Crime Analysis in European Countries in Year 2019

Clementine Surya

Task 1: Manipulation

1. Load dataset

In the code below, I load the dataset of library `EurostatCrime2019.csv` :

```
data=read.csv("EurostatCrime2019.csv",row.names = 1)
```

2. Size and Structure of Dataset

The dataset contains 41 rows and 13 columns and the structure of this dataset is shown below:

```
nrow(data)
```

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

```
ncol(data)
```

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

```
str(data)
```

```
## '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 0.9
1 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 33.33 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 ...
```

3. Produce appropriate commands

(i) Remove Columns of Rape & Sexual.assault

For most countries, sexual violence figures are the sum of rape and sexual assault. Thus, I remove the columns Rape and Sexual.assault with below code:

```
data1=subset(data, select = -c(Rape, Sexual.assault))
```

(ii) Remove Columns of Theft and Burglary

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, I remove the columns involving theft and burglary with below code:

```
data2=subset(data1,select=-c(Burglary,Theft,Burglary.of.private.residential.premises,Theft.of.a.motorized.la
nd.vehicle))
```

(iii) Add column for overall offences

I add a column containing the overall record of offences for each country (per hundred thousand inhabitants) and the new column name is “totaloffense”:

```
#I use 'rowSums' to add every rows and obtain the total offense:
totaloffense  = rowSums(data2)
#I combine "totaloffense" column to the dataset with below code
data3=cbind(data2,totaloffense)
```

4. List the countries that contain any missing data

Below are list of countries that contain any missing data:

```
names(which(rowSums(is.na(data3))>0))
```

```
##  [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

In below code, I remove the countries with missing data from the dataframe. The new dataset is called “data4”. *dataset4* will be used in Analysis part

```
data4 = na.omit(data3)
```

6. Observations and variables on the new dataframe

After I remove the countries with missing data, the new dataframe has 22 rows and 8 columns (22 observations of 8 variables).

```
nrow(data4)
```

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

```
ncol(data4)
```

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

```
str(data4)
```

```
## '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 39.54 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 ...
##  $ totaloffense                                         : num  90 621 146 335 150 ...
##  - attr(*, "na.action")= 'omit' Named int [1:19] 3 4 9 10 11 13 16 17 22 27 ...
##   ..- attr(*, "names")= chr [1:19] "Belgium" "Bosnia and Herzegovina" "Denmark" "England and Wales" ...
```

```
#I can show the dimension using dim as well:
dim(data4)
```

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

Task 2: Analysis

1. Three most common crimes in Ireland in 2019

With below code, I generate **the 3 most common crimes in Ireland in 2019**:

```
#Firstly, I need to exclude "totaloffense" column
excl_offense = subset(data4, select = -c(totaloffense) )
#Secondly, I want to obtain Ireland's crimes only using below code
ireland_crime=subset(excl_offense,rownames(excl_offense)=="Ireland")
#Lastly, I want to know the 3 most common crimes with below code
colnames(sort(ireland_crime, decreasing = TRUE)[1:3])
```

```
## [1] "Unlawful.acts.involving.controlled.drugs.or.precursors"
## [2] "Assault"
## [3] "Sexual.violence"
```

2. Proportion of the overall crimes was due to Assault in Ireland in 2019

Below is the code for **proportion of the overall crimes that was due to Assault in Ireland in 2019**:

```
#I want to obtain Ireland's crimes dataset only with below code:
ireland_only<-subset(data4,rownames(data4)=="Ireland")
#To generate the proportion of the overall crimes that was due to Assault in Ireland, I use below code
ireland_only$Assault/ireland_only$totaloffense
```

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

3. Country with the highest record of kidnapping in 2019

With below code, I generate **the highest record of kidnapping in 2019** (per hundred thousand inhabitants):

```
highest_kidnapping<-rownames(data4[which.max(data4$Kidnapping),])
highest_kidnapping
```

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

4. Country with the lowest overall record of offences in 2019

And **the lowest overall record of offences** in 2019 (per hundred thousand inhabitants):

```
highest_offence<- rownames(data4[which.min(data4$totaloffense),])
highest_offence
```

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

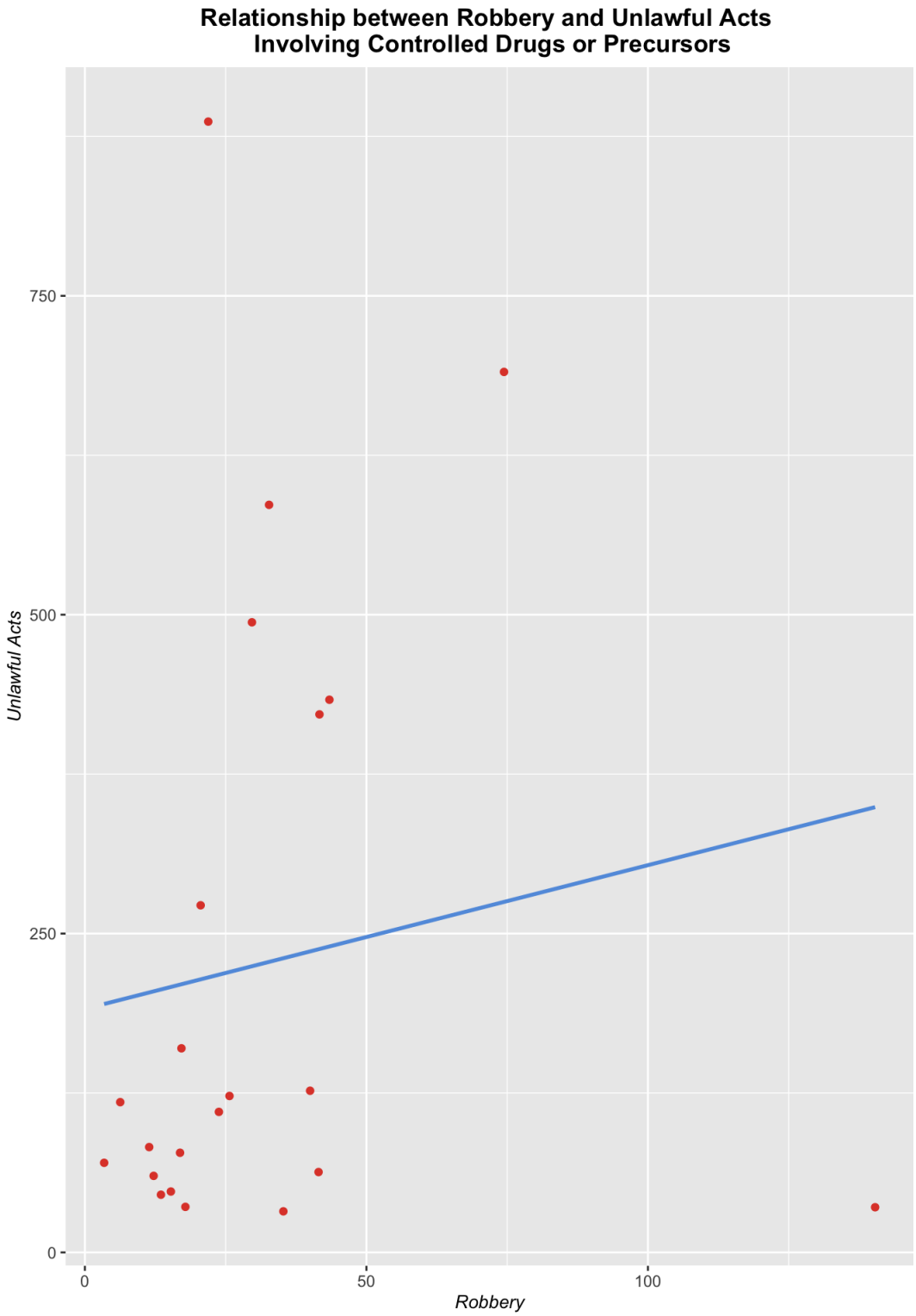
5. Plot displaying the relationship between robbery and unlawful acts involving controlled drugs or precursors

Here, I show a plot displaying the relationship between robbery and unlawful acts. First, I load ggplot2 `library` function

```
library("ggplot2")
```

With below code, I generate a scatterplot and regression line

```
ggplot(data = data4, aes(x = Robbery, y = Unlawful.acts.involving.controlled.drugs.or.precursors)) +
        geom_point(color = "#DF4635") +
        geom_smooth(method = "lm", se = FALSE, col = "#629CDF") +
        labs(title = "Relationship between Robbery and Unlawful Acts \n Involving Controlled Drugs or Precur
sors",
             x = "Robbery",y = "Unlawful Acts")+
        theme(plot.title = element_text(face="bold",hjust = 0.5),
        axis.title.x = element_text(face="italic", size = 10),
        axis.title.y = element_text(face="italic",size= 10))
```



I would like to know the correlation between Robbery and Unlawful Acts. Thus, I generate below code and obtain a weak relationship between Robbery and Unlawful Acts with correlation of **0.1328638**.

```
cor(data4$Robbery, data4$Unlawful.acts.involving.controlled.drugs.or.precursors)
```

```
## [1] 0.1328638
```

Task 3: Creativity

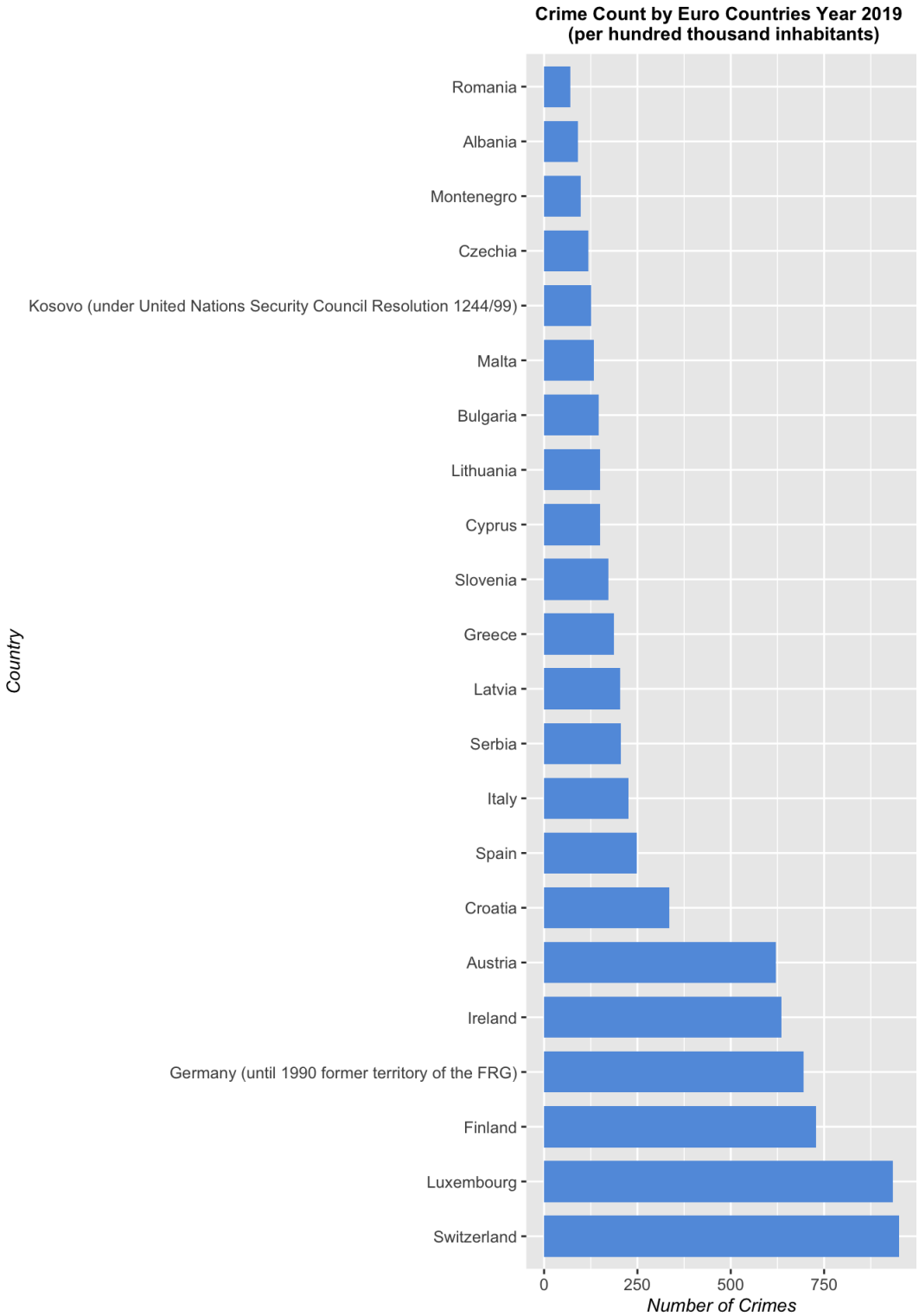
For creativity part, I create 3 plots:

Plot 1

```
#I create a data frame called "df2" that consist of 22 observations and 2 variables: Country and Total Offen
se with below code:
df <- data4
df$country_name <- row.names(df)
df2=data.frame(df$country_name,df$totaloffense)
colnames(df2) <- c("Country", "TotalOffense")
#The structure of "df2" shown below:
str(df2)
```

```
## 'data.frame':    22 obs. of  2 variables:
##  $ Country     : chr  "Albania" "Austria" "Bulgaria" "Croatia" ...
##  $ TotalOffense: num  90 621 146 335 150 ...
```

```
#I create a ascending barplot with below code:
plot1=ggplot(data = df2, aes(x = reorder(Country, -TotalOffense),
                             y = TotalOffense, fill = Country)) +
  geom_bar(stat = "identity", width = 0.75,fill= "#629CDF")+
  coord_flip() +
  labs(x = "Country",
       y = "Number of Crimes",
       title = "Crime Count by Euro Countries Year 2019 \n (per hundred thousand inhabitants)") +
  theme(plot.title = element_text(face="bold",hjust = 0.5, size= 10),
        axis.title.x = element_text(face="italic", size = 10),
        axis.title.y = element_text(face="italic", size = 10))
plot1
```



The first plot shows overall record of offences in 2019 by the Euro Countries. It has X label as Number of Crimes (per hundred thousand inhabitants) and Y label as Countries.

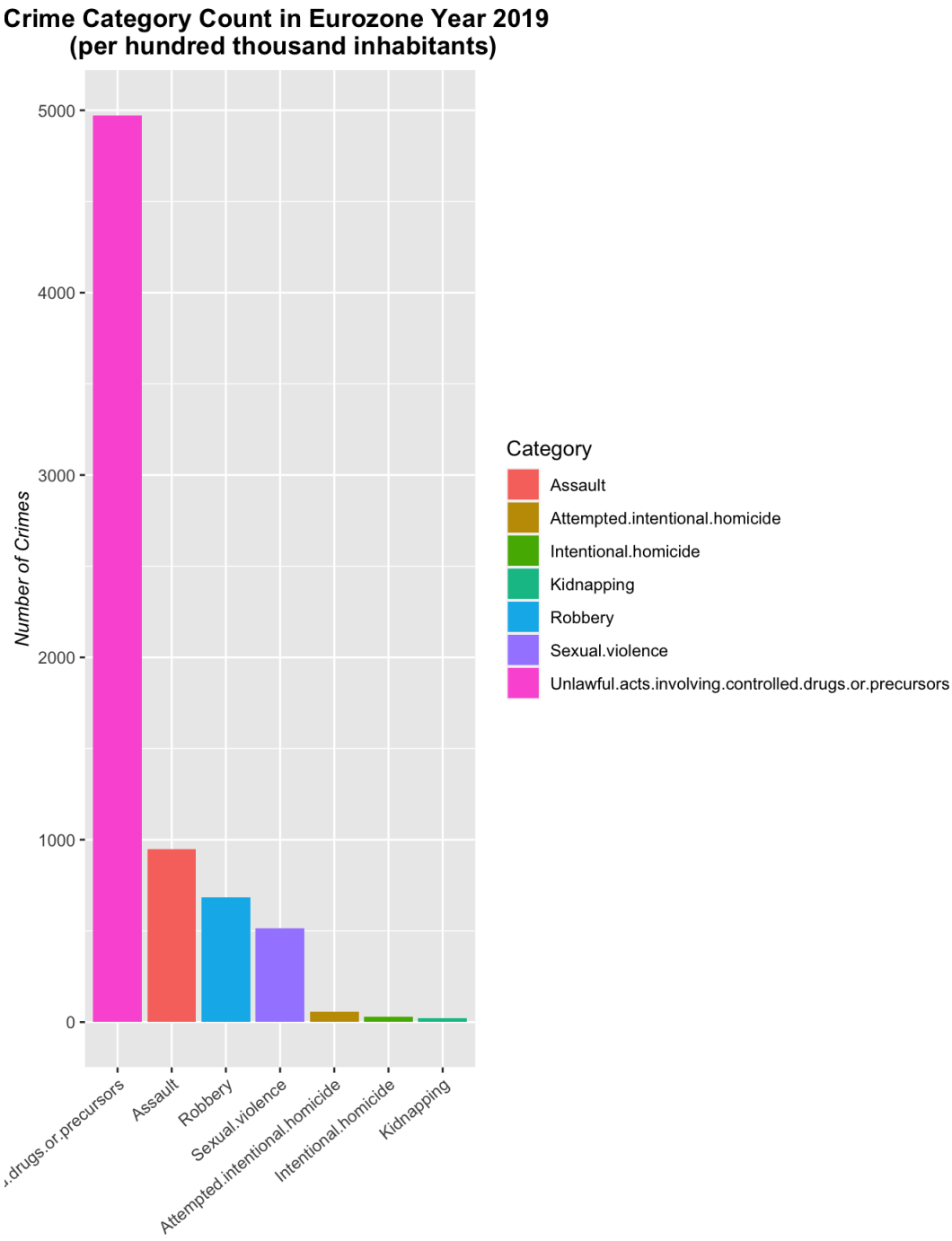
From the graph, we can conclude that the top 3 highest record of offences were in **1) Switzerland, 2) Luxembourg, and 3) Finland**. And the 3 lowest record of offences were in **1) Romania, 2) Albania, and 3) Montenegro**.

Plot 2

```
#I create a data frame called "df5" that consist of 7 observations of 2 variables: Crime Category and TotalO
ffense with below code:
df3<-excl_offense
tdf3<-as.data.frame(t(as.matrix(df3)))
totaloffense1<-rowSums(tdf3)
df4<-cbind(tdf3,totaloffense1)
df4$crime_name <- row.names(df4)
df5=data.frame(df4$crime_name,df4$totaloffense1)
colnames(df5) <- c("Category", "TotalOffense")
#The structure of df5 is shown below:
str(df5)
```

```
## 'data.frame':    7 obs. of  2 variables:
##  $ Category    : chr  "Intentional.homicide" "Attempted.intentional.homicide" "Assault" "Kidnapping" ...
##  $ TotalOffense: num  29.4 56.1 946.7 22.6 513.4 ...
```

```
#I create a desceding barplot with below code:
plot2= ggplot(data = df5, aes(x = reorder(Category, -TotalOffense),
                              y = TotalOffense, fill = Category)) +
  geom_bar(stat = "identity") +
  labs(x = "Category of Crime", y = "Number of Crimes", title = "Crime Category Count in Eurozone Year 2019
\n (per hundred thousand inhabitants)") +
  theme(plot.title = element_text(face="bold",hjust = 0.5),
        axis.title.x = element_text(face="italic", size = 10),
        axis.title.y = element_text(face="italic", size = 10))
plot2 + theme(axis.text.x = element_text(angle = 40, vjust = 1, hjust=1))
```



The second plot shows the overall record of *Crime Category* in Europe zones in 2019 where the X label is the Crime's Category and Y label is the Number of Crimes (per hundred thousand inhabitants). From the graph, we can conclude that the highest record of offences in Euro zones on 2019 was **Unlawful Acts Involving Controlled Drugs or Precursors** while the lowest record of offences was **Kidnapping**

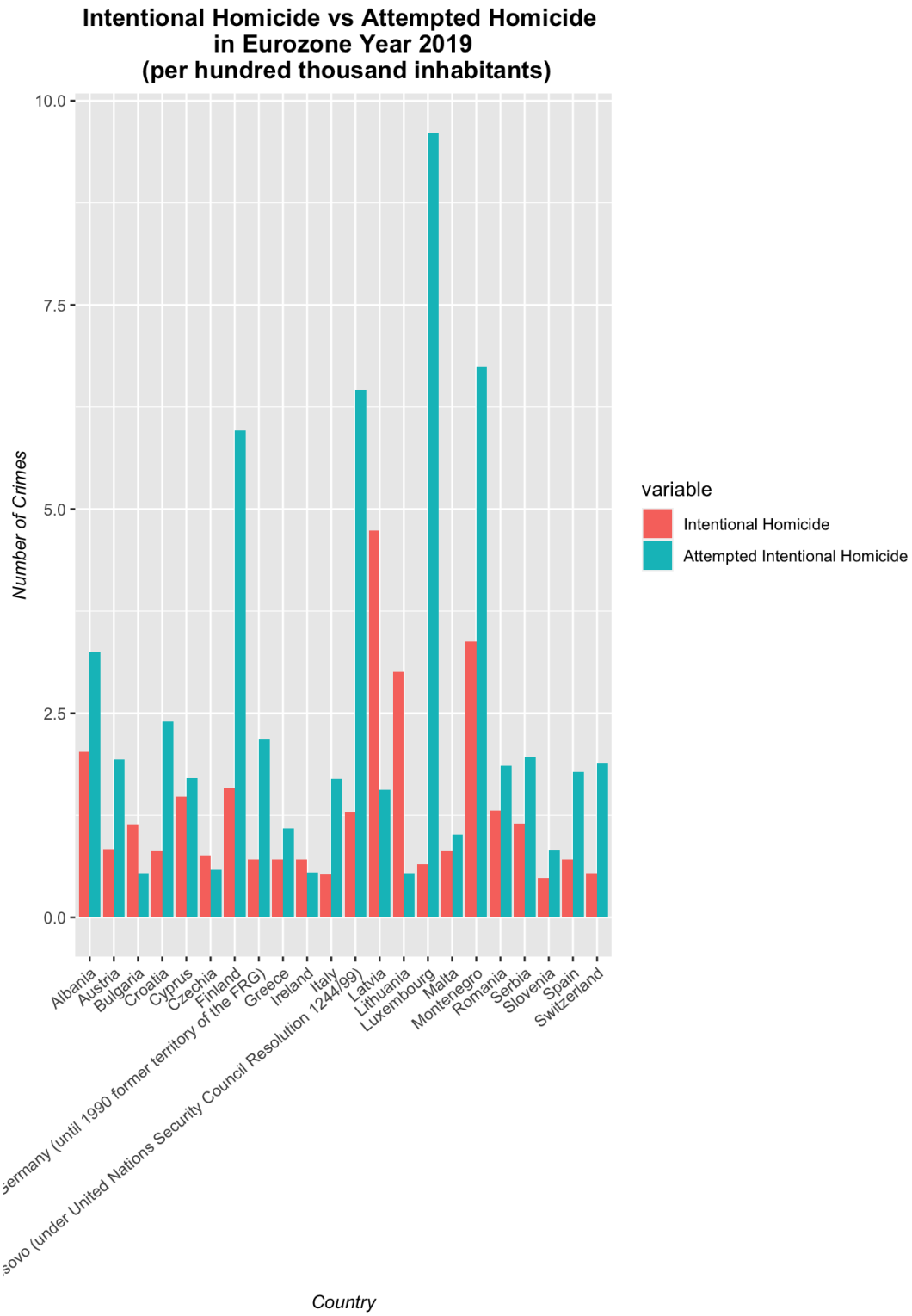
Plot 3

```
#I load library reshape2 because I want to use the melt function
library(reshape2)
```

```
#I create a data frame called "df7melt" that consist of 44 observations of 3 variables: Country, Homicide Va
riable and Number of Homicide Crime with below code:
df6=subset(data4,select=-c(Assault,Kidnapping,Sexual.violence,Robbery,Unlawful.acts.involving.controlled.dru
gs.or.precursors,totaloffense))
df6$countryname1 <- row.names(df6)
df7= data.frame(df6$countryname1,df6$Intentional.homicide,df6$Attempted.intentional.homicide)
colnames(df7) <- c("CountryName", "Intentional Homicide","Attempted Intentional Homicide")
df7melt<-melt(df7,id="CountryName")
#The structure of df7melt shown below:
str(df7melt)
```

```
## 'data.frame':    44 obs. of  3 variables:
##  $ CountryName: chr  "Albania" "Austria" "Bulgaria" "Croatia" ...
##  $ variable   : Factor w/ 2 levels "Intentional Homicide",..: 1 1 1 1 1 1 1 1 1 1 ...
##  $ value      : num  2.03 0.84 1.14 0.81 1.48 0.76 1.59 0.71 0.71 0.71 ...
```

```
#I create a comparison barplot with below code:
plot3=ggplot(data = df7melt, aes(x = CountryName,
                       y = value, fill = variable)) +
  geom_col(position = "dodge") +
  labs(x = "Country", y = "Number of Crimes", title = "Intentional Homicide vs Attempted Homicide \n in Euro
zone Year 2019 \n (per hundred thousand inhabitants)") +
  theme(plot.title = element_text(face="bold",hjust = 0.5),
        axis.title.x = element_text(face="italic", size = 10),
        axis.title.y = element_text(face="italic", size = 10))
plot3 + theme(axis.text.x = element_text(angle = 40, vjust = 1, hjust=1))
```



In the third plot, I would like to compare the number of crime between Intentional Homicide and Attempted Intentional Homicide Crimes in Eurozones in 2019. The X label is Country in Eurozone and Y label is Number of Crimes (per hundred thousand inhabitants). We can see in the graph that in **Most** Euro countries the crime of Attempted Intentional Homicide is more than Intentional Homicide. Definition on Attempted Intentional Homicide and Intentional Homicide are shown in the reference section below.

Reference:

Link: 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)

Intentional Homicide is defines as unlawful death inflicted upon a person with the intent to cause death or serious injury. It is also include murder, honour killing, serious assault leading to death, death as a result of terrorist activities, dowry-related killings, femicide, infanticide, voluntary manslaughter, extrajudicial killings, killings caused by excessive use of force by law enforcement/state officials.

Attempted Intentional Homicide is defines as attempt to inflict unlawful death upon a person with the intent to cause death or serious injury. Data on attempted intentional homicide should also include attempted murder, attempt to inflict death as a result of terrorist activities, attempted infanticide, attempted femicide and exclude conspiracy to procure or commit illegal feticide.