Missing Migrants

- Kazi Jewel Rana
- Utkal Sirikonda

Page: https://utkal97.github.io/missing-migrants-visualization/
Source Code: https://github.com/Utkal97/missing-migrants-visualization
Link to video:
https://www.youtube.com/watch?v=nz61ownSiQU&feature=youtu.be

Overview :

Migrants often face perilous conditions and challenges along the way, which can lead to tragic outcomes. The Missing Migrants Project has diligently collected data on such incidents worldwide since 2014. This project aims to shed light on this critical issue and provide valuable insights into the patterns of migration and the challenges migrants face during their journeys. With 13021 incidents and 17 relevant attributes, we have a robust dataset that can be leveraged to gain meaningful insights.

Related Work:

Our idea is mainly inspired by the Missing Migrants Project, where the people involved collected all the data regarding the incidents taking place from 2014 and posted it online. Even though the visualization provided by them gives an overview of the number of incidents that take place, we wanted to delve deeper and provide more insights on the trend of causes through the years in each region. This is not provided in the Missing Migrants Project.

Question we want to solve:

One of the primary motivations for this project is to visualize and analyze incidents of missing migrants and deaths globally. By aggregating these incidents, we can identify patterns in migration routes and understand the challenges migrants encounter during their journeys. This information can help inform policies and interventions to ensure the safety and well-being of migrants.

We started with a representation of the world map as a Flow chart by directly showing arrows from source region to incident regions, which would make a cluttered view. When distributing it into a Proportional Symbol map and a Sankey chart, we were still unable to display and analyze the trends we wanted to observe. Over work is an extension to the Missing Migrants Project by showing trends for each cause and region.

Data and Processing:

We have taken the dataset from the Missing Migrants Project (https://missingmigrants.iom.int/data), by the International Organization of Migration. The dataset has recorded incidents of missing people from 2014 - current year. The datafile can be retrieved from the below link as well:
https://www.kaggle.com/datasets/nelgiriyewithana/global-missing-migrants-dataset

As a part of processing of data, we had a lot of incidents (rows) in our data without main information fields like location. We have eliminated all the incidents without latitude and longitude values for accurately representing the useful data.

Also, "the *coordinates*", *"number of deaths"* fields were in string format which were changed to float and integer values respectively to be used in our script.

Since a region is very vast, for each region, we calculated the centroid of all the incidents that took place in that region to place a circle around it. This gives a better picture of where the incidents are actually happening.

Exploratory Data Analysis:

We had initially thought of using a choropleth map to show the count of incidents, but this wasn't actually answering our questions much, then we went to use a point map and flow map. As per our observation, there was a huge clutter in the Mediterranean Region due to the high amount of incidents. The graph was so heavy that our browser hanged. This was not an option to us since the links in the map have occluded some regions nearby, and did not make a good visualization.

Hence, we pieced down the map into two:
1. Proportional Symbol Map - to show the amount of incidents,
2. Sankey Chart - to display the flow among different regions

Design Evolution:

Initially we only had an expectation to offer three elements: A world map with incidents plotted, a bar chart to show the number of deaths over the years, and a trend chart to describe deaths due to different causes over the years.

World Map:

We wanted to visualize a world map as a flow map, but it created a lot of clutter in some regions like the Mediterranean and Central America. We obtained something like this initially, where the links have occluded some incidents. Also, this was not helpful in filtering out a complete region.
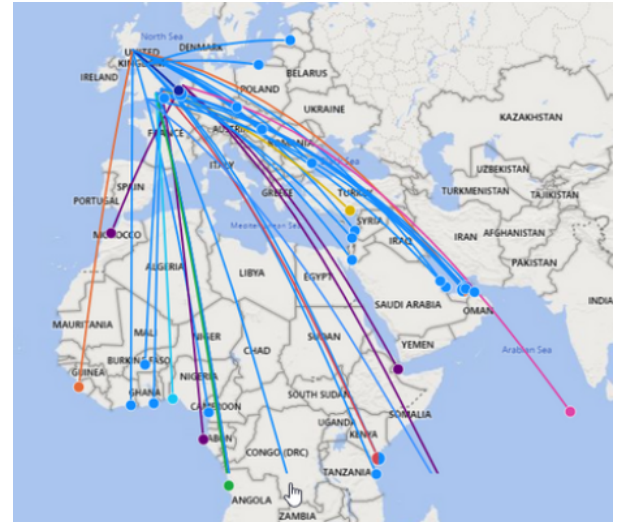
Then we went on to remove the links and just represent the incidents as points on the map.We have colored each point according to the origin region of people for each incident. This again created a clutter and was not useful for filtering out regions as a whole, since there are a huge number of small sized points.
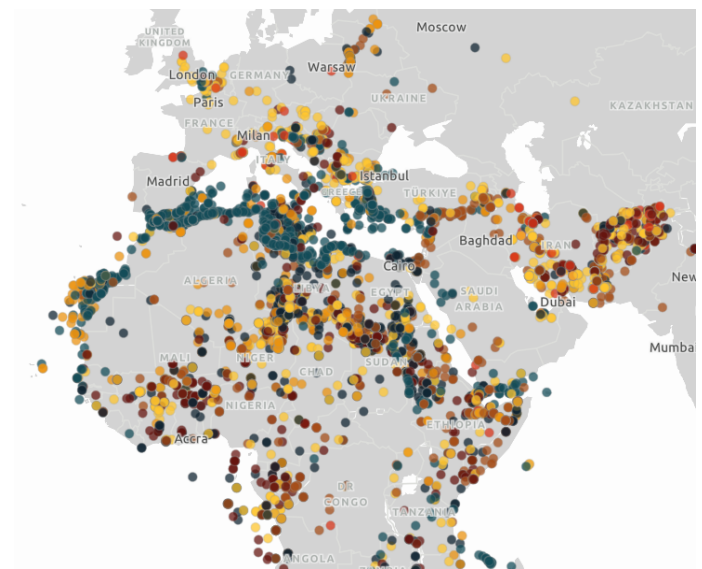
After a long thought, we finally divided the visualization into two graphs:
1. A Proportional Symbol Map - describes the amount of deaths that took place in the region.
2. A Sankey Chart - describes where the people originated from (alternative to flow map)

The map now can display all the incidents accumulated as one and can easily filter out the rest of visualizations. The Sankey chart gives a clear picture of the number of people using each migration route.

We also thought of displaying a Stacked Area Chart to show the trends of cause of deaths which was in our proposal, but we changed it to a Stacked Bar Chart since the data we have is discrete and the visualization create with Area chart was not looking nice when the values became small (unfortunately we do not have a screenshot to show it).

We filter Stacked Bar chart and Pie Chart with the world map. To filter one region, we just need to click on it. **To remove the region filter, we need to again click on the same circle or a different circle (please note, I missed this point in my video)**.

Design choices:
   1. Proportional Symbol Map -
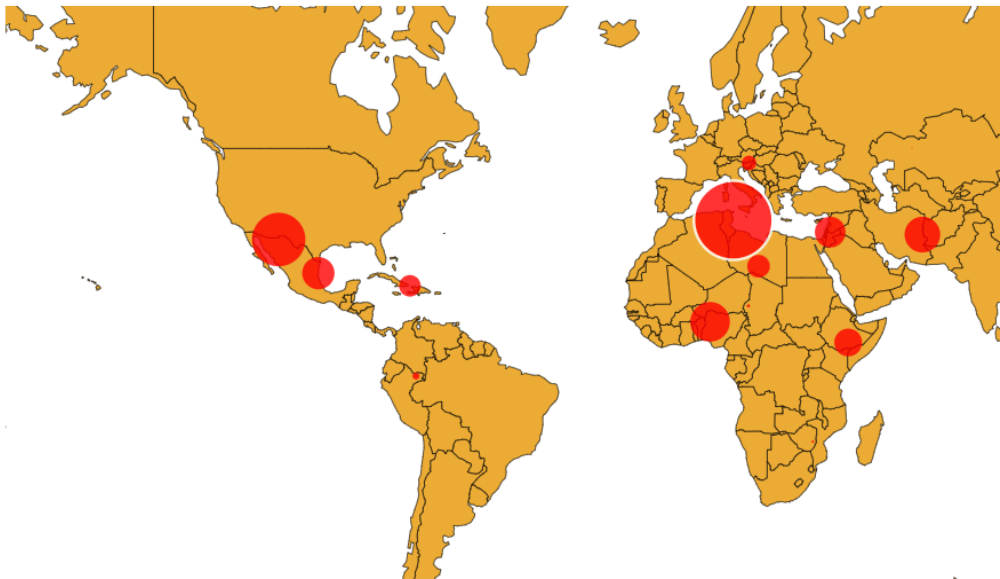
Mediterranean region with 27655 fatalities over the years

-

   ● Each symbol (circle) is scaled according to the number of deaths taken place. But we faced a challenge, where some regions

(~400) had very few incidents while the Mediterranean region had a high number (>25000). So our initial map had point sized circles for some regions. We applied a **non-linear scaling** such that the smaller numbers pump up better and bigger ones are lessened.
- Rather than color, we opted to choose area as our channel since count is Quantitative.
- We didn't want to rely on tooltips as suggested, so we displayed a description text below the map about the region and the count of deaths.
- When clicked on a mark (circle), the region is selected and the mark is highlighted and filters out the rest of the graphs.

2. Sankey Chart -

Sankey Chart for Migration Movements
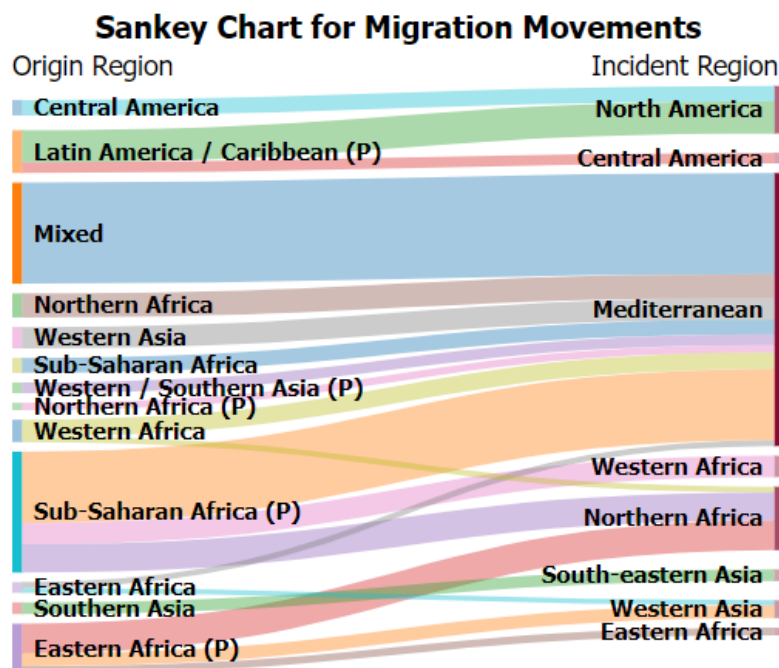
- Since each region is Categorical, we have chosen color as channels for links.
- Length and Color Saturation are channels for number of incidents taken place for each region (mark)