R Project on

**Since 1992**

Analysis of journalists Deaths

**By: Nikeeta Akbari & Kanishka Sunnam**

**Guided By: Dr. Shilpa Balan**

**A. Data set URL**: <https://data.world/journalism/journalist-deaths>

**Data set Format:** CSV Format

**Project Description:**

This dataset contains information about journalists who were killed/murdered while reporting news from the year 1992 to 2018. News keeps us informed of the various occasions, issues, and events happening around us. Journalists are responsible for bringing this information to the public through various mediums. This dataset holds information about the journalists who were killed or murdered, the organization he/she worked for, the nationality of the journalists, the country in which they were killed etc. The dataset contains the medium for which journalists worked such as Print, Radio, Television, Internet etc. The type of news on which journalists were working which got them killed/murdered. By this analysis, we plan to shed some light on what type of news reporting made the reporters vulnerable and whether that location needs more security or not. We are also keen to know whether nationality affects the number of deaths. The dataset also includes information about whether the journalists were subjected to violence such as torturing, threatening, kidnapping etc.

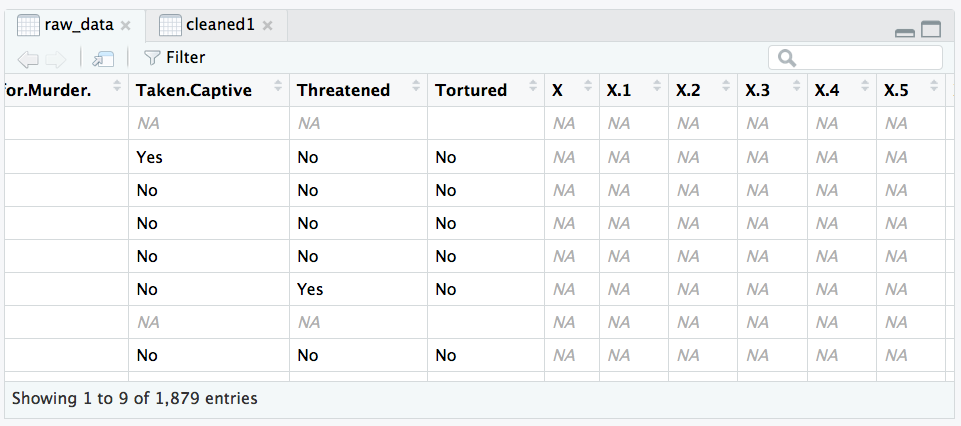
Journalists risk their lives in the process of reporting news. Basically, we rely on journalists to bring us the stories/news that helps shape our society. Through the data analysis we aim to identify the significant factors that impact these deaths such as race, type of news, location of news, organization for which reporters were working, etc., the number of journalists being killed every year, and try to increase awareness regarding this issue so that we can mitigate the risk of murder rates. International journalist associations and NGOs also provide free or low-cost training to journalists in developing countries. [1] By providing information related to vulnerable location/area, proper training can be provided to journalists so that they can handle the situation, which can in turn help in reducing the count of death eventually.

**B. Data Cleaning:**

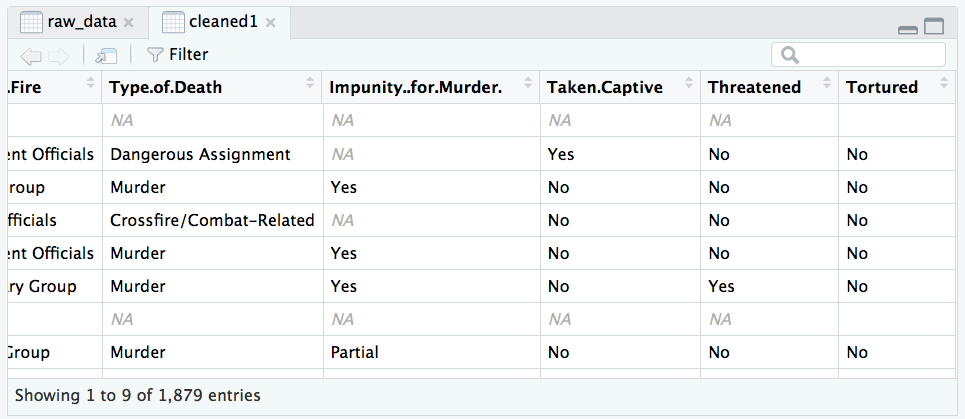
**1. Removing unwanted columns:**

Dataset contains some unwanted columns such as X, X.1, X.2, etc. which are removed by selecting necessary columns from dataset.

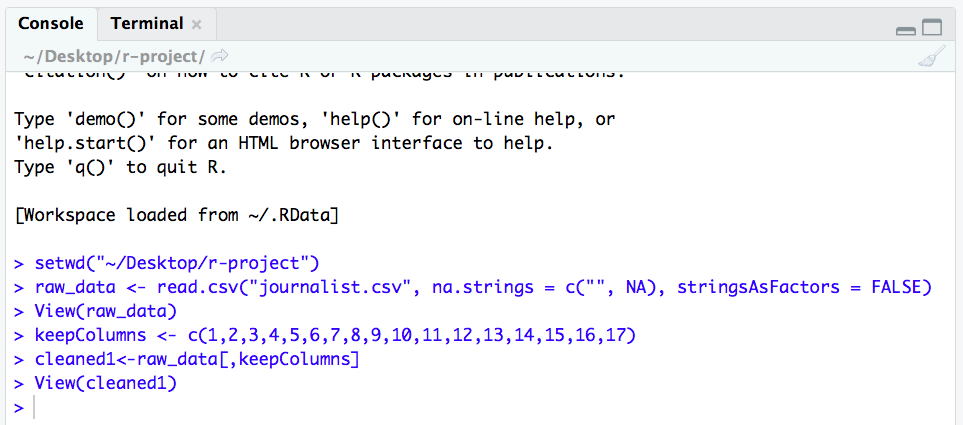
**Before:** Raw Data



**After:**



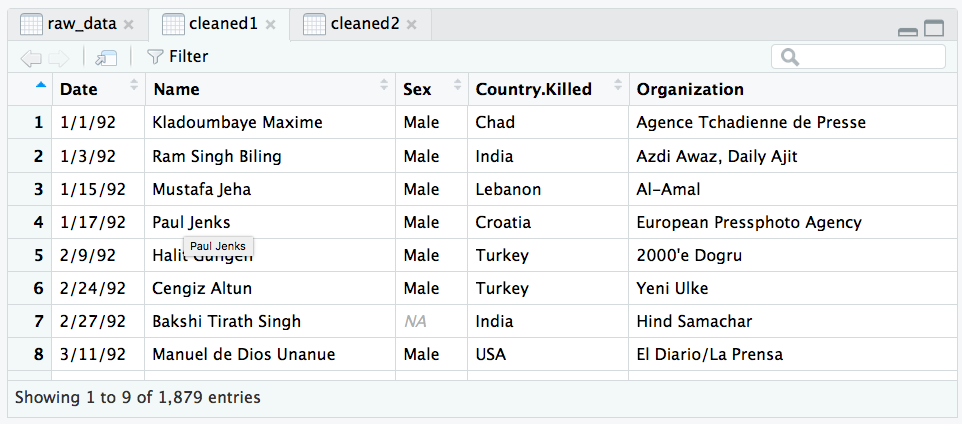
**Code for Removing Unwanted Columns:**



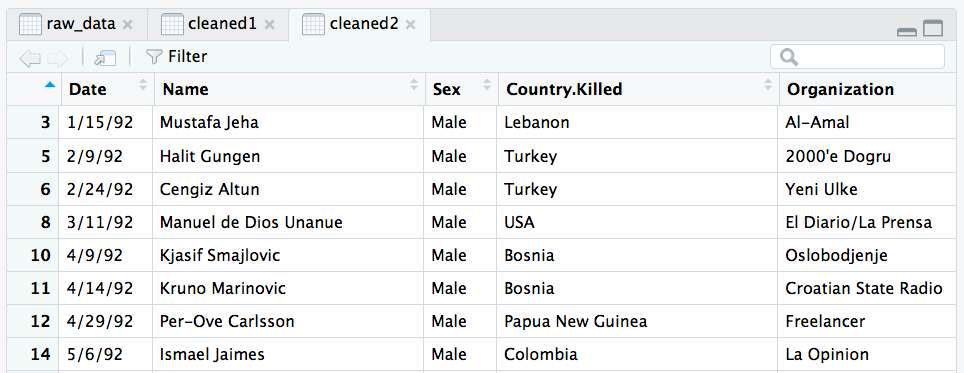
**2. Removing null values:**

Dataset contains lots of null values which are removed using omit() function.

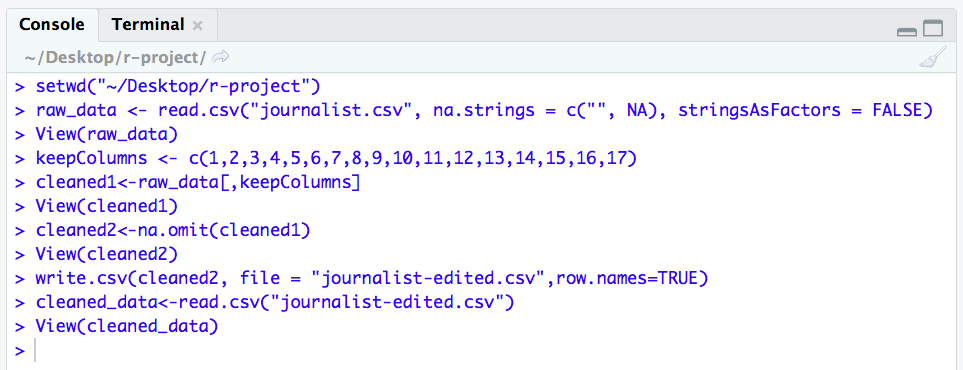
**Before:**



**After:**



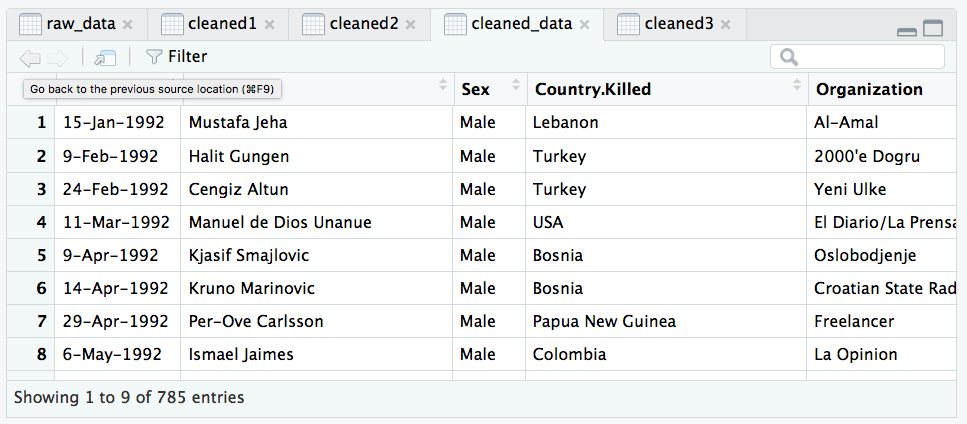
**Code for Removing Null Values:**



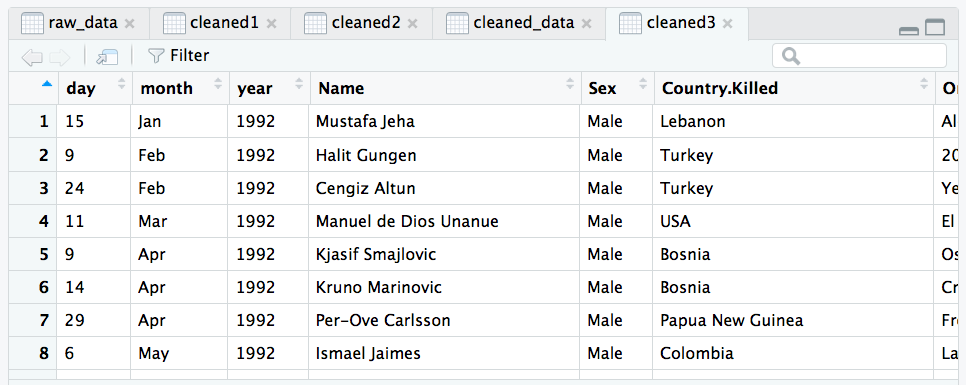
**3. Splitting Columns:**

In order to get separate year column, we split ‘Date’ column from our dataset into three subsets such as Day, Month and Year using separate function.

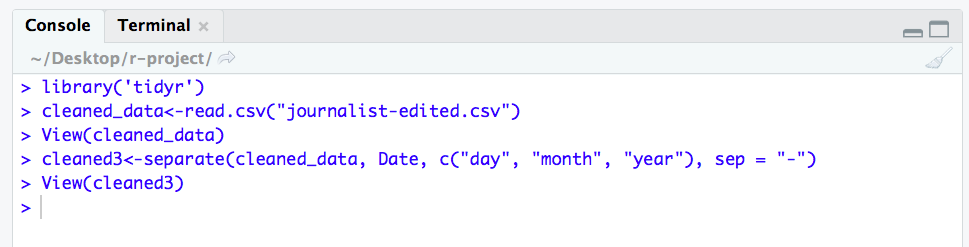
**Before:**



**After:**



**Code for splitting Columns:**

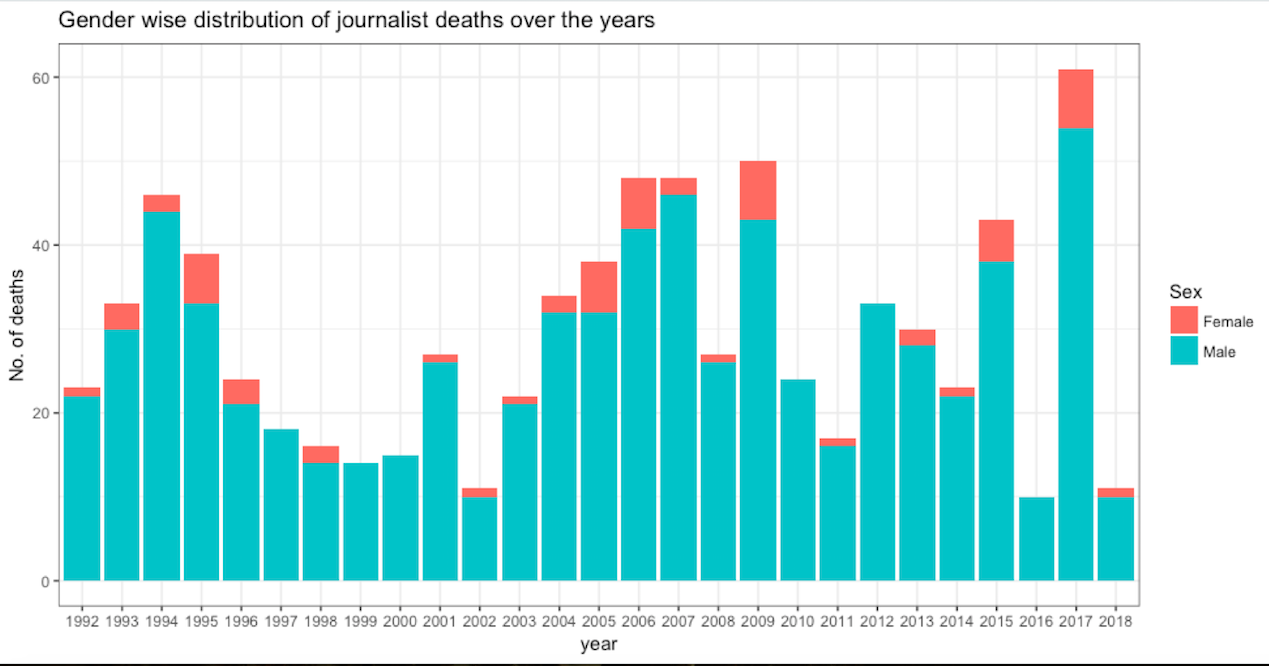


**C. Analysis & Visualizations:**

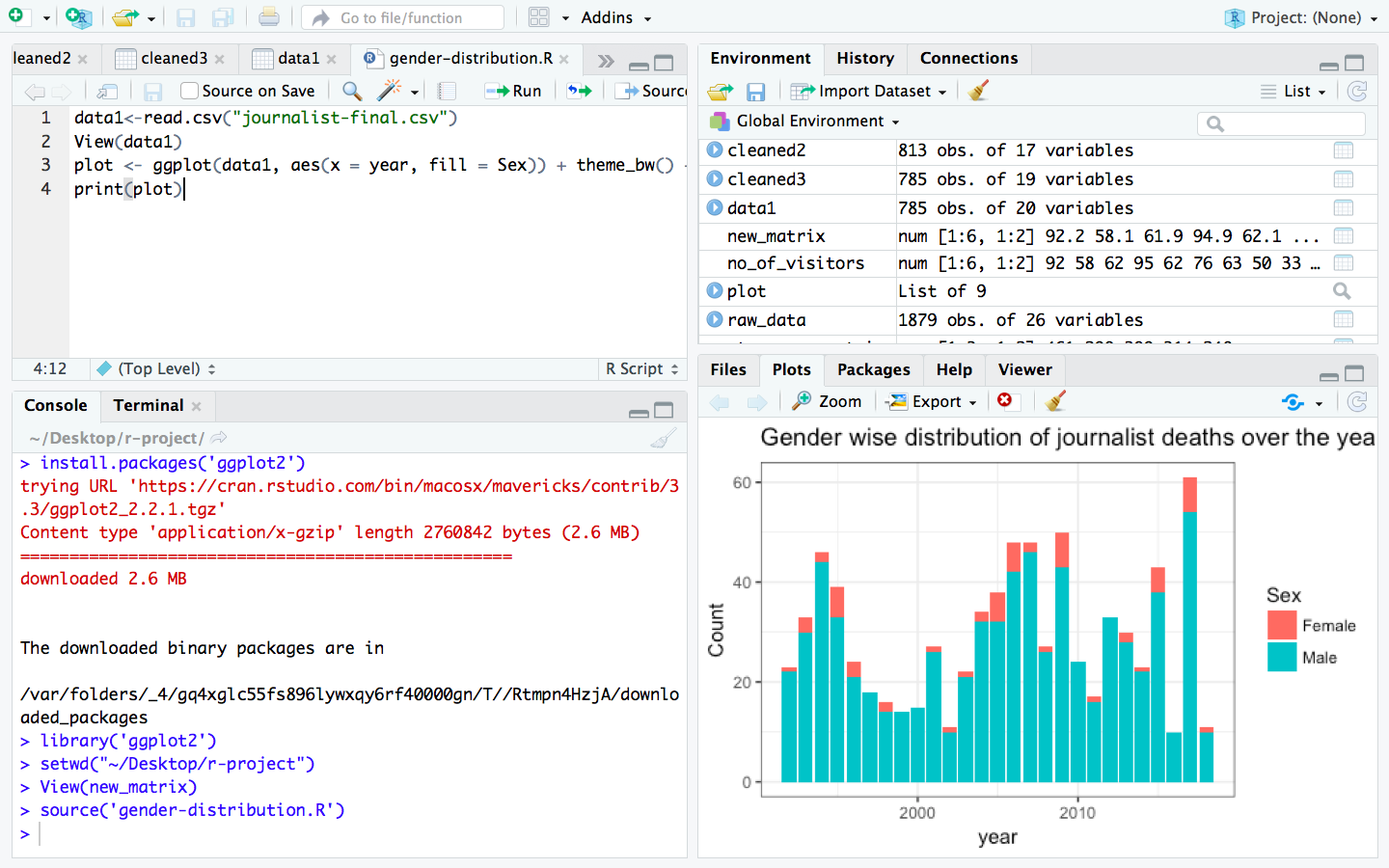
**Question 1:** What is the death rate per year according to gender?

**Description:** According to a [landmark survey](http://www.iwmf.org/our-research/global-report/) in 2011, more than 500 media companies worldwide found that women made up only about one-third of the journalism workforce.[2] Above statement is regarding gender difference in field of journalism, our aim was to check whether there was a difference in number of deaths with respect to gender. Using ggplot2 package and ggplot function we have created following bar chart. The visualization clearly shown that female death count is less as compared to male death. In Year 2017, highest number of journalists’ deaths were reported. Following Bar chart includes death count from the year 1992 to 2018.

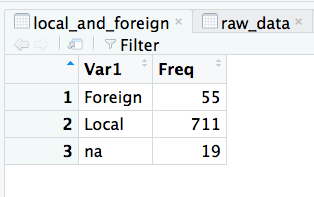
**Visualization Screenshot:**

****

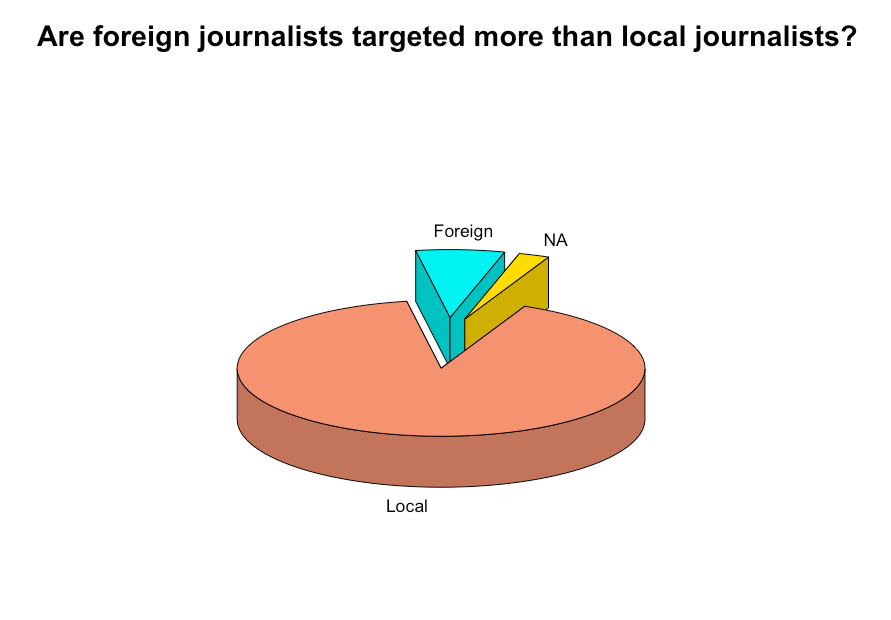
**Code Screenshot of Bar Chart:**



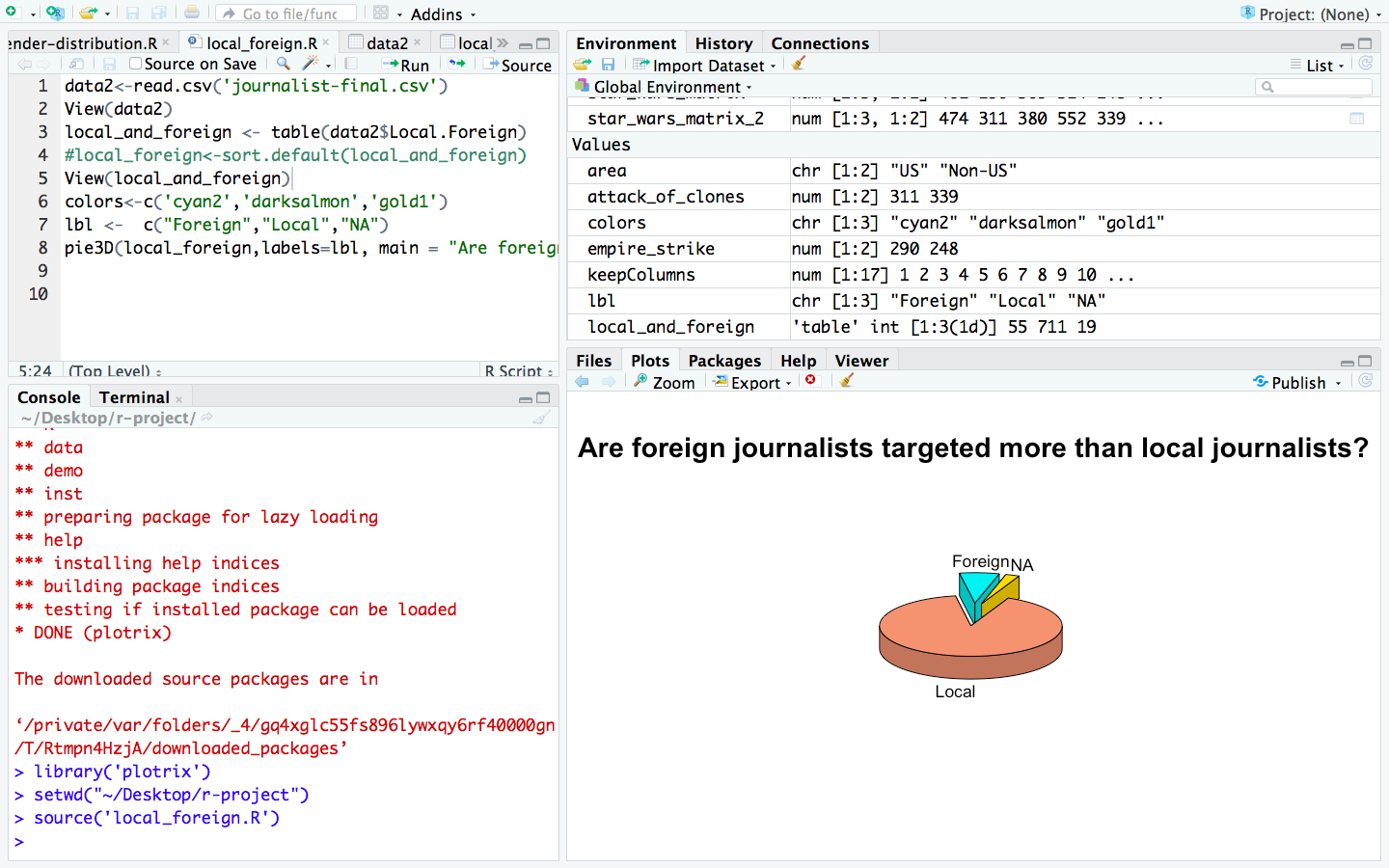
**Question 2: Are foreign journalists targeted more than local journalists?**

**Description:** Dataset contains a column named local.foreign which includes information about whether journalist is local or foreign. By this analysis we can determine that local journalists are targeted more than foreign journalists. We have used pie3D function from the plotrix package for the visualization. The pic chart shows that 80% of the journalists were local. The table on the right shows that 785 journalists were local and only 55 were foreign.

**Visualization Screenshot:**



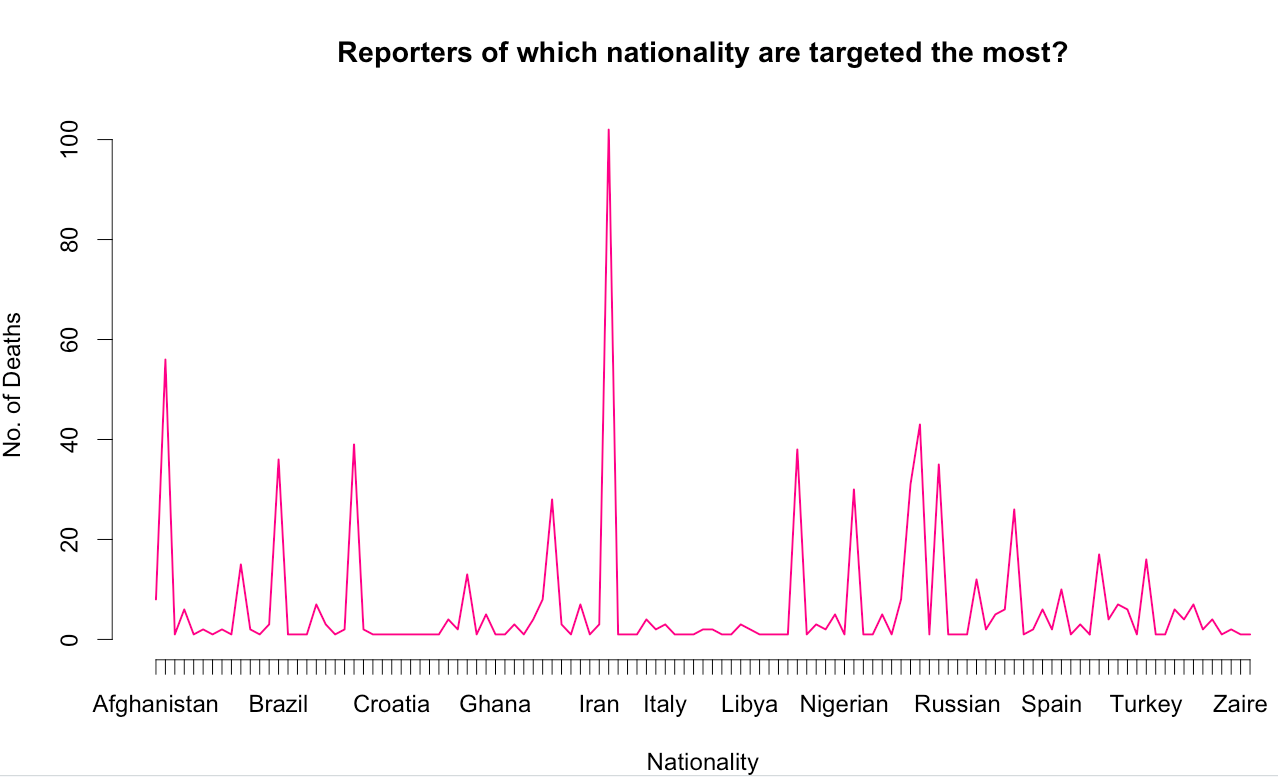
**Code Screenshot of Pie Chart:**



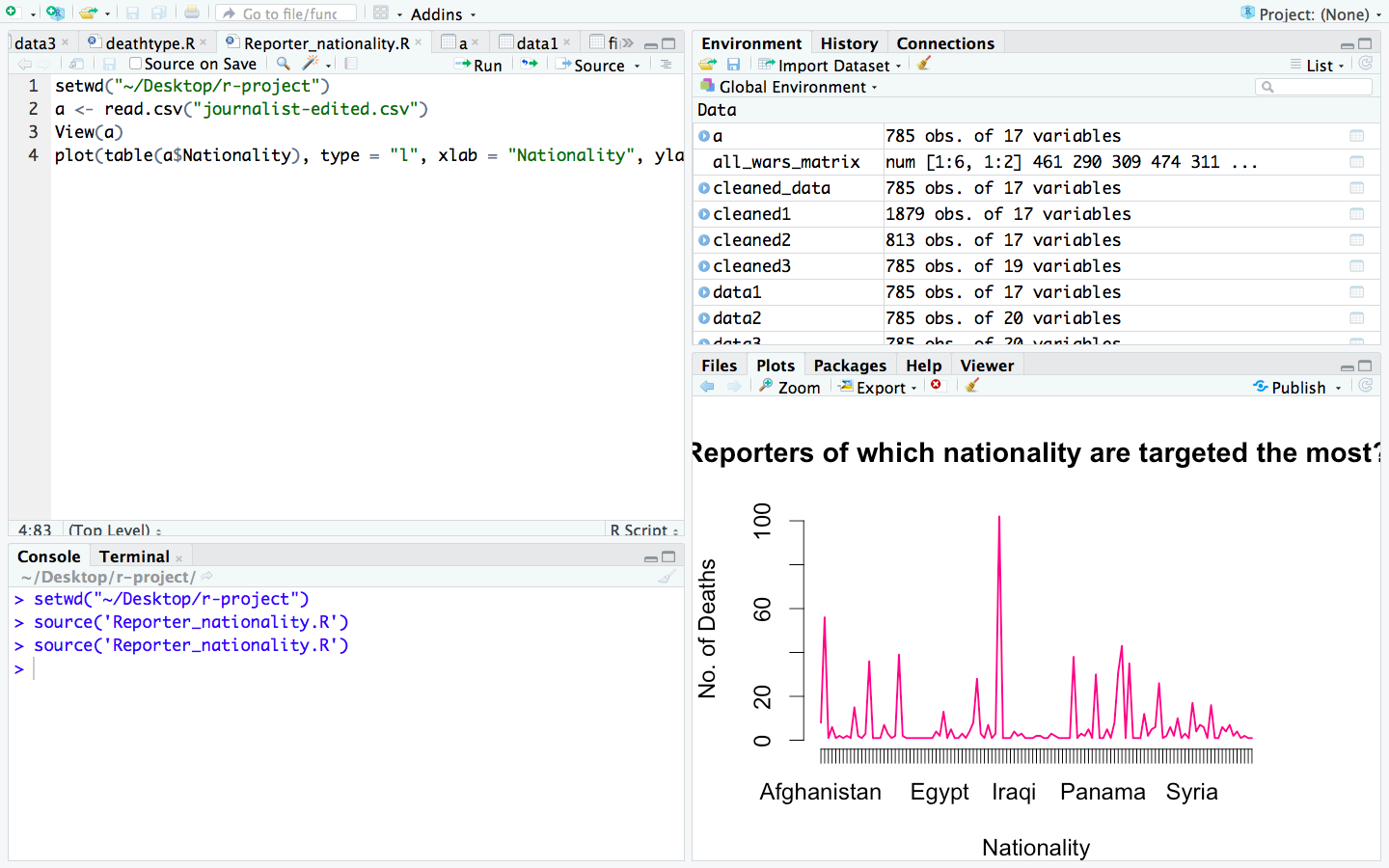
**Question 3: Reporters of which country were targeted the most?**

**Description:** Here we want to analyze which country is the deadliest for both local and foreign journalist. Some reports stated that Iran has the maximum number of journalists in jail than any other country [3]. After analyzing our data, from the Line chart we can confirm that Iran is the deadliest country for journalists. Afghanistan, Brazil, Russian etc. follow Iran for death count. We have used the plot function in order to get following visualization.

**Visualization Screenshot:**

****

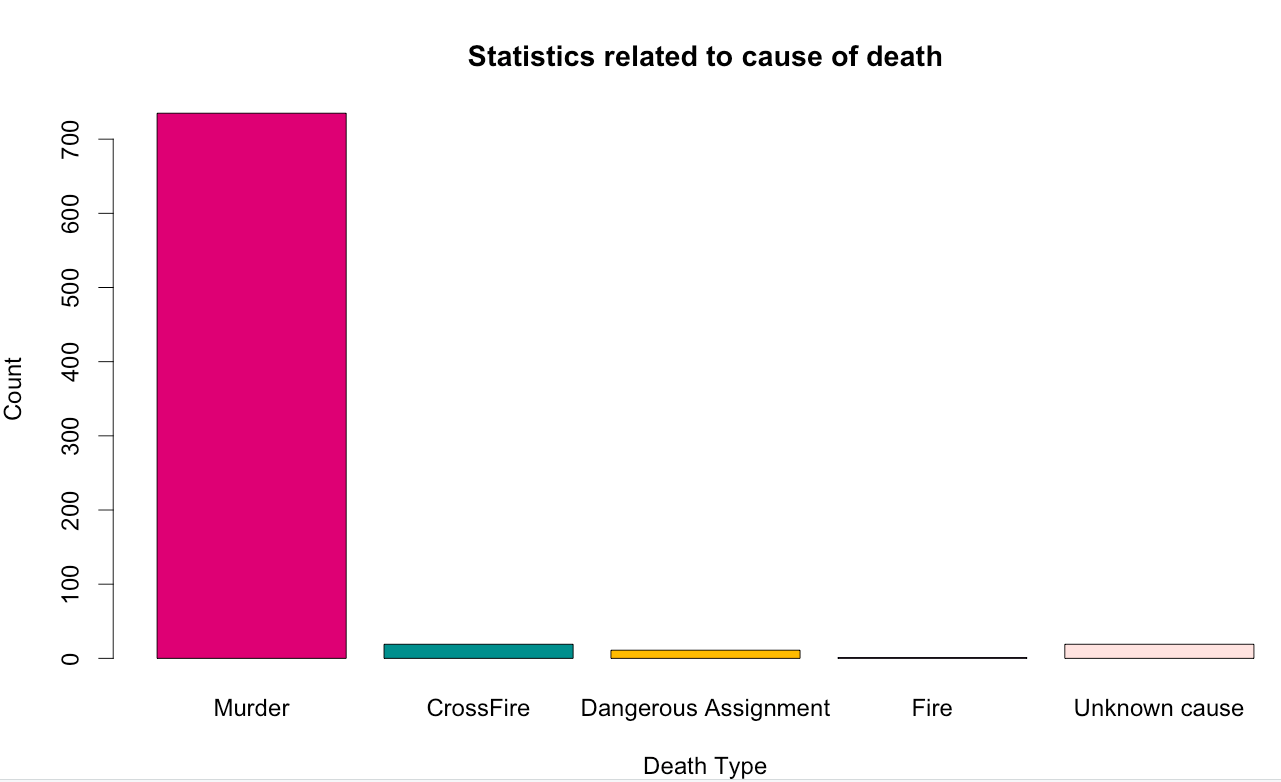
**Code Screenshot of Line Chart:**



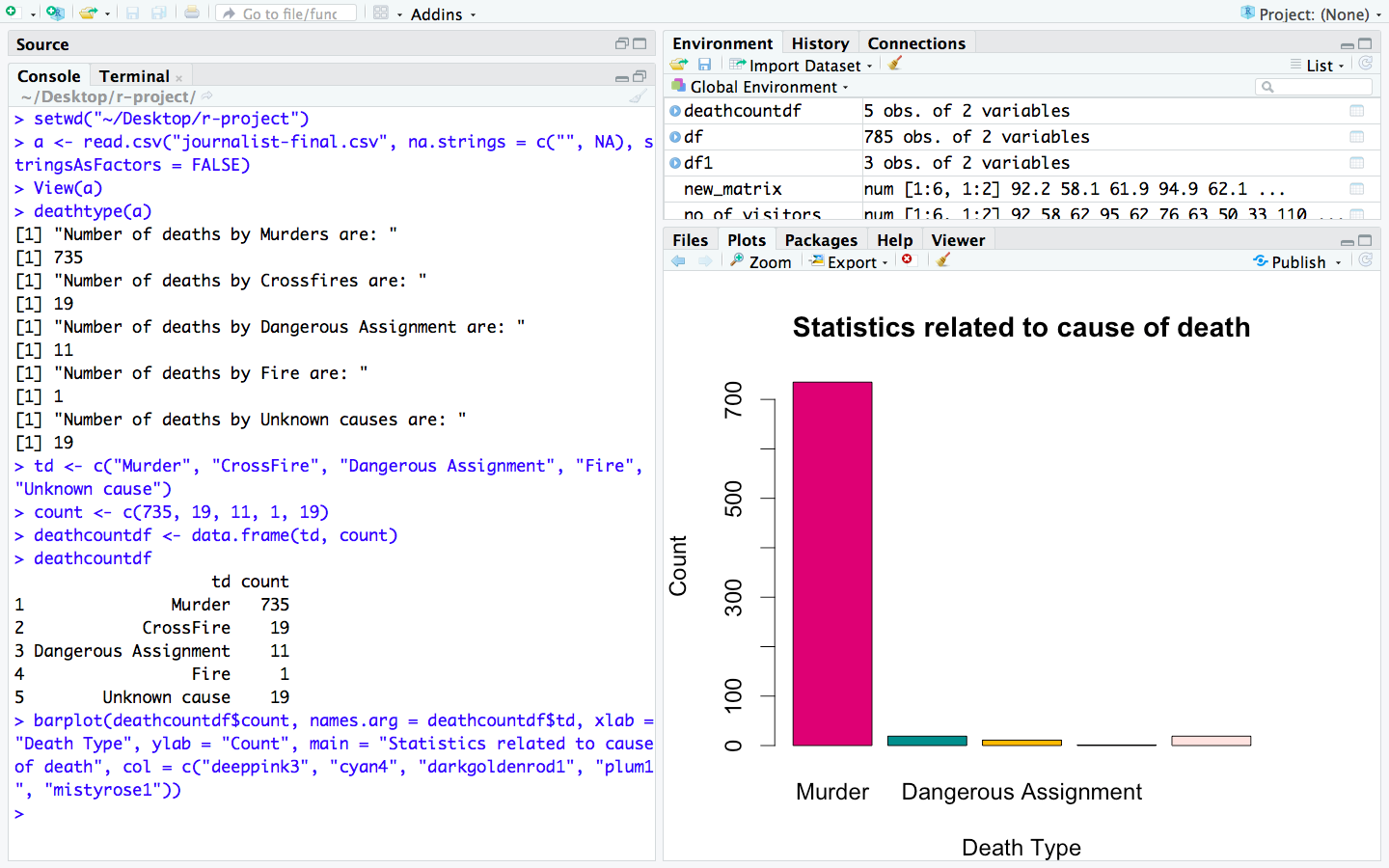
**D. Statistical Summary and Functions:**

**Description of User Defined Function deathtype():** we have created Used Defined Function named deathtype() which is used to calculate the type of death of the journalists. Our dataset contains information about the type of death such are murder, crossfire, dangerous assignment, etc. We have created bar plot after getting all the death count as shown in following visualization. We can see that most of them were murdered on duty.

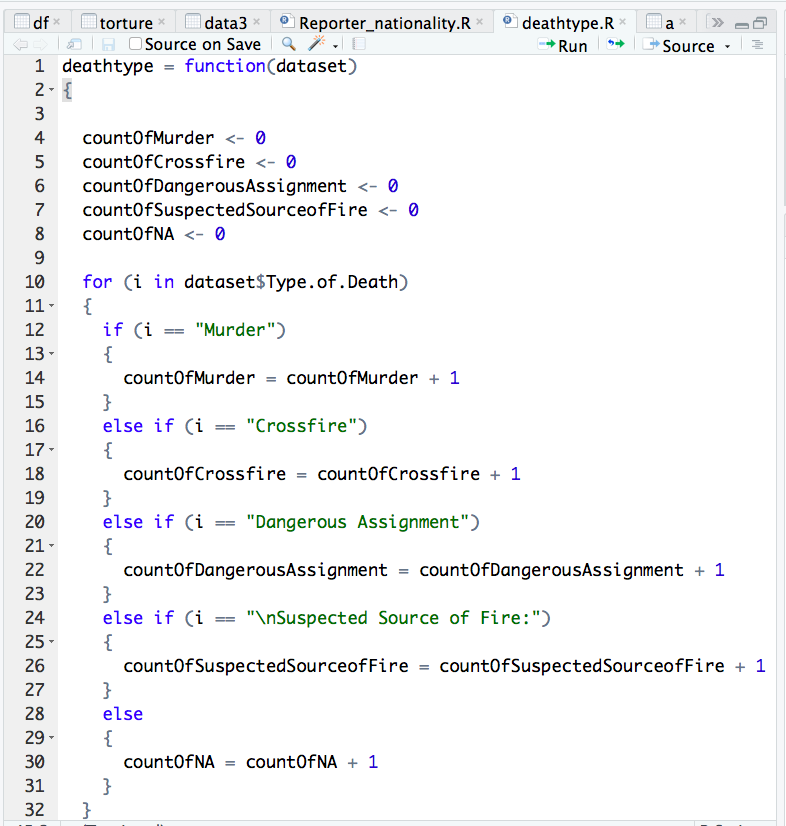
**Visualization Screenshot:**

****

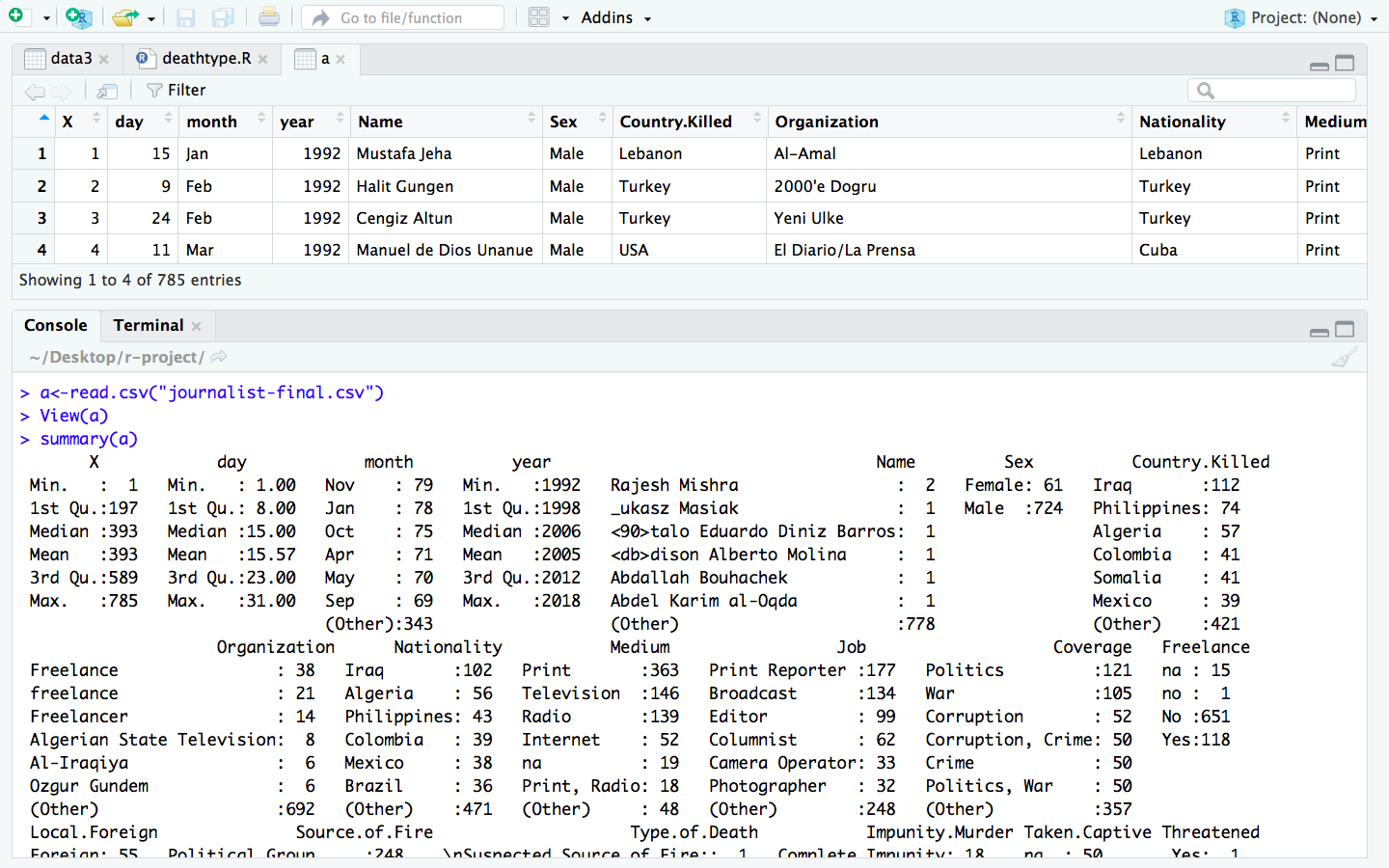
**Console Code Screenshot:**

****

**Code Screenshot of UDF:**

****

**Statistical Summary Screenshot:**

****

**E. Code of entire project:**

**1. Code for Removing unwanted columns:**

> setwd("~/Desktop/r-project")

> raw\_data<-read.csv("journalist.csv", na.strings = c("", NA), stringsAsFactors = FALSE)

> View(raw\_data)

> keepColumns <- c(1,2,3,4,5,6,7,8,9,10,11,12,13,14,15,16,17)

> cleaned1<-raw\_data[,keepColumns]

> View(cleaned1)

**2. Code for removing Null values:**

> setwd("~/Desktop/r-project")

> raw\_data<-read.csv("journalist.csv", na.strings = c("", NA), stringsAsFactors = FALSE)

> View(raw\_data)

> keepColumns <- c(1,2,3,4,5,6,7,8,9,10,11,12,13,14,15,16,17)

> cleaned1<-raw\_data[,keepColumns]

> View(cleaned1)

> cleaned2<-na.omit(cleaned1)

> View(cleaned2)

> write.csv(cleaned2, file = "journalist-edited.csv",row.names=TRUE)

> cleaned\_data<-read.csv("journalist-edited.csv")

> View(cleaned\_data)

**3. Code for splitting column:**

> install.packages('tidyr')

> library('tidyr')

> cleaned\_data<-read.csv("journalist-edited.csv")

> View(cleaned\_data)

> cleaned3<-separate(cleaned\_data, Date, c("day", "month", "year"), sep = "-")

> View(cleaned3)

> write.csv(cleaned3, file = "journalist-final.csv",row.names=TRUE)

**4. Code for Bar Chart(Visualization 1):**

Console Code:

> install.packages('ggplot2')

> library('ggplot2')

> setwd("~/Desktop/r-project")

> View(new\_matrix)

> source('gender-distribution.R')

Script 1: gender-distribution.R

data1<-read.csv("journalist-final.csv")

View(data1)

plot <- ggplot(data1, aes(x = year, fill = Sex)) + theme\_bw() + geom\_bar() + labs(y = "Count", title = "Gender wise distribution of journalist deaths over the years")

print(plot)

**5. Code for Pie Chart(Visualization 2):**

Console Code:

> install.packages('plotrix')

> library('plotrix')

> setwd("~/Desktop/r-project")

> source('local\_foreign.R')

Script 2: local\_foreign.R

data2<-read.csv('journalist-final.csv')

View(data2)

local\_and\_foreign <- table(data2$Local.Foreign)

View(local\_and\_foreign)

colors<-c('cyan2','darksalmon','gold1')

lbl <- c("Foreign","Local","NA")

pie3D(local\_foreign,labels=lbl, main = "Are foreign journalists targeted more than local journalists?", col = colors, explode=0.2, radius=0.8,start= 1.3,labelcex = 1.1)

**6. Code for Line Chart(Visualization 3):**

Console Code:

> setwd("~/Desktop/r-project")

> source('reporter\_nationality.R')

Script 3: reporter\_nationality.R

setwd("~/Desktop/r-project")

setwd("~/Desktop/r-project")

a <- read.csv("journalist-edited.csv")

View(a)

plot(table(a$Nationality), type = "l", xlab = "Nationality", ylab = "No. of Deaths", main = "Reporters of which nationality are targeted the most?", col = "violetred2")

**7. Code for User Defined Function:**

Console Code:

> setwd("~/Desktop/r-project")

> a <- read.csv("journalist-final.csv", na.strings = c("", NA), stringsAsFactors = FALSE)

> View(a)

> deathtype(a)

**#Based on Output of above user defined function, data frame was created and inserted into barplot as following.**

> td <- c("Murder", "CrossFire", "Dangerous Assignment", "Fire", "Unknown cause")

> count <- c(735, 19, 11, 1, 19)

> deathcountdf <- data.frame(td, count)

> deathcountdf

> barplot(deathcountdf$count, names.arg = deathcountdf$td, xlab = "Death Type", ylab = "Count", main = "Statistics related to cause of death", col = c("deeppink3", "cyan4", "darkgoldenrod1", "plum1", "mistyrose1"))

Script 4:deathtype()

deathtype = function(dataset)

{

countOfMurder <- 0

countOfCrossfire <- 0

countOfDangerousAssignment <- 0

countOfSuspectedSourceofFire <- 0

countOfNA <- 0

for (i in dataset$Type.of.Death)

{

if (i == "Murder")

{

countOfMurder = countOfMurder + 1

}

else if (i == "Crossfire")

{

countOfCrossfire = countOfCrossfire + 1

}

else if (i == "Dangerous Assignment")

{

countOfDangerousAssignment = countOfDangerousAssignment + 1

}

else if (i == "\nSuspected Source of Fire:")

{

countOfSuspectedSourceofFire = countOfSuspectedSourceofFire + 1

}

else

{

countOfNA = countOfNA + 1

}

}

print("Number of deaths by Murders are: ")

print(countOfMurder)

print("Number of deaths by Crossfires are: ")

print(countOfCrossfire)

print("Number of deaths by Dangerous Assignment are: ")

print(countOfDangerousAssignment)

print("Number of deaths by Fire are: ")

print(countOfSuspectedSourceofFire)

print("Number of deaths by Unknown causes are: ")

print(countOfNA)

}

**References:**

[1] (2017, February 21). Journalists face rising threats in trouble spots: Watchdog. Retrieved from <http://www.dailymail.co.uk/wires/afp/article-4245608/Journalists-face-rising-threats-trouble-spots-watchdog.html>

[2] Gender disparities and journalism: Research perspectives. (2015, June 08). Retrieved from <https://journalistsresource.org/studies/society/news-media/female-journalists-media-sexism-emerging-trends>

[3] Rights group: Iran jails most journalists. (2010, January 06). Retrieved from http://www.cnn.com/2010/WORLD/meast/01/06/iran.press.freedom/index.html