

Eurostat 2019 Analysis

Jason Ballantyne

11/10/2021

Task 1

1.

```
# Import the crime data
CrimeStats <- read.csv(file = "C:\\Users\\jason\\OneDrive - University College Dublin\\Documents\\MSO
head(CrimeStats)
```

```
##              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
```

2.

```
# Getting the size (number of rows and columns)
```

```
# Getting the number of rows
nrow(CrimeStats)
```

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

```
# Getting the number of columns
ncol(CrimeStats)
```

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

```
# An alternative method for checking the number of rows and columns
# dim(CrimeStats)
```

```
# Getting the structure of the dataset
str(CrimeStats)
```

```
## 'data.frame': 41 obs. of 13 variables:
## $ Intentional.homicide : num 2.03 0.84 1.27 NA 1.14 0.81 1.48 0
## $ Attempted.intentional.homicide : num 3.25 1.93 8.87 NA 0.54 2.4 1.71 0
## $ 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 0.1
## $ 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 ...
```

As per the above output:

- The number of rows are 41
- The number of columns are 13
- The structure of the dataset is a data frame.

3.

(i)

```
# Removing the columns Rape and Sexual Assault
CrimeStats <- CrimeStats[, !(names(CrimeStats) %in% c("Rape", "Sexual.assault"))]
```

(ii)

```
# Removing the columns Theft,
# Theft.of.a.motorized.land.vehicle,
# Burglary, Burglary.of.private.residential.premises

CrimeStats <- CrimeStats[, !(names(CrimeStats)
                             %in% c("Theft",
                                     "Theft.of.a.motorized.land.vehicle",
                                     "Burglary",
                                     "Burglary.of.private.residential.premises"))]

# Checking that all the columns from (i) & (ii) have been successfully removed
head(CrimeStats)
```

```
##              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
```

(iii)

```
# Adding a column that contains the overall record of offences for each country
CrimeStats$Overall.Offences <- rowSums(Filter(is.numeric, CrimeStats), na.rm = TRUE)

# Checking that the new column with a record of the overall offences has been added
```

```
head(CrimeStats)
```

```
##              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.Offences
## Albania             90.00
## Austria            620.75
## Belgium           1331.83
## Bosnia and Herzegovina      0.00
## Bulgaria           145.93
## Croatia            335.06
```

4.

```
# Listing the countries that contain missing data
```

```
# We will firstly add any row with an NA to the dataframe CountryList
```

```
# Then we will print the list of countries that contain missing data
```

```
CountryList <- CrimeStats[rowSums(is.na(CrimeStats)) > 0,]
rownames(CountryList)
```

```
## [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"
```

The above output tells us that the countries that contain missing data include:

- Belgium

- Bosnia and Herzegovina
- Denmark
- England and Wales
- Estonia
- France
- Hungary
- Iceland
- Liechtenstein
- Netherlands
- North Macedonia
- Northern Ireland (UK)
- Norway
- Poland
- Portugal
- Scotland
- Slovakia
- Sweden
- Turkey

5.

```
# Removing the countries with missing data
CrimeStats <- na.omit(CrimeStats)

# An alternative method to removing the countries with missing data
# CrimeStats <- CrimeStats[rowSums(is.na(CrimeStats)) == 0, ]

# Summing the number of NA's to prove that missing data
# has been removed from the dataframe
sum(is.na(CrimeStats))
```

```
## [1] 0
```

Based on the above, there is 0 data missing, which shows we have successfully removed the countries with missing data from the dataframe.

6.

```
# Checking how many observations and variables are in the new dataframe

# Checking the number of observations
nrow(CrimeStats)
```

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

```
# Checking the number of variables
ncol(CrimeStats)
```

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

```
# Alternative method for checking the observations and variables
# dim(CrimeStats)
```

- There are 22 observations in the new dataset
- There are 8 variables in the new dataset

Task 2

1.

```
# Creating a new dataframe that contains all crimes in Ireland for 2019

IrelandCrimes <- CrimeStats["Ireland",
                             c("Intentional.homicide",
                                "Attempted.intentional.homicide",
                                "Assault", "Kidnapping",
                                "Sexual.violence", "Robbery",
                                "Unlawful.acts.involving.controlled.drugs.or.precursors" )]

# Sorting the dataframe in Descending order
IrelandCrimes <- sort(IrelandCrimes, decreasing = TRUE)

# Extracting the first three columns to find the 3 most common crimes in Ireland
IrelandCrimes[,1:3]

##           Unlawful.acts.involving.controlled.drugs.or.precursors Assault
## Ireland                                           421.84  102.18
##           Sexual.violence
## Ireland                                           67.86
```

The 3 most common crimes in Ireland in 2019 based on the above output were:

1. Unlawful acts involving controlled drugs or precursors
2. Assault
3. Sexual violence

2.

```
# Creating a new dataframe that contains Assault and Overall.offences
# figures in Ireland for 2019
Ireland_Assault <- CrimeStats["Ireland", c("Assault", "Overall.Offences")]

# Dataframe that calculates the proportion/percent of the
# overall crimes that was due to assault
Assault_Percent <- (Ireland_Assault$Assault/Ireland_Assault$Overall.Offences) * 100
Assault_Percent

## [1] 16.05316
```

Based on the above output, 16.0531649% of all crime in Ireland in 2019 was due to assault.

3.

```
# Creating a new dataframe that contains the row with the maximum/highest  
# record of kidnapping in 2019  
Kidnapping_Max <- CrimeStats[which.max(CrimeStats$Kidnapping),]  
  
# Extracting the first column to display which country had the  
# highest record of kidnapping in 2019  
rownames(Kidnapping_Max)
```

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

- Based on the output above, Luxembourg had the highest record of kidnapping in 2019.

4.

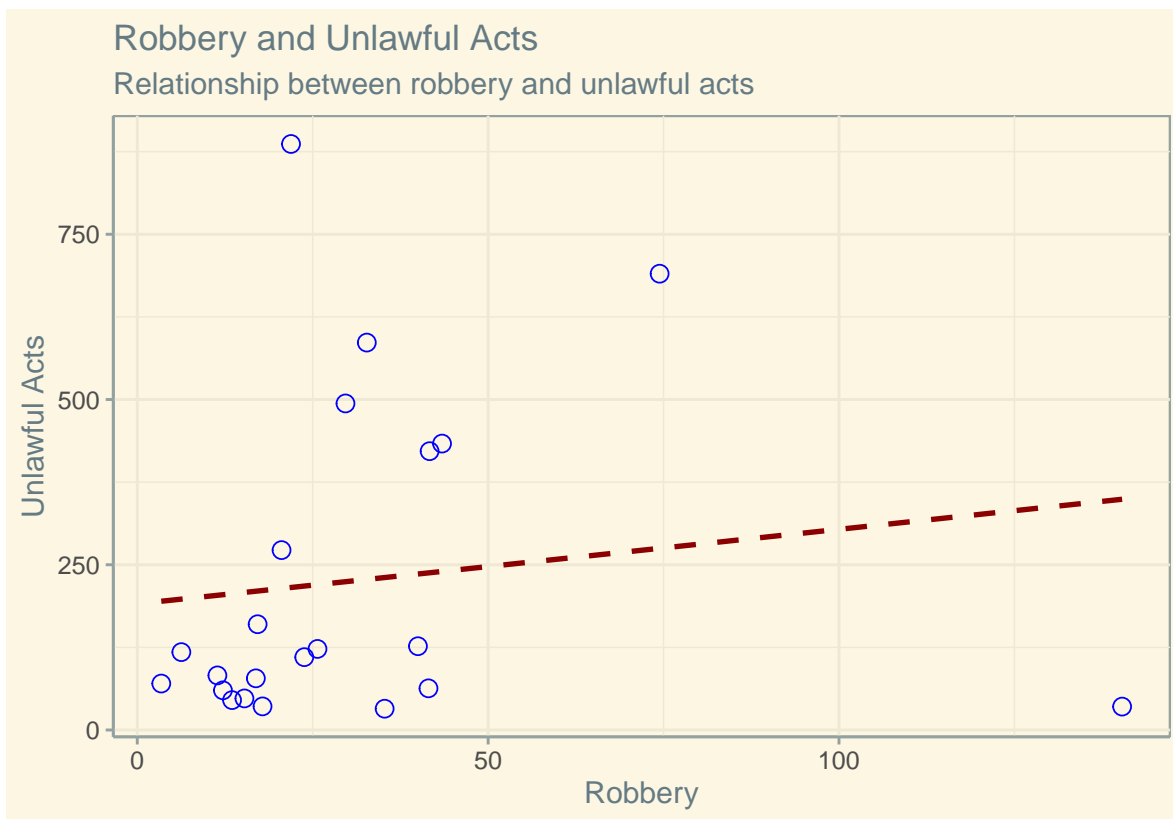
```
# Creating a new dataframe that contains the row with the lowest overall  
# recorded offences in 2019  
Lowest_Offence <- CrimeStats[which.min(CrimeStats$Overall.Offences),]  
  
# Extracting the first column to display which country had the  
# lowest overall record of offences in 2019  
rownames(Lowest_Offence)
```

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

- Based on the output above, Romania had the lowest overall record of offences in 2019.

5.

```
# load packages  
library(ggplot2)  
library(ggthemes)  
library(dplyr)  
  
# Creating a plot to display the relationship between robbery and unlawful acts  
# Plot was made to look "nice" through the use of ggplot, ggthemes and dplyr  
# A regression line was added to show the trend line of my scatter plot at a glance  
# Axis labels, a title, subtitle, and coloured shapes have all been included  
  
CrimeStats %>%  
  ggplot(aes(x = Robbery, y =  
              Unlawful.acts.involving.controlled.drugs.or.precursors)) +  
  geom_point(size=3, shape=21, color="blue") +  
  geom_smooth(formula = y ~ x,  
              method=lm, se=FALSE, linetype="dashed", color="darkred") +  
  labs(title = "Robbery and Unlawful Acts",  
        subtitle = "Relationship between robbery and unlawful acts",  
        x = "Robbery",  
        y = "Unlawful Acts") +  
  theme_solarized() +  
  theme(axis.title = element_text())
```



The plot shows that robbery and unlawful acts involving controlled drugs or precursors have a weak positive correlation.

Task 3

```
# Plotting overall offences on the European map
# to quickly identify the high offending countries
# We will combine ggplot2's map_data with our raw EurostatCrime2019 data

# Loading packages
library(ggplot2)
library(tidyverse)
library(cowplot)

# Using the original data set to get a complete list of the European countries
# We are not using row.names = 1 in this instance
# as we need the countries as a column to preform a join with ggplot2's map data
CreativeStats <- read.csv(file = "C:\\Users\\jason\\OneDrive - University College Dublin\\Documents\\CreativeStats.csv")

# Creating a column containing the overall record of offences for each country
CreativeStats$offences <- rowSums(Filter(is.numeric, CreativeStats), na.rm = TRUE)

# Extracting two columns to a new database
# which include the country and overall offences
EuropeMap <- CreativeStats[, c("X", "offences")]
head(EuropeMap)
```

```
## X offences
```



```
## 1          Albania    315.75
## 2          Austria   2725.98
## 3          Belgium   4447.04
## 4 Bosnia and Herzegovina 0.00
## 5          Bulgaria   720.36
## 6          Croatia   1015.93
```

```
# Changing the column name to match ggplot2's in order to preform a left join
colnames(EuropeMap)[colnames(EuropeMap) == "X"] <- "region"
```

```
# Renaming mismatching Eurostat country
# names to align with ggplot2's country names
EuropeMap$region[EuropeMap$region
  %in% c("Czechia",
        "England and Wales",
        "Germany (until 1990 former territory of the FRG)",
        "Kosovo (under United Nations
        Security Council Resolution 1244/99)",
        "Northern Ireland (UK)")] <-
  c("Czech Republic",
    "UK", "Germany", "Kosovo", "Northern Ireland")
```

```
head(EuropeMap)
```

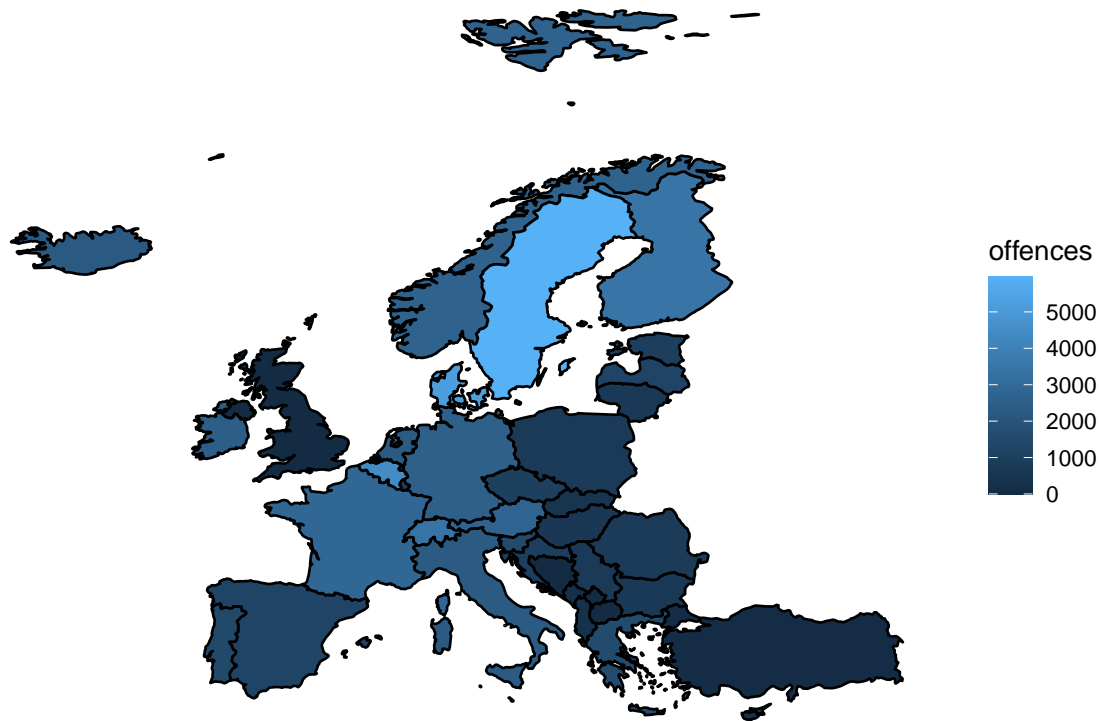
```
##          region offences
## 1          Albania    315.75
## 2          Austria   2725.98
## 3          Belgium   4447.04
## 4 Bosnia and Herzegovina 0.00
## 5          Bulgaria   720.36
## 6          Croatia   1015.93
```

```
# Creating a new dataframe to store ggplot2's mapdata
mapdata <- map_data("world")
View(mapdata)
```

```
# Left joining our crime data on the column region
mapdata <- left_join(mapdata, EuropeMap, by="region")
View(mapdata)
```

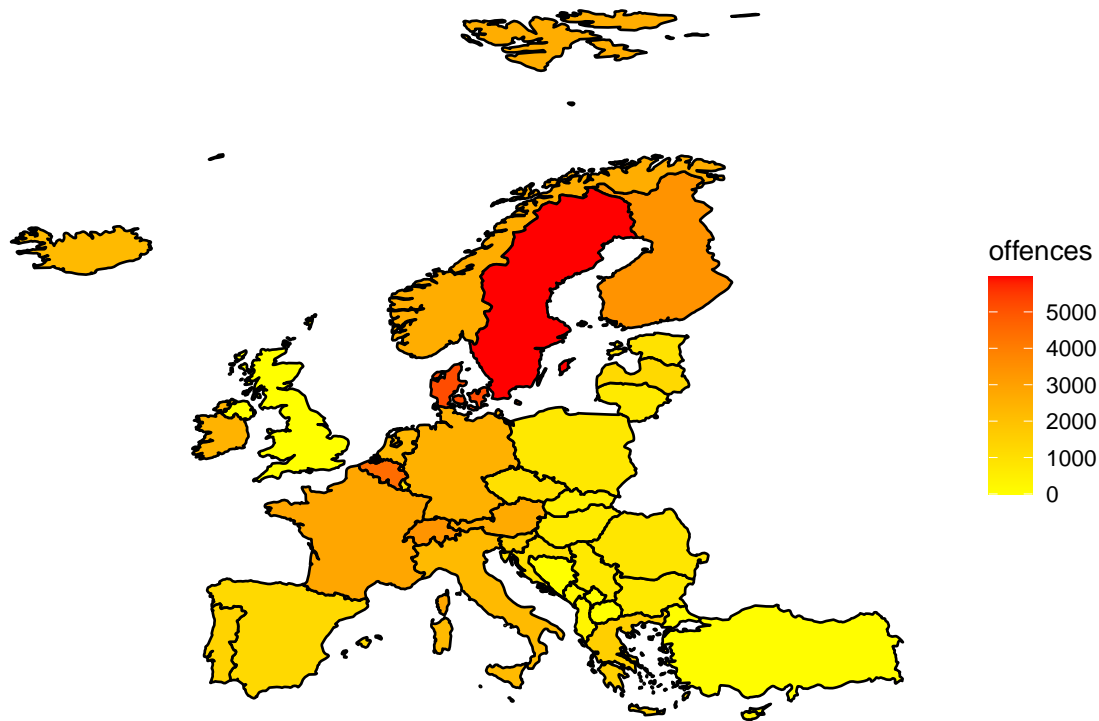
```
# Filtering NA's and storing in new dataframe
mapdata1<-mapdata %>% filter(!is.na(mapdata$offences))
View(mapdata1)
```

```
# Plotting our first map
map1<-ggplot(mapdata1, aes( x = long, y = lat, group=group)) +
  geom_polygon(aes(fill = offences), color = "black")+
  theme(axis.text.x = element_blank(),
        axis.text.y = element_blank(),
        axis.ticks = element_blank(),
        axis.title.y=element_blank(),
        axis.title.x=element_blank(),
        rect = element_blank())
map1
```



```
# Plotting our second map
map2 <- map1 + scale_fill_gradient(name = "offences",
                                   low = "yellow",
                                   high = "red",
                                   na.value = "blue")+

  theme(axis.text.x = element_blank(),
        axis.text.y = element_blank(),
        axis.ticks = element_blank(),
        axis.title.y=element_blank(),
        axis.title.x=element_blank(),
        rect = element_blank())
map2
```



It is interesting to see that overall offences do not seem to correlate with countries that are considered less wealthy as one might expect. For example, Sweden is listed as having the highest overall offences but is listed as 5th on Europe's GDP per capita in 2019 according to Eurostat's research. Similarly, Denmark is ranked second in overall offences and is listed as the 4th highest GDP per capita according to the same source.

On the other end of the scale, many eastern European countries such as Albania, Montenegro & Kosovo have some of the lowest offences and while being on the lower end of the GDP per capita scale for Europe.

This goes against the preconception that offences are linked to wealth. This is immediately apparent from the two plots where we can see that richer Scandinavian countries are in the higher end of the offences scale whereas less wealthy eastern European countries are in the lower end. This is a fascinating observation that we have not yet discovered in the dataset thus far.