Europe Crime Offense 2019 Analysis

Author : Vedanti Borate

Data Cleaning

Loading dataset

Below we fetch the locally saved dataset and review the dataset structure.

```
#set path
  setwd("~/Documents/GitCodes/RProgramming")

#read csv file as table and stores in dataframe
  crime2019<-read.csv("2_EurostatCrime2019.csv",header = TRUE,row.names = 1 )

# summarcrime2 and structure of data
  nrow(crime2019)
```

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

```
  ncol(crime2019)
```

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

```
  str(crime2019)
```

```
## '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 33.3
3 NA 1.87 ...
##  $ Sexual.assault                                    : num  2.69 26.64 44.1
2 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 N
A 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 ...
```

Observations: Dataframe with 41 rows and 13 columns. This shows that the object is a dataframe. It again repeats the number of rows and columns, and shows that all columns are numerical recordings of crime rates. Some NAs are visible too.

Data Manipulations

Most countries sexual violence figures are the sum of rape and sexual assault. Removing these non required columns from dataset. For some countries Theft includes also burglary, and theft of motorised land vehicle, in others they are recorded separately. In order to compare the different countries, removing the columns involving theft and burglary as well.

```
crime2 <- subset(crime2019,
            select = -c(Rape,
            Sexual.assault))

crime2019 <- subset(crime2019,
                select = -c(Theft,
                        Theft.of.a.motorized.land.vehicle,
                        Burglary,
                        Burglary.of.private.residential.premises))
str(crime2019)
```

```
## 'data.frame':    41 obs. of  9 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 33.3
3 NA 1.87 ...
##  $ Sexual.assault                                   : num   2.69 26.64 44.1
2 NA NA ...
##  $ 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 ...
```

Observations: Removed all non required or non consistent data for unbiased data analysis for each country.

Adding new column that holds total of all offences record countrywise.

```
crime2019$TotalOffence <- rowSums(crime2019, na.rm = FALSE)
```

Observations: New Column TotalOffence holds total per hundred thousand inhabitants for different countries. If any record value is NA then sum would NA.

Next, checking if there are any missing data and printing them.

```
checkNA <- crime2019[!complete.cases(crime2019), ]
rownames(checkNA)
```

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

```
#alternative method
# checkNA <- apply(crime2019, 1, function(x) any(is.na(x)))
# rownames(crime2019)[checkNA]
```

Observations: There are 22 records of countries where there is missing information.

Removing missing data

```
crime2019 <- crime2019[complete.cases(crime2019), ]
str(crime2019)
```

```
## 'data.frame':    19 obs. of  10 variables:
##  $ Intentional.homicide                             : num  2.03 0.84 0.81
1.48 0.76 1.59 0.71 0.71 0.71 1.28 ...
##  $ Attempted.intentional.homicide                   : num  3.25 1.93 2.4
1.71 0.58 5.96 2.18 1.09 0.55 6.46 ...
##  $ Assault                                          : num  5.52 43.29 18.0
6 20.09 43.98 ...
##  $ Kidnapping                                       : num  0.14 0.07 0.02
0.91 0.11 0.02 5.44 0.66 1.71 1.17 ...
##  $ Sexual.violence                                  : num  5.38 50.9 21.05
1.94 14.65 ...
##  $ Rape                                             : num  2.69 18.92 11.5
8 1.83 6.41 ...
##  $ Sexual.assault                                   : num  2.69 26.64 8.61
0.11 8.12 ...
##  $ Robbery                                          : num  3.42 29.67 20.5
6 6.28 13.51 ...
##  $ Unlawful.acts.involving.controlled.drugs.or.precursors: num  70.3 494.1 272.
2 117.8 45.2 ...
##  $ TotalOffence                                     : num  95.4 666.3 355.
2 152.2 133.4 ...
```

```
#alternative
# crime2 <- crime2019[checkNA != 1, ]
# str(crime2)
```

```
dim(crime2019)
```

```
## [1] 19 10
```

Observations: New Data has 22 observations (rows) and 8 variables (columns).

Analysis Task

Find 3 most common crimes in Ireland country

```
# select the row containing data for Ireland and remove the last column which cont
ains the Total
Ireland <- crime2019["Ireland", -ncol(crime2019)]
# sort the crimes in Ireland in decreasing order
Ireland <- sort(Ireland, decreasing = TRUE)
```

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

```
# select the names of the top three crimes

#alternative
#names(sort(crime2019["Ireland", -ncol(crime2019)], decreasing = TRUE)[1:3])
```

Observations: Highest amount of offenses in Ireland are Unlawful acts involving controlled drugs or precursors, Assault,Sexual.violence.

Finding Specifics of certain crime offenses.

```
crime2019["Ireland",]$Assault / crime2019["Ireland",]$Total
```

```
## [1] 0.1472716
```

Observations: Proportion of Assault in total offenses of Ireland is 0.1605316

```
rownames(crime2019)[which.max(crime2019$Kidnapping)]
```

```
## [1] "Germany (until 1990 former territory of the FRG)"
```

Luxembourg is the country with the highest record of kidnapping amongst all countries recorded.

```
rownames(crime2019)[which.min(crime2019$Total)]
```

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

Romania is the country with the lowest overall record of offences.

```
plot(crime2019$Robbery,
     crime2019$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))
```
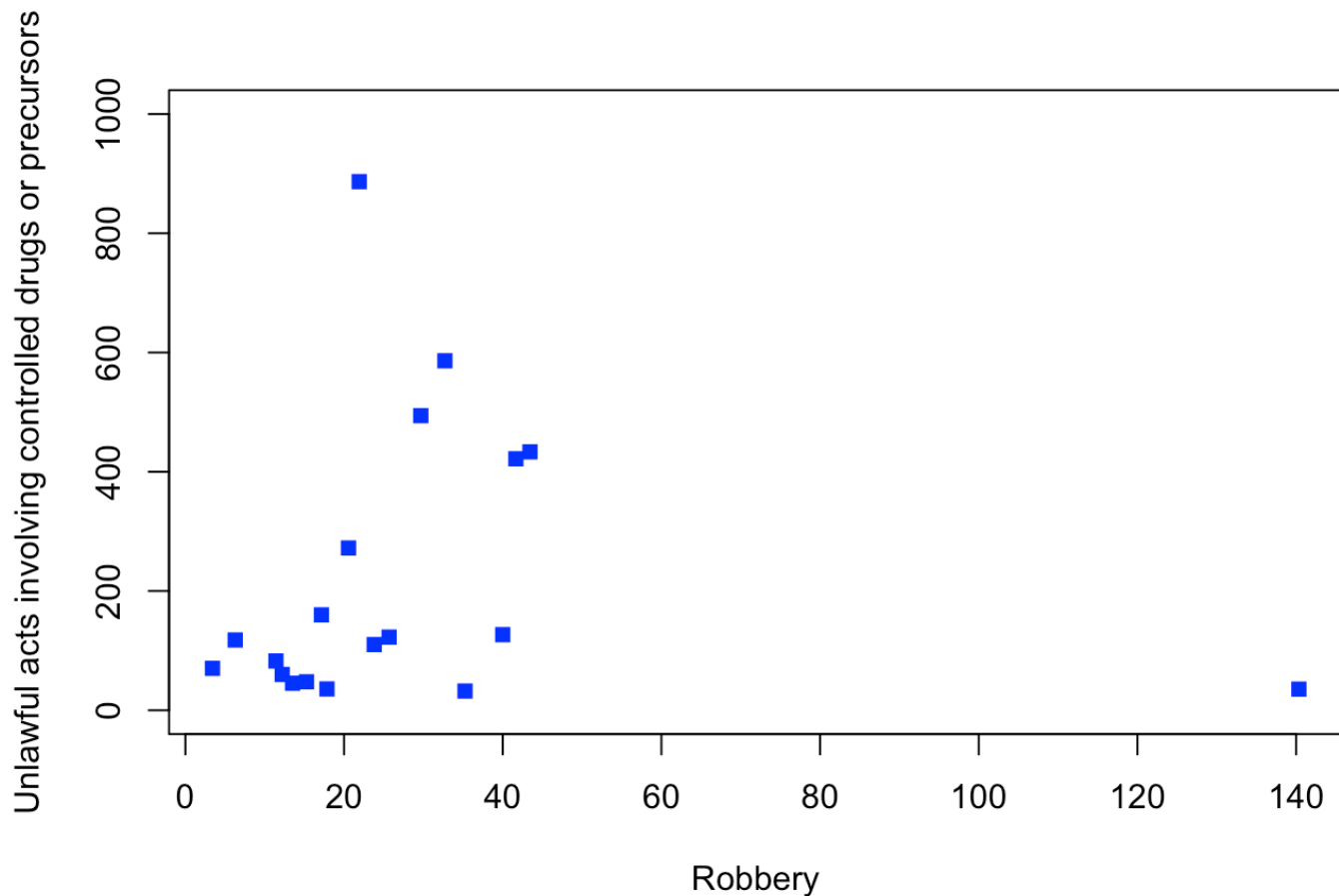
Observations: This scatter plot shows a strong positive correlation between the number of unlawful acts involving controlled drugs or precursors and the number of robberies. The graph shows that as the number of unlawful acts involving controlled drugs or precursors increases, so does the number of robberies. The graph also shows that there are a few outliers, which are data points that fall far from the trendline.

Creativity Task

```
ggpairs(crime2019, columns=(5:10))
```
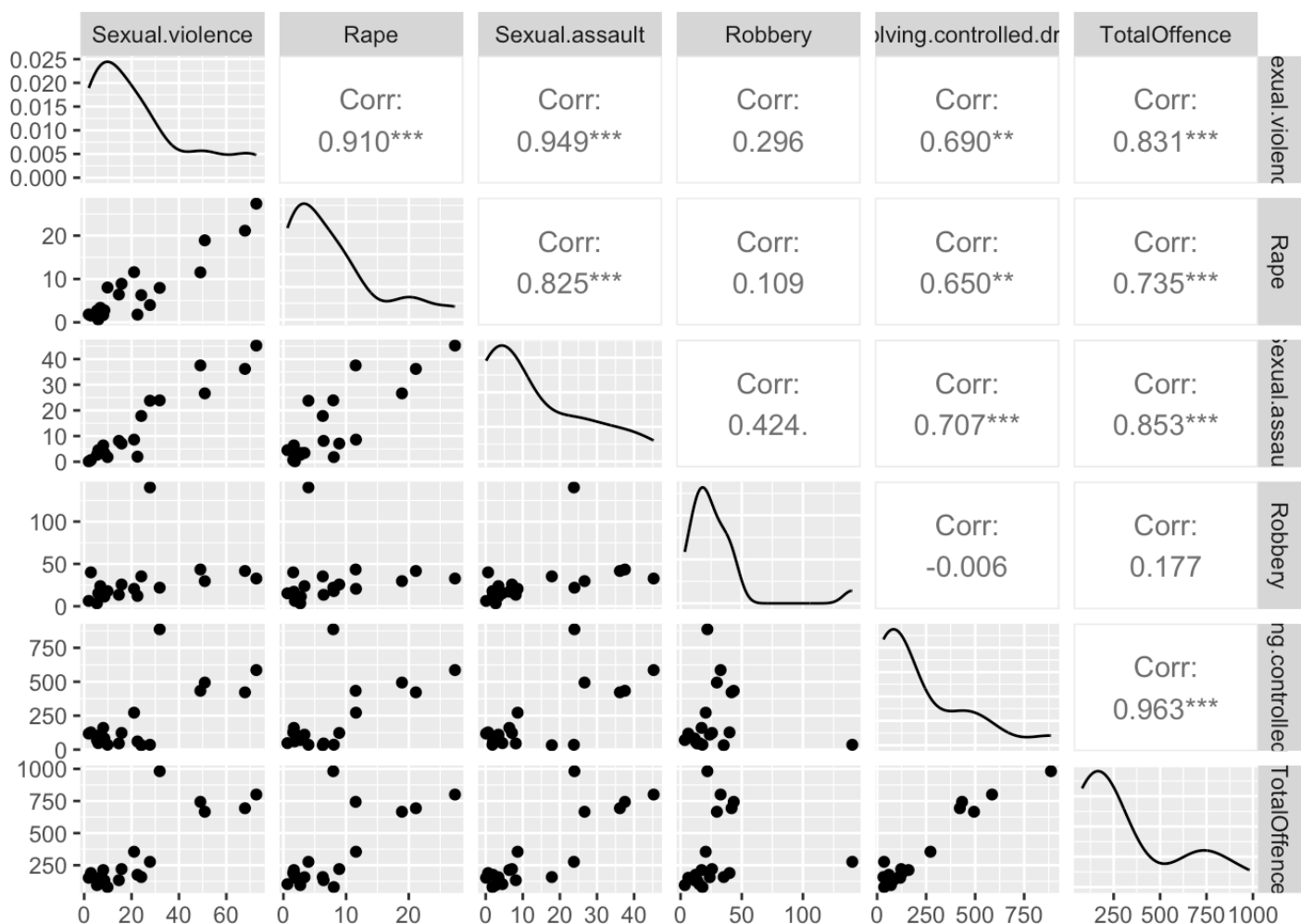
Observations: The graph shows the correlation between the number of cases of sexual violence and other crimes. The correlation between sexual violence and other crimes is high, with a correlation coefficient of 0.910 for rape, 0.949 for sexual assault, 0.831 for robbery, and 0.825 for total offenses.

This means that as the number of cases of sexual violence increases, so does the number of cases of other crimes. The graph also shows that there is a strong correlation between the different types of sexual violence. such as the correlation between rape and sexual assault is 0.963. This means that the two crimes are often committed together.This information could be useful for law enforcement agencies when developing strategies to combat sexual violence and other crimes.

```
plot(crime2019$Sexual.violence,crime2019$Sexual.assault, xlab="Violence ",      yla
b="Assault",type="p")
title("Sexual Crimes data collected")
```

Sexual Crimes data collected
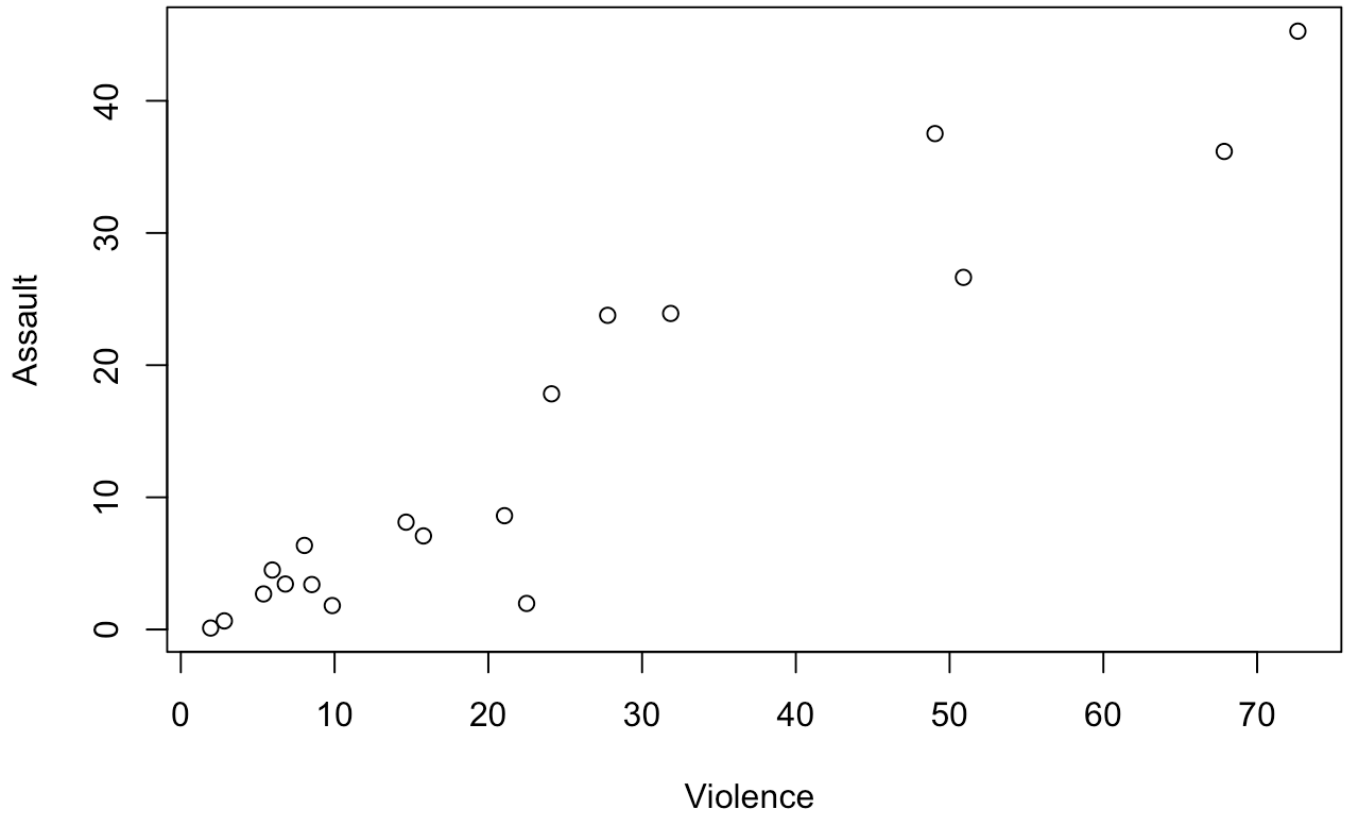
Observations: The scatter plot shows the relationship between the number of sexual crimes committed and the number of violence committed. There is a positive correlation between sexual crimes and violence crimes. This means that as the number of sexual crimes increases, so does the number of violence crimes.