Probability and Statistics (IE 6200) - Sec 7 - Final Project Report

Omkar Kalange, Ayush Pandey

Analysis of Missing Migrants across Geographical Borders

1. Objective

This project is to shed some light into deaths of hundreds of migrants that are fleeing from countries engulfed in warfare and conflicts, or simply trying to find a better place to live in. Hopefully, we will get a better understanding of the circumstances of their deaths and get to know overall patterns from that tragedies. We hope that monitoring these tragic events and studying the data relating to them can help to avoid a recurrence of these. We are all human beings.

2. Dataset Description

The dataset comes from the Missing Migrant Project that tracks the deaths of migrants that have perished on their way to their destination. The dataset contains observations that have been collected from all over the world. The description of the variables, you can find below.

Variable name	Description
Web ID	An automatically generated ID of WEB
Region of accident	The region in which an incident took place
Reported date	Estimated date of death or finding a body
Reported year	The year in which the incident occurred
Reported month	The month in which the incident occurred
Number dead	The total number of people confirmed dead in one incident
Minimum Estimated Number of Missing	The total number of those who are missing and are thus assumed to be dead (minimum)
Total dead and missing	The sum of the 'number dead' and 'Minimum Estimated Number of Missing' variables
Number of Survivors	The number of survivors
Number of Females	The number of females
Number of Males	The number of males
Number of children	The number of children
Cause of Death	The cause of death
Location Description	Where the incident happened
Information Source	The source of information
Location Coordinates	The Coordinates of location
Migration Route	The route of migration
URL	The URL of a site where information about a given incident is written
UNSD Geographical Grouping	The grouping of incidents based on their geo-location made by the UNSD
Source Quality	How much a source of information is reliable

Fig 1: Data set and associated columns.

3. Assumption

Missing Migrants Project records deaths of migrants, including refugees and asylum-seekers, who have died or gone missing in the process of migration towards an international destination. These data represent minimum estimates, as many deaths during migration go unrecorded. We have dealt with missing values in the entire process.

4. Data Visualization

4.1 Causes of deaths at different geographical regions

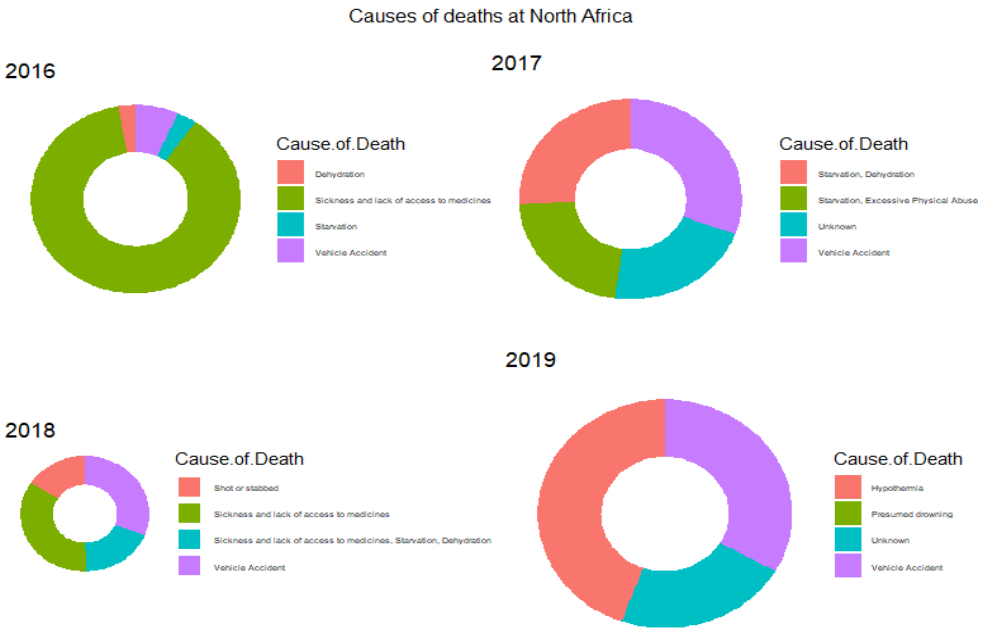


Fig 2: Causes of death at North Africa

It can be observed that -

1. Most of the people died due to sickness in 2016.
2. More deaths were due to vehicle accidents and drowning in 2017.
3. In 2019, the major cause of death was Hypothermia.

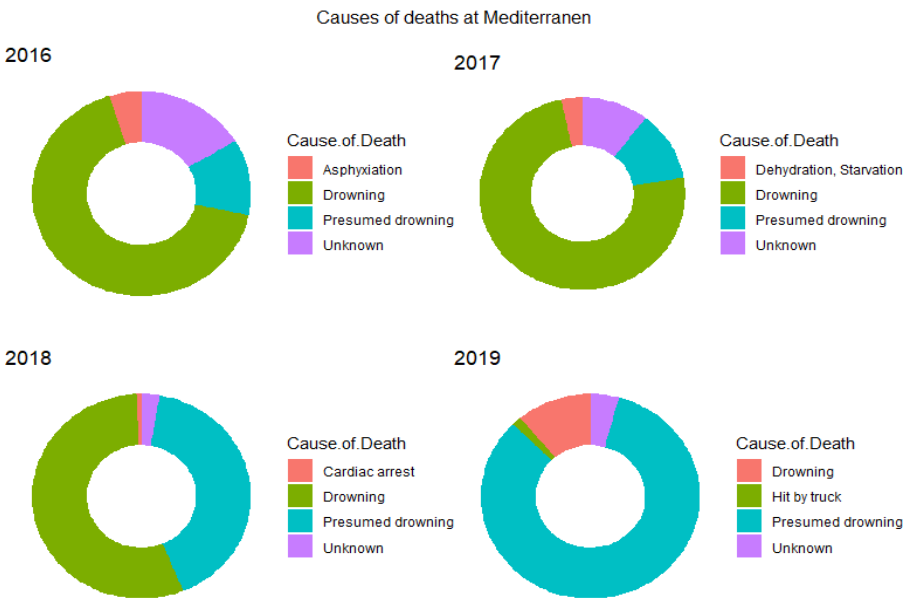


Fig 3: Causes of death at Mediterranean

It can be implied from this graph that most of the deaths at Mediterranean occurred due to drowning in all those years, this is because of presence of oceans and seas at the migration routes.

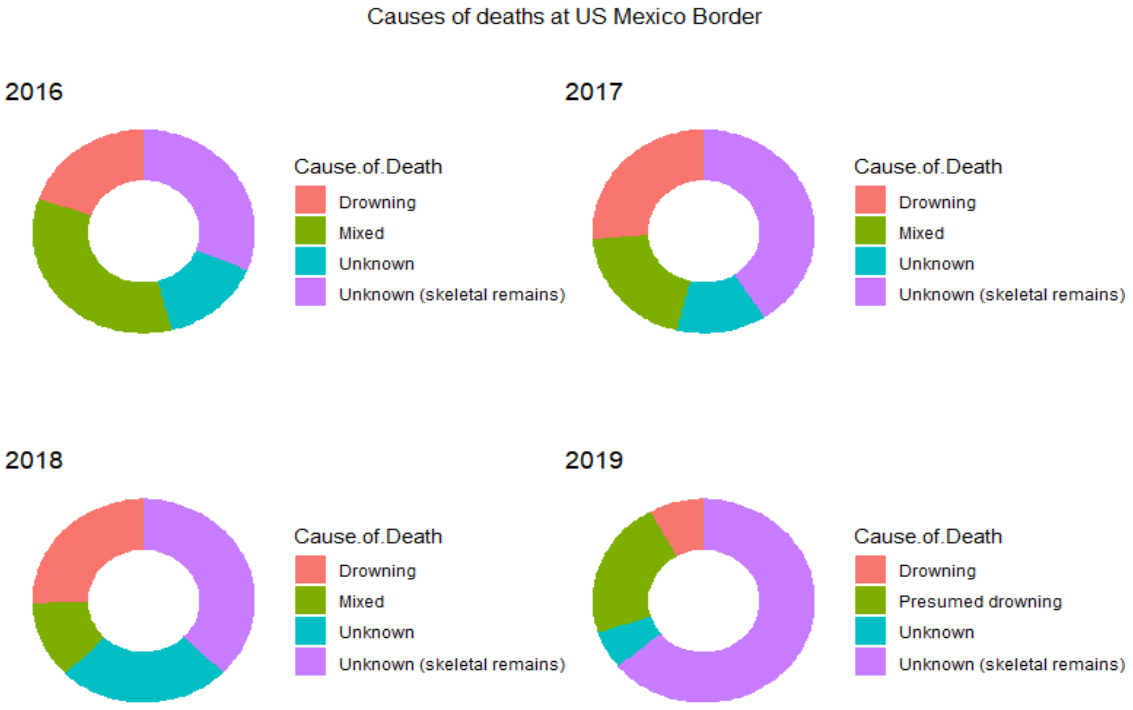


Fig 4: Causes of deaths at US-Mexico Border

It can be inferred from the donut graphs the reason of the occurrence of most deaths was unknown and their skeletons were found. Although there were few cases where people died due to drowning.

4.2 Visualizing the number of incidents on prominent regions –

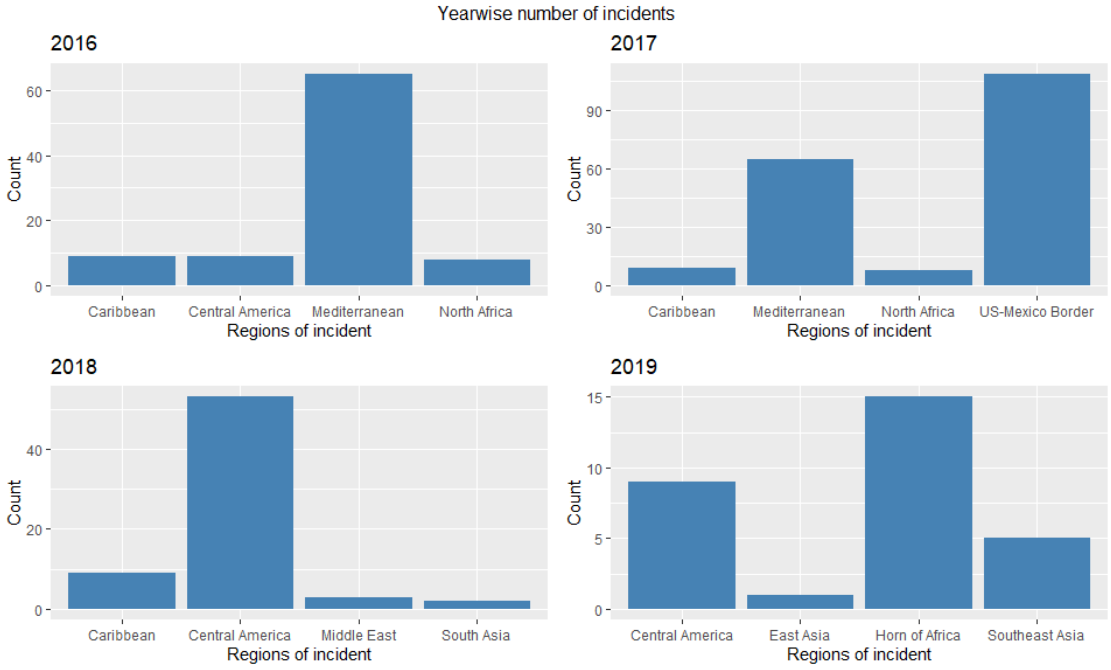


Fig 5: Number of incidents on prominent regions

From the bar chart, we can state few observations that -

In 2016, most of the incidents were reported in Europe and Horn of Africa. However in 2017 the rise of incidents around Mediterranean increased rapidly and most incidents took place on the same migration route. This trend was followed in 2018 and 2019.

4.3 Visualizing the number of deaths per year over different regions -

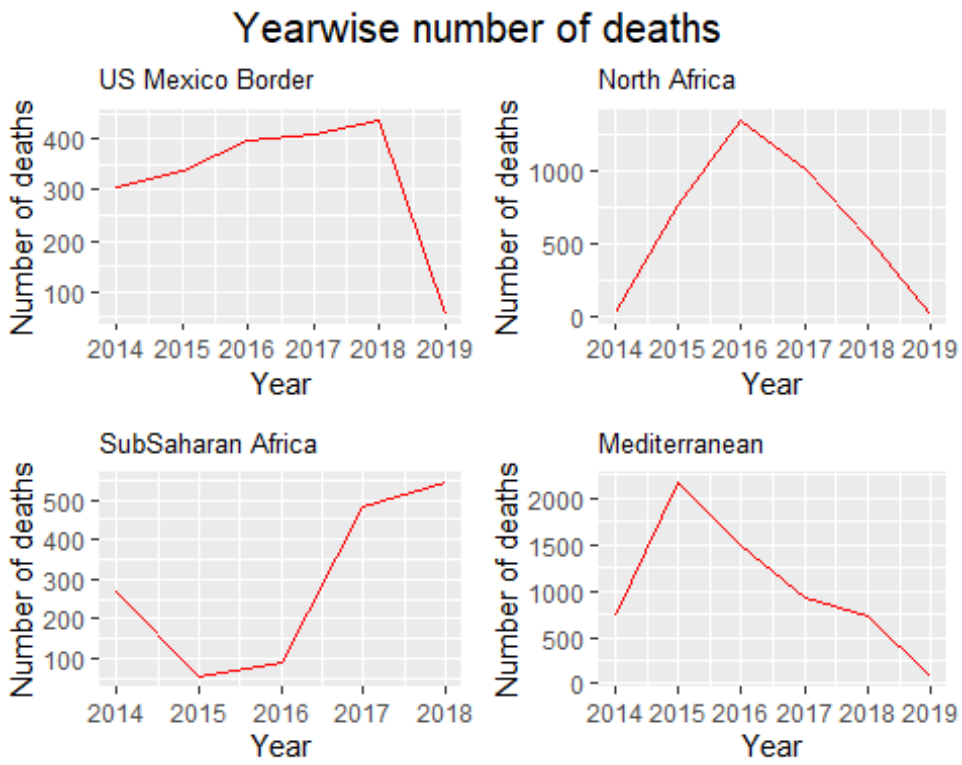


Fig 6: Year wise number of deaths for different regions

It can be observed that number of deaths kept increasing till 2018 at US Mexico border and sharply decreasing till 2019. The highest number of deaths occurred in 2016 at North Africa region. Sub Saharan Africa region recorded no deaths in 2019 after having maximum number of deaths in 2018. The Mediterranean region deaths also significantly reduced from 2015 to 2019.

4.4 Visualizing and Comparing Male/Female deaths across different regions -

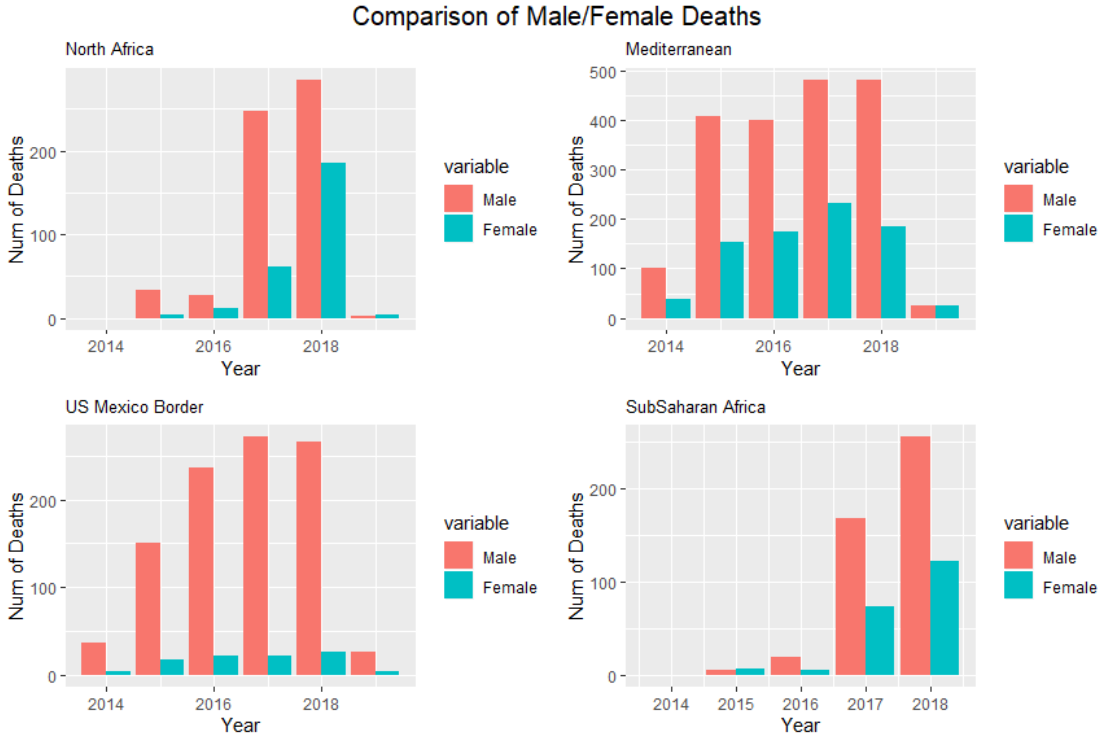


Fig 7: Comparison of Male/Female deaths across regions

Across Different geographic regions, it can be noted that the death count of males in comparison to females was more for every year. However in Sub Saharan region, during 2015 more females died.

4.5 Geographical view of number of incidents

4.5.1 North Africa

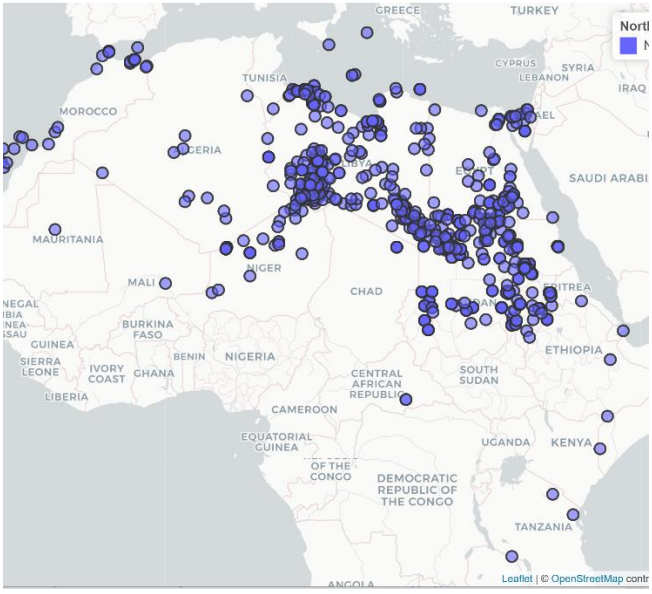


Fig 7 : Incidents near North Africa

Apparanetly, Sudan is a "transit country" (especially from Somalia, Ethiopia and Dem. Rep. Congo), through which masses of people travel, with the hope of reaching Europe. Unfortunately, many of them die, from the harsh weather conditions, on their way there. It can be seen on the map below.

4.5.2 US Mexico Border

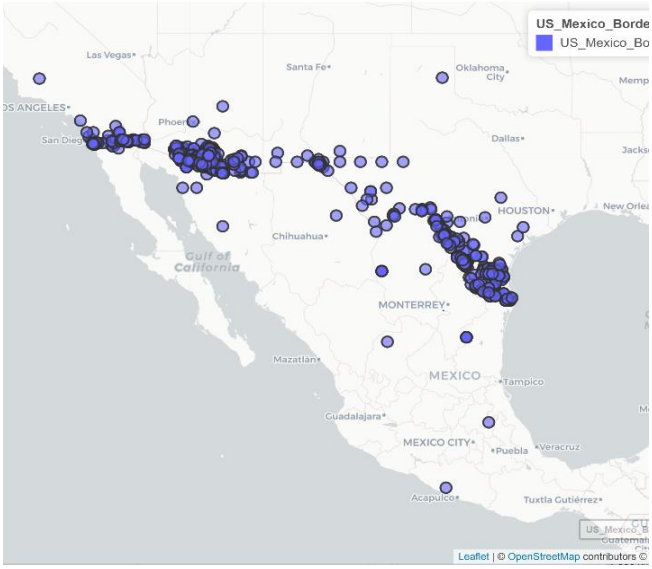


Fig 8 : Incidents near US Mexico Border

Migrants planning to cross the US-Mexico border are more inclined to take a safer ruote, steering away from the desert.
Therefore, ironically, on the deserts are less reported deaths, as there are less migrants in those areas.
The map below shows it well.

4.5.3 Sub Saharan No. of Incidents

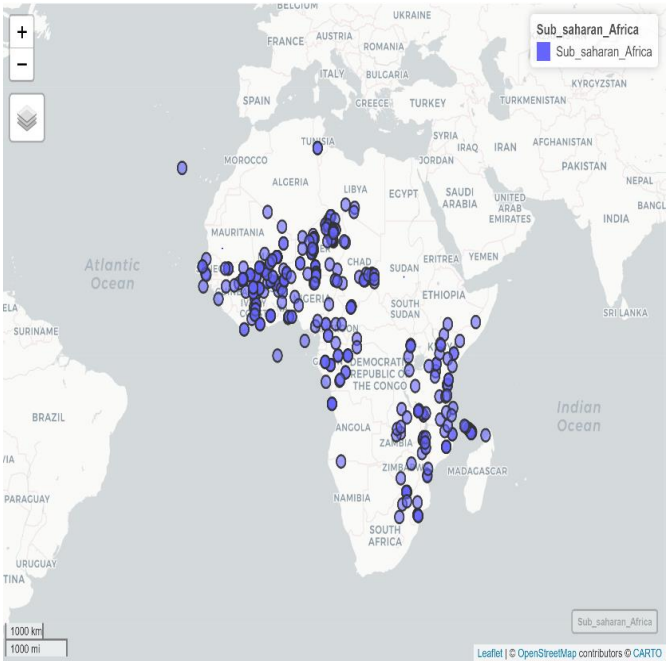


Fig 9 : Incidents near Sub Saharan Africa

The increased desire of North Africans to migrate from their country permanently is understandable given the increasingly difficult economic conditions in the region.

4.5.4 Mediterranean No. of Incidents

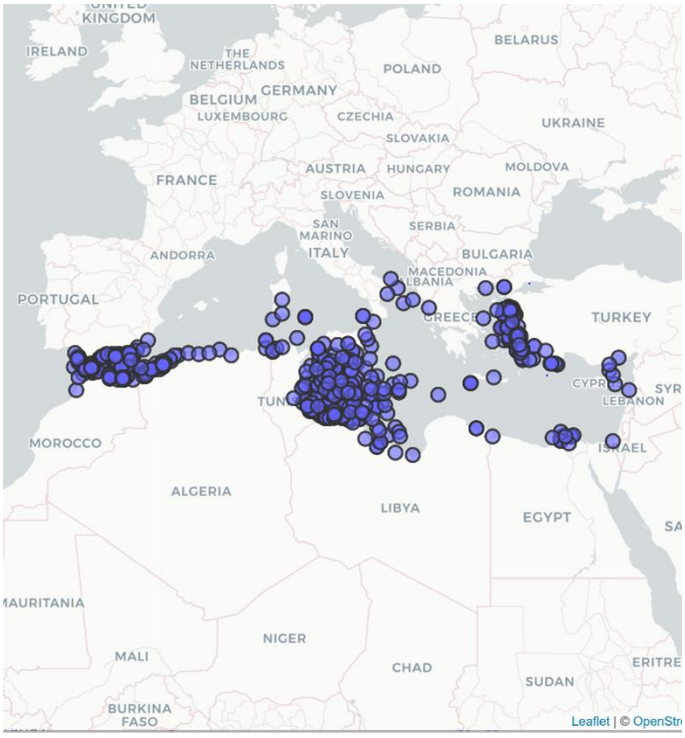


Fig 10 : Incidents near Mediterranean

Those routes have the shortest distance on the see; that's why migrants take them.
Other than those, asylum seekers are fleeing from the Middle East, where a few years ago a war broke out (especially in Syria). They migrate through Turkey and then through countries like Hungary, Bulgaria, Austria, Serbia, heading towards the Western Europe.
Some of them die falling from the train under La Manche Canal, trying to get to the United Kingdom from the Continental Europe.

5. Statistical Analysis

Hypothesis Testing -

Hypothesis Testing for mean of people missing at US Mexico border vs North africa

Null Hypothesis: The mean of missing people at US border is equal to mean of missing people at North Africa borders.

Alternate Hypothesis: The mean of missing people at US border is not equal to mean of missing people at North Africa borders.

```
t.test(samUS$Missingnum,samNA$Missingnum)

##
##  Welch Two Sample t-test
##
## data:  samUS$Missingnum and samNA$Missingnum
```

```
## t = -5.5682, df = 8.8879, p-value = 0.0003642
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -2.511649 -1.058459
## sample estimates:
##  mean of x  mean of y
## -0.3232912  1.4617633
```

Conclusion - The number of missing people at US Mexico Border are not same as those of North Africa.

Hypothesis Testing for mean number of people who survived in US Mexico 2017 and 2018

Null Hypothesis : Mean number of people survived in 2016 is same as 2017

Alternative Hypothesis : Mean number of 2016 survivors are not equal to 2017

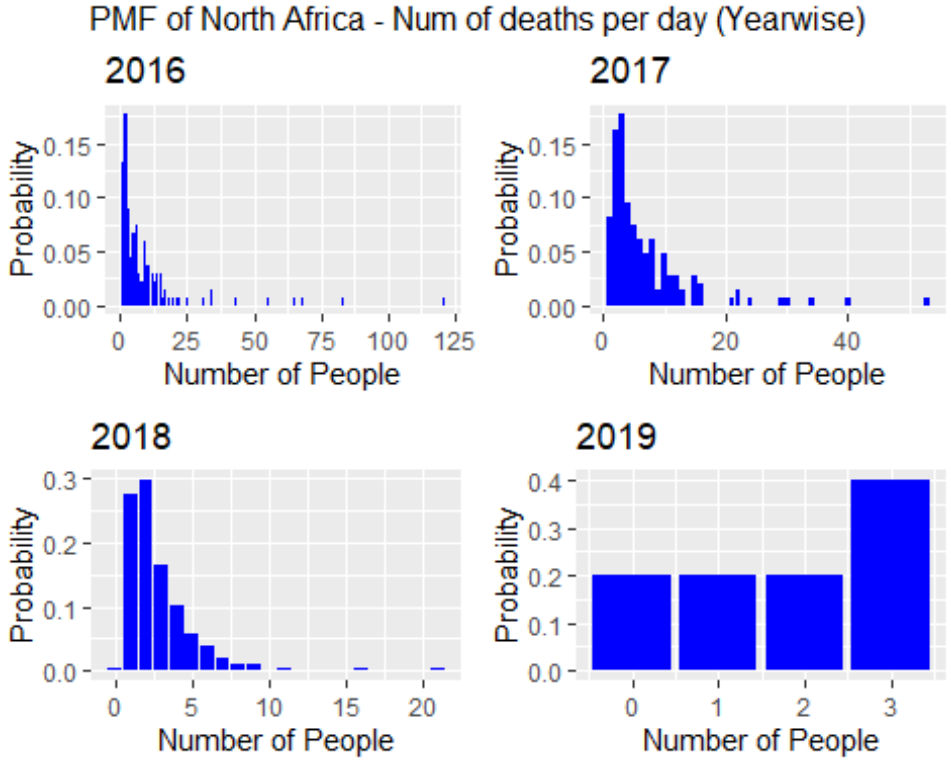
```
t.test(sam16$Survivors,sam17$Survivors)

##
##  Welch Two Sample t-test
##
## data:  sam16$Survivors and sam17$Survivors
## t = -0.1446, df = 8.0834, p-value = 0.8886
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -0.7294669  0.6432376
## sample estimates:
## mean of x mean of y
## 0.6704155 0.7135302
```

Conclusion - The mean number of people who survived in US-Mexico border in 2016 is same as those who survived in 2017.

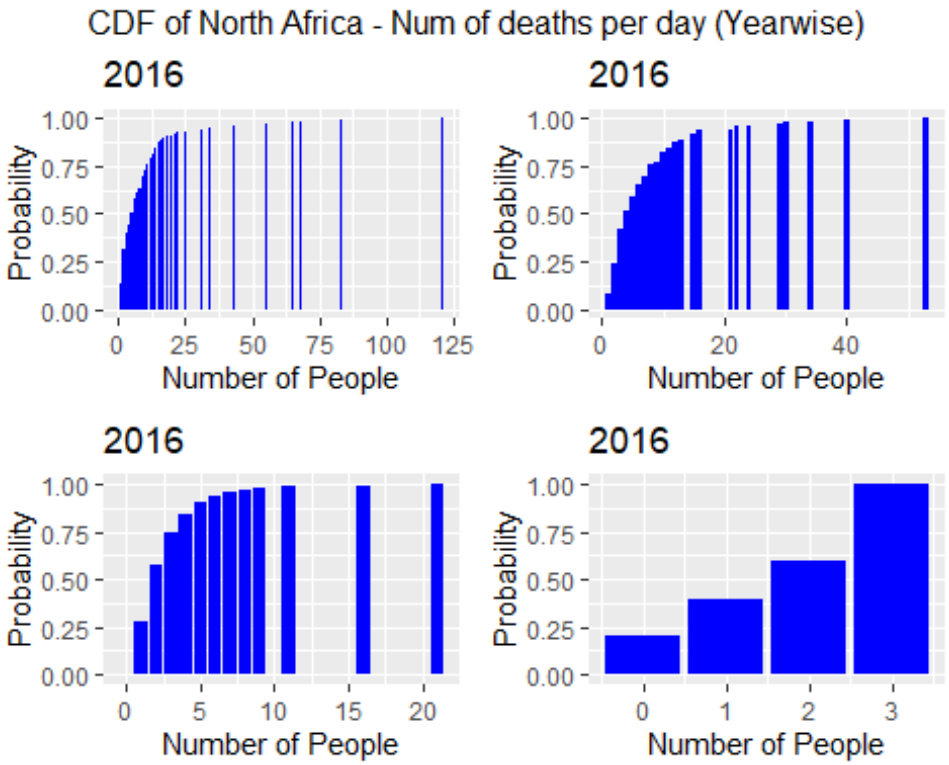
Probability and Statistics (IE 6200) - Sec 7 - Final Project Report

Omkar Kalange, Ayush Pandey

Probability Distributions for number of deaths -

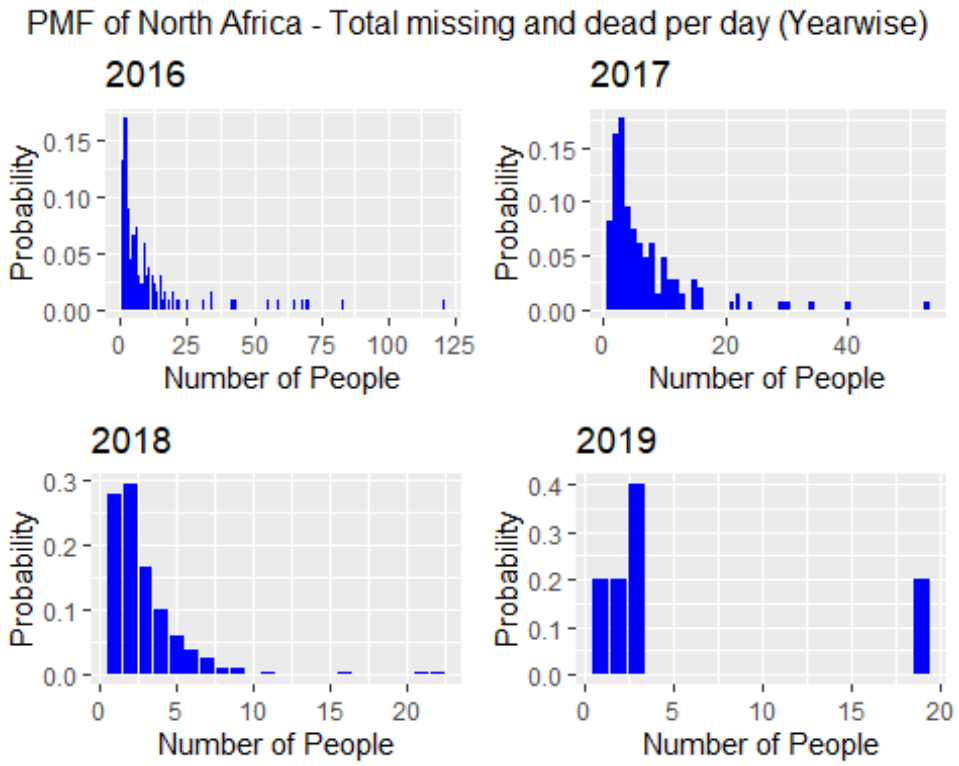


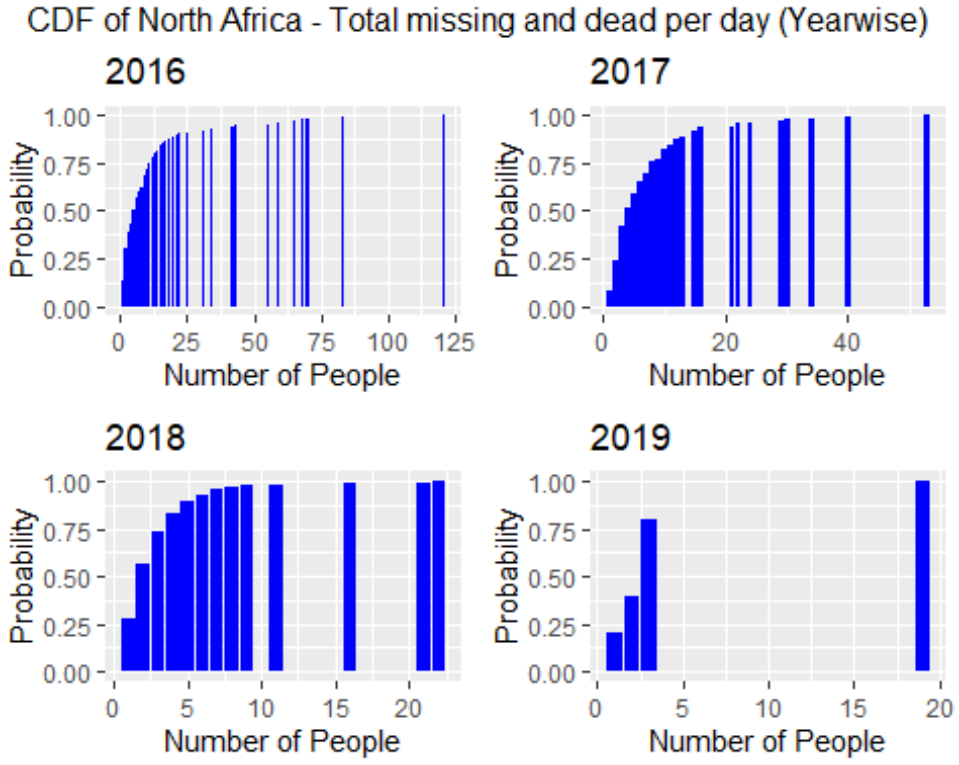
CDF of North Africa region deaths -



Probability and Statistics (IE 6200) - Sec 7 - Final Project Report

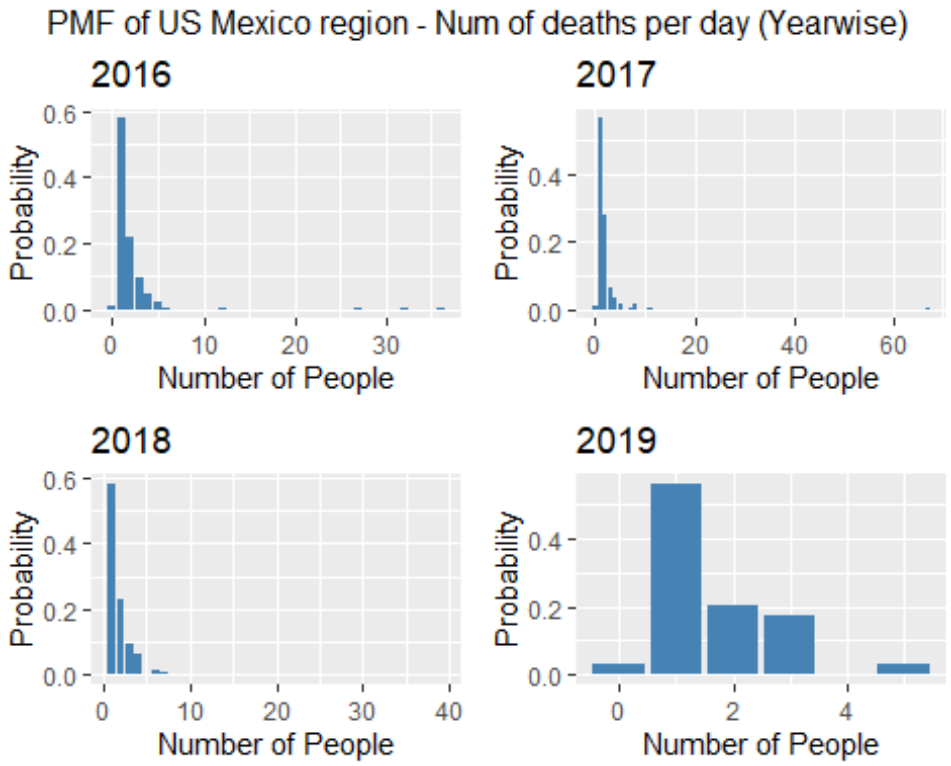
Omkar Kalange, Ayush Pandey

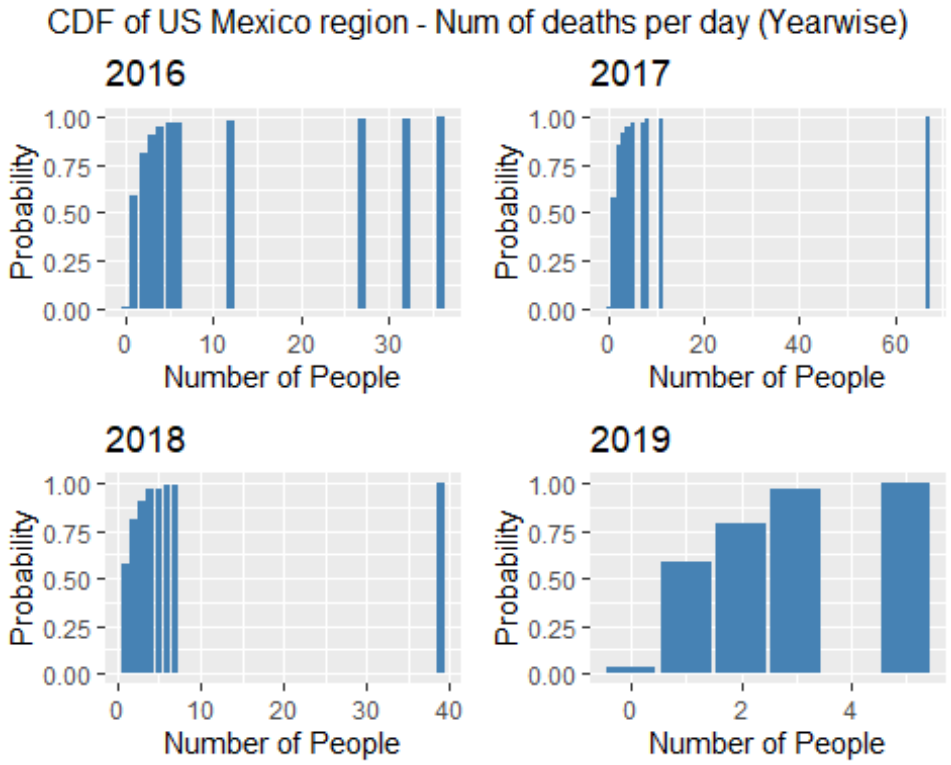




Probability and Statistics (IE 6200) - Sec 7 - Final Project Report

Omkar Kalange, Ayush Pandey

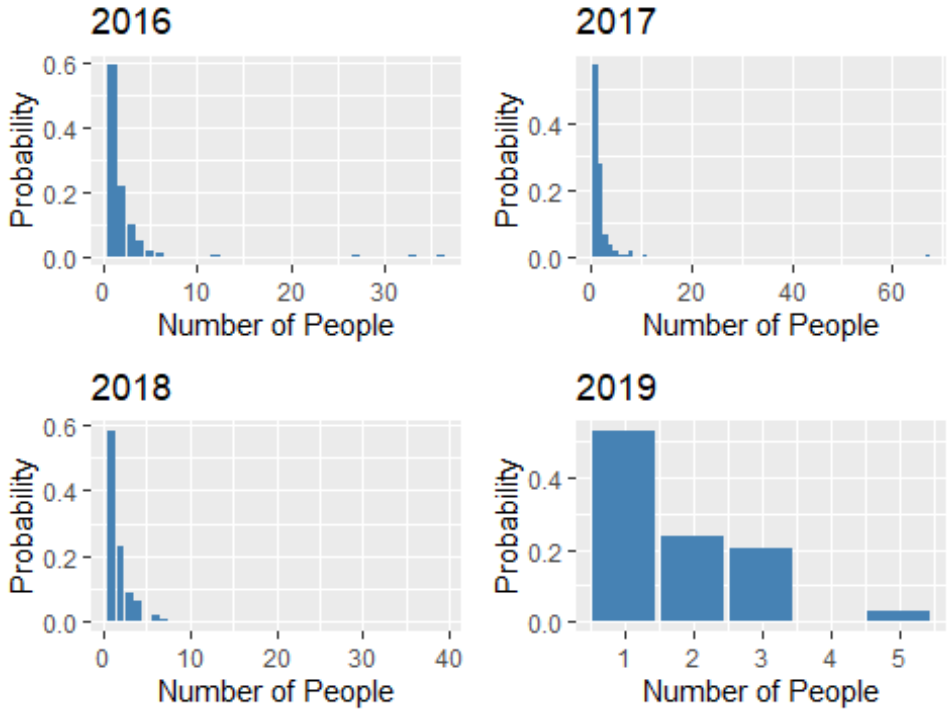




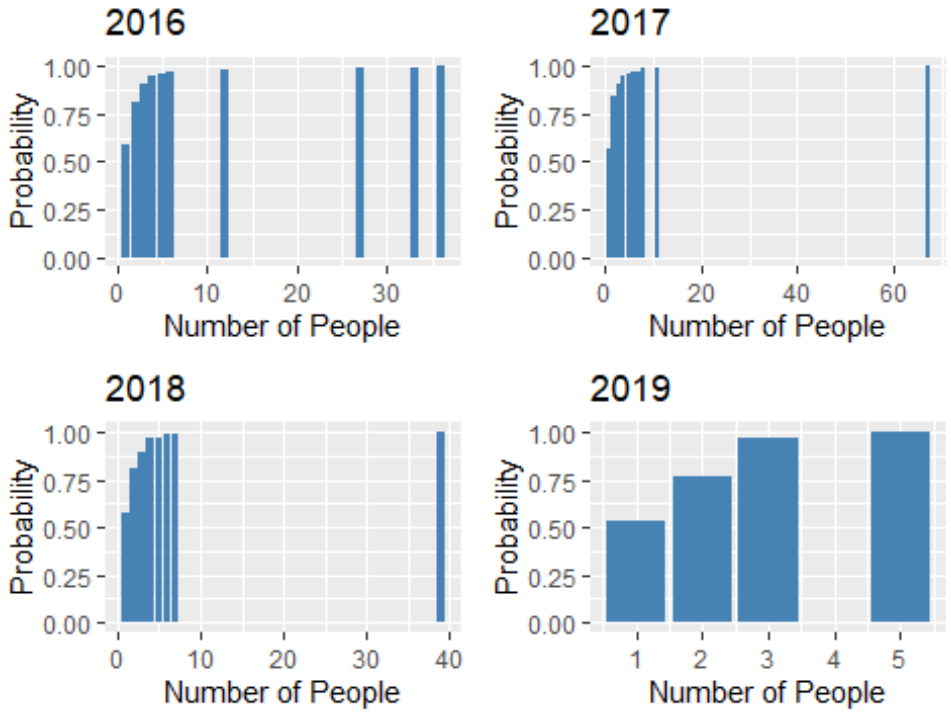
Probability and Statistics (IE 6200) - Sec 7 - Final Project Report

Omkar Kalange, Ayush Pandey

'MF of US Mexico region - Total deaths and missing per day (Yearwise



:DF of US Mexico region - Total deaths and missing per day (Yearwise



Probability and Statistics (IE 6200) - Sec 7 - Final Project Report

Omkar Kalange, Ayush Pandey

Correlation Statistics

Correlation between 2016 and 2017 deaths at US Mexico border -

```
cor.test(head(US_M_2016$freqmU1,6),head(US_M_2017$freqmU2,6))

##
##  Pearson's product-moment correlation
##
## data:  head(US_M_2016$freqmU1, 6) and head(US_M_2017$freqmU2, 6)
## t = 14.846, df = 4, p-value = 0.0001199
## alternative hypothesis: true correlation is not equal to 0
## 95 percent confidence interval:
##  0.9171329 0.9990650
## sample estimates:
##       cor
## 0.9910478
```

This suggests that the death count in 2016 is highly correlated to the death count in 2017.

Correlation between 2017 and 2018 deaths at US Mexico border –

```
cor.test(head(US_M_2017$freqmU2,6),head(US_M_2018$freqmU3,6))

##
##  Pearson's product-moment correlation
##
## data:  head(US_M_2017$freqmU2, 6) and head(US_M_2018$freqmU3, 6)
## t = 16.787, df = 4, p-value = 7.379e-05
## alternative hypothesis: true correlation is not equal to 0
## 95 percent confidence interval:
##  0.9344743 0.9992673
## sample estimates:
##       cor
## 0.9929779
```

So, the number of deaths in 2017 are also on similiar lines with those in 2018.

Correlation between 2017 US Mexico deaths vs North Africa –

```
cor.test(head(US_M_2017$freqmU2,6),head(NA_M_2017$freqM2,6))

##
##  Pearson's product-moment correlation
##
## data:  head(US_M_2017$freqmU2, 6) and head(NA_M_2017$freqM2, 6)
## t = 0.098878, df = 4, p-value = 0.926
## alternative hypothesis: true correlation is not equal to 0
## 95 percent confidence interval:
##  -0.7940012  0.8277682
## sample estimates:
##        cor
## 0.04937892
```

The number of deaths in 2017 at US-Mexico have very less correlation to those at North Africa.

6. Conclusion

We hope that monitoring these tragic events and studying the data relating to them can help to avoid a recurrence of these.