Analiza kriminala i socio-ekonomskih faktora

Bruno Ćorić, Filip Škrlec, Jelena Matečić, Iva Zekić

11/6/2020

Učitavanje podataka

Imamo dva skupa podataka kriminala i socio-ekonomskih faktora za grad Chicago.

```
crimeDataset <- read.csv("crime_datasets/Crimes_-_One_year_prior_to_present.csv"
                         ,stringsAsFactors = F,na.strings = "")

povertyDataset <- read.csv("crime_datasets/Chicago_poverty_and_crime.csv"
                           ,stringsAsFactors = F, na.strings="")

head(crimeDataset)
```

```
##      CASE.    DATE..OF.OCCURRENCE               BLOCK IUCR PRIMARY.DESCRIPTION
## 1 JD388829 10/04/2020 08:31:00 PM 086XX S CARPENTER ST 0560             ASSAULT
## 2 JD346990 08/26/2020 01:33:00 PM  011XX N DEARBORN ST 0890               THEFT
## 3 JD403530 10/18/2020 03:50:00 PM     049XX W ADAMS ST 0460             BATTERY
## 4 JD141525 02/05/2020 02:54:00 PM   030XX N HALSTED ST 0860               THEFT
## 5 JD366829 08/26/2020 02:19:00 AM 021XX W CULLERTON ST 0890               THEFT
## 6 JD205528 04/09/2020 02:00:00 PM   029XX S ARCHER AVE 1320     CRIMINAL DAMAGE
##   SECONDARY.DESCRIPTION LOCATION.DESCRIPTION ARREST DOMESTIC BEAT WARD FBI.CD
## 1                SIMPLE            APARTMENT      N        N  613   21    08A
## 2          FROM BUILDING            APARTMENT      N        N 1824    2     06
## 3                SIMPLE               STREET      N        N 1533   28    08B
## 4           RETAIL THEFT           DRUG STORE      N        N 1933   44     06
## 5          FROM BUILDING            APARTMENT      N        N 1234   25     06
## 6             TO VEHICLE               STREET      N        N  913   11     14
##   X.COORDINATE Y.COORDINATE LATITUDE LONGITUDE                    LOCATION
## 1      1170827      1847522 41.73707 -87.64972 (41.737074199, -87.64972468)
## 2           NA           NA       NA        NA                        <NA>
## 3           NA           NA       NA        NA                        <NA>
## 4           NA           NA       NA        NA                        <NA>
## 5           NA           NA       NA        NA                        <NA>
## 6      1168260      1885596 41.84161 -87.65803 (41.841609341, -87.65803375)
```

```
head(povertyDataset)
```

```
##   Community.Area Community.Area.Name Assault..Homicide. Firearm.related
## 1              1         Rogers Park                7.7             5.2
## 2              2          West Ridge                5.8             3.7
## 3              3              Uptown                5.4             4.6
## 4              4      Lincoln Square                5.0             6.1
## 5              5        North Center                1.0             1.0
## 6              6           Lake View                1.4             1.8
##   Below.Poverty.Level Crowded.Housing Dependency No.High.School.Diploma
```

```
## 1                22.7             7.9       28.8                    18.1
## 2                15.1             7.0       38.3                    19.6
## 3                22.7             4.6       22.2                    13.6
## 4                 9.5             3.1       25.6                    12.5
## 5                 7.1             0.2       25.5                     5.4
## 6                10.5             1.2       16.5                     2.9
##   Per.Capita.Income Unemployment
## 1             23714          7.5
## 2             21375          7.9
## 3             32355          7.7
## 4             35503          6.8
## 5             51615          4.5
## 6             58227          4.7
```

```
str(crimeDataset)
```

```
## 'data.frame':    216032 obs. of  17 variables:
##  $ CASE.                : chr  "JD388829" "JD346990" "JD403530" "JD141525" ...
##  $ DATE..OF.OCCURRENCE  : chr  "10/04/2020 08:31:00 PM" "08/26/2020 01:33:00 PM" "10/18/2020 03:50:0
##  $ BLOCK                : chr  "086XX S CARPENTER ST" "011XX N DEARBORN ST" "049XX W ADAMS ST" "030X
##  $ IUCR                 : chr  "0560" "0890" "0460" "0860" ...
##  $ PRIMARY.DESCRIPTION  : chr  "ASSAULT" "THEFT" "BATTERY" "THEFT" ...
##  $ SECONDARY.DESCRIPTION: chr  "SIMPLE" "FROM BUILDING" "SIMPLE" "RETAIL THEFT" ...
##  $ LOCATION.DESCRIPTION : chr  "APARTMENT" "APARTMENT" "STREET" "DRUG STORE" ...
##  $ ARREST               : chr  "N" "N" "N" "N" ...
##  $ DOMESTIC             : chr  "N" "N" "N" "N" ...
##  $ BEAT                 : int  613 1824 1533 1933 1234 913 312 1924 2422 333 ...
##  $ WARD                 : int  21 2 28 44 25 11 20 44 49 7 ...
##  $ FBI.CD               : chr  "08A" "06" "08B" "06" ...
##  $ X.COORDINATE         : int  1170827 NA NA NA NA 1168260 1180028 NA NA NA ...
##  $ Y.COORDINATE         : int  1847522 NA NA NA NA 1885596 1862391 NA NA NA ...
##  $ LATITUDE             : num  41.7 NA NA NA NA ...
##  $ LONGITUDE            : num  -87.6 NA NA NA NA ...
##  $ LOCATION             : chr  "(41.737074199, -87.64972468)" NA NA NA ...
```

```
str(povertyDataset)
```

```
## 'data.frame':    77 obs. of  10 variables:
##  $ Community.Area        : int  1 2 3 4 5 6 7 8 9 10 ...
##  $ Community.Area.Name   : chr  "Rogers Park" "West Ridge" "Uptown" "Lincoln Square" ...
##  $ Assault..Homicide.    : num  7.7 5.8 5.4 5 1 1.4 0.7 3.7 0 4.7 ...
##  $ Firearm.related       : num  5.2 3.7 4.6 6.1 1 1.8 2.3 3.2 7.1 8.7 ...
##  $ Below.Poverty.Level   : num  22.7 15.1 22.7 9.5 7.1 10.5 11.8 13.4 5.1 5.9 ...
##  $ Crowded.Housing       : num  7.9 7 4.6 3.1 0.2 1.2 0.6 2 0.6 2.3 ...
##  $ Dependency            : num  28.8 38.3 22.2 25.6 25.5 16.5 20.4 23.3 36.6 40.6 ...
##  $ No.High.School.Diploma: num  18.1 19.6 13.6 12.5 5.4 2.9 4.3 3.4 8.5 13.5 ...
##  $ Per.Capita.Income     : int  23714 21375 32355 35503 51615 58227 71403 87163 38337 31659 ...
##  $ Unemployment          : num  7.5 7.9 7.7 6.8 4.5 4.7 4.5 5.2 7.4 7.3 ...
```

Faktorizirat ćemo podatke koje bi bilo logično faktorizirati kao što su podaci u stupcu Arrest,Domestic.

```
crimeDataset$ARREST <- as.factor(crimeDataset$ARREST)
crimeDataset$DOMESTIC <- as.factor(crimeDataset$DOMESTIC)
```

Provjeravamo fale li nam neki podaci u najbitnim kategorijama u oba dataseta.

```r
s <-c(1,2,3,4,5,6,8,9)
sum(is.na(crimeDataset[s]))
```

```
## [1] 0
```

```r
sum(is.na(povertyDataset))
```

```
## [1] 0
```

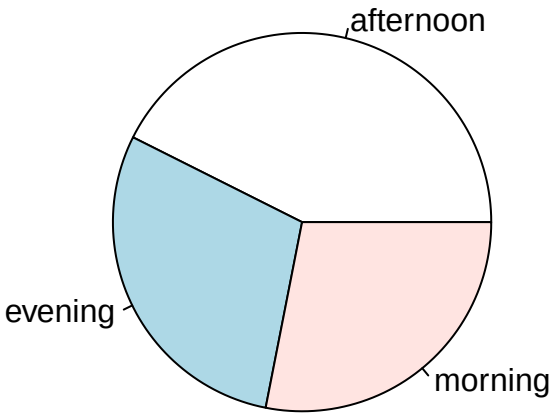
Razlika učestalosti zločina ovisno o tome koje je doba dana

Podijelit ćemo dan na 3 dijela. Od 5 do 13 će biti prvi dio dana. Od 13 do 21 drugi dio dana, a od 21 do 5 treći dio dana.

```r
timeOfTheDay <- mdy_hms(crimeDataset$DATE..OF.OCCURRENCE) %>% hour
timeOfTheDay <- sapply(timeOfTheDay, function(x) {
  if(x >= 5 & x < 13) {
    "morning"
  } else if(x >= 13 & x < 21) {
    "afternoon"
  } else {
    "evening"
  }
},simplify="vector")
timeOfTheDay <- as.factor(timeOfTheDay)
crimeDataset$TIME.OF.DAY <- timeOfTheDay
head(crimeDataset[c("DATE..OF.OCCURRENCE","TIME.OF.DAY")])
```

```
##      DATE..OF.OCCURRENCE TIME.OF.DAY
## 1 10/04/2020 08:31:00 PM   afternoon
## 2 08/26/2020 01:33:00 PM   afternoon
## 3 10/18/2020 03:50:00 PM   afternoon
## 4 02/05/2020 02:54:00 PM   afternoon
## 5 08/26/2020 02:19:00 AM     evening
## 6 04/09/2020 02:00:00 PM   afternoon
```

```r
pie(table(crimeDataset$TIME.OF.DAY))
```



Napravit ćemo test o homogenosti u kojem želimo viditi postoji li razlika u količini zločina s obzirom na doba dana. Napravit ćemo test homogenosti u kojem ćemo provjeriti je li broj zločina opasnih po život jednak za sva 3 doba dana. Zločine koje smo uzeli da su opasni po život nalaze se u varijabli `dangCrimes`.

Var1	Freq
ARSON	545
ASSAULT	18603
BATTERY	43158
BURGLARY	9155
CONCEALED CARRY LICENSE VIOLATION	153
CRIM SEXUAL ASSAULT	233
CRIMINAL DAMAGE	25119
CRIMINAL SEXUAL ASSAULT	960
CRIMINAL TRESPASS	4717
DECEPTIVE PRACTICE	14498
GAMBLING	29
HOMICIDE	724
HUMAN TRAFFICKING	4
INTERFERENCE WITH PUBLIC OFFICER	808
INTIMIDATION	159
KIDNAPPING	140
LIQUOR LAW VIOLATION	162
MOTOR VEHICLE THEFT	9446
NARCOTICS	8258
NON-CRIMINAL	2
OBSCENITY	53
OFFENSE INVOLVING CHILDREN	1898
OTHER NARCOTIC VIOLATION	10
OTHER OFFENSE	12910
PROSTITUTION	395
PUBLIC INDECENCY	8
PUBLIC PEACE VIOLATION	1365
ROBBERY	7890
SEX OFFENSE	1009
STALKING	183
THEFT	45747
WEAPONS VIOLATION	7691

dangerousCrimes	Freq
dangerous	103354
less dangerous	112678

	dangerous	less dangerous
afternoon	41532	50498
evening	35107	28222
morning	26715	33958

```
##
##  Pearson's Chi-squared test
##
## data:  dangerous
## X-squared = 2088.1, df = 2, p-value < 2.2e-16
```