CMP5352: Data Visualisation Report

Dikshant Madai (23140738)

Saturday 15$^{\text{th}}$ June, 2024

Subject: CMP5352: Data Visualisation

Word count:188

Contents

Abstract

This research examines data from the International Organization for Migration's Missing Migrants Project, presenting a graphic portrayal of the global tragedy of migrant fatalities and disappearances between 2014 and 2023. The data suggests that the Mediterranean Sea remains the most dangerous route, with a consistently high number of missing migrants. 2016 saw the largest number of missing migrants and deaths in a single year. Despite this sad reality, the number of survivors has fluctuated over time, with the highest number recorded in 2016. A considerable gender disparity was seen across regions, with more males reporting missing than females.The Central Mediterranean route, particularly between Libya and Italy, has been highlighted as the most perilous area for migrants, with the highest number of missing people. Finally, the research ranks Afghanistan as the nation of origin with the highest number of migrant deaths, followed by Ethiopia, Mali, Algeria, and Côte d'Ivoire. This report aims to expose the global epidemic of missing migrants while also informing initiatives to improve safety and solve the challenges that individuals confront on their journeys.

1 Introduction

This dataset contains a detailed record of missing migrants and their tragic journeys to overseas destinations, gathered by the Missing Migrants Project, an initiative launched by the overseas Organization for Migration (IOM) in 2014. The collection tracks deaths and disappearances, shedding insight on the difficulties migrants confront on their journeys. Please keep in mind that because of the complexity of data collecting, the figures reported are most likely an undercount. The dataset is a monument to the individuals who died, as well as the families and communities affected by their absence. This report uses data visualization to analyze patterns in migrant deaths and disappearances recorded by the International Organization for Migration's Missing Migrants Project. It aims to raise awareness of this global issue and inform efforts to improve migrant safety. The report explores trends by region, location, migration route, and country of origin. By visualizing the data, we can identify areas with the highest incidence of missing migrants and understand how these patterns have evolved over time. Missing Migrants Project collects data through a variety of sources, including:The Missing

- Migrants Project collects data through a variety of sources
- Media Reports
- Witness Accounts
- Government and NGO Reports
- Community Reports

2 Motivation and Objectives

The Missing Migrants Project is a crucial resource for understanding the dangers faced by migrants worldwide. This report utilizes the project's data to answer the following questions:

- Which regions have the highest number of migrant deaths and disappearances, and how do these numbers vary over time?
- What insights can we gain from analyzing the total number of missing migrants each year? which year is the highest number of the missing migrants and death.
- Trend in the Number of Survivors Over the Years. which year is the more Survivors?
- What is the gender ratio of missing migrants across different regions?
- Which is the top impact location of death on number of missing migrants?

- Which migration route has the highest incidence of missing persons, and how does it compare to others?

- which is the most countries of origin by number of migrant deaths ?

The data provides valuable insights to guide policymakers and humanitarian organizations in their efforts to address migrant safety.

3 Data Description

It has the record of missing migrants from 2014 to 2023 Here is the information about the dataset :

- **Incident Type**: Type of migration incident

- **Incident Year**: Year when the incident happened

- **Reported Month**: Month when the incident was reported.

- **Region of Origin**: Geographical region where the migrants originated.

- **Region of Incident:**: The region where the incident occurred

- **Country of Origin:**:Country from where the migrants came

- **Number of Dead**: Number of confirmed dead migrants

- **Minimum Estimated Number of Missing**: Minimum estimated number of missing migrants.

- **Total Number of Dead and Missing**: Total number of deceased and missing migrants

- **Number of Survivors**:The number of migrants that survived the incident.

- **Number of Females**: Number of female migrants involved.

- **Number of Males**: Number of male migrants involved.

- **Number of Children**: Number of children migrants involved. participated.

- **Cause of Death**: Cause of death among migrants.

- **Migration Route**: Route used by migrants during their journey (if available)

- **Location of Death:**: Description of the location where the incident occurred.

- **Information Source**: Source of information about the incident.

- **Coordinates**: Approximately where the incident occurred.

- **UNSD Geographical Grouping**: Geographic classification according to the United Nations Statistics Division

Data source: `https://www.kaggle.com/datasets/nelgiriyewithana/global-missing-migrants-datas` (click here)

4 Visualization

```
> # Libraries
> library(ggplot2)
> library(reshape2)
> library(dplyr)
> library(maps)
> library(tidyr)
> library(bbplot)
> library(lubridate)
> library(readr)
> library(rmarkdown)
```

4.1 After cleaning this data set i have loaded it.

```
> # Loading Dataset and information about it
> df <- read_csv("D:/Data science/final_visulization_code/cleandataset/migrants_clean.csv")
> # Dealing qith missing values
> # Get the number of rows and columns
> num_rows <- nrow(df)
> num_cols <- ncol(df)
> cat("Number of rows:", num_rows, "\n")

Number of rows: 13020

> cat("Number of columns:", num_cols, "\n")

Number of columns: 19

> # Calculate missing values for each column
> missing_values <- sapply(df, function(x) sum(is.na(x)))
> print(missing_values)
```

	Incident Type		Incident Year
	0		0
	Reported Month		Region of Origin
	0		0
	Region of Incident		Country of Origin
	0		0
	Number of Dead	Minimum Estimated	Number of Missing
	0		0
Total Number of Dead and Missing			Number of Survivors
	0		0
	Number of Females		Number of Males
	0		0
	Number of Children		Cause of Death
	0		0

```
                 Migration Route              Location of Death
                               0                             0
               Information Source                  Coordinates
                               0                             0
          UNSD Geographical Grouping
                               0
> # bbc_style customize theme
> bbc_style <- function() {
+   font <- "Helvetica"
+
+   ggplot2::theme(
+     plot.title = ggplot2::element_text(
+       family = font, size = 15, face = "bold", color = "#222222", hjust = 0.5),
+     plot.subtitle = ggplot2::element_text(
+       family = font, size = 18, margin = ggplot2::margin(9, 0, 9, 0)),
+     plot.caption = ggplot2::element_blank(),
+
+     legend.position = "top",
+     legend.text.align = 0,
+     legend.background = ggplot2::element_blank(),
+     legend.title = ggplot2::element_blank(),
+     legend.key = ggplot2::element_blank(),
+     legend.text = ggplot2::element_text(
+       family = font, size = 10, color = "#222222", hjust = 0.70),
+
+     axis.title = ggplot2::element_blank(),
+     axis.text = ggplot2::element_text(
+       family = font, size = 11, color = "#222222"),
+     axis.text.x = ggplot2::element_text(margin = ggplot2::margin(5, b = 10)),
+     axis.ticks = ggplot2::element_blank(),
+     axis.line = ggplot2::element_blank(),
+
+     panel.grid.minor = ggplot2::element_blank(),
+     panel.grid.major.x = ggplot2::element_blank(),
+     panel.background = ggplot2::element_blank(),
+
+     strip.background = ggplot2::element_rect(fill = "white"),
+     strip.text = ggplot2::element_text(size = 20, hjust = 0)
+   )
+ }

> # Dealing with missing values
> # Get the number of rows and columns
> num_rows <- nrow(df)
> num_cols <- ncol(df)
> cat("Number of rows:", num_rows, "\n")
```

```
Number of rows: 13020

> cat("Number of columns:", num_cols, "\n")

Number of columns: 19
```

4.2   Heatmap of Correlation Matrix

```
> # 1). Heatmap of Correlation Matrix
>
> # Filter numeric columns
> numeric_df <- df %>% select_if(is.numeric)
> # Compute correlation matrix
> correlation_matrix <- cor(numeric_df, use = "complete.obs")
> # Melt the correlation matrix for plotting
> melted_correlation <- melt(correlation_matrix)
> heatmap_plot <- ggplot2::ggplot(melted_correlation,
+                                   aes(Var1, Var2, fill = value)) +
+   ggplot2::geom_tile(color = "white") +
+   ggplot2::scale_fill_gradient2(low = "blue", high = "red", mid = "white",
+                     midpoint = 0, limit = c(-1, 1), space = "Lab",
+                     name = "Correlation") +
+   ggplot2::theme_minimal() +
+   ggplot2::theme(plot.title = ggplot2::element_text(
+     size = 19, face = "bold", color = "#222222"),
+         axis.text.x = ggplot2::element_text(
+           angle = 45, vjust = 1, size = 10, hjust = 1, face="bold"),
+         axis.text.y = ggplot2::element_text(
+           face="bold", size = 10),
+        panel.grid.major = ggplot2::element_blank(),
+        panel.grid.minor = ggplot2::element_blank(),
+        legend.justification = c(1, 0),
+        legend.direction = "vertical") +
+   ggplot2::labs(title = "Correlation Heatmap",
+       x = "",
+       y = "")
> # Print the plot
> print(heatmap_plot)
```

Correlation Heatmap
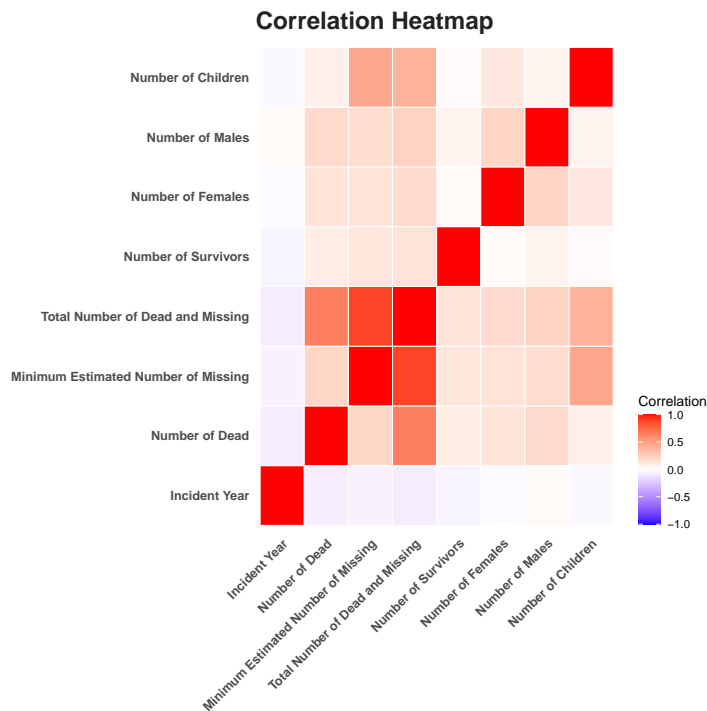
The heatmap shows the correlation between different variables related to migration incidents.There is also a strong positive correlation between the number of dead and the total number of dead and missing. The heatmap shows that there is a strong correlation between the number of deaths and the number of missing people. This suggests that as the number of deaths increases, the number of missing people also tends to increase.

4.3 Regions with highest number of migrant deaths and disappearances over time

```
> #2) Which regions have the highest number of migrant deaths
> #and disappearances,and how do these numbers vary over time?
> #How do these numbers vary over time, particularly in the most affected regions?
>
> #==>
> # Splitting 'Coordinates' column into 'latitude' and 'longitude'
> df <- separate(df, Coordinates, into = c("latitude", "longitude"),
+                sep = ", ", convert = TRUE)
> # Convert latitude and longitude to numeric
> df$latitude <- as.numeric(df$latitude)
> df$longitude <- as.numeric(df$longitude)
> # Download world map data
```

```
> world_map <- map_data("world")
> # Creating the scatterplot on a world map
> ggplot2::ggplot() +
+   ggplot2::geom_map(data = world_map, map = world_map,
+             aes(long, lat, map_id = region),
+             color = "green", fill = "lightyellow") +
+
+   ggplot2::geom_point(data = df,
+               aes(x = longitude, y = latitude,
+                   color = `Total Number of Dead and Missing`,
+                   size = `Total Number of Dead and Missing`)) +
+
+   # Title and axis labels
+   ggplot2::labs(
+     title = "Migration Incidents on World Map",
+     x = "Longitude",
+     y = "Latitude",
+     color = "Total Number of Dead and Missing",
+     size = "Total Number of Dead and Missing"
+   ) +
+
+   ggplot2::scale_color_gradient(low = "blue", high = "red") +
+
+   bbc_style() +
+   ggplot2::theme(
+     text = ggplot2::element_text(family = "noto"),
+     plot.title = ggplot2::element_text(size = 16, hjust = 0.5),
+     legend.position = "bottom",
+     panel.grid.major = ggplot2::element_blank(),
+     panel.grid.minor = ggplot2::element_blank()
+   )
>
```

Migration Incidents on World Map
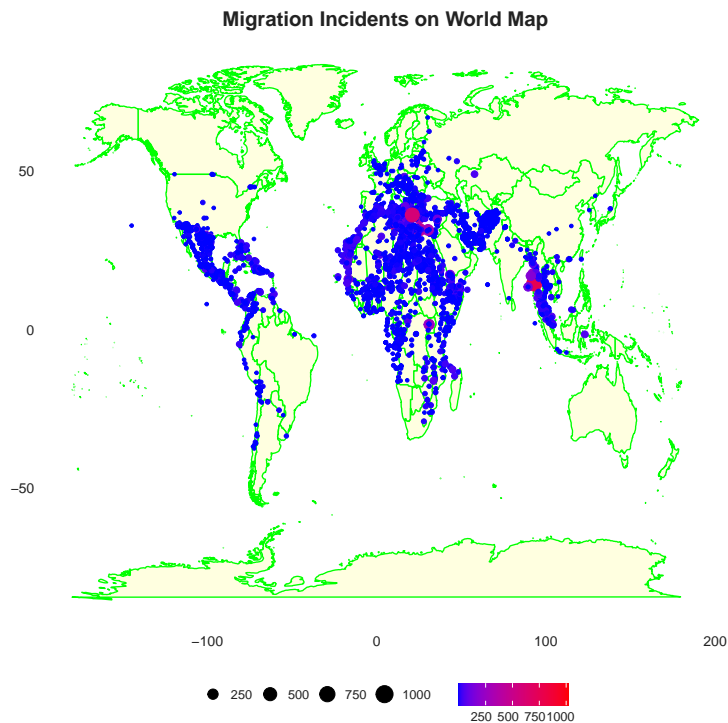
The map shows the locations of migration incidents around the world. The color of the dots represents the number of incidents at each location. The red dots represent the highest number of incidents, while the blue dots represent the lowest number of incidents. The map shows that the majority of migration incidents occur in the Mediterranean Sea, the South China Sea and more.

4.4   Region with Highest Number of Incident

```
> # 3). Which regions have the highest number of incidents?
>
>
> # Create a date column for easier grouping by year
> df$date <- as.Date(paste0(df$`Incident Year`, "-01-01"))
> # 1. Grouping by region to find total missing migrants
> region_data <- df %>%
+  group_by(`Region of Incident`) %>%
+  summarise(total_missing = sum(`Total Number of Dead and Missing`,
+                          na.rm = TRUE)) %>%
+  arrange(desc(total_missing))
> # Create the bar chart
> ggplot2::ggplot(region_data, aes(x = reorder(`Region of Incident`,
+                                        total_missing),
+                                   y = total_missing)) +
```

```
+  ggplot2::geom_bar(stat = "identity", fill = "darkred") +
+  ggplot2::coord_flip() + # Flip coordinates for better readability
+  ggplot2::geom_text(aes(label = total_missing), hjust = -0.09,
+                     color = "black", size=2.5) +
+  ggplot2::labs(
+    title = "Regions with the Highest Number of Missing Migrants",
+    x = NULL,
+    y = "Total Missing Migrants"
+  ) +
+  bbc_style() +
+  ggplot2::theme(
+    panel.grid.major = ggplot2::element_blank(),
+    panel.grid.minor = ggplot2::element_blank(),
+    axis.title.x = ggplot2::element_blank(),
+    axis.text.x = ggplot2::element_blank(),
+    axis.ticks.x = ggplot2::element_blank(),
+    axis.text.y = ggplot2::element_text(margin = ggplot2::margin(r = -15))
+
+  )
```

Regions with the Highest Number of Missing Migrants

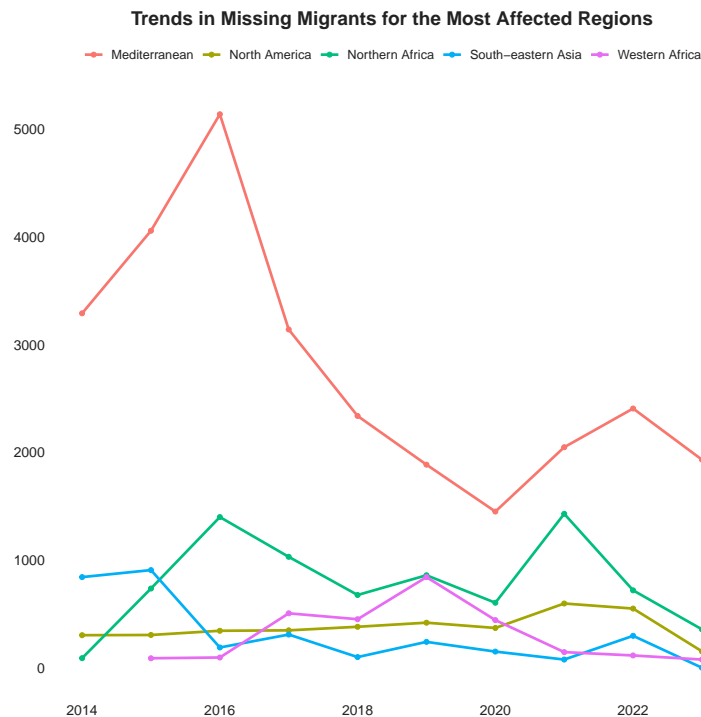| Region	Total Missing Migrants
Mediterranean	27675
Northern Africa	7899
North America	3770
South–eastern Asia	3114
Western Africa	2765
Southern Asia	2494
Central America	2253
Western Asia	2142
Eastern Africa	1937
Caribbean	1503
Europe	1019
South America	467
Middle Africa	214
Southern Africa	161
Central Asia	52
Eastern Asia	48
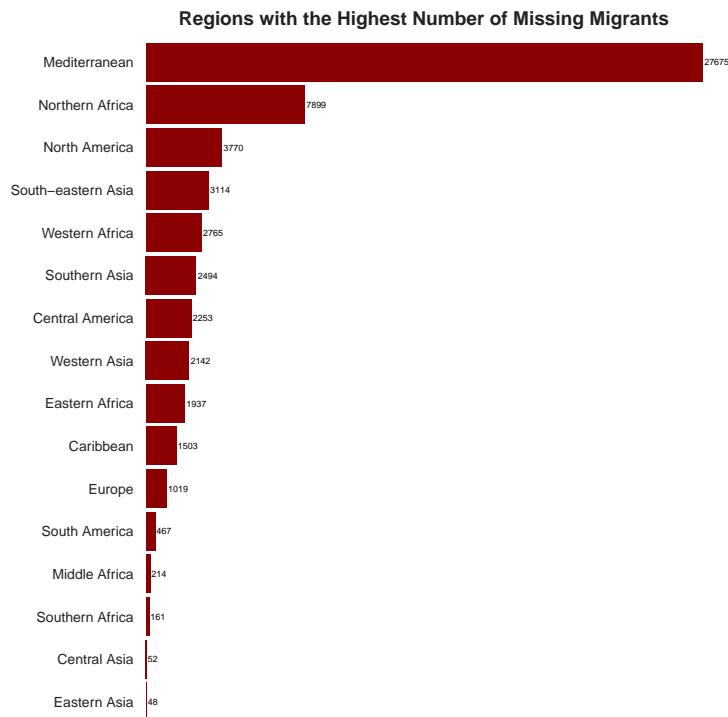
The graph depicts the number of missing migrants according on region. The Mediterranean region has the highest number of missing migrants, followed by North Africa and North America. The amount of missing migrants reduces as

13

you go down the list, with Eastern Asia having the fewest. The graph indicates that missing migrants have a disproportionate impact in the Mediterranean region. This is most likely owing to the area's proximity to war zones and the large number of individuals attempting to cross the Mediterranean Sea to reach Europe.

4.5   Most Affected Regions

```
> # Filtering for the top 5 most affected regions
> top_regions <- region_data %>%
+  top_n(5) %>%
+  pull(`Region of Incident`)
> # Filtering the main dataset to include only the top regions
> df_top_regions <- df %>%
+  filter(`Region of Incident` %in% top_regions)
> # Calculating yearly trends for each of the top regions
> region_trends <- df_top_regions %>%
+  group_by(year = `Incident Year`, `Region of Incident`) %>%
+  summarize(total_missing = sum(`Total Number of Dead and Missing`,
+                               na.rm = TRUE))
> # Plotting the trends over time
> ggplot2::ggplot(region_trends, aes(x = year, y = total_missing,
+                                    color = `Region of Incident`)) +
+  ggplot2::geom_line(size = 1) +
+  ggplot2::geom_point() +
+  ggplot2::labs(title = "Trends in Missing Migrants for the Most Affected Regions",
+       x = "Year", y = "Total Missing", color = "Region") +
+  bbc_style()
>
```

14

**Trends in Missing Migrants for the Most Affected Regions**

The Mediterranean region has the most missing migrants, followed by North Africa and North America. The number of missing migrants in the Mediterranean region has decreased since 2016, although it remains the region with the most missing people. The number of missing migrants in South-east Asia has been decreasing since 2014, whereas the number of missing migrants in Western Africa has increased.

4.6 What insights can we gain from analyzing the total number of missing migrants each year?

```
> # 2. What insights can we gain from analyzing the total number of missing
> #migrants each year?
> total_missing_per_year <- df %>%
+   group_by(year = year(date)) %>%
+   summarize(total_missing = sum(`Total Number of Dead and Missing`,
+                                 na.rm = TRUE))
> # Plotting total number of missing migrants over time
> ggplot2::ggplot(total_missing_per_year, aes(x = year, y = total_missing)) +
+   ggplot2::geom_line(color = "green") +
+   ggplot2::geom_point() +
+   ggplot2::geom_text(aes(label = total_missing), vjust = -0.5, size = 3) +
+   ggplot2::labs(title = "Total Number of Missing Migrants Over Time",
```

```
+         x = "Year", y = "Total Missing") +
+  bbc_style()
```

Total Number of Missing Migrants Over Time
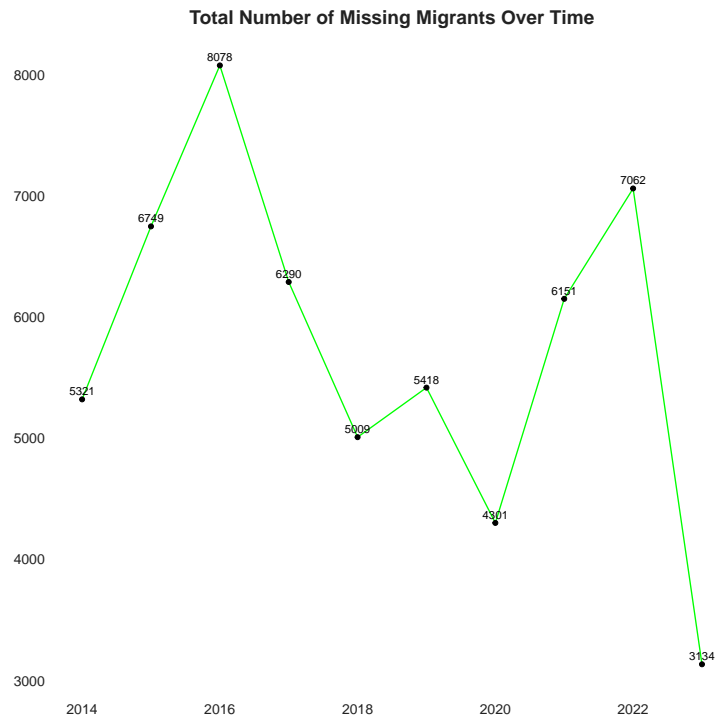
The graph shows the total number of missing migrants over time. The number of missing migrants increased from 2014 to 2016, reaching a peak of 8078 in 2016. The number then decreased until 2020, when it began to increase again. In 2022, the number of missing migrants reached 7062. The graph shows that the number of missing migrants is fluctuating, but it is generally increasing over time

4.7   What is the trend in the number of survivors over the years?

```
> #3.What is the trend in the number of survivors over the years?
> # Group by Incident year and calculate the sum of Number of Survivors
> survivors_trend <- df %>%
+  group_by(`Incident Year`) %>%
+  summarize(total_survivors = sum(`Number of Survivors`, na.rm = TRUE))
> # Plotting the trend of survivors
> ggplot2::ggplot(survivors_trend, aes(x = `Incident Year`,
+                                      y = total_survivors)) +
+  ggplot2::geom_line(color = "green") +
+  ggplot2::geom_point() +
```

```
+   ggplot2::geom_text(aes(label = total_survivors), vjust = -0.5, size = 3) +
+   ggplot2::labs(
+     title = "Trend in the Number of Survivors Over the Years",
+     x = "Year",
+     y = "Total Survivors"
+   ) +
+   bbc_style()
>
>
```

Trend in the Number of Survivors Over the Years
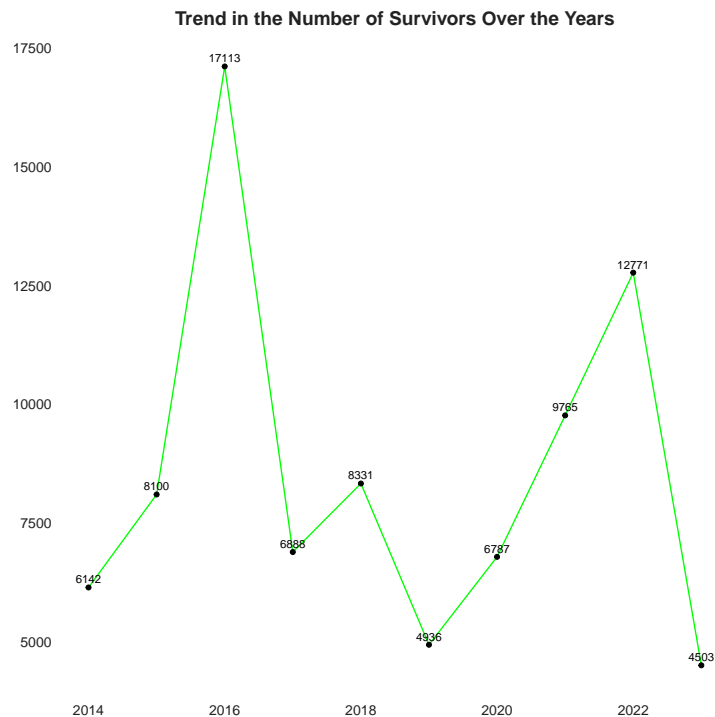
The graph shows the trend in the number of survivors over the years. The number of survivors increased from 2014 to 2016, reaching a peak of 17113 in 2016. The number then decreased until 2019, when it began to increase again. In 2022, the number of survivors reached 12771. The graph shows that the number of survivors is fluctuating, but it is generally increasing over time.

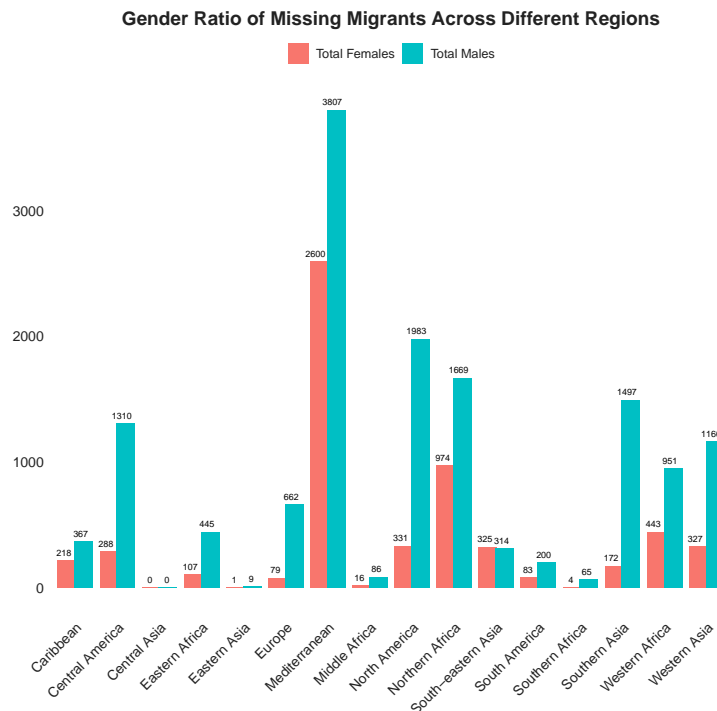## 4.8 What is the gender ratio of missing migrants across different regions

```
> #4. What is the gender ratio of missing migrants across different regions
> # Group by Region of Incident and calculate the sum of Number
> #of Females and Males
> gender_ratio <- df %>%
```

```
+   group_by(`Region of Incident`) %>%
+   summarize(total_females = sum(`Number of Females`, na.rm = TRUE),
+             total_males = sum(`Number of Males`, na.rm = TRUE)) %>%
+   gather(key = "Gender", value = "Count", total_females, total_males)
> ggplot2::ggplot(gender_ratio, aes(x = `Region of Incident`, y = Count,
+                                    fill = Gender)) +
+   ggplot2::geom_bar(stat = "identity",
+                     position = position_dodge(width = 0.8)) +
+   ggplot2::geom_text(aes(label = Count, y = Count, group = Gender),
+                     position = position_dodge(width = 0.8),
+                     vjust = -0.7, hjust = 0.7, size = 2.5, color = "black") +
+   ggplot2::labs(
+     title = "Gender Ratio of Missing Migrants Across Different Regions",
+     x = "Region",
+     y = "Count",
+     fill = "Gender"
+   ) +
+   bbc_style() +
+   ggplot2::theme(axis.text.x = ggplot2::element_text(angle = 45,
+                                         vjust = 1, hjust = 1)) +
+   scale_fill_discrete(labels = c("Total Females", "Total Males"))
```


The graph shows the gender ratio of missing migrants across different regions.

The Mediterranean region has the highest number of missing migrants, followed by Northern Africa and North America.
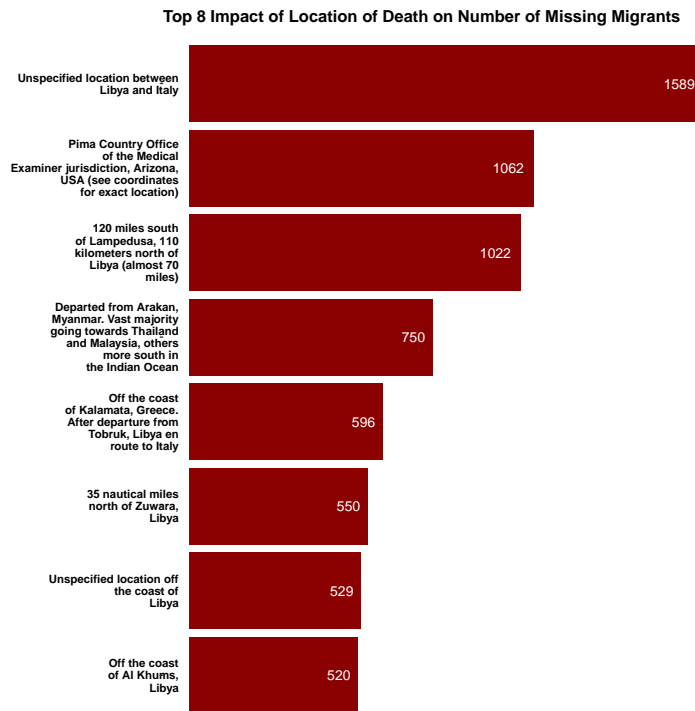
4.9 How does the number of missing migrants vary across different locations of death? Are certain locations significantly associated with higher numbers of missing migrants?

```
> library(dplyr)
> library(ggplot2)
> # Function to insert newline characters after every 6 words
> insert_newlines <- function(text, words_per_line = 3) {
+  words <- unlist(strsplit(text, " "))
+  lines <- lapply(seq(1, length(words), words_per_line),
+                  function(i) {
+    paste(words[i:min(i + words_per_line - 1, length(words))],
+          collapse = " ")
+  })
+  return(paste(lines, collapse = "\n"))
+ }
> # Apply the function to mutate the dataframe
> df <- df %>%
+  mutate(`Location of Death` = sapply(`Location of Death`,
+                                      insert_newlines))
> # Filter the top 8 locations of death
> top_locations <- df %>%
+  group_by(`Location of Death`) %>%
+  summarise(total_missing = sum(`Total Number of Dead and Missing`)) %>%
+  top_n(8, total_missing)
> # Plotting the impact of location of death for the top 8 locations
> ggplot2::ggplot(top_locations, aes(x = reorder(`Location of Death`,
+                                                 total_missing),
+                                    y = total_missing)) +
+  ggplot2::geom_bar(stat = "identity", fill = "darkred") +
+  ggplot2::geom_text(aes(label = total_missing), hjust = 1.3,
+                     color = "white") +
+  ggplot2::coord_flip() +
+  ggplot2::labs(
+    title = "Top 8 Impact of Location of Death on Number of Missing Migrants",
+    x = "Location of Death",
+    y = "Total Number of Dead and Missing"
+  ) +
+  ggplot2::theme(
+    plot.title = ggplot2::element_text(face = "bold"),
+    panel.grid.major = ggplot2::element_blank(),
```

```
+     panel.grid.minor = ggplot2::element_blank(),
+     axis.title.x = ggplot2::element_blank(),
+     axis.text.x = ggplot2::element_blank(),
+     axis.ticks.x = ggplot2::element_blank(),
+     axis.text.y = ggplot2::element_text(face = "bold",
+                                         margin = ggplot2::margin(r = -15),
+                                         color = "black"),
+     axis.title.y = ggplot2::element_blank(),         # Remove y-axis title
+     panel.background = ggplot2::element_rect(fill = "white", color = NA)
+   )
>
>
```

Top 8 Impact of Location of Death on Number of Missing Migrants

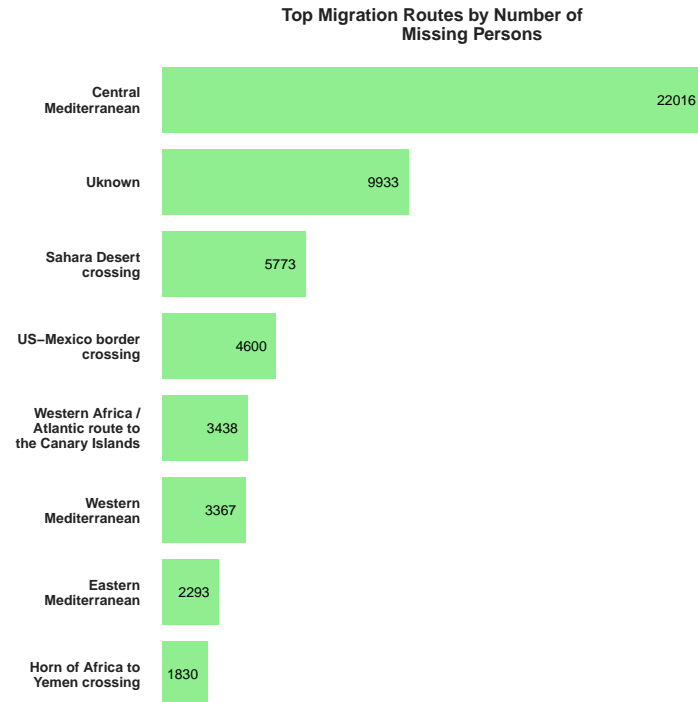| Location	Number
Unspecified location between Libya and Italy	1589
Pima Country Office of the Medical Examiner jurisdiction, Arizona, USA (see coordinates for exact location)	1062
120 miles south of Lampedusa, 110 kilometers north of Libya (almost 70 miles)	1022
Departed from Arakan, Myanmar. Vast majority going towards Thailand and Malaysia, others more south in the Indian Ocean	750
Off the coast of Kalamata, Greece. After departure from Tobruk, Libya en route to Italy	596
35 nautical miles north of Zuwara, Libya	550
Unspecified location off the coast of Libya	529
Off the coast of Al Khums, Libya	520

The graph shows the top 8 locations where missing migrants were found dead. The most common location is unspecified location between Libya and Italy, followed by Pima County, Arizona, USA and 120 miles south of Lampedusa. The data shows that the Mediterranean sea is the most dangerous route for migrants.

4.10 Which migration route has the highest incidence of missing persons, and how does it compare to others

```
> #6) "Which migration route has the highest incidence of missing persons,
> #and how does it compare to others?"
> # Group by Migration route
> route_comparison <- df %>%
+  group_by(`Migration Route`) %>%
+  summarize(total_missing = sum(`Total Number of Dead and Missing`,
+                                na.rm = TRUE)) %>%
+  filter(!is.na(`Migration Route`))
> library(ggplot2)
> library(ggthemes)
> # Calculate total missing per route
> route_comparison <- df %>%
+  group_by(`Migration Route`) %>%
+  summarize(total_missing = sum(`Total Number of Dead and Missing`,
+                                na.rm = TRUE)) %>%
+  filter(!is.na(`Migration Route`))
> # Sort routes by total missing in descending order
> route_comparison <- route_comparison[order(route_comparison$total_missing,
+                                       decreasing = TRUE), ]
> # Create the first bar chart for the top 8 routes
> ggplot2::ggplot(route_comparison[1:8, ],
+               aes(x = reorder(stringr::str_wrap(`Migration Route`,
+                                               width = 20), total_missing),
+                   y = total_missing)) +
+  ggplot2::geom_bar(stat = "identity", fill = "lightgreen",
+                    width = 0.8) + # Adjust width for spacing
+  ggplot2::geom_text(aes(label = total_missing), hjust = 1.3,
+                     color = "black", face="bold") +
+  ggplot2::coord_flip() +
+  ggplot2::labs(title = "Top Migration Routes by Number of
+               Missing Persons",
+       x = "Route", y = "Total Missing") +
+  bbc_style() +
+  ggplot2::theme(
+    panel.grid.major = ggplot2::element_blank(),
+    panel.grid.minor = ggplot2::element_blank(),
+    axis.title.x = ggplot2::element_blank(),
+    axis.text.x = ggplot2::element_blank(),
+    axis.ticks.x = ggplot2::element_blank(),
+    text = ggplot2::element_text(family = "noto"),
+    plot.title = ggplot2::element_text(size = 14, hjust = 0.5),
+    axis.text.y = ggplot2::element_text(face = "bold",
+                                       margin = ggplot2::margin(r = -7))
```

21

```
+   )
```

Top Migration Routes by Number of Missing Persons

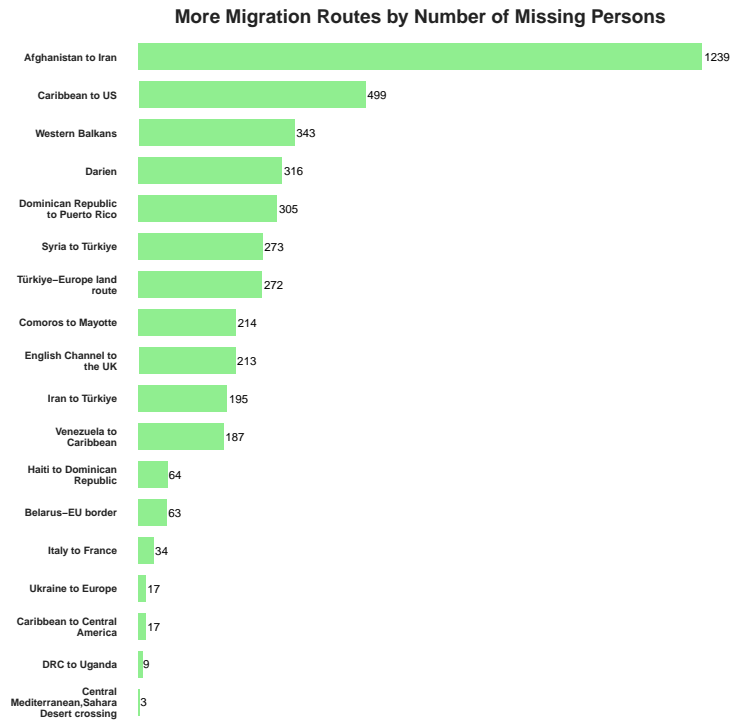| Route	
Central Mediterranean	22016
Uknown	9933
Sahara Desert crossing	5773
US–Mexico border crossing	4600
Western Africa / Atlantic route to the Canary Islands	3438
Western Mediterranean	3367
Eastern Mediterranean	2293
Horn of Africa to Yemen crossing	1830

```
> library(ggplot2)
> library(stringr)
> ggplot2::ggplot(route_comparison[9:nrow(route_comparison), ],
+                 aes(x = reorder(stringr::str_wrap(`Migration Route`,
+                                                    width = 20), total_missing),
+                     y = total_missing)) +
+   ggplot2::geom_bar(stat = "identity", fill = "lightgreen",
+                     width = 0.7) +
+   ggplot2::geom_text(aes(label = total_missing), hjust = -0.1,
+                      color = "black", size = 3) +
+   ggplot2::coord_flip() +
+   ggplot2::labs(title = "More Migration Routes by Number of Missing Persons",
+        x = "Route", y = "Total Missing") +
+   bbc_style() +
+   ggplot2::theme(
+     panel.grid.major = ggplot2::element_blank(),
+     panel.grid.minor = ggplot2::element_blank(),
+     axis.title.x = ggplot2::element_blank(),
+     axis.text.x = ggplot2::element_blank(),
+     axis.ticks.x = ggplot2::element_blank(),
```

```
+      text = ggplot2::element_text(family = "noto"),
+      plot.title = ggplot2::element_text(size = 15, hjust = 0.5),
+      axis.text.y = ggplot2::element_text(margin = ggplot2::margin(r = -8),
+                                          size=8, face = "bold")
+   )
```

More Migration Routes by Number of Missing Persons

Route	Value
Afghanistan to Iran	1239
Caribbean to US	499
Western Balkans	343
Darien	316
Dominican Republic to Puerto Rico	305
Syria to Türkiye	273
Türkiye–Europe land route	272
Comoros to Mayotte	214
English Channel to the UK	213
Iran to Türkiye	195
Venezuela to Caribbean	187
Haiti to Dominican Republic	64
Belarus–EU border	63
Italy to France	34
Ukraine to Europe	17
Caribbean to Central America	17
DRC to Uganda	9
Central Mediterranean,Sahara Desert crossing	3

The graphs show the top migration routes by the number of missing peo-
ple. The Central Mediterranean route is the most dangerous route, with 22018
missing people. The Sahara Desert crossing has 5773 missing people, and the
US-Mexico border crossing has 4500 missing people. The other routes have
fewer missing people. The graph shows that the Central Mediterranean route is
by far the most dangerous route for migrants, with a significantly higher number
of missing people than any other route.

4.11 Top 10 countries of origin by number of migrant deaths

```
> library(knitr)
> knitr::opts_chunk$set(echo=F, eval=T, message = F, warning=F, cache = F, fig=TRUE)

> #7 Top 10 countries of origin by number of migrant deaths
> library(ggplot2)
> library(dplyr)
> library(ggthemes)
> library(tidyr)
> library(stringr)
> # Split multiple countries into separate columns
> df_separated <- df %>%
+  separate(`Country of Origin`, into = paste0("Country", 1:10),
+           sep = ",", fill = "right")
> # Exclude Unknown/Mixed
> df_filtered <- df_separated %>%
+  filter(!(`Country1` %in% c("Unknown", "Mixed")) &
+          !(`Country2` %in% c("Unknown", "Mixed")) &
+          !(`Country3` %in% c("Unknown", "Mixed")) &
+          !(`Country4` %in% c("Unknown", "Mixed")) &
+          !(`Country5` %in% c("Unknown", "Mixed")) &
+          !(`Country6` %in% c("Unknown", "Mixed")) &
+          !(`Country7` %in% c("Unknown", "Mixed")) &
+          !(`Country8` %in% c("Unknown", "Mixed")) &
+          !(`Country9` %in% c("Unknown", "Mixed")) &
+          !(`Country10` %in% c("Unknown", "Mixed")))
> # Calculate top 10 countries of origin by number of deaths
> top_countries <- df_filtered %>%
+  gather(key = "Country_Index", value = "Country",
+         Country1:Country10, na.rm = TRUE) %>%
+  group_by(Country) %>%
+  summarise(total_deaths = sum(`Number of Dead`, na.rm = TRUE)) %>%
+  arrange(desc(total_deaths)) %>%
+  head(10)
> ggplot2::ggplot(top_countries, aes(x = reorder(Country, total_deaths),
+                                     y = total_deaths)) +
+  ggplot2::geom_bar(stat = "identity", fill = "darkred") +
+  ggplot2::geom_text(aes(label = total_deaths), hjust = 2, color = "white",) +
+  ggplot2::coord_flip() +
+  ggplot2::labs(title = "Top 10 Countries of Origin by Number of Migrant Deaths",
+        x = "Country of Origin",
+        y = "Total Deaths") +
+  bbc_style() +
+  ggplot2::theme(
```
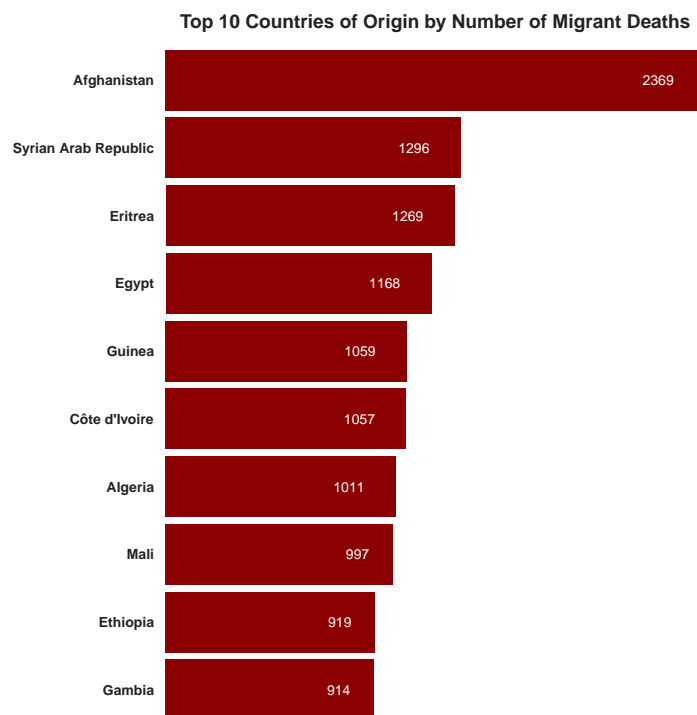
```
+    panel.grid.major = ggplot2::element_blank(),
+    panel.grid.minor = ggplot2::element_blank(),
+    axis.title.x = ggplot2::element_blank(),
+    axis.text.x = ggplot2::element_blank(),
+    axis.ticks.x = ggplot2::element_blank(),
+    text = ggplot2::element_text(family = "noto"),
+    plot.title = ggplot2::element_text(size = 15, hjust = 0.5),
+    axis.text.y = ggplot2::element_text(face = "bold",
+                                        margin = ggplot2::margin(r = -15))
+  )
```

Top 10 Countries of Origin by Number of Migrant Deaths

Country	Deaths
Afghanistan	2369
Syrian Arab Republic	1296
Eritrea	1269
Egypt	1168
Guinea	1059
Côte d'Ivoire	1057
Algeria	1011
Mali	997
Ethiopia	919
Gambia	914

The graph shows the top 10 countries of origin by number of migrant deaths. The country with the highest number of migrant deaths is Afghanistan, with 2369 deaths. This is followed by the Syrian Arab Republic with 1296 deaths and Eritrea with 1289 deaths. The graph shows that the majority of migrant deaths are from countries in the Middle East and Africa. These countries are often affected by conflict, poverty, and lack of opportunities, which can lead people to seek a better life in other countries.

5 Summary

1. Which regions have the highest number of migrant deaths and disappearances, and how do these numbers vary over time?

1.1. expectation

Data from 2014-2023 suggests the Mediterranean Sea route has consistently been the deadliest, with the highest number of deaths and disappearances. Over time there is a more number of missing migrants records [1]

1.2. What insights i get from my visualization

Over the time Mediterranean Sea has the highest number of death over the time.

2. What insights can we gain from analyzing the total number of missing migrants each year? which year is the highest number of the missing migrants and death.

2.1. Expectation

The total number of missing migrants each year provides insights into the scale and trends of migrant deaths and disappearances worldwide. The highest number of missing migrants and deaths in a single year was recorded in 2023, with at least 8,565 deaths.[2]

2.2. What insights i get from my visualization

The total number of missing migrants each year provides insights into the scale and trends of migrant deaths and disappearances worldwide and found that the highest number of missing migrants and deaths in a single year was recorded in 2016, with at least 8078 deaths.

3. Trend in the Number of Survivors Over the Years. which year is the more Survivors?

3.1. Expectation

Number of survivors has varied over the years. The highest number of survivors was recorded in 2023.[2]

3.2. What insights i get from my visualization

The number of survivors has varied over the years and found that the highest number of survivors was recorded in 2016 by 17113 survivors.

4. What are the gender ratios of missing migrants in different regions, and how do they vary across the globe?

4.1. Expectation

There are more male migrants reported missing than female migrants. This imbalance is particularly pronounced in certain regions, such as Southern Asia, where there are significantly more male migrants reported missing compared to female migrants and there is no record of missing migrants of central Asia. Number of survivors has varied over the years. The highest number of survivors was recorded in 2023.[2]

4.2. What insights i get from my visualization

Yes,There are more male migrants reported missing than female migrants. This imbalance is particularly pronounced in certain regions, such as Southern Asia, where there are significantly more male migrants reported missing compared to female migrants and there is no record of missing migrants of central Asia.

5. Which is the top impact location of death on number of missing migrants?

5.1. Expectation

The exact location with the highest number of missing migrants is the Mediterranean Sea, particularly the route between Libya and Italy. According to the International Organization for Migration (IOM), between 2014 and 2023, there were over 3129 reported deaths and disappearances of migrants attempting to cross the Mediterranean Sea, with the majority occurring on the Central Mediterranean route between Libya and Italy. Number of survivors has varied over the years. The highest number of survivors was recorded in 2023.[3]

5.2. What insights i get from my visualization

yes, The exact location with the highest number of missing migrants is the Mediterranean Sea, particularly the route between Libya and Italy but it has the 1589 missing migrants.

6. Which migration route has the highest incidence of missing persons, and how does it compare to others?

6.1. Expectation

Central Mediterranean Sea highest incidence of missing persons, mostly the route between Libya and Italy, has the highest incidence of missing persons.[3]

6.2. What insights i get from my visualization

yes Central Mediterranean Sea highest incidence of missing persons, mostly the route between Libya and Italy, has the highest incidence of missing persons.

7. Which is the most countries of origin by number of migrant deaths ?

7.1. Expectation

The countries with the highest number of migrant deaths were those experiencing conflict and instability, particularly Afghanistan, Rohingya refugees, and countries in Africa.[2]

7.2. What insights i get from my visualization

Yes, the most migrant death countries of origin by number of migrant deaths are in Afghanistan.

This report shows us the sad reality of people trying to find a better life but facing danger and even death along the way. The Mediterranean Sea is the most dangerous place for these journeys, and there are more men missing than women and many more It's important to help people on these journeys and make sure they are safe. We need to understand why people are leaving their homes and work together to make the world a fairer place for everyone.

References

[1] Map tracking migrant deaths and disappearances. https://gmdac.iom. int/map-tracking-migrant-deaths-and-disappearances. Accessed: 14 June 2024.

[2] Global Migration Data Analysis Centre. Map tracking migrant deaths and disappearances. Global Migration Data Analysis Centre, no date.

[3] 10 of the coolest visualizations of migration data. Migration data portal, no date.