UNIVERSITY OF UTAH, SCHOOL OF COMPUTING, DEPARTMENT OF COMPUTER SCIENCE
VISUALIZATION FOR DATA SCIENCE

Missing Migrants

William Tang, u1169036, u1169036@utah.edu
Yu-Ching Hsu, u1264420, ruby841009@gmail.com

1 BACKGROUND AND MOTIVATION

Migration has always been an important issue, especially in recent years. For the majority of refugees and asylum-seekers, they are often willing to take risk to choose remote or difficult routes to avoid detection. Thus, most deaths stemmed from the irregular nature of migratory journeys. Countless bodies were never discovered and the officials were rarely aware of these incidents, resulting in severe psychological trauma to the people who lost their loved one since there could never be a closure for them.

Our project motivation is to raise public awareness of the risky migration routes and the deaths that they have caused. We hope, by this visualization project, we can help reduce the number of migration routes being taken and thus reduce the number of missing or death counts.

2 PROJECT OBJECTIVES

1. Highlight risky migration routes
2. Categorize the cause of death by taking the risky routes
3. Analyze the most popular migration destinations
4. Does the nationality relate to the rate of death or disappearance

3 Data Collection

We collected two files in this project:

1. Missing Migrants Dataset:Kaggle: https://www.kaggle.com/jmataya/missingmigrants.
2. World Map:Kaggle: https://www.kaggle.com/ktochylin/world-countries

The data format for the Missing Migrants Dataset is CSV format as shown in Fig. 3.1. In the data we observed that there are some blank fields exist, and that would put the difficulties during data aggregation. Hence, the detail of how we deal with these missing data would bring up in the Sec. 4.

	A	B	C	D	E	F	G	H	I	J	K	L
1	id	cause_of_deat	region_origin	affected_natio	missing	dead	incident_region	date	source	reliability	lat	lon
2	1	Presumed drov	Middle East	Iraq	1	1	Mediterranean	5/11/2015	IOM Greec	Verified	36.8915	27.2877
3	3	Fell from train	Central America &	Honduras		1	Central Americ	3/11/2015	La Jornada	Partially Ve	15.9564	-93.6631
4	4	Presumed drov	Middle East			1	Mediterranean	3/11/2015	Hellenic Cc	Verified	36.50439	27.36325
5	6	Drowning	MENA		6	4	Mediterranean	1/11/2015	Reuters	Partially Ve	37.2856	27.0866
6	7	Vehicle accide	South East Asia	Cambodia		4	Southeast Asia	1/11/2015	Phnom Per	Partially Ve	13.3611	100.985
7	8	Drowning	MENA			11	Mediterranean	1/11/2015	Hellenic Cc	Verified	37.75181	26.96029
8	9	Drowning	MENA			1	Mediterranean	1/11/2015	Hellenic Cc	Verified	38.35754	26.45216
9	10	Drowning	MENA			1	Mediterranean	1/11/2015	Hellenic Cc	Verified	39.27619	26.38344
10	11	Drowning			0	3	Mediterranean	28/10/2015	AP	Partially Ve	37.4578	26.9721
11	12	Died of unknov	MENA	Syria		1	Mediterranean	28/10/2015	IOM Greec	Verified	39.37082	26.17647
12	13	Drowning	MENA	Iran		1	Mediterranean	26/10/2015	IOM Greec	Verified	39.28098	26.38413
13	14	Presumed drov	South Asia	Afghanistan	7	3	Mediterranean	23/10/2015	IOM Greec	Verified	39.27301	26.42327
14	15	Presumed drov	Middle East/ South Asia			1	Mediterranean	1/11/2015	Hellenic Cc	Verified	39.27407	26.38619
15	16	Drowning		Afghanistan		4	Mediterranean	31/10/2015	Hurriyet Da	Partially Ve	39.27992	26.38894
16	17	Exposure. Died	South Asia	Afghan		1	Europe	31/10/2015	AP	Partially Ve	45.2604	19.1706
17	18	Drowning				1	Mediterranean	30/10/2015	AP	Partially Ve	39.27407	26.38825
18	19	Drowning	Sub-Saharan Africa	Likely Comorian		8	Sub-Saharan Afr	30/10/2015	LINFO	Partially Ve	-11.8028	43.8519
19	20	Drowning	Middle East/ South Asia			1	Mediterranean	30/10/2015	Hellenic Cc	Verified	37.4578	26.9721

Figure 3.1: Caption


```
{"type":"FeatureCollection","features":[
{"type":"Feature","properties":{"name":"Afghanistan"},"geometry":{"type":"Polygon","coordinates":[[[61.210817,35.650072],
{"type":"Feature","properties":{"name":"Angola"},"geometry":{"type":"MultiPolygon","coordinates":[[[[16.326528,-5.87747],
{"type":"Feature","properties":{"name":"Albania"},"geometry":{"type":"Polygon","coordinates":[[[20.590247,41.855404],[20.
{"type":"Feature","properties":{"name":"United Arab Emirates"},"geometry":{"type":"Polygon","coordinates":[[[51.579519,24
{"type":"Feature","properties":{"name":"Argentina"},"geometry":{"type":"MultiPolygon","coordinates":[[[[-65.5,-55.2],[-66
{"type":"Feature","properties":{"name":"Armenia"},"geometry":{"type":"Polygon","coordinates":[[[43.582746,41.092143],[44.
{"type":"Feature","properties":{"name":"Antarctica"},"geometry":{"type":"MultiPolygon","coordinates":[[[[-59.572095,-80.0
{"type":"Feature","properties":{"name":"French Southern and Antarctic Lands"},"geometry":{"type":"Polygon","coordinates":
{"type":"Feature","properties":{"name":"Australia"},"geometry":{"type":"MultiPolygon","coordinates":[[[[145.397978,-40.79
{"type":"Feature","properties":{"name":"Austria"},"geometry":{"type":"Polygon","coordinates":[[[16.979667,48.123497],[16.
```


Figure 3.2: Caption

4 Data Pre-processing

4.1 Data Clean Up for World Map

After reading the Missing Migrant Dataset by d3.csv(), the original data set contains the keys in Tab. **??**. As mentioned, there exist some missing in the data. To deal with it, we use .filter() to filter out the data if it's "$longitude$" and "$region\_origin$" are "", which is blank.

Additionally, the "*region_origin*" and "*incident_region*" are all classified by region, which included multiple countries. Since some regions have the dispute in country of composition, we use Wikipedia as the reference to define these regions and compute the relative representative center. The regions and their obtained center is list in Fig. 4.2.

```
/*
0: "Middle East : 402.57421888, 25.9580601"
1: "Central America & Mexico : 264.12890397, 16.77088981" (lon, lat)
2: "MENA( Middle East and North Africa) 43.03124785, 28.64431799"
3: "South East Asia: 466.87500335, 13.54775471"
4: "": 0,0
5: "South Asia: 438.01171996, 22.95031305"
6: "Middle East/ South Asia: 24.774265, 46.738586"
7: "Sub-Saharan Africa: 383.73046674, 3.65378603"
8: "Horn of Africa: 403.7695346, 6.42013471"
9: "Caribbean: 283.64062071, 20.33638895"
10: "South America: 300.9375, -11.62641141"
11: "East Asia: 469.26561534, 35.17380481"
12: "Central America:  274.32421982, 13.27417061"
13: "North Africa: 377.2265625, 24.59108635"
14: "Southern Europe: 13.53515625, 41.56120985"
15: "Mixed: 0,0"
16: "Horn of Africa (P): 403.7695346, 6.42013471"
 */
```

Figure 4.1: Caption

4.2 Data cleanup for ranking table

1. We use d3's nest and rollup functions to group data and extract the necessary data columns for the visualization on the ranking table. The nationality, death and missing count are grouped by using this technique. However, incident region and cause of death appear to be more complicated for extraction. We plan to focus on extracting individual subset of the entire data to achieve efficient loading for different the interaction between the world map and the ranking table.

2. Some of the data entries are not consistent. There are a lot of unorganized data such as the name of the nationality, the cause of death and the incident region. They are mostly data entries made by the people who discovered the bodies of the migrants. Therefore, we are working on removing the data outliners.

5 Visualization Design

In our visualization design, we decide to use 3 different type of representations to visualize our data. In Figure 1, we display the origin of the migration route and the incident location of the migrants who went missing or found dead on a world map. Through this visualization, users will be able to see the patterns of dangerous routes being taken. The number of routes can be adjusted for precise selection.

Figure 4.2: Caption

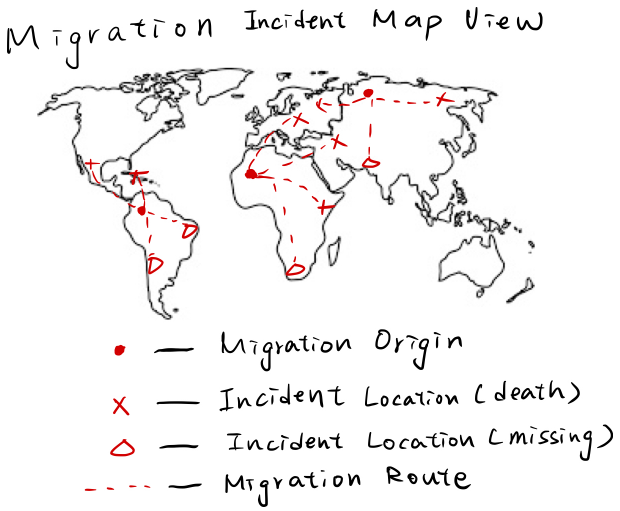


Figure 5.1: Migration Incident Map View

Upon clicking/hovering the area on the map or migrate routes, the pie chart in Figure 2 will be updated to reflect the weighted percentage of cause of the death and the nationality of the affected individuals. There will be two pie charts in total. Each is responsible for displaying the cause of death and nationality percentage.

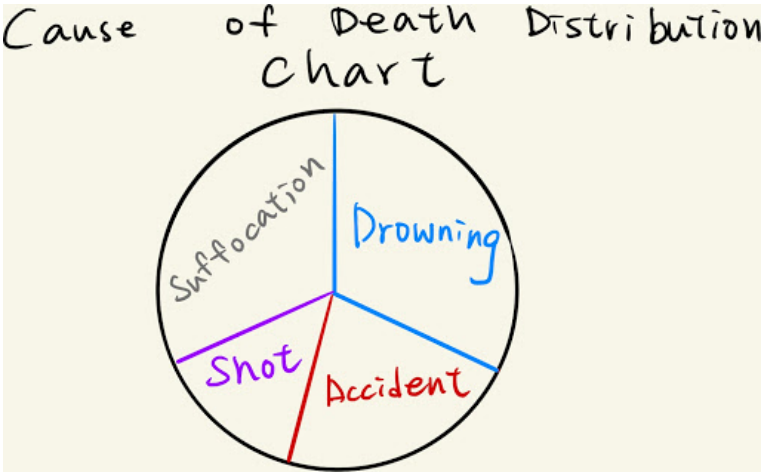


Figure 5.2: Cause of Death Distribution

Finally, in our 3rd visualization as shown in Figure 3, an overall ranking for the following categories are displayed:

- Nationality of the affected individuals
- Death count
- Missing count
- Incident region
- Cause of death

Each category title can be used to perform sorting in ascending or descending order. In addition, each data row is expandable to further display an data entry that details the specifics of the incident report. This is the most important visualization as it details the ranking of the most affected area and the total count of death/missing as well as the report for the individual incident. It allows the users to quickly find which incident region is the most dangerous one.

6 Must-Have Features

- Click:
  - Click on migration route as shown in Fig. 5.1
    - <chart> show ratio of cause of death in pie chart

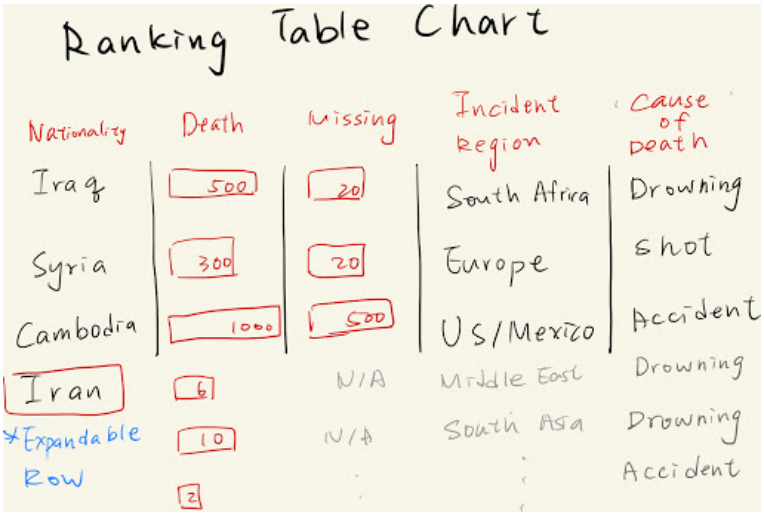


Figure 5.3: Ranking Table

    - <text> show amount of death/missing
  - Click on region as shown in the Fig. 5.1
    - <chart> show ratio of top 5 nationality that death/missing in the region in pie chart
  - Click on header in the ranking table shown in the Fig. 11.2:
    - Update the whole table based on the selected header (ascending/ descending order)

- Link two charts:
  - Click on region in the world map and highlight the corresponding rows in the ranking table for which incident region equal to the selected region

- Dropdown:
  - By clicking on the nationality in the ranking table, dropdown and show each data of death/missing/incident region/cause of death that come up to the selected nationality.

7 Optional Features

1. Scroll bar: Shows the migration routes with top 50 missing/death by default, and let the user scroll the number of routes to show

2. Migration route animation: Creates an animation for each individual route that shows the migration pattern.

8 PROJECT SCHEDULE

- Week 1(10/3): Project topic discussing
  - S386: H1B/Green card issue
  - Collecting data
  - Visualization design
  - Motivation and background discussion
- Week 2: Project topic discussing
  - World missing migrants
  - Collecting data
  - Motivation and background discussion
  - Project objectives confirm
  - Visualization design (draft)
- Week 3 (Project Proposal Deadline 10/25): World missing migrants
  - Visualization design (draft)
  - Project proposal finished
- Week 4: Implementation
  - Word map with migration route implementation
  - Ranking table implementation
  - Process book start
- Week 5 (Project Milestone Deadline 11/08):
  - Process book continue
  - Code organize
  - Create release in Github
- Week 6: Implementation
  - Process book continue
  - Pie chart interaction
  - Mouse events and data interaction
  - Optional feature confirm
- Week 7: Implementation finish
  - Process book almost finish
  - Code and design adjustment

- Week 8: Prepare for final submission
  - two-minute screen-cast with narration video
  - Code and design adjustment
  - Project website
- Week 9 (Final Project Deadline 11/27):
  - Integrated the works
  - Peer assessment

9 Related Work

10 Project Peer Feedback

- Feedback group: Unemployment rate across USA
- Peer review members: Mingxuan Luo, Qing Ye, Yulan Wang

	Feedback	Solution
1	Estimated route misleading	Replace by straight line
2	Tooltip active region unclear	Highlight the region that allow interact with pie chart (exist data) when hover
3	Tooltip: add Date data when hover	Accepted
4	Additional storytelling	Survey if any news affect the migrants behavior, or reporting the related topic
5	Target audience	Hang to the local government authority for better rescue activities
6	Animation in route	Footprint disappear

Table 10.1: Caption

10.1 General Questions

- Are the objectives interesting to the target audience?

  They are interested in the idea of displaying how many lives were lost during the allegedly illegal migration and where the incidents happened. Mingxuan pointed out that the target audience can be government officials as well as law enforcement because these are the organizations that have the power to pursue the potential migrants from engaging in such risky business.

- Is the scope of the project appropriate? If not, suggest improvement.

  The all agree that the project scope is appropriate. They were concerned about the data size being small since we only have access to the incidents from 2013-2016. However, we do have substaintial data set to show. It should not be a problem.

- Is the split between optional and must-have features appropriate? Why?

  Yes, Mingxuan also suggested we should not use estimated route as it could be misleading

- Does the visualization scale to the used dataset? Could it handle larger but similar datasets?

  They were concerned about the routes shown on the world map would have overlapping issues if the data size is big.

- Is the project plan detailed enough? Is a path to the final project clear?

  They all agreed and liked the schedule for our final project.

- Is an interesting story told?

  They suggested that we should include TV news report on certain incidents as part of our storytelling to make it more interesting.

10.2 Visual Encoding

- Does the visualization follow the principles used in class?

  Yes. They also remind us to use clear channel to emphasize what we really want to show. We are working on the data attribute, mark and the channel as we work toward the final phase.

- What is the primary visual encoding? Does it match to the most important aspect of the data?

  They pointed out the lines and the circles on the world map being the primary encoding. It does match the most important aspect of the data as it clearly shows which area is most impacted.

- What other visual variables are used? Are they effective?

  The ranking table that shows the distribution of different aspect of the data also helps the audience understand the entire data.

- Is color sensibly used? If not, suggest improvements.

  We agreed that the color encoding is the most important problem we need to figure out in our design. As we work toward the final phase, we will put more effort in this issue.

10.3 Interaction and Animation

- Is the interaction meaningful? If not, suggest improvements.

  Mingxuan suggested we should add an creative animation in displaying the migration routes since our design currently lacks animation. They were also concerned about the default display for the donut chart and world map. We should try to address the default status for all of our interaction designs. Other than that, the overall interaction design is meaningful.

- If multiple views, are they coordinated? If not, would it be meaningful?

  They agreed that all three views/charts are coordinated and any interaction with the world map will update the table and donut chart accordingly.

- Is there any animation planned? Is it clear? Is it intuitive?

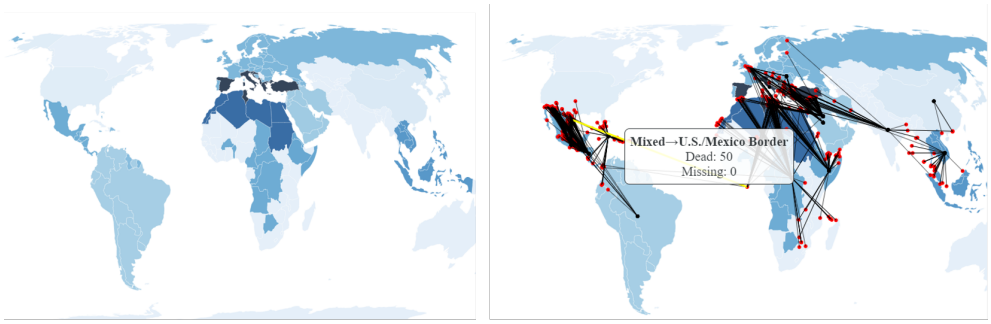
  As we mentioned, the animation for the migration route will likely be added in our design. However, we did not present it in the proposal. We will try our best to accommodate this.

11 Implementation

11.1 World Map

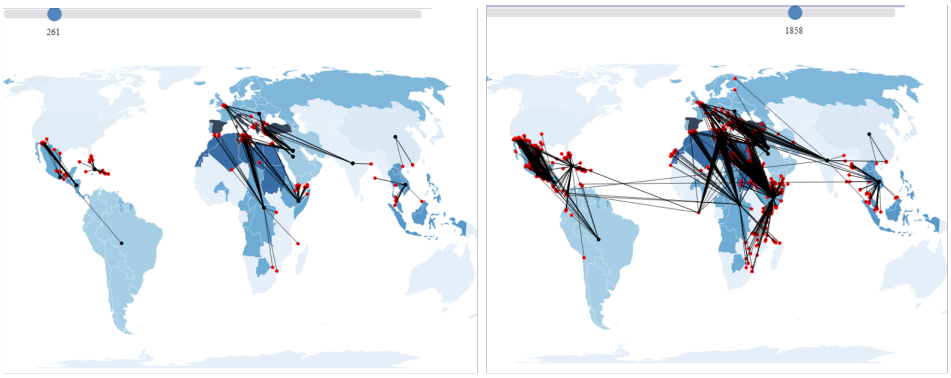
First, we color code by the total death amount based on region. The higher amount of death occurred ($incident\_region$), the darker the blue is. Secondly, plot the origin region in black circle and the incident region in red circle, and link the pair in black line.

For the hover tooltip: we show the detail information of the hovered route (change color in yellow) including "from where to the end" (A place →B place), the actual mount of death and missing people.



(a) (b)

Figure 11.1: ddd



(a) (b)

Figure 11.2: ddd