

VDP Report:

Migration Flux throughout the years in Portugal
(From 2008 to 2021)

Master's degree in Data Science & Engineering – FEUP

Visualization and Data Preparation

November 2022

Authors: Cátia Teixeira, Sónia Ferreira, Vasco Bartolomeu



Introduction

*"History in its broadest aspect is a record of man's migrations from one environment to another" -
Ellsworth Huntington*

Starting from the first humans to spread from Africa, humans have been roaming from one place to another all over the globe. It is the process that allowed human species to be so ubiquitous at the present moment and may as well be the reason the species survived at all throughout periods of both selective and environmental evolutionary pressures (Lewontin, 1982).

Portugal is well known for having a tradition of producing, a large number, of emigrants. It is also true that for some decades already, there is also an increase in immigration to this country (Góis, P. and Marques, J.C., 2009).

In fact, in 1999 the number of immigrants residing in Portugal represented less than 2 per cent of the total resident population. These were overwhelmingly from Portuguese-speaking countries (either Brazil or PALOP). This scenario changed drastically with sudden migratory wave from Eastern Europe around 2001, particularly from Ukraine. Within a few years, the foreign resident population rose from 208 198 to 460 293 people (approximately 5 per cent of total population in 2005). This period marked a significant feminization and a diversification in countries of origin (Fassmann, H., 2009).

As it is visible, the social and political significance of migration has increased. Migration flows and dynamics have become more mixed in an interconnected worldwide. Therefore, this work aims to study migration fluxes in Portugal from the period of 2008 to 2021. The objective is to access trends, patterns and a better understand how the migration flux may have changed through this timeframe in Portugal.

This study was prepared with the support of data visualization, a powerful tool that allows people to analyze and examine large data sets, which would otherwise be difficult to understand.

Presently, data visualization is used across all industries, to increase sales, promote trends, in a journalistic piece as part of the communication process, etc. Humans need data visualization, because the human brain is not equipped to consume so much raw, unorganized information and turn it in something usable and understandable quickly. It's here that graphs, and charts come into play to communicate data findings, so humans can identify patterns and relations between data being analyzed and highlight interesting details to gain insights and make better decisions faster (Munzner, T., 2014).

Given, the information above, the main goal of this project is to answer these two questions:

- How did the migration flux in Portugal evolved throughout the years between 2008 and 2021?
- How is this flux characterized throughout this timeframe?
 - How is migration distributed between age groups?
 - How is immigration distributed between different origin countries?

Methodology

For the data preparation and data visualization, R language was used through RStudio software.

Data Source

This study used three datasets about the topic migration, collected from the website "[PORDATA](#): dados estatísticos sobre Portugal e europa". The below table provides, a brief description of the available data in each dataset.

Table 1 - Dataset's description

Dataset	Dataset Description
Permanent emigrants per age group¹	Gives the information per year (from 2008 to 2021) and age group ¹ , for each person (national or foreign) who, in the reference period, having stayed in the country for a continuous period of at least one year, left it with the intention of residing in another country for a continuous period of at least one year or more. (metadata - INE)
Dataset	Dataset Description
Permanent immigrants per age group¹	Gives the information per year (from 2008 to 2021) and age group ¹ , for each person (national or foreign) who, having resided abroad for a continuous period equal to or greater than one year, in the reference period, entered the country with the intention of staying here for a period of at least one year or more (metadata - INE).
Nationalities of legal immigrants in Portugal	Information per year (from 2008 to 2009) and main origins of the foreign population with legal residence status, by total. This refers to people of non-Portuguese nationality, that have authorization or residence card, in accordance with the legislation on foreigners in force. It does not include foreigners with a regular situation under the terms of the granting of residence permits, short-stay visas, study, work or temporary visas, as well as foreigners with an irregular situation (metadata - INE).

¹ Age range, in years, in which the individual fits, according to the reference moment. (metadata - INE)

Pre-processing steps

Since the data was raw and not organized in a way that would allow us to create immediately the data visualizations that we needed, the steps described below were taken as data pre-processing after the datasets were uploaded.

Datasets “Permanent emigrants per age group” and “Permanent immigrants per age group”

- Step 1: columns name and type weren't suitable to explore the dataset, due to the symbol in the name (e.g. “15-19”) and the fact the type of almost all columns being char.

Therefore, columns needed to be renamed and the type of all columns was changed to numeric;

The data inside the dataset wasn't fitting for the needed operations, due to the exiting spaces;

Therefore, the spaces on the data were also removed;

- Step2: in order to use these two datasets in a combined plot, a merged dataset of the two original datasets was created – total. For this data frame, a factor variable named “type” was also created in order to allow the separation of the two types of data stored (emigration and immigration);
- Step3: In order to create a new categorical variable, for the datasets “emigration” and “immigration” it was removed the column “type”;

After the differences between the two datasets were calculated into “diffs”, this was later used on the first plot to represent the (positive/negative migration balance).

Heatmaps:

Additionally to the previous steps, the data was reshaped for the two heatmaps accordingly. For the balance heatmap the data was also normalized using min-max normalization between -1 to 1.

Dataset “Nationalities of legal immigrants in Portugal”

For this dataset the data was transformed from wide data into long data.

Results

Plot 1 - Emigration and immigration throughout the years in Portugal

In this first graph we wanted to show how the total number of emigrants and immigrants were distributed throughout the years. In order to do this, it was chosen to use a scatter plot connecting the distributions with lines, facilitating the user to follow its behavior. After plotting these two distributions it was necessary to show the flux migration direction, so the solution was to add bar plots with the difference between immigration and emigration. This way it is possible to show how the migration flux evolved and the behavior of the variation in the number of migrators in Portugal.

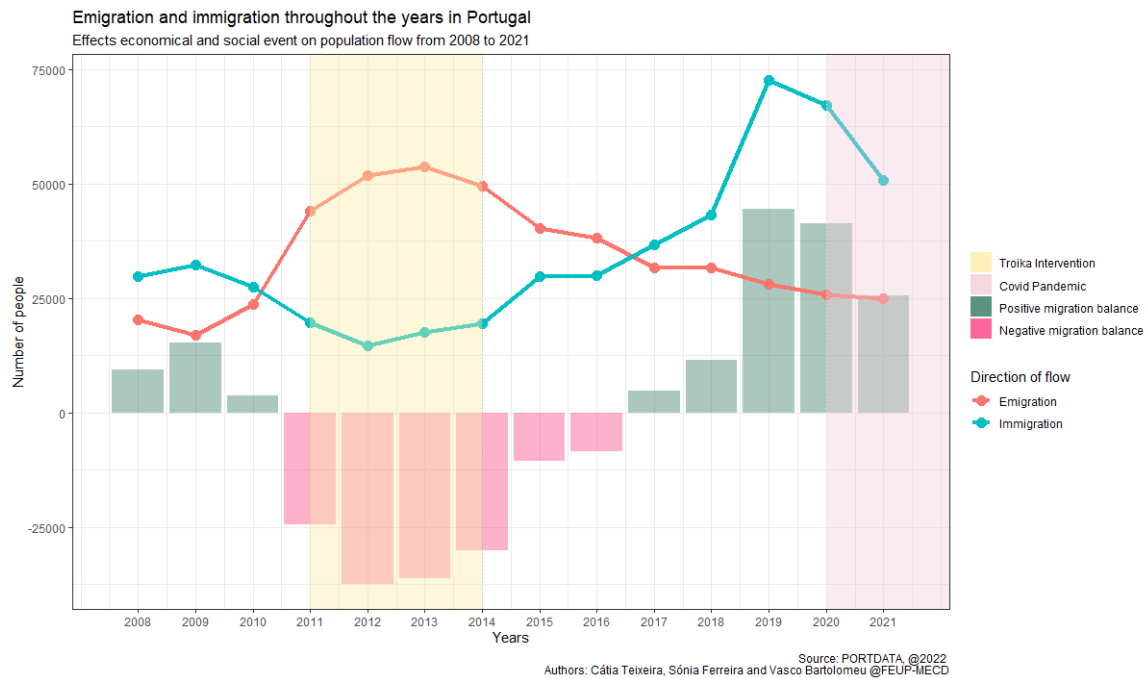


Figure 1 - Emigration and immigration visualization

On a first glance at this plot, one of the things that strikes immediately, is the rising peak on emigration around the years that Troika intervened in Portugal and the decrease on immigration on that same period. From 2011 to 2016, Portugal could be considered as an emigration country, since more people emigrated from Portugal than those who immigrated to Portugal. But this trend changed after 2017 where a positive migration balance is observed again.

During the period of 2011 to 2016, Portugal was suffering with an economic and financial crisis, therefore, this scenario is not the most attractive for immigration. The lack of labor opportunities during this period led to a massive emigration for the pursuit of new and better opportunities, leading to a negative migration balance.

Since 2017, Portugal has again a positive migration balance, due to decrease of permanent outflow emigrants and increase inflow of immigrants.

Due to the covid-19 pandemic, in 2020 the migration balance decreased, this is led to a slight decrease in the permanent entries into the country (immigrations) and the permanent departures (emigrations)

In this first graph we wanted to show how the total number of emigrants and immigrants were distributed throughout the years. In order to do that we decided to go for a scatter plot connecting the distributions with lines to facilitate the user to follow its behavior. After plotting these two distributions we also wanted to show the flux migration direction, so our solution was to add bar plots with the difference between immigration and emigration. This way we are able to show how the migration flux evolved and the behavior of the variation in the number of migrators in Portugal.

Plot 2 & Plot 3 – Migration flow and Migration balance of Portugal throughout the years, respectively

A heatmap allows a person to spot a trend or issues with data at a glance, thanks to the color-coded nature, since our goal was to spot trends over the years per age group² and not the actual numerical values, this type of chart was considered the most appropriated for this analysis.

For plot “Migration Flow”, we considered that placing the two heatmaps side-by-side would be easier to see the differences between the distributions. We opted to use more saturated colors to define higher values and less saturated to lower values.

For plot “Migration Balance of Portugal through the years” we used a color palette that represents negative differences with colder colors and positive differences with warmer colors. We were also careful using a palette that represented the zero as white so it would be easier for the user to distinguish the values.

In terms of fonts and other details chosen to encode this plot, we try to opt for semi-formal details, adapted to the type of subject that we are studying. As for colors, the palette approaches the “powerful color scheme”. With two different types of scales and two different type of data to represent, it was necessary to choose two different palettes.

Migration flow of Portugal throughout the years

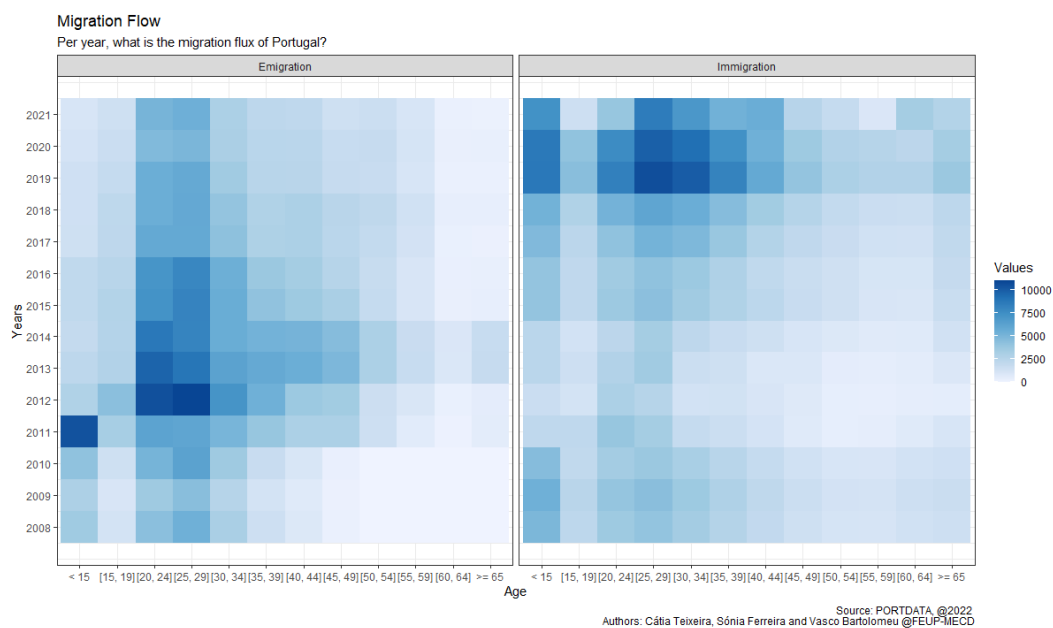


Figure 2 - Migration flow by age group visualization

This plot aims to provide an overview per age group², throughout the years in analysis in this study. The goal is to observe in which years there was more emigration and immigration for each age group² and perceive if there is a tendency to migrate more for some of these groups.

On the previous plot it was observed that during the years of 2011 to 2016, there was a significant increase on the emigration flux and the same can be seen here as the plot tends to be more saturated in these years, with people from all age groups² leaving the country.

Throughout the years of this study, it seems that the predominant age of emigration is until the age of 44. However, there is a period after 2011, where suddenly this behavior changes and we can observe that people above 44 years old start to emigrate as well, coinciding with the economic and financial crisis in Portugal.

As for the immigrant's inflow, as previous concluded the biggest peak since 2008 was in the two years before the Covid pandemic (2020), which can be observed as the more saturated area in this plot. In terms of age groups² over the years there is a small predominancy in the groups until the age of 39, however in the years of 2019 and 2020, this seems to have changed and we can observe a big predominancy of immigrants under the age of 44 when compared with the remaining age groups².

Migration Balance of Portugal throughout the years

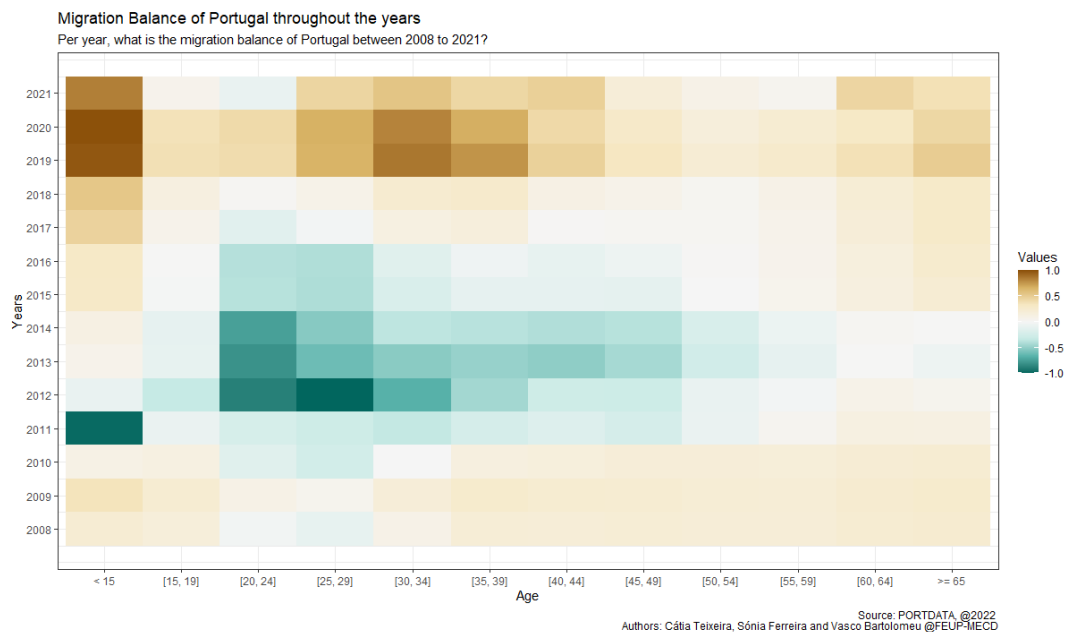


Figure 3 - Migration Balance visualization

² Age range, in years, in which the individual fits, according to the reference moment. (metadata - INE)

This plot aims to support the conclusions reached on the plot “Emigration and Immigration throughout the years in Portugal” with the detail of age per group². As it is visible on the bluer part of this plot there was a negative balance observed during the years of 2011 to 2014 for almost all age groups², this tendency started to revert after 2015. As this process reverted, we can see a browner shade from 2019 with a considerable increase on the emigration balance. This is probably due to the fact that the permanent flow of emigrants slowed down, and even returned through the covid pandemic.

Plot 4 – Nationalities of legal immigrants in Portugal throughout the years

The main purpose of this graphic is to visualize the countries of origin flow for the immigration already previously discussed.

In horizon graphics, the darker color represents higher absolute values, and we can also read the information regarding the trend lines. The data is split into trends using a uniform interval, and the color is based on how far above they are from the zero axis.

The color palette used for this graphic aimed to follow the same scheme as the other visualizations in this project that roughly follows the “powerful color encoding”. Another important factor that played a part in choosing the palette was the good distinguishability between the intervals in order to make the interpretation easier on the viewer, as well as with a good contrast with the theme background.

In this visualization we are then able to compare the trends and volumes of immigration for each country throughout the years. In this way, we can easily compare countries either between themselves or by year.

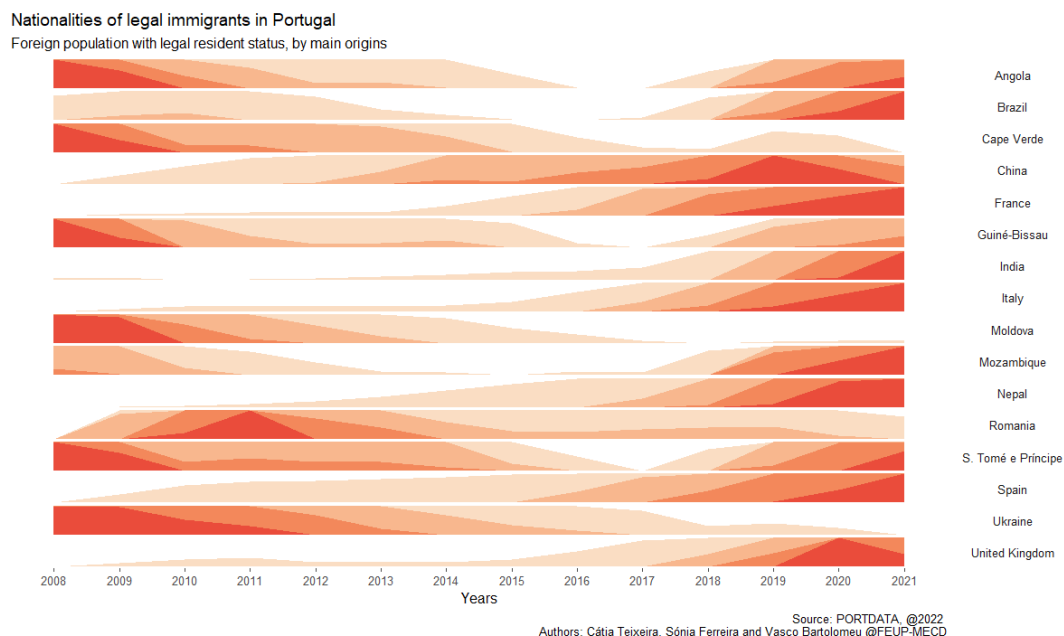


Figure 4 - Nationalities of immigrants visualization

Similarly to the previous graphics, we can clearly see a decrease in immigration trend around the years of Troika intervention (lighter colors positioned in the middle of the x axis). For the majority of origin countries, the trend to increase started again from 2018.

There are some special cases to be noticed. For example, in the case of Ukraine and Moldova it is noticeable the decrease from 2008 onwards in the migratory wave that took place from 2001 from Eastern Europe previously. Countries from America and Africa had a significant decrease during Troika intervention and rose again shortly after. Asian countries began their rise after the Troika intervention had its effect diminished, perhaps due to the effects of “Golden Visas” policy reactivation after a short suspension in 2015.

In 2017 there were changes made to Portuguese immigration law (seasonal work, intra-company transfers, etc.) including tax exceptions and a booming in tech sector. This was probably an influence on the rise of immigration from European countries.

Conclusions

The purpose of this project was to find meaningful and perceptive visualizations to help answer questions in a simple way. After plotting all the graphics, it was clear what each graphic aimed to do and it was easy to extract conclusions from them.

The first graphic was clear to show the influence of social and economic events in the migration flux, such as, the Troika intervention in Portugal, and recently the Covid-19 pandemic. Not only was it able to show how emigration and immigration shifted but also was able to indicate the direction of the flow.

What this graph was missing was the behavior of migration considering the age of the population to answer the second question, so for that purpose it was plotted the migration flux heatmap. This visualization was able to describe clearly how migration flux varies according to different events and how it affects different age groups. During the Troika intervention Portugal had severe measures to respond to the economic crisis that had a huge impact on the labor class, the reason why the peak of emigration was pursued by people in the age class of 20 to 50 years old. Outside this range we see that the flux was really close to zero in this period.

On the other hand, after 2019 the world suffered one of the biggest pandemics yet and it affected economics but more severely the social stability in the society. During this period Portugal was able to respond accordingly to the pandemic and this made it become attractive to immigrants, but this time it affected age groups more evenly.

To support the behavior explained before it was plotted a parallel heatmap to compare the trend on age groups between emigration and immigration. It is clear to see that they behave the exact opposite, having on the emigration a vertical region of interest in the young labor age class (with more effect in the Troika period) and immigration a horizontal region of interest after the Covid-19 pandemic (having affected age classes more evenly as referred before).

As for the horizon graphic, it was possible to dive more deeply into the specific cases of immigration by country (since each country has different pressures that lead to immigration). The general trend of

Troika intervention is once again very clear in the visualization, but it was possible to see each individual case and its fluctuation throughout the years. It was possible to observe a different behavior of immigration happening for each country (being roughly identical between themselves inside each continent) depending on the combination of different socio-economical events on the origin countries and Portugal.

Having this said we can conclude that all plots were able to respond accordingly to the proposed questions and they were able to clearly show trends and share information.

Future work

Looking at the migration fluxes throughout the past years in Portugal, by comparing the totals into the age groups and countries of origin of immigrants, provided a powerful image of the current state of the migration flux.

But, in order to have a better overview of the impact of the migration fluxes in the social and economics of Portugal, it would be interesting to study the trends of the migrations, like for instance:

- Emigrations and immigrations by sex;
- Annual population balances: total, natural and migratory.

Focusing on the analysis data of this study can be a good starting point, but in order to have a better overview and a board understanding of the topic more data should be included.

References

- 1 Lewontin, R. (1982). *Human Diversity*. Scientific American Books - W. H. Freeman & Co.
- 2 Góis, P. and Marques, J.C. (2009), *Portugal as a Semi-peripheral Country in the Global Migration System*. *International Migration*, 47: 21-50. HYPERLINK: <https://doi.org/10.1111/j.1468-2435.2009.00523.x>
- 3 Fassmann, H. (2009). *European migration: Historical overview and statistical problems*. *Statistics and reality. Concepts and measurements of migration in Europe*, 21-44.
- 4 Munzner, T. (2014). *Visualization Analysis and Design* (1st ed.). A K Peters/CRC Press. <https://doi.org/10.1201/b17511>
- 5 PORDATA, *Migrações*. Retrieved October 28th 2022 from <https://www.pordata.pt/>