Assignment 1

Rohit Jyotiba Chougule & 19200240

10/10/2019

```
knitr::opts_chunk$set(echo = TRUE)
```

Task 1: Manipulation

1. Load the dataset EurostatCrime2017.csv. Notice that the first column of the csv file contains the names of the countries that must be read as row names [Hint: Load in the file using the function read.csv]

```
# Reading the csv file from the question to the variable- df
df <- read.csv("EurostatCrime2017.csv")
```

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

```
# Getting the dimensions of the data as number of rows and columns respectively:
dim(df)
```

```
## [1] 41 12
```

```
# Structure of the dataset is as seen below:
str(df, strict.width="wrap")
```

```
## 'data.frame':    41 obs. of  12 variables:
## $ ï..Country : Factor w/ 41 levels "Albania","Austria",..: 3 5 8 9 14 11
##     18 15 38 13 ...
## $ Intentional_homicide : num 1.7 1.34 0.62 1.06 0.89 2.2 0.86 0.72 0.66
##     1.41 ...
## $ Attempted_intentional_homicide : num 8.47 0.44 0.72 3.69 2.18 1.22 0.27
##     1.39 1.76 3.77 ...
## $ 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 ...
```

```
#strict.width=wrap is useful here to wrap around the text in the output for the str()

# renaming the column to Country as it has some unwanted characters at the start of column name
names(df)[names(df)=="ï..Country"] <- "Country"
```

The number of rows are 41 and columns are 12 as seen from the output for the code chunk Task1Q2

3. Produce appropriate commands to answer the following questions:

(i) For some countries Theft includes also burglary, and theft of motorised land vehicle, in others they are recorded separately. Add a new column called All Theft which contains the sum of all the crimes that have a theft component: • Theft, • Theft of a motorized land vehicle, • Burglary, • Burglary of private residential premises Please consider NA values as 0 in this case. [Hint: you may want to use the function apply or rowSums, check their help files to see how to deal with missing values.] [1]

(ii) Remove the columns: Theft, Theft of a motorized land vehicle, Burglary, and Burglary of private residential premises.

```
# Adding the column All_Totals to perform the sum of the columns:
All_Totals = rowSums(df[8:11], na.rm = TRUE)
df$All_Thefts <- All_Totals
All_Totals
```

```
##  [1]    0.00  609.92 1129.79 5382.97 2050.69  679.30 1956.29 2146.60
##  [9] 1022.46    0.00  717.69 2329.51  500.12 1174.13  724.35    0.00
## [17]  819.95 2393.46 2393.99    0.00  548.30 1199.99  716.42 1584.54
## [25]  450.75 2602.38 5005.47 4048.98 2171.66 1695.72 1693.82    0.00
## [33] 1868.20 2503.48  230.89    0.00  228.22  620.14    0.00    0.00
## [41]  670.41
```

(ii)

```
# Removing the columns as suggested in the question:
df <- df[-c(8:11)]
colnames(df)
```

```
## [1] "Country"
## [2] "Intentional_homicide"
## [3] "Attempted_intentional_homicide"
## [4] "Assault"
## [5] "Kidnapping"
## [6] "Sexual.violence"
## [7] "Robbery"
## [8] "Unlawful_acts_involving_controlled_drugs_or_precursors"
## [9] "All_Thefts"
```

4. Work with the dataset you created in question 3ii, and list the countries that contain any missing data

```
# Identifying all the rows with NA values in the dataset
test<-df[!complete.cases(df),]

#Printing the list of countries having NA values
test[1]
```

```
##              Country
## 4            Denmark
## 10            France
```

2

```
## 11                    Croatia
## 17                    Hungary
## 19                Netherlands
## 20                    Austria
## 21                     Poland
## 22                   Portugal
## 27                     Sweden
## 28          England_and_Wales
## 31                    Iceland
## 32              Liechtenstein
## 33                     Norway
## 36            North_Macedonia
## 39                     Turkey
## 40    Bosnia_and_Herzegovina
```

5. Remove the countries with missing data from the dataframe.

```r
# Usage of omit function to remove the NA values rows from the data set
new_df <- na.omit(df, invert=FALSE)

# Changing the row numbers to arrange the row numbers in a sequence 1 to n
rownames(new_df) <- 1:nrow(new_df)
```

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

```r
# Dimension of the new dataset after having removed NA values
dim(new_df)
```

```
## [1] 25  9
```

The number of rows are 25 and columns are 9 in the new data set.

Task 2: Analysis

Work with the dataset produced in question 6. Produce appropriate commands to answer the following questions:

1. According to these data what were the 3 most common crimes in Ireland in 2017?

```r
# Identifying country Ireland from the entire dataset for further processsing
ireland_df <- new_df[which(new_df$Country=="Ireland"),]

# Assigning sequence to the rows for new subset data frame
rownames(ireland_df) <- 1:nrow(ireland_df)

#ireland_df is the row or a small version of the data frame
# which has records for the crimes in Ireland

# Taking the Column values of the subset data frame into a Vector for sorting
test <- unlist((ireland_df))
```

```r
# Identifying the position of crimes in descending order.
pos <- order(test, decreasing = TRUE )

# Defining another vector crime_ire to append the actual crime
# values for top3 crimes in Ireland

crime_ire <- c()

# Loop to append the values from Crime values in the crime_ire

for(i in 1:3){
 crime_ire[i] <- c(test[pos[i]])
}

#typeof(ireland_df)

# Searching for the crime name for top three crimes from
# crime_ire to print their names from the data frame


# Given below is the list of top 3 crimes in Ireland in 2017
for(i in 1:3){
  print((ireland_df[which(ireland_df==crime_ire[i])]))
}
```

```
##   All_Thefts
## 1    1956.29
##   Unlawful_acts_involving_controlled_drugs_or_precursors
## 1                                                 351.58
##   Assault
## 1   84.59
```

All type of theft, Unlawful acts involving drugs and precursors, and Assault are the top 3 crimes in Ireland

2. Which country has the highest overall record of offences (per hundred thousand inhabitants)?

```r
Total_Offence = rowSums(new_df[2:9], na.rm = TRUE)

new_df$TotalCrime <- Total_Offence

Max_Crime = max(Total_Offence)

Country <- subset(new_df, TotalCrime==Max_Crime)

# The country with highest overall record of offences is:
Country[1]
```

```
##       Country
## 21 Switzerland
```

Switzerland is the country with highest overall offence.

Task 3: Creativity

Do something interesting with these data! Create a table or a plot which shows something we have not discovered above already and outline your findings.

```r
# Taking the sum of all the columns except the column names to find the total crime
# Taking a transpose of the sum
Total_Crime <- t(colSums(new_df[2:9]))

# Taking the data into the dataframe to get it into a readable format
All_Crime_df <- as.data.frame((Total_Crime))

# The below Table/dataframe shows most number of crimes in all the dataset
print(All_Crime_df, row.names = FALSE)
```

```
##   Intentional_homicide Attempted_intentional_homicide Assault Kidnapping
##                  36.84                          80.31  1638.8       44.9
##   Sexual.violence Robbery
##            881.67  964.65
##   Unlawful_acts_involving_controlled_drugs_or_precursors All_Thefts
##                                                  5735.18   30191.23
```

```r
# Thefts is the crime which occurs most across all the countries from the given dataset
print(All_Crime_df[which.max(All_Crime_df)], row.names = FALSE)
```

```
##   All_Thefts
##     30191.23
```

```r
# Intentional homicide is the least occured crime in all the countries from the given dataset

print(All_Crime_df[which.min(All_Crime_df)], row.names = FALSE)
```

```
##   Intentional_homicide
##                  36.84
```

```r
# The barplot as well shows the maximum number of crime overall
barplot(Total_Crime, beside = TRUE, las =2)
```