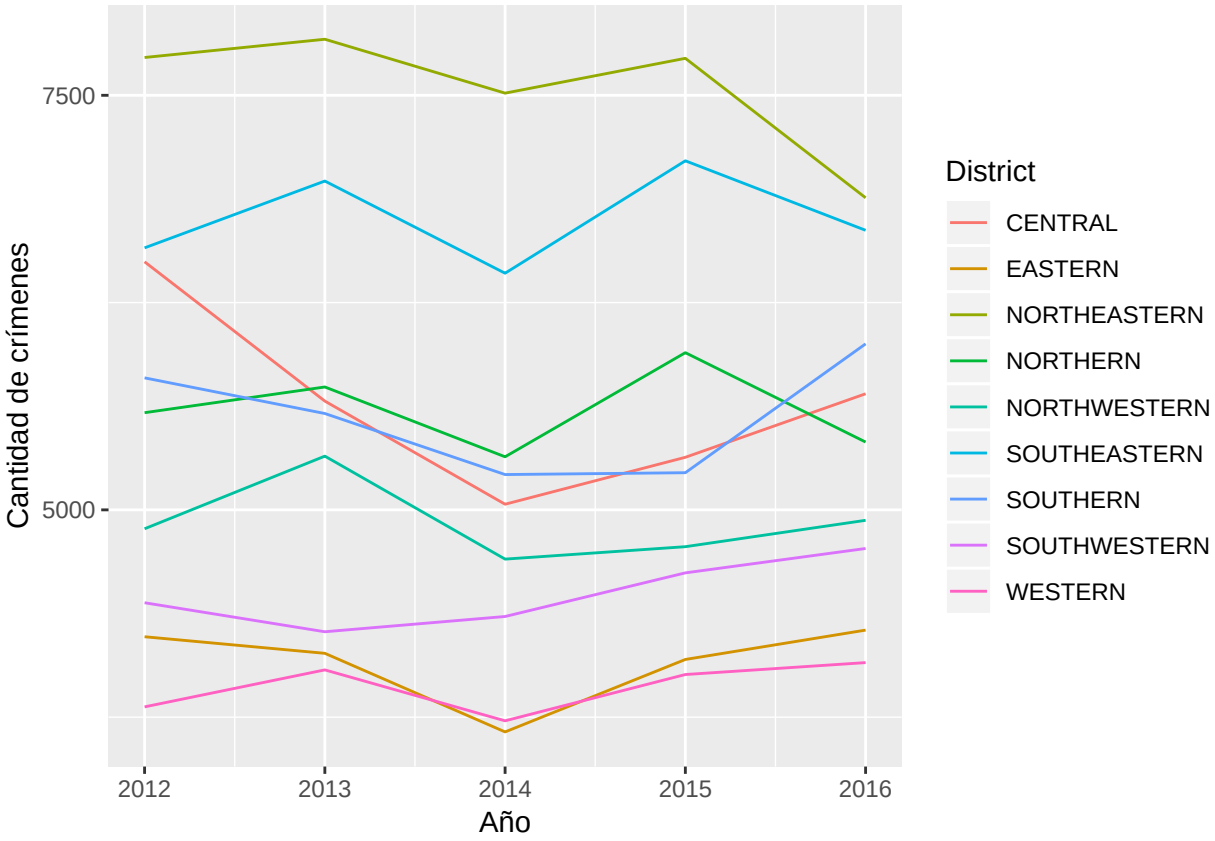


Vemos que el "arma" más usada son las propias manos seguido de armas de fuego y cuchillos.

¿Qué zona de la ciudad de Baltimore es la más conflictiva? ¿Cuál es la más segura?

```
by_discrict = crimes %>% filter(!is.na(District) & YearOfCrime != 2017)

ggplot(by_discrict, aes(x = YearOfCrime, color = District)) + geom_line(stat="count") +
  scale_x_continuous(breaks = seq.int(2012, 2017, 1)) +
  scale_y_continuous(breaks = seq.int(2500, 50000, 2500)) +
  labs(x = "Año", y = "Cantidad de crímenes")
```



De aquí podemos extraer que el distrito NORESTE es el más peligroso, con más de 7500 crímenes en 2012, 2013, 2014 y 2015 siendo el más conflictivo en esos años.

De media, el distrito más seguro sería el distrito OESTE ya que contiene la tasa de crímenes baja a lo largo de los años.

Conclusiones

Los crímenes en Baltimore no conducen a un patrón concluyente en cuanto hora y lugar. Existen muchos casos en el que no se usa ningún tipo de arma. Los distritos más seguros de mayor a menor en 2016 (el año más cercano al actual) son:

- OESTE
- ESTE
- SUROESTE
- NOROESTE
- NORTE
- CENTRAL
- SUR
- NORESTE

Por lo que el distrito más seguro para vivir es el OESTE.

Las horas más seguras son las primeras horas de la mañana y de madrugada.

Las calles de Baltimore son generalmente seguras, pero habría que evitar ciertas zonas (como en todas las ciudades).

En cuanto las armas no ha sido posible obtener información más concreta, tan solo que el “arma” más usada son las manos del propio agresor.