Data Visualisa tion

Assignment 2

A dive into Forced Migration

Monday 8:00

Semester 1, 2024

Word count:

Nguyen Duy Anh - 104212294

Lai Gia Khanh- 104222015

[link]

Table of Contents

Contents

[1. Introduction 3](#_Toc161039019)

[1.1. Background and Motivation 3](#_Toc161039020)

[1.2. Visualisation Purpose 3](#_Toc161039021)

[1.3. Project Schedule 3](#_Toc161039022)

[2. Data 3](#_Toc161039023)

[2.1. Data Source 3](#_Toc161039024)

[2.2. Data Processing 3](#_Toc161039025)

[3. Requirements 3](#_Toc161039026)

[3.1. Must-Have Features 3](#_Toc161039027)

[3.2. Optional Features 3](#_Toc161039028)

[4. Visualisation Design 3](#_Toc161039029)

[5. Validation 3](#_Toc161039030)

[6. Conclusion 3](#_Toc161039031)

[References 4](#_Toc161039032)

# Introduction

## Background and Motivation

## Visualisation Purpose

## Project Schedule

# Data

## Data Source

All the data for this project come from the UN, particularly, the UNHCR and UN Department of Economic and Social Affairs. There are two datasets used for this project which can be found here <https://www.un.org/development/desa/pd/content/international-migrant-stock>

And here <https://www.unhcr.org/refugee-statistics/insights/explainers/forcibly-displaced-flow-data.html>

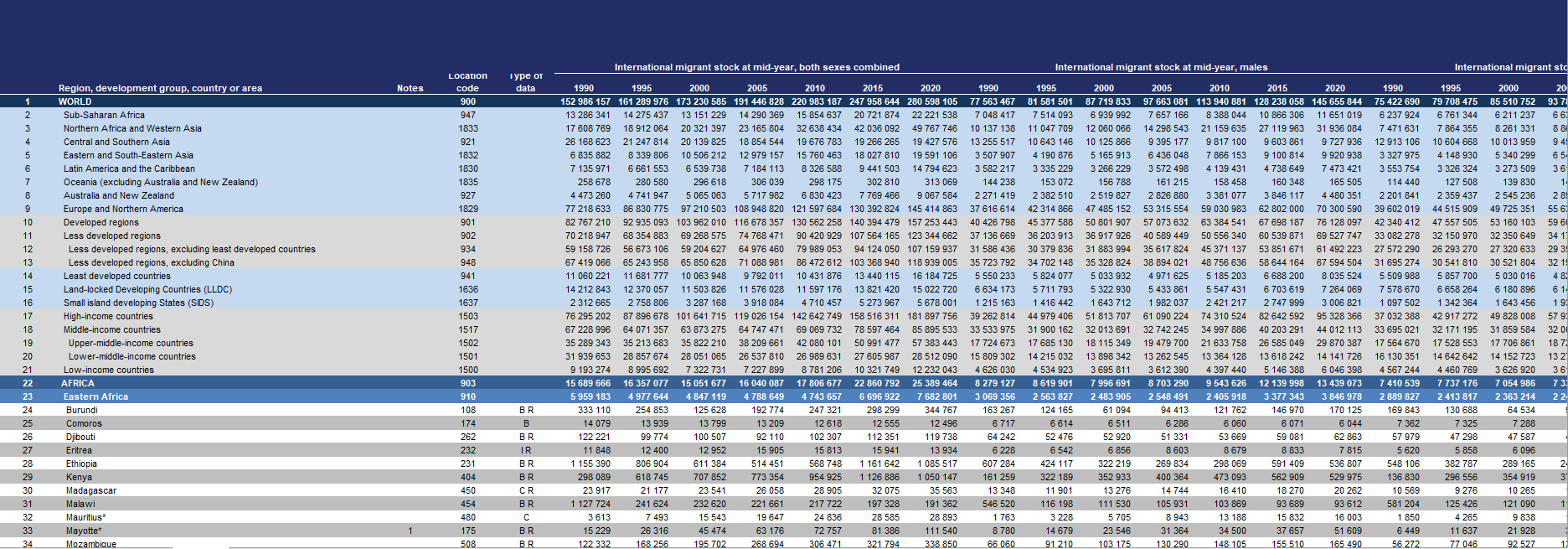
The dataset is structured as a table, containing categorical attributes such as Region, development group, country or area, Notes, Location code, and Type of data. These attributes offer descriptive insights into the regions, countries, and the nature of the data provided. Additionally, the dataset incorporates numerical values representing years from 1990 to 2020, serving as interval/quantitative data. These temporal values facilitate the examination of trends and variations over time in the international migrant stock across different regions and categories. The dataset presents estimates of international migrants disaggregated by age, sex, and country or area of origin, drawing from national statistics including population censuses, registers, and surveys. Covering 232 countries and areas globally, it provides a comprehensive overview based on official statistics concerning the foreign-born or foreign population. The dataset contains multiple tables but my team will only be using the International migrant stock at mid-years, both sexes combine to get the fullest data for the line chart.

For the second dataset, it contains 3 tables: Summary of migrant by years, Summary of Asylum and the last one contain the total data of migrant across the world from the origin country to destination country and amount of migrant from 1962 to 2023. For the second visualization, we will use the last table as our dataset. This dataset comprises information on asylum seekers originating from different countries seeking refuge in various destination countries. It includes categorical attributes such as the country of origin, its ISO code, and name, along with corresponding details for the destination country. Additionally, asylum region, type of protection sought, year of asylum application, and the count of asylum seekers are provided. These data, spanning from 1962 onwards, offer insights into the flow of asylum seekers across different regions globally, aiding in the analysis of migration patterns and refugee movements over time.

## Data Processing

Substantial data clean up are require for our datasets.

Regarding the first table, the pre-cleanup dataset has a lot of unnecessary data and stretching 294 lines, this is a problem as it will make the code run slow and messy to dealt with



*Fig1. Pre-cleanup dataset 1*

The cleanup process starts with extracting the necessary data to a new excel file.

We realize that the simplicity of the dataset will help us a lot during the coding process. In the first visualization, we only need the Years and Count adjacent to the year so we extract those two to a new table. After cleanup, this is the new dataset. This will be turn into csv file to use for the first visualization. If you notice, the dataset from UN only provided the year from 1990 to 2020, pre 1990 data are not online anymore so we have to do some more research for the full dataset. Additional data can be found here: <https://www.statista.com/statistics/679702/international-migrant-stock-1960-to-2015/>



*Fig2. Dataset 1 after cleanup*

The second dataset are much more complicated. As mentioned before, it contains comprises information on asylum seekers originating from different countries seeking refuge in various destination countries on the span of 61 years.



*Fig3. Pre-cleanup dataset 2*

We decided to only focus on 3 countries: Ukraine, Venezuela and Syria as prime example of forces migration. In the new dataset, we exclude the redundant column such as country code of origin and destination country. Luckily, there are filter built into the data table of the UN so everything is much easier. For each targeted country, we choose a time span that is eventful and saw and influx of migrant. For example, we take data from 2010 to 2023 for Ukraine. This requires filter country to only Ukraine and limit the time from 2010 forward. Moreover, we examine the data to determined the most popular destination for Ukraine refugee and put a focus on those country. After the clean up process, we have the final dataset that will be turned to csv file for coding



*Fig4. Post-cleanup Ukraine dataset*

The process will be repeat again for other countries.

# Requirements

## Must-Have Features

For our project, the requirement is to have a one-page website with HTML, CSS and d3.js for the visualization. The webpage will include an introduction, 2 type of visualization:

* Line chart: Use to show the trend of migration through history. This visualization will have a hover features. The aim of this visualization is to help user have a better idea of the ongoing increase in amount of people migrating.
* Stacked Area Chart of selected countries: This visualization main purpose is to talk about forced migration, aka people forced to runaway from their home country by either war, economic crisis or both.

In addition to charts, explanation will also be provided on why we chose this data, what happened in the period of time in the dataset and what we found out after analyzing

## Optional Features

* Connection Map: The Connection Map is a powerful tool designed to enhance the presentation and understanding of immigration patterns across the globe. This feature provides users with a visual representation of the interconnectedness of different regions through immigration flows. By leveraging a world map interface, users can explore and analyze the movement of people across borders, gaining valuable insights into migration trends and patterns.

# Visualisation Design

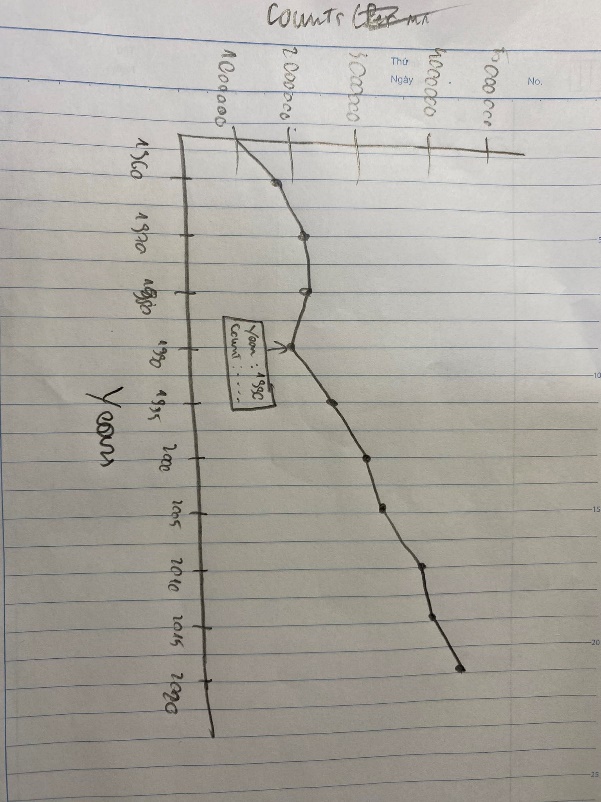
## Line Graph

### Design

The idea of the line graph is to show the evolution of migration number of the world throughout history

* X axis will represent years from 1960 to 2020
* Y axis will represent Count(people)
* Dots will be use to indicate the migration number for each year when hover mouse over, allow user to know the data better.

Sketch:



Final Design:

# Validation

# Conclusion

# References