

A thick dark blue vertical bar runs down the left side of the page. A dark red arrow points to the right from the bar, containing the text 'R Project on'. Below the bar, several thin, curved lines in shades of blue and grey sweep upwards and to the right.

R Project on

Analysis of journalists Deaths

Since 1992

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

or.Murder.	Taken.Captive	Threatened	Tortured	X	X.1	X.2	X.3	X.4	X.5
	NA	NA		NA	NA	NA	NA	NA	NA
	Yes	No	No	NA	NA	NA	NA	NA	NA
	No	No	No	NA	NA	NA	NA	NA	NA
	No	No	No	NA	NA	NA	NA	NA	NA
	No	No	No	NA	NA	NA	NA	NA	NA
	No	Yes	No	NA	NA	NA	NA	NA	NA
	NA	NA		NA	NA	NA	NA	NA	NA
	No	No	No	NA	NA	NA	NA	NA	NA

Showing 1 to 9 of 1,879 entries

After:

.Fire	Type.of.Death	Impunity..for.Murder.	Taken.Captive	Threatened	Tortured
	NA	NA	NA	NA	
nt Officials	Dangerous Assignment	NA	Yes	No	No
roup	Murder	Yes	No	No	No
fficials	Crossfire/Combat-Related	NA	No	No	No
nt Officials	Murder	Yes	No	No	No
ry Group	Murder	Yes	No	Yes	No
	NA	NA	NA	NA	
Group	Murder	Partial	No	No	No

Showing 1 to 9 of 1,879 entries

Code for Removing Unwanted Columns:

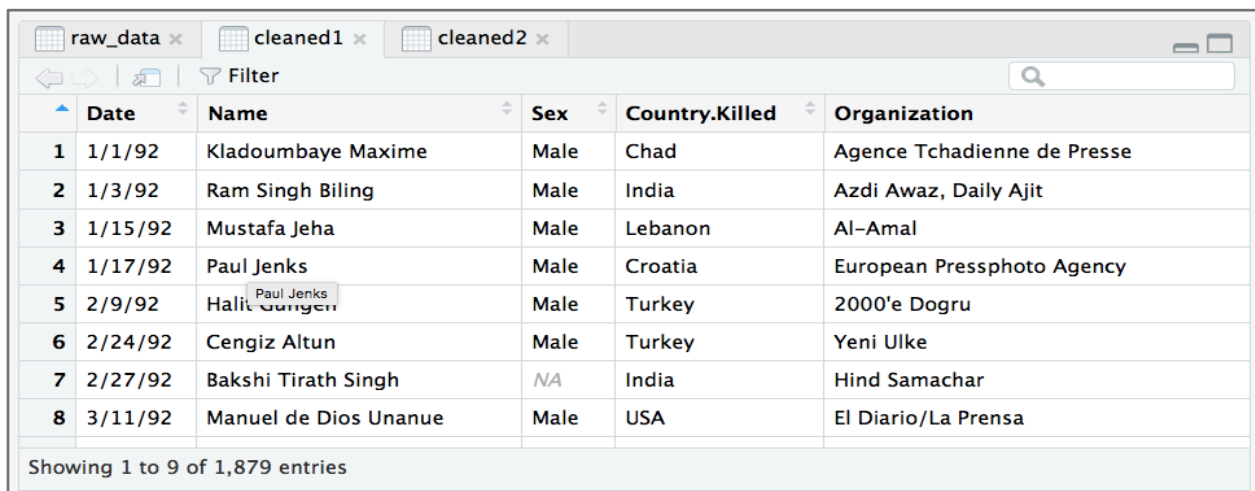
```
[Workspace loaded from ~/.RData]

> 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. Removing null values:

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

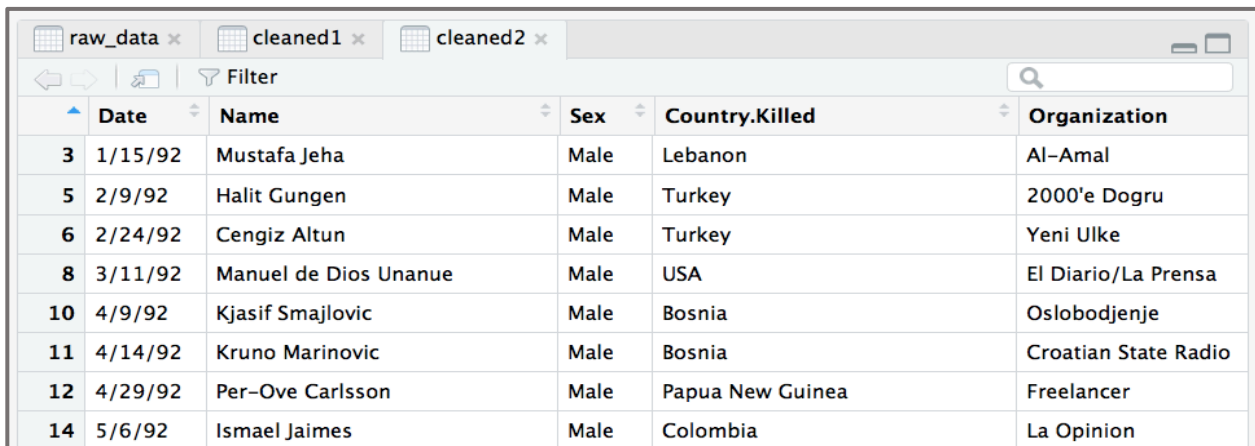
Before:



	Date	Name	Sex	Country.Killed	Organization
1	1/1/92	Kladoumbaye Maxime	Male	Chad	Agence Tchadienne de Presse
2	1/3/92	Ram Singh Biling	Male	India	Azdi Awaz, Daily Ajit
3	1/15/92	Mustafa Jeha	Male	Lebanon	Al-Amal
4	1/17/92	Paul Jenks	Male	Croatia	European Pressphoto Agency
5	2/9/92	Halit Gungen	Male	Turkey	2000'e Dogru
6	2/24/92	Cengiz Altun	Male	Turkey	Yeni Ulke
7	2/27/92	Bakshi Tirath Singh	NA	India	Hind Samachar
8	3/11/92	Manuel de Dios Unanue	Male	USA	El Diario/La Prensa

Showing 1 to 9 of 1,879 entries

After:



	Date	Name	Sex	Country.Killed	Organization
3	1/15/92	Mustafa Jeha	Male	Lebanon	Al-Amal
5	2/9/92	Halit Gungen	Male	Turkey	2000'e Dogru
6	2/24/92	Cengiz Altun	Male	Turkey	Yeni Ulke
8	3/11/92	Manuel de Dios Unanue	Male	USA	El Diario/La Prensa
10	4/9/92	Kjasif Smajlovic	Male	Bosnia	Oslobodjenje
11	4/14/92	Kruno Marinovic	Male	Bosnia	Croatian State Radio
12	4/29/92	Per-Ove Carlsson	Male	Papua New Guinea	Freelancer
14	5/6/92	Ismael Jaimes	Male	Colombia	La Opinion

Code for Removing Null Values:

```
Console Terminal x
~/Desktop/r-project/ ↗
> 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. 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:

	Sex	Country.Killed	Organization
1	Male	Lebanon	Al-Amal
2	Male	Turkey	2000'e Dogru
3	Male	Turkey	Yeni Ulke
4	Male	USA	El Diario/La Prensa
5	Male	Bosnia	Oslobodjenje
6	Male	Bosnia	Croatian State Rad

After:

	day	month	year	Name	Sex	Country.Killed	O
1	15	Jan	1992	Mustafa Jeha	Male	Lebanon	Al
2	9	Feb	1992	Halit Gungen	Male	Turkey	20
3	24	Feb	1992	Cengiz Altun	Male	Turkey	Ye
4	11	Mar	1992	Manuel de Dios Unanue	Male	USA	El
5	9	Apr	1992	Kjasif Smajlovic	Male	Bosnia	Os
6	14	Apr	1992	Kruno Marinovic	Male	Bosnia	Cr

Code for splitting Columns:

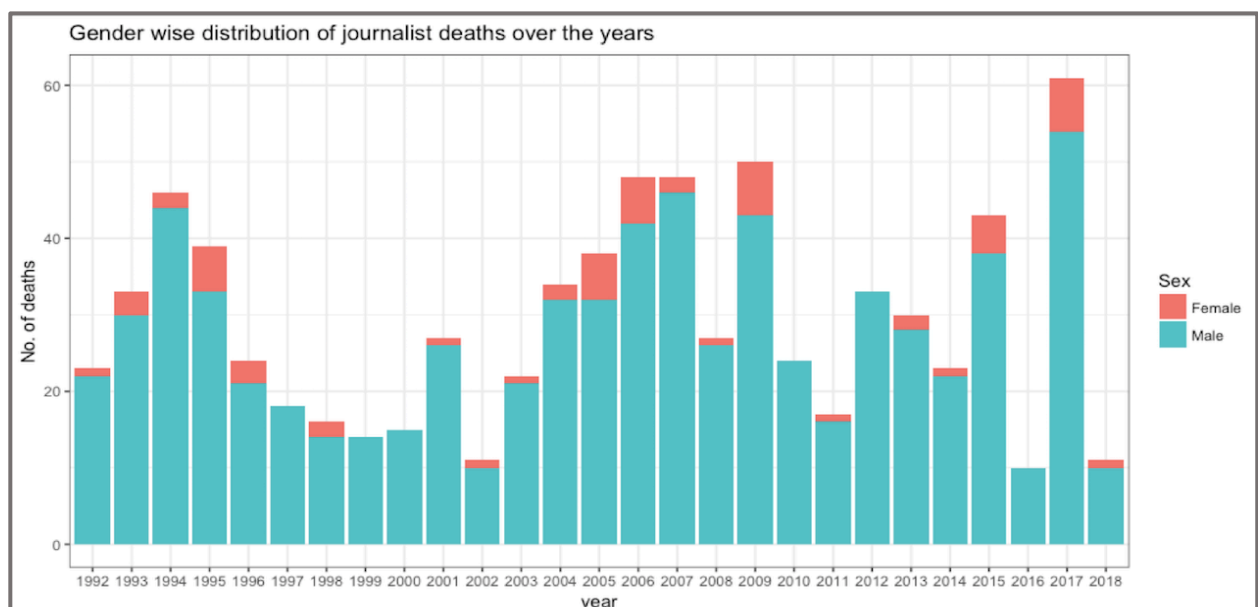
```
Console Terminal x
~/Desktop/r-project/
> library('tidyr')
> cleaned_data<-read.csv("journalist-edited.csv")
> View(cleaned_data)
> cleaned3<-separate(cleaned_data, Date, c("day", "month", "year"), sep = "-")
> View(cleaned3)
>
```

C. Analysis & Visualizations:

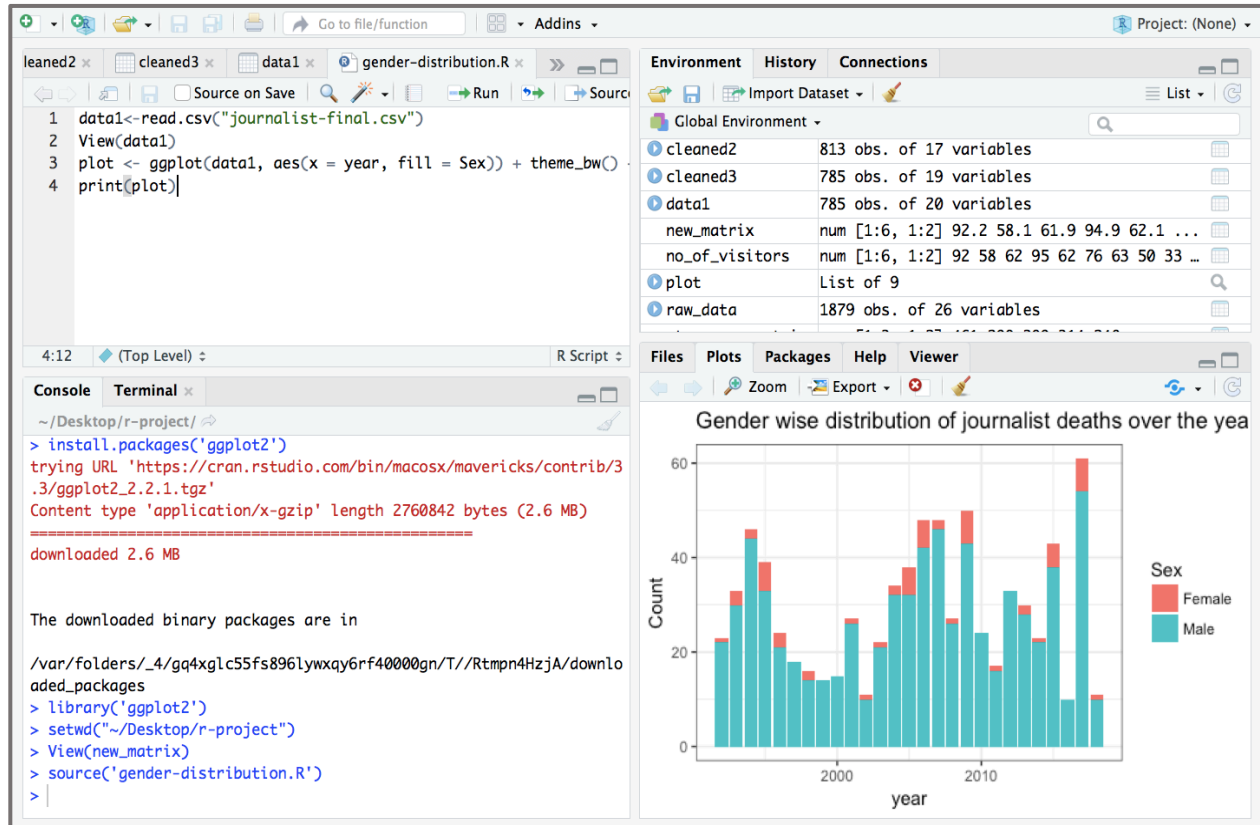
Question 1: What is the death rate per year according to gender?

Description: According to a landmark survey 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:

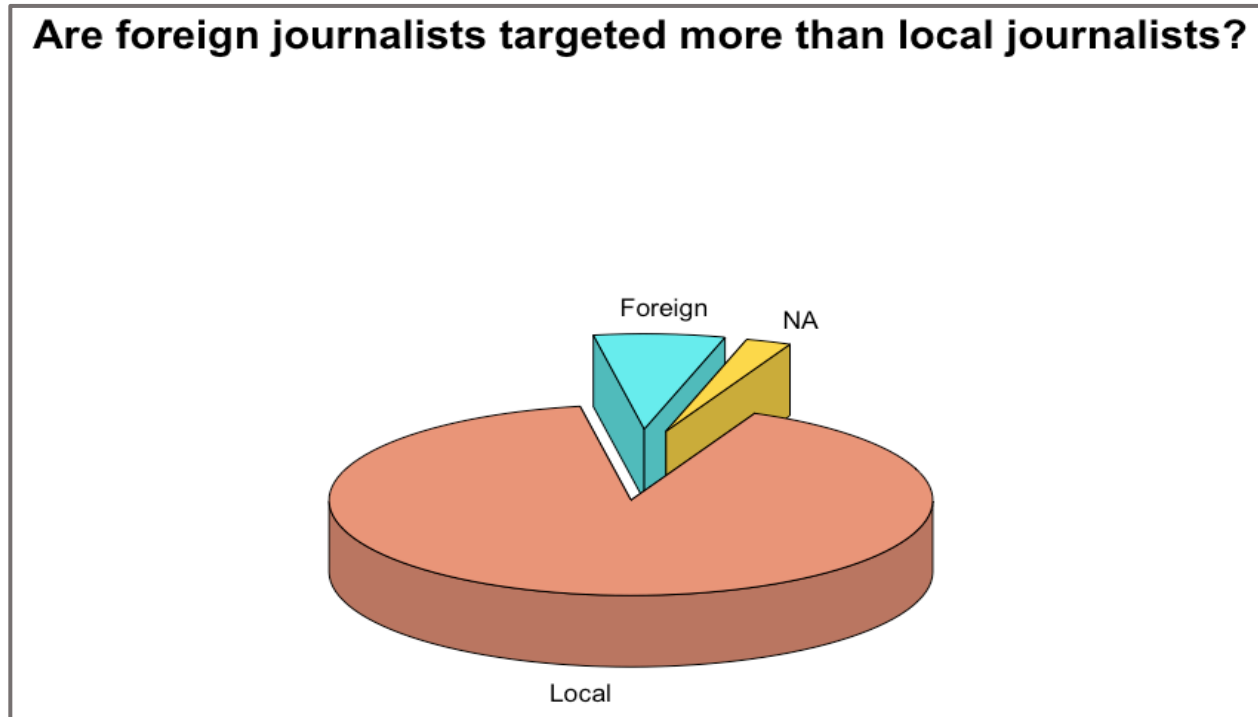


Question 2: Are foreign journalists targeted more than local journalists?

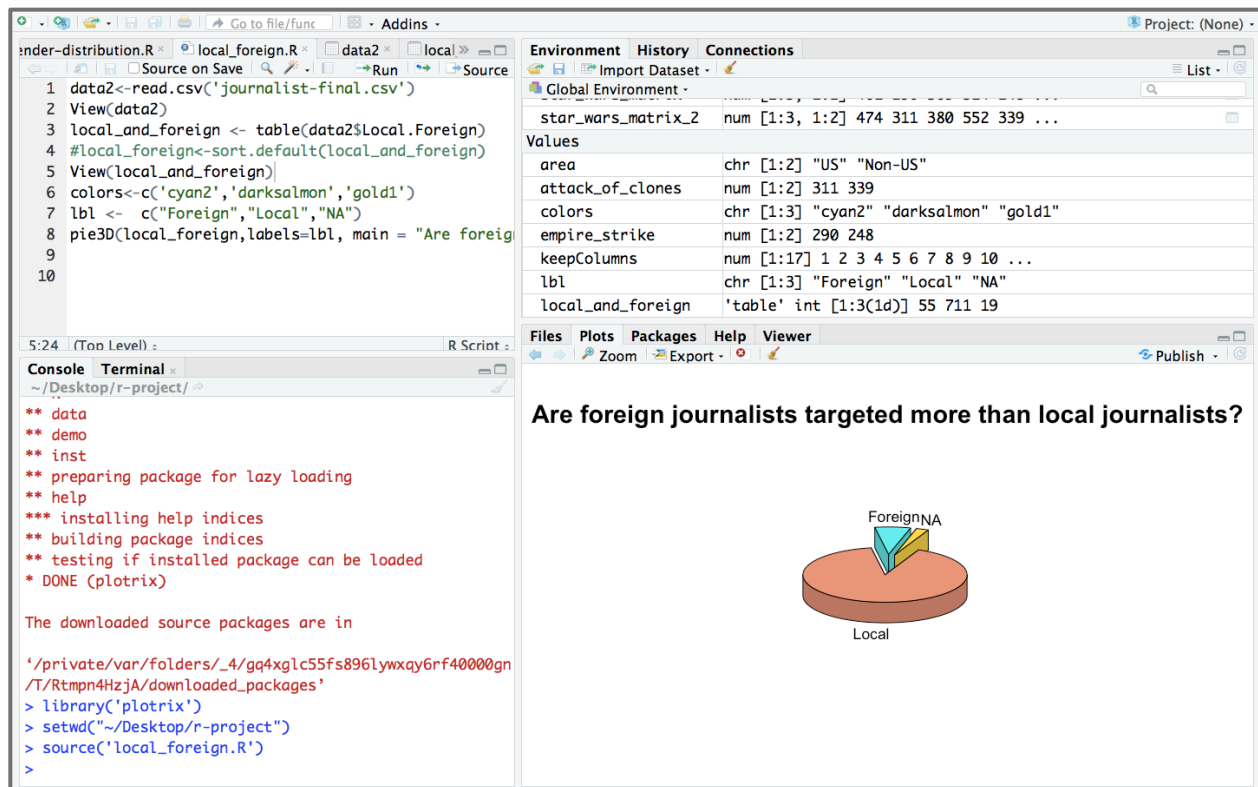
Description: Dataset contains a column named `local.foreign` which includes information about whether journalist is local or foreign. By this analysis we can determine that local journalists are targeted more than foreign journalists. We have used `pie3D` function from the `plotrix` package for the visualization. The pie 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.

local_and_foreign		
Filter		
	Var1	Freq
1	Foreign	55
2	Local	711
3	na	19

Visualization Screenshot:



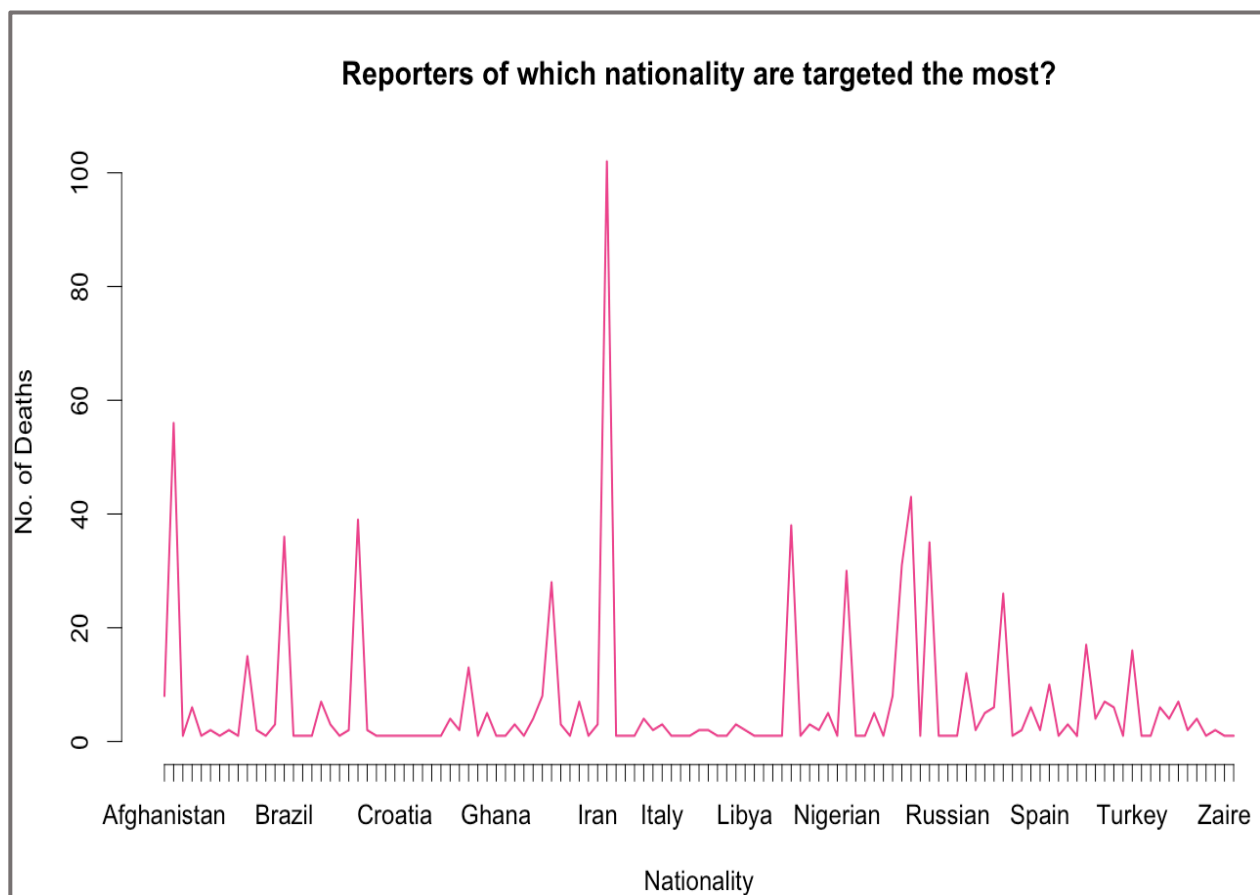
Code Screenshot of Pie Chart:



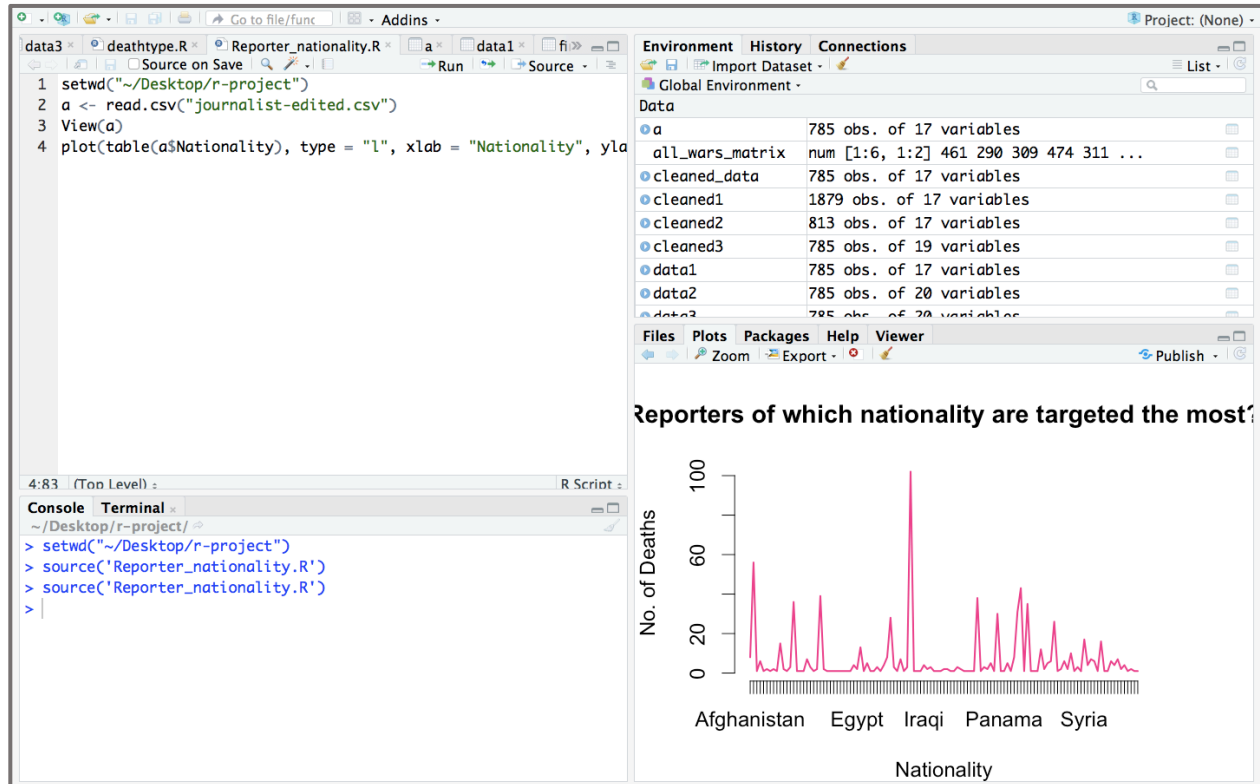
Question 3: Reporters of which country were targeted the most?

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:



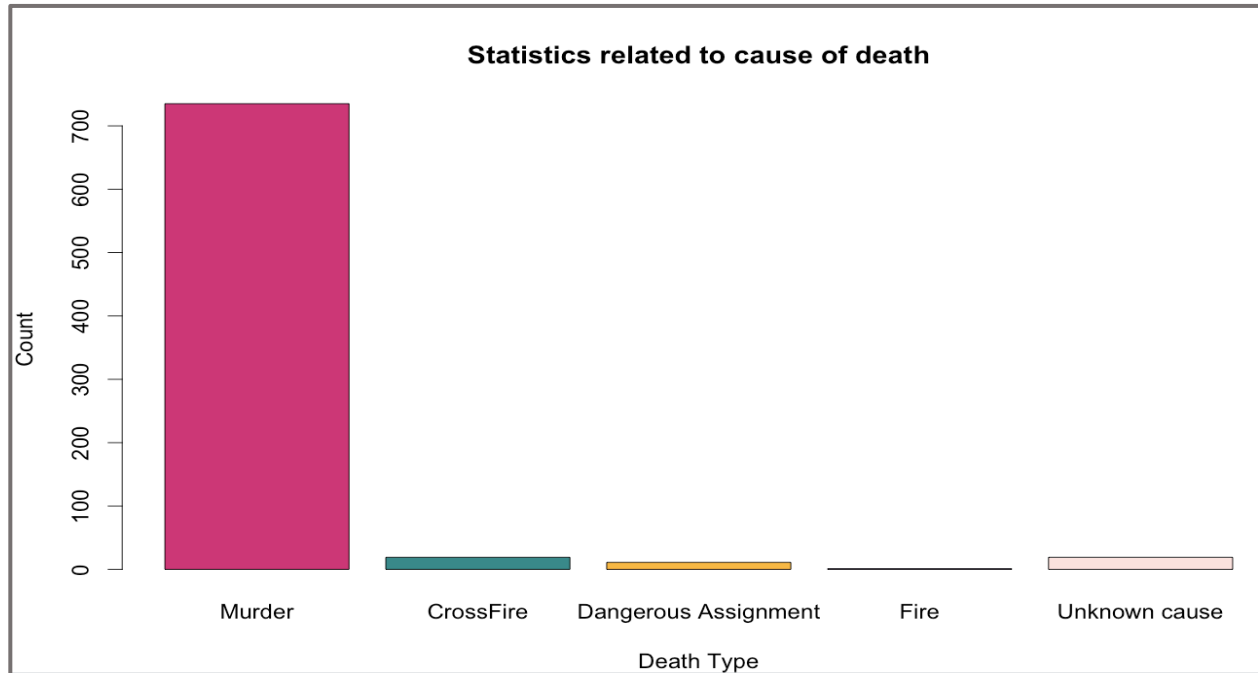
Code Screenshot of Line Chart:



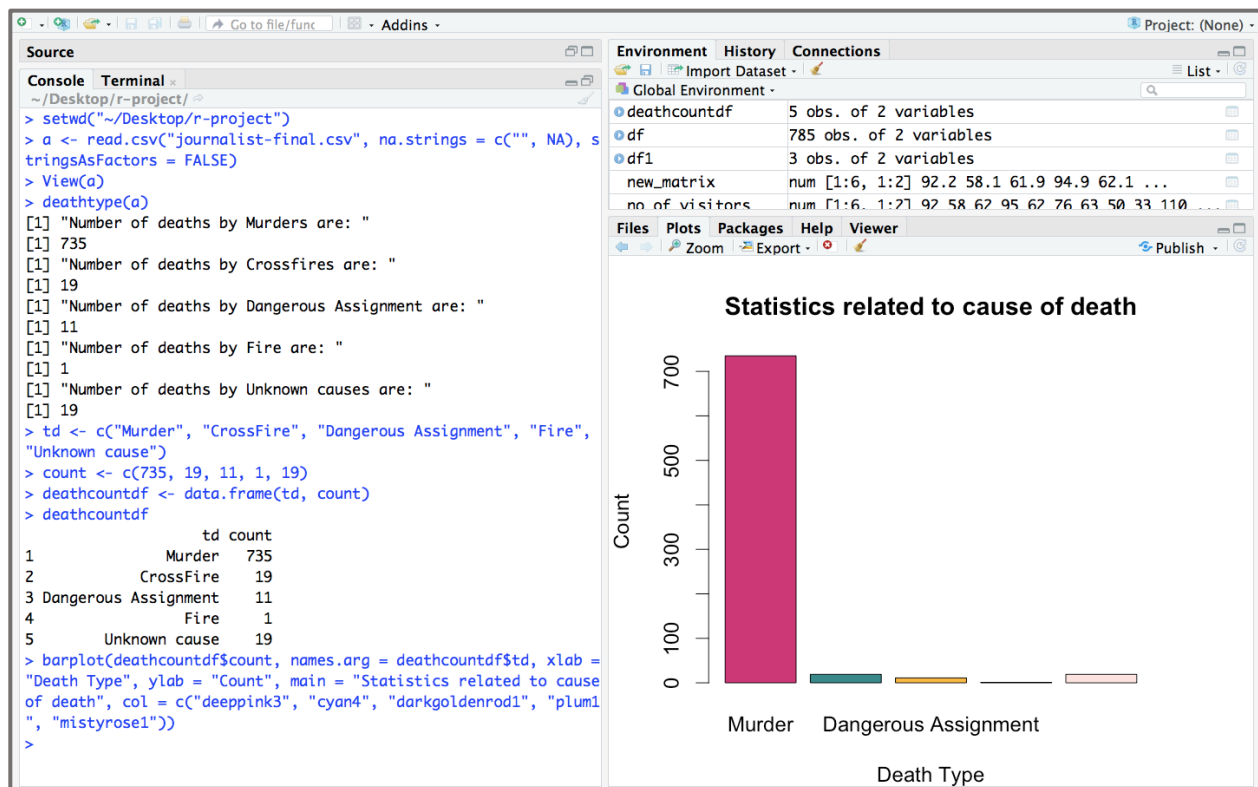
D. Statistical Summary and Functions:

Description of User Defined Function deathtype(): we have created User 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 as 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.

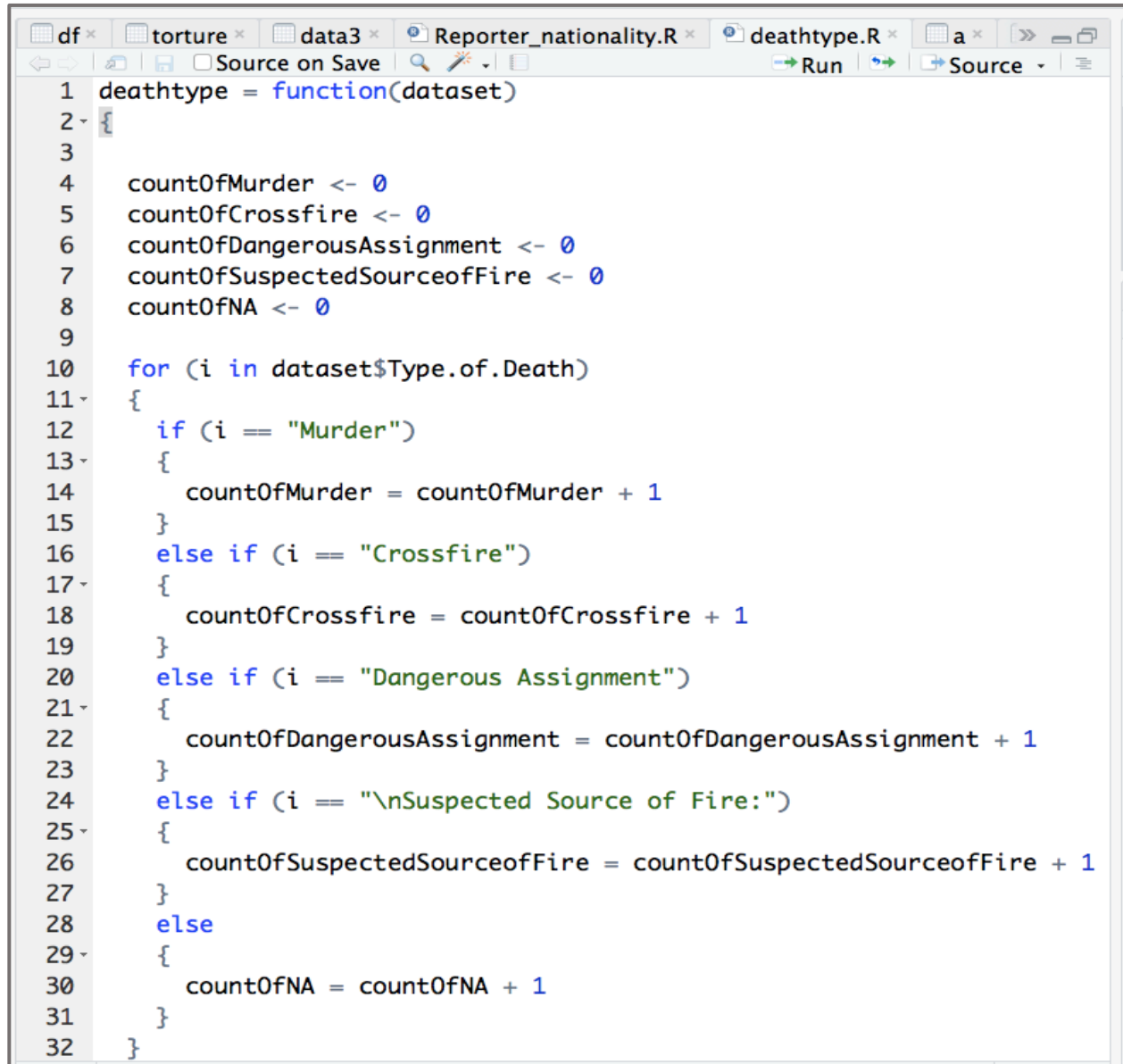
Visualization Screenshot:



Console Code Screenshot:



Code Screenshot of UDF:

The image shows a screenshot of an R Studio editor window. The title bar at the top displays several open files: 'df', 'torture', 'data3', 'Reporter_nationality.R', 'deathtype.R', and 'a'. The 'deathtype.R' file is the active one. The editor area contains R code for a function named 'deathtype' that takes a 'dataset' as input. The function initializes five counters to zero: 'countOfMurder', 'countOfCrossfire', 'countOfDangerousAssignment', 'countOfSuspectedSourceofFire', and 'countOfNA'. It then iterates through 'dataset\$Type.of.Death' using a 'for' loop. Inside the loop, it uses 'if' and 'else if' statements to check the death type and increment the corresponding counter by 1. The possible death types are 'Murder', 'Crossfire', 'Dangerous Assignment', and '\nSuspected Source of Fire:'. The loop ends with a closing brace. The code is as follows:

```
1 deathtype = function(dataset)
2 {
3
4   countOfMurder <- 0
5   countOfCrossfire <- 0
6   countOfDangerousAssignment <- 0
7   countOfSuspectedSourceofFire <- 0
8   countOfNA <- 0
9
10  for (i in dataset$Type.of.Death)
11  {
12    if (i == "Murder")
13    {
14      countOfMurder = countOfMurder + 1
15    }
16    else if (i == "Crossfire")
17    {
18      countOfCrossfire = countOfCrossfire + 1
19    }
20    else if (i == "Dangerous Assignment")
21    {
22      countOfDangerousAssignment = countOfDangerousAssignment + 1
23    }
24    else if (i == "\nSuspected Source of Fire:")
25    {
26      countOfSuspectedSourceofFire = countOfSuspectedSourceofFire + 1
27    }
28    else
29    {
30      countOfNA = countOfNA + 1
31    }
32  }
```

Statistical Summary Screenshot:

data3 x deathtype.R x a x

Filter

	X	day	month	year	Name	Sex	Country.Killed	Organization	Nationality	Medium
1	1	15	Jan	1992	Mustafa Jeha	Male	Lebanon	Al-Amal	Lebanon	Print
2	2	9	Feb	1992	Halit Gungen	Male	Turkey	2000'e Dogru	Turkey	Print
3	3	24	Feb	1992	Cengiz Altun	Male	Turkey	Yeni Ulke	Turkey	Print
4	4	11	Mar	1992	Manuel de Dios Unanue	Male	USA	El Diario/La Prensa	Cuba	Print

Showing 1 to 4 of 785 entries

Console Terminal x

~/Desktop/r-project/

```
> a<-read.csv("journalist-final.csv")
> View(a)
> summary(a)
```

X: 1, day: 1.00, month: 79, year: 1992, Name: Rajesh Mishra, Sex: 2, Country.Killed: 112, 1st Qu.:197, 1st Qu.: 8.00, Jan: 78, 1st Qu.:1998, Lukasz Masiak: 1, Male: 724, Philippines: 74, Median:393, Median:15.00, Oct: 75, Median:2006, <90>talo Eduardo Diniz Barros: 1, Algeria: 57, Mean:393, Mean:15.57, Apr: 71, Mean:2005, <db>dison Alberto Molina: 1, Colombia: 41, 3rd Qu.:589, 3rd Qu.:23.00, May: 70, 3rd Qu.:2012, Abdallah Bouhachek: 1, Somalia: 41, Max.:785, Max.:31.00, Sep: 69, Max.:2018, Abdel Karim al-Oqda: 1, Mexico: 39, (Other):343, (Other):778, (Other):421, Organization: 38, Iraq: 102, Print: 363, Print Reporter: 177, Politics: 121, na: 15, freelance: 21, Algeria: 56, Television: 146, Broadcast: 134, War: 105, no: 1, Freelancer: 14, Philippines: 43, Radio: 139, Editor: 99, Corruption: 52, No: 651, Algerian State Television: 8, Colombia: 39, Internet: 52, Columnist: 62, Corruption, Crime: 50, Yes: 118, Al-Iraqiya: 6, Mexico: 38, na: 19, Camera Operator: 33, Crime: 50, Ozgur Gundem: 6, Brazil: 36, Print, Radio: 18, Photographer: 32, Politics, War: 50, (Other): 692, (Other): 471, (Other): 48, (Other): 248, (Other): 357, Local.Foreign: Source.of.Fire: Type.of.Death: Impunity.Murder Taken.Captive Threatened, Foreign: 55, Political Group: 248, \nSuspected Source of Fire: 1, Complete Impunity: 18, na: 50, Yes: 1

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>