NOVA IMS Information Management School

Data Visualization Project

GLOBAL MIGRATION PATTERNS

An Overview

Carlos Pereira | M20190426
Cátia Duro | M20190394
João Miguel Lopes | M20190465
Marta Faria | M20190178

Dashboard available at: https://dv-project-migration.herokuapp.com/

TABLE OF CONTENTS

1. INTRODUCTION

Migration refers to the movement of people from place to place. People migrate for many different reasons, which can be classified as economic, social, political or environmental.

- **Economic migration** is related to finding work or better economic opportunities;
- **Social migration** refers to the search of a better quality of life or to be closer to family and friends;
- **Political migration** occurs when people is moving to escape conflict, political persecution, terrorism, or human rights violations; and,
- **Environmental** causes of migration include the adverse effects of climate change, natural disasters, and other environmental factors (BBC, 2020).

Over the last decades, migration has become a key issue for countries all over the world. Today, more people than ever live in a country other than the one in which they were born. Estimates indicate that, in 2019, the worldwide number of migrants reached 272 million, which comprises 3.5% of the global population. Even if most of these migrants, migrate by choice, a considerable share migrates by necessity. At the end of 2018, the number of forcibly displaced people reached 70 million (United Nations, 2020).

In this context, this project objective is to explore migration patterns and their underlying causes.

2. DATASET DESCRIPTION

Given the need to understand what drives migration, our choice was to build a dataset by combining information from several data sources, namely data on political, economic and social indicators and data on the migration flows. The objective was to provide an overview of the migration patterns (where do people come from and where do they go to) while simultaneously giving a clear idea of the underlying causes that justify people leaving/entering in a specific country. We have supported our core information in three international sites (OECD, The Global Economy and Our World in Data) that make available key country indicators and further detailed information.

The main dataset was dived in two subsets: the first subset consists of a collection of political, economic and social indicators while the second regards the migration flows, i.e. the number of people entering in a country (inflow), the number of people leaving the country (outflow) and the net-migration (the difference between inflow and outflow). Besides the migration flows, the migration dataset includes both the origin country and the destination country.

Data for both subsets is analyzed for a 10-year period, on a yearly basis (from 2008 to 2017), including 204 countries. The time range to be displayed was chosen having in mind data consistency, reliability and quality while providing enough data to the viewer to have a clear perspective of the problem and to perceive recent years evolution. Even if some data was available for 2018, migration data is not easy to collect, and it largely depends on countries reporting their numbers. Therefore, it takes some time to centralize the information and to make it available. For this reason, most of the countries had no data available for 2018, leading to the exclusion of this year.

From the support indicators, that will help to contextualize the visualization and help the viewer to understand some subjacent motives for leaving/entering a specific country, we chose to display data on the number of deaths due to conflicts and terrorism, GDP per capita, political stability and health spending per capita. The number of deaths due to conflicts and terrorism gives an idea of the country safety, the GDP per capita is the most commonly used economic indicator, the political stability index measures the likelihood that a government will be destabilized or overthrown by unconstitutional or violent means, including politically-motivated violence and terrorism (The Global Economy, 2020b) and, finally, the health spending per capita (including estimates on healthcare goods and services consumed during each year) provides an idea of the country quality of life (The Global Economy, 2020a).

In terms of the attribute types, most of the attributes used in this work are quantitative being the exception the country names, which are categorical.

3. DATA VISUALIZATION

3.1. Visualization and interaction

The theme chosen was inspired by a topic that we have explored in the previous assignment, by presenting a paper related to gulls' migration. Having addressed a similar subject, made us realize that the study of movement data is recent and is showing rapid innovations and evolution in visualizations. Additionally, our work was also inspired on the Open Migration Dashboard (AK Foundation, 2017). This dashboard presents a visualization on migrations on the

Mediterranean area which are presented to the viewer through charts and maps, complemented with some context information, key numbers and some statistics that help the viewer to follow the story. In this context, we have tried to accomplish our objectives by combining both interactive maps and some context data that helps the reader understand the subject under study.

Bearing in mind the importance of interactivity on users' engagement and information understanding, we provide the viewer with three types of interaction in our visualization. Firstly, we give the possibility to the user to choose which variable to display in the choropleth map using radio items. The user may choose among three variables: i) net-migration, ii) inflow and iii) outflow. Such choice allows the user to better perceive which are the main movements in each country while having a global perspective of worldwide migration flows. Besides choosing the variable to display the viewer may also interactively view the migration flows evolution between 2008 and 2017.

After giving the viewer a global perspective, our understanding was that the user would want to have a detailed overview of some specific country or year. Therefore, we chose bar and line graphs, along with text boxes with key numbers to display context information and data comparison on a specific country and year. The user is given the possibility to choose for which country and year should the information be displayed. Overall, Ben Scheiderman's influential mantra of *overview first, zoom and filter, details on demand* was followed in the development of our project (Hamp Data Visualization, 2016).

3.2. Reading the visualization

To guide our line of thought, we have structured this visualization project as presented in Figure 1. The structure was thought to lead the path of the user, from left to right, top to bottom, starting in the first box with a brief introduction to the migration issue as well as the main types/causes of migration. The context is presented through text, and it is important to link all the indicators presented in the app. For the frame color of the visualization, we chose a very neutral color, grey, so the visualizations pop to the user, leading to a clean presentation. Regarding the text itself, we made it blend with our background, to keep the focus on the choropleth graph, the most important visualization in the first zone of visualization and the starting point that directs the users to the following visualizations.

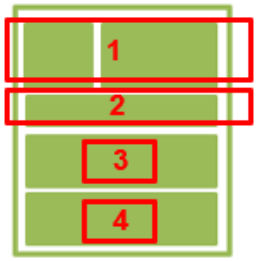

Figure 1. Structure of the visualization

Our data is spatial in its essence, since we are dealing with data for each country. Therefore, our preference was to display the global overview in a map. The choropleth map (Figure 2) was chosen. Since the core of our visualization are the migration flows (inflow, outflow and net-migration) we decided to give the users the possibility to display each of the three variables in the choropleth. The selection options were implemented by using radio components right above the choropleth to be interpreted as part of the choropleth itself. For the colors we have chosen a gradient of color from low to high in a greenish range of colors.

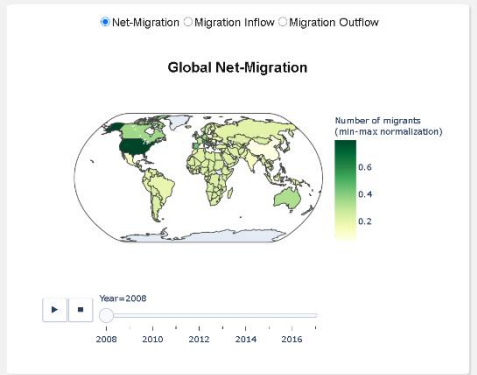

Figure 2. Choropleth map with interactivity

To divide our visualization, we created an intermediary box (Figure 3) between the first and the third and fourth lines of visualization. In this box we included the selections that will define all the graphs below. For the selections we have two variables, the country selection through a simple dropdown and a year selection through a slider. Moreover, we thought it would be helpful to the users to have some context information on the chosen country and year, so in the right side we inserted some boxes with key numbers. These numbers reflect the total number of inflows in the country, the total number of outflows and the main migration flow for that selection.

Figure 3. Interactive items and some indicators for the chosen country and year

In the third part (Figure 4), we start by displaying the top 10 countries in terms of outflow and inflow, i.e. which are the main countries from which people migrate to the chosen country and for which countries do people from the chosen country migrate to. Different colors were chosen as a preattentive feature, allowing to clearly distinguish that the two charts display opposing information. This divergence is also put on focus by the usage of opposing directions in the bars. In the outflow, the direction is from right to left, to show the negative aspect (people leaving the country), and the red color reinforces the message; on the second bar the opposite direction was applied as well as the use of green color to transmit the positive aspect. Our goal was that the viewer can visually perceive the net value of this combination. To complement this visual we have added a graph line comparing, for the country selected the same indicators, but displaying the evolution over the previous 4 years (considering the selected year).

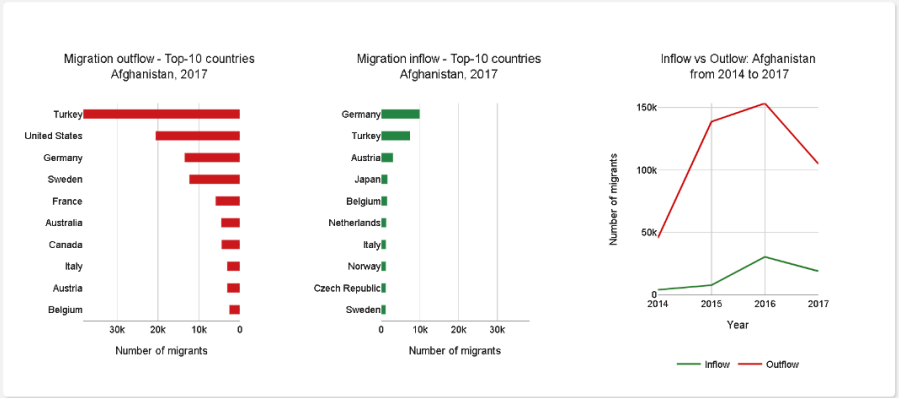

Figure 4. Detailed country perspective on migration patterns

Finally, in the fourth line of visualization (Figure 5), we have decided to give the viewer some key indicators about the selected country, based on the categories presented in the introduction. These indicators provide further understanding on the migration motives.

The chosen indicators (deaths due to conflicts and terrorism, GDP per capita, political stability and health spending per capita) are displayed by using bar graphs, complemented with a trend line corresponding to the global annual average (all countries average for that year). This line provides the user with a baseline case for comparison, allowing the viewer to compare the selected country with the world average. The chosen color is in line with the color palette chosen for the overall visualization, to keep the consistency in the visualization. Similarly, to the line graph, these bar graphs are also affected the year slider presenting a range of 4 years, from the selected year backwards.

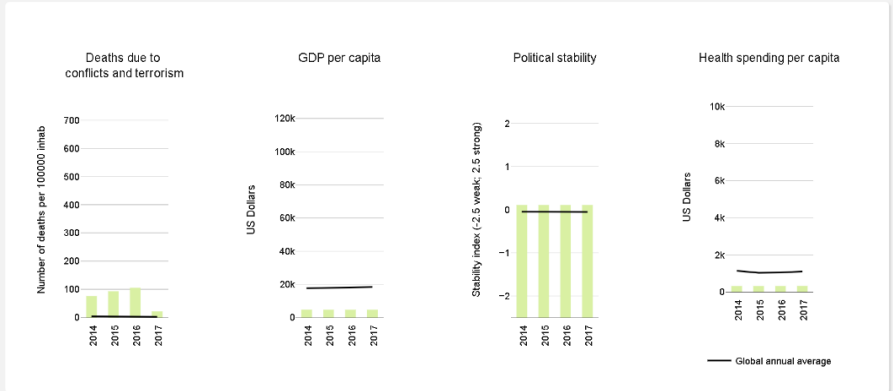

Figure 5. Political, economic and social indicators

3.3. Technical aspects

Regarding the technical aspects, we have followed the requirements and procedures adopted in the practical classes, by using Plotly Dash and PyCharm for code development, for debugging and to have an overview of the final visualization. After code development, the project folder was placed in GitHub and deployed by using the Heroku app. The project folder includes the complete code (in the app.py file), the CSS file with some formatting configurations, the requirements file which contains the packages needed to run the code and the dataset files (two xlsx files). Notice that due to some Plotly bug, the last version of Plotly does not run the code for the choropleth map. Therefore, in the requirements file we have set the Plotly version equal to 3.6.0. Moreover, we have also included in the project folder a print screen of the app final appearance to assure the layout is similar to what would be expectable.

The code building is divided in four main stages: we started by reading the data from the xlsx files, the second step regards the data pre-processing stage, afterwards we have defined the interactive elements and the app layout, finally we defined the callbacks and respective text boxes and plots (bar and line charts), all of which change based on the user's inputs.

In the data pre-processing stage, we have normalized the variables that feed the choropleth graph. Some countries have much higher migration flows than others what would lead to some countries dominating over the remaining. Moreover, since the net-migration may be positive, null or negative the usage of the log scale was not possible. Therefore, we selected the min-max normalization, which is commonly used in data mining projects. This normalization converts the values in a 0 to 1 scale. This allowed to have a better color discrimination in the output.

After the data pre-processing, we defined our interactive elements, we have defined a Country dropdown (dcc.Dropdown), a year slider(dcc.Slider) and a radio selection for the choropleth(dcc.RadioItems). The next step was to define the app layout, using html.Divs. In addition to the use of Plotly built-in options for styling we have also applied a set of css styles. The css components allow us to apply styling in a systematic and easier manner. The last step was to build the app, by defining the callbacks and functions that will provide the dynamic functionality to our plots creating the core of the visualization project.

> **Code available at:** https://github.com/MartaFaria/DV_Project_Migration

4. Discussion

4.1. Limitations and future work

The ability to perceive the countries of interest in the choropleth and selecting them to better understand the numbers presented either in the outflow or in the inflow, worked as planned, with the graphs provided we can have an idea of the situation of each individual country evaluating the story of migrations in and out through the years as well as the growth or decrease in the indicators. The drill-down possibility in the indicators by year also allows us to perceive some patterns in data. A step that we had some limitation was the connection between the choropleth and the selection. In future work we would like to have the possibility to link the countries in the choropleth to the remaining graphs so when we select a country in the map it changes automatically the information bellow, avoiding the need for the country selection in the dropdown. Another improvement would be the reverse if a year bar was selected it should show the country so we can a have a greater visual impact.

Also, we would like to proceed to a general validation from random users to help us understand the visual journey they take when looking to our visualization and give us a better understanding of correctness in our display, giving us the opportunity to make some adaptions on it.

In the validation process, we would like to have some feedback on the difficulty of understanding and reading the report, for the normal user as well as the more analytical user used to reading indicators.

In this work, to assure the coherence and completeness of data, we have only selected data from the OCDE list of countries. This means that not all migration movements were included in our visualization. Therefore, even if we acknowledge the difficulty in accessing a detailed and complete migration dataset, for future analysis we would like to have a more comprehensive data set, having information about all the countries worldwide, for the timeframe that we showed.

REFERENCES

AK Foundation. (2017). Dashboard ⁄ Open Migration. Retrieved June 14, 2020, from https://openmigration.org/en/dashboard/

BBC. (2020). Why do people migrate? - Migration trends - GCSE Geography Revision - BBC Bitesize. Retrieved June 10, 2020, from https://www.bbc.co.uk/bitesize/guides/z8x6wxs/revision/2

Hamp Data Visualization. (2016). Schneiderman's Mantra | Data Visualization. Retrieved June 14, 2020, from https://hampdatavisualization.wordpress.com/2016/02/26/schneidermans-mantra/

The Global Economy. (2020a). Health spending per capita by country, around the world | TheGlobalEconomy.com. Retrieved June 14, 2020, from https://www.theglobaleconomy.com/rankings/health_spending_per_capita/

The Global Economy. (2020b). Political stability by country, around the world | TheGlobalEconomy.com. Retrieved June 14, 2020, from https://www.theglobaleconomy.com/rankings/wb_political_stability/

United Nations. (2020). Migration | United Nations. Retrieved June 10, 2020, from https://www.un.org/en/sections/issues-depth/migration/index.html