

# Migration Data Visualization Dashboard

## User Study

Neetha Reddy  
IMT2018050  
neetha.reddy@iiitb.ac.in

International Institute of Information Technology Bangalore

## 1 Introduction

According to the most recent UN Population Division estimates, the number of foreign migrants worldwide was 280.6 million (or 3.6 per cent of the world's population) as of mid-2020[1]. Although some people have the option to relocate, most are forced to do so owing to external forces (such as wars, lack of employment, poor economic conditions, or natural catastrophes). Understanding the causes and implications of migration, as well as forecasting future flows, is critical for the development of national public policy and urban resource planning. Visualization aids in the interpretation of vast volumes of migrant flow data and provides insights into how a country's policies influence migrant mobility. They are, in fact, necessary for extracting information from unstructured data. For example, the dataset I chose has almost 68,000 lines of raw data, and we cannot derive useful inferences from it without employing visualisation tools. I used the UNHCR migration datasets from 2000 to 2016 for the scope of visualisations in this tool [2].

## 2 Tool Overview

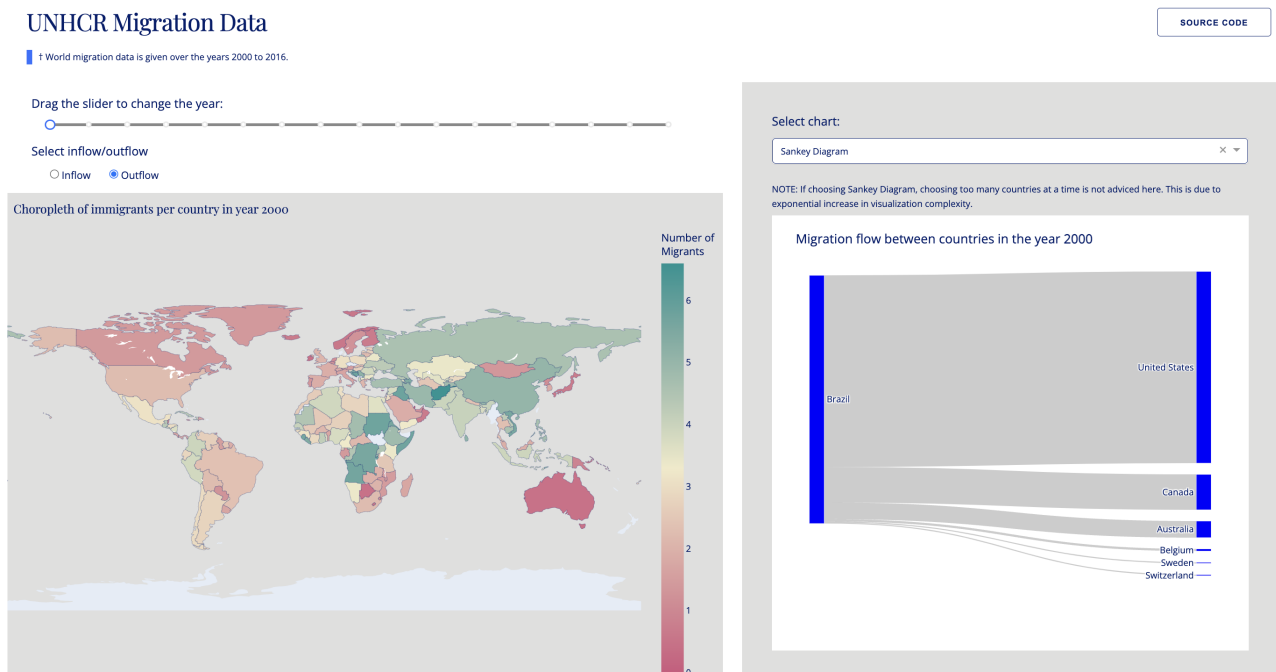


Figure 1: Dashboard starting page

This visualisation tool was created to track international migration flows into and out of each country. The dashboard has two synchronised spatio-temporal visualisation displays that are juxtaposed with one other. The user should be able to determine the inflow/outflow of migrants from the chosen regions, evaluate the composition of migrants/immigrants, and examine interconnection across nations at any given snapshot (Fig. 1.).

### 3 User Study

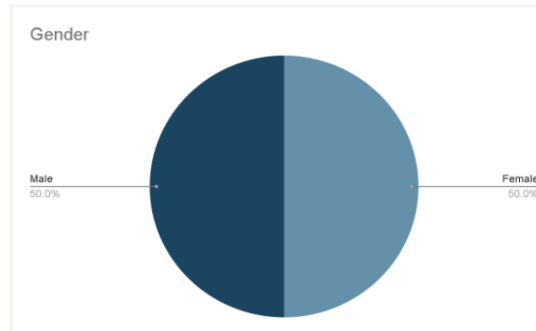


Figure 2: Gender ratio

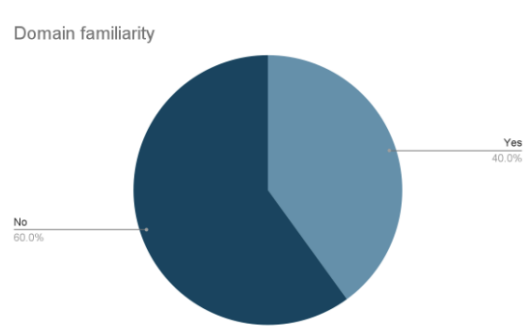


Figure 3: Having Visualization background (formal education)

This section covers the tool's user study specifics, including design, experiments, implementation, results, and findings. These processes are detailed below:

#### 3.1 Design

For the user study, semi-structured interviews (video+google form) were conducted with 10 participants (students):

- 5 male and 5 female.
- 4 having visualization background while the rest did not have any formal education in the domain.

All the participants belonged to the age group 20-22 years old. The exact age distribution is given below in Figure 2.

Participant tasks were divided into two categories:

- Dashboard-based tasks.
- Inference-based tasks.

These tasks were chosen to mimic how users might use the tool outside this experiment. The above mentioned tasks were then followed by a feedback section. Each user study took place for about 15-20 minutes.

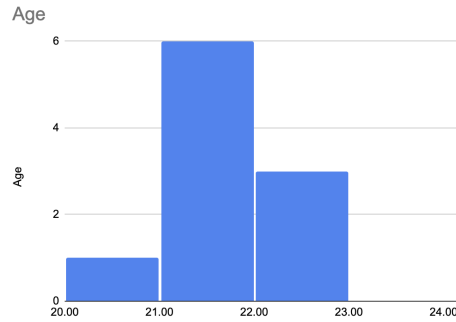


Figure 4: Age Distribution

### 3.2 Implementation

Each user study begins with a walk-through of the tool and its functionalities. Users are also provided with an additional user guide and given about five minutes time to get accustomed to the dashboard. They then have to fill a form regarding their age, gender, educational qualifications, etc. after which they are redirected to the tasks. The study contains 4 questions – 2 each from dashboard-based and inference-based tasks respectively. These questions are framed in such a way that in order to answer a question, the user has to perform certain tasks. The above mentioned tasks were fixed based on their difficulty levels and span all the visualizations used in the tool. The questions posed to the user are:

- **Dashboard-based:**
  - Select India using lasso select/box select, inflow, Sankey diagram for the year 2016. What are the countries from which highest migrant flow is observed? (Write the top 3 countries separated by comma)
  - Select Greenland, Canada and the US (outflow) and find the common migration hubs for the year 2011 using the node-link diagram.
- **Inference-based:**
  - What was the outflow from Australia like in the year 2011? (1-3 sentences)
  - What are the origin countries (Top 3) of migrants going to Greenland in 2011? Can you speculate why this flow was observed? (1-3 sentences)

After answering the above mentioned questions, users are taken to a visualization feedback section. In this section, every visualization used in the dashboard is provided along with the motivation behind each one and a 5-point satisfaction Likert scale (Figure 3.) (ranging from "Extremely unsatisfied" to "Extremely satisfied"). Apart from this, an additional option to provide additional feedback is also given.



Figure 5: Age Distribution

### 3.3 Response Data

The tool was designed in a way that minimizes ambiguity and built focusing on accuracy. Therefore, the user accuracy was observed to be 100 percent in this study. However, this might be attributed to the tailored questions and small subset of users. But by choosing such an implementation came with time trade-offs i.e. time taken by a user to navigate the dashboard and answer the question increased.

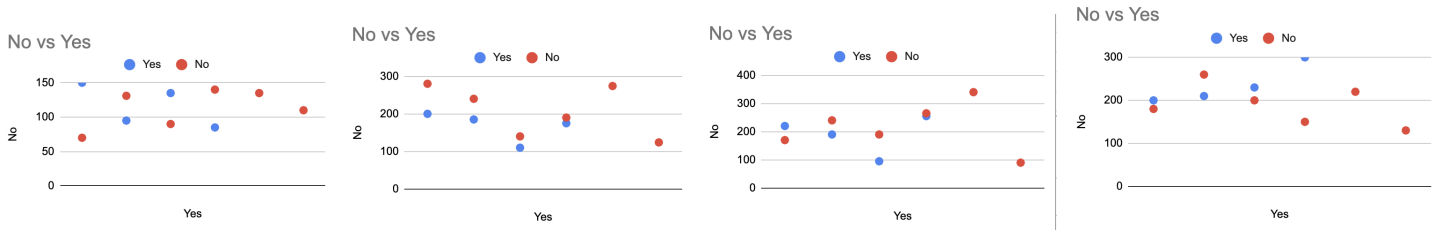


Figure 6: Having domain knowledge

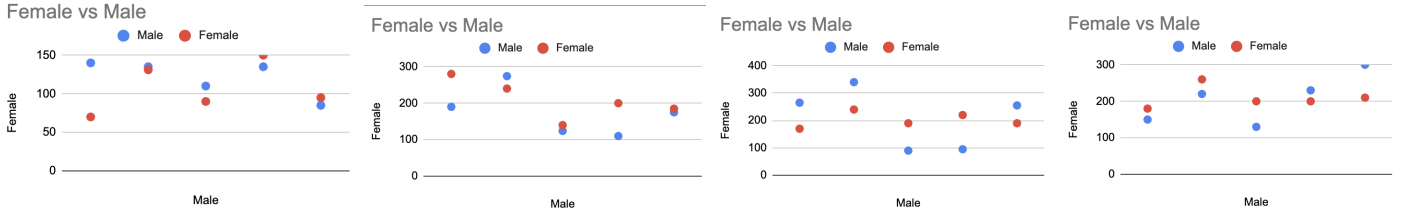


Figure 7: Male vs Female

The total average time taken for users with prior domain knowledge to complete all the questions stands at 177s which is slightly lower time compared to everyday users 181s.

Further it is observed that male participants on an average outperformed their female counterparts by about 5-6s.

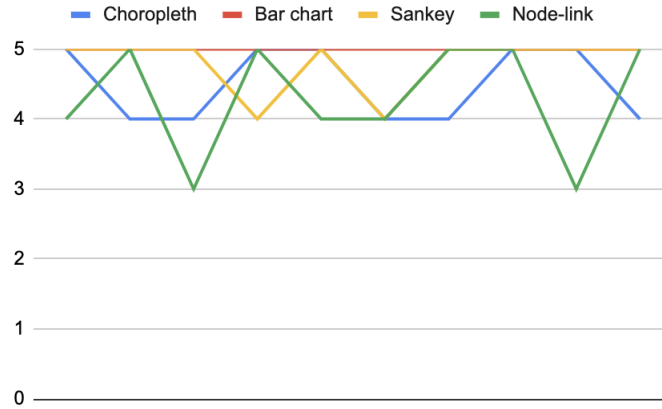


Figure 8: Feedback preferences

Most users found bar chart and sankey diagram to be the easiest to interpret followed by choropleth map and node-link diagram being the least preferred. While analysing the video footage, I also observed that users with no visualization background found it little difficult to understand and navigate the choropleth and node-link diagram.

### 3.4 Additional experiments and Challenges

This experiment was also conducted on users belonging to different age groups and educational backgrounds to mimic the everyday user. It was observed from these additional users that this tool was designed assuming the user has basic geographical knowledge and “tech-savviness”. So based on the user knowledge on international geography and their comfort levels with using technology, the time taken and accuracies per task varied. Further, when studying the additional feedback provided by users who participated in the user study, I observed that each user had suggested a different customization to make it easier for them to

infer data. This is in line with the paradox of InfoVis Tools where designers are rewarded for "generality" while users prefer specificity.

### **3.5 Future Work**

Based on feedback received from users and additional experiments conducted, I realised it would be more user-friendly to provide an alternative method for selecting specific countries. This can be done using search-guided multi-select drop-downs. Since many users faced issues understanding the node-link diagram, there is room for improvement for making this visualization more interpretable by adding additional visual encodings. Further, additional interesting visualizations can also be introduced. Another interesting suggestion is to add a widget containing all the major historic events that might be responsible for migrant flow for that year (This was actually one of the suggestions from the additional feedback section!)

## **4 References**

1. International Migration 2020 Highlights
2. UNHCR Statistical Dataset