

Assignment 1

Ranjith_Dhanaraj_19200041

10/9/2019

Loading the dataset ‘EurostatCrime2017.csv’

```
csv_file <- read.csv("EurostatCrime2017.csv", row.names = 1, header = TRUE)
```

Finding the size and structure of the dataset

```
dim(csv_file)
```

```
## [1] 41 11
```

```
str(csv_file)
```

```
## 'data.frame': 41 obs. of 11 variables:
##   $ Intentional_homicide : num 1.7 1.34 0.62 1.06 0.89 2.2 0.86 0.71 ...
##   $ Attempted_intentional_homicide : num 8.47 0.44 0.72 3.69 2.18 1.22 0.27 1 ...
##   $ Assault : num 611 39.6 45.1 33.1 166.1 ...
##   $ Kidnapping : num 10.31 1.44 0.16 NA 5.6 ...
##   $ Sexual.violence : num 63.22 9.19 13.37 83.41 42.19 ...
##   $ Robbery : num 167 21.9 15 35.5 47.1 ...
##   $ Burglary : num NA 125 228 955 443 ...
##   $ Burglary_of_private_residential_premises : num NA NA 68.4 702.6 141.2 ...
##   $ Theft : num NA 452 632 3721 1401 ...
##   $ Theft_of_a_motorized_land_vehicle : num NA 33.36 201.84 3.79 65.58 ...
##   $ Unlawful Acts_involving_controlled_drugs_or_precursors: num 506.6 70.2 52.9 481.6 400.6 ...
```

Normalize and add a new column called AllTheft

```
normalized_data <- csv_file
normalized_data[is.na(csv_file)] <- 0
csv_file$AllTheft <- normalized_data$Theft +
  normalized_data$Theft_of_a_motorized_land_vehicle + normalized_data$Burglary +
  normalized_data$Burglary_of_private_residential_premises
colnames(csv_file)
```

```
## [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"
## [12] "AllTheft"
```

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

```
csv_file[, c('Theft', 'Theft_of_a_motorized_land_vehicle',
            'Burglary_of_private_residential_premises',
            'Burglary')] <- NULL
colnames(csv_file)

## [1] "Intentional_homicide"
## [2] "Attempted_intentional_homicide"
## [3] "Assault"
## [4] "Kidnapping"
## [5] "Sexual.violence"
## [6] "Robbery"
## [7] "Unlawful Acts_involving_controlled_drugs_or_precursors"
## [8] "AllTheft"
```

Finding countries with NA value

```
countries_with_na <- csv_file[rowSums(is.na(csv_file)) > 0, ]
rownames(countries_with_na)
```

```
## [1] "Denmark"                  "France"
## [3] "Croatia"                  "Hungary"
## [5] "Netherlands"              "Austria"
## [7] "Poland"                   "Portugal"
## [9] "Sweden"                   "England_and_Wales"
## [11] "Iceland"                  "Liechtenstein"
## [13] "Norway"                   "North_Macedonia"
## [15] "Turkey"                   "Bosnia_and_Herzegovina"
```

Removing Countries with missing data

```
csv_file_new <- na.omit(csv_file)
csv_file_new
```

	Intentional_homicide	Attempted_intentional_homicide
## Belgium	1.70	8.47
## Bulgaria	1.34	0.44
## Czechia	0.62	0.72
## Germany	0.89	2.18
## Estonia	2.20	1.22
## Ireland	0.86	0.27
## Greece	0.72	1.39
## Spain	0.66	1.76
## Italy	0.61	1.81
## Cyprus	0.82	1.17
## Latvia	5.59	0.92
## Lithuania	3.97	0.56
## Luxembourg	0.34	12.53
## Malta	1.96	1.30
## Romania	1.46	1.95

## Slovenia		0.92		1.98
## Slovakia		1.47		1.47
## Finland		1.25		6.32
## Scotland		1.09		4.52
## Northern_Ireland_UK		1.29		5.30
## Switzerland		0.53		2.27
## Montenegro		1.77		8.19
## Albania		1.81		4.48
## Serbia		1.12		1.58
## Kosovo		1.85		7.51
##	Assault	Kidnapping	Sexual.violence	Robbery
## Belgium	611.03	10.31	63.22	166.97
## Bulgaria	39.58	1.44	9.19	21.94
## Czechia	45.06	0.16	13.37	14.98
## Germany	166.09	5.60	42.19	47.08
## Estonia	5.78	0.00	19.69	15.28
## Ireland	84.59	1.59	53.11	45.71
## Greece	14.02	0.72	4.21	39.66
## Spain	38.88	0.15	25.13	143.53
## Italy	108.56	0.33	8.44	51.44
## Cyprus	16.85	4.21	3.86	13.69
## Latvia	33.69	0.21	13.44	30.97
## Lithuania	6.25	0.00	7.94	38.20
## Luxembourg	99.04	7.79	61.29	76.86
## Malta	40.41	0.00	16.95	43.88
## Romania	1.38	1.58	7.81	16.10
## Slovenia	67.67	0.05	20.67	11.52
## Slovakia	30.21	1.67	12.97	8.63
## Finland	28.73	0.02	55.49	29.80
## Scotland	72.84	4.30	214.75	28.74
## Northern_Ireland_UK	56.41	3.59	179.95	30.38
## Switzerland	6.96	0.04	30.82	20.74
## Montenegro	23.62	0.00	3.70	18.96
## Albania	4.35	0.07	3.93	6.40
## Serbia	18.24	0.17	4.45	25.98
## Kosovo	18.56	0.90	5.10	17.21
##	Unlawful Acts involving controlled drugs or precursors			
## Belgium			506.65	
## Bulgaria			70.25	
## Czechia			52.93	
## Germany			400.60	
## Estonia			441.46	
## Ireland			351.58	
## Greece			118.82	
## Spain			27.85	
## Italy			63.28	
## Cyprus			111.02	
## Latvia			153.27	
## Lithuania			92.07	
## Luxembourg			454.23	
## Malta			67.35	
## Romania			25.24	
## Slovenia			78.61	
## Slovakia			27.23	

```

## Finland          505.61
## Scotland        596.43
## Northern_Ireland_UK 336.11
## Switzerland     951.05
## Montenegro      37.12
## Albania          72.55
## Serbia           119.41
## Kosovo           74.46
## AllTheft
## Belgium         0.00
## Bulgaria        609.92
## Czechia         1129.79
## Germany         2050.69
## Estonia          679.30
## Ireland          1956.29
## Greece          2146.60
## Spain            1022.46
## Italy             2329.51
## Cyprus            500.12
## Latvia            1174.13
## Lithuania         724.35
## Luxembourg       0.00
## Malta            2393.46
## Romania          716.42
## Slovenia         1584.54
## Slovakia          450.75
## Finland          2602.38
## Scotland          2171.66
## Northern_Ireland_UK 1695.72
## Switzerland       2503.48
## Montenegro        230.89
## Albania           228.22
## Serbia            620.14
## Kosovo            670.41

```

Finding the number of observations and variables in this new dataframe

```
dim(csv_file_new)
```

```
## [1] 25  8
```

Finding the 3 most crimes in Ireland with the new dataset

```

ireland_info <- csv_file_new["Ireland", ]
crimes_sorted <- sort(ireland_info, decreasing = TRUE)
top_three_crimes <- c(crimes_sorted[1:3])
top_three_crimes

```

```

## $AllTheft
## [1] 1956.29
##
## $Unlawful_acts_involving_controlled_drugs_or_precursors

```

```
## [1] 351.58
##
## $Assault
## [1] 84.59
```

Finding country with highest overall offences

```
csv_file_new$all_offense <- rowSums(csv_file_new)
rownames(csv_file_new[which.max(csv_file_new$all_offense), ])
```

```
## [1] "Switzerland"
```

We can still figure out how dangerous a country is. Theft and burglary are mostly non-violent, if we categorize crimes into 3 categories : Violent, Non-violent and Drug Abuse, we can get a better picture of which country is being plagued by a particular type of crime.

Graph 1 gives us the count as well the type without scaling.

```
library(ggplot2)
country_names = c(row.names(csv_file_new))

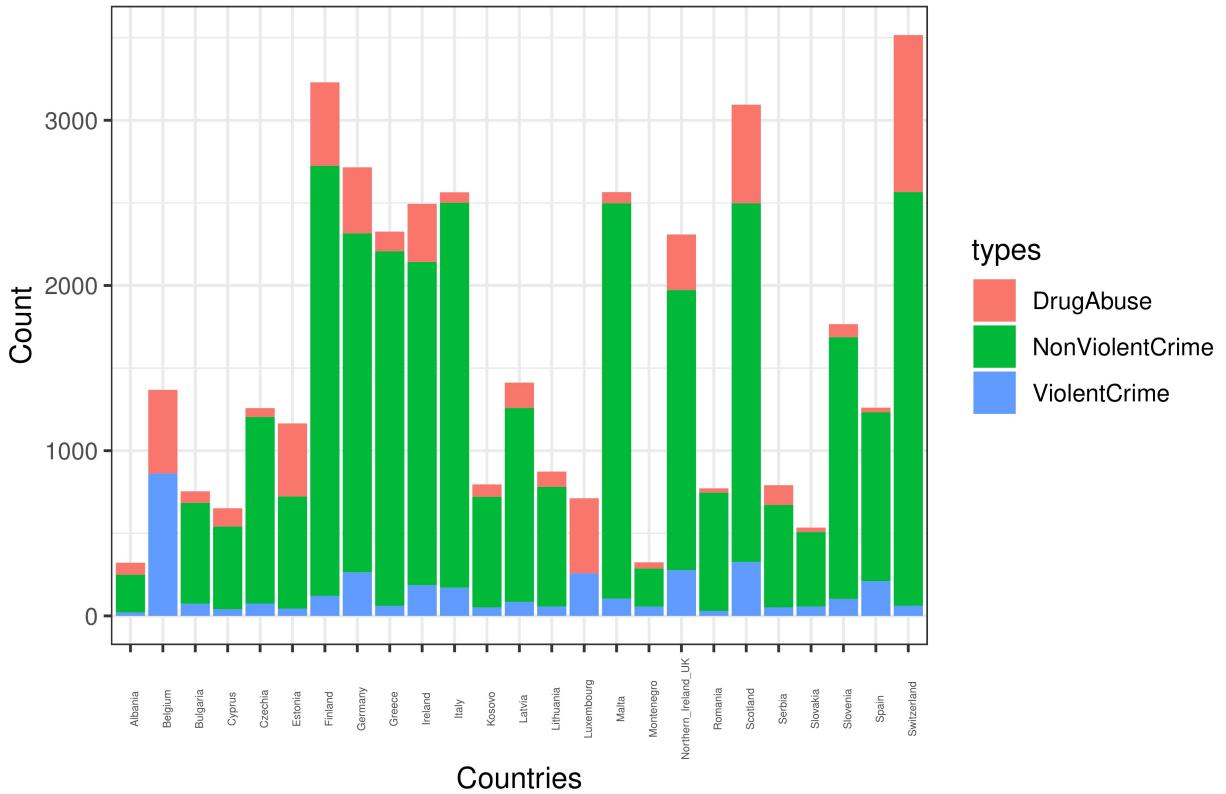
AllCrimes <- csv_file_new$all_offense
ViolentCrime <- csv_file_new$all_offense -
  csv_file_new$AllTheft -
  csv_file_new$Unlawful_acts_involving_controlled_drugs_or_precursors
NonViolentCrime <- csv_file_new$AllTheft
DrugAbuse <- csv_file_new$Unlawful_acts_involving_controlled_drugs_or_precursors
crimes <- c(ViolentCrime, NonViolentCrime, DrugAbuse)
types <- c(rep("ViolentCrime", length(ViolentCrime)),
           rep("NonViolentCrime", length(NonViolentCrime)),
           rep("DrugAbuse", length(DrugAbuse)))

plotting_data <- data.frame(country_names, crimes, types)

plt <- ggplot(plotting_data, aes(country_names, crimes))

plt + geom_bar(stat = "identity", aes(fill = types)) + xlab("Countries") + ylab("Count") +
  ggtitle("Types of Crimes by Country") + theme_bw() + theme(axis.text.x = element_text(size = 4, angle = 90))
```

Types of Crimes by Country

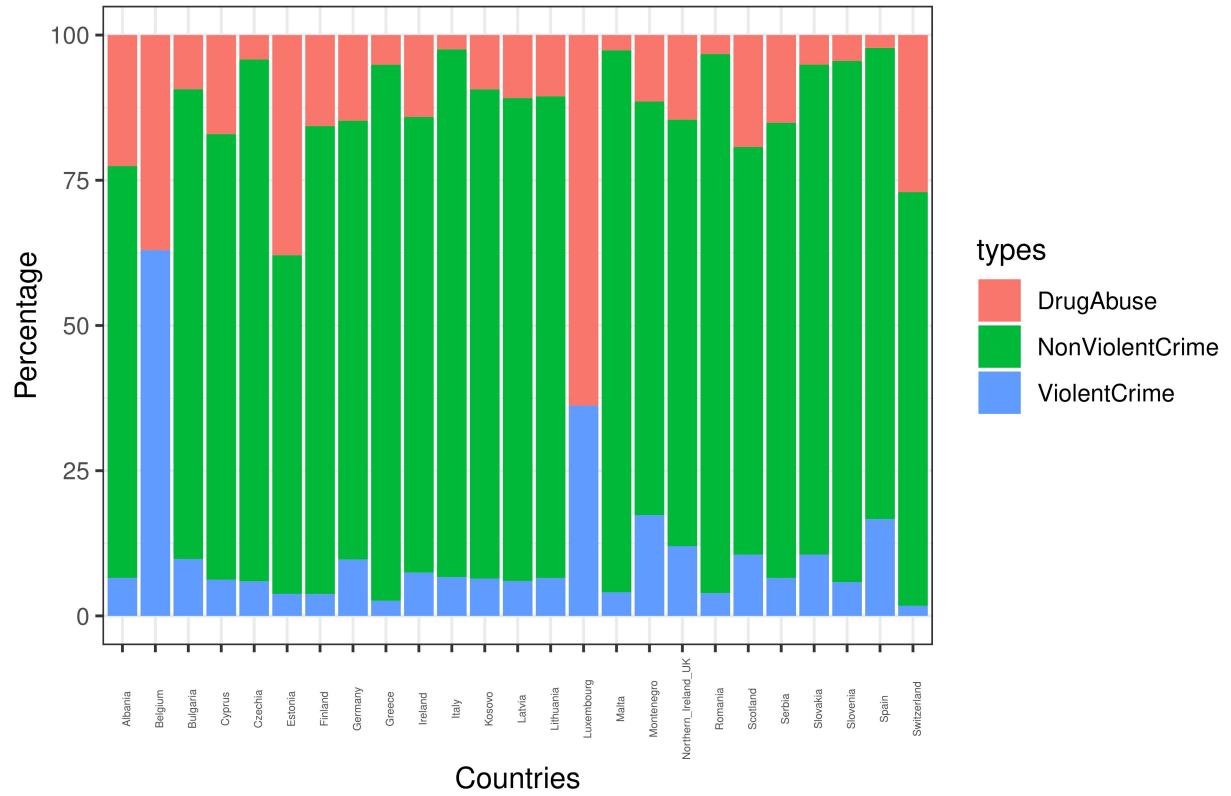


Graph 2, where everything is scaled to 100, to get better representation of types of crimes

```
plt <- ggplot(plotting_data, aes(country_names, (crimes/AllCrimes * 100)))

plt + geom_bar(stat = "identity", aes(fill = types)) +
  xlab("Countries") + ylab("Percentage") +
  ggtitle("Types of Crimes by Country") + theme_bw() + theme(axis.text.x = element_text(size = 4, angle = 90))
```

Types of Crimes by Country



We can see here that even though countries like Finland, where lots of crimes occur, most of them tend to be nonviolent unlike in countries like Belgium, where even though the amount of crime is far less compared to Finland, the type of crime seems to be mostly violent. This would make Belgium more dangerous than Finland purely because of the type of crime that occurs in this country. The same can be said for switzerland as well, where most of the crimes are non violent but there is an alarming number of crimes related to drugs as well.

Belgium and Luxemborg seems to be the most violent.

Estonia, Belgium and Luxemborg also seem to have problems with cases related to drugs.

Provided we get the population of each nation we can get an accurate picture of the crime rate since sheer numbers alone isn't enough to represent the crime statistics of a nation.