The background of the slide features a close-up, slightly blurred image of two hands holding a white sheet of paper. On the paper, a line graph is visible, showing a series of connected points that form a curve. The lighting is soft, and the colors are muted, with a focus on the white of the paper and the skin tones of the hands.

Reveal the Truth About Migration

COM-480 Data Visualisation

Jonathan Doenz
Aleksandar Hrusanov
Asli Yörüsün



SWISS FEDERAL INSTITUTE OF TECHNOLOGY IN LAUSANNE, SWITZERLAND

[HTTPS://COM-480-DATA-VISUALIZATION.GITHUB.IO/COM-480-PROJECT-VIZKHALIFA/](https://com-480-data-visualization.github.io/com-480-project-vizkhalifa/)

This report was done under the supervision of teaching assistant Kristina Gligoric with the support of the Professor Laurent Gilles Marie Vuillon within a total of 14 weeks, from February 17th to May 29th of 2020.

May 2020



we all
are
migrants

Contents

1	Introduction	5
1.1	Overview	5
1.2	Motivation	5
2	Discovering what to do	7
2.1	Related Works	7
2.1.1	What Have Others Done on Related Topics	7
2.1.2	Why is our approach original?	7
2.2	Explanatory Data Analysis	7
2.3	Brainstorming	8
3	Implementation	11
3.1	Project Architecture	11
3.2	Challenges	11
3.3	Design and Interaction Elements	12
3.3.1	World Migration on the Map, <i>see Fig.2.3</i>	12
3.3.2	Diverging Horizontal Bar Chart, <i>see Fig.2.4</i>	12
3.3.3	Sankey Diagram, <i>see Fig.2.5</i>	12
4	Further Evaluation and Peer Assessment	13
4.1	Further Evaluation	13
4.2	Peer Assessment	13



1. Introduction

1.1 Overview

Debates on migration often appear in the media. They usually focus on a single event or a specific group of people and aim at eliciting emotional reactions from the reader. In addition, many of available resources have a very narrow scope as well (e.g. limited to certain region of the world, having shorter time frame etc). It is, therefore, hard to gain comprehensive insight into the complex topic of migration and to build an objective idea about migration as a global phenomenon. **Through our visualization, we try to present a wider set of information, gathered from various sources, in an interactive and intuitive way. We hope to raise awareness and spark interest for further exploration.**

1.2 Motivation

A recent such example is the European refugee crisis. Our attention was focused on this particular group of people migrating in the European Union from across the Mediterranean Sea or through Southeast Europe. While the media often presented this as a local one-off occurrence, the reality is way more constant and widespread. While this is a more specific example, it still is a big part of global migration.

The main goal of our project is to raise awareness to the global scale of *migration*, as well as to put forward some fine grain details about it. Many of the resources and data we found were specialized or localized, so we wanted to gather a collection of

datasets to build a bigger picture. We also wanted to provide our users with a more engaging way to explore the data, so we aimed at having a dynamic and an interactive variety of visualizations.

From a global perspective, our visualizations present the migration flow between any two countries averaged over 5 years by addressing the biased information issue on the news about the global migration. So the next time someone claims that there are too many immigrants from this particular country to that particular one, you have a mean to approve or disprove the claim. Besides, you can have an overview of what are the countries that have the highest (and lowest) numbers of immigrants overall and with respect to their total populations. Furthermore, it presents, comparatively, countries' development levels and income levels. We also wanted to help analyse some more specific subtopics of migration. So, we provided intuitive ways of exploring age and gender distribution across international migrant stock, as well as common reasons and locations for the death of migrants and refugees along the main migration routes, hoping to help reveal any common patterns.



2. Discovering what to do

2.1 Related Works

2.1.1 What Have Others Done on Related Topics

Professor Guy Abel has made very significant work in studying global migration as well as visualizing it. The two following papers of him are based on data sets we use.

- [Estimates of Global Bilateral Migration Flows by Gender Between 1960 and 2015](#)
- [Bilateral international migration flow estimates for 200 countries](#)

Visualizations by Prof. Guy Abel:

- [The global flow of people](#)
- [Migration stock as chord diagrams](#)

Researchers from University of Washington and The Nature Conservancy modeled potential habitat for 2954 species using climate change projections and the climatic needs of each species, which we used as an inspiration for a map visualization:

- [Animal migration](#)

The other important work done by [1] shows fascinating visualisations about Mapping internal connectivity through human migration in malaria endemic countries.

- [Internal migration flows in low and middle income countries](#). Additionally, the related data set can be found [here](#) [2], [3].

We particularly liked the layout of this [website](#) and used it as further inspiration material for our initial website layout. Additionally, we found this [website](#) that represents migration flows on the global level.

We used it as inspiration material for our first visualization map.

2.1.2 Why is our approach original?

We were unable to find visualizations showing the migration flows directly on a world map and at the granularity of single countries. The migration movement is just as important as the final raw numbers of migrants in a given country. We provide [a more global picture of migration movement as well as to allow to easily compare the magnitude of migration flow between any countries in one place](#). On top of that, we incorporate [information about development and income level of countries to give additional layer of analysis of the global migration movement](#). Our detailed gender-and-age based visualization of international migration stock can give rise to patterns in terms of common destination countries for male/female migrants, as well as give a distribution of their age ranges. Lastly, we provide [a very intuitive way to investigate how dangerous the main migration routes are and what are the most common death causes, allowing for analysis of safest routes](#). We present a unique combination of information, enabling for a more extensive and informed analysis.

2.2 Explanatory Data Analysis

The data set that we collected for our purpose can be found [here](#). In this exploratory data analysis, we studied the Migration Flow dataset, which contains

the migration flows between any two countries aggregated over 5 years intervals from 1960 to 2015, for women and men separately. The details of the analyses performed can be seen in the file `com480project_vizkhalifa/migration_datasets_EDA.ipynb` where we mainly focused on the recent time period: 2010-2015. Here we display a subset of the results obtained.

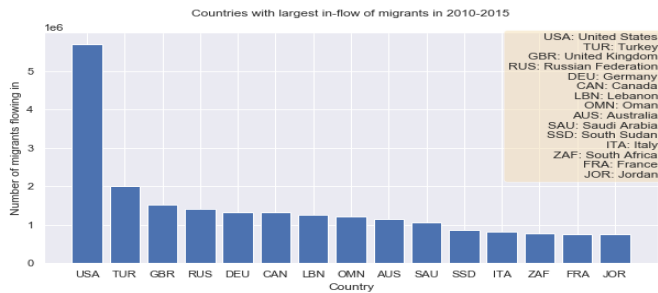


Figure 2.1: Countries with largest inflow of migrants in 2010-2015

We first looked at the countries with the highest overall migrant inflows without separating women and men. We observed that the USA is the country with the largest inflow of migrants by far as can be seen in the Fig.2.1.

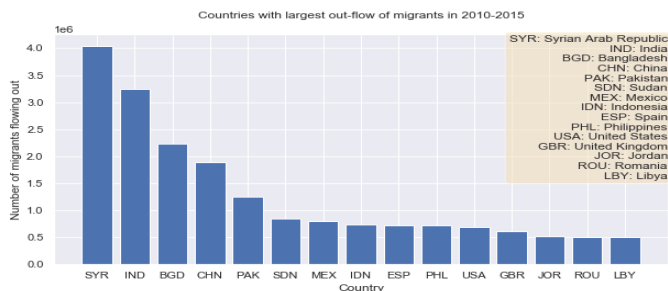


Figure 2.2: Countries with largest outflow of migrants in 2010-2015

Looking at the largest migrants' outflows in Fig.2.2, we noticed the vast impact of the civil war undergoing in Syria since 2011.

2.3 Brainstorming

When we settled on the topic of migration, we knew that part of our visualization would be on the world map. We wanted to show the global scale of migration at one place and to visually represent the migration *movement* in an intuitive way, making it easy to

analyse major origin and destination countries. For this purpose, we envisioned having something like Fig.2.3- a *choropleth map* representing the magnitude of inflow/outflow of migrants, given a user-selected country, direction representation, connecting origin and destination country pairs, and a set of possible filters for the user to explore different variables (e.g. time-period, gender).

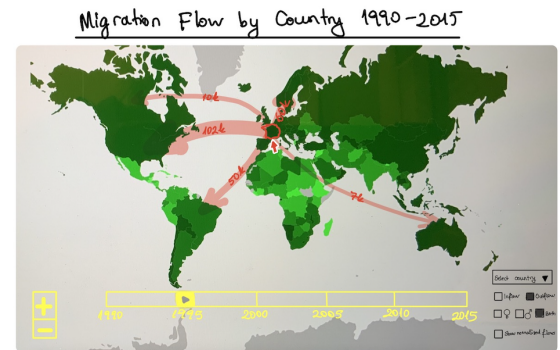


Figure 2.3: Migration Flow by Country 1990-2015

We were inspired by our topic and initial idea, but we wanted to expand and provide more fine grained details on the matter. We extracted data on *International Migration Stock* based on gender and age-range. We thought that by visualizing the distribution across age and gender, users could gain some insight into common patterns of migrants arriving at certain countries. Given that, for each destination country, we have two genders (male and female), and a handful of age-ranges, we decided that the best way to represent this data is with a *two-sided diverging bar chart*, see Fig.2.4.

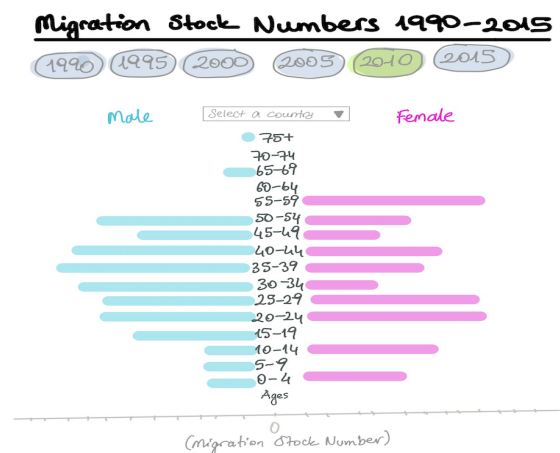


Figure 2.4: Migration Stock Numbers 1990-2015

We also learnt about a group of common *migration routes* - global routes the majority of migrants travel through. We thought it would be insightful to explore commonalities within each of those routes. After searching for datasets related to migration routes, we decided to use a *sankey diagram* to visualize migrant and refugee deaths based on migration route, cause of death, and location where the event occurred, showing information about most and least dangerous routes (see Fig.2.5).

✖ Extra

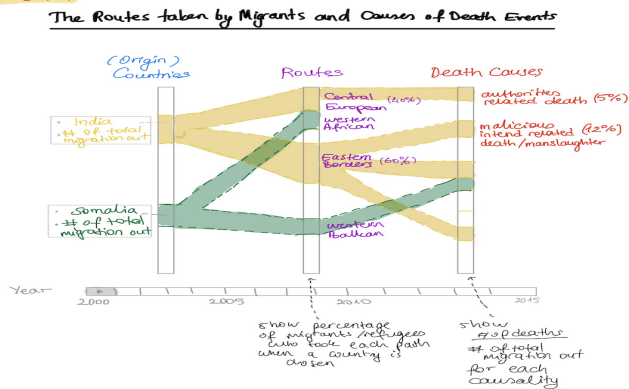


Figure 2.5: The Routes Taken by Migrants and Causes of Death Events

We extracted categorical information about countries' level of development and income level. We envisioned representing this information in some sort of a *sunburst diagram*, having country names around the inner-most circle and representing in outer layers the total number of migrants coming from/going to countries of different development and income levels (see Fig.2.6).

✖ Extra

Migration Stock Numbers Based on Countries' Development Level

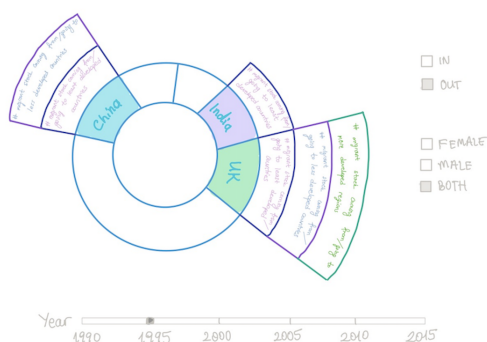


Figure 2.6: Migration Stock Numbers Based on Countries' Development Level

One major subgroup of migrants are *refugees*. We thought of dedicating a separate visualization to them. The only somewhat complete data we found was about refugee population by country. The most meaningful way to represent it seemed a map, showing the magnitude of refugee population for each country (e.g. by circles of different radius, see Fig.2.7).

Refugee Population by Country 1990-2015

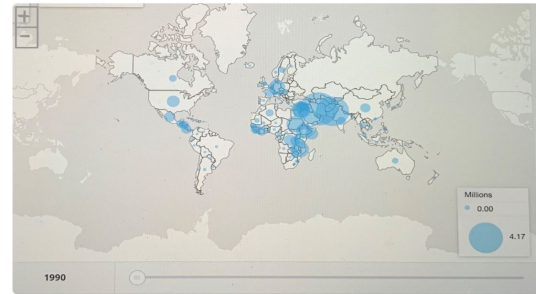


Figure 2.7: Refugee Population by Country 1990-2015

Lastly, we thought of having some brief introductory plots at the beginning of our website (e.g. migration flow w.r.t gender, migration flow w.r.t. inflow/outflow numbers), see Fig.2.8.

Migration Flow 1990-2015

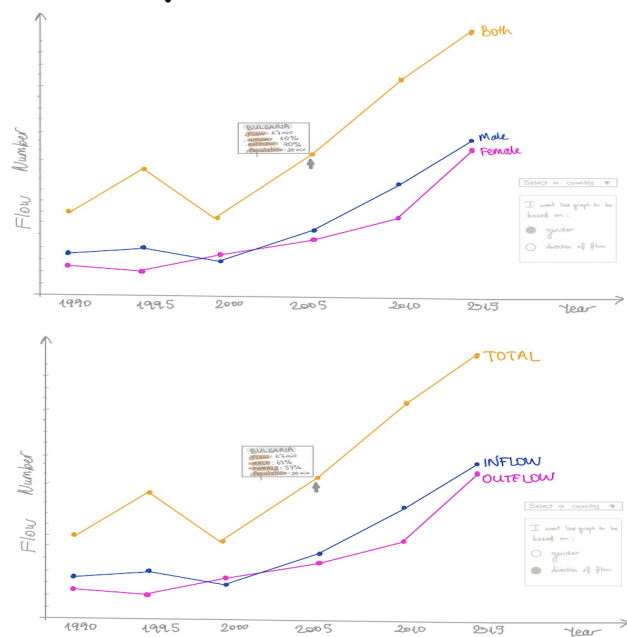


Figure 2.8: Migration Flow w.r.t. Gender (above) and w.r.t. In/Out-Flow Numbers (below)



3. Implementation

3.1 Project Architecture

The main github repository for our study can be found in Fig.3.1 as a tree structure. The html content of our website is in `index.html`. All the files related to CSS-styling are in the `css` folder. The pre-processed datasets are in folder `data`, and all the scripts needed to make our visualizations are in the `script` folder. We also have an `images` folder containing the images and gif animations that we use.

```
├── ProcessBook.pdf
├── README.md
├── com-480-project-vizkhalifa.iml
├── css
│   ├── bootstrap.css
│   ├── bootstrap.min.css
│   ├── main.css
│   └── style.css
├── data
│   ├── International_and_totalStock_GenderAge.csv
│   ├── RefugeeMigrant_Deaths.csv
│   ├── countries_development_info.csv
│   ├── country_codes_and_names.json
│   ├── filteredsankey.csv
│   ├── migflows_gender_separated_1990_2015_filtered_without0flows.csv
│   ├── pop.csv
│   └── world.json
├── favicon/
├── figures/
├── images
│   ├── index.html
│   ├── milestone2_report.pdf
│   └── notebook
│       ├── Json_CSVCreator.ipynb
│       ├── excelParser.ipynb
│       ├── migration_datasets_EDA.ipynb
│       └── migration_datasets_transformation_to_use_in_js.ipynb
├── scripts
│   ├── constants.js
│   ├── display_world_map.js
│   ├── filter_panel.js
│   ├── helpers.js
│   ├── migration_stock_chart.js
│   ├── sankey_filter.js
│   ├── sankey_helper.js
│   └── slider.js
├── sketches
│   ├── Images_pdf.pdf
│   ├── sketch1.jpg
│   ├── sketch2.jpg
│   ├── sketch3.jpg
│   ├── sketch4.jpg
│   ├── sketch5.jpg
│   └── sketch6.jpg
```

Figure 3.1: Project Github Repository Architecture

3.2 Challenges

The initial challenge we faced was finding appropriate data. We felt strongly about the topic we had

chosen, so we were determined to find good data sets. We found plenty of resources - government organizations, university public libraries, NGOs. The problem was that the majority of those resources offered only partial or strongly localized and specialized data. We had to gather and combine multiple data sets in order to comprise what we actually needed for the visualizations we wanted to create. After some hard work, we successfully synthesized rich data sets to that fit our project needs.

A very challenging part of the project for us was that none of us had prior experience with D3.js and little such with JavaScript. Hence, it was at first overwhelming to capture what's going on in this world and was a steep learning curve throughout the semester, especially when we reached the project part. We spent a lot of time experimenting with D3.js and browsing for additional resources. We learnt a lot during the implementation of our interactive visualisations but we realized that we have set ourselves an ambitious plan and, due to time restriction, we might not be able to execute everything we had planned (e.g. the sketches proposed under the section 2.3). As a result, we decided to discard some of the visualizations and to reorganize others. In the next section, you will read only about the visualizations that we ended up implemented in our website and the changes that we made onto them with respect to the original plan.

Another big challenge we faced had to do with the D3.js library itself and the different versions it supports. As mentioned above, we explored some

additional D3.js resources online during our project work. We often faced a common issue - different examples supported different version of the library (e.g. v.3, v.4 being the most common ones). It seems to be the case that, in newer versions, besides the newly introduced functions, there are some changes made to old functions' names, signatures, and syntax. Adapting a given function call to a different version proved to be painful on various occasions. After some significant time spent on such activities, we found out we can support multiple D3.js version within the same script file which saved us a lot of time and effort for the rest of the project work.

The last significant challenge for us was related to the fact that none of us had much prior web development experience. We knew the basics of HTML and CSS which helped us create a simple website. The obstacles appeared when we had to tackle more advanced styling and position of different elements. We had a vision in mind about what we would want our website to look like, but quickly found out that it will take us more time than we have to spare to learn how to actually achieve it. Finally, we decided to keep the website look minimal but functional and easy to navigate. In that way we would be sure to have enough time to focus on the visualizations and make them as good as we intended to.

3.3 Design and Interaction Elements

At the end, we decided to implement three main visualizations with a diverse range of interactive elements.

3.3.1 World Migration on the Map, see Fig.2.3

Our first visualization, the *World Migration on the Map*, contains three main parts which can be toggled through a drop-down menu above the map, to the right. For all three of them, we use the same projection to transform the TopoJSON information into a 2D world map. From the *Map Type* drop-down list on top right, you can see the following functionalities onto the map:

- *Migration Flows* represents flows of migrants given a origin/destination country. It gives users the option to select a country of interest either from a drop-down list of countries, or simply by clicking on the country representation on the map. Once a country is selected, the map is colored (e.g. a choropleth map), representing the magnitude of the respective

countries' migration flow, flowing into or flowing out of the selected countries (depending on user selection). Animated curves trace the distance between each country pair. We also provide an option to show the flows of women or men separately. And flows normalized by the population of the selected country.

- *Development Levels* is a choropleth map, showing the development level of each country, from *least developed* to *more developed*.
- *Income Levels* is a choropleth map, showing the income level of each country, from *high income* to *low income*.

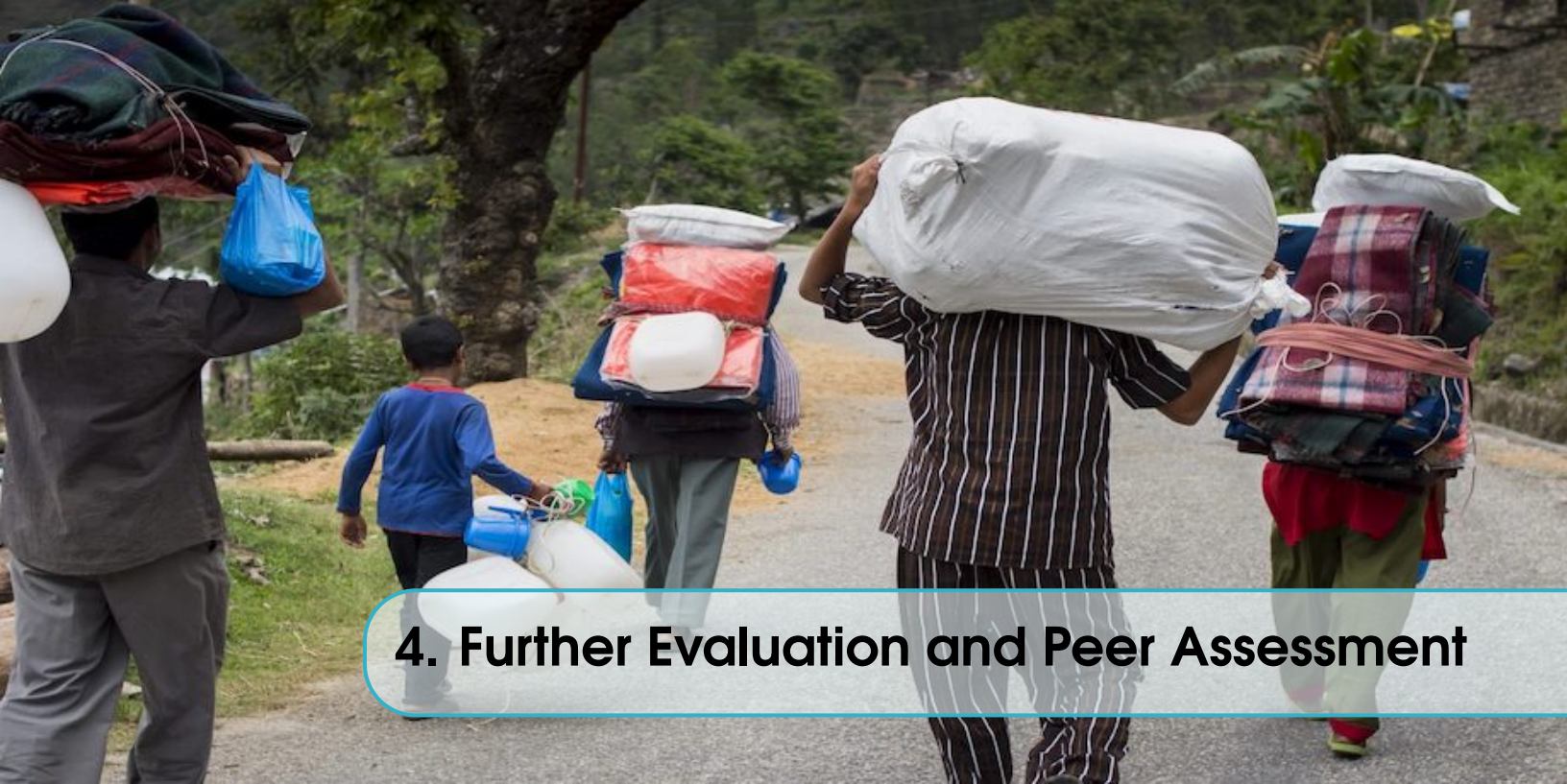
In addition, when the user hovers over a specific country, they can see information about the country for the selected time period, such as population, total inflow and total outflow count, and the respective percentages of the population.

3.3.2 Diverging Horizontal Bar Chart, see Fig.2.4

Our next visualization is a diverging bar chart, visualizing international migration stock based on gender and age groups. Destination country and year can again be selected from a drop-down menu and a time slider, respectively. For the given selection, the bar chart represents the number of male (to the left) and female (to the right) migrants for each of the 15 consecutive age groups of increments by 5. Hovering over a specific bar a pop-up presents comparative information with previous time period data (if applicable), as well as percentage ratio with respect to gender for the current year's data.

3.3.3 Sankey Diagram, see Fig.2.5

Lastly, we implemented a *sankey diagram* with D3 elements; 3 columns as being *migration route*, *event (death / missing) causality*, and *specific reason of the event*. There are two dropdown list: the migration route and the event year. When the page first loaded, it shows only 3 columns as mentioned above. If the visualization is filtered by choosing a specific migration route from the dropdown list, it is updated with an additional column to show the exact location along the chosen route, where the death occurred. It is also possible to click over each node in each column of the sankey diagram to show its distribution which is the number of the event happened in our study. This will give an insight into which routes tend to be more dangerous and which routes are safer.



4. Further Evaluation and Peer Assessment

4.1 Further Evaluation

We enjoyed a lot throughout our journey in Data Visualisation project by discovering the world of D3.js. However, we are upset about not being able to visualise all the sketches we had initially. Hence, for further improvement of our website, we plan to implement those ideas that are currently not presented on the website. We will also, keep updating the data we used for our study to be consistent with our aim at being unbiased and giving more accurate data to our readers. Last but not least, we would want to further improve the overall style of the website and to make it look more unique, better fitting the quality of the visualization it hosts.

4.2 Peer Assessment

From the very beginning of the semester, we had brainstorming sessions both in person and online on the topic that we are going to implement for our data visualisation project. We had various ideas and we all listened to each other, discussed the feasibility of the ideas and the preferences of our team members, until we reached a consensus about the work plan. We all cared about the migration topic and every one of us spent so much time to find comprehensive data to help better express our thoughts and to find additional resources to guide us through the realization of our ideas. We all had strong and weak areas of expertise, but the team dynamics were great and work distribution happened seamlessly, resulting both in

efficient work and learning experience for everyone. Jonathan took part in the Explanatory Data Analysis by using our main data set which is Migration Flow. Asli did the data handling & pre-processing of International Migration Stock, Migrant and Refugee Death Causalities, and also did some preprocessing in Migration Flow data set to using as input data files to create our interactive graphs as well as the screencast. Aleksandar built the skeleton of our website and implemented the Diverging Horizontal Bar Chart, and World Migration on the Map. Jonathan implemented the World Migration on the Map. Asli implemented the Sankey Diagram and curated the content of our website. Throughout the course of the project, we all helped each other improve the visualizations by doing code review, code refactoring, and styling. We also performed continuous quality assurance testing, locating and fixing bugs, in order to present our best visualisations to our readers.



Bibliography

- [1] A. Sorichetta, T. Bird, and N. e. a. Ruktanonchai, “Mapping internal connectivity through human migration in malaria endemic countries,” *Sci Data* 3, vol. 160066, 2016.
- [2] “International migration flows to and from selected countries: The 2015 revision (pop/db/mig/flow/rev.2015).”
- [3] “International migrant stock 2019 (united nations database, pop/db/mig/stock/rev.2019).”
- [4] G. J. Abel and J. E. Cohen, “Bilateral international migration flow estimates for 200 countries,” *Scientific data*, vol. 6, no. 1, pp. 1–13, 2019.
- [5] G. J. Abel, “Estimates of global bilateral migration flows by gender between 1960 and 20151,” *International Migration Review*, vol. 52, no. 3, pp. 809–852, 2018.