Assignment 1

Conor Heffron 23211267

Task 1: Manipulation

1. Load dataset

- Load the dataset EurostatCrime2021.xlsx. Notice that the data starts in row 6, with row 7 containing the variable names, and that the missing values are represented by ":". [Note: do not modify the original file EurostatCrime2021.xlsx directly (by opening it with Excel, for example) as we need to be able to render your .Qmd file and reproduce your final document using the original dataset.

```
library(readxl)
# EurostatCrime2021 <- read_excel("/Users/conorheffron/Google Drive/UCD/MSc in AI for Medi



EurostatCrime2021 <- read_excel("/Users/conorheffron/Google Drive/UCD/MSc in AI for Medici
```

2. Size / Dimensions of data set

- What is the size (number of rows and columns) and the structure of this dataset?
- Answer: 41 observations / records of 18 variables / columns

```
dim(EurostatCrime2021)
```

```
[1] 41 18
```

3. Remove fraud columns X2

- Remove the columns Fraud and Money laundering (they contain no data).

```
library(dplyr)
```

```
Attaching package: 'dplyr'

The following objects are masked from 'package:stats':

    filter, lag

The following objects are masked from 'package:base':

    intersect, setdiff, setequal, union
```

```
df <- select(EurostatCrime2021, -Fraud, -"Money laundering")
```

4. Remove theft/burglary related columns X2

- For some countries Theft includes also Theft of a motorized vehicle or parts thereof, Burglary, and Burglary of private residential premises, in others they are recorded separately. To compare different countries, remove the columns involving theft and burglary: • Theft, • Theft of a motorized vehicle or parts thereof, • Burglary, • Burglary of private residential premises

```
df2 <- subset(select(df, -Theft, -"Theft of a motorized vehicle or parts thereof", -Burgla
```

5. Add a column

- Add a column containing the overall record of offences for each country (per hundred thousand inhabitants) [Hint: there is a function in base R that allow you to do this].

```
df2_num <- select(df2, -Country)
df2[colnames(df2_num)] <- sapply(df2[colnames(df2_num)], as.numeric)
```

```
Warning in lapply(X = X, FUN = FUN, ...): NAs introduced by coercion

Warning in lapply(X = X, FUN = FUN, ...): NAs introduced by coercion

Warning in lapply(X = X, FUN = FUN, ...): NAs introduced by coercion

Warning in lapply(X = X, FUN = FUN, ...): NAs introduced by coercion

Warning in lapply(X = X, FUN = FUN, ...): NAs introduced by coercion

Warning in lapply(X = X, FUN = FUN, ...): NAs introduced by coercion

Warning in lapply(X = X, FUN = FUN, ...): NAs introduced by coercion

Warning in lapply(X = X, FUN = FUN, ...): NAs introduced by coercion

Warning in lapply(X = X, FUN = FUN, ...): NAs introduced by coercion

Warning in lapply(X = X, FUN = FUN, ...): NAs introduced by coercion

Warning in lapply(X = X, FUN = FUN, ...): NAs introduced by coercion
```

```
df2$offences <- rowSums(select(df2, -Country))
```

6. List countries with missing data

- Work with the dataset you just created, and write some code to list the countries that contain any missing data.

```
print(df2[which.max(df2$offences),])
```

```
# A tibble: 1 x 13
  Country `Intentional homicide` Attempted intentional homic~1 `Serious assault`
  <chr>                    <dbl>                         <dbl>             <dbl>
1 Finland                    1.7                          6.76              30.0
# i abbreviated name: 1: `Attempted intentional homicide`
# i 9 more variables: Kidnapping <dbl>, Rape <dbl>, `Sexual assault` <dbl>,
#   `Sexual exploitation` <dbl>, Robbery <dbl>,
#   `Unlawful acts involving controlled drugs or precursors` <dbl>,
#   Corruption <dbl>, `Acts against computer systems` <dbl>, offences <dbl>
```

```r
countries<-list()
for(i in 1:nrow(df2)) {
 if(any(is.na(df2[i,]))) {
  countries <- append(countries, df2[[i,"Country"]])
 }
}
print(length(countries))
```

```
[1] 29
```

```r
print(countries)
```

```
[[1]]
[1] "Belgium"

[[2]]
[1] "Bulgaria"

[[3]]
[1] "Denmark"

[[4]]
[1] "Germany"

[[5]]
[1] "Estonia"

[[6]]
[1] "France"

[[7]]
[1] "Italy"

[[8]]
[1] "Luxembourg"

[[9]]
[1] "Hungary"

[[10]]
```

```
[1] "Malta"

[[11]]
[1] "Netherlands"

[[12]]
[1] "Poland"

[[13]]
[1] "Portugal"

[[14]]
[1] "Slovenia"

[[15]]
[1] "Slovakia"

[[16]]
[1] "Sweden"

[[17]]
[1] "Iceland"

[[18]]
[1] "Liechtenstein"

[[19]]
[1] "Norway"

[[20]]
[1] "Switzerland"

[[21]]
[1] "England and Wales"

[[22]]
[1] "Scotland"

[[23]]
[1] "Northern Ireland (UK)"

[[24]]
[1] "Bosnia and Herzegovina"
```

```
[[25]]
[1] "Montenegro"

[[26]]
[1] "North Macedonia"

[[27]]
[1] "Serbia"

[[28]]
[1] "Türkiye"

[[29]]
[1] "Kosovo (under United Nations Security Council Resolution 1244/99)"
```

7. Remove the countries with missing data.

```
df2 <- na.omit(df2)
```

8. How many observations and variables are in this new dataset?

```
dim(df2)
```

```
[1] 12 13
```

Task 2: Analysis

- Work with the dataset produced at the end of Task 1.

1. Which country has the highest overall record of offences in 2021 (per hundred thousand inhab- itants)? To get full marks you must also provide the R code that returns that country name only. [1]

```
print(df2[which.max(df2$offences),"Country"])
```

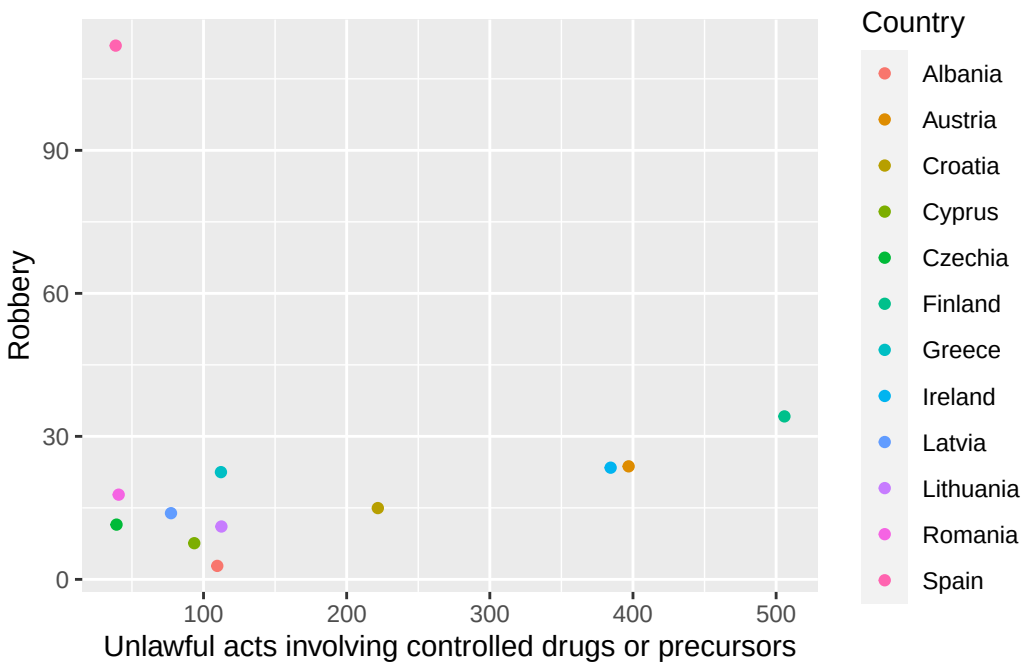
```
# A tibble: 1 x 1
  Country
  <chr>
1 Finland
```

2. Produce a table showing the countries and the proportion of the overall crimes due to acts against computer system, sort the rows in ascending order of proportions, and display only the first three decimal digits. [2]

```
# country
# (crimes due to acts against computer ) / offences * 100

df5 <- select(df2, Country, offences, "Acts against computer systems")
df5$proportion <- (df2$"Acts against computer systems" / df2$offences) * 100

df5[,'proportion']=round(df5[,'proportion'],3)

df6 <- df5[order(as.integer(df5$proportion),decreasing = FALSE), ]
```

3. Create a plot displaying the relationship between robbery and unlawful acts involving controlled drugs or precursors. Make the plot "nice" i.e., show country names, change size of the plot, change axis labels, etc.

```
# plot.window(xlim = c(0, 120), ylim = c(0, 510))
# plot(df2)
# plot(y=df2$"Robbery", x=df2$"Unlawful acts involving controlled drugs or precursors", ty
# points(df2$Country, y= Null, typer = "p")
# text(df2$"Robbery", df2$"Unlawful acts involving controlled drugs or precursors", labels

library(ggplot2)
# d = data.frame(x=runif(50),y=runif(50),z=runif(50))
colnames(df2) <- gsub(" ", "_", colnames(df2))
df2 |>
  ggplot(mapping = aes(x = Unlawful_acts_involving_controlled_drugs_or_precursors, y = Rob
  geom_point(aes(colour = Country)) +
  labs(x = "Unlawful acts involving controlled drugs or precursors")
```




```
# dplyr group by country and then plot with above pts
```

Task 3: Creativity

- Do something interesting with these data! Create two plots showing something we have not discov- ered above already and outline your findings. For this Task you can decide if you want to use the original dataset from Task 1 Question 1 or the modified one. [4]