

# Assignment 1

Michael Whelan 17338833

01/10/2022

The dataset EurostatCrime2019.csv records offences (values per hundred thousand inhabitants) by offence category in 41 European Countries in 2019. 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).

```
#Calling libraries to perform different functions
library(dplyr)
library(tidyr)

#Setting Working Directory
#setwd("/Users/Michael/Desktop/GitHub/Assignments/European Crime Stats 2019")
```

## Task 1: Manipulation

1. Load the dataset EurostatCrime2019.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]. [1 marks]

```
#Reading the csv file into R and turning the countries into row names so it is not treated as a variable
df <- read.csv("EurostatCrime2019 copy.csv", header = T, row.names = 1)
print(head(df)) #Getting an understanding for the data
```

```
##              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 Rape Sexual.assault
## Albania          5.52      0.14          5.38  2.69          2.69
## Austria         43.29      0.07         50.90 18.92         26.64
## Belgium        556.36      NA          77.45 33.33         44.12
## Bosnia and Herzegovina    NA      NA          NA  NA          NA
## Bulgaria        39.54      1.03          8.64  1.87          NA
## Croatia         18.06      0.02         21.05 11.58          8.61
##              Robbery Burglary
## Albania          3.42      NA
## Austria         29.67  613.22
## Belgium        140.14  565.92
## Bosnia and Herzegovina    NA      NA
```

```
## Bulgaria          16.90    79.81
## Croatia           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
```

To explain the data, take Albania's Intentional Homicides which is 2.03. This means for every 100,000 people there are 2.03 intentional homicides.

2. What is the size (number of rows and columns) and the structure of this dataset? [0.5 marks]

```
str(df) #Analysing the structure of the data
```

```
## 'data.frame':   41 obs. of  13 variables:
## $ Intentional.homicide      : num  2.03 0.84 1.27 NA 1.14 0.81 1.48 0.7
## $ Attempted.intentional.homicide : num  3.25 1.93 8.87 NA 0.54 2.4 1.71 0.58
## $ 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
## $ 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 ...
```

This data frame is a 2-dimensional structure made up of rows and columns. There are 13 variables (Columns) and 41 observations (Rows). Each object in the table is a `num` which means they are numeric. This means they can be real numbers, integers, floating point numbers etc.

3. Produce appropriate commands to do the following actions:

- For most countries sexual violence figures are the sum of rape and sexual assault. Remove the columns `Rape` and `Sexual.assault`. [0.5 marks]

```
colnames(df) #Checking what numbers the columns are
```

```
## [1] "Intentional.homicide"  
## [2] "Attempted.intentional.homicide"  
## [3] "Assault"  
## [4] "Kidnapping"  
## [5] "Sexual.violence"  
## [6] "Rape"  
## [7] "Sexual.assault"  
## [8] "Robbery"  
## [9] "Burglary"  
## [10] "Burglary.of.private.residential.premises"  
## [11] "Theft"  
## [12] "Theft.of.a.motorized.land.vehicle"  
## [13] "Unlawful.acts.involving.controlled.drugs.or.precursors"
```

```
df <- df[-c(6,7)] #Taking out column 6,7 i.e Rape and Sexual.assault  
colnames(df) #Checking Answer
```

```
## [1] "Intentional.homicide"  
## [2] "Attempted.intentional.homicide"  
## [3] "Assault"  
## [4] "Kidnapping"  
## [5] "Sexual.violence"  
## [6] "Robbery"  
## [7] "Burglary"  
## [8] "Burglary.of.private.residential.premises"  
## [9] "Theft"  
## [10] "Theft.of.a.motorized.land.vehicle"  
## [11] "Unlawful.acts.involving.controlled.drugs.or.precursors"
```

- 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, remove the columns involving theft and burglary:

```
colnames(df) #Checking what numbers the columns are
```

```
## [1] "Intentional.homicide"  
## [2] "Attempted.intentional.homicide"  
## [3] "Assault"  
## [4] "Kidnapping"  
## [5] "Sexual.violence"  
## [6] "Robbery"  
## [7] "Burglary"  
## [8] "Burglary.of.private.residential.premises"  
## [9] "Theft"  
## [10] "Theft.of.a.motorized.land.vehicle"  
## [11] "Unlawful.acts.involving.controlled.drugs.or.precursors"
```

```
df <- df[-c(7:10)] #Removing columns 7-10 i.e any columns to do with theft and burglary
colnames(df) #Checking Answer
```

```
## [1] "Intentional.homicide"
## [2] "Attempted.intentional.homicide"
## [3] "Assault"
## [4] "Kidnapping"
## [5] "Sexual.violence"
## [6] "Robbery"
## [7] "Unlawful.acts.involving.controlled.drugs.or.precursors"
```

- Add a column containing the overall record of offences for each country (per hundred thousand inhabitants)? [1 marks]

```
df$"Overall Record of Offences" <- rowSums(df, na.rm = F) #Summing the rows together in a new column ca
head(df) #Checking Answer
```

```
##              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
##              Overall Record of Offences
## Albania                                90.00
## Austria                               620.75
## Belgium                               NA
## Bosnia and Herzegovina                  NA
## Bulgaria                              145.93
## Croatia                              335.06
```

```
#Double Checking Answer e.g Albania
df[1,] #Albania
```

```
##              Intentional.homicide Attempted.intentional.homicide Assault Kidnapping
## Albania                2.03                      3.25         5.52         0.14
```

```
## Sexual.violence Robbery
## Albania 5.38 3.42
## Unlawful.acts.involving.controlled.drugs.or.precursors
## Albania 70.26
## Overall Record of Offences
## Albania 90
```

```
#Adding all of Albania's rows = 90 which is the first entry of Overall Record of offences
c(2.03 + 3.25 + 5.52 + 0.14 + 5.38 + 3.42 + 70.26)
```

```
## [1] 90
```

4. Work with the dataset you just created, and list the countries that contain any missing data. [1.5 marks]

```
#Making a new variable dfna that is looking for any rows with NA. If you add a row with even 1 NA in it
```

```
dfna <- df[rowSums(is.na(df)) > 0,]
NA_Countries <- c(rownames(dfna)) #Containing the rownames of dfna in a new variable
NA_Countries
```

```
## [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. [1 marks]

```
df<-df[complete.cases(df),] #Returning a vector that has no missing vales
head(df) #Checking answer
```

```
## 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
## Albania 70.26
## Austria 494.05
## Bulgaria 78.14
## Croatia 272.16
```

```
## Cyprus 117.82
## Czechia 45.25
## Overall Record of Offences
## Albania 90.00
## Austria 620.75
## Bulgaria 145.93
## Croatia 335.06
## Cyprus 150.23
## Czechia 118.84
```

There are now no NA values.

6. How many observations and variables are in this new dataframe? [0.5 marks]

```
str(df) #Checking structure of new data frame
```

```
## '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
## $ Attempted.intentional.homicide : num 3.25 1.93 0.54 2.4 1.71 0.58 5.96 2.
## $ 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
## $ 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 ...
## $ Overall Record of Offences : num 90 621 146 335 150 ...
```

Now there are 22 Observations (Rows) and 8 Variables (Columns) in the data frame now.

## Task 2: Analysis

1. According to these data what were the 3 most common crimes in Ireland in 2019? [2 marks]

```
ire <- unlist(df[10,]) #Gives me back numbers I can manipulate i.e atomic components
tail(sort(ire), 4) #4 because Overall Record of offences dosent count and I can sort now
```

```
## Sexual.violence
## 67.86
## Assault
## 102.18
## Unlawful.acts.involving.controlled.drugs.or.precursors
## 421.84
## Overall Record of Offences
## 636.51
```

The 3 most common crimes in Ireland in 2019 were:

- Unlawful Acts Involving Controlled Drugs or Precursors
- Assault
- Sexual Assault

2. What proportion of the overall crimes was due to Assault in Ireland in 2019? [1.5 marks]

```
print(ire) #Analysing the column indexes
```

```
##              Intentional.homicide
##                      0.71
##      Attempted.intentional.homicide
##                      0.55
##                      Assault
##                      102.18
##                      Kidnapping
##                      1.71
##                      Sexual.violence
##                      67.86
##                      Robbery
##                      41.66
## Unlawful.acts.involving.controlled.drugs.or.precursors
##                      421.84
##                      Overall Record of Offences
##                      636.51
```

```
ire[3] / ire[8] #Dividing column 3 and 8 in ire
```

```
## Assault
## 0.1605316
```

The proportion of overall crimes due to Assault in Ireland was ~ 16%

3. Which country had the highest record of kidnapping in 2019 (per hundred thousand inhabitants)? [1 marks]

```
max <- which(df$Kidnapping == max(df$Kidnapping)) #Which row in Kidnapping is the highest value
rownames(df[max,]) #Displaying row 15 and no columns
```

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

Luxembourg had the highest record of kidnapping in 2019 (per hundred thousand inhabitants) at 7.17

4. Which country had the lowest overall record of offences in 2019 (per hundred thousand inhabitants)? [1 marks]

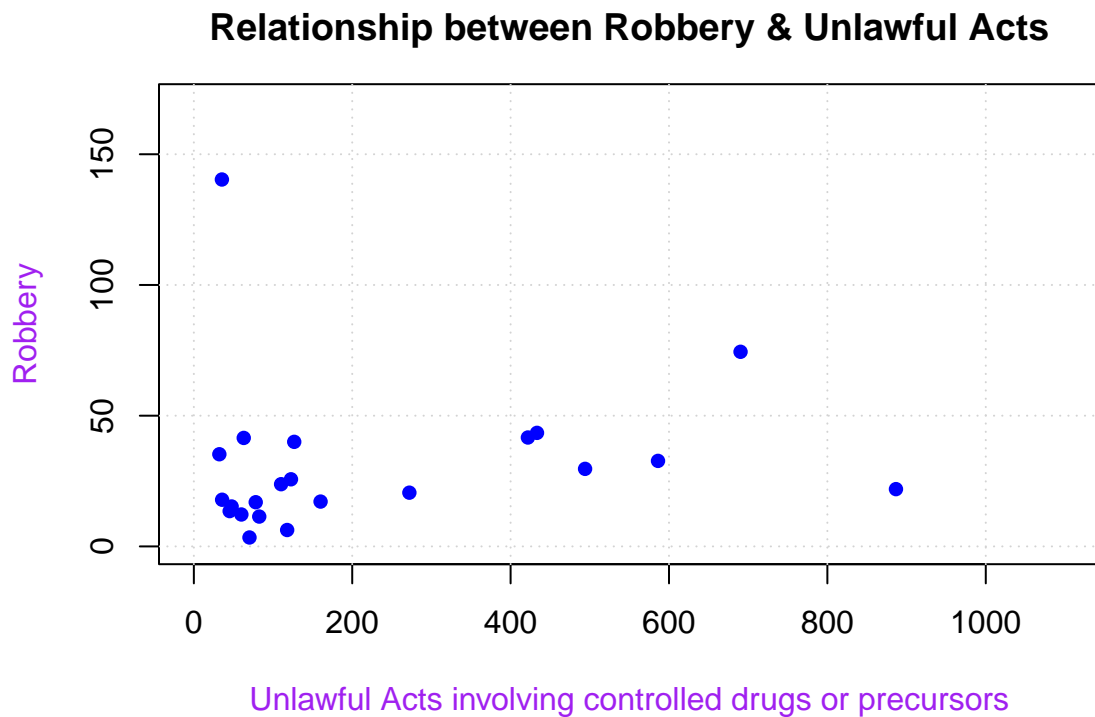
```
min <- which(df$`Overall Record of Offences` == min(df$`Overall Record of Offences`) ) #Which row in Overall Record of Offences is the lowest value
rownames(df[min,]) #Displaying the row with no columns
```

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

Romania had the lowest overall offences in 2019 (per hundred thousand inhabitants) at 70.06

5. Create a plot displaying the relationship between robbery and unlawful acts involving controlled drugs or precursors. Make the plot look “nice” i.e. change axis labels etc. [2 marks]

```
#Plotting the relationship between Robbery and Unlawful Acts ... and designing the graph
par(mar=c(7,5,3,3))
plot(x = df$Unlawful.acts.involving.controlled.drugs.or.precursors, y = df$Robbery, xlab = "Unlawful Acts", ylab = "Robbery",
grid())
```



### Task 3: Creativity

Do something interesting with these data (either the original dataset or the modified one)! Create a nice plot which shows something we have not discovered above already and outline your findings.

I will start this analysis by retrieving the original csv file. During the previous exercises rather than renaming df I kept overwriting it so I am starting again with the original csv file.

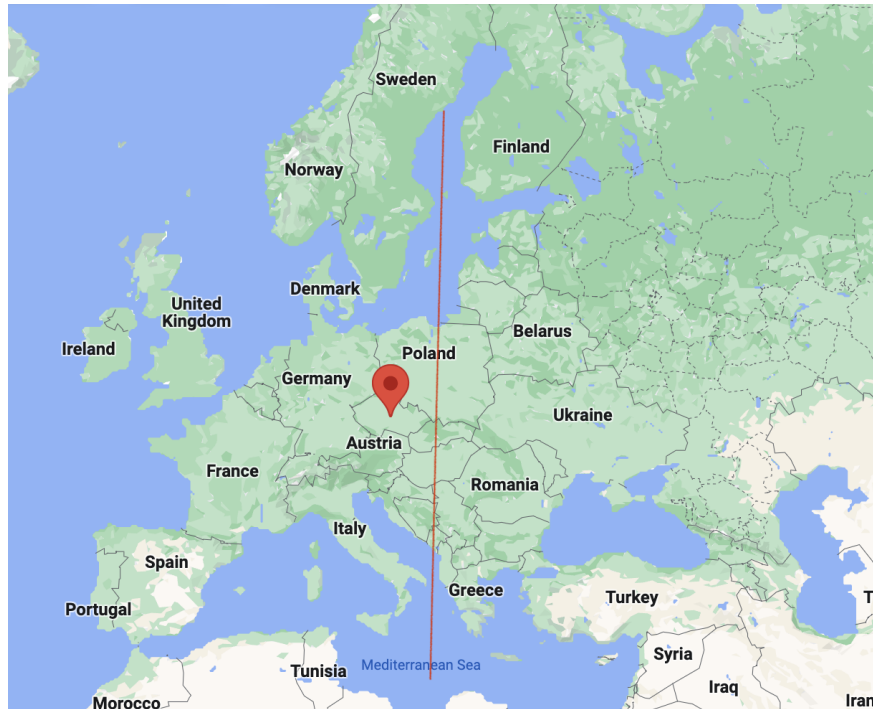
```
df2 <- read.csv("EStatCrime2019_V2 copy.csv", header = T, row.names = 1) #Reading the same csv file but
head(df2) #Analysing the data
```

```
##              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 Rape Sexual.assault
## Albania             5.52         0.14          5.38  2.69         2.69
## Austria            43.29         0.07         50.90 18.92        26.64
## Belgium           556.36         NA          77.45 33.33        44.12
## Bosnia and Herzegovina  NA         NA           NA   NA         NA
## Bulgaria          39.54         1.03          8.64  1.87         NA
```



## Croatia	18.06	0.02	21.05	11.58	8.61
##	Robbery	Burglary			
## Albania	3.42	NA			
## Austria	29.67	613.22			
## Belgium	140.14	565.92			
## Bosnia and Herzegovina	NA	NA			
## Bulgaria	16.90	79.81			
## Croatia	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		

I thought to look into which has a higher crime rate in Europe the East or the West. If we divide Europe into two from the space between Sweden and Finland down to the gap between Italy and Greece.



Let's divide our data set into 2 parts East and West

```
df2$"Overall Record of Offences" <- rowSums(df2, na.rm = T ) #Adding up the row names and putting them
df2$'Location' <- c('E','W','W','E','W','W','E','W','W','W','E','E','W','W','E','E','W','W','W','E','E')

df2 %>% select(Location) #Analysing If I input the location values correctly
```

	Location
##	
## Albania	E
## Austria	W
## Belgium	W
## Bosnia and Herzegovina	E
## Bulgaria	W
## Croatia	W
## Cyprus	E
## Czechia	W
## Denmark	W
## England and Wales	W
## Estonia	E
## Finland	E
## France	W
## Germany (until 1990 former territory of the FRG)	W
## Greece	E
## Hungary	E
## Iceland	W
## Ireland	W
## Italy	W

```
## Kosovo (under United Nations Security Council Resolution 1244/99)      E
## Latvia                                                                E
## Liechtenstein                                                         W
## Lithuania                                                             E
## Luxembourg                                                            W
## Malta                                                                 W
## Montenegro                                                            E
## Netherlands                                                           W
## North Macedonia                                                       E
## Northern Ireland (UK)                                                W
## Norway                                                                W
## Poland                                                                E
## Portugal                                                              W
## Romania                                                               E
## Scotland                                                              W
## Serbia                                                                E
## Slovakia                                                              E
## Slovenia                                                              E
## Spain                                                                 W
## Sweden                                                                W
## Switzerland                                                          W
## Turkey                                                                E
```

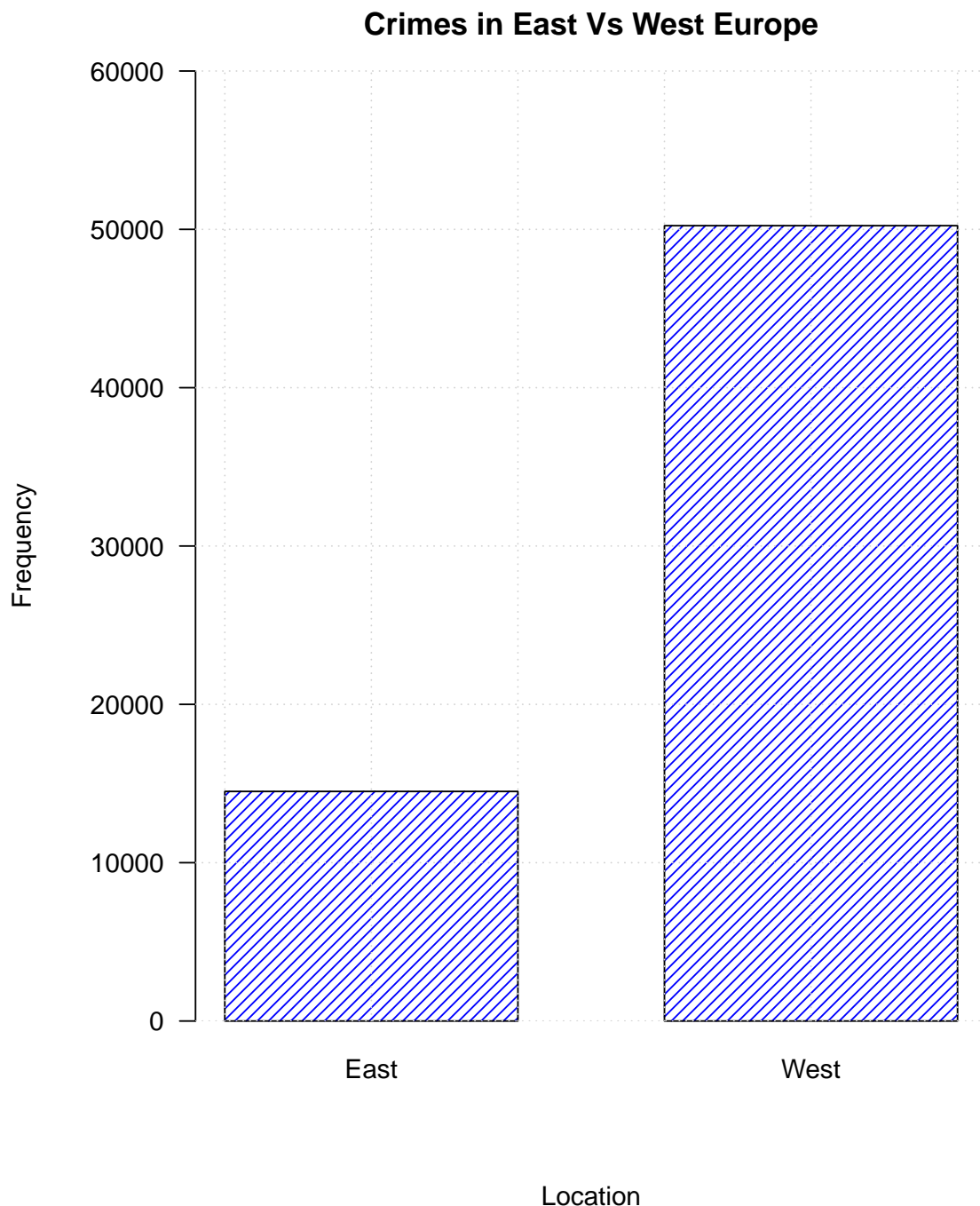
```
E <- subset(df2$`Overall Record of Offences`, df2$Location == 'E') #Taking out the eastern countries
sum(E) #Summing their overall record of offences
```

```
## [1] 14500.48
```

```
W <- subset(df2$`Overall Record of Offences`, df2$Location == 'W') #Taking out the western countries
sum(W) #Summing their overall record of offences
```

```
## [1] 50232.57
```

```
par(mar=c(7,8,3,2))
bar <- barplot(c("East" = sum(E), "West" = sum(W)), space = 0.5, density = 20, angle = 45, col = "blue"
grid()
```



According to the barplot, a lot more crimes happen in the West. This does not mean the West is more dangerous as this information may not be very accurate there are a few factors to take into account.

The first being there is more countries in the West than East therefore more people. In this data set ~ 43% of countries are Eastern meaning ~ 53% are Western.

Another factor could be how I divided the map and I feel this is the biggest factor. It could be inaccurate for example, Poland was split in the middle so it could have been either East or West depending on who you ask. I took it as East because the capital city Warsaw was more Eastern. It was difficult to decide where

this line went because I tried to divide as many countries in 2 as possible while keeping the line as vertical as possible too.

A possible factor could be what countries identify as Eastern, maybe it is not a matter of land mass but public opinion. Maybe some countries feel half of them are Eastern and half are Western which I did not take into account here i.e each country was either fully Western or fully Eastern

There are also some countries that are not in this data set such as Austria, Andorra and Ukraine which could have an affect too.

Based on this data alone it does make sense. Look at the top 6 highest rates in Europe below. All but Finland are Western (according to my map).

```
#Looking at the first 6 overall rates
overall <- c(order(df2$`Overall Record of Offences`[1:41], decreasing = T))
df3 <- data.frame(df2[overall,0], sort(df2$"Overall Record of Offences", decreasing = T))
head(df3)
```

```
##                sort.df2..Overall.Record.of.Offences...decreasing...T.
## Sweden                                5954.69
## Denmark                               5168.83
## Belgium                               4447.04
## Finland                               3401.86
## Switzerland                           3362.60
## Liechtenstein                         3043.42
```