Assignment 1

K.Saketh Sai Nigam 22201204

2022-09-27

Task 1: Manipulation

```
#1. Load the dataset EurostatCrime2019.csv. Notice that the first column of the cs
v filecontains the names of the countries that must be read as row names [Hint: Lo
ad in the file using the function read.csv].

#Loading the dataset
D <- read.csv("/Users/saketh/Desktop/R /Ass1/EurostatCrime2019.csv",row.names = 'X
')

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

#No:of Rows of Dataset 'D'
print(nrow(D))
```

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

```
#No:of Columns of Dataset 'D'
print(ncol(D))
```

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

```
#Structure of Dataset 'D'
print(str(D))
```

```
## '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 ...
## NULL
```

```
#3.(i) For most countries sexual violence figures are the sum of rape and sexual a
ssault. Remove the columns Rape and Sexual.assault.

#Removing the columns Rape and Sexual.assault.
D <- subset(D,select = -c(Rape,Sexual.assault))
head(D)
```

```
##                         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 Robbery Burglary
## Albania                    5.52       0.14            5.38    3.42       NA
## Austria                   43.29       0.07           50.90   29.67   613.22
## Belgium                  556.36         NA           77.45  140.14   565.92
## Bosnia and Herzegovina       NA         NA              NA      NA       NA
## Bulgaria                  39.54       1.03            8.64   16.90    79.81
## Croatia                   18.06       0.02           21.05   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
```

```
#3.(ii) For some countries Theft includes also burglary, and theft of motorised la
nd vehicle, in others they are recorded separately. In order to compare the differ
ent countries, remove the columns involving theft and burglary: 1.Theft,2. Theft.o
f.a.motorized.land.vehicle,3. Burglary,4. Burglary.of.private.residential.premises
.

#Removing the columns Theft,Theft.of.a.motorized.land.vehicle,Burglary,Burglary.of
.private.residential.premises.
D <- subset(D,select = -c(Theft,Theft.of.a.motorized.land.vehicle,Burglary,Burglar
y.of.private.residential.premises))
head(D)
```

```
##                         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 Robbery
## Albania                    5.52       0.14            5.38    3.42
## Austria                   43.29       0.07           50.90   29.67
## Belgium                  556.36         NA           77.45  140.14
## Bosnia and Herzegovina       NA         NA              NA      NA
## Bulgaria                 39.54       1.03            8.64   16.90
## Croatia                  18.06       0.02           21.05   20.56
##                         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
```

```
#3.(iii) Add a column containing the overall record of offences for each country ( per hundred thousand inhabitants)?

#Creating the column called OverAllRecords
D$OverAllRecords = D$Intentional.homicide+D$Attempted.intentional.homicide+D$Assault+D$Kidnapping+D$Sexual.violence+D$Robbery+D$Unlawful.acts.involving.controlled.drugs.or.precursors
head(D)
```

```
##                         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 Robbery
## Albania                    5.52       0.14            5.38    3.42
## Austria                   43.29       0.07           50.90   29.67
## Belgium                  556.36         NA           77.45  140.14
## Bosnia and Herzegovina       NA         NA              NA      NA
## Bulgaria                 39.54       1.03            8.64   16.90
## Croatia                  18.06       0.02           21.05   20.56
##                         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
##                         OverAllRecords
## Albania                         90.00
## Austria                        620.75
## Belgium                            NA
## Bosnia and Herzegovina             NA
## Bulgaria                       145.93
## Croatia                        335.06
```

```
#4. Work with the dataset you just created, and list the countries that contain an
y missing data

#Printing the countries which have the missing data
CountriesOfMissingData <- names(which(rowSums(is.na(D))>0))
print(CountriesOfMissingData)
```

```
##  [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"
```

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

#Removing the countries which have misssing data from the Dataset 'D'
D <- D[!(row.names(D) %in% CountriesOfMissingData),]
head(D)
```

```
##          Intentional.homicide Attempted.intentional.homicide Assault Kidnapping
## Albania                  2.03                            3.25    5.52       0.14
## Austria                  0.84                            1.93   43.29       0.07
## Bulgaria                 1.14                            0.54   39.54       1.03
## Croatia                  0.81                            2.40   18.06       0.02
## Cyprus                   1.48                            1.71   20.09       0.91
## Czechia                  0.76                            0.58   43.98       0.11
##          Sexual.violence Robbery
## Albania             5.38    3.42
## Austria            50.90   29.67
## Bulgaria            8.64   16.90
## Croatia            21.05   20.56
## Cyprus              1.94    6.28
## Czechia            14.65   13.51
##          Unlawful.acts.involving.controlled.drugs.or.precursors OverAllRecords
## Albania                                                  70.26          90.00
## Austria                                                 494.05         620.75
## Bulgaria                                                 78.14         145.93
## Croatia                                                 272.16         335.06
## Cyprus                                                  117.82         150.23
## Czechia                                                  45.25         118.84
```

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

#No:of Rows of new Dataset 'D'
print(nrow(D))
```

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

```
#No:of Columns of new Dataset 'D'
print(ncol(D))
```

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

```
#Structure of new Dataset 'D'
print(str(D))
```

```
## '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 39.5
4 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 ...
##  $ OverAllRecords                                          : num  90 621 146 335
150 ...
## NULL
```

Task 2: Analysis

```
#1. According to these data what were the 3 most common crimes in Ireland in 2019?
D1<-D[(row.names(D) %in% "Ireland"),]
D1<-t(D1)
D2<-data.frame(D1[rev(order(as.numeric(as.character(D1)))), ,drop=FALSE])
D3<-D2[-c(1:2),,drop = FALSE]

#The Three most common crimes in Ireland
row.names(D3[c(1:3),,drop = FALSE])
```

```
## [1] "Assault"         "Sexual.violence" "Robbery"
```

```
#2. What proportion of the overall crimes was due to Assault in Ireland in 2019?

D["Ireland","Assault"]/D["Ireland","OverAllRecords"]
```

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

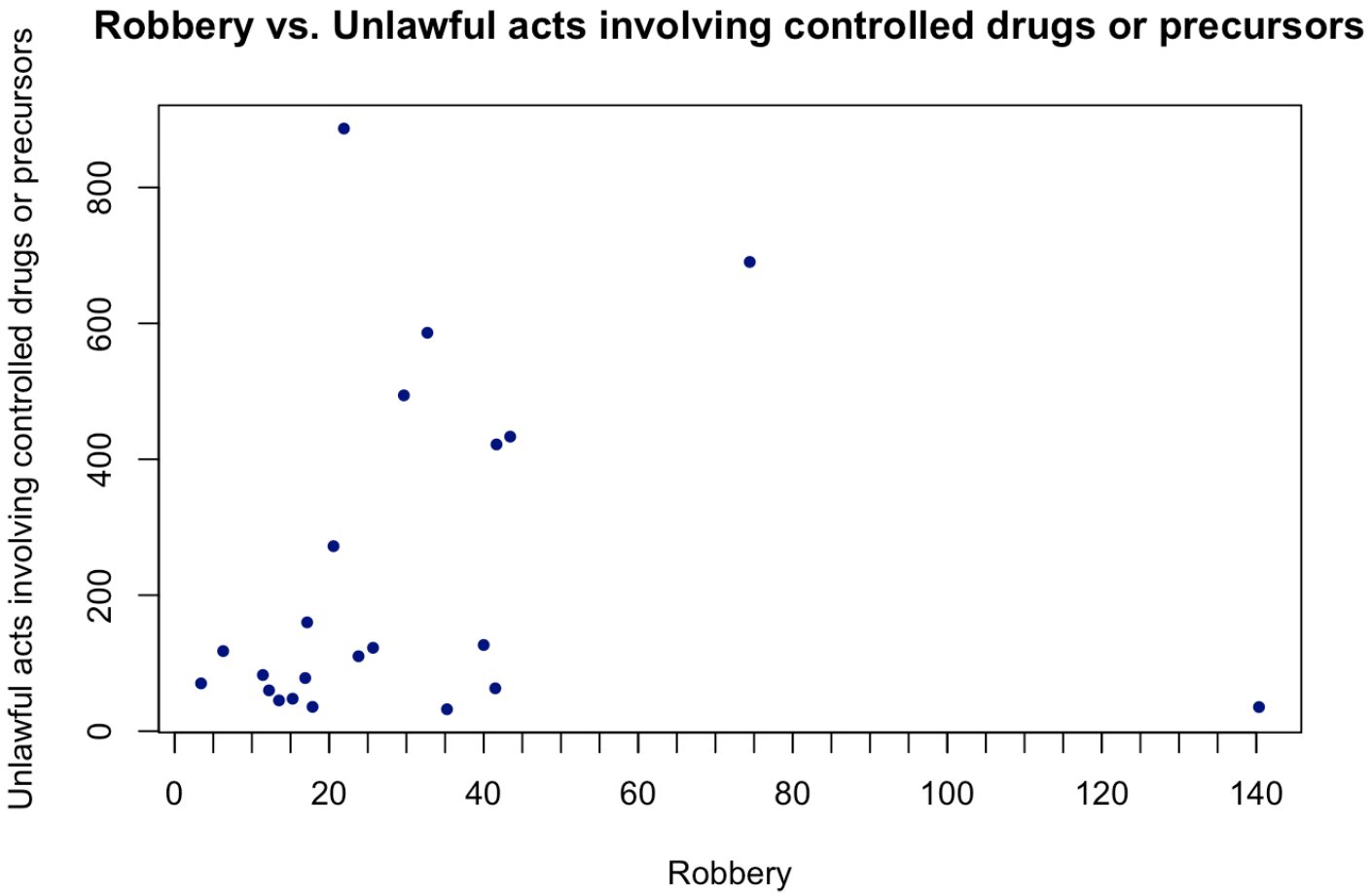
```
#3. Which country had the highest record of kidnapping in 2019 (per hundred thousa
nd inhabitants)?

HigestRecordOfKidnapping <- D[order(-D$Kidnapping),,drop = FALSE]
row.names(HigestRecordOfKidnapping)[1]
```

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

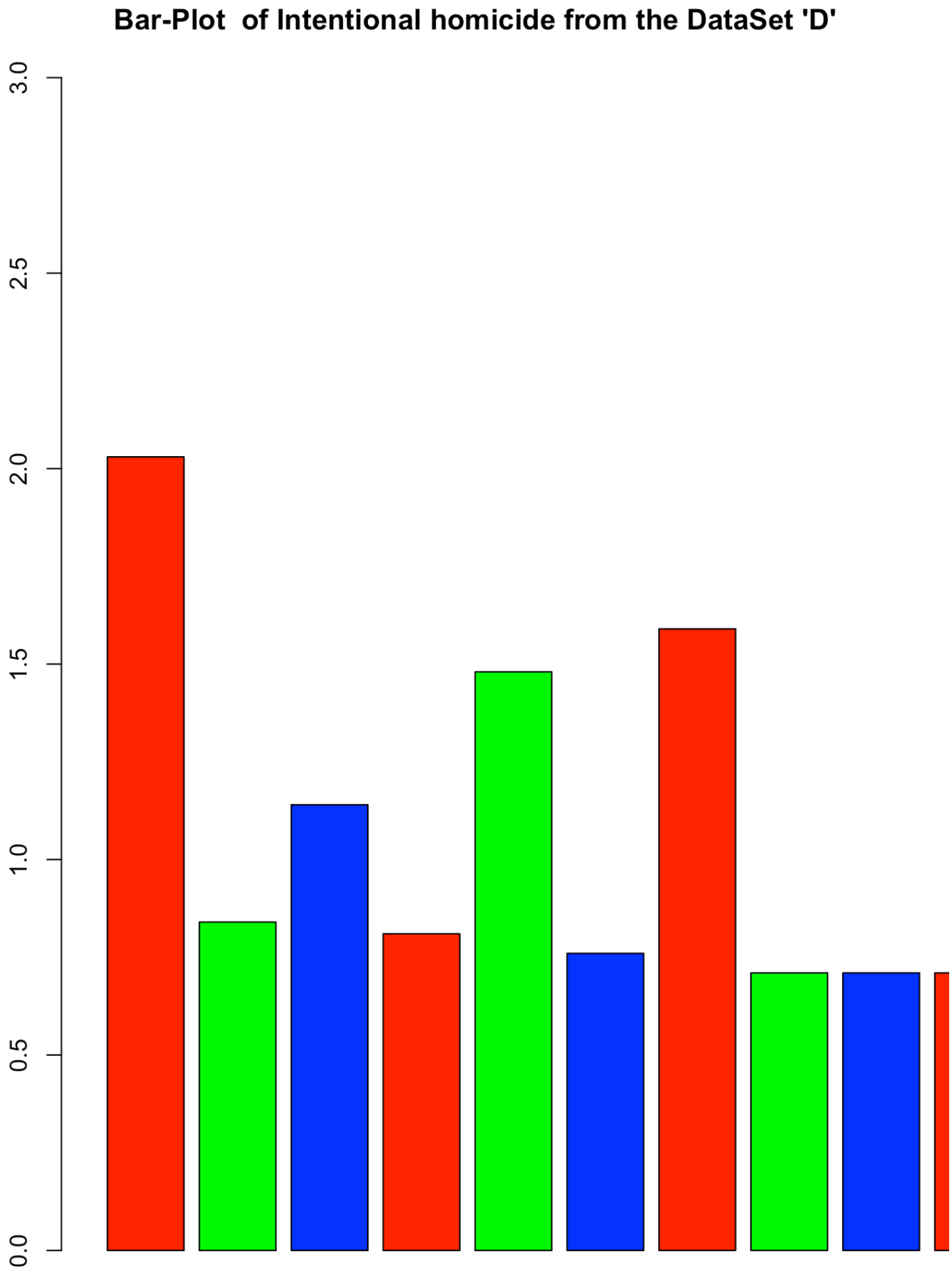
```
#4. Which country had the lowest overall record of offences in 2019 (per hundred t
housand inhabitants)?

LowestOfOverAllRecords<-D[order(-D$OverAllRecords), , drop = FALSE]

row.names(LowestOfOverAllRecords)[length(row.names(LowestOfOverAllRecords))]
```

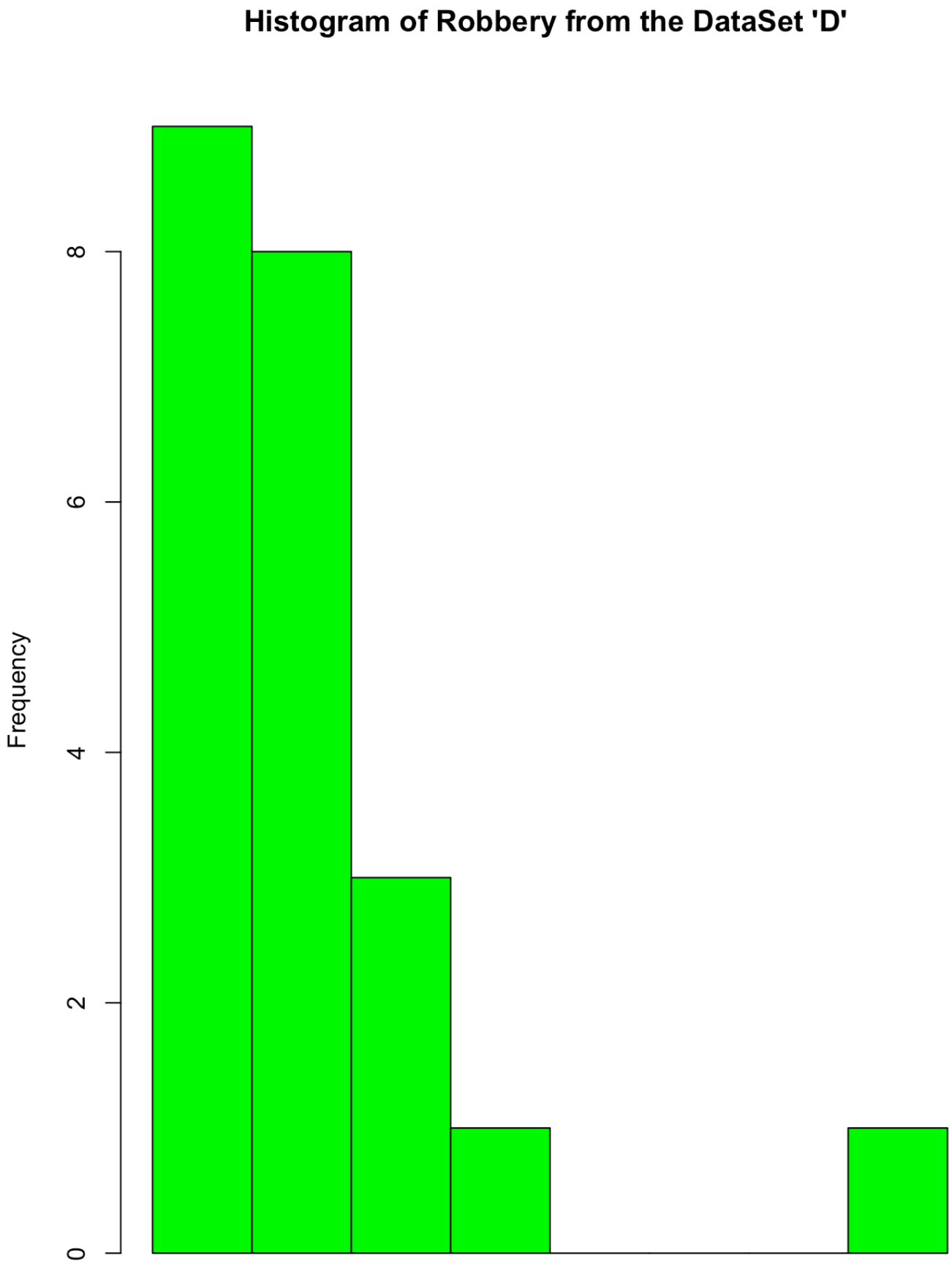
```
## [1] "Romania"
```

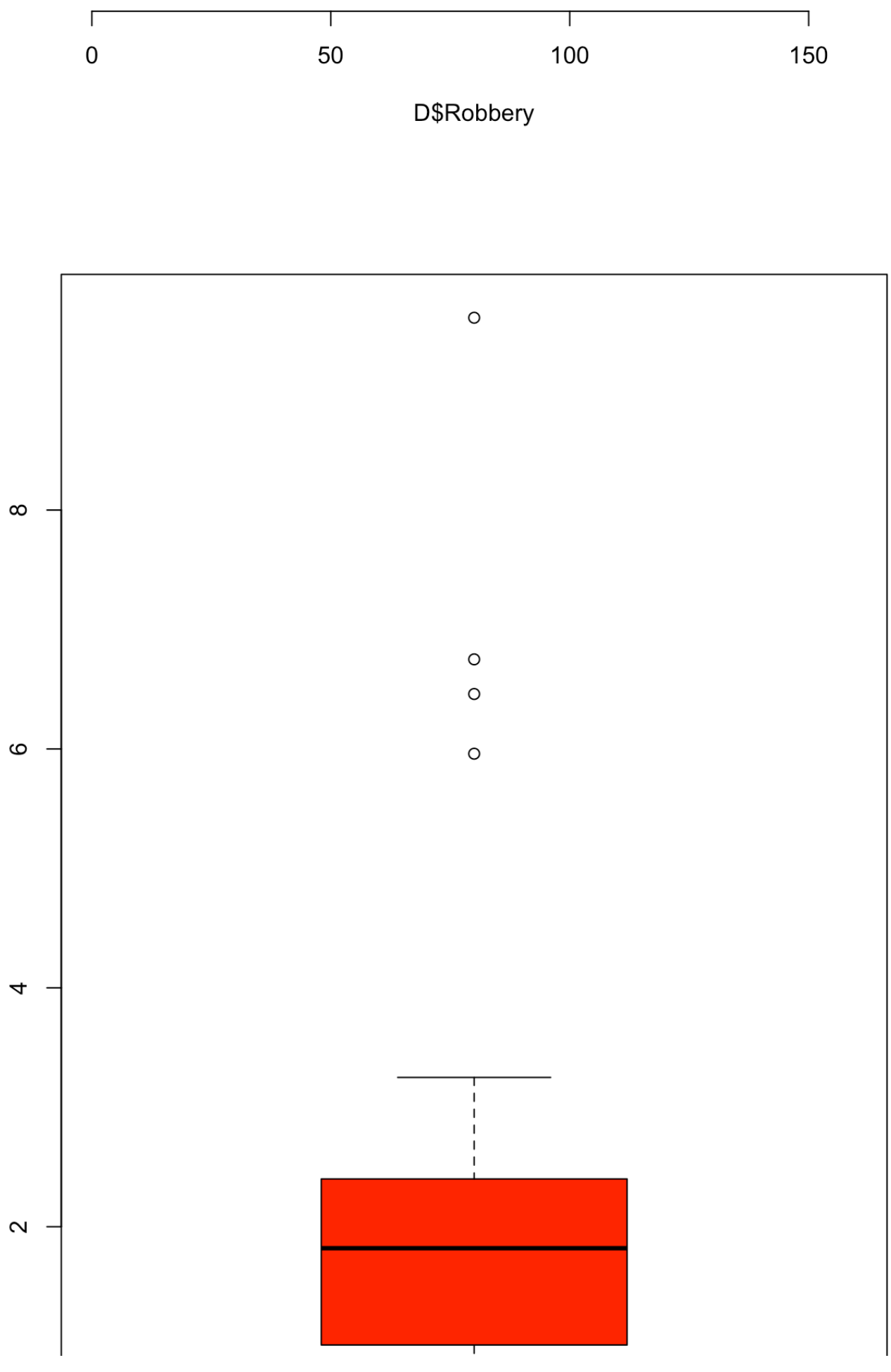
```
#5. Create a plot displaying the relationship between robbery and unlawful acts in
volving controlled drugs or precursors. Make the plot look "nice" i.e. change axis
labels etc.?

plot(D$Robbery,D$Unlawful.acts.involving.controlled.drugs.or.precursors,main="Robb
ery vs. Unlawful acts involving controlled drugs or precursors",xlab="Robbery", yl
ab="Unlawful acts involving controlled drugs or precursors",col="navy",pch=20)
axis(side = 1, at = seq(0, 140, by = 5), labels = FALSE)
```

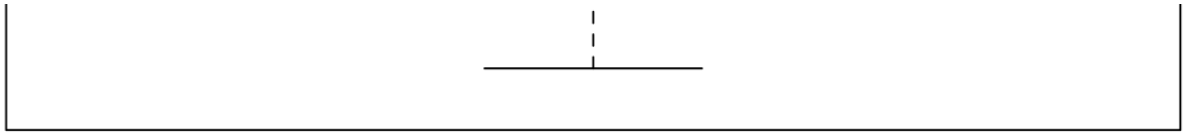


Task 3: Creativity




Histogram of Robbery from the DataSet 'D'
Frequency
0
2
4
6
8


0
50
100
150
D$Robbery
2
4
6
8



Note that the `echo = FALSE` parameter was added to the code chunk to prevent printing of the R code that generated the plot.