

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



Table of Contents

[Introduction - 3 -](#_Toc118154376)

[Methodology - 3 -](#_Toc118154377)

[Data Source - 3 -](#_Toc118154378)

[Pre-processing - 3 -](#_Toc118154379)

[Results - 4 -](#_Toc118154380)

[Conclusions - 4 -](#_Toc118154381)

[References - 4 -](#_Toc118154382)

# Introduction

“History in its broadest aspect is a record of man's migrations from one environment to another” - [Ellsworth Huntington](https://www.azquotes.com/author/7083-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 Brasil 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).

# To continue : De preferência usar referências se possível – algo que é valorizado. A seguir devíamos fazer ligação com o que foi dito e com o que fomos estudar neste trabalho em específico. Este ultimo parágrafo faz uma boa relação com o gráfico das nacionalidades. Nota: Ligados à VPN da FEUP temos acesso a artigos científicos

As its visible the social and political significance of migration has increased. Migration flows and dynamics have become more mixed in an interconnected world. Therefore, this work aims to study migration fluxes in Portugal from the period of 2008 to 2021, in order to access trends, patterns and a better understanding how the migration flux may have changed through the years in Portugal.

This study will be 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 in 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

If you torture the data long enough, it will tell you anything.

John W. Tukey

You have to prepare a final report that:

 states the question(s) you aim to answer

 describes the dataset(s) used and briefly explains the pre-processing steps

 provides a justification for your visualization design (type of chart, encoding, colors, fonts,

etc.)

 cannot exceed 10 pages with an 11 pt. font

# Methodology

This chapter describes the methodology applied in the course of this project.

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

## Data Source

This study will use three datasets about the topic migration, collected from the website “[PORDATA](https://www.pordata.pt/): dados estatísticos sobre Portugal e europa”. The below table provides, a briefly description of the available data in each dataset.

|  |  |
| --- | --- |
| Dataset | Dataset Description |
| Permanent emigrants per age group[[1]](#footnote-1) | Gives the information per year (from 2008 to 2021) and age group1, 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) |
| Permanent immigrants per age group1 | Gives the information per year (from 2008 to 2021) and age group1, 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) |

## 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.

* Step2: in order to use these two datasets in a combined plot, a merged dataset of the two original datasets was created (total)

|  |  |
| --- | --- |
| Step’s | Step Descriptions |
| Step 1 | Rename the columns name, to not contain symbols |
| Step2 |  |
|  |  |

# Ultimo gráfico: pivot da tabela

# Results

Plot 1

Chart, line chart

Description automatically generated

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 intervein 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 and therefore was not the most attractive country to immigrate, plus the lack of opportunities during this period led to the decision to emigrate in 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)

provides a justification for your visualization design (type of chart, encoding, colors, fonts,

etc.)

In terms of choosing this type data visualization

Plot 2

A picture containing background pattern

Description automatically generated

This plot aims to provide an overview per age group[[2]](#footnote-2), 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 certain groups.

On the previous plot it was observed that during the years of 2011 to 2016, there was as an exponential increase on the emigration flux and the same can be seen here as the plot tends to be bluer 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 emigrate as well, this coincides 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 bluer 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 remain age groups2.

provides a justification for your visualization design (type of chart, encoding, colors, fonts,

etc.)

A heatmap allows a person to spot a trend or issues with data at a glance thanks to the color-coded natured, since our goal was to spot trends over the years per age group2, this type of chart was the most appropriated. For the plot “Migration Flow” the more bluer the heat map

In terms of choosing this type data visualization

PLOT3

# Conclusions

# Look back and ahead / Future works?

# References

Lewontin, R. (1982). *Human Diversity.* Scientific American Books - - W. H. Freeman & Co.

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" https://doi.org/10.1111/j.1468-2435.2009.00523.x

Fassmann, H. (2009). *European migration: Historical overview and statistical problems*. Statistics and reality. Concepts and measurements of migration in Europe, 21-44.

1. Age range, in years, in which the individual fits, according to the reference moment. (metadata - INE) [↑](#footnote-ref-1)
2. Age range, in years, in which the individual fits, according to the reference moment. (metadata - INE) [↑](#footnote-ref-2)