STAT40730 - Assignment 1

Daniel Williams - 21203054

14/10/2021

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

```
EurostatCrime2019 <- read.csv("EurostatCrime2019.csv", header = TRUE, row.names = 1)
head(EurostatCrime2019, 3)
```

	Intentional.homicide	Attempted.intentional.homicide	Assault	Kidnapping
	<dbl>	<dbl>	<dbl>	<dbl>
Albania	2.03	3.25	5.52	0.14
Austria	0.84	1.93	43.29	0.07
Belgium	1.27	8.87	556.36	*NA*

3 rows | 1-6 of 14 columns

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

```
str(EurostatCrime2019)
```

```
## 'data.frame':    41 obs. of  13 variables:
##  $ Intentional.homicide                                 : num  2.03 0.84 1.27 NA 1.14 0.8
1 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 33.33 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 dataset has 41 rows, and 13 columns (in addition to the the rownames and colheaders)

3i. For most countries sexual violence figures are the sum of rape and sexual assault. Remove the columns Rape and Sexual.assault.

```
EurostatCrime2019_2 <- EurostatCrime2019[,-c(6,7)]
str(EurostatCrime2019_2)
```

```
## 'data.frame':    41 obs. of  11 variables:
##  $ Intentional.homicide                         : num  2.03 0.84 1.27 NA 1.14 0.8
1 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
 ...
##  $ 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
 ...
```

3ii. Removing 4 columns related to theft & burglary

```
EurostatCrime2019_3 <- EurostatCrime2019_2[,-c(7:10)]
str(EurostatCrime2019_3)
```

```
## 'data.frame':    41 obs. of  7 variables:
##  $ Intentional.homicide                         : num  2.03 0.84 1.27 NA 1.14 0.8
1 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
 ...
##  $ Robbery                                      : num  3.42 29.67 140.14 NA 16.9
 ...
##  $ Unlawful.acts.involving.controlled.drugs.or.precursors: num  70.3 494.1 547.7 NA 78.1
 ...
```

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

```
EurostatCrime2019_4 <- cbind(EurostatCrime2019_3, rowSums(EurostatCrime2019_3, na.rm = TRUE))
head(EurostatCrime2019_4, 3)
```

	Intentional.homicide	Attempted.intentional.homicide	Assault	Kidnapping
	<dbl>	<dbl>	<dbl>	<dbl>
Albania	2.03	3.25	5.52	0.14
Austria	0.84	1.93	43.29	0.07
Belgium	1.27	8.87	556.36	*NA*

3 rows | 1-6 of 9 columns

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

```
EurostatCrime2019_5 <- EurostatCrime2019_4[!complete.cases(EurostatCrime2019_4),]
countries_with_missing_data <- rownames(EurostatCrime2019_5)
countries_with_missing_data
```

```
##  [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.

```
EurostatCrime2019_6 <- EurostatCrime2019_4[complete.cases(EurostatCrime2019_4),]
```

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

```
str(EurostatCrime2019_6)
```

```
## '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 ...
##  $ rowSums(EurostatCrime2019_3, na.rm = TRUE)        : num  90 621 146 335 150 ...
```

Task 2 - Analysis

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

```
EurostatCrime2019_7 <- EurostatCrime2019_6['Ireland',-8]
EurostatCrime2019_8 <-  t(apply(EurostatCrime2019_7[-1], 1, function(x) head(sort(-x), 3)))
most_common_crimes_in_ireland <- colnames(EurostatCrime2019_8)
most_common_crimes_in_ireland
```

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

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

```
EurostatCrime2019_9 <- EurostatCrime2019_6['Ireland',]
EurostatCrime2019_10 <- cbind(EurostatCrime2019_9, EurostatCrime2019_9$Assault / EurostatCrim
e2019_9$`rowSums(EurostatCrime2019_3, na.rm = TRUE)`)
Ireland_Assault_pc <- EurostatCrime2019_10[,9] * 100
Ireland_Assault_pc
```

```
## [1] 16.05316
```

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

```
EurostatCrime2019_11 <- EurostatCrime2019_6[order(EurostatCrime2019_6$Kidnapping, decreasing
= TRUE), ]
Highest_Kidnap <- head(EurostatCrime2019_11,1)
Highest_Kidnap
```

	Intentional.homicide <dbl>	Attempted.intentional.homicide <dbl>	Assault <dbl>	Kidnappin <db
Luxembourg	0.65	9.61	103.76	7.1

1 row | 1-6 of 9 columns

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

```
EurostatCrime2019_12 <- EurostatCrime2019_6 [order(EurostatCrime2019_6$`rowSums(EurostatCrime
2019_3, na.rm = TRUE)`), ]
Lowest_Overall <- head(EurostatCrime2019_12,1)
Lowest_Overall
```

|	Intentional.homicide <dbl>	Attempted.intentional.homicide <dbl>	Assault <dbl>	Kidnapping <dbl>
Romania	1.31	1.86	1.46	2.01

1 row | 1-6 of 9 columns

5. Create a plot displaying the relationship between robbery and unlawful acts involving controlled drugs or precursors. Make the plot look " i.e. change axis labels etc.

```
par(bg = "beige", mar = c(4,4,3,4), mgp = c(3,0.4,0), las =1)
plot(EurostatCrime2019_6$Robbery, EurostatCrime2019_6$Unlawful.acts.involving.controlled.drug
s.or.precursors,
    type = "p",
    xlab = 'Robbery offences / 100k',
    ylab = 'Unlawful acts involving controlled drugs or precursors / 100k',
    main = 'Relationship between Robbery & Drug related \n crimes within EU contries in 2019'
,
    pch = 16,
    cex = 1.5,
    col = 'blue'
     )
text(100,180,"Spain has a significantly higher \n number of Robberies than \n other EU countr
ies")
grid()
```


Relationship between Robbery & Drug related crimes within EU contries in 2019
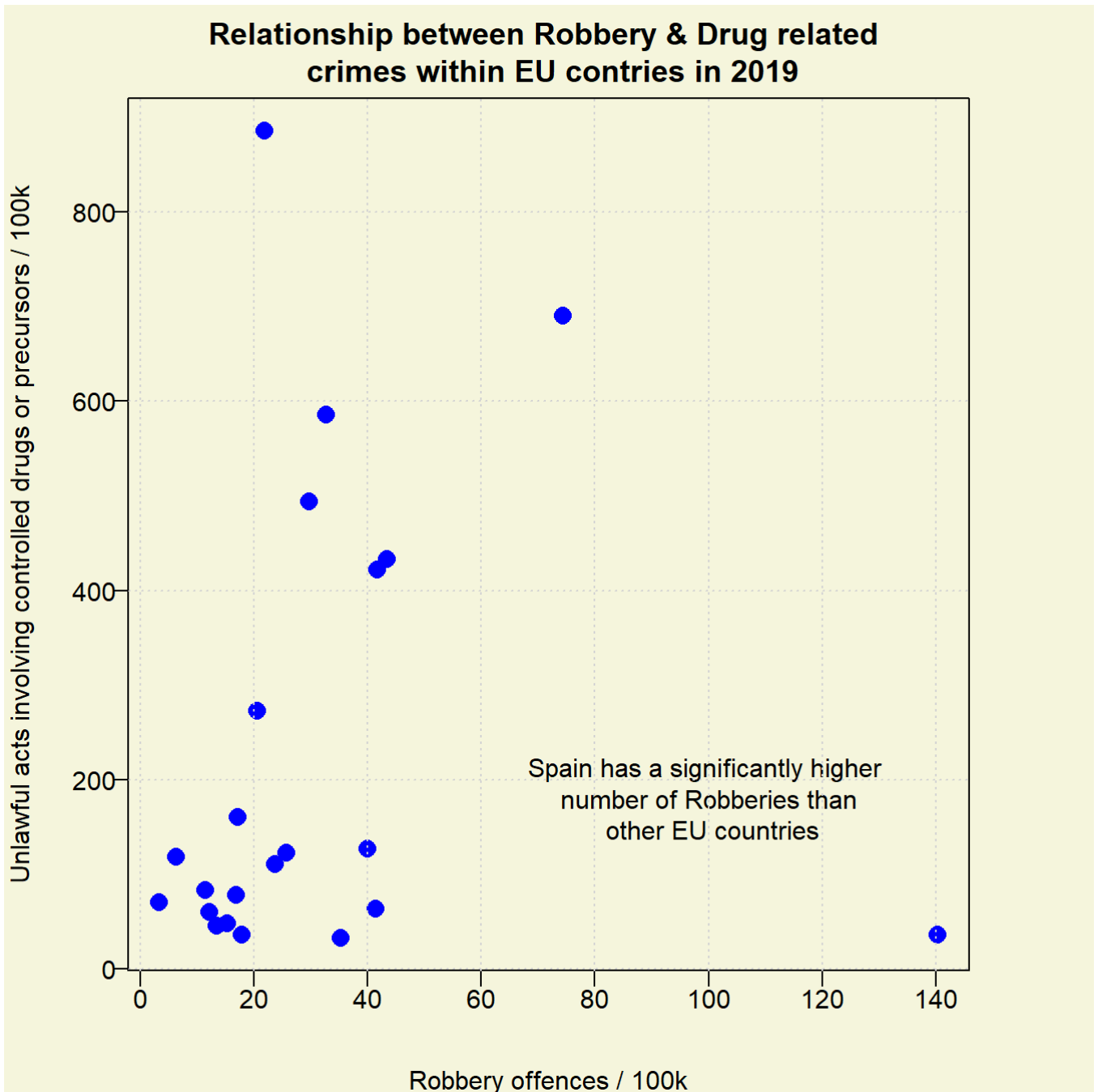
Task 3 - Creativity

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.

```
hist(EurostatCrime2019_6$`rowSums(EurostatCrime2019_3, na.rm = TRUE)`,
    breaks = 10,
    xlab = 'Total crime per 100k inhabitants',
    main = 'Histogram of total crime statistics for EU countries in 2019 (per 100k)',
    col = 'gray51',
    border = 'grey88',
    )
text(750, 5, "6 out of the 22 countries \n recorded >600 crimes \n per 100k inhabitants", font = 3, family = 'mono', col = 'red')
grid()
```

Histogram of total crime statistics for EU countries in 2019 (per 100k)
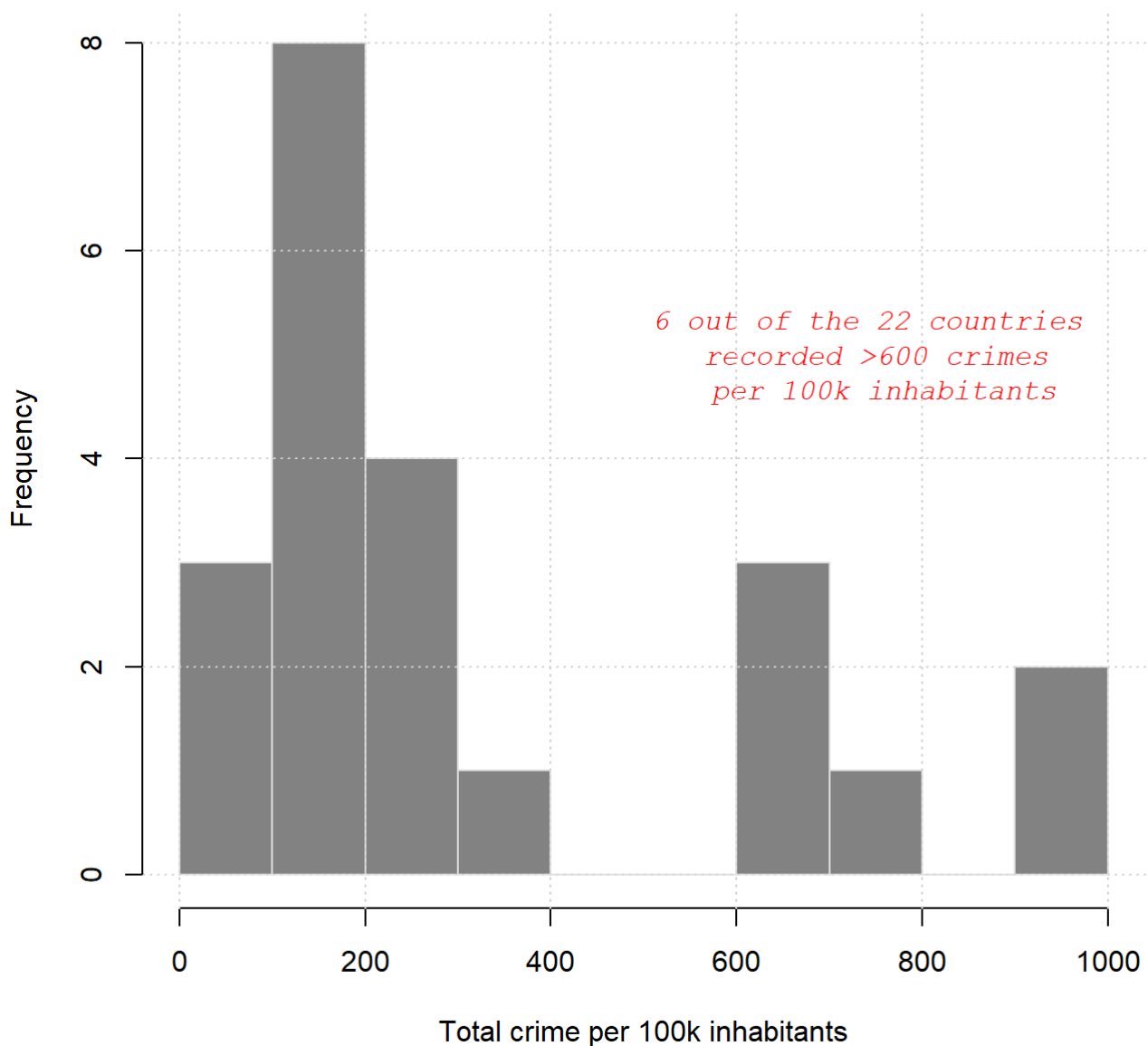
It is clear that the majority of countries record a total level of crime lower than 400 incidents per 100k inhabitants