

# R Assignment 1

*Ashtami Bhuleskar 18201912*

*October 11, 2018*

## Data Programming with R

## Task 1: Manipulation

### 1.1: Load the dataset EurostatCrime

```
mydata <- read.csv(file = "C:\\Users\\Ashtami\\Documents\\R\\EurostatCrime2015.csv", head=TRUE, sep=",")  
colnames(mydata)[colnames(mydata)=="X"] <- "Country"  
mydata
```

```
## Country
## 1 Albania
## 2 Austria
## 3 Belgium
## 4 Bosnia and Herzegovina
## 5 Bulgaria
## 6 Croatia
## 7 Cyprus
## 8 Czech Republic
## 9 Denmark
## 10 England and Wales
## 11 Estonia
## 12 Finland
## 13 Former Yugoslav Republic of Macedonia, the
## 14 France
## 15 Germany
## 16 Greece
## 17 Hungary
## 18 Iceland
## 19 Ireland
## 20 Italy
## 21 Kosovo (under United Nations Security Council Resolution 1244/99)
## 22 Latvia
## 23 Liechtenstein
## 24 Lithuania
## 25 Luxembourg
## 26 Malta
## 27 Montenegro
## 28 Netherlands
## 29 Northern Ireland (UK)
## 30 Norway
## 31 Poland
## 32 Portugal
## 33 Romania
## 34 Scotland
## 35 Serbia
## 36 Slovakia
## 37 Slovenia
## 38 Spain
## 39 Sweden
## 40 Switzerland
## 41 Turkey
## Assault Intentional.homicide Rape Robbery Sexual.assault
## 1 NA NA NA NA NA
## 2 40.36 0.49 13.18 39.83 27.39
## 3 603.26 1.96 25.50 196.68 65.92
## 4 NA NA NA NA NA
## 5 34.99 1.79 1.65 27.02 6.72
## 6 19.03 0.88 6.11 31.03 8.21
## 7 16.65 1.42 2.36 10.98 9.45
## 8 148.69 0.80 5.67 19.19 7.79
## 9 25.80 0.81 18.57 35.94 19.88
## 10 744.32 NA 62.07 88.27 NA
```

## 11	7.45	3.19	12.24	25.63	9.35
## 12	28.22	1.61	19.23	28.33	31.74
## 13	NA	NA	NA	NA	NA
## 14	367.19	1.53	19.49	157.79	30.06
## 15	156.90	0.81	8.65	55.01	33.55
## 16	14.96	0.79	1.12	39.75	3.52
## 17	127.80	1.00	3.84	14.64	2.45
## 18	25.83	0.91	54.09	16.10	NA
## 19	321.48	1.32	11.62	55.63	34.74
## 20	105.34	0.77	NA	57.68	6.58
## 21	NA	NA	NA	NA	NA
## 22	26.89	4.08	3.02	39.22	10.62
## 23	329.18	0.00	2.68	8.03	42.82
## 24	7.33	5.75	5.31	54.43	7.46
## 25	108.00	0.89	12.08	98.41	51.16
## 26	42.62	0.93	5.36	56.37	18.87
## 27	22.50	2.73	0.80	25.08	3.38
## 28	282.21	NA	7.07	56.89	36.45
## 29	65.29	1.25	38.66	43.85	116.89
## 30	NA	NA	NA	NA	NA
## 31	14.52	0.75	3.24	21.42	1.40
## 32	4.52	0.96	3.61	149.13	21.24
## 33	1.50	1.46	5.11	16.90	3.24
## 34	NA	NA	NA	NA	145.04
## 35	16.05	1.28	0.86	42.59	3.91
## 36	35.05	0.89	1.60	9.94	10.29
## 37	74.65	0.97	2.04	11.25	10.47
## 38	62.55	0.65	2.65	139.03	18.60
## 39	47.52	1.15	56.88	86.80	120.79
## 40	7.48	0.69	6.46	39.80	26.44
## 41	NA	NA	NA	NA	NA

##	Sexual.violence	Theft
## 1	NA	NA
## 2	40.57	1586.92
## 3	91.42	1660.42
## 4	NA	NA
## 5	8.37	531.99
## 6	14.32	320.62
## 7	11.81	108.38
## 8	13.47	1319.87
## 9	38.45	3436.13
## 10	NA	2215.82
## 11	21.60	863.51
## 12	50.97	1781.22
## 13	NA	NA
## 14	49.54	1846.91
## 15	42.20	1646.84
## 16	4.64	923.72
## 17	6.28	1031.67
## 18	NA	1225.16
## 19	46.36	1500.60
## 20	NA	1719.49
## 21	NA	NA
## 22	13.64	976.14

```
## 23      45.50  516.51
## 24      12.77  688.78
## 25      63.24 1650.74
## 26      24.22 2015.40
## 27       4.18  132.94
## 28      43.52 3219.39
## 29     155.54 1300.20
## 30         NA     NA
## 31       4.64  363.54
## 32      24.86  832.95
## 33       8.35  545.72
## 34         NA     NA
## 35       4.76  317.71
## 36      11.90  444.37
## 37      12.51 1105.16
## 38      21.25  442.96
## 39     177.67 3828.01
## 40      32.90 1772.66
## 41         NA     NA
```

## 1.2: Size and the structure of this dataset

```
str(mydata)
```

```
## 'data.frame':  41 obs. of  8 variables:
## $ Country      : Factor w/ 41 levels "Albania","Austria",...: 1 2 3 4 5 6 7 8 9 10 ...
## $ Assault      : num  NA 40.4 603.3 NA 35 ...
## $ Intentional.homicide: num  NA 0.49 1.96 NA 1.79 0.88 1.42 0.8 0.81 NA ...
## $ Rape         : num  NA 13.18 25.5 NA 1.65 ...
## $ Robbery      : num  NA 39.8 196.7 NA 27 ...
## $ Sexual.assault : num  NA 27.39 65.92 NA 6.72 ...
## $ Sexual.violence : num  NA 40.57 91.42 NA 8.37 ...
## $ Theft        : num  NA 1587 1660 NA 532 ...
```

```
class(mydata)
```

```
## [1] "data.frame"
```

```
dim(mydata)
```

```
## [1] 41  8
```

## 1.3: Produce appropriate commands

i) Add a new column called Sex.crime

```
mydata$Sex.crime <- mydata$Rape + mydata$Sexual.assault + mydata$Sexual.violence
```

## ii) Remove the columns Rape, Sexual.assault and Sexual.violence

```
mydata$Rape <- NULL  
mydata$Sexual.assault <- NULL  
mydata$Sexual.violence <- NULL  
mydata
```

##	Country
## 1	Albania
## 2	Austria
## 3	Belgium
## 4	Bosnia and Herzegovina
## 5	Bulgaria
## 6	Croatia
## 7	Cyprus
## 8	Czech Republic
## 9	Denmark
## 10	England and Wales
## 11	Estonia
## 12	Finland
## 13	Former Yugoslav Republic of Macedonia, the
## 14	France
## 15	Germany
## 16	Greece
## 17	Hungary
## 18	Iceland
## 19	Ireland
## 20	Italy
## 21	Kosovo (under United Nations Security Council Resolution 1244/99)
## 22	Latvia
## 23	Liechtenstein
## 24	Lithuania
## 25	Luxembourg
## 26	Malta
## 27	Montenegro
## 28	Netherlands
## 29	Northern Ireland (UK)
## 30	Norway
## 31	Poland
## 32	Portugal
## 33	Romania
## 34	Scotland
## 35	Serbia
## 36	Slovakia
## 37	Slovenia
## 38	Spain
## 39	Sweden
## 40	Switzerland
## 41	Turkey

##	Assault	Intentional.homicide	Robbery	Theft	Sex.crime
## 1	NA	NA	NA	NA	NA
## 2	40.36	0.49	39.83	1586.92	81.14
## 3	603.26	1.96	196.68	1660.42	182.84
## 4	NA	NA	NA	NA	NA
## 5	34.99	1.79	27.02	531.99	16.74
## 6	19.03	0.88	31.03	320.62	28.64
## 7	16.65	1.42	10.98	108.38	23.62
## 8	148.69	0.80	19.19	1319.87	26.93
## 9	25.80	0.81	35.94	3436.13	76.90
## 10	744.32	NA	88.27	2215.82	NA

## 11	7.45	3.19	25.63	863.51	43.19
## 12	28.22	1.61	28.33	1781.22	101.94
## 13	NA	NA	NA	NA	NA
## 14	367.19	1.53	157.79	1846.91	99.09
## 15	156.90	0.81	55.01	1646.84	84.40
## 16	14.96	0.79	39.75	923.72	9.28
## 17	127.80	1.00	14.64	1031.67	12.57
## 18	25.83	0.91	16.10	1225.16	NA
## 19	321.48	1.32	55.63	1500.60	92.72
## 20	105.34	0.77	57.68	1719.49	NA
## 21	NA	NA	NA	NA	NA
## 22	26.89	4.08	39.22	976.14	27.28
## 23	329.18	0.00	8.03	516.51	91.00
## 24	7.33	5.75	54.43	688.78	25.54
## 25	108.00	0.89	98.41	1650.74	126.48
## 26	42.62	0.93	56.37	2015.40	48.45
## 27	22.50	2.73	25.08	132.94	8.36
## 28	282.21	NA	56.89	3219.39	87.04
## 29	65.29	1.25	43.85	1300.20	311.09
## 30	NA	NA	NA	NA	NA
## 31	14.52	0.75	21.42	363.54	9.28
## 32	4.52	0.96	149.13	832.95	49.71
## 33	1.50	1.46	16.90	545.72	16.70
## 34	NA	NA	NA	NA	NA
## 35	16.05	1.28	42.59	317.71	9.53
## 36	35.05	0.89	9.94	444.37	23.79
## 37	74.65	0.97	11.25	1105.16	25.02
## 38	62.55	0.65	139.03	442.96	42.50
## 39	47.52	1.15	86.80	3828.01	355.34
## 40	7.48	0.69	39.80	1772.66	65.80
## 41	NA	NA	NA	NA	NA

## 1.4: List the countries that contain any missing data

```
countryList <- mydata$Country[complete.cases(mydata)==FALSE]
countryList
```

```
## [1] Albania
## [2] Bosnia and Herzegovina
## [3] England and Wales
## [4] Former Yugoslav Republic of Macedonia, the
## [5] Iceland
## [6] Italy
## [7] Kosovo (under United Nations Security Council Resolution 1244/99)
## [8] Netherlands
## [9] Norway
## [10] Scotland
## [11] Turkey
## 41 Levels: Albania Austria Belgium Bosnia and Herzegovina ... Turkey
```

## 1.5: Remove the countries with missing data from the dataframe

```
mydata_new <- na.omit(mydata)
mydata_new
```



##	Country	Assault	Intentional.homicide	Robbery	Theft
## 2	Austria	40.36	0.49	39.83	1586.92
## 3	Belgium	603.26	1.96	196.68	1660.42
## 5	Bulgaria	34.99	1.79	27.02	531.99
## 6	Croatia	19.03	0.88	31.03	320.62
## 7	Cyprus	16.65	1.42	10.98	108.38
## 8	Czech Republic	148.69	0.80	19.19	1319.87
## 9	Denmark	25.80	0.81	35.94	3436.13
## 11	Estonia	7.45	3.19	25.63	863.51
## 12	Finland	28.22	1.61	28.33	1781.22
## 14	France	367.19	1.53	157.79	1846.91
## 15	Germany	156.90	0.81	55.01	1646.84
## 16	Greece	14.96	0.79	39.75	923.72
## 17	Hungary	127.80	1.00	14.64	1031.67
## 19	Ireland	321.48	1.32	55.63	1500.60
## 22	Latvia	26.89	4.08	39.22	976.14
## 23	Liechtenstein	329.18	0.00	8.03	516.51
## 24	Lithuania	7.33	5.75	54.43	688.78
## 25	Luxembourg	108.00	0.89	98.41	1650.74
## 26	Malta	42.62	0.93	56.37	2015.40
## 27	Montenegro	22.50	2.73	25.08	132.94
## 29	Northern Ireland (UK)	65.29	1.25	43.85	1300.20
## 31	Poland	14.52	0.75	21.42	363.54
## 32	Portugal	4.52	0.96	149.13	832.95
## 33	Romania	1.50	1.46	16.90	545.72
## 35	Serbia	16.05	1.28	42.59	317.71
## 36	Slovakia	35.05	0.89	9.94	444.37
## 37	Slovenia	74.65	0.97	11.25	1105.16
## 38	Spain	62.55	0.65	139.03	442.96
## 39	Sweden	47.52	1.15	86.80	3828.01
## 40	Switzerland	7.48	0.69	39.80	1772.66

## Sex.crime

## 2	81.14
## 3	182.84
## 5	16.74
## 6	28.64
## 7	23.62
## 8	26.93
## 9	76.90
## 11	43.19
## 12	101.94
## 14	99.09
## 15	84.40
## 16	9.28
## 17	12.57
## 19	92.72
## 22	27.28
## 23	91.00
## 24	25.54
## 25	126.48
## 26	48.45
## 27	8.36
## 29	311.09

```
## 31      9.28
## 32     49.71
## 33     16.70
## 35      9.53
## 36     23.79
## 37     25.02
## 38     42.50
## 39    355.34
## 40     65.80
```

## 1.6: What is the size of this new dataframe?

```
dim(mydata_new)
```

```
## [1] 30  6
```

## Task 2: Analysis

### 2.1: What was the most common crime in Ireland in 2015?

```
irishmax <- max(mydata[which(mydata$Country=="Ireland"),2:ncol(mydata)])
irishmax
```

```
## [1] 1500.6
```

```
colnames(mydata)[which(mydata == irishmax, arr.ind = TRUE)[2]]
```

```
## [1] "Theft"
```

### 2.2: Three least common crimes in Ireland in 2015?

```
leastIrish <- sort(mydata[which(mydata$Country=="Ireland"),2:ncol(mydata)])
leastIrish[1:3]
```

```
##      Intentional.homicide Robbery Sex.crime
## 19                1.32    55.63    92.72
```

### 2.3: Which country have the highest record of offences

```
maxCountry <- mydata$Country[which.max(rowSums(mydata[-1],na.rm = TRUE))]  
maxCountry
```

```
## [1] Sweden  
## 41 Levels: Albania Austria Belgium Bosnia and Herzegovina ... Turkey
```

### 3

## Correlation Plot

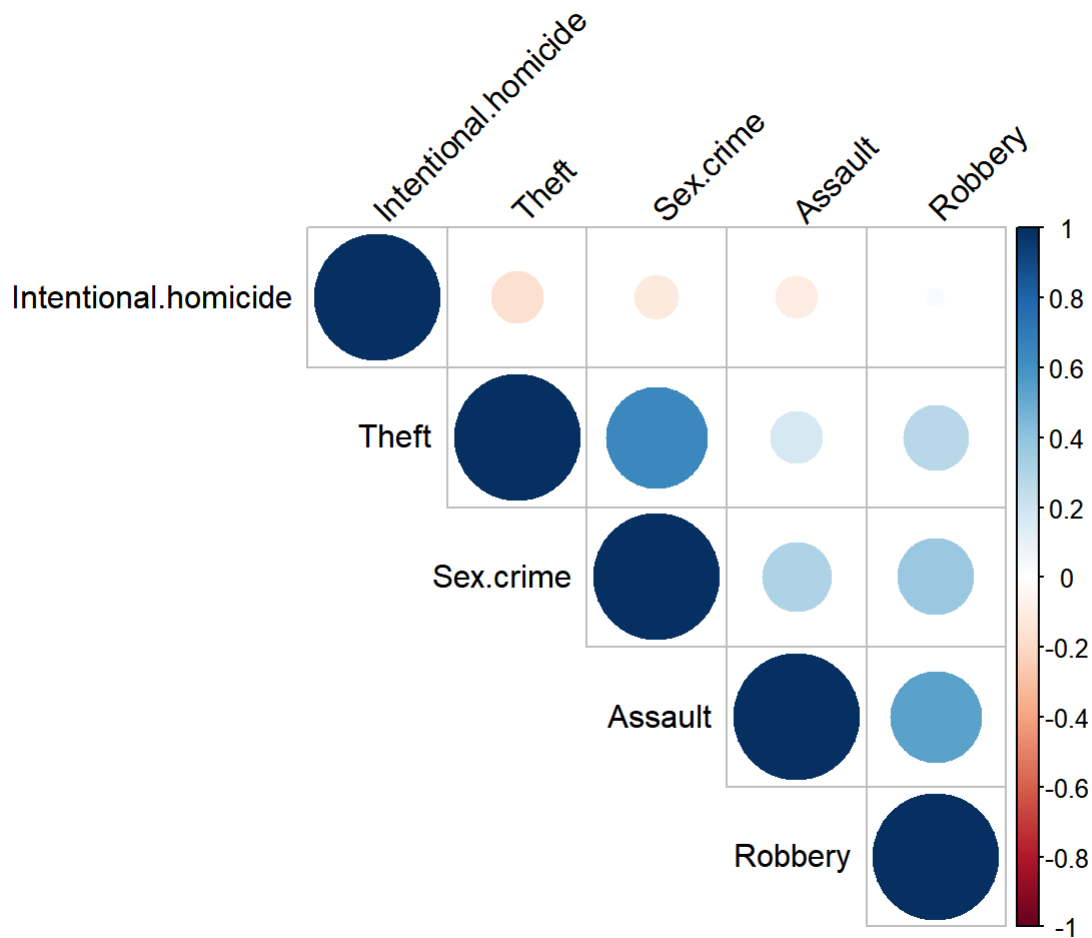
```
correlat <- mydata_new[,c(2,3,4,5,6)]  
corr_matrix <- cor(correlat)  
corr_matrix <- round(corr_matrix,2)
```

```
library(corrplot)
```

```
## Warning: package 'corrplot' was built under R version 3.4.4
```

```
## corrplot 0.84 loaded
```

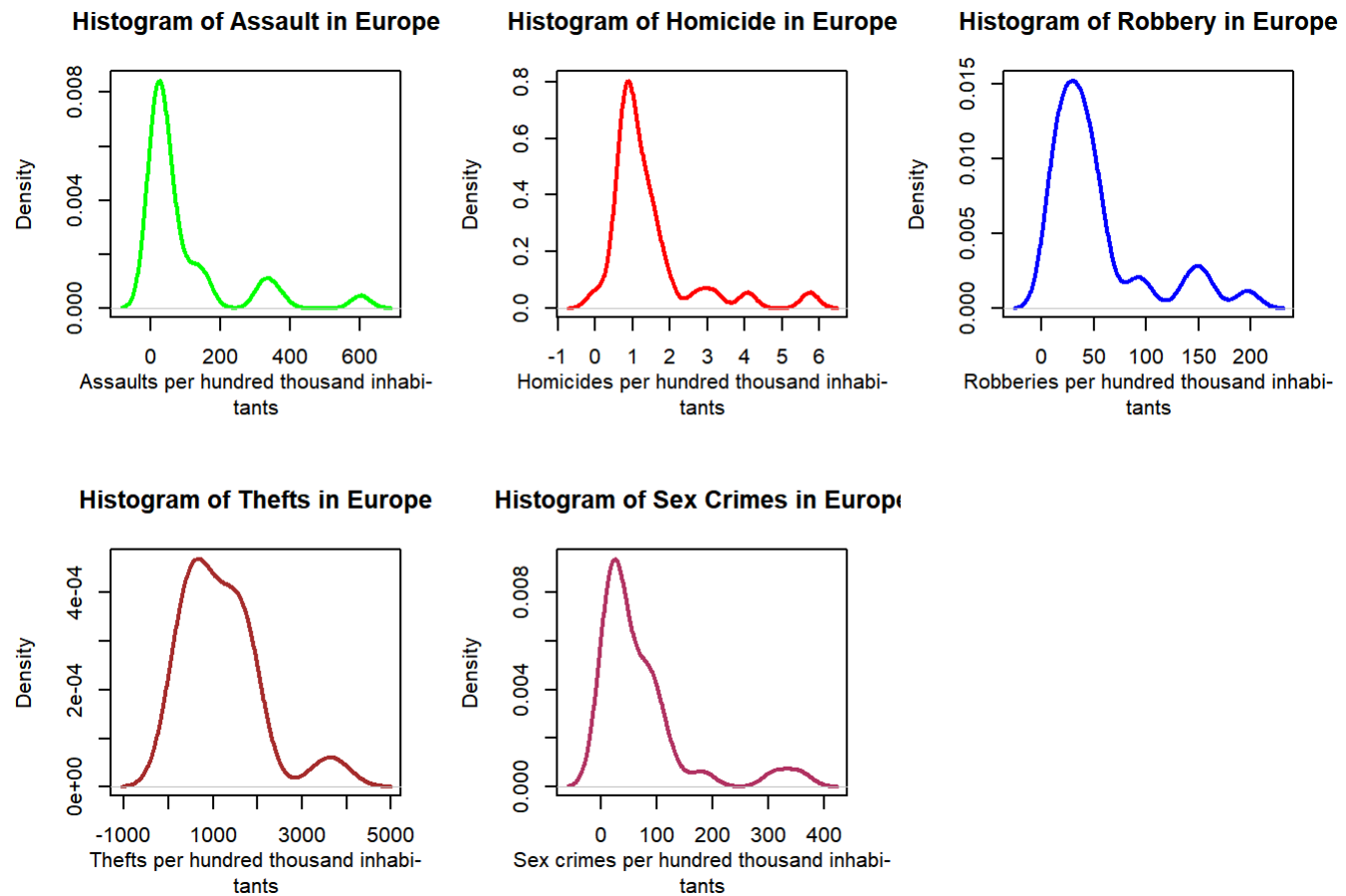
```
corrplot(corr_matrix, type = "upper", order = "hclust",  
         tl.col = "black", tl.srt = 45)
```



The correlation graph displays correlation between different offenses like Robbery, Assault, Sex Crime, Theft and International Homicide. From the plot, we can analyse that there is a strong positive correlation between Theft and Sex crime. We can infer from this that the countries having high occurrences of Theft are likely to have considerably high occurrences of Sex Crime. Other offenses show less correlation.

## Density Plot

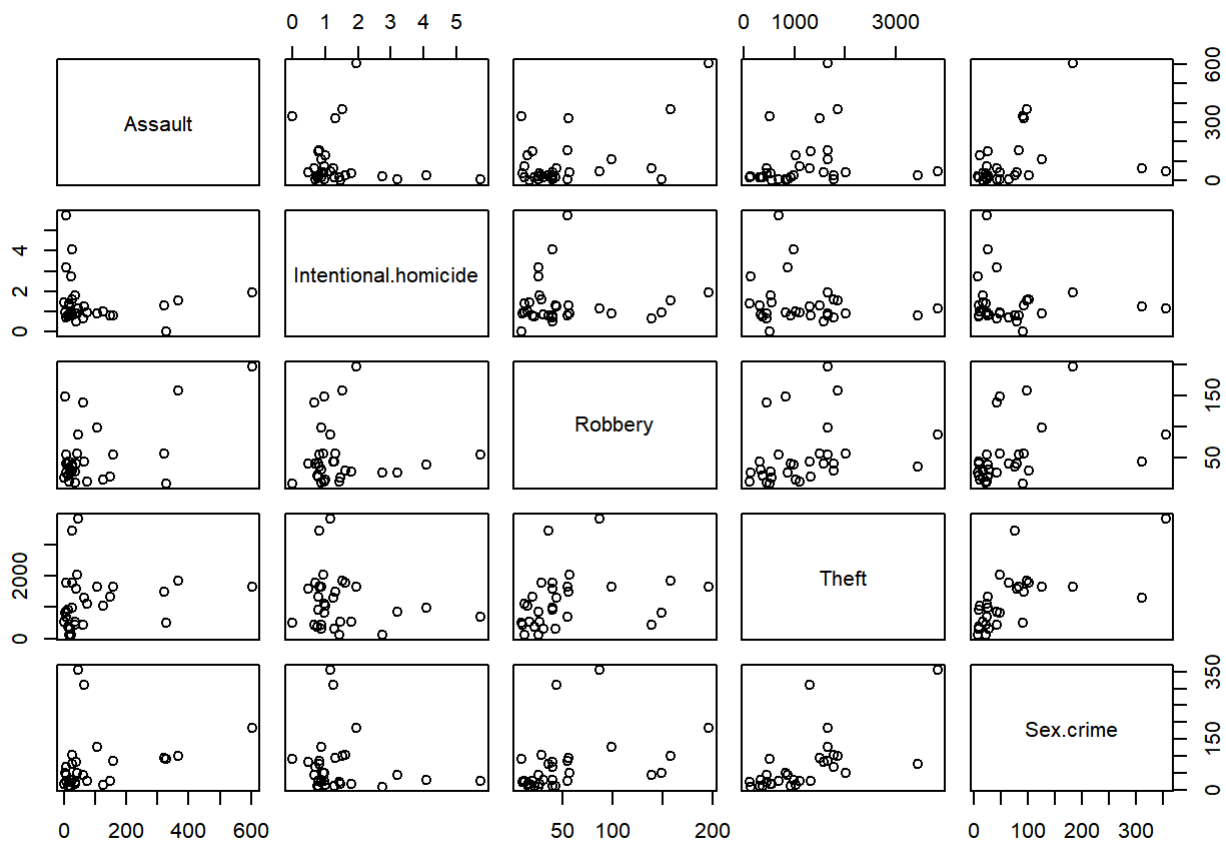
```
par(mfrow=c(2,3))
plot(density(mydata_new$Assault), xlab = "Assaults per hundred thousand inhabitants", main = "Histogram of Assault in Europe", col = "green", lwd = 2)
plot(density(mydata_new$Intentional.homicide), xlab = "Homicides per hundred thousand inhabitants", main = "Histogram of Homicide in Europe", col = "red", lwd = 2)
plot(density(mydata_new$Robbery), xlab = "Robberies per hundred thousand inhabitants", main = "Histogram of Robbery in Europe", col = "blue", lwd = 2)
plot(density(mydata_new$Theft), xlab = "Thefts per hundred thousand inhabitants", main = "Histogram of Thefts in Europe", col = "brown", lwd = 2)
plot(density(mydata_new$Sex.crime), xlab = "Sex crimes per hundred thousand inhabitants", main = "Histogram of Sex Crimes in Europe", col = "maroon", lwd = 2)
```



The density plots demonstrate a graph of the density of values for each offense. The peak of the graph shows the region with the highest number of a particular offense. The tail of the graph is the region with lower density of crime for a particular offense.

## Scatterplot Matrix

```
pairs(mydata_new[-1])
```



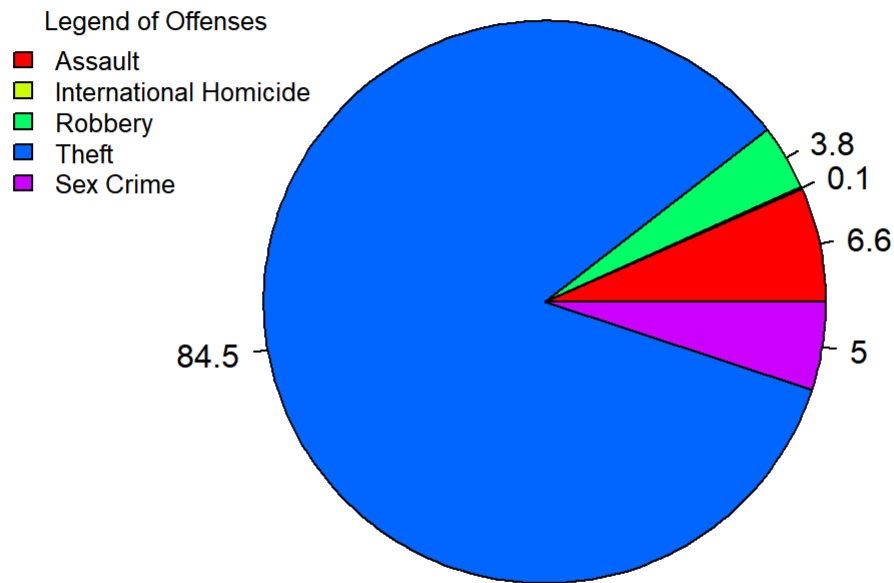
The graph is the scatterplot between all pairs of types of offenses under consideration. The extremely scattered data points in the graphs can be thought of as outliers. There is no evident linear pattern in any of the paired graphs.

## Pie Chart

```
par(mfrow=c(1,1))
total <- colSums(mydata_new[-1])
percent <- round(100*total/sum(total), 1)
pie(total, main= "Total Offense Percentage", labels = percent, cex= 1, col=rainbow(length(total)),radius = 1)

legend("topleft", c("Assault","International Homicide","Robbery","Theft", "Sex Crime"), cex = 0.8,
      fill = rainbow(length(total)), bty = "n", title = "Legend of Offenses")
```

## Total Offense Percentage



The pie chart illustrates different offenses in Europe in terms of percent occurrence. By analysing the chart, we can see that the maximum offense in entire europe is Theft with 84.5% and the lowest of all is International Homicide having 0.1% only.