Eurostat2019 analysis

Akolade Sofiyyah Iwalewa,22201441

2022-11-11

```r
# Loading the data into R
df <- read.csv("EurostatCrime2019.csv",
    header = TRUE, # first row contains column names
    row.names = 1) # first column contains country names
```

```r
# Having an overview of the data set
head(df)
```

```
##                         Intentional.homicide Attempted.intentional.homicide
## Albania                                 2.03                           3.25
## Austria                                 0.84                           1.93
## Belgium                                 1.27                           8.87
## Bosnia and Herzegovina                    NA                             NA
## Bulgaria                                 1.14                           0.54
## Croatia                                  0.81                           2.40
##                         Assault Kidnapping Sexual.violence  Rape Sexual.assault
## Albania                    5.52       0.14            5.38  2.69           2.69
## Austria                   43.29       0.07           50.90 18.92          26.64
## Belgium                  556.36         NA           77.45 33.33          44.12
## Bosnia and Herzegovina      NA         NA              NA    NA             NA
## Bulgaria                  39.54       1.03            8.64  1.87             NA
## Croatia                   18.06       0.02           21.05 11.58           8.61
##                         Robbery Burglary
## Albania                    3.42       NA
## Austria                   29.67   613.22
## Belgium                  140.14   565.92
## Bosnia and Herzegovina      NA       NA
## Bulgaria                  16.90    79.81
## Croatia                   20.56   265.73
##                         Burglary.of.private.residential.premises    Theft
## Albania                                                    40.42   168.84
## Austria                                                    99.31  1302.92
## Belgium                                                   410.12  1951.96
## Bosnia and Herzegovina                                        NA       NA
## Bulgaria                                                      NA   473.88
## Croatia                                                    78.53   291.00
##                         Theft.of.a.motorized.land.vehicle
## Albania                                             11.11
## Austria                                             44.22
## Belgium                                            109.76
## Bosnia and Herzegovina                                 NA
## Bulgaria                                            18.87
## Croatia                                             25.42
##                         Unlawful.acts.involving.controlled.drugs.or.precursors
## Albania                                                                  70.26
## Austria                                                                 494.05
```

```
## Belgium                                                  547.74
## Bosnia and Herzegovina                                       NA
## Bulgaria                                                   78.14
## Croatia                                                   272.16
```

```
# Checking the dimension of the data set
dim(df)
```

```
## [1] 41 13
```

> The dataset contains 41 observations and 13 varaibles

```
# Checking the structure of the data set
str(df)
```

```
## 'data.frame':    41 obs. of  13 variables:
##  $ Intentional.homicide                             : num  2.03 0.84 1.27 NA 1.14 0.81 1.48 0.76 0.91 NA ...
##  $ Attempted.intentional.homicide                   : num  3.25 1.93 8.87 NA 0.54 2.4 1.71 0.58 2.57 NA ...
##  $ Assault                                          : num  5.52 43.29 556.36 NA 39.54 ...
##  $ Kidnapping                                       : num  0.14 0.07 NA NA 1.03 0.02 0.91 0.11 NA NA ...
##  $ Sexual.violence                                  : num  5.38 50.9 77.45 NA 8.64 ...
##  $ Rape                                             : num  2.69 18.92 33.33 NA 1.87 ...
##  $ Sexual.assault                                   : num  2.69 26.64 44.12 NA NA ...
##  $ Robbery                                          : num  3.42 29.67 140.14 NA 16.9 ...
##  $ Burglary                                         : num  NA 613.2 565.9 NA 79.8 ...
##  $ Burglary.of.private.residential.premises         : num  40.4 99.3 410.1 NA NA ...
##  $ Theft                                            : num  169 1303 1952 NA 474 ...
##  $ Theft.of.a.motorized.land.vehicle                : num  11.1 44.2 109.8 NA 18.9 ...
##  $ Unlawful.acts.involving.controlled.drugs.or.precursors: num  70.3 494.1 547.7 NA 78.1 ...
```

> All the variables are numeric

```r
#For most countries sexual violence figures are the sum of rape and sexual assault, so I will remove the columns Rape and Se
xual.assault.
df <- subset(df,
              select = -c(Rape,
                            Sexual.assault))
```

```r
# Checking if the columns has been removed
str(df)
```

```
## 'data.frame':    41 obs. of  11 variables:
##  $ Intentional.homicide                            : num  2.03 0.84 1.27 NA 1.14 0.81 1.48 0.76 0.91 NA ...
##  $ Attempted.intentional.homicide                  : num  3.25 1.93 8.87 NA 0.54 2.4 1.71 0.58 2.57 NA ...
##  $ Assault                                         : num  5.52 43.29 556.36 NA 39.54 ...
##  $ Kidnapping                                      : num  0.14 0.07 NA NA 1.03 0.02 0.91 0.11 NA NA ...
##  $ Sexual.violence                                 : num  5.38 50.9 77.45 NA 8.64 ...
##  $ Robbery                                         : num  3.42 29.67 140.14 NA 16.9 ...
##  $ Burglary                                        : num  NA 613.2 565.9 NA 79.8 ...
##  $ Burglary.of.private.residential.premises        : num  40.4 99.3 410.1 NA NA ...
##  $ Theft                                           : num  169 1303 1952 NA 474 ...
##  $ Theft.of.a.motorized.land.vehicle               : num  11.1 44.2 109.8 NA 18.9 ...
##  $ Unlawful.acts.involving.controlled.drugs.or.precursors: num  70.3 494.1 547.7 NA 78.1 ...
```

```r
# For some countries Theft also include burglary, and theft of motorized land vehicle, in others they are recorded separatel
y. I will remove the columns involving theft and burglary.
df <- subset(df,
              select = -c(Theft,
                            Theft.of.a.motorized.land.vehicle,
                            Burglary,
                            Burglary.of.private.residential.premises))
```

```r
# Checking if the columns has been removed
str(df)
```

```
## 'data.frame':    41 obs. of  7 variables:
##  $ Intentional.homicide                        : num  2.03 0.84 1.27 NA 1.14 0.81 1.48 0.76 0.91 NA ...
##  $ Attempted.intentional.homicide              : num  3.25 1.93 8.87 NA 0.54 2.4 1.71 0.58 2.57 NA ...
##  $ Assault                                     : num  5.52 43.29 556.36 NA 39.54 ...
##  $ Kidnapping                                  : num  0.14 0.07 NA NA 1.03 0.02 0.91 0.11 NA NA ...
##  $ Sexual.violence                             : num  5.38 50.9 77.45 NA 8.64 ...
##  $ Robbery                                     : num  3.42 29.67 140.14 NA 16.9 ...
##  $ Unlawful.acts.involving.controlled.drugs.or.precursors: num  70.3 494.1 547.7 NA 78.1 ...
```

```r
# Adding a new column that contains overall record of each country
df$Total <- rowSums(df, na.rm = FALSE)
```

```r
# Checking the countries with missing data
checkNA <- df[!complete.cases(df), ]
rownames(checkNA)
```

```
##  [1] "Belgium"             "Bosnia and Herzegovina" "Denmark"
##  [4] "England and Wales"   "Estonia"             "France"
##  [7] "Hungary"             "Iceland"             "Liechtenstein"
## [10] "Netherlands"         "North Macedonia"     "Northern Ireland (UK)"
## [13] "Norway"              "Poland"              "Portugal"
## [16] "Scotland"            "Slovakia"            "Sweden"
## [19] "Turkey"
```

```r
# Removing the countries with missing data
df <- df[complete.cases(df), ]
str(df)
```

```
## 'data.frame':    22 obs. of  8 variables:
##  $ Intentional.homicide                      : num  2.03 0.84 1.14 0.81 1.48 0.76 1.59 0.71 0.71 0.71 ...
##  $ Attempted.intentional.homicide            : num  3.25 1.93 0.54 2.4 1.71 0.58 5.96 2.18 1.09 0.55 ...
##  $ Assault                                   : num  5.52 43.29 39.54 18.06 20.09 ...
##  $ Kidnapping                                : num  0.14 0.07 1.03 0.02 0.91 0.11 0.02 5.44 0.66 1.71 ...
##  $ Sexual.violence                           : num  5.38 50.9 8.64 21.05 1.94 ...
##  $ Robbery                                   : num  3.42 29.67 16.9 20.56 6.28 ...
##  $ Unlawful.acts.involving.controlled.drugs.or.precursors: num  70.3 494.1 78.1 272.2 117.8 ...
##  $ Total                                     : num  90 621 146 335 150 ...
```

```
#Checking the dimension of my newly formed data set
dim(df)
```

```
## [1] 22  8
```

```
#Determining the 3 most common crimes in Ireland in 2019
names(sort(df["Ireland", -ncol(df)], decreasing = TRUE)[1:3])
```

```
## Warning in xtfrm.data.frame(x): cannot xtfrm data frames
```

```
## [1] "Unlawful.acts.involving.controlled.drugs.or.precursors"
## [2] "Assault"
## [3] "Sexual.violence"
```

The 3 most common crimes in Ireland in 2019 are: - Unlawful acts involving controlled drugs or precursors -
Assault
- Sexual violence

```
# Determining the proportion of the overall crimes that was due to Assault in Ireland in 2019
df["Ireland",]$Assault / df["Ireland",]$Total
```

```
## [1] 0.1605316
```

The proportion of the overall crimes that was due to Assault in Ireland in 2019 is 0.1605316

```
# Determining Which country had the highest record of kidnapping in 2019
rownames(df)[which.max(df$Kidnapping)]
```

```
## [1] "Luxembourg"
```

The country that had the highest record of kidnapping in 2019 was Luxembourg.

```
# Determining Which country had the lowest overall record of offences in 2019
rownames(df)[which.min(df$Total)]
```

```
## [1] "Romania"
```

The country that had the lowest overall record of offences in 2019 was Romania.

```
# plotting a chart that displays the relationship between robbery and unlawful acts involving controlled drugs or precursors.
plot(df$Robbery,
     df$Unlawful.acts.involving.controlled.drugs.or.precursors,
     xlab = "Robbery",
     ylab = "Unlawful acts involving controlled drugs or precursors",
     pch = 15, col = "blue", ylim = c(0,1000))
```
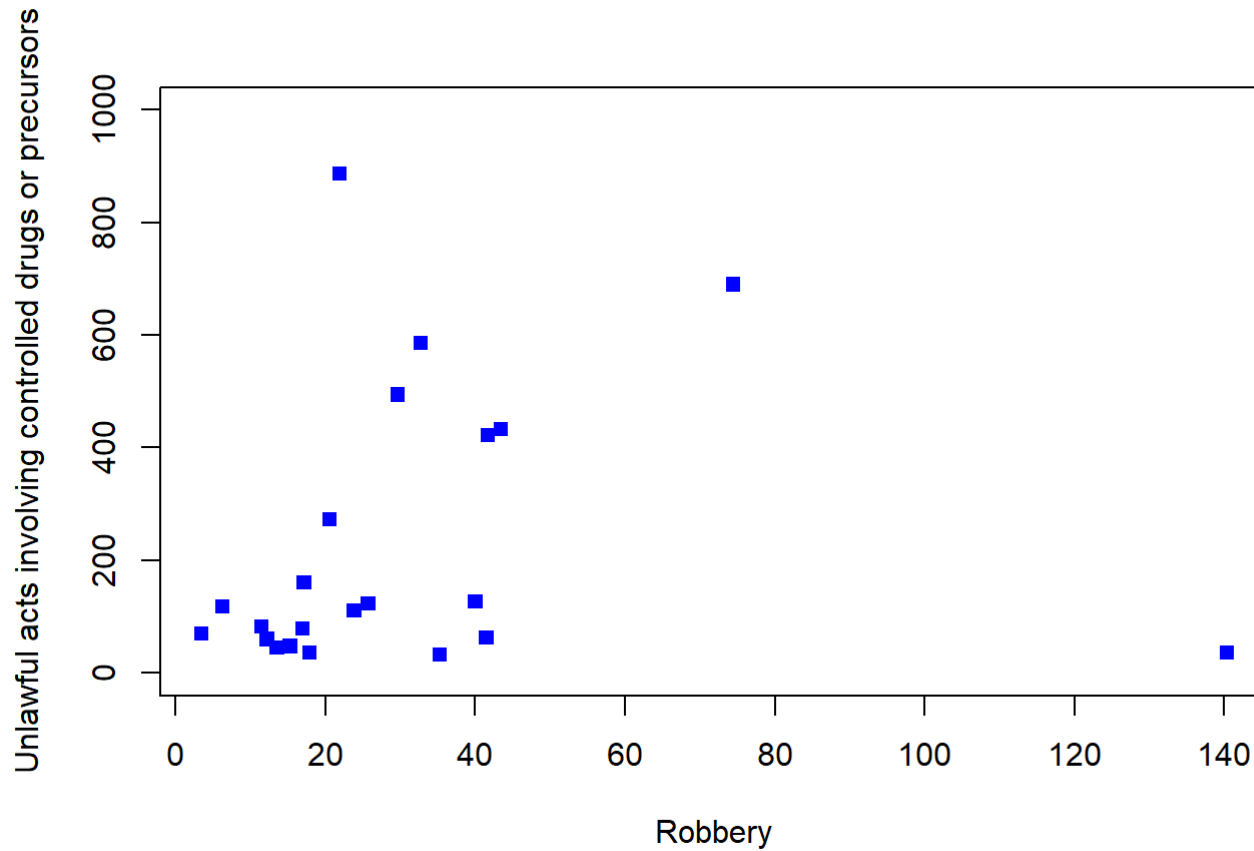
Unlawful acts involving controlled drugs or precursors

Robbery

The relationship between robbery and unlawful acts involving controlled drugs or precursors is not strong enough.