



Analysis of recorded offences by offence category (2019)

Idelson Mindo

2022-10-12

Contents

0.1	Introduction	2
0.2	Method	2
0.2.1	Load data	2
0.2.2	The size (number of rows and columns)	2
0.2.3	The structure of the dataset	3
0.2.4	Remove the columns Rape and Sexual.assault	3
0.2.5	Remove the columns involving theft and burglary	3
0.2.6	Overall record of offences for each country (per hundred thousand inhabitants)	4
0.2.7	Missing data	4
0.2.8	Remove missing data	4
0.2.9	Final dataframe	5
0.3	Analysis and Results	5
0.3.1	The 3 most common crimes in Ireland in 2019	5
0.3.2	Proportion of the overall crimes due to Assault in Ireland in 2019	5
0.3.3	The highest record of kidnapping per country in 2019	6
0.3.4	The lowest overall record of offences per country in 2019	6
0.3.5	Relationship between robbery and unlawful acts involving controlled drugs or precursors	6
0.3.6	Overall offences per countries	7

0.3.7	Offences correlation	8
0.4	Conclusion	10
0.5	Material and data source	10

0.1 Introduction

The present assignment aims to manipulate and analyse police recorded offences by crime in country level for European Union Member States, EFTA countries, EU Candidate countries, and EU Potential Candidates based on homicide, assault, kidnapping, sexual violence, rape, sexual assault, robbery, burglary, burglary of private residential premises, theft, theft of a motorized land vehicle and drug crimes variables.

The remainder of this report is organised as follows: In Section 2, Methods - data source and data manipulation as well as details of the code. In section 3, present the data analyses and results. In section 4, conclusion and section 5, Material and data source.

0.2 Method

0.2.1 Load data

The function `read.csv()` is utilized to load the dataset as showed in the code below:

```
# import csv file and set the first column as row name
esc<- read.csv("data/EurostatCrime2019.csv", row.names = 1)
head(esc[, (colnames(esc[,1:3]))]) # print rows and 3 columns
```

	Intentional.homicide	Attempted.intentional.homicide	Assault
Albania	2.03	3.25	5.52
Austria	0.84	1.93	43.29
Belgium	1.27	8.87	556.36
Bosnia and Herzegovina	NA	NA	NA
Bulgaria	1.14	0.54	39.54
Croatia	0.81	2.40	18.06

In the table above is possible to have a basic demonstration of our data composed by different variables in columns and observations in rows. The have this preview of the table was utilized the `head()` function.

0.2.2 The size (number of rows and columns)

The `dim()` function was used to determine the size of rows and columns within the dataset

```
dim(esc) # dataset size
```

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

```
#dim(esc)[2]
```

This dataset is composed by 41 rows and 13 columns.

0.2.3 The structure of the dataset

The function `str()` was used to determine the structure of the dataset. This dataset is composed by 41 observations and 13 variables.

```
str(esc) # dataset structure and data type
```

```
## 'data.frame':    41 obs. of  13 variables:
## $ Intentional.homicide           : num  2.03 0.84 1.27 NA 1.14 0.81
## $ Attempted.intentional.homicide : num  3.25 1.93 8.87 NA 0.54 2.4
## $ Assault                        : num  5.52 43.29 556.36 NA 39.54
## $ Kidnapping                    : num  0.14 0.07 NA NA 1.03 0.02
## $ 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 ...
```

0.2.4 Remove the columns Rape and Sexual.assault

For most countries sexual violence figures are the sum of rape and sexual assault so this data is not longer important for further analyses. So they will be removed from the dataset. The function `names()` from base r will be used to perform the process.

```
esc_remove2var <- esc[ , ! names(esc) %in% c("Rape", "Sexual.assault")]
```

0.2.5 Remove the columns involving theft and burglary

For some countries Theft includes also burglary, and theft of motorized land vehicle, in others they are recorded separately. In order to compare the different countries will be removed the following variables: - Theft, - Theft.of.a.motorized.land.vehicle, - Burglary, - Burglary.of.private.residential.premises

The function `names()` from base r will be used to perform the process.

```
esc_remove4var <- esc_remove2var[ , ! names(esc_remove2var)
%in% c("Theft", "Theft.of.a.motorized.land.vehicler",
      "Burglary", "Burglary.of.private.residential.premises")]
```

0.2.6 Overall record of offences for each country (per hundred thousand inhabitants)

To have the overall record of offences for each country will be added a column. and to perform this task will be used the `apply()` function.

```
esc_remove4var$overall <- apply(esc_remove4var[, (1:7)], 1, sum)
head(esc_remove4var[8])
```

	overall
Albania	90.00
Austria	620.75
Belgium	NA
Bosnia and Herzegovina	NA
Bulgaria	145.93
Croatia	335.06

0.2.7 Missing data

From the new dataset created, the `rownames()` function, combined with `which()` and `complete.cases()` function was used to list the countries that contain any missing data. Therefore was identifies 19 countries with missing data.

```
rownames(esc_remove4var)[which(complete.cases(esc_remove4var) == FALSE)]
```

```
## [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"
```

0.2.8 Remove missing data

The process bellow aims to remove the countries with missing data from the dataframe. Therefore, was used the `na.omit` function to perform the task.

```
esc_removeNA <- na.omit(esc_remove4var)
```

0.2.9 Final dataframe

After the detailed data cleaning process, the final dataframe contain 22 observations and 8 variables. The `dim()` function was called again to perform this task. therefore renamed from `esc_removeNA` to `crimeData`.

```
crimeData <- esc_removeNA
dim(crimeData)
```

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

0.3 Analysis and Results

0.3.1 The 3 most common crimes in Ireland in 2019

According to these data, Unlawful.acts.involving.controlled.drugs.or.precursors with 421.84 records, Assault (102.18 records)and Sexual.violence with 67.86 records were the 3 most common crimes in Ireland in 2019.

```
cIreland <- crimeData["Ireland",] # filter the Ireland from the countries row name

row.names(cIreland) <- LETTERS[1] # create a colmun name
cIreland_transpose <- t(cIreland) # transpose the cIreland dataframe
# turn the row.name column to a non row.name column
cIreland1 <- cbind(rownames(cIreland_transpose),
                  data.frame(cIreland_transpose, row.names=NULL))
data_new1 <- cIreland1[order(cIreland1$A, decreasing = TRUE), ] # sort the values
colnames(data_new1)[1] ="crime"
colnames(data_new1)[2] ="value" # rename the columns
data_new1[2:4,] # select the top 3 values
```

	crime	value
7	Unlawful.acts.involving.controlled.drugs.or.precursors	421.84
3	Assault	102.18
5	Sexual.violence	67.86

0.3.2 Proportion of the overall crimes due to Assault in Ireland in 2019

16.05% was the proportion of the overall crimes due to Assault in Ireland in 2019.

```
prop_assault_Irl <- cIreland$Assault/cIreland$overall*100 # proportion
prop_assault_Irl
```

```
## [1] 16.05316
```

0.3.3 The highest record of kidnapping per country in 2019

Luxembourg had the highest record of kidnapping in 2019 (per hundred thousand inhabitants) with 7.17 records.

```
rownames(crimeData)[which.max(crimeData$Kidnapping)] # indetify the country
```

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

```
max(crimeData$Kidnapping) # get the value
```

```
## [1] 7.17
```

0.3.4 The lowest overall record of offences per country in 2019

Romania had the lowest overall in 2019 (per hundred thousand inhabitants) with 70.06 records.

```
rownames(crimeData)[which.min(crimeData$overall)] #identify the country
```

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

```
min(crimeData$overall) #get the value
```

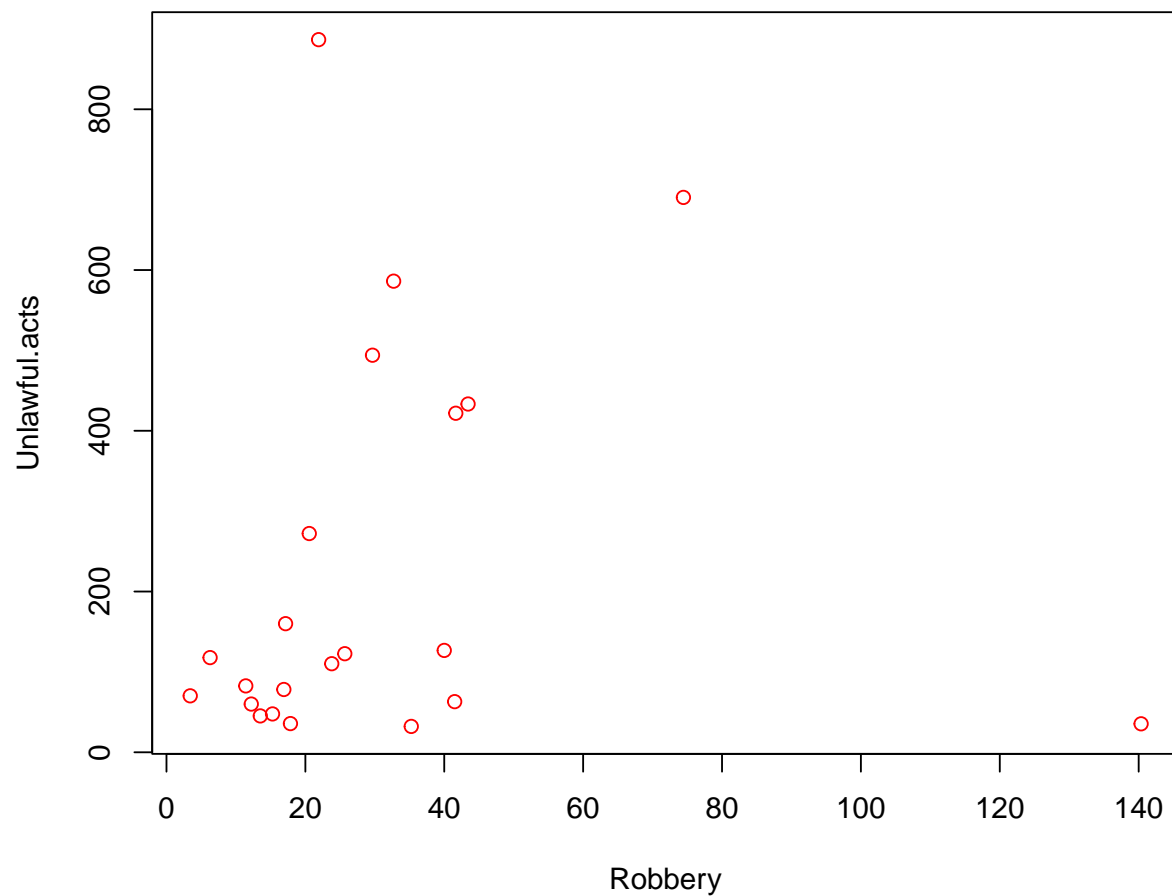
```
## [1] 70.06
```

0.3.5 Relationship between robbery and unlawful acts involving controlled drugs or precursors

The plot bellow displaying the negative correlation between robbery and unlawful acts involving controlled drugs or precursors.

```
plot(
  crimeData$Robbery,
  crimeData$Unlawful.acts.involving.controlled.drugs.or.precursors,
  main="Robbery VS Unlawful acts involving controlled drugs or precursors",
  xlab="Robbery", ylab="Unlawful.acts", col = "red")
```

Robbery VS Unlawful acts involving controlled drugs or precursors

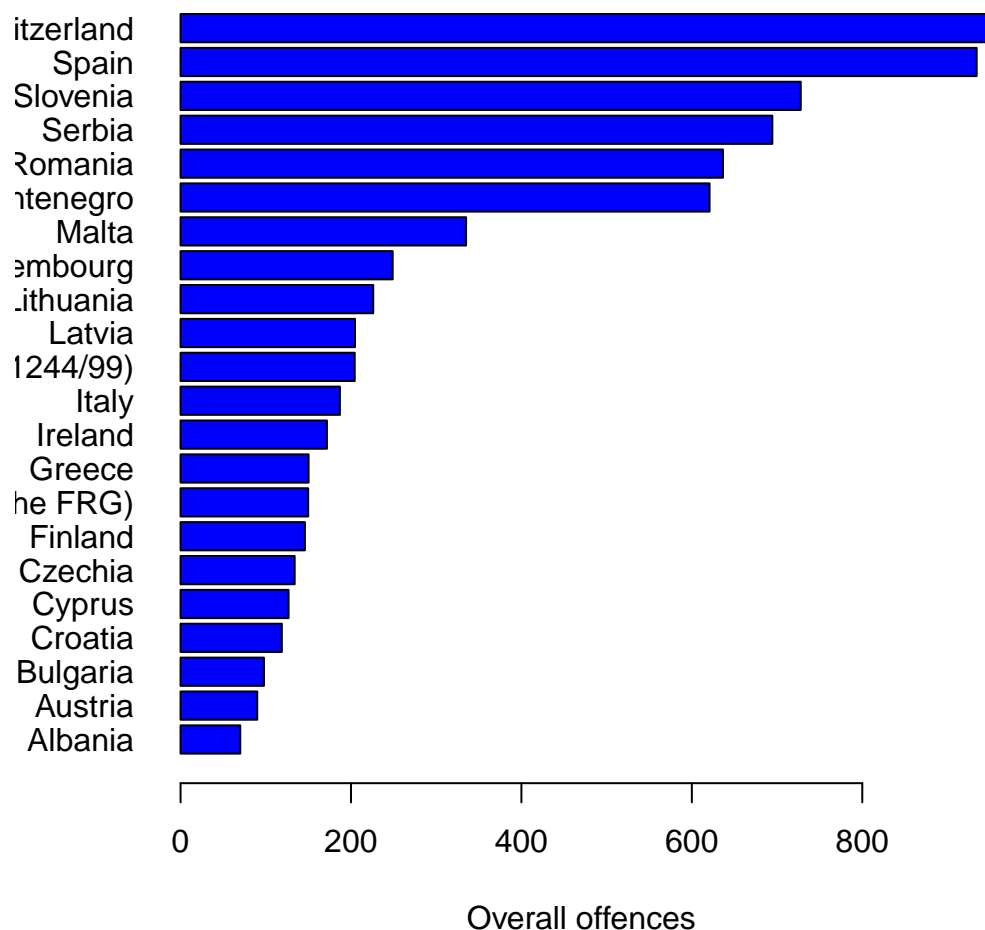


0.3.6 Overall offences per countries

Based on data obtained, Switzerland, Spain and Slovenia were the countries with higher overall records registered in 2019. However Albania, Austria and Bulgaria were the countries with lowest overall offences records registered.

```
barplot(
  sort(crimeData$overall, decreasing = FALSE), names.arg = rownames(crimeData), #sort
  horiz = TRUE, main = "Overall offences per countries",
  xlab = "Overall offences", las = 1,
  col = "blue")
```

Overall offences per countries



0.3.7 Offences correlation

Variables:

- I.homicide = Intentional homicide
- A.I.homicide = Attempted intentional homicide
- Assault = Assault
- Kidnapping = Kidnapping
- Sex.violence = Sexual violence
- Robbery = Robbery
- Unlawful_AICDOP = Unlawful acts involving controlled drugs or precursors

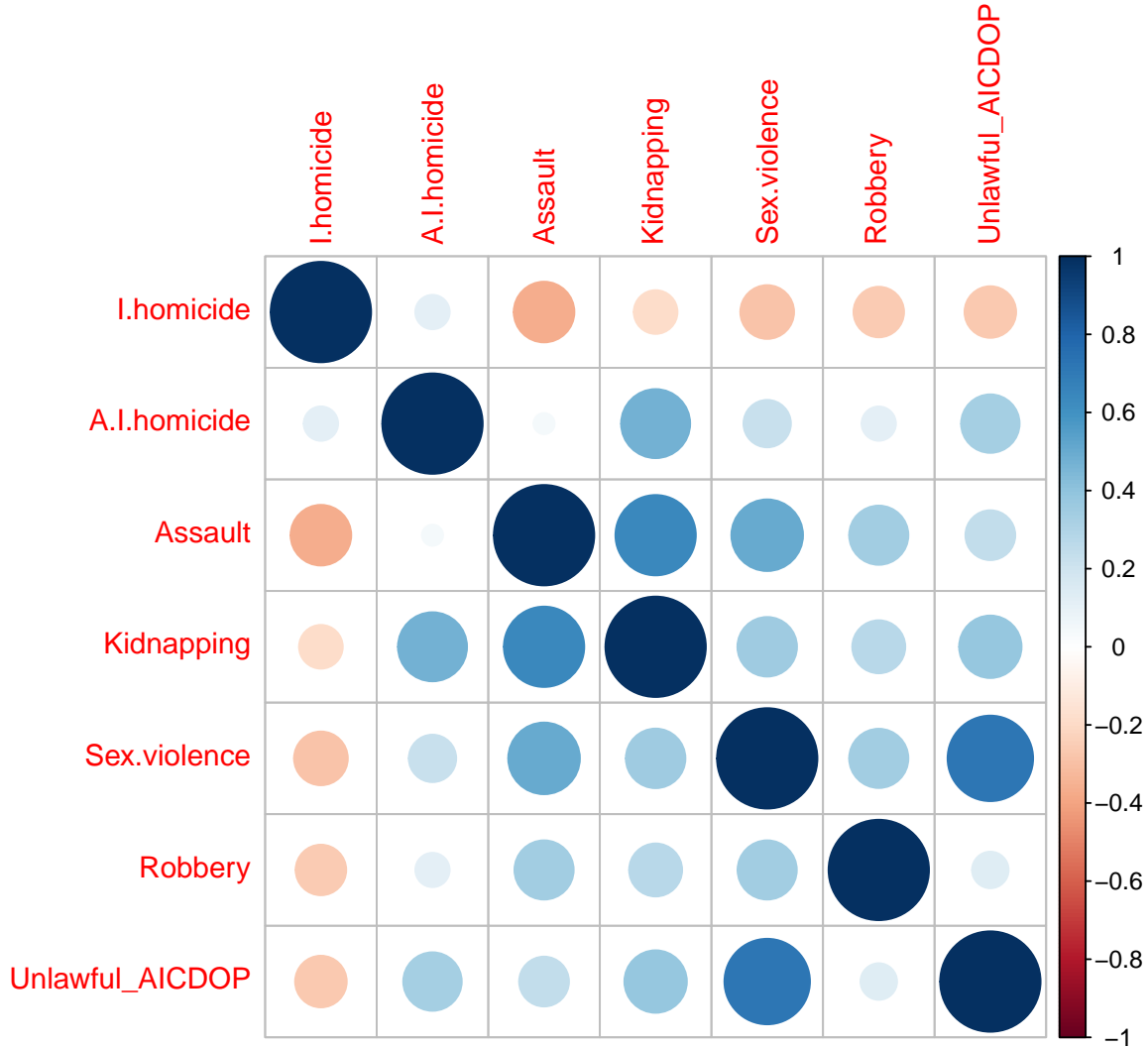
The method is Pearson correlation was applied to infer the relationship among the offences variables and a Correlogram was generated. Positive correlations are displayed in a blue scale while negative correlations are displayed in a red scale.


```
colnames(crimeData) <- c("I.homicide",
                        "A.I.homicide", "Assault",
                        "Kidnapping", "Sex.violence",
                        "Robbery", "Unlawful_AICDOP") #column rename
Correl<-cor(crimeData[,1:7])
Correl
```

```
##           I.homicide A.I.homicide      Assault Kidnapping Sex.violence
## I.homicide      1.0000000    0.11776758 -0.36533263 -0.1898979  -0.2868692
## A.I.homicide    0.1177676    1.00000000  0.04430472  0.4731022   0.2233189
## Assault        -0.3653326    0.04430472  1.00000000  0.6404096   0.5093042
## Kidnapping     -0.1898979    0.47310218  0.64040962  1.0000000   0.3524807
## Sex.violence   -0.2868692    0.22331892  0.50930419  0.3524807   1.0000000
## Robbery        -0.2537762    0.11693840  0.34798732  0.2787385   0.3480115
## Unlawful_AICDOP -0.2644103    0.33965090  0.24667535  0.3872509   0.7254062
##           Robbery Unlawful_AICDOP
## I.homicide      -0.2537762    -0.2644103
## A.I.homicide    0.1169384     0.3396509
## Assault         0.3479873     0.2466754
## Kidnapping      0.2787385     0.3872509
## Sex.violence    0.3480115     0.7254062
## Robbery         1.0000000     0.1328638
## Unlawful_AICDOP 0.1328638     1.0000000
```

The variables are strongly correlated with each other (i.e.: Assault is positively correlated with assaults). The increase of Unlawful acts involving controlled drugs or precursors is correlated with Sexual violence (0.7254062), The Kidnaps are correlated with Assault (0.64040962) and Kidnaps are correlated with Attempted intentional homicide (0.47310218).

```
#install.packages("corrplot")
library(corrplot)
corrplot(Correl) # person correlation
```



0.4 Conclusion

This paper presents an analysis via data analyses to understand the crimes/offences occurred in country level for European Union Member States, EFTA countries, EU Candidate countries, and EU Potential Candidates. The crime category varies across the countries and they are not similar. Countries as Switzerland, Spain and Slovenia had registered higher number of offences records. The offences records in all countries was related each other, for example there is a relation between kidnaps and assaults so when a kidnap occur there is a larger chance that the person kidnapped to be assaulted as well.

0.5 Material and data source

- R website: <https://www.rstudio.com/>
- R Markdown website: <https://rmarkdown.rstudio.com/>
- data source: https://ec.europa.eu/eurostat/cache/metadata/en/crim_off_cat_esms.htm#stat_process1654673830093