

# EuroStat Crime Analysis

Khusboo Satpute

09/10/2019

## EUROCRIME STAT ANALYSIS

The dataset EurostatCrime2017.csv records offences (values per hundred thousand inhabitants) by offence category in 41 European Countries in 2017. Full information on the dataset is available here: [https://ec.europa.eu/eurostat/cache/metadata/en/crim\\_off\\_cat\\_esms.htm](https://ec.europa.eu/eurostat/cache/metadata/en/crim_off_cat_esms.htm). Completed using R Markdown, check that all the output and code are correctly shown in your final document. Knit your document frequently to fix errors. Once completed, submit the Rmd file and the resulting pdf or word document which shows all your code.

### Task 1: Manipulation

#### 1. Load the dataset EurostatCrime2017.csv. Notice that the first column of the csv file

contains the names of the countries that must be read as row names. 2. What is the size (number of rows and columns) and the structure of this dataset? 3. Produce appropriate commands to answer the following questions: (i) For some countries Theft includes also burglary, and theft of motorised land vehicle, in others they are recorded separately. Add a new column called All Theft which contains the sum of all the crimes that have a theft component: Theft, Theft of a motorized land vehicle, Burglary, Burglary of private residential premises. Please consider NA values as 0 in this case. (ii) Remove the columns: Theft, Theft of a motorized land vehicle, Burglary, and Burglary of private residential premises.

```
df=read.csv("EurostatCrime2017.csv", stringsAsFactors = FALSE, row.names = 1)
```

#### 2. Size & structure of this dataset

```
nrow(df)

## [1] 41

ncol(df)

## [1] 11

str(df)
```

```
## 'data.frame': 41 obs. of 11 variables:
## $ Intentional_homicide : num 1.7 1.34 0
.62 1.06 0.89 2.2 0.86 0.72 0.66 1.41 ...
## $ Attempted_intentional_homicide : num 8.47 0.44
0.72 3.69 2.18 1.22 0.27 1.39 1.76 3.77 ...
## $ Assault : num 611 39.6 4
5.1 33.1 166.1 ...
## $ Kidnapping : num 10.31 1.44
0.16 NA 5.6 ...
## $ Sexual.violence : num 63.22 9.19
13.37 83.41 42.19 ...
## $ Robbery : num 167 21.9 1
5 35.5 47.1 ...
## $ Burglary : num NA 125 228
955 443 ...
## $ Burglary_of_private_residential_premises : num NA NA 68.4
702.6 141.2 ...
## $ Theft : num NA 452 632
3721 1401 ...
## $ Theft_of_a_motorized_land_vehicle : num NA 33.36 2
01.84 3.79 65.58 ...
## $ Unlawful_acts_involving_controlled_drugs_or_precursors: num 506.6 70.2
52.9 481.6 400.6 ...
```

### 3. Produce appropriate commands to answer the following questions:

#### 3.1 Add a new column called All Theft

```
Ncol=c("Burglary", "Theft", "Theft_of_a_motorized_land_vehicle", "Burglary_of
_private_residential_premises")
df$Burglary[is.na(df$Burglary)] <- 0
df$Burglary_of_private_residential_premises[is.na(df$Burglary_of_private_resi
dential_premises)]<- 0
df$Theft<-0
df$Theft_of_a_motorized_land_vehicle<-0

df$Alltheft<-rowSums(df[,Ncol])

df$Alltheft

## [1] 0.00 124.57 296.44 1657.97 583.75 73.35 399.24 811.56
## [9] 602.67 0.00 398.32 323.20 277.38 373.36 90.94 0.00
## [17] 0.00 527.92 727.22 0.00 233.94 341.17 212.05 564.26
## [25] 135.50 488.32 1233.12 1172.66 480.59 365.52 426.49 0.00
## [33] 0.00 777.14 145.09 0.00 39.80 302.83 0.00 0.00
## [41] 320.60
```

### 3.2 Remove the other columns

```
df<- subset(df,select = -c(Burglary,Theft,
                           Theft_of_a_motorized_land_vehicle,
                           Burglary_of_private_residential_premises))
```

### 4. List the countries that contain any missing data

```
country<-list()
for(i in 1:nrow(df))
{
  if(any(is.na(df[i,])))
  {
    country<- c(country,rownames(df[i,]))
  }
}

print(country)

## [[1]]
## [1] "Denmark"
##
## [[2]]
## [1] "France"
##
## [[3]]
## [1] "Croatia"
##
## [[4]]
## [1] "Hungary"
##
## [[5]]
## [1] "Netherlands"
##
## [[6]]
## [1] "Austria"
##
## [[7]]
## [1] "Poland"
##
## [[8]]
## [1] "Portugal"
##
## [[9]]
## [1] "Sweden"
##
## [[10]]
## [1] "England_and_Wales"
##
## [[11]]
## [1] "Iceland"
```

```
##
## [[12]]
## [1] "Liechtenstein"
##
## [[13]]
## [1] "Norway"
##
## [[14]]
## [1] "North_Macedonia"
##
## [[15]]
## [1] "Turkey"
##
## [[16]]
## [1] "Bosnia_and_Herzegovina"
```

## 5. Remove the countries with missing data from the dataframe.

```
df <- na.omit(df)
```

## 6. How many observations and variables are in this new dataframe?

```
dim(df)
```

```
## [1] 25 8
```

## Task 2: Analysis

### 1. The 3 most common crimes in Ireland in 2017?

```
ireland_df <- subset(df, rownames(df) == "Ireland")
top_crime <- colnames(sort(ireland_df, decreasing = TRUE)[1:3])
top_crime

## [1] "Alltheft"
## [2] "Unlawful_acts_involving_controlled_drugs_or_precursors"
## [3] "Assault"
```

### 2. Which country has the highest overall record of offences

```
df$offence <- apply(df, 1, sum)
H_offence <- rownames(df[which.max(df$offence),])
H_offence

## [1] "Switzerland"
```

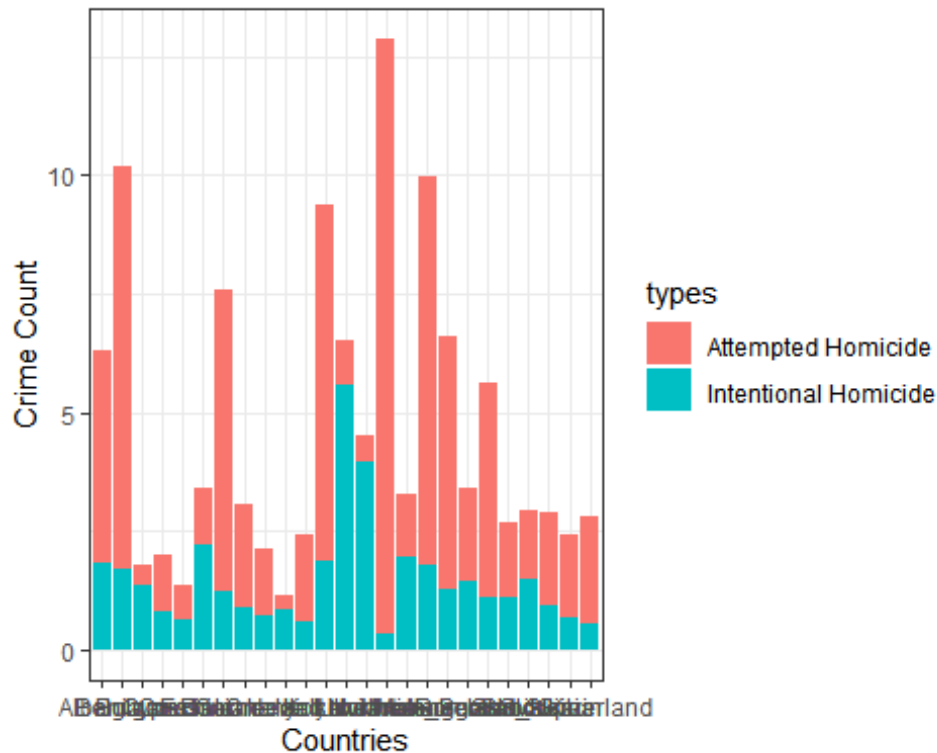
## Task 3: Creativity

```
library(ggplot2)
crimes <- c(df$Intentional_homicide, df$Attempted_intentional_homicide)
countries <- rownames(df)
types <- c(rep("Intentional Homicide", length(df$Intentional_homicide)),
```

```

    rep("Attempted Homicide", length(df$Attempted_intentional_homicide
)))
plot_data <- data.frame(countries, crimes, types)
plt <- ggplot(plot_data, aes(countries, crimes))
plt + geom_bar(stat = 'identity', aes(fill = types)) +
  xlab("Countries") + ylab("Crime Count") + theme_bw()

```



The Plot describe the crimecount for Homicide (Attempted Homicide and Intentional Homicide )for various countries in Euro states. The Graph has X label as Countries and Y label Crime Count from the graph. Also, we can infer the total homicide in a particular country as well as we can also observe that attempted Homicide cases are more in every country indicated by red block than Intentional Homicide indicated by green in color.