



FINAL REPORT

MIGRANTS DATASET

DSC 465
**DATA
VISUALIZATION**
GROUP 3 – DATA PIRATES
Ayesha Anjum, Ayesha Zafar,
Swathi Babu, Umme E Hani

DSC 465 – DATA VISUALIZATION

FINAL REPORT

INTRODUCTION:

The data is part of a specific project called the Missing Migrants Project which tracks deaths of migrants, including refugees, who have gone missing along mixed migration routes worldwide. This data has been compiled from a variety of sources. Furthermore, these sources vary depending on the region and broadly include data from national authorities, such as Coast Guards and Medical Examiners; media reports; NGOs; and interviews with survivors of shipwrecks. The data set records incidents of missing persons and deaths of migrants.

The dataset has 2420 observations and 12 variables, 5 numerical, 6 categorical and 1 ordinal variable, where numerical variables are ID, Missing Persons, Dead, Longitude, Latitude and categorical variables are Cause of Death, Region of Origin, Nationality, Incident Region, Source, Source Reliability and Date is the Ordinal variable.

We mostly worked with the variables Missing Persons and Dead Persons to show the pattern with respect to region of origin, incident region and cause of death.

EXPLORATORY VISUALIZATIONS:

We cleaned the dataset first by choosing appropriate names for the cause of deaths, region of origin, incident region and nationality. We filled in the NA values with 0 for missing migrants and for dead migrants, we filled in with mode based on regions as less data was missing compared to the “missing” migrants count.

Checking the summary of the numerical variables.

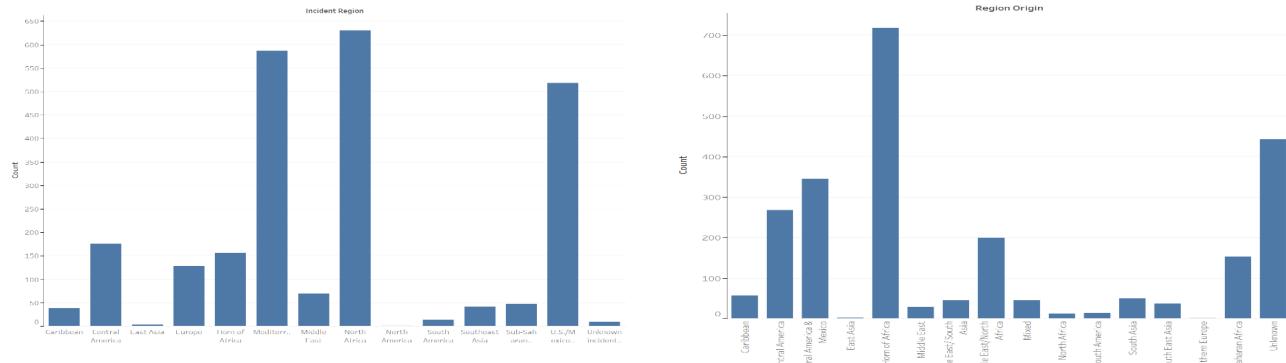
```
> summary(mig)

  id      cause_of_death    region_origin    nationality
Min.   : 1  Length:2420        Length:2420        Length:2420
1st Qu.: 28571 Class :character  Class :character  Class :character
Median :121178 Mode  :character  Mode  :character  Mode  :character
Mean   : 95926
3rd Qu.:144678
Max.   :184750

  missing      dead      incident_region      date
Min.   : 0.000  Min.   : 0.000  Length:2420        Length:2420
1st Qu.: 0.000  1st Qu.: 1.000  Class :character  Class :character
Median : 0.000  Median : 1.000  Mode  :character  Mode  :character
Mean   : 4.441  Mean   : 4.572
3rd Qu.: 0.000  3rd Qu.: 3.000
Max.   :750.000  Max.   :750.000

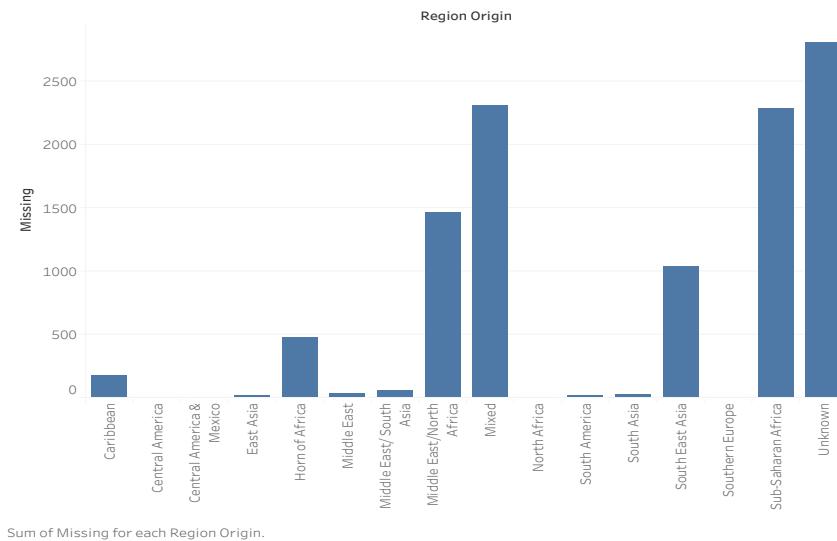
  source      reliability      lat          lon
Length:2420  Length:2420        Min.   :-26.22  Min.   :-117.07
Class :character  Class :character  1st Qu.: 19.54  1st Qu.: -97.04
Mode  :character  Mode  :character  Median : 29.35  Median :  14.47
                                         Mean   : 26.90  Mean   : -14.00
                                         3rd Qu.: 34.04  3rd Qu.:  32.01
                                         Max.   : 66.97  Max.   : 116.22
                                         NA's   :4       NA's   :4
```

We can see that a lot of data in the missing column is 0. This is mainly because we chose to replace the NA with zero to make analysis easier. So, the zeros do not necessarily mean that there weren't any missing people, they just did not have data. The dead column is better since we replaced NA values with the mode of each individual region.



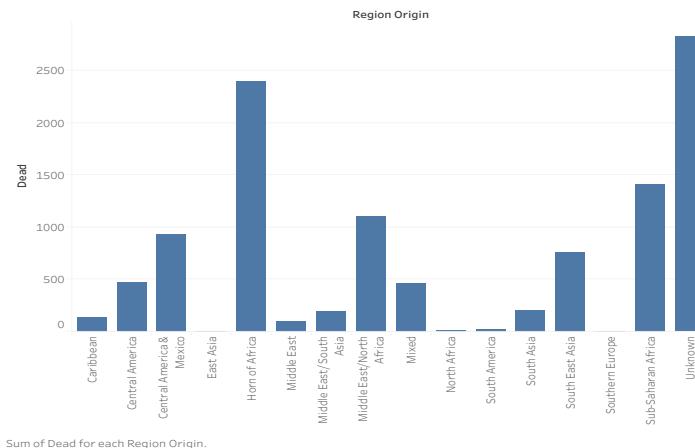
We can see that North Africa has appeared with the highest number as the incident region of the migrants, followed by the Mediterranean region, whereas the Horn of Africa has occurred with the highest number as the region of origin of the migrants.

Total missing migrants from each region of origin

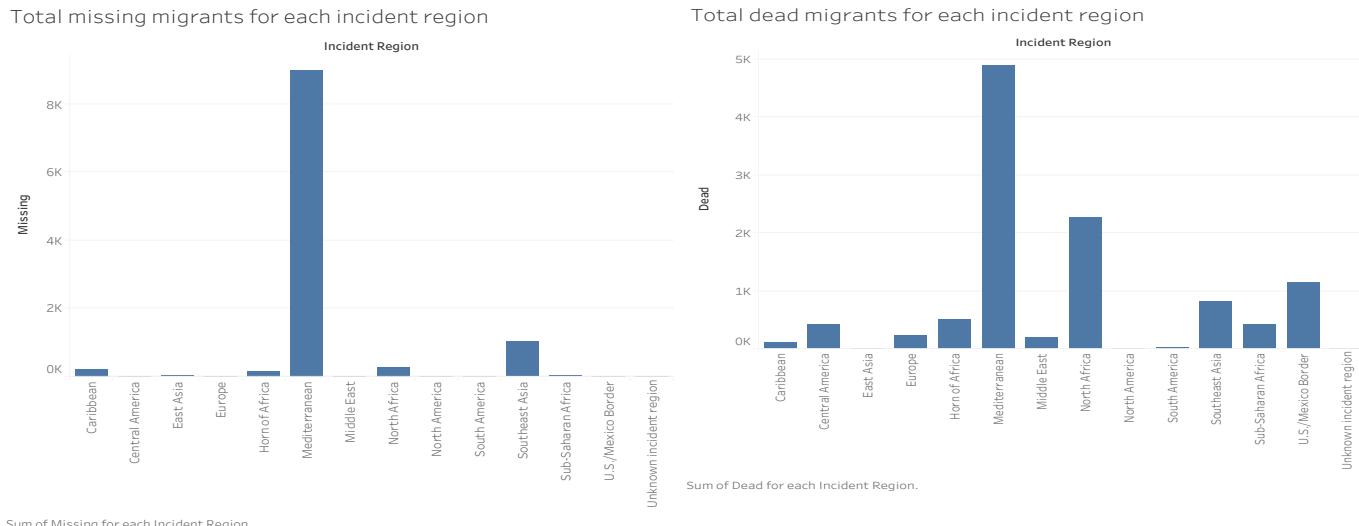


For missing, apart from mixed and unknown regions of origin, most migrants were missing from Sub-Saharan Africa, followed by the Middle East/ North Africa.

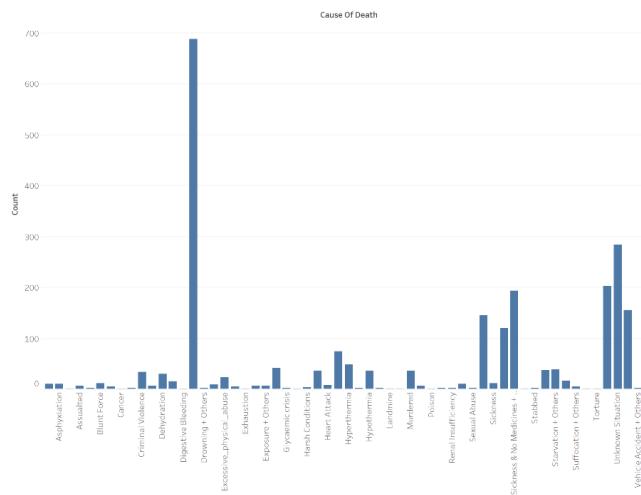
Total dead migrants from each region of origin



For the deceased, apart from the unknown region of origin, the Horn of Africa has the highest count, followed by Sub-Saharan Africa.



These bar graphs depict that the Mediterranean region has the highest number of missing and dead migrants.



The cause of death for most migrants is Drowning.

From this information, we found that we can explore the missing and dead variable for specific regions—especially for regions that had high missing or death rate by glyph maps with color/size, bar plots, heat maps etc. We also thought to look at how date/month/year plays a role(a trend) in the number of missing and dead migrants using line graph.

EXPLANATORY VISUALIZATIONS AND THEIR ANALYSIS:

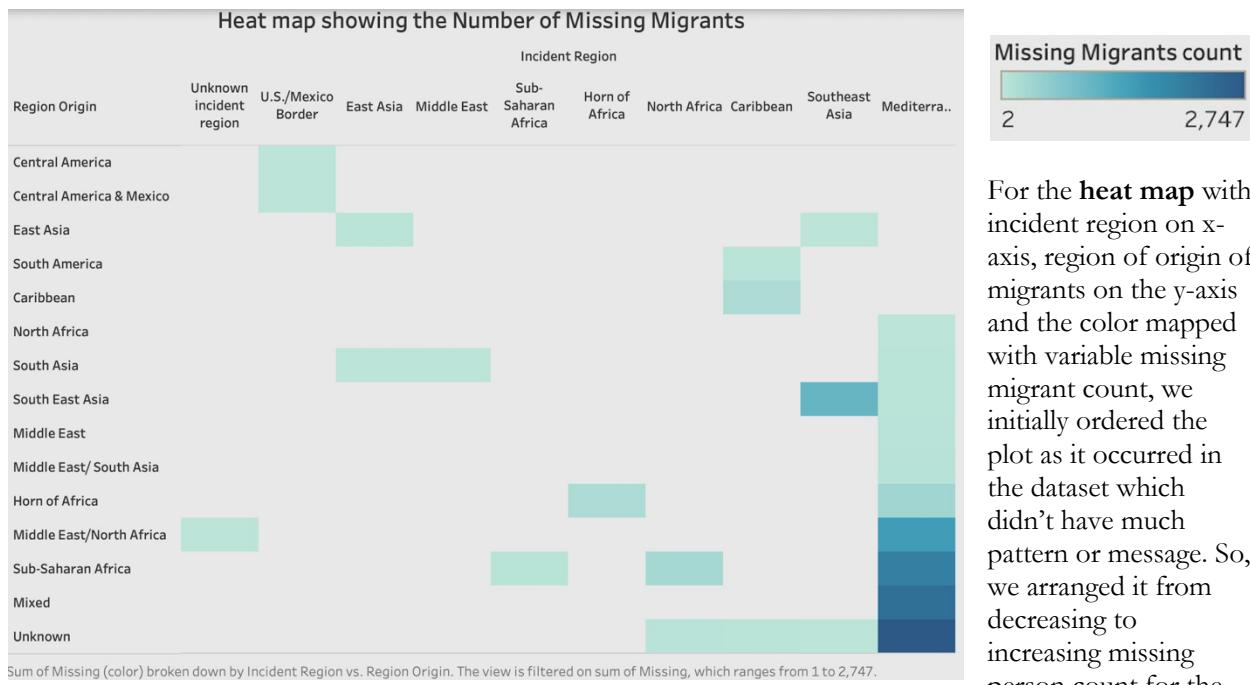
We tried to show the pattern for two important variables – Missing migrants and Dead migrants. We believed we could convey the message better when they are looked at separately.

Missing Migrants –



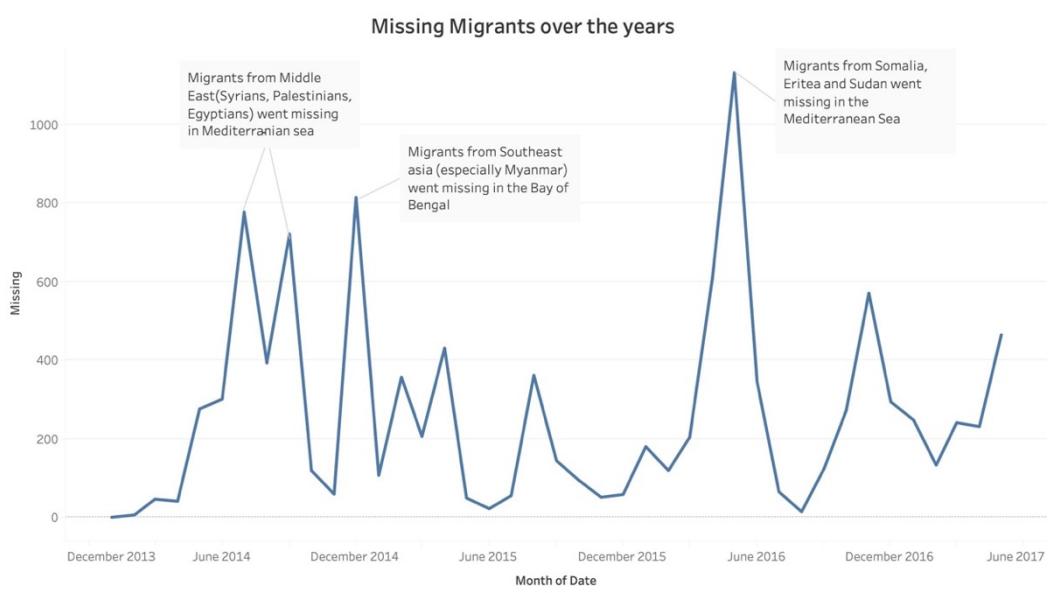
We used Tableau to create this **Map Plot** for the count of Missing Migrants over the incident region, the count as in number is clearly represented on the map. The scale shows the smallest count starts with 5 and goes over as high as 8989 in the Mediterranean region, followed by Southeast Asia with a count of 1032 lastly, the least number has been recorded in U.S/Mexico with a count of 5, followed by Middle East with a count of 11. The number of deceased are represented by size. We eliminated the Unknown regions from the plot as it was difficult to decode which region comes under the “unknown” region.

By doing some research over the internet, we concluded that these numbers could be high because This project gave us clarity as to how migrants crossing borders over the Mediterranean Sea, died during their travel and how their bodies went unidentified, missing or even sank down into the sea, hence giving rise to a huge number of missing migrants from the Mediterranean Sea. For the Southeast Asia, research on the ongoing project stated that, the majority of missing migrants, recorded on migratory routes in Asia where migrants were trying to reach safety and better opportunities on sea migration routes to other countries. There could be many primary reasons for this including sickness or physical assault by smugglers or traffickers.



For the **heat map** with incident region on x-axis, region of origin of migrants on the y-axis and the color mapped with variable missing migrant count, we initially ordered the plot as it occurred in the dataset which didn't have much pattern or message. So, we arranged it from decreasing to increasing missing person count for the pattern to show up clearer.

We used tableau to create this heat map and we decided to use sequential color scheme as this variable was continuous and we think that the two end extremes need not be emphasized with divergent color scheme. In the heat map, we can see that the highest trend with the count of 2747 for the incident region Mediterranean region are migrants originating from unknown region. The next highest missing count seems to be migrants originating from mixed regions and they also went missing in the Mediterranean region. We can see from the heat map that migrants originating from various regions in the dataset go missing in the Mediterranean region. There are also migrants originating from Southeast Asia who went missing in Southeast Asia itself. The least number of missing migrants with the least count of 2 is from many origins few of them being, middle east, central America, north Africa and so on is shown using the lightest shade of color.

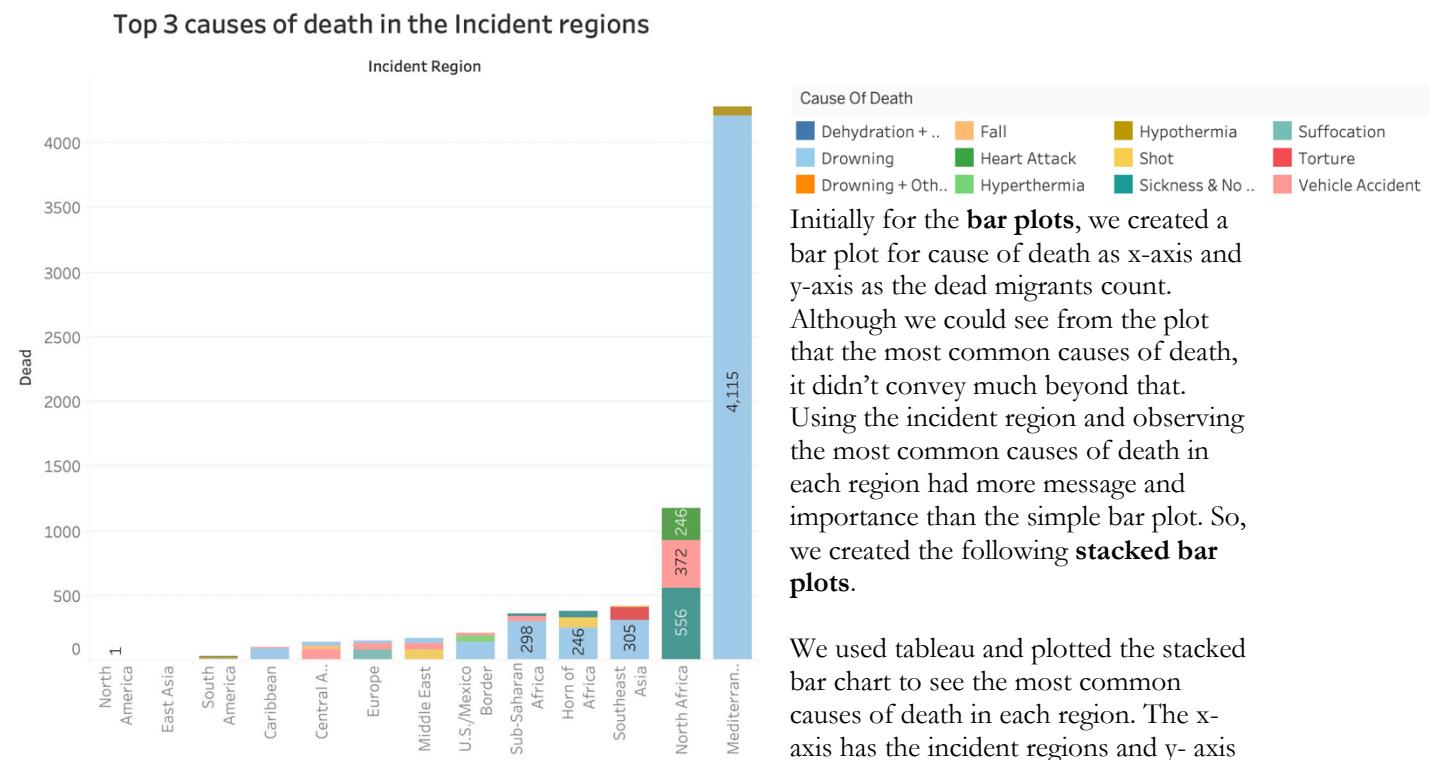


We first created the **line plot** with each peak marked with each missing count as text for all points and also marked each month on the axis (many minor and major tick marks). This made the graph too cluttered. So, we reduced the tick marks in the x-axis and removed the text from all

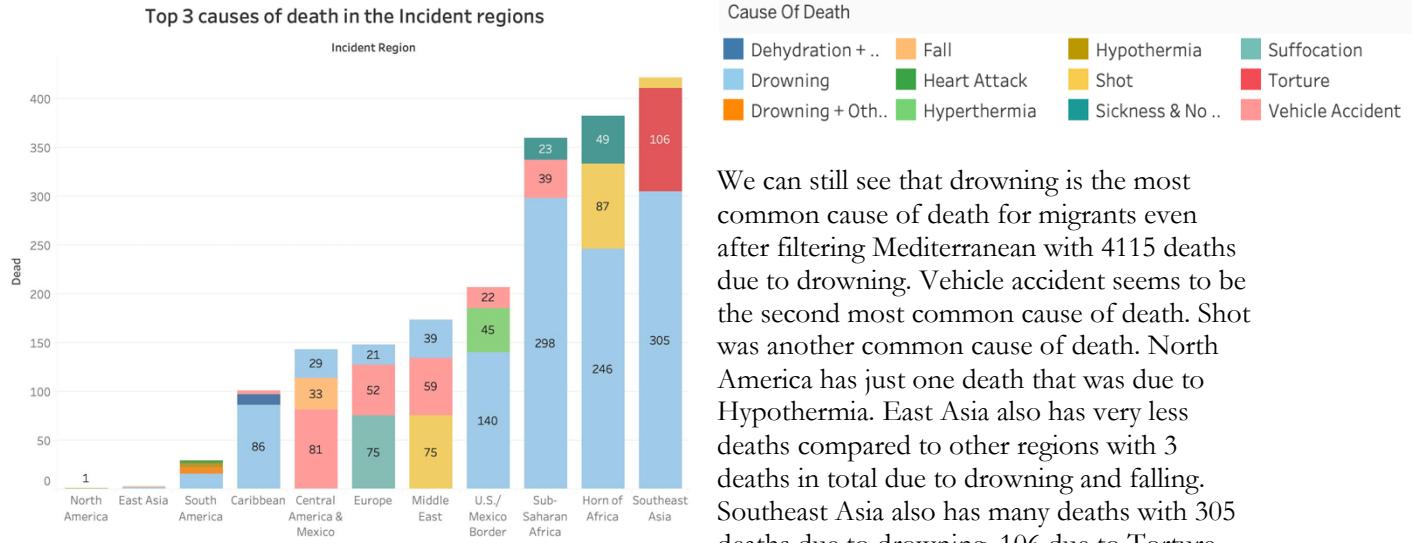
peaks. Also, doing some extra research using the internet, we were able to come up with specific events for major peaks in the plot.

We used tableau to create this line plot to show the trend of the number of missing migrants over the years with x-axis as months of the years 2013-2017 and y-axis as the sum of missing migrants. The specific events were annotated for prominent peaks. There are peaks in the number of missing migrants during certain months over the years – July 2014, September 2014, December 2014, and May 2016. This dataset is not exhaustive. There are more missing and dead migrants than mentioned here. Keeping that in mind, these peaks correspond to some events that took place over the years. July 2014 and September 2014 in this dataset are migrants originating from Middle east went missing in the Mediterranean Sea. Also, migrants from Sub-Saharan Africa went missing here as well. December 2014 , migrants from Southeast Asia went missing in the Bay of Bengal. May 2016, migrants from mixed regions went missing in the Mediterranean Sea.

Dead Migrants –

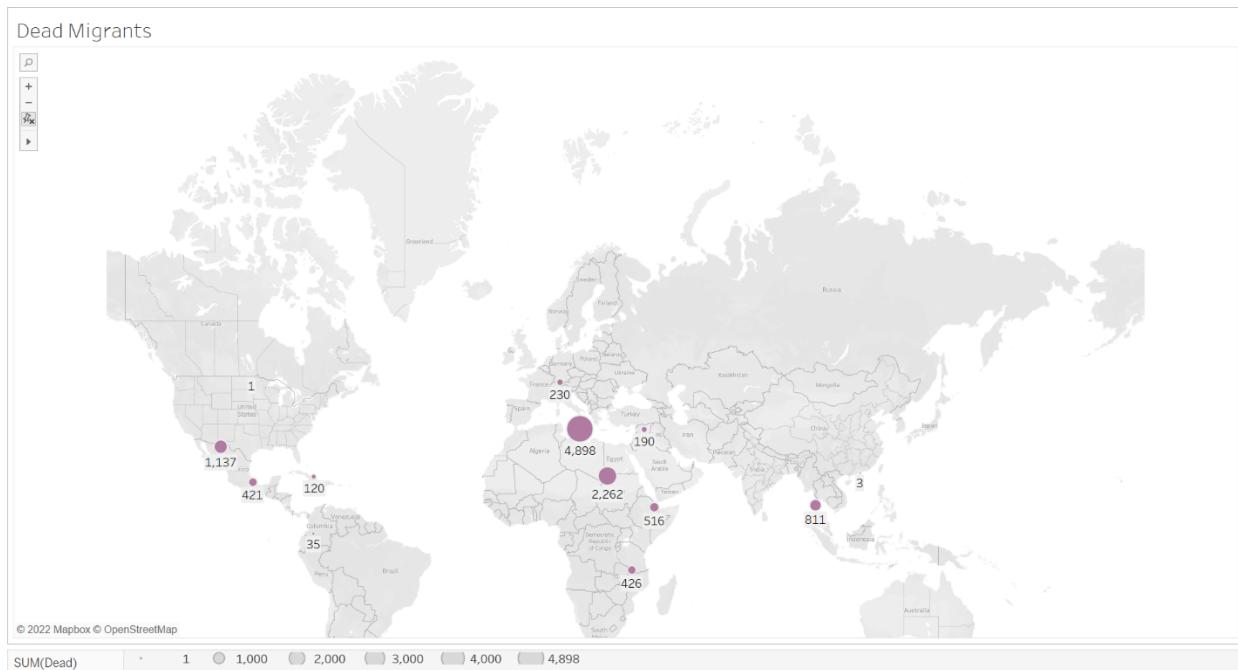


We also noted that we couldn't observe the common deaths of the other regions. So filtering the last two regions North Africa and Mediterranean, we can see the stacked bar plot below.



We can still see that drowning is the most common cause of death for migrants even after filtering Mediterranean with 4115 deaths due to drowning. Vehicle accident seems to be the second most common cause of death. Shot was another common cause of death. North America has just one death that was due to Hypothermia. East Asia also has very less deaths compared to other regions with 3 deaths in total due to drowning and falling. Southeast Asia also has many deaths with 305 deaths due to drowning, 106 due to Torture

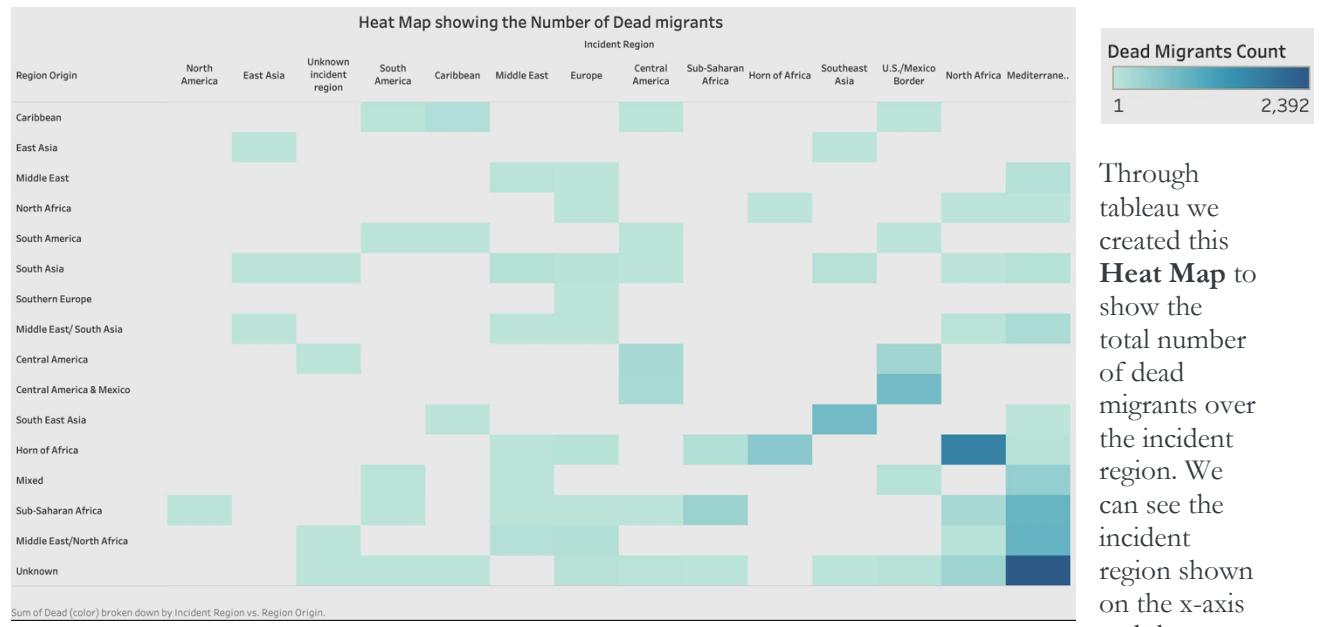
and 10 shot. It is also important to note that Southeast Asia is the only region with torture as one of the common causes of death.



We used Tableau to create this **Map Plot** for the count of Dead Migrants over the incident region, the count as in number is clearly represented on the map. The scale shows the smallest count starts with 1 and goes over as high as 4898 in the Mediterranean region, followed by North Africa with a count of 2262, lastly, the least number has been recorded in North America with a count of 1, followed by East Asia with a count of 3. The number of deceased are represented by size. We eliminated the Unknown regions from the plot as it was difficult to decode which region comes under the “unknown” region.

By doing some research over the internet for this plot too, we concluded that these numbers could be high because there is a long history of migration via the Mediterranean. Human mobility in all directions across the Mediterranean has occurred for thousands of years. More recently, since at least the mid-1990s, thousands of people each year have crossed the Mediterranean by boat from the northern coasts of Africa and Turkey to seek asylum or to migrate to Europe if they do not have the documentation required by the countries of

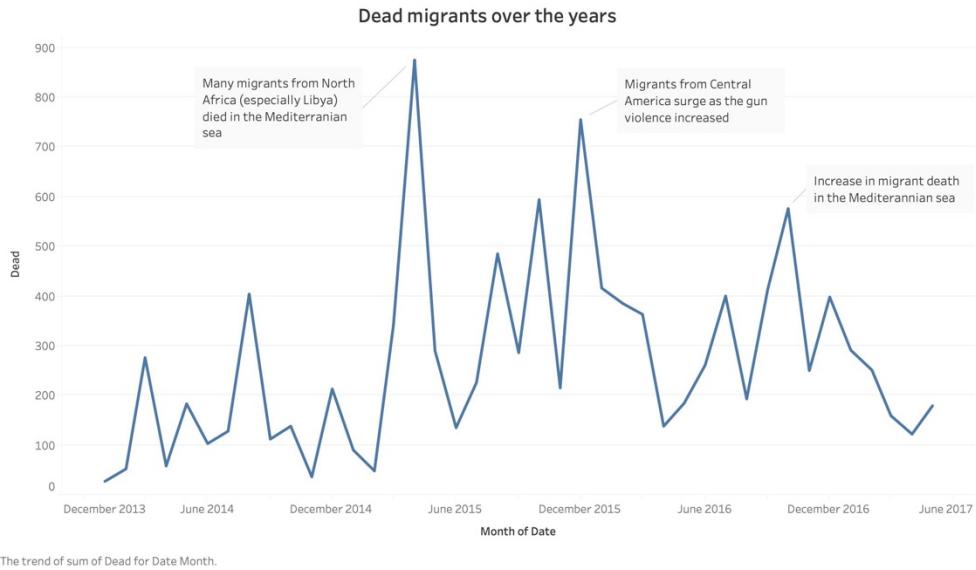
destination. Due to this main reason, Mediterranean Sea has the highest number of deaths leading to unnatural deaths of migrants due to unforeseen weather conditions. Another reason for Africa recording the second highest number is that, the migratory journeys prior to these Mediterranean crossings were also highly risky, as they often involved crossing remote terrains such as the Sahara Desert, in North Africa and residing, at least temporarily, in countries such as Libya where conditions for migrants are quite dangerous, hence leading to many deaths.



Through tableau we created this **Heat Map** to show the total number of dead migrants over the incident region. We can see the incident region shown on the x-axis and the

region for the origin of the migrants could be seen on the y-axis.

At first when we ordered the map through our data it didn't show proper representation of what we wanted. We then rearranged the regions - rearranging the rows and columns; bringing the Caribbean closer to the Middle East, Horn of Africa shifted below and so on. The map shown above is the changed final representation showing clearly the data we want to convey. Variables being continuous we went ahead with the sequential color scheme. In the heat map we can see that the highest number of dead migrants recorded is from the incident region being an unknown region and their origin being from the Mediterranean with the total count of 2392. The second highest number of dead migrants seems to be from the origin of North Africa and the incident region being the Horn of Africa, can be seen through a slightly lighter shade of color in the heat map. The least number of dead migrants with the count of 2 spread around many different origin areas and incident regions is shown using the lightest shade of color in the heat map. For instance, origins being from South America, Middle East, Sub-Saharan and incident regions being Caribbean, Mixed, and Horn of Africa. As shown in one of the bar graphs in our project and after doing some research online we discovered that the main cause of the migrants death is through drowning which is understandable if we consider the source of transportation as air or water, there are higher chances of mishaps that could take place during the travel of long distances and long hours between the origin area and the incident region of the migrants.



We first created the **line plot** with each peak marked with the actual dead migrant value as text and also marked each month on the axis (many minor and major tick marks). This made the graph too cluttered just like the missing time plot. So, following the same steps as missing line plot, we got the above plot.

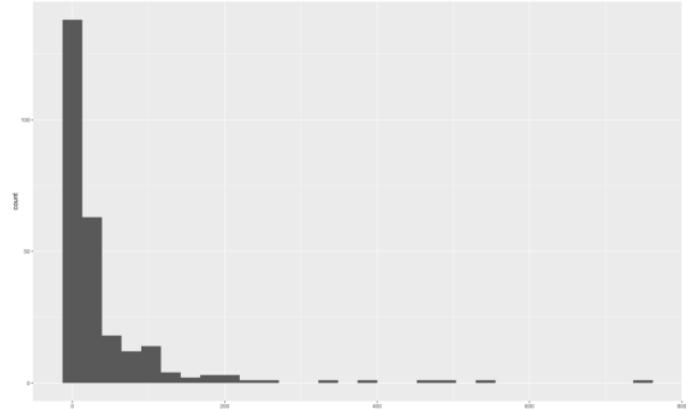
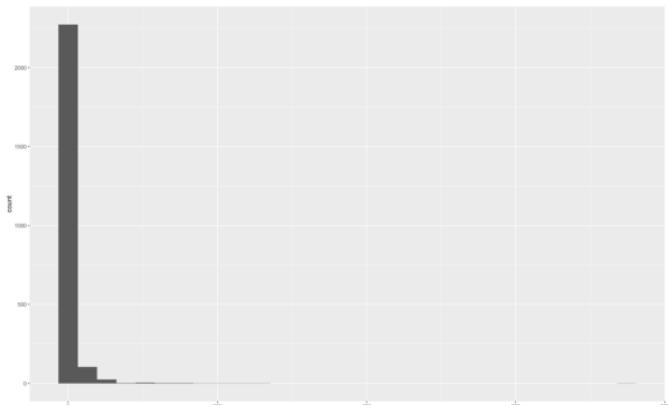
We used tableau to create this line plot to show the trend of the number of dead migrants over the years with x-axis as months of the years 2013-2017 and y-axis as the sum of dead migrants. The specific events were annotated for prominent peaks. There are peaks in the number of dead migrants during certain months over the years – April 2015, December 2015, and October 2016. In April 2015, migrants from North Africa died in the Mediterranean Sea. In December 2015, migrants from Central America died at the US/Mexico border. There were migrants from Horn of Africa who died in North Africa. In October 2016, migrants from various regions died at the Mediterranean Sea (like Sub-Saharan Africa)

Conclusion:

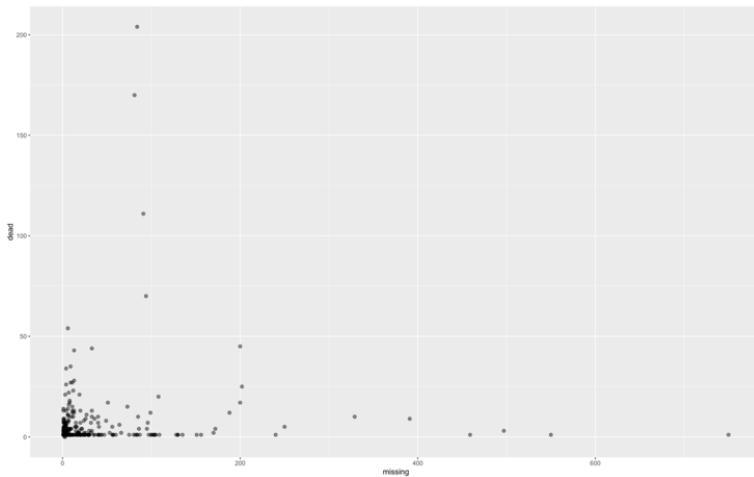
As we conclude our project on data visualization, we can see that the visualizations have a wide range of potential applications in several fields, while also being exposed to their practical and ethical complexity. Developing an effective and ethical data visualization is a complex yet interesting procedure. In our data pirates project, we dealt with missing and dead migrants data over different regions. With a total of 2420 observations and 12 variables we developed Heat maps, Line plots, Map plots and Bar plots; representing the proper visual comprehension we wanted to convey. For instance, if we look at the explanatory plots from our project, we can see that Mediterranean region has the highest number as being the incident region for the migrants missing and the highest number of death migrants was from the Mediterranean region. Through this procedure we were also able to figure out that drowning was the most common cause of this, also looking precisely through the years/months using line graphs. Interestingly, Heat maps and map plots represent the count of the missing and dead migrants over the origin and incident regions respectively using shade of colors helping us see from which region the highest count of migrants were reported. We were able to visualize the data over a specific period of time, which helps us understand during which year the missing/dead migrants reported were at their peak. We became conscious of the fact that the slightest change and rearrangement of data can allow us to represent more accurate visualization. By the end we were able to learn how to plot multiple aspects, understanding the sorting and arranging it from decreasing to increasing the person's count to get the pattern to visualize accurately. Through our project we were able to analyze the origin and incident area, highest peak time(reported records) of the missing and dead migrants. If we had more time, we would have tried to include an interactive plot where we can observe missing or dead migrants by varying different parameters like Region of origin, Incident region and cause of death.

APPENDIX:

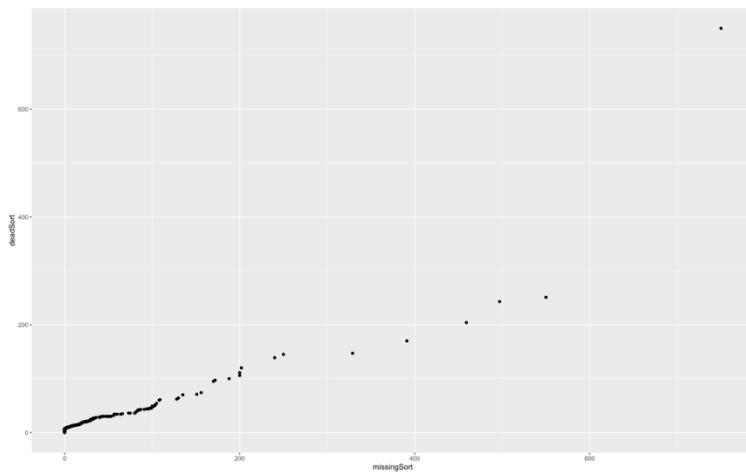
(1) Rest of the exploratory analysis results –



The variables dead and missing are skewed to the left.



This is a scatter plot for missing and dead migrants. We can see that these variables do not have a linear relationship and low correlation.



From this quantile-quantile plot, we can see that they do not have a normal linear relationship with respect to each other.

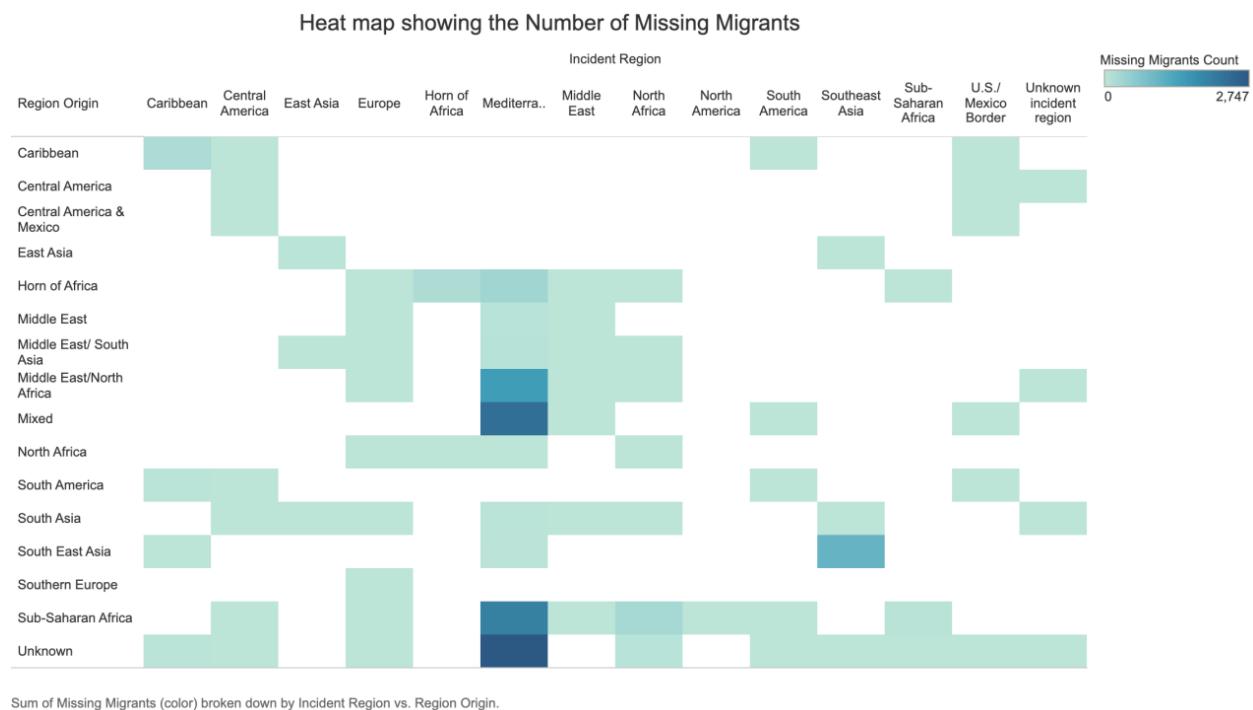
(2) Individual Reports –

Ayesha Anjum

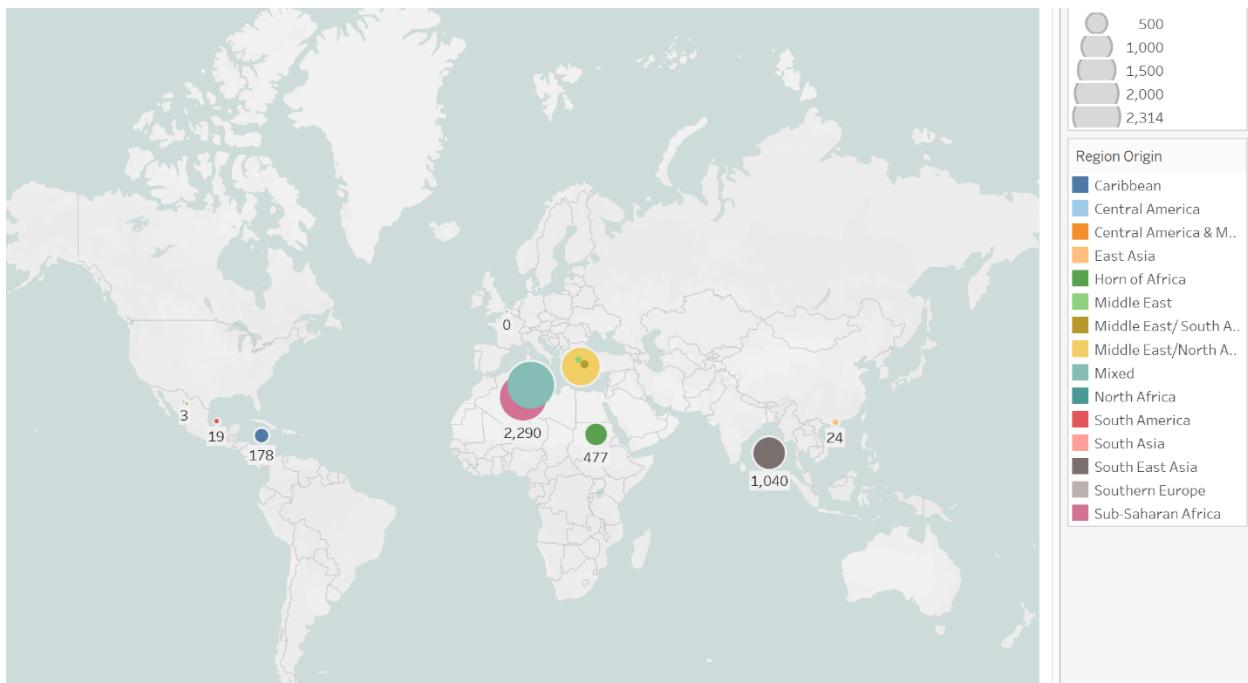
After choosing the Missing migrants dataset, I began going through it, learning each of the variables, and their possible relationships.

In milestone 2, I performed cleaning on the dataset and performed exploratory analysis, which consisted of a bar graph that stated the cause of death of the migrants. Drowning is the highest, followed by unknown situation being second highest.

For the exploratory part, I performed initial drafting for the heat map of missing migrants over the region of origin and incident region, which depicted that the highest number of migrants were found missing in the Mediterranean region whose origin was unknown. Later the heat map was redrafted for a better message.



Also drafted a geographic map for the missing migrants based on the region of origin. The highest number of missing migrants was from Sub-Saharan Africa, followed by Southeast Asia being the second highest.



For the group presentation, I redrafted the geographic plot for dead migrants by combining the incident regions so that it could give a better understanding. Also revised the heat map for the dead migrants. The plots concluded that the Mediterranean region is the deadliest migration route globally, as nearly 4898 deaths were recorded from 2014-to 2017.

This course taught me about visualizing data in Tableau and R studio. I learned how to design visuals that deliver the message to their audience. It improved my grasp of visual encodings, color schemes, and RStudio packages such as ggplot, tidyverse, etc. It provided a wonderful chance to model my results and examine all the considerations involved in selecting a visualization. I'm glad I got to be a part of Data Pirates, and I'm grateful to every team member who helped this project go smoothly.

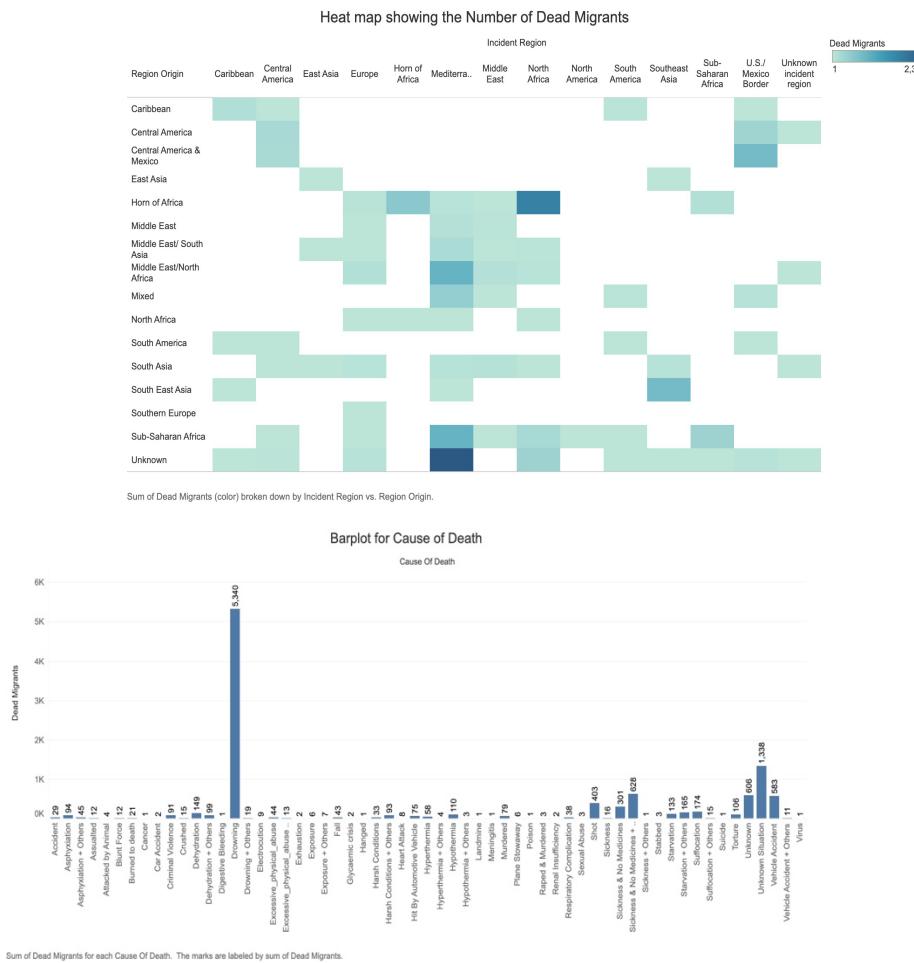
Ayesha Zafar

I started the Missing Migrants Project by first doing the exploratory part, which were two different bar plots. I chose bar plots because they are easy to understand by the target audience and wouldn't consume one's time in trying to figure out what's being portrayed in the plot.

One was a bar plot which displayed information about total number of Dead Migrants from Each Region of Origin. This bar plot was very clear enough in showing that Horn of Africa had the highest number recorded for deaths in that region. The second highest number of deaths were recorded in Sub Saharan Africa where the number was close to 1500. The least number was recorded in East Asia. Another Bar plot done by me illustrated the total number of dead migrants for each incident region, where the highest number was recorded in the Mediterranean Region with a count of almost 5k. This was followed by North Africa with a count of almost 2.5k.

I also worked on the explanatory part of the project, where I initially drafted the heat map for Dead migrants, over the region of origin and Incident Region. This map showed which region had the highest number of dead migrants and in which region, The shade chart on the right shows that the region with highest number of deaths is in the darkest shade, which was Unknown. Later, as the unknown region didn't make sense, we had to change this plot to a much better version. I have also created the Bar Plot which shows the different causes of death. In this plot, we see that the highest number of deaths were recorded due to drowning, where

the figure stood at 5340, followed by an unknown situation for the death being caused. This bar plot was later modified into a much clear and understandable plot.



I contributed to the presentation, where I gave the introduction to our project and, spoke about the Cleaning of the project, the important variables involved in the project and why this project is significant. I then talked about the Bar Plot in the presentation, stating reasons as to why the number could be high in the category of drowning.

Finally, I drafted the map plot for missing and dead migrants, making necessary changes as mentioned by the professor, to adjust the legend and remove the unknown values from the plots so that they are much easier to understand by the target audience.

From this Course (Data Visualization), I learned a lot about different types of graphs, and how to create and visualize them. Tableau was a very helpful tool where we can use different types of variables and create various types of graphs. Also, color choices schemes help a lot to visualize graphs easily. Moreover, R studio is also a nice tool where we can create plots like ggplot, mosaic plot and so on. The Professor created a discussion tab to expand our data visualization where every student must post their beautiful visualization and comment on each other's post which also helped me a lot to understand other people's perspectives and their visualization. In this group project also, I learnt a lot from my teammates by applying their ideas and suggestions, likewise sharing my ideas with them. However, initially it was hard for

me to decide which type of visualization is best for the audience to understand but now after finishing up the course, I learned many different types of techniques that to create better visualization.

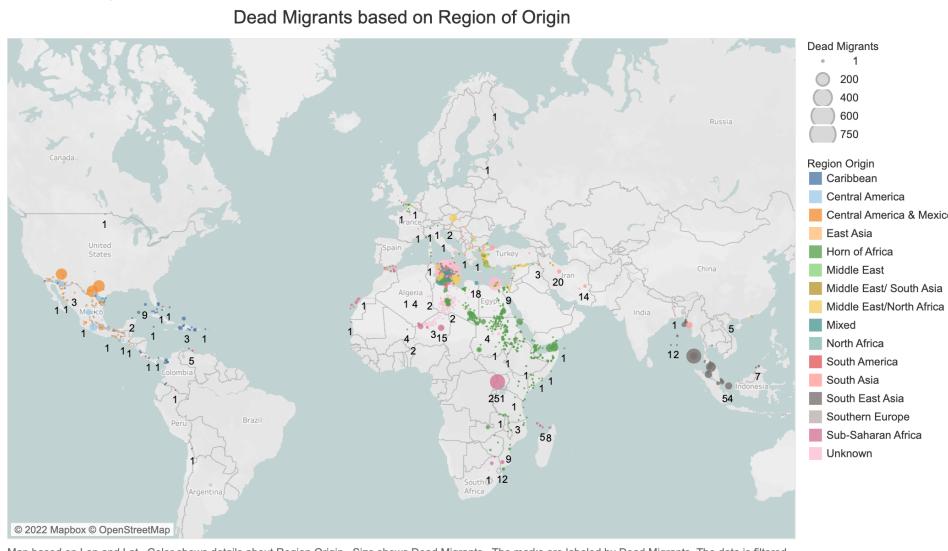
Swathi Babu

After deciding on the dataset, I cleaned half the dataset, particularly the cause of death, missing persons, dead persons and the reliability columns. Ayesha Anjum cleaned the rest of columns.

For the exploratory analysis, I came up with the summary of the columns in the dataset (in the Exploratory section) and noted down my insights from the output about how skewed the missing and dead variables are. I also provided the reasons for the skewness. It was due to our decision to change all 'NA' values in missing to 0 as around 88.8% of the missing column was NA. We did not want to make any assumptions by using mean/median/mode to fill the column as such high amount of data was missing. We did not want to delete such high amount of data either so we filled it in with 0. For dead variable since only 4.2% of the column data was missing, we filled it with mode based on the region.

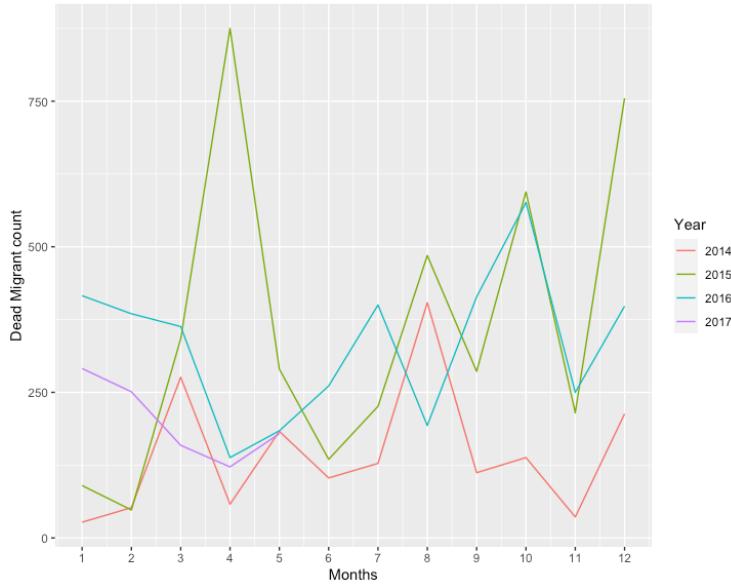
To further look into this, I created the two histograms to see the distributions of the numerical variables dead and missing(in the Appendix above). I also plotted the scatter plot between the 2 important numerical variables dead and missing count(in the Appendix above). To compare the distributions of these variables, I plotted the Quantile – Quantile plot to see that the relationship between them is not a linear normal relationship.

For the milestone 3 that is the explanatory plots, I worked on the initial drafts of the time plots for missing variable and for the dead variable. I chose to do these plots to show the trend of how the number of missing/dead migrants increases and decreases over the years. I also created the initial draft of the glyph map plot plotting dead migrants count on the map using the latitude and longitudes variables. I used size to represent the number of dead migrants in that location and used color to signify the region of origin (Plot is as follows)



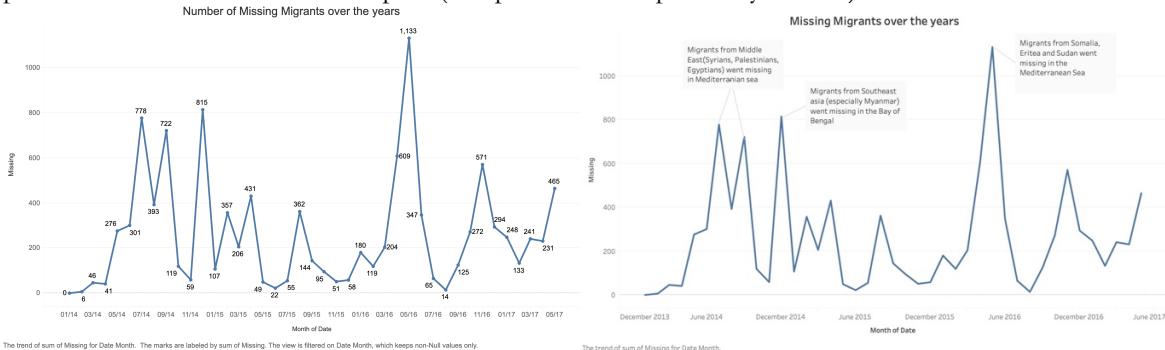
For homework 4 , I did an animated line plot showing the change in the dead migrants over the months where each year was represented by a separate line. This plot was not included in the final report as it did not convey as much information as the time plots already part of the report. That showed the trend over the years better than the plot below.

Line Plot to show the Dead migrant count over the years



Based on the feedback received for the explanatory plots, I made the appropriate changes to the line plots. I deleted the annotations of the dead and missing migrants value that I added at each point. I also researched online to find events that correspond with the line plots peaks. I tried to compare our dataset and real-life incidents and tried to find the event/cause of the prominent peaks in both the line plots. Then using the insights from the plot and the extra research, I presented these 2 plots for the final presentation.

After listening to the feedbacks on our presentation, I realized that some of the plots I created are quite cluttered. Sometimes to make sure that the information is conveyed accurately, I overcompensate by adding texts, many tick marks in x-axis and y- axis etc. Keeping this point in mind, I redrafted the line plots for the final report, reducing the clutter as much as possible. Removed some of the major and minor ticks marks on the x-axis and just labelled few months of each year rather than listing all months like I previously did. I have attached one of the original drafts and the final draft below to show how much reducing the clutter in the line plots made it better in the final report (line plots in the explanatory section)



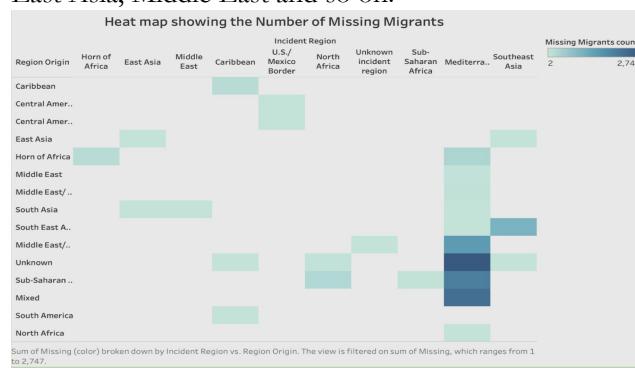
With the suggestions given for the bar plot, for the final report, I used tableau and created the hierarchical (stacked) bar plot that shows the three common causes death in each region of origin. This was easier in tableau as filtering is extremely simple in tableau. I also created the final draft of the heat map for the dead migrants. Previously, patterns weren't very clear. So, I rearranged it to make the pattern useful and also moved the legend down to decrease the white space.

I learnt a lot about organizing visualizations from this project. What plots to put together and what kind of plots to use for what kind of data. I tried around 5-10 different methods/plots before deciding on one plot

for a set of variables. That helped me learn about a lot of different plots. I also realized all the common mistakes I make when I create visualizations. I've been making visualizations since the first quarter and I noticed a lot of mistakes that I did back then. I also noticed that I clutter the plots sometimes and this project definitely helped me with identifying clutters and removing them better now. I also learned the software R and Tableau very well, learnt how to manipulate data and which software is best for what circumstances.

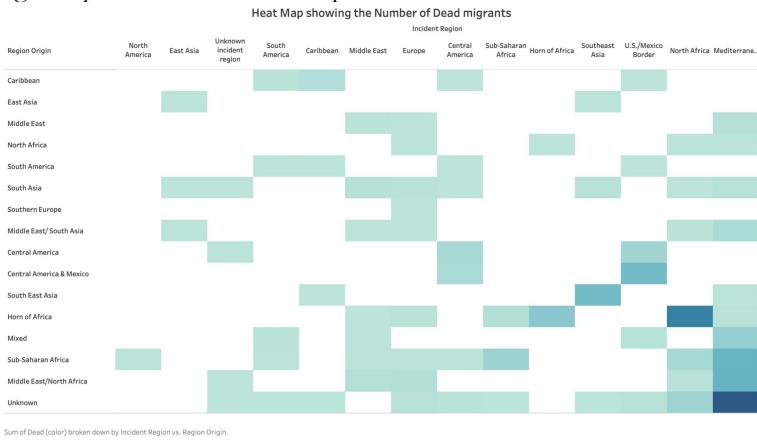
Umme E Hani

From finding the dataset to writing the conclusion, it has been an interesting learning experience with data visualization. As soon as the data set was finalized we discussed how we wanted to portray the data that is comprehensive to the reading audience. After understanding the data cleaning and exploratory plots, I moved forward with my part in the project by researching how a specific data is visualized by other people, where I learned that When we click on a certain region on the interactive dashboard, all of the data in terms of trends and other attributes are displayed in tabular or other visualization charts. It can be a real-time interactive database connection, with any new database addition shown in the visualization dashboard. This seemed like a really efficient way to cover all data without overcrowding in just one plot. I participated in writing the description of the visualization 1 where I explained through the heat map, line graph and the map plot the number of missing migrants with respect to origin region and incident region - Mediterranean and unknown region being the one with highest count, number of missing migrants over the years - 05/16 being the peak time with the total count of 1,133 and count of missing migrants based on their region of origin - with highest number of missing migrants from the Middle East / South Asia with the total count of 750. In the project video presentation I contributed in explaining about the heat map showing the number of missing migrants. After making the required changes according to the feedback provided by the professor and re-ordering the rows and columns and an attempt to make the background color less white, we can see that highest number of missing migrants with a count of 2,747 is from Mediterranean being the incident region and Unknown being the origin region and the least number with a count of 2 being from multiple regions of origins such as South Asia, East Asia, South America and so on and the incident regions being Caribbean, East Asia, Middle East and so on.



After the presentation, for the final project report, after making the changes in the heat map showing the number of dead migrants, I wrote the description about what the heat map visualized. Below is the changed heat map shown. After rearranging and making the necessary changes as suggested in the project feedback we can see the changed data representation, with the highest count of dead migrants now being 2,392 and the

lowest count of dead migrants being 1. Apart from the heat map, I added the conclusion and added the



reference links to the final project.

Data visualization course gave me an opportunity to understand and visualize the datasets through various aspects. I was able to understand that visualization is really a comprehensive and an interactive procedure to provide details about the data you want to represent. Creating and understanding new methods to represent the data really opens a wide range of information to a keen observer. Learning about exploratory and different plots was an interesting part, heat maps being the most I was drawn towards, I found it fascinating how you can visualize the data using shades of colors that represent any aspect you want. It being my first time using Tableau, with the help of this project I was able to really understand the data visualization. It helped me realize how critical it is to show the visualizations in such a way that the audience can understand, also allowing us to make the visualizations to be presented in a beautiful way. Professor Eli Brown's feedback was of immense help in making our project more understandable and accurate. I believe that the most important thing that I learned during the project is not to clutter the visualization we are trying to represent and create an accurate visualization.

(3) R Code –

```
# DATA PIRATES -Group 3 - Missing Migrants Dataset
library(ggplot2)
library(dplyr)
library(base)
#Importing the datatset
mig = read.csv("/Users/swathib/Desktop/Spring Term 2021-22/DSC 465 Data Visualization/Final Project/Final Project extra files/MissingMigrantsProject.csv")

#changing column name
colnames(mig)
names(mig)[names(mig) == "affected_nationality"] = "nationality"

#Cleaning
#Cause of death
mig$cause_of_death[mig$cause_of_death=='AH1N1 influenza virus, while stuck at border'] =
'Virus'
mig$cause_of_death[mig$cause_of_death=='Asphyxiation (Silica sand inhalation)'] = 'Asphyxiation'
mig$cause_of_death[mig$cause_of_death=='Asphyxiation and crushing'] = 'Asphyxiation + Others'
mig$cause_of_death[mig$cause_of_death=='Assaulted by smugglers']= 'Assualted'
```

mig\$cause_of_death[mig\$cause_of_death=='Attacked by hippopotamus']= 'Attacked by Animal'
mig\$cause_of_death[mig\$cause_of_death=='Beat-up and killed']= 'Assaulted'
mig\$cause_of_death[mig\$cause_of_death=='Beat-up and thrown into river']= 'Assaulted'
mig\$cause_of_death[mig\$cause_of_death=='Beaten to death on train']= 'Assaulted'
mig\$cause_of_death[mig\$cause_of_death=='Beating/shot by traffickers']= 'Assaulted'
mig\$cause_of_death[mig\$cause_of_death=='Blunt force head trauma']= 'Blunt Force'
mig\$cause_of_death[mig\$cause_of_death=='Blunt force injuries']= 'Blunt Force'
mig\$cause_of_death[mig\$cause_of_death=='Boat fire']= 'Burned to death'
mig\$cause_of_death[mig\$cause_of_death=='Bronchial aspiration']= 'Respiratory Complication'
mig\$cause_of_death[mig\$cause_of_death=='Bronchopneumonia']= 'Respiratory Complication'
mig\$cause_of_death[mig\$cause_of_death=='Burned to death hiding in truck']= 'Burned to death'
mig\$cause_of_death[mig\$cause_of_death=='Burns and Suffocation']= 'Burned to death'
mig\$cause_of_death[mig\$cause_of_death=='Burns from cooking gas explosion in connection house in Libya']= 'Burned to death'
mig\$cause_of_death[mig\$cause_of_death=='Car accident']= 'Vehicle Accident'
mig\$cause_of_death[mig\$cause_of_death=='Cervix cancer']= 'Cancer'
mig\$cause_of_death[mig\$cause_of_death=='Clubbed/beaten to death']= 'Assaulted'
mig\$cause_of_death[mig\$cause_of_death=='Criminal Violence']= 'Criminal Violence'
mig\$cause_of_death[mig\$cause_of_death=='Crushed']= 'Crushed'
mig\$cause_of_death[mig\$cause_of_death=='Crushed / drowning']= 'Crushed'
mig\$cause_of_death[mig\$cause_of_death=='Crushed by bus on ferry']= 'Crushed'
mig\$cause_of_death[mig\$cause_of_death=='Crushed by pallets']= 'Crushed'
mig\$cause_of_death[mig\$cause_of_death=='Crushed to death']= 'Crushed'
mig\$cause_of_death[mig\$cause_of_death=='Cut in half by train']= 'Accident'
mig\$cause_of_death[mig\$cause_of_death=='Dehydration']= 'Dehydration'
mig\$cause_of_death[mig\$cause_of_death=='Dehydration
Harsh_weather_lack_of_adequate_shelter']= 'Dehydration + Others'
mig\$cause_of_death[mig\$cause_of_death=='Dehydration Harsh_weather_lack_of_adequate_shelter Suffocation Excessive_physical_abuse Sexual_abuse']= 'Dehydration + Others'
mig\$cause_of_death[mig\$cause_of_death=='Dehydration Suffocation Vehicle_Accident']= 'Dehydration + Others'
mig\$cause_of_death[mig\$cause_of_death=='Dehydration Vehicle_Accident Excessive_physical_abuse']= 'Dehydration + Others'
mig\$cause_of_death[mig\$cause_of_death=='Dehydration and exposure to the elements']= 'Dehydration + Others'
mig\$cause_of_death[mig\$cause_of_death=='Dehydration, Asphyxiation']= 'Dehydration + Others'
mig\$cause_of_death[mig\$cause_of_death=='Dehydration, Presumed drowning']= 'Dehydration + Others'
mig\$cause_of_death[mig\$cause_of_death=='Dehydration, Starvation']= 'Dehydration + Others'
mig\$cause_of_death[mig\$cause_of_death=='Died of unknown cause in hospital shortly after rescue']= 'Unknown Situation'
mig\$cause_of_death[mig\$cause_of_death=='Digestive bleeding']= 'Digestive Bleeding'
mig\$cause_of_death[mig\$cause_of_death=='Drowning after being thrown overboard by other passengers']= 'Drowning'
mig\$cause_of_death[mig\$cause_of_death=='Drowning due to forced disembarkation']= 'Drowning'
mig\$cause_of_death[mig\$cause_of_death=='Drowning or suffocation in hull']= 'Drowning'
mig\$cause_of_death[mig\$cause_of_death=='Drowning, Asphyxiation']= 'Drowning'
mig\$cause_of_death[mig\$cause_of_death=='Drowning, Other']= 'Drowning'
mig\$cause_of_death[mig\$cause_of_death=='Drowning, Trampling']= 'Drowning + Others'
mig\$cause_of_death[mig\$cause_of_death=='Drowning. Boat collided with ferry']= 'Drowning'
mig\$cause_of_death[mig\$cause_of_death=='Electrocuted on train']= 'Electrocution'
mig\$cause_of_death[mig\$cause_of_death=='Electrocution on railway']= 'Electrocution'

mig\$cause_of_death[mig\$cause_of_death=='Excessive_physical_abuse']=
'Excessive_physical_abuse'
mig\$cause_of_death[mig\$cause_of_death=='Excessive_physical_abuse
Sexual_abuse']= 'Excessive_physical_abuse + Others'
mig\$cause_of_death[mig\$cause_of_death=='Excessive_physical_abuse
Shot_or_Stabbed']= 'Excessive_physical_abuse + Others'
mig\$cause_of_death[mig\$cause_of_death=='Exposure']= 'Exposure'
mig\$cause_of_death[mig\$cause_of_death=='Exposure, Hyperthermia']= 'Exposure + Others'
mig\$cause_of_death[mig\$cause_of_death=='Exposure, Hypothermia']= 'Exposure + Others'
mig\$cause_of_death[mig\$cause_of_death=='Exposure, hypothermia']= 'Exposure + Others'
mig\$cause_of_death[mig\$cause_of_death=='Exposure. Died upon entry to refugee camp.']=
'Exposure'
mig\$cause_of_death[mig\$cause_of_death=='Fall from cliff']= 'Fall'
mig\$cause_of_death[mig\$cause_of_death=='Fell from boat']= 'Fall'
mig\$cause_of_death[mig\$cause_of_death=='Fell from train']= 'Fall'
mig\$cause_of_death[mig\$cause_of_death=='Fell from truck']= 'Fall'
mig\$cause_of_death[mig\$cause_of_death=='Fell from wall']= 'Fall'
mig\$cause_of_death[mig\$cause_of_death=='Found hanged']= 'Hanged'
mig\$cause_of_death[mig\$cause_of_death=='Fuel Inhalation']= 'Respiratory Complication'
mig\$cause_of_death[mig\$cause_of_death=='Fuel burns']= 'Burned to death'
mig\$cause_of_death[mig\$cause_of_death=='Glycemic crisis (Lack of Insuline Treatment)']=
'Glycaemic crisis'
mig\$cause_of_death[mig\$cause_of_death=='Gylcemic crisis (Diabetic, medicine thrown
overboard)']= 'Glycaemic crisis'
mig\$cause_of_death[mig\$cause_of_death=='Harsh conditions']= 'Harsh Conditions'
mig\$cause_of_death[mig\$cause_of_death=='Harsh_weather_lack_of_adequate_shelter']= 'Harsh
Conditions + Others'
mig\$cause_of_death[mig\$cause_of_death=='Harsh_weather_lack_of_adequate_shelter
Excessive_physical_abuse']= 'Harsh Conditions + Others'
mig\$cause_of_death[mig\$cause_of_death=='Harsh_weather_lack_of_adequate_shelter
Excessive_physical_abuse Sexual_abuse]= 'Harsh Conditions + Others'
mig\$cause_of_death[mig\$cause_of_death=='Harsh_weather_lack_of_adequate_shelter Other']=
'Harsh Conditions + Others'
mig\$cause_of_death[mig\$cause_of_death=='Harsh_weather_lack_of_adequate_shelter
Suffocation']= 'Harsh Conditions + Others'
mig\$cause_of_death[mig\$cause_of_death=='Harsh_weather_lack_of_adequate_shelter Suffocation
Excessive_physical_abuse Sexual_abuse]= 'Harsh Conditions + Others'
mig\$cause_of_death[mig\$cause_of_death=='Harsh_weather_lack_of_adequate_shelter Suffocation
Vehicle_Accident]= 'Harsh Conditions + Others'
mig\$cause_of_death[mig\$cause_of_death=='Harsh_weather_lack_of_adequate_shelter
Vehicle_Accident]= 'Harsh Conditions + Others'
mig\$cause_of_death[mig\$cause_of_death=='Harsh_weather_lack_of_adequate_shelter, Suffocation,
Excessive_physical_abuse, Sexual_abuse]= 'Harsh Conditions + Others'
mig\$cause_of_death[mig\$cause_of_death=='Head injury']= 'Blunt Force'
mig\$cause_of_death[mig\$cause_of_death=='Head injury from fall']= 'Blunt Force'
mig\$cause_of_death[mig\$cause_of_death=='Head trauma (hit by boat propeller)']= 'Blunt Force'
mig\$cause_of_death[mig\$cause_of_death=='Heart Attack']= 'Heart Attack'
mig\$cause_of_death[mig\$cause_of_death=='Heart attack']= 'Heart Attack'
mig\$cause_of_death[mig\$cause_of_death=='Hi by truck']= 'Vehicle Accident'
mig\$cause_of_death[mig\$cause_of_death=='Hit by Vehicle']= 'Vehicle Accident'
mig\$cause_of_death[mig\$cause_of_death=='Hit by car']= 'Vehicle Accident'
mig\$cause_of_death[mig\$cause_of_death=='Hit by train']= 'Vehicle Accident'

`mig$cause_of_death[mig$cause_of_death=='Hit by truck']= 'Vehicle Accident'`
`mig$cause_of_death[mig$cause_of_death=='Hit by vehicle']= 'Vehicle Accident'`
`mig$cause_of_death[mig$cause_of_death=='Homicide, likely by asphyxiation']= 'Asphyxiation'`
`mig$cause_of_death[mig$cause_of_death=='Hunger, fatigue']= 'Starvation'`
`mig$cause_of_death[mig$cause_of_death=='Hyperthermia']= 'Hyperthermia'`
`mig$cause_of_death[mig$cause_of_death=='Hyperthermia, Abandoned by smugglers in the desert']= 'Hyperthermia'`
`mig$cause_of_death[mig$cause_of_death=='Hyperthermia, Dehydration']= 'Hyperthermia + Others'`
`mig$cause_of_death[mig$cause_of_death=='Hyperthermia, starvation']= 'Hyperthermia + Others'`
`mig$cause_of_death[mig$cause_of_death=='Hypothermia']= 'Hypothermia'`
`mig$cause_of_death[mig$cause_of_death=='Hypothermia, Exhaustion']= 'Hypothermia + Others'`
`mig$cause_of_death[mig$cause_of_death=='Hypothermia, Malnutrition']= 'Hypothermia + Others'`
`mig$cause_of_death[mig$cause_of_death=='Inhalation of toxic fumes from boat engine']= 'Respiratory Complication'`
`mig$cause_of_death[mig$cause_of_death=='Injured from a fight']= 'Blunt Force'`
`mig$cause_of_death[mig$cause_of_death=='Injuries caused by boat motor']= 'Blunt Force'`
`mig$cause_of_death[mig$cause_of_death=='Killed']= 'Murdered'`
`mig$cause_of_death[mig$cause_of_death=='Landmine']= 'Landmine'`
`mig$cause_of_death[mig$cause_of_death=='Likely drowning']= 'Drowning'`
`mig$cause_of_death[mig$cause_of_death=='Likely suffocation (found dead in a truck)']= 'Suffocation'`
`mig$cause_of_death[mig$cause_of_death=='Lung infection']= 'Respiratory Complication'`
`mig$cause_of_death[mig$cause_of_death=='Meningitis']= 'Meningitis'`
`mig$cause_of_death[mig$cause_of_death=='Mixed']= 'Unknown Situation'`
`mig$cause_of_death[mig$cause_of_death=='Mixed - mostly drownings or shootings']= 'Drowning + Others'`
`mig$cause_of_death[mig$cause_of_death=='Mixed. Migrants were stranded on boats']= 'Unknown Situation'`
`mig$cause_of_death[mig$cause_of_death=='Mostly starvation, dehydration, and beatings by crew members']= 'Starvation + Others'`
`mig$cause_of_death[mig$cause_of_death=='Multiple blunt force injuries']= 'Blunt Force'`
`mig$cause_of_death[mig$cause_of_death=='Murdered']= 'Murdered'`
`mig$cause_of_death[mig$cause_of_death=='Murdered (bandits)']= 'Murdered'`
`mig$cause_of_death[mig$cause_of_death=='Murdered (head wound)']= 'Murdered'`
`mig$cause_of_death[mig$cause_of_death=='Murdered (militia)']= 'Murdered'`
`mig$cause_of_death[mig$cause_of_death=='Murdered by gang members']= 'Murdered'`
`mig$cause_of_death[mig$cause_of_death=='On board violence']= 'Criminal Violence'`
`mig$cause_of_death[mig$cause_of_death=='Other']= 'Unknown Situation'`
`mig$cause_of_death[mig$cause_of_death=='Other Shot_or_Stabbed']= 'Shot'`
`mig$cause_of_death[mig$cause_of_death=='Pending']= 'Unknown Situation'`
`mig$cause_of_death[mig$cause_of_death=='Pima County (see spreadsheet for exact location)']= 'Unknown Situation'`
`mig$cause_of_death[mig$cause_of_death=='Plane Stowaway']= 'Plane Stowaway'`
`mig$cause_of_death[mig$cause_of_death=='Plane stowaway']= 'Plane Stowaway'`
`mig$cause_of_death[mig$cause_of_death=='Poison']= 'Poison'`
`mig$cause_of_death[mig$cause_of_death=='Presumed Dehydration']= 'Dehydration'`
`mig$cause_of_death[mig$cause_of_death=='Presumed Drowning']= 'Drowning'`
`mig$cause_of_death[mig$cause_of_death=='Presumed asphyxiation']= 'Asphyxiation'`
`mig$cause_of_death[mig$cause_of_death=='Presumed dehydration']= 'Dehydration'`
`mig$cause_of_death[mig$cause_of_death=='Presumed drowning']= 'Drowning'`
`mig$cause_of_death[mig$cause_of_death=='Presumed exposure']= 'Exposure'`

`mig$cause_of_death[mig$cause_of_death=='Presumed hyperthermia']= 'Hyperthermia'`
`mig$cause_of_death[mig$cause_of_death=='Presumed hypothermia']= 'Hypothermia'`
`mig$cause_of_death[mig$cause_of_death=='Presumed shot']= 'Shot'`
`mig$cause_of_death[mig$cause_of_death=='Presumed violence']= 'Criminal Violence'`
`mig$cause_of_death[mig$cause_of_death=='Probable drowning']= 'Drowning'`
`mig$cause_of_death[mig$cause_of_death=='Pulmonary complications']= 'Respiratory Complication'`
`mig$cause_of_death[mig$cause_of_death=='Pulmonary edema']= 'Respiratory Complication'`
`mig$cause_of_death[mig$cause_of_death=='Pulmonary edema, Kidney failure']= 'Respiratory Complication'`
`mig$cause_of_death[mig$cause_of_death=='Raped and Murdured']= 'Raped & Murdered'`
`mig$cause_of_death[mig$cause_of_death=='Raped and murdered']= 'Raped & Murdered'`
`mig$cause_of_death[mig$cause_of_death=='Renal insufficiency']= 'Renal Insufficiency'`
`mig$cause_of_death[mig$cause_of_death=='Renal insufficiency and pulmonary edema']= 'Renal Insufficiency'`
`mig$cause_of_death[mig$cause_of_death=='Respiratory illness']= 'Respiratory Complication'`
`mig$cause_of_death[mig$cause_of_death=='Respiratory problem']= 'Respiratory Complication'`
`mig$cause_of_death[mig$cause_of_death=='Road accident']= 'Accident'`
`mig$cause_of_death[mig$cause_of_death=='Severe exhaustion and dehydration']= 'Exhaustion'`
`mig$cause_of_death[mig$cause_of_death=='Sexual_abuse']= 'Sexual Abuse'`
`mig$cause_of_death[mig$cause_of_death=='Shot']= 'Shot'`
`mig$cause_of_death[mig$cause_of_death=='Shot by Apache helicopter']= 'Shot'`
`mig$cause_of_death[mig$cause_of_death=='Shot_or_Stabbed']= 'Shot'`
`mig$cause_of_death[mig$cause_of_death=='Sickness']= 'Sickness'`
`mig$cause_of_death[mig$cause_of_death=='Sickness, Harsh conditions']= 'Sickness + Others'`
`mig$cause_of_death[mig$cause_of_death=='Sickness_and_lack_of_access_to_medicines']= 'Sickness & No Medicines'`
`mig$cause_of_death[mig$cause_of_death=='Sickness_and_lack_of_access_to_medicines, Dehydration']= 'Sickness & No Medicines + Others'`
`mig$cause_of_death[mig$cause_of_death=='Sickness_and_lack_of_access_to_medicines Dehydration Excessive_physical_abuse']= 'Sickness & No Medicines + Others'`
`mig$cause_of_death[mig$cause_of_death=='Sickness_and_lack_of_access_to_medicines Dehydration Excessive_physical_abuse Sexual_abuse']= 'Sickness & No Medicines + Others'`
`mig$cause_of_death[mig$cause_of_death=='Sickness_and_lack_of_access_to_medicines Dehydration Harsh_weather_lack_of_adequate_shelter']= 'Sickness & No Medicines + Others'`
`mig$cause_of_death[mig$cause_of_death=='Sickness_and_lack_of_access_to_medicines Dehydration Harsh_weather_lack_of_adequate_shelter Excessive_physical_abuse']= 'Sickness & No Medicines + Others'`
`mig$cause_of_death[mig$cause_of_death=='Sickness_and_lack_of_access_to_medicines Dehydration Harsh_weather_lack_of_adequate_shelter Excessive_physical_abuse Sexual_abuse']= 'Sickness & No Medicines + Others'`
`mig$cause_of_death[mig$cause_of_death=='Sickness_and_lack_of_access_to_medicines Dehydration Shot_or_Stabbed']= 'Sickness & No Medicines + Others'`
`mig$cause_of_death[mig$cause_of_death=='Sickness_and_lack_of_access_to_medicines Dehydration Suffocation']= 'Sickness & No Medicines + Others'`
`mig$cause_of_death[mig$cause_of_death=='Sickness_and_lack_of_access_to_medicines Dehydration Vehicle_Accident']= 'Sickness & No Medicines + Others'`
`mig$cause_of_death[mig$cause_of_death=='Sickness_and_lack_of_access_to_medicines Excessive_physical_abuse']= 'Sickness & No Medicines + Others'`

mig\$cause_of_death[mig\$cause_of_death=='Sickness_and_lack_of_access_to_medicines, Starvation, Harsh_weather_lack_of_adequate_shelter, Excessive_physical_abuse, Sexual_abuse']= 'Sickness & No Medicines + Others'
mig\$cause_of_death[mig\$cause_of_death=='Sickness_and_lack_of_access_to_medicines Starvation Harsh_weather_lack_of_adequate_shelter Vehicle_Accident Excessive_physical_abuse']= 'Sickness & No Medicines + Others'
mig\$cause_of_death[mig\$cause_of_death=='Stabbed']= 'Stabbed'
mig\$cause_of_death[mig\$cause_of_death=='Starvation']= 'Starvation'
mig\$cause_of_death[mig\$cause_of_death=='Starvation, Dehydration']= 'Starvation + Others'
mig\$cause_of_death[mig\$cause_of_death=='Starvation Dehydration Excessive_physical_abuse']= 'Starvation + Others'
mig\$cause_of_death[mig\$cause_of_death=='Starvation Dehydration Sexual_abuse Shot_or_Stabbed']= 'Starvation + Others'
mig\$cause_of_death[mig\$cause_of_death=='Starvation Dehydration Sexual_abuse Shot_or_Stabbed']= 'Starvation + Others'
mig\$cause_of_death[mig\$cause_of_death=='Starvation Dehydration Harsh_weather_lack_of_adequate_shelter']= 'Starvation + Others'
mig\$cause_of_death[mig\$cause_of_death=='Starvation Dehydration Harsh_weather_lack_of_adequate_shelter]= 'Starvation + Others'
mig\$cause_of_death[mig\$cause_of_death=='Starvation Dehydration Harsh_weather_lack_of_adequate_shelter Excessive_physical_abuse']= 'Starvation + Others'
mig\$cause_of_death[mig\$cause_of_death=='Starvation Dehydration Harsh_weather_lack_of_adequate_shelter Suffocation']= 'Starvation + Others'
mig\$cause_of_death[mig\$cause_of_death=='Starvation Dehydration Harsh_weather_lack_of_adequate_shelter Suffocation Excessive_physical_abuse']= 'Starvation + Others'
mig\$cause_of_death[mig\$cause_of_death=='Starvation Dehydration Other']= 'Starvation + Others'
mig\$cause_of_death[mig\$cause_of_death=='Starvation Dehydration Sexual_abuse Shot_or_Stabbed']= 'Starvation + Others'
mig\$cause_of_death[mig\$cause_of_death=='Starvation Dehydration Sexual_abuse Vehicle_Accident']= 'Starvation + Others'
mig\$cause_of_death[mig\$cause_of_death=='Starvation Excessive_physical_abuse']= 'Starvation + Others'
mig\$cause_of_death[mig\$cause_of_death=='Starvation Excessive_physical_abuse Sexual_abuse']= 'Starvation + Others'
mig\$cause_of_death[mig\$cause_of_death=='Starvation Harsh_weather_lack_of_adequate_shelter']= 'Starvation + Others'
mig\$cause_of_death[mig\$cause_of_death=='Starvation Harsh_weather_lack_of_adequate_shelter Suffocation']= 'Starvation + Others'
mig\$cause_of_death[mig\$cause_of_death=='Starvation Harsh_weather_lack_of_adequate_shelter Suffocation Excessive_physical_abuse']= 'Starvation + Others'
mig\$cause_of_death[mig\$cause_of_death=='Starvation Suffocation Excessive_physical_abuse']= 'Starvation + Others'
mig\$cause_of_death[mig\$cause_of_death=='Starvation Suffocation Excessive_physical_abuse Sexual_abuse']= 'Starvation + Others'
mig\$cause_of_death[mig\$cause_of_death=='Starvation Suffocation Shot_or_Stabbed']= 'Starvation + Others'
mig\$cause_of_death[mig\$cause_of_death=='Starvation Dehydration']= 'Starvation + Others'
mig\$cause_of_death[mig\$cause_of_death=='Starvation, Dehydration, Exhaustion']= 'Starvation + Others'
mig\$cause_of_death[mig\$cause_of_death=='Starvation, Dehydration, Suffocation']= 'Starvation + Others'

mig\$cause_of_death[mig\$cause_of_death=='Starvation, Suffocation']= 'Starvation + Others'
mig\$cause_of_death[mig\$cause_of_death=='Starvation, sickness']= 'Starvation + Others'
mig\$cause_of_death[mig\$cause_of_death=='Suffocation Excessive_physical_abuse']= 'Suffocation + Others'
mig\$cause_of_death[mig\$cause_of_death=='Suffocation Vehicle_Accident']= 'Suffocation + Others'
mig\$cause_of_death[mig\$cause_of_death=='Suffocation, Other']= 'Suffocation + Others'
mig\$cause_of_death[mig\$cause_of_death=='Suffocation, Trampled']= 'Suffocation + Others'
mig\$cause_of_death[mig\$cause_of_death=='Suicide']= 'Suicide'
mig\$cause_of_death[mig\$cause_of_death=="Tekeze River, near Himora, Ethiopia']= 'Unknown Situation'
mig\$cause_of_death[mig\$cause_of_death=="Tortured to death"]= 'Criminal Violence'
mig\$cause_of_death[mig\$cause_of_death=="Toxic fumes/asphyxiation"]= 'Asphyxiation'
mig\$cause_of_death[mig\$cause_of_death=="Train accident"]= 'Accident'
mig\$cause_of_death[mig\$cause_of_death=="Truck crash (was clinging to bottom of truck)]= 'Blunt Force'
mig\$cause_of_death[mig\$cause_of_death=="Unclear, possibly related to earlier beating by truck driver']= 'Blunt Force'
mig\$cause_of_death[mig\$cause_of_death=="Undernourished, dehydration']= 'Starvation + Others'
mig\$cause_of_death[mig\$cause_of_death=="Unknow (skeletal remains)]= 'Unknown Situation'
mig\$cause_of_death[mig\$cause_of_death=="Unknown"]= 'Unknown Situation'
mig\$cause_of_death[mig\$cause_of_death=="Unknown (Mummified remains)]= 'Unknown Situation'
mig\$cause_of_death[mig\$cause_of_death=="Unknown (Skeletal remains)]= 'Unknown Situation'
mig\$cause_of_death[mig\$cause_of_death=="Unknown (bodies found decomposed)]= 'Unknown Situation'
mig\$cause_of_death[mig\$cause_of_death=="Unknown (bodies recovered from boat)]= 'Unknown Situation'
mig\$cause_of_death[mig\$cause_of_death=="Unknown (bodies recovered on boat)]= 'Unknown Situation'
mig\$cause_of_death[mig\$cause_of_death=="Unknown (body found in advanced state of decomposition)]= 'Unknown Situation'
mig\$cause_of_death[mig\$cause_of_death=="Unknown (body found in desert)]= 'Unknown Situation'
mig\$cause_of_death[mig\$cause_of_death=="Unknown (body recovered from boat)]= 'Unknown Situation'
mig\$cause_of_death[mig\$cause_of_death=="Unknown (decomposed remains)]= 'Unknown Situation'
mig\$cause_of_death[mig\$cause_of_death=="Unknown (found dead near train tracks)]= 'Unknown Situation'
mig\$cause_of_death[mig\$cause_of_death=="Unknown (found dead on dinghy)]= 'Unknown Situation'
mig\$cause_of_death[mig\$cause_of_death=="Unknown (found dead on top of train)]= 'Unknown Situation'
mig\$cause_of_death[mig\$cause_of_death=="Unknown (found on motorway)]= 'Unknown Situation'
mig\$cause_of_death[mig\$cause_of_death=="Unknown (mummified remains)]= 'Unknown Situation'
mig\$cause_of_death[mig\$cause_of_death=="Unknown (postmortem decomposition)]= 'Unknown Situation'
mig\$cause_of_death[mig\$cause_of_death=="Unknown (remains)]= 'Unknown Situation'

```

mig$cause_of_death[mig$cause_of_death=="Unknown (skeletal and mummified remains)"]="Unknown Situation"
mig$cause_of_death[mig$cause_of_death=="Unknown (skeletal remains and mummified)"]="Unknown Situation"
mig$cause_of_death[mig$cause_of_death=="Unknown (skeletal remains)"]="Unknown Situation"
mig$cause_of_death[mig$cause_of_death=="Unknown, Skeletal Remains"]="Unknown Situation"
mig$cause_of_death[mig$cause_of_death=="Unknown, plane stowaway"]="Plane Stowaway"
mig$cause_of_death[mig$cause_of_death=="Unknown, torture involved"]="Torture"
mig$cause_of_death[mig$cause_of_death=="Unspecified deaths on 'La Bestia' train"]="Unknown Situation"
mig$cause_of_death[mig$cause_of_death=="Unspecified location between North Africa and Italy. Body brought to Calabria."]="Unknown Situation"
mig$cause_of_death[mig$cause_of_death=="Van accident"]="Vehicle Accident"
mig$cause_of_death[mig$cause_of_death=="Various"]="Unknown Situation"
mig$cause_of_death[mig$cause_of_death=="Vehicle accident"]="Vehicle Accident"
mig$cause_of_death[mig$cause_of_death=="Vehicle incident"]="Vehicle Accident"
mig$cause_of_death[mig$cause_of_death=="Vehicle_Accident"]="Vehicle Accident"
mig$cause_of_death[mig$cause_of_death=="Vehicle_Accident Other"]="Vehicle Accident + Others"
mig$cause_of_death[mig$cause_of_death=="Vehicle_Accident Shot_or_Stabbed"]="Vehicle Accident + Others"
mig$cause_of_death[mig$cause_of_death=="Violence"]="Criminal Violence"
mig$cause_of_death[mig$cause_of_death=="Violence during riot"]="Criminal Violence"
mig$cause_of_death[mig$cause_of_death=="Violent robbery"]="Criminal Violence"
mig$cause_of_death[mig$cause_of_death=="heart attack"]="Heart Attack"
mig$cause_of_death[mig$cause_of_death=="drowning"]="Drowning"
mig$cause_of_death[mig$cause_of_death=="gang violence (body dismembered)"]="Criminal Violence"
mig$cause_of_death[mig$cause_of_death=="mixed"]="Unknown Situation"
mig$cause_of_death[mig$cause_of_death=="unknown (corpses recovered from boats)"]="Unknown Situation"
mig$cause_of_death[mig$cause_of_death=="shot"]="Shot"
mig$cause_of_death[is.na(mig$cause_of_death)]="Unknown"
mig$cause_of_death[mig$cause_of_death=="]"]="Unknown"

```

```

#Region of origin
length(unique(mig$region_origin))
View(unique(mig$region_origin))
mig$region_origin[mig$region_origin==""]的文化="Unknown"
mig$region_origin[mig$region_origin=="Horn of Africa (P)"]="Horn of Africa"
mig$region_origin[mig$region_origin=="MENA"]="Middle East/North Africa"

```

```

#Nationality
length(unique(mig$nationality))
View(unique(mig$nationality))
mig$nationality[mig$nationality==""]的文化="Unknown Nationality"
mig$nationality[mig$nationality=="1 Honduran, 3 Mexican"]="Honduras, Mexico"
mig$nationality[mig$nationality=="1 Nigerian, others unknown. Survivors all from Sub-Saharan Africa"]="Nigeria"
mig$nationality[mig$nationality=="1 Venezuelan, 1 unknown"]="Venezuela"

```

mig\$nationality[mig\$nationality=='13 Cuba, 1 Dominican Republic, 1 Colombia']= 'Cuba, Dominican Republic, Colombia'

mig\$nationality[mig\$nationality=='15 dead from Palestine. Missing are from Palestine, Syria, and Egypt']= 'Palestine'

mig\$nationality[mig\$nationality=='2 from Niger']= 'Niger'

mig\$nationality[mig\$nationality=='2 Senegal, 2 Guinea, 1 Ghana']= 'Senegal, Guinea, Ghana'

mig\$nationality[mig\$nationality=='20 Unknown, 1 Bangladesh, 1 Senegal']= 'Bangladesh, Senegal'

mig\$nationality[mig\$nationality=='Afghan']= 'Afghanistan'

mig\$nationality[mig\$nationality=='Afghanistan, Syrian Arab Republic']= 'Afghanistan, Syria'

mig\$nationality[mig\$nationality=='African']= 'Africa'

mig\$nationality[mig\$nationality=='Algerian']= 'Algeria'

mig\$nationality[mig\$nationality=='At least 1 from Mexico']= 'Mexico'

mig\$nationality[mig\$nationality=='at least 4 Syrian']= 'Syria'

mig\$nationality[mig\$nationality=='Bangladeshi, Rohingya']= 'Bangladesh, Myanmar'

mig\$nationality[mig\$nationality=='Brazil (12), Dominican Republic (5), Cuba (2)']= 'Brazil, Dominican Republic, Cuba'

mig\$nationality[mig\$nationality=='Brazilian']= 'Brazil'

mig\$nationality[mig\$nationality=='Cameroon (2) Gambia (1), Mauritius (1), Ivory Coast (2)']= 'Cameroon, Gambia, Mauritius, Ivory Coast'

mig\$nationality[mig\$nationality=='Cameroon, Senegal and Ivory Coast']= 'Cameroon, Senegal, Ivory Coast'

mig\$nationality[mig\$nationality=='Cameroonian']= 'Cameroon'

mig\$nationality[mig\$nationality=='China (ethnic Uighur)']= 'China'

mig\$nationality[mig\$nationality=='Comoran']= 'Comoros'

mig\$nationality[mig\$nationality=='Congo']= 'Democratic Republic of Congo'

mig\$nationality[mig\$nationality=='Costa rica']= 'Costa Rica'

mig\$nationality[mig\$nationality=='Cuba or Dominican Republic']= 'Cuba'

mig\$nationality[mig\$nationality=='Democratic Republic of the Congo']= 'Democratic Republic of Congo'

mig\$nationality[mig\$nationality=='Domican Republic']= 'Dominican Republic'

mig\$nationality[mig\$nationality=='Ecuadorian']= 'Ecuador'

mig\$nationality[mig\$nationality=='Ecuator']= 'Ecuador'

mig\$nationality[mig\$nationality=='Egypt (est.80), Ethiopia (est.150), Somalia (est.190), Sudan, Syria']= 'Egypt, Ethiopia, Somalia, Sudan, Syria'

mig\$nationality[mig\$nationality=='Egypt, Syrian Arab Republic']= 'Egypt, Syria'

mig\$nationality[mig\$nationality=='Egyptian, Eritrean, Sudanese, Syrian']= 'Egypt, Eritrea, Sudan, Syria'

mig\$nationality[mig\$nationality=='Eritrea (2), Syria (1)']= 'Eritrea, Syria'

mig\$nationality[mig\$nationality=='Eritrea or Somalia']= 'Eritrea'

mig\$nationality[mig\$nationality=='Eritrean']= 'Eritrea'

mig\$nationality[mig\$nationality=='Ethiopia and Somalia']= 'Ethiopia, Somalia'

mig\$nationality[mig\$nationality=='Ethiopian']= 'Ethiopia'

mig\$nationality[mig\$nationality=='Ethnic Rohingya']= 'Myanmar'

mig\$nationality[mig\$nationality=='Ethnic Yazidis (Iraq)']= 'Iraq'

mig\$nationality[mig\$nationality=='Gambia (2) Guinea Bissau (1)']= 'Gambia, Guinea Bissau'

mig\$nationality[mig\$nationality=='Gambia, Nigeria and Senegal']= 'Gambia, Nigeria, Senegal'

mig\$nationality[mig\$nationality=='Guatemala and El Salvador']= 'Guatemala, El Salvador'

mig\$nationality[mig\$nationality=='Guatemalan']= 'Guatemala'

mig\$nationality[mig\$nationality=='Guinea Conakry']= 'Guinea'

mig\$nationality[mig\$nationality=='Guinean']= 'Guinea'

mig\$nationality[mig\$nationality=='Haitian']= 'Haiti'

mig\$nationality[mig\$nationality=='Honduran']= 'Honduras'

mig\$nationality[mig\$nationality=='Honduras and Guatemala']='Honduras, Guatemala'
mig\$nationality[mig\$nationality=='Honduras or El Salvador']='Honduras'
mig\$nationality[mig\$nationality=='Iran and Iraq']='Iran, Iraq'
mig\$nationality[mig\$nationality=='Iraq, Algeria and Syria (initial reports)']='Iraq, Algeria, Syria'
mig\$nationality[mig\$nationality=='Iraqi']='Iraq'
mig\$nationality[mig\$nationality=='Iraqui']='Iraq'
mig\$nationality[mig\$nationality=='Ivory Coast (1) Unknown (16)']='Ivory Coast'
mig\$nationality[mig\$nationality=='Ivory Coast (15), Mali (7), Senegal (5), Guinea (1), Mauritania (1)']='Ivory Coast, Mali, Senegal, Guinea, Mauritania'
mig\$nationality[mig\$nationality=='Ivory Coast and Guinea Conakry']='Ivory Coast, Guinea'
mig\$nationality[mig\$nationality=='Lebanese']='Lebanon'
mig\$nationality[mig\$nationality=='Likely Comorian']='Comoros'
mig\$nationality[mig\$nationality=='likely Comoros']='Comoros'
mig\$nationality[mig\$nationality=='Likely Eritrea']='Eritrea'
mig\$nationality[mig\$nationality=='Likely Rohingya']='Myanmar'
mig\$nationality[mig\$nationality=='Mahgreb']='Maghreb'
mig\$nationality[mig\$nationality=='Malagasy']='Madagascar'
mig\$nationality[mig\$nationality=='Malawian']='Malawi'
mig\$nationality[mig\$nationality=='Mexican']='Mexico'
mig\$nationality[mig\$nationality=='Mixed']='Mixed Nationality'
mig\$nationality[mig\$nationality=='Moroccan']='Morocco'
mig\$nationality[mig\$nationality=='Mostly Ethiopian. Others were Somalian and 2 Yemeni crew']='Ethiopia, Somalia, Yemen'
mig\$nationality[mig\$nationality=='mostly from Sudan, Bangladesh']='Sudan, Bangladesh'
mig\$nationality[mig\$nationality=='Mostly from Syria, possibly some from Iraq']='Syria, Iraq'
mig\$nationality[mig\$nationality=='Mozambican']='Mozambique'
mig\$nationality[mig\$nationality=='Myanmar (Rohingya)']='Myanmar'
mig\$nationality[mig\$nationality=='Niger (3), Mali(3), Senegal(3), Guinea(3), Ivory Coast(2), CAR(1), Liberia(1)']='Niger, Mali, Senegal, Guinea, Ivory Coast, Central African Republic, Liberia'
mig\$nationality[mig\$nationality=='Nigeria (2) Cameroon (2)']='Nigeria, Cameroon'
mig\$nationality[mig\$nationality=='Nigeria, others']='Nigeria'
mig\$nationality[mig\$nationality=='Nigerian']='Nigeria'
mig\$nationality[mig\$nationality=='Pakistani']='Pakistan'
mig\$nationality[mig\$nationality=='Peru (1), Unknown (1)']='Peru'
mig\$nationality[mig\$nationality=='Reported as \"mostly\" Senegal']='Senegal'
mig\$nationality[mig\$nationality=='Reported as unspecified national of Africa']='Africa'
mig\$nationality[mig\$nationality=='Reported as unspecified national of Central America']='Central America'
mig\$nationality[mig\$nationality=='Reported as unspecified national of Sub-Saharan Africa']='Africa'
mig\$nationality[mig\$nationality=='Reported as unspecified nationals of Africa']='Africa'
mig\$nationality[mig\$nationality=='Reported as unspecified nationals of Central America']='Central America'
mig\$nationality[mig\$nationality=='Reported as unspecified nationals of Honduras']='Honduras'
mig\$nationality[mig\$nationality=='Reported as unspecified nationals of Honduras and Guatemala']='Honduras, Guatemala'
mig\$nationality[mig\$nationality=='Reported as unspecified nationals of Horn of Africa']='Africa'
mig\$nationality[mig\$nationality=='Reported as unspecified nationals of Rohingya']='Myanmar'
mig\$nationality[mig\$nationality=='Reported as unspecified nationals of Somalia, Sudan and Nigeria']='Somalia, Sudan, Nigeria'
mig\$nationality[mig\$nationality=='Reported as unspecified nationals of Sub-Saharan Africa']='Africa'

mig\$nationality[mig\$nationality=='Reported as unspecified nationals of Syria, Afghanistan, Iraq, Iran']= 'Syria, Afghanistan, Iraq, Iran'
mig\$nationality[mig\$nationality=='Reported as unspecified nationals of the Horn of Africa']= 'Africa'
mig\$nationality[mig\$nationality=='Reported as unspecified nationals of West Africa']= 'Africa'
mig\$nationality[mig\$nationality=='Rohingya']= 'Myanmar'
mig\$nationality[mig\$nationality=='Salvadoran']= 'El Salvadoran'
mig\$nationality[mig\$nationality=='Senegal, Ivory Coast, Gambia, Guinea, Niger, Mali and Mauritania']= 'Senegal, Ivory Coast, Gambia, Guinea, Niger, Mali, Mauritania'
mig\$nationality[mig\$nationality=='Somalia (1) Eritrea (1)']= 'Somalia, Eritrea'
mig\$nationality[mig\$nationality=='Sub-Saharan Africa']= 'Africa'
mig\$nationality[mig\$nationality=='Sudan, Ethiopia, Eritrea']= "
mig\$nationality[mig\$nationality=='Sudanese']= 'Sudan'
mig\$nationality[mig\$nationality=='Survivors from Burkina Faso, Malia, Guinea, and the Ivory Coast']= 'Burkina Faso, Malia, Guinea, Ivory Coast'
mig\$nationality[mig\$nationality=='Survivors from Nigeria, Ghana, Niger']= 'Nigeria, Ghana, Niger'
mig\$nationality[mig\$nationality=='Survivors from Sub-Saharan Africa']= 'Africa'
mig\$nationality[mig\$nationality=='Survivors from Syria, Iraq, and Somalia']= 'Syria, Iraq, Somalia'
mig\$nationality[mig\$nationality=='Survivors were from Gambia, Ghana and Mali']= 'Gambia, Ghana, Mali'
mig\$nationality[mig\$nationality=='Syria (mostly)']= 'Syria'
mig\$nationality[mig\$nationality=='Syrian']= 'Syria'
mig\$nationality[mig\$nationality=='Syrian and Iraqi']= 'Syria, Iraq'
mig\$nationality[mig\$nationality=='Syrian Arab Republic']= 'Syria'
mig\$nationality[mig\$nationality=='Syrian Arab Republic, Egypt, Sudan']= 'Syria, Egypt, Sudan'
mig\$nationality[mig\$nationality=='Syrian Arab Republic, Eritrea, Somalia, Cameroon']= 'Syria, Eritrea, Somalia, Cameroon'
mig\$nationality[mig\$nationality=='Syrian Kurds']= 'Syria'
mig\$nationality[mig\$nationality=='Syrian, Egyptian, other African']= 'Syria, Egypt, Africa'
mig\$nationality[mig\$nationality=='Syrian, Iraqi, Afghan']= 'Syria, Iraq, Afghanistan'
mig\$nationality[mig\$nationality=='Syrian. An infant was among the dead']= 'Syria'
mig\$nationality[mig\$nationality=='Unknown']= 'Unknown Nationality'
mig\$nationality[mig\$nationality=='Unknown (5 reported as nationals from Sub-Saharan Africa and 1 from Morocco)']= 'Africa, Morocco'
mig\$nationality[mig\$nationality=='Unknown (Kurdish)']= 'Iran'
mig\$nationality[mig\$nationality=='Unknown (skeletal remains)']= 'Unknown Nationality'
mig\$nationality[mig\$nationality=='Unknown. Survivors all from Sub-Saharan African nations.']= 'Africa'
mig\$nationality[mig\$nationality=='Unknown. Survivors from Bangladesh, Burma, or are ethnic Rohingya']= 'Bangladesh, Myanmar'
mig\$nationality[mig\$nationality=='Unkown']= 'Unknown Nationality'
mig\$nationality[mig\$nationality=='Unnnknown']= 'Unknown Nationality'
mig\$nationality[mig\$nationality=='Unspecified nationals of Western Africa']= 'Africa'
mig\$nationality[mig\$nationality=='Unspecified national of Sub-Saharan Africa']= 'Africa'
mig\$nationality[mig\$nationality=='Unspecified nationalities of Sub-Saharan Africa']= 'Africa'
mig\$nationality[mig\$nationality=='Unspecified nationalities of Subsaharan Africa']= 'Africa'
mig\$nationality[mig\$nationality=='Unspecified nationality of North Africa']= 'Africa'
mig\$nationality[mig\$nationality=='Unspecified nationality of Sub-Saharan Africa']= 'Africa'
mig\$nationality[mig\$nationality=='Unspecified nationals of Africa']= 'Africa'
mig\$nationality[mig\$nationality=='Unspecified nationals of Sub-Saharan Africa']= 'Africa'
mig\$nationality[mig\$nationality=='Unspecified nationals of West Africa']= 'Africa'
mig\$nationality[mig\$nationality=='Various']= 'Unknown Nationality'
mig\$nationality[mig\$nationality=='Venezulean']= 'Venezuela'

```

length(unique(mig$nationality))#134

#Incident Region
length(unique(mig$incident_region))
View(unique(mig$incident_region))

mig$incident_region[mig$incident_region==""]='Unknown incident region'
mig$incident_region[mig$incident_region=='Central America incl. Mexico']= 'Central America'

#Missing number of people
#Filling the number of missing with missing value as 0
mig$missing[is.na(mig$missing)]= 0

#Dead number of people
#Filling in the missing data with mode based on region
find_mode <- function(v) {
  uniqv <- unique(v)
  uniqv[which.max(tabulate(match(v, uniqv)))]
}

x = find_mode(mig$dead)#Found the mode for all regions(individually) using the above function to
be 1
mig$dead[is.na(mig$dead)]= x

#Relialiblty
mig$reliability[is.na(mig$reliability)]= "Unknown"
mig$reliability[mig$reliability==""]= "Unknown"
mig$reliability[mig$reliability=="Partially verified"]= "Partially Verified"

#Cleaning for the glyph map plot
mig$map_region = mig$incident_region
mig$map_region[mig$map_region==""]='Unknown'
mig$map_region[mig$map_region=='Central America incl. Mexico']= 'Central America'
mig$map_region[mig$map_region=='Middle East']= 'Middle East'
mig$map_region[mig$map_region=='Middle East ']='Middle East'
mig$map_region[mig$map_region=='Central America']= 'Central America & Mexico'

#Exporting the dataset -- for using in tableau
library(writexl)
write_xlsx(mig, "//Users/swathib/Desktop/Spring Term 2021-22/DSC 465 Data
Visualization/Final Project/Final Project extra files/migrants.xlsx")

#EXPLORATORY ANALYSIS
summary(mig)

#Boxplots - doesn't seem to work since there are too many zeroes
ggplot(mig, aes(missing))+ geom_boxplot()

```

```

ggplot(mig, aes(dead)) + geom_boxplot()

#Histogram - not great either
mig %>% filter(mig$missing!=0) %>%
ggplot(aes(missing)) + geom_histogram() # definitely skewed to the left

mig %>% filter(mig$dead!=0) %>%
ggplot(aes(dead)) + geom_histogram() #definitely skewed to the left

#Bar plots
ggplot(mig, aes(cause_of_death)) + geom_bar() #too many categories. Maybe specific regions would
be better
ggplot(mig, aes(nationality)) + geom_bar() #too many
ggplot(mig, aes(region_origin)) + geom_bar() #Horn of africa - that's where most migrants
originated from
ggplot(mig, aes(incident_region)) + geom_bar()#Most incidents happened at North America
followed by Mediterranean and US/Mexico border
ggplot(mig, aes(reliability)) + geom_bar() #Majority - partially verified

#Univariate scatterplots - Not sure what interesting thing to say about these
ggplot(mig, aes(x=factor(region_origin), y =dead)) +
  geom_point(alpha = 0.5)
ggplot(mig, aes(x=factor(region_origin), y =missing)) +
  geom_point(alpha = 0.5)
ggplot(mig, aes(x=factor(incident_region), y =dead)) +
  geom_point(alpha = 0.5)
ggplot(mig, aes(x=factor(incident_region), y =missing)) +
  geom_point(alpha = 0.5)

#Scatterplot
mig %>% filter(mig$missing!=0) %>%
  ggplot(aes(missing,dead)) +
  geom_point(alpha = 0.5, size = 2)

#Quantile-Quantile plot
migQQ = mig %>% drop_na() %$%
  data.frame(missingSort=sort(missing),
             deadSort=sort(dead))
ggplot(migQQ,aes(missingSort,deadSort)) + geom_point()

#EXPLANATORY PLOTS – drafts we tried out that didn't work
#(1)TIME VS MISSING ..... Not working out well.
class(mig$date)
mig$date= as.Date(mig$date,"%d/%m/%Y")

#MonthYear VS missing
mig$mmyy = format(mig$date, "%m%y")

migTrans <- mig %>%
  group_by(mmyy) %>%
  summarise_at(vars(missing), list(sumMiss = sum))
migTrans$mmyy = as.numeric(migTrans$mmyy)

```

```

ggplot(migTrans,aes(x = mmyy, y = sumMiss))+  

  geom_line()  
  

#Year VS Missing  

mig$yy = format(mig$date, "%Y")  

migTrans1 <- mig %>%  

  group_by(yy) %>%  

  summarise_at(vars(missing), list(sumMiss = sum))  

migTrans1$yy = as.numeric(migTrans1$yy)  
  

ggplot(migTrans1,aes(x = yy, y = sumMiss))+  

  geom_line()  

str(migTrans1)  
  

#(2)  

library(gganimate)  

library(gifski)  

library(plotly)  

mig[is.na(mig$yy),]  

mig1 = mig[-c(460,489,540,594,640,1133,1134,1135,1214),]  
  

p <- ggplot(mig1, aes(x = missing, y=dead, colour = region_origin)) +  

  geom_point(alpha = 0.7) +  

  scale_color_viridis_d() +  

  scale_size(range = c(2, 12)) +  

  labs(x = "Missing migrants", y = "Dead migrants")  

p  

ggplotly(p)  
  

#(3)  

mig1$yy = as.numeric(mig1$yy,na.rm = TRUE)  

p + transition_time(yy) +  

  labs(title = "Year: {frame_time}")

```

#Rest of plots were done in Tableau as that was easier to manipulate for our plots where we had to filter quite a bit to achieve readable plots

References:

<https://www.kaggle.com/datasets/jmataya/missingmigrants>

<https://www.kaggle.com/code/olgaminguett/eda-w-shap-missing-migrants>

<https://bookdown.org/rdpeng/exdata/exploratory-graphs.html>

<https://www.migrationdataportal.org/themes/migrant-deaths-and-disappearances>

<https://missingmigrants.iom.int/methodology>

<https://www.bbc.com/news/world-europe-34131911>

<https://www.amnesty.org/en/latest/news/2015/05/why-are-the-rohingya-fleeing-myanmar/>

<https://www.cnn.com/2016/05/29/europe/migrant-deaths/index.html>

<https://www.iom.int/news/over-3770-migrants-have-died-trying-cross-mediterranean-europe-2015>

<https://images.app.goo.gl/6mXpnYndd2KvHKXq9>