A Proposal on

Understanding Missing Migrants: A Visual Data Representation

Group #: 06
- *Kazi Jewel Rana, C98989458*
krana@g.clemson.edu ,
- *Utkal Sirikonda, C33980158*
usiriko@g.clemson.edu

Link to page: https://utkal97.github.io/missing-migrants-visualization/
Link to repo: https://github.com/Utkal97/missing-migrants-visualization

Background:

The phenomenon of missing migrants and recorded deaths during their journeys to reach settlements in other countries is a pressing global issue. 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.

Motivation:

Analyzing Migration Patterns and Challenges: 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.

Utilizing the Global Missing Migrants Dataset: The choice of the Global Missing Migrants Dataset as the basis for this project stems from the rich and comprehensive nature of the data it provides. With 13021 incidents and 17 relevant attributes, we have a robust dataset that can be leveraged to gain meaningful insights.

Humanitarian and Ethical Considerations: The project is motivated by a strong sense of empathy and respect for the individuals who have lost their lives during migration journeys and the families and communities affected by their absence. By working with this dataset, the project aims to raise awareness about the human cost of migration and contribute to honoring the lives lost.

Diverse Analysis Opportunities: The dataset offers a wide range of variables, allowing for diverse analysis possibilities. It provides opportunities to explore various aspects of migration,

including demographic factors, survival rates, causes of death, temporal trends, and geospatial patterns. These analyses can answer critical questions about migration and provide valuable insights for policymakers and humanitarian organizations.

Communication and Reporting: Effective data visualization is essential for communicating findings to governments and other stakeholders. Visualizations can help governments understand the migration landscape in less vigilant regions and address issues that require attention. This project recognizes the importance of clear and impactful communication in driving positive change.

Therefore, this project can both represent a good visualization of missing migrants effectively and raise awareness and promote empathy for the individuals and communities affected by these migration tragedies.

Objectives:

Our main objective is to clearly visualize the records of migrants that went missing over the years.

Specific objectives are:

1. to identify the dangerous migration routes
2. to analyze temporal trends in missing migrant incidents
3. to determine the common causes of fatalities
4. to visualize the geographic distribution of incidents
5. to explore demographic characteristics of migrants
6. to calculate survival and mortality rates
7. to find out the locations where high number of fatalities caused
8. to know how the causes are changing across the years.
9. to identify which countries, receive most migrants and which countries have most people who want to emigrate.

Data:

Dataset is obtained 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

Data Processing:

Since there are missing values, miscoded and misrepresented values, we need to process the data before analysis. The processing steps can be as below:

1. Data Cleanup:

Handling Missing Data: We can check for missing values in the dataset to impute missing values, remove rows with missing data, or use other statistical value assigning techniques based on the specific attributes.

Outlier Detection: We may need to identify and handle outliers in the data, especially in numeric fields like the number of dead and missing migrants. Outliers can significantly affect the analysis and visualizations.

Data Type Conversion: We will have to ensure that data types are consistent and appropriate for each attribute.

2. Data Derivation:

Calculating Derived Metrics: Depending on the research questions and analysis objectives, we may need to derive additional metrics from the existing data. For example, we might calculate mortality rates, create time series data, timestamp or aggregate data by specific regions or routes.

Categorization: We can categorize certain attributes to simplify analysis. For example, grouping incident types into broader categories or aggregating countries into regions for easier visualization and analysis.

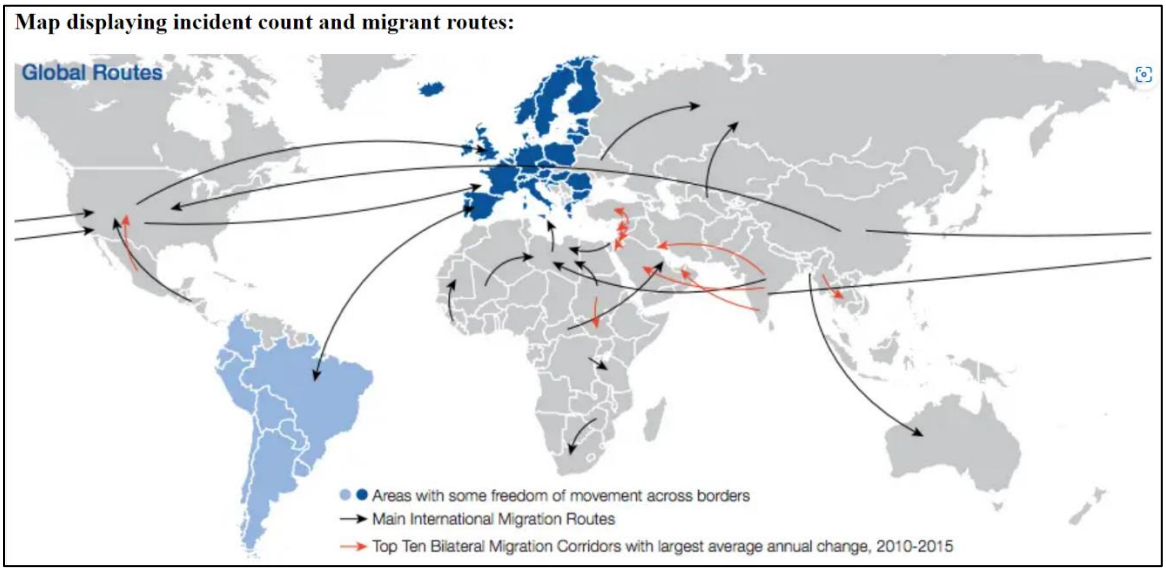
3. Data Processing Implementation:

Data Processing Tools: We will utilize data processing packages from programming languages like Python, R, or use data analysis software like Excel to perform the necessary data cleaning and transformation tasks.

Visualization Aligned with Objectives:

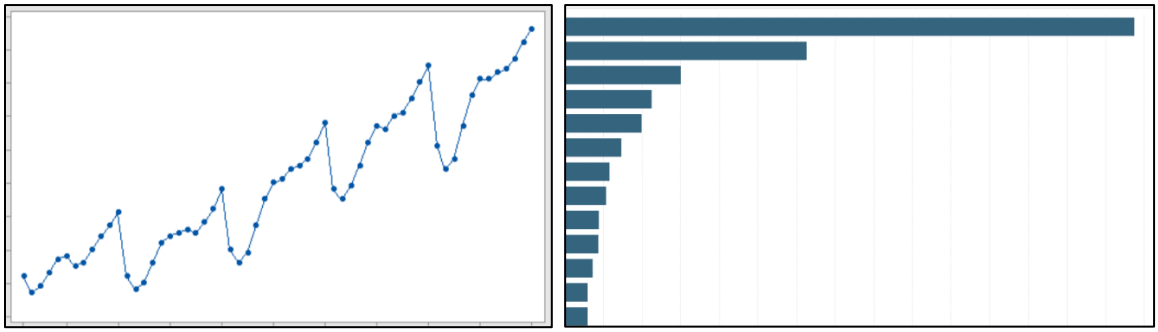
1. **Identifying the dangerous migration routes:**

- Directive lines (marks) in the world map. Position (Geographic Coordinates), Color, thickness of the directive lines can be used as channels. Example is:

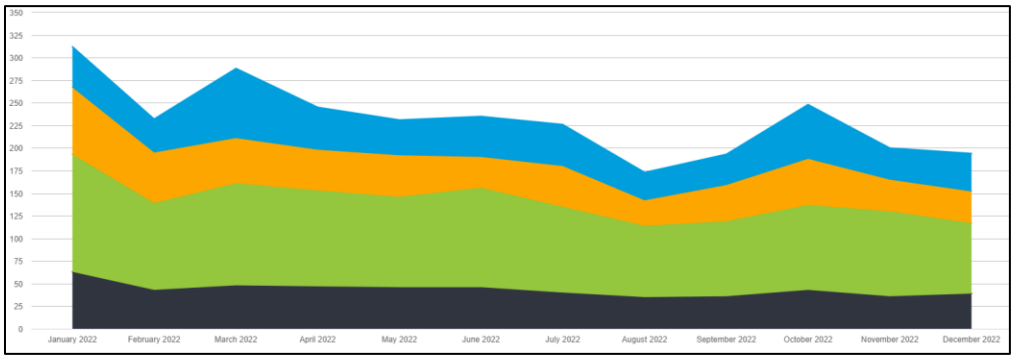


2. **to analyze temporal trends in missing migrant incidents:**

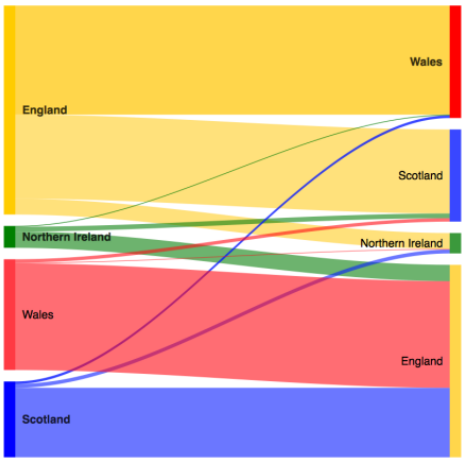
- Line charts to show trends over time. Points will be the marks and position in axis, color, and size of the points will be channels.
- Bar charts to compare incidents across years or months. Each bar represents marks and position of the bar in X and Y-axis, color, Gridlines, labels, etc. can be channels. Example is as below:



3. **to determine the causes of fatalities:**
    - Area chart can do that. The area under the line, which is filled with color, or a pattern is the mark. Axial positions, color, boundary lines are channels. Example is:

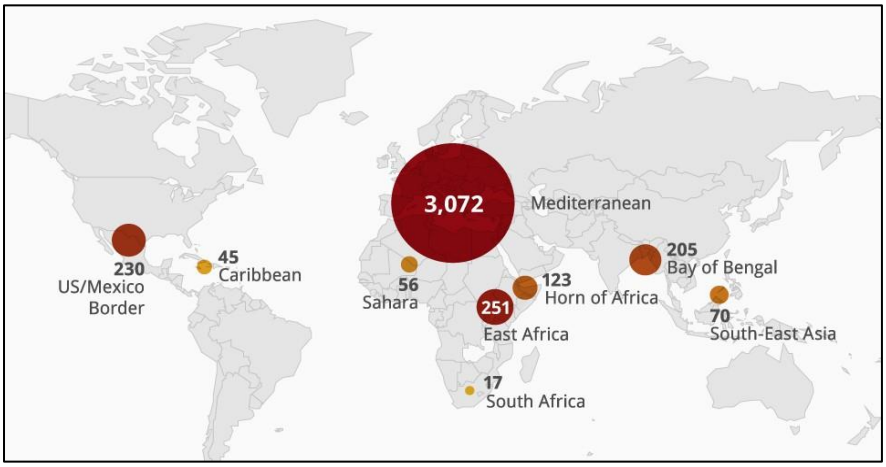


4. **Which countries receive most migrants and which countries have most people who want to emigrate:**
    - Sankey Chart. Example:



5. to visualize the geographic distribution of incidents:

- Bubble maps to visualize incidents by location with varying bubble sizes.

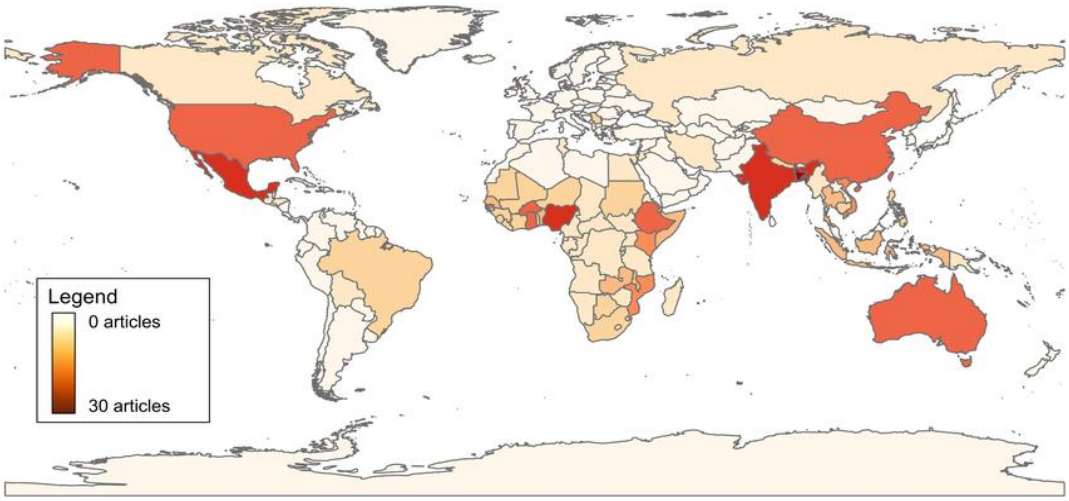


It has points as **marks** and **size** as channels to represent the fatality at locations of incidents.

Drawbacks :

1. It fails to show the migration path,
2. And fails to show where the victims came from.

- **Heatmaps to display geographic hotspots of incidents.**



The heat map shows the count of incidents in that country.

Drawbacks:

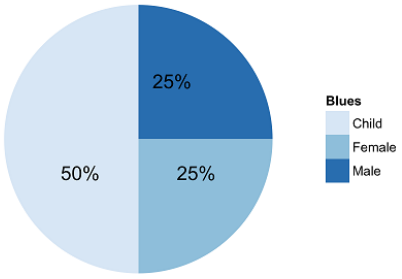
1. Fails to show the exact locations of incidents taking place,
2. Shows the whole country as a dangerous one.

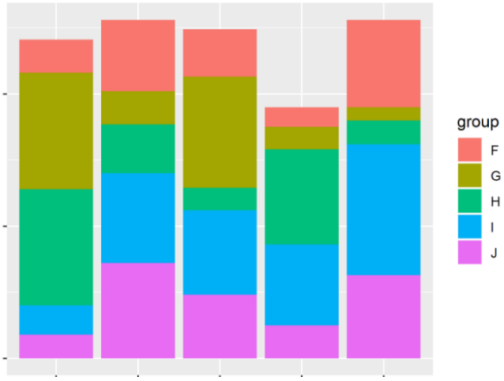
Our choice: We show area marks like in option 2 and also the migration routes in the world map. We do not have an example image as of now, but the world map will have lines(marks) to represent the migration paths and also points (marks).

Here, we try to overcome the above-mentioned drawbacks.

6. **to explore demographic characteristics of migrants:**
   - Stacked bar charts to show gender and age group breakdowns.
   - Pie charts to represent gender or age group distributions.

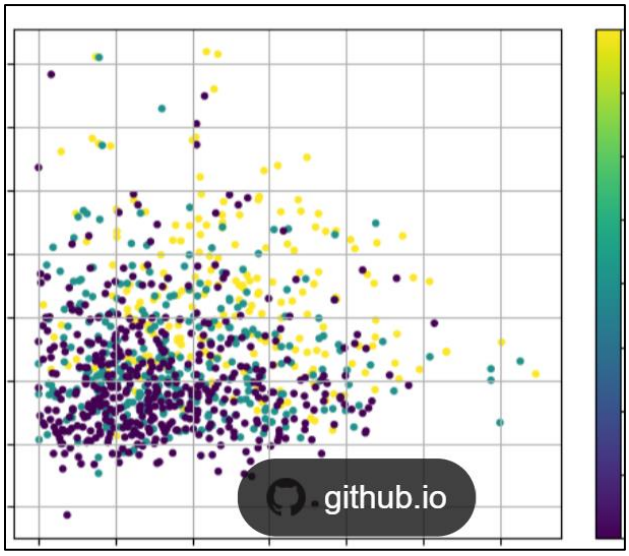
Example-





7. **to calculate survival and mortality rates:**
   - Scatter plots to explore correlations between variables.

Example:



Must-Have Features:

Here are some must-have features without which we would consider our project to be a failure:

- **Interactive Visualizations:** Interactive elements such as tooltips, filters, and interactive charts are crucial for enabling users to explore the data on their own. Without interactivity, the visualization may lack depth and engagement.

- **Temporal Analysis:** The ability to analyze migration incidents over time is fundamental. Time-based visualizations, such as time series charts or animated maps, should be included to reveal temporal trends, seasonality, and changes over the years.

- **Geospatial Mapping:** Given the dataset's geolocation coordinates, a map-based visualization is essential. It should allow users to see the geographic distribution of incidents, pinpoint specific locations, and potentially overlay other data like migration routes or regions.

- **Multiple Chart Types:** Different types of charts and graphs should be used to address various research questions. For example, bar charts for demographic breakdowns, heatmaps for geographic patterns, and line charts for temporal trends.

- **Filters and Drill-Down:** Users should have the ability to filter data based on attributes like region, migration route, or cause of death. Additionally, they should be able to drill down into specific details of incidents for a more granular view.

- **Contextual Information:** Provide context for the data by including background information, explanations of visualization techniques, and definitions of terms used in the dataset. This ensures that users can interpret the visualizations accurately.

- **Comparative Analysis:** Enable users to compare different aspects of the data. This could include side-by-side comparisons of regions, routes, or incident types to identify patterns and differences.

- **Data Transparency:** Ensure that users have access to the raw data or detailed information about each incident, including source references. Transparency is vital for trustworthiness and research integrity.

- **Mobile Responsiveness:** In today's digital age, it's essential to make the visualization mobile-friendly to reach a broader audience. Responsive design ensures that users can access and interact with the project on various devices.

- **Narrative or Storytelling:** There should be a narrative or storytelling element to guide users through the visualization and help them understand the significance of the data. This can involve annotations, captions, or a guided tour.

- **Accessibility:** The visualization should be accessible to all users, including those with disabilities. This involves providing alternative text for images, using readable fonts, and adhering to accessibility guidelines.

- **Feedback Mechanism**: We should include a way for users to provide feedback or report issues with the visualization to help improve the project and correct any inaccuracies or errors.

- **User Guidance:** We should also provide guidance on how to use the visualization, especially if it includes complex features. Tutorials, tooltips, or a user guide can assist users in navigating the visualization effectively.

These must-have features will contribute to the project's success by making the data accessible, understandable, and actionable for users.

Optional Features:

We can consider the below features to be nice to have, but not critical:

- **Advanced Machine Learning or Predictive Models:** Developing machine learning models for predictive analysis or forecasting can add depth to the project. However, these models may be considered an enhancement in the future rather than a core requirement.

- **Customization Options:** Allowing users to customize aspects of the visualization, such as color schemes or chart types, can be valuable for personalization but may not be essential for basic functionality.

- **Social Sharing Integration:** Integrating social sharing buttons or options to export visualizations can make it easier for users to share findings with others but may not be a project's primary focus.

- **User Accounts and Personal Dashboards**: Providing user accounts with the ability to save preferences or create personal dashboards can enhance user engagement but may add complexity to the project.

- **Dynamic Data Updates:** Real-time or periodic data updates can keep the visualization current, but this feature may not be crucial if the dataset is relatively static.

- **Cross-Platform Compatibility:** While mobile responsiveness is a must-have, additional efforts for compatibility with specific platforms or devices may be considered a nice-to-have based on the project's target audience.

- **Advanced Data Export Options:** Offering advanced data export features, such as CSV downloads or integration with data analysis tools, can be helpful for power users but may not be required for a basic understanding of the data.

- **Multi-Language Support:** If the project aims to reach a global audience, multi-language support can be beneficial. However, it might not be essential if the primary audience is English-speaking.

- **Collaboration Features:** Features that enable collaboration among users, such as commenting or sharing annotations, can enhance the project's social and collaborative aspects but may not be the main focus.

- **Custom Visual Themes:** Allowing users to select different visual themes or skins can enhance the aesthetics of the visualization but may not be a top priority.

- **Data Export as Visuals:** Enabling users to export specific visualizations or charts as image files or PDFs can be useful for reports or presentations but may not be required for all users.

- **User Feedback Integration:** While feedback mechanisms are important, advanced features like automated feedback analysis or sentiment analysis may be considered enhancements.

Project Schedule:

The project timeline is as below: