Assignment-1

Aniket Guha Roy-19200164

10/11/2019

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 [Hint: Load in the file using the function read.csv]

```
setwd("C:/Users/anike/Desktop/R module/Assignments/Assignment")
EuroCrime = read.csv("EurostatCrime2017.csv",header = T, row.names = 1)
```

2. What is the size (number of rows and columns) and the structure of this dataset?
number of rows: 41 number of columns : 11

```
dim(EuroCrime)
```

```
## [1] 41 11
```

```
str(EuroCrime)
```

```
## 'data.frame':    41 obs. of  11 variables:
##  $ Intentional_homicide                                : num  1.7 1.34 0.62 1.06 0.89 2.2 0.86 0.7
##  $ Attempted_intentional_homicide                      : num  8.47 0.44 0.72 3.69 2.18 1.22 0.27 1
##  $ Assault                                             : num  611 39.6 45.1 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 15 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 33.36 201.84 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:

(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. [Hint: you may want to use the function apply or rowSums, check their help files to see how to deal with missing values.]

```
EuroCrime$AllTheft = rowSums(EuroCrime[,c("Theft","Theft_of_a_motorized_land_vehicle",
                      "Burglary","Burglary_of_private_residential_premises")],na.rm = T)
```

(ii)Remove the columns: Theft, Theft of a motorized land vehicle, Burglary, and Burglary of private residential premises

```
EuroCrime[ ,c("Theft","Theft_of_a_motorized_land_vehicle","Burglary",
              "Burglary_of_private_residential_premises")]=NULL
head(EuroCrime)
```

```
##          Intentional_homicide Attempted_intentional_homicide Assault
## Belgium                  1.70                           8.47  611.03
## Bulgaria                 1.34                           0.44   39.58
## Czechia                  0.62                           0.72   45.06
## Denmark                  1.06                           3.69   33.12
## Germany                  0.89                           2.18  166.09
## Estonia                  2.20                           1.22    5.78
##          Kidnapping Sexual.violence Robbery
## Belgium       10.31           63.22  166.97
## Bulgaria       1.44            9.19   21.94
## Czechia        0.16           13.37   14.98
## Denmark          NA           83.41   35.52
## Germany        5.60           42.19   47.08
## Estonia        0.00           19.69   15.28
##          Unlawful_acts_involving_controlled_drugs_or_precursors AllTheft
## Belgium                                                  506.65     0.00
## Bulgaria                                                  70.25   609.92
## Czechia                                                   52.93  1129.79
## Denmark                                                  481.56  5382.97
## Germany                                                  400.60  2050.69
## Estonia                                                  441.46   679.30
```

4. Work with the dataset you created in question 3ii, and list the countries that contain any missing data
   Ans: Denmark" “France” “Croatia” “Hungary”
   “Netherlands” “Austria” “Poland” “Portugal”
   “Sweden” “England_and_Wales” “Iceland” “Liechtenstein”
   "Norway

```
  row.names(EuroCrime[!complete.cases(EuroCrime),])
```

```
##  [1] "Denmark"                "France"
##  [3] "Croatia"                "Hungary"
##  [5] "Netherlands"            "Austria"
##  [7] "Poland"                 "Portugal"
##  [9] "Sweden"                 "England_and_Wales"
## [11] "Iceland"                "Liechtenstein"
## [13] "Norway"                 "North_Macedonia"
## [15] "Turkey"                 "Bosnia_and_Herzegovina"
```

5. Remove the countries with missing data from the dataframe

```
EuroCrime=EuroCrime[complete.cases(EuroCrime),]
```

6. How many observations and variables are in this new dataframe
   number of rows: 25 number of columns:8

```r
dim(EuroCrime)
```

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

Analysis

1.According to these data what were the 3 most common crimes in Ireland in 2017?
Ans: 1:All_Theft 2.Unlawful Acts 3.Assault

```r
sort(EuroCrime[c("Ireland"),],decreasing = T)[1:3]
```

```
##         AllTheft Unlawful_acts_involving_controlled_drugs_or_precursors
## Ireland  1956.29                                                351.58
##         Assault
## Ireland   84.59
```

2.Which country has the highest overall record of offences (per hundred thousand inhabitants)? Ans: Switzerland

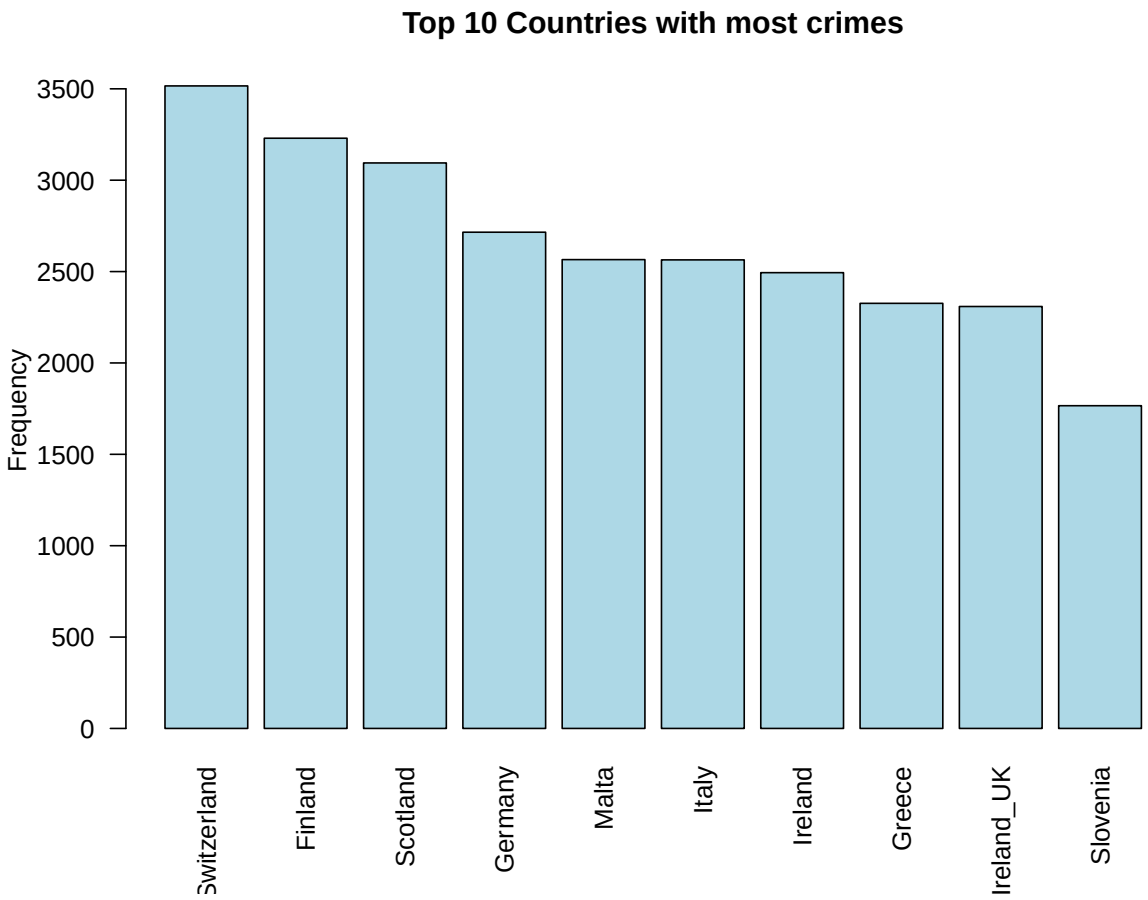
```r
rownames(EuroCrime)[rowSums(EuroCrime) == max(rowSums(EuroCrime))]
```

```
## [1] "Switzerland"
```

Creativity

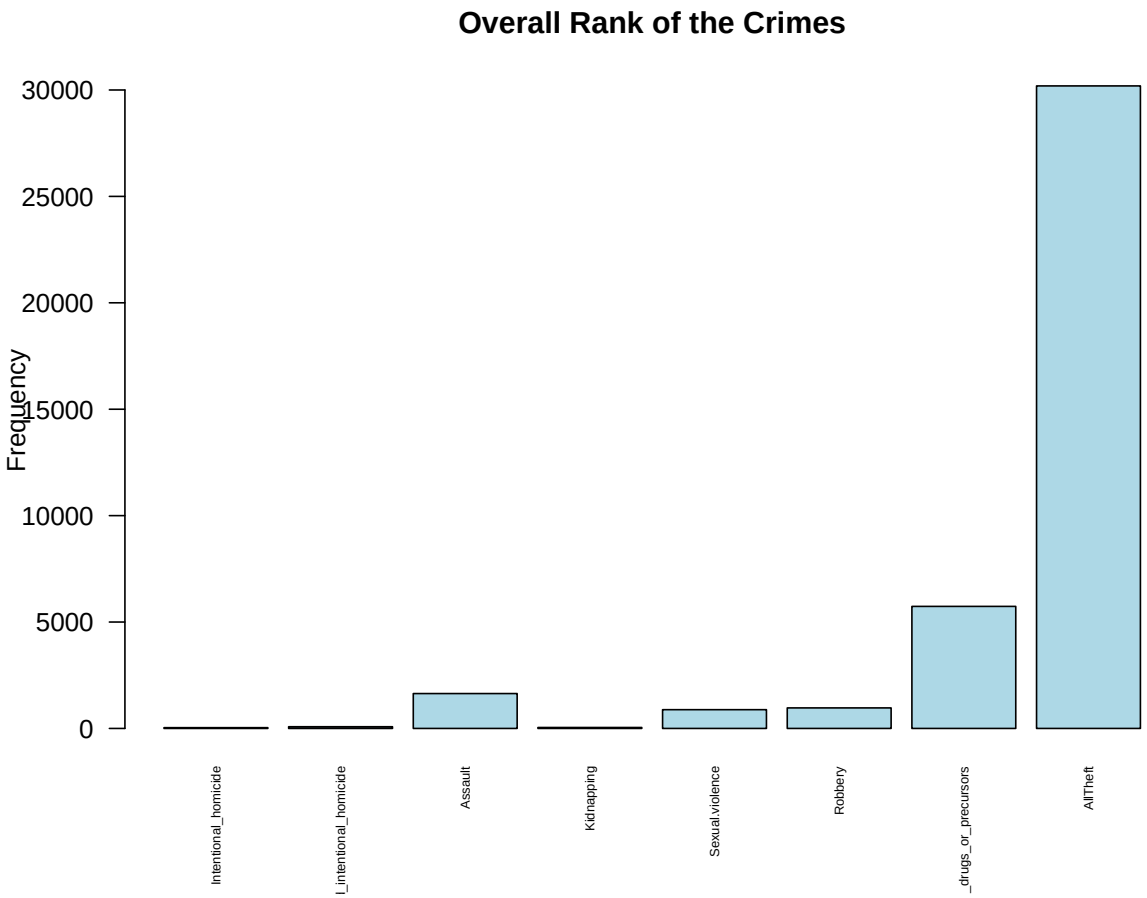
1.First we are going to take a look at the top countries having highest overall record of offences:
We already knew Switzerland was the country with higest offenses recorded but we get an idea of other top countries and how they match up to SWitzerland as well

```r
barplot(sort(rowSums(EuroCrime),decreasing = T)[1:10],ylab = "Frequency",
        main = "Top 10 Countries with most crimes",las=2, col = "lightblue")
```



2. Rank of the Crimes
   This graph gives us an idea of which is the most common committed crime, i.e. Theft/Burglary. Next most common crime is Unlawful_acts_involving_controlled_drugs_or_precursors followed by Assault.But Unlawful acts is quite less than Theft. Hence overall if we could bring down Thefts, the crime rate might get reduced overall.

```
barplot(colSums(EuroCrime), main = "Overall Rank of the Crimes",
        ylab = "Frequency",las=2, cex.names = 0.5,col = "lightblue")
```



3. All_Thefts
   This graph gives an idea of the most committed crime-All Thefts

- We infer that there are no outliers.
- All the observations lie between 1st and 3rd quadrants
- Skewed right or positively skewed
- Median slightly above 1000
  INFERENCE:
- There are no countries in the data which have a frequency of Theft less than approx 600 and more than roughly 2100.
- Hence we get to know most of the countries in the data have an occurence of thefts/ burglary more than 1000 times.

```
boxplot(EuroCrime$AllTheft, main = "ALL_Thefts", ylab = "frequency")
```


ALL_Thefts
frequency
0
500
1000
2000