Using population census data to assist in sampling of survey on discrimination.

Afshin Ashofteh[1,2], Joao Lopes[2], Pedro Campos[2,3]

1 – Nova Information Management School, Nova University of Lisbon.

2 – Statistics Portugal.

3 – Faculty of Economics, University of Porto.

Abstract

The European Union dictates that every person should enjoy their freedom and the chance to participate equally in society, irrespective of their race or ethnicity. To assess the impact of racism or ethnic discrimination, Statistics Portugal implemented the Survey on Living Conditions, Origins and Trajectories of Population Resident in Portugal (ICOT). One of the main challenges in data collection of minority populations is obtaining representativeness. To overcome this limitation, Statistic Portugal used population census data to efficiently target marginal population and ethnic minorities. Accordingly, we propose a methodology to use socio-demographic data from population census in assisting sampling via the creation of groups of parishes with significant representation of the target population. The identification of these groups and their subsequent geospatial analysis provides a framework to use sampling schemes with oversampling. We identified six groups of parishes with marked differences in their characteristics and geographic locations: cluster A with population with low education; cluster B with population of migrants working mostly in the primary sector; cluster C with wealthy, mostly European migrants; cluster D with mostly afrodescendants migrants in an urban context; cluster E with mostly first or second generation of emigrants that returned to Portugal from other European countries; and cluster F with mostly European migrants living in rural areas.

Keywords: Ethnic minorities, Discrimination and racism, Geospatial information, Hierarchical clustering, Coverage error, Population census.

Defining clusters of parishes with ethnic minorities at risk of discrimination.

1. Introduction

The European Union (EU) recommended policy solutions to assess the impact of racism or other form of discrimination by collecting and analysing racial and ethnic origin data (European Commission, 2020). The main challenges in data collection are four-fold: Confidentiality and Anonymity; Representativeness; Comparability; and Categorization (Farkas, 2017). Limitations on representativeness are mostly due to small numbers of national minority members and respondents in statistical samples. To overcome this limitation, national statistical offices can use external data to assist in targeting immigration groups more efficiently (Statistics Canada 2015). A similar approach is used in EU countries (i.e. Croatia, Norway, Switzerland, Romania, Germany, and Turkey) (Morales et al, 2020).

In 2023, Statistics Portugal implemented the Survey on Living Conditions, Origins and Trajectories of Population Resident in Portugal (Inquérito às Origens e Trajetórias da População Residente, ICOT) (Statistics Portugal, 2023). The aim of this survey was to provide information on the diversity of the resident population regarding their origins, ethnic belonging/identification, generational trajectories and living conditions. Survey ICOT addressed several complex and sensitive issues: family, education, working life, languages spoken and learned, housing conditions, mobility, health and limitations, income and social mobility, discrimination, and social and family relations. For this reason, the survey can be of utmost importance in defining public policies. The strategy for data collection followed by Statistic Portugal was similar to the one used by Statistics Canada and by National Statistical Offices of other EU countries, in that, population census data assisted in efficiently targeting marginal population and ethnic minorities (Statistics Portugal, 2022a).

In this work, we are going to characterize all 3092 parishes of Portugal using socio-demographic indicators collected in population census from 2021, amounting to a total of 25 variables divided in 5 themes: Sex-ratio, Naturality, Literacy, Education-level, and Religion. This information will then be used to perform a hierarchical clustering analysis to identify groups of parishes composed by marginal population and ethnic minorities. For this purpose, we devised a methodology to assess the significance of each considered variable according to their distribution on all the parishes. Following the identification of selected clusters, we will enrich the data with geographical information. The identification of these groups and their subsequent geospatial analysis can provide a framework for a sampling scheme used in the context of the survey ICOT.

Defining clusters of parishes with ethnic minorities at risk of discrimination.

2. Data and Methods

We collected six socio-demographic indicators from population census 2021 on five main themes: Sex-ratio - variable C0200 (sex); Naturality - variables C0700 (naturality at parish-level), and C0720 (naturality at country-level); Literacy - variable C1100 (literacy level); Education - variable C1300 (education level); and Religion - variable C3000 (religious faith). We then calculated the percentages of the indicators at parish-level for each of the 3092 parishes in Portugal. This resulted in a database with one observation per parish with a variable for each category of the considered census indicators. Since the aim of the study was to identify groups of parishes with a significant presence of marginal populations vulnerable to discrimination, we chose the final set of variables in accordance. In particular, we omitted information related to Portuguese naturality, to education levels above completion of Primary Education, and to Christian Catholic religion. Thus, obtaining 25 variables belonging to the five main themes:

- Sex-ratio – Proportion of females (Female).
- Naturality – Brazilian (BR); population from Portuguese-speaking African countries (PALOP); population from European Union or Great Britain (E27 + GB); Venezuelan (VE); Eastern Europeans (East-Eur); Swiss (CH); Southern Asians (South-Asian); Chinese (CN); South-African (ZA); North Americans (North-Amer); population from other less represented countries (Other).
- Literacy – Proportion of illiterates (Illiterate).
- Education – Early childhood education (“Nenhum”); Uncompleted primary education (“Básico 1”).
- Religion – Christian Orthodox (Orthodox); Christian Protestant or Evangelical (Protestant/Evangelical); Jehovah’s witness (Jehovah’s witness); other Christian denominations (Other Christian); Buddhists (Buddhist); Hindu (Hindu); Jewish (Jewish); Muslin (Muslin); other religion (Other religion); no religion (Non-religious).

Using the variables considered, we performed a hierarchical clustering analysis to define groups of parishes. We used Euclidean distances and Ward’s minimum variance (Ward, 1963) as the clustering criteria. To select the number of clusters to use, we calculated various indexes using three main criteria to select the optimal number of clusters: minimization criteria – McClain (McClain and Rao 1975), SD index (Halkidi et al. 2000); ‘index > 1’ criteria – Frey (Frey and Van Groenewoud 1972); and maximization criteria – Ptbiserial (Milligan 1980, 1981),

Silhouette (Rousseeuw 1987). After defining the groups, we devised a methodology to characterize them according to the average values of the variables for each group of parishes.

For the characterization of the clusters, first we calculated several thresholds for each variable according to their percentile of the distribution across the 3092 parishes: 1% (P01), 5% (P05), 10% (P10), 90% (P90), 95% (P95) and 99% (P99). Then, we calculated the averages of each variable for each group of parishes and determined their level of significance by considering three levels of negative (left-tail) significance (< P10, -; < P05, - -; < P01, - - -) and three levels of positive (right-tail) significance (> P90, +; > P95, ++; > P99, +++).

To understand the dispersion and location of the groups of parishes, we represented them across the geography of mainland Portugal. For the final definition of the groups of parishes, we took a flexible and interactive approach. We set several possible optimal numbers of clusters according to the different criteria. Then, for each possible level of cut on the hierarchical clustering, we characterize all the defined groups of parishes using both significance of the variables and geospatial analyses. In this way, we could define different depth of cuts depending on the branch of the hierarchical clustering so that the resulting clusters could be more informative.

All analyses were performed in environment R x64 4.3.1. mainly using packages supplied with the distribution (R Core Team 2023). Package "ROracle" was used to retrieve Census 2021 data (Mukhin et al 2021), "tidyverse" was used to perform data analysis (Wickham et al 2019), "NbClust" was used to calculate indexed for the optimal number of clusters (Charrad et al, 2014), “sf” (Pebesma 2018, Pebesma & Bivand 2023) and “scales” (Wickham & Seidel 2022) were used to create the spatial plots, and “dendextend” (Galili 2015) and “RColorBrewer” (Neuwirth 2022) were used to create the hierarchical clustering analysis plots.

3. Results

3.1. Creating clusters of parishes

Using various previously defined indexes, we selected the optimal number of clusters of parishes to consider. Figure 1 shows the selection of three values: 5, 9 and 21 clusters.

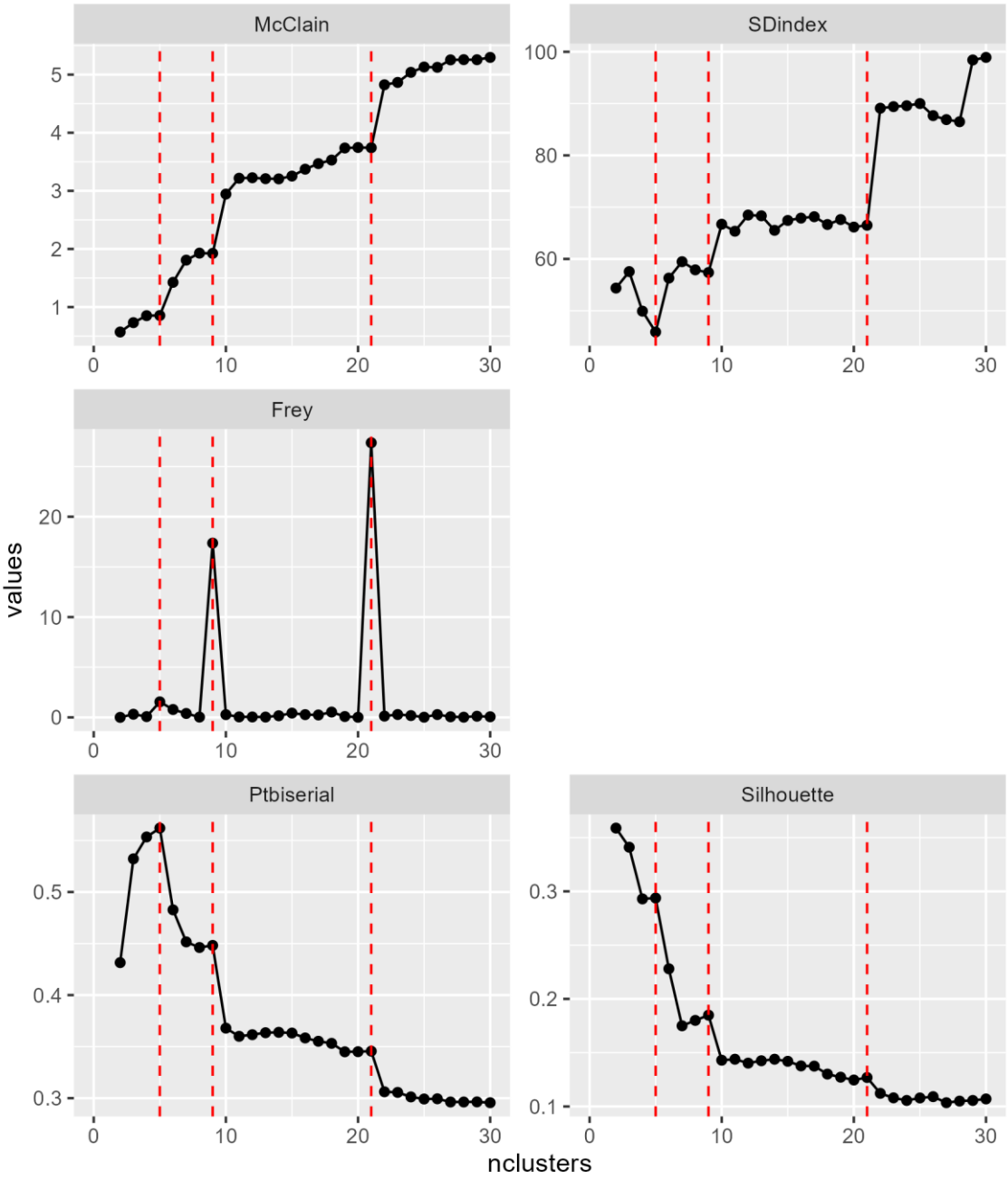

Figure 1. Index values for optimizing the number of clusters using minimization criteria (McClain, SD index), "index greater than 1" criteria (Frey), and maximization criteria (Ptbiserial,

Silhouette) methodologies. Red dashed lines show the number of clusters selected: 5, 9 and 21.

The cut-off for the creation of clusters were considered after the hierarchical clustering analyses (Figure 2). In this figure, we also represented the 6 main cluster of parishes selected after the characterization of all the clusters created (red, purple, light green, dark green, light blue, and dark blue). Notice that the cluster at the leftmost side (red) was considered only at the lowest depth, equivalent to the cut-off on 5 clusters, whereas all the others were considered at the cut-off on 21 clusters.

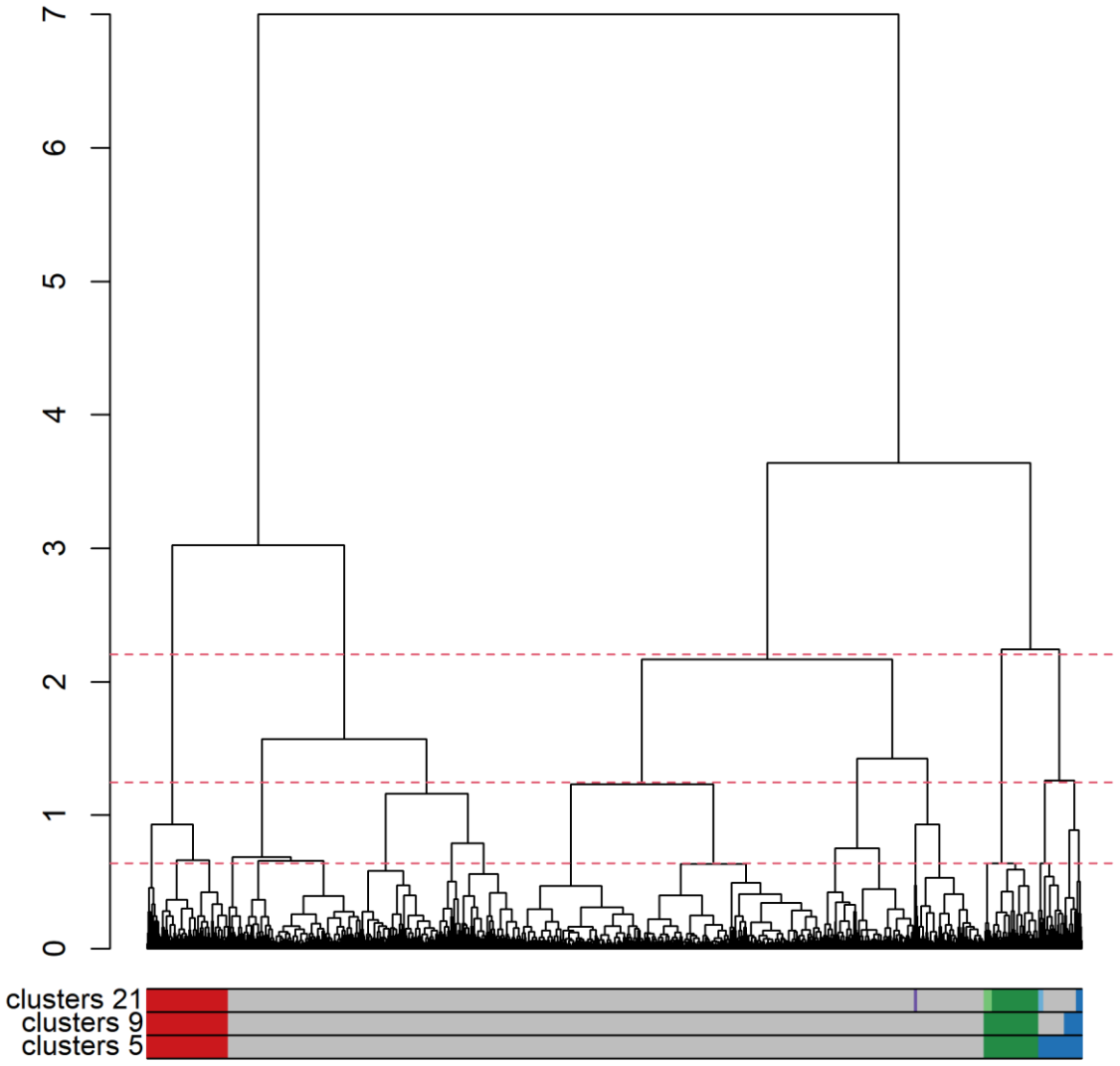

Figure 2. Hierarchical clustering of parishes. Red dashed lines show the cut-offs selected (equivalent to optimal number of clusters): 5, 9 and 21. Coloured horizontal bars show

selected clusters for each of the three cut-offs (A, red; B, purple; C, light green; D, dark green; E, light blue; and F, dark blue). Clusters of parishes not selected are represented in grey.

3.2. Characterization of selected clusters of parishes

Considering the three significance levels for the leftmost tail and three significance levels for the rightmost tail of the distributions, we selected 6 clusters of parishes than can be divided in two main types:

1) Composed of populations with incomplete primary education (i.e. Cluster A);
2) Composed of marginal populations and ethnic minorities at risk of discrimination (i.e. Clusters B, C, D, E, and F).

Table 1. Characterization of selected clusters of parishes.

Variable	**A**	**B**	**C**	**D**	**E**	**F**
n	270	11	26	153	18	16
Female		- - -				-
BR			+	++		
PALOP			+	++		
E27 + GB			++		+++	+++
VE	- -					
East-Eur	- -	++	++			
CH						
South-Asia	-	+++				
CN	- - -				- - -	
ZA	- -				+	
North-Amer						
Other		+++	+			++
Illiterate						
“Nenhum”						
“Básico 1”	++		-	- -		
Orthodox	- -	++	++			+
Protestant/Evangelical			++	+		++
Jehovah’s witness						
Other Christian			++	+	+	++
Buddhist	- -	+++				+
Hindu	- - -	+++				
Jewish						+
Muslin	- -	+++		+		
Other religion	- -	+++				++
Non-religious			+	++		+++

n, number of parishes in cluster; - - -, < P01; - -, < P05, -, < P10; +, > P90; ++, P95; +++, P99.

Defining clusters of parishes with ethnic minorities at risk of discrimination.

We represented the selected six clusters of parishes using a geospatial analysis by considering the geography of mainland Portugal, observing strong geographical structure of the clusters (Figure 3).

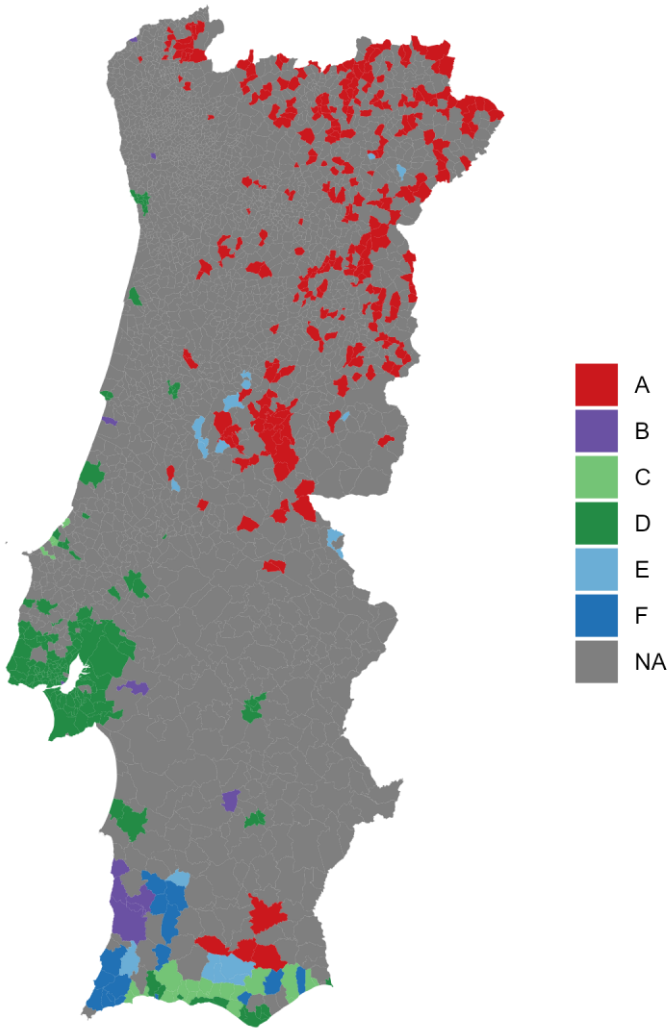

Figure 3. Geographical distribution of selected clusters of parishes in mainland Portugal.

Parishes in Cluster A are mainly located inland in the north and centre of Portugal. Cluster B parishes are located mainly in the regions Metropolitan Area of Lisbon and "Alentejo". Parishes in Cluster C are mainly located in the region of "Algarve" and on the west coast of mainland Portugal. Cluster D parishes are mainly located in the Lisbon Metropolitan Area and in various municipal capitals throughout the country. Parishes in Cluster E are located inland in the regions of "Algarve" and "Baixo Alentejo" and scattered throughout the north and centre

of the country. Finally, Cluster F parishes are located mostly in rural areas in the region of “Algarve” and in the coastal area of “Alentejo”.

Defining clusters of parishes with ethnic minorities at risk of discrimination.

4. Discussion

4.1. Choosing selected clusters of parishes

Using the parishes database, we constructed a hierarchical clustering tree and chose the cut-offs according to various indexes of optimal number of clusters. From visual analyses of the indexes, we select cut-offs corresponding to 5, 9 and 21 clusters. Using three depth levels for the cut-offs allowed us to choose a different cut-off depending on the tree branch considered. The selection of the clusters was done after characterizing each one using the socio-demographic variables. We devised a methodology to assess the significance of each variable used according to their full distribution on the 3092 parishes. This was performed by using commonly used levels of significance for each variable (1%, 5%, 10%, 90%, 95% and 99%). Considering the thresholds of significance for each of the 25 variables, along with the groups of clusters resulting from the three chosen depth levels for the hierarchical tree cut-offs, we selected 6 clusters of parishes. Interestingly, these clusters corresponded to two main types:

1) Composed of populations with high level of illiteracy and incomplete primary education (i.e. Cluster A);
2) Composed of marginal populations and ethnic minorities at risk of discrimination (i.e. Clusters B, C, D, E, and F).

Notice that cluster A was considered only at the lowest depth, equivalent to the cut-off at 5 clusters. The other selected clusters, representing parishes with a high percentage of ethnic minorities, were considered at the deepest cut-off level of the tree, equivalent to 21 clusters. Parishes from non-selected clusters were considered unimportant, having non-significant levels of interesting socio-demographic variables.

4.2. Characterization of selected clusters of parishes

Cluster A represents parishes whose population have a low level of education, potentially increasing the discriminatory pressure on minority ethnic groups (Santos, 2019b). Has seen before, parishes in this cluster are mainly located inland in the north and centre of Portugal, but there are also four inland parishes in the region of “Baixo Alentejo” (Cavaco, 2018).

Cluster B groups parishes characterized by a gender ratio skewed towards males, and with populations from Eastern Europe, Southern Asia, and other less represented countries. The most represented minority religions are Christian Orthodox, Buddhist, Hindu, Muslim, as well as other less represented religions. Parishes are located mostly in “Alentejo” (Odemira and

Ferreira do Alentejo); and in Lisbon Metropolitan Area (Montijo and Santa Maria Maior). The characterization of these parishes follows an expected pattern of regions with large population of male immigrants (Statistics Portugal, 2022b) working in the primary sector. In fact, Odemira, Ferreira do Alentejo and Montijo are regions with strong agriculture/fishing activity with a need for labour force that has been fulfilled by immigrants (Carvalho, 2021; Agostinho, 2022).

Cluster C groups parishes composed by minority populations from Brazil, PALOP countries, EU countries and the UK, and other less represented countries. These parishes are also characterized by having a significant presence of a moderately well-educated population. The most represented minority religions are Christian Orthodox, Christian Protestant/Evangelical and other Christian faiths, as well as people declaring non-religious beliefs. This parishes are mainly located in the region of "Algarve" and the so called "Silver Coast" of Portugal (i.e. Óbidos, Caldas da Rainha and Alcobaça). The characterization of this cluster and its geographic location suggested parishes with high touristic activity targeting wealthy population (Santos & Godinho, 2018). The presence of immigrants from EU countries and the UK, generally perceived as having a good impact in the development of the country, may indicate low levels of discrimination towards migrants in these parishes (Oliveira, 2023).

Cluster D represents parishes composed by minority populations from Brazil and PALOP countries, and with populations with high levels of education. The most represented minority religions are Christian Protestant/Evangelical and other Christian faiths, Muslim, as well as people declaring non-religious beliefs. These parishes are mainly located in the Lisbon and Porto Metropolitan Areas and in various municipal capitals (e.g. Albufeira, Aveiro, Beja, Caldas da Rainha, Coimbra, Faro, Portimão, Santarém and Aveiro). This cluster seems to be related to urban centres with high concentration of migrants. Despite having a population with high levels of education, the composition of the migrant population being mostly from Brazil and PALOP countries creates a high potential for the presence of historical racism towards afrodescentants (Santos, 2019a). These results suggest that parishes in this cluster may be of particular importance for surveying discrimination towards ethnic minorities.

Cluster E is made up of parishes with minority populations from the EU and the UK, and from South Africa. These parishes are also characterized by having significant presence of population of both Christian Orthodox and Christian Protestant/Evangelical faiths. These parishes are located inland in the regions of "Algarve" and "Baixo Alentejo" and scattered throughout the north and centre of the country. The characterization and the geographical

location of these parishes suggest a population of migrants returned from historical emigration outflows to EU countries and UK (Oliveira & Neves, 2017). As in cluster C, in this cluster there seems to be low potential for discrimination towards migrant populations given the similar ethnicity to the Portuguese population.

Cluster F groups parishes characterized by a gender ratio somewhat skewed towards Males and composed by minority populations from the EU and the UK. The most represented minority religions are Christian Orthodox, Christian Protestant/Evangelical and other Christian faiths, Buddhist, Jewish, other less represented religions, as well as people declaring no religious beliefs. These parishes are located in the region of “Algarve” and in the coastal area of “Alentejo”, and include the top ones in terms of proportion of population from EU countries and UK. Notoriously, these are populations that potentially benefit from beneficial fiscal conditions (Santos & Godinho, 2018). From Typology of Urban Areas (Tipologia de áreas urbanas, TIPAU), we observe that, contrary to cluster C, these parishes are mostly rural. Nevertheless, and similarly to cluster C, the general opinion that these migrants have a good impact in the development of the country suggests low relevance for surveying discrimination and racism (Oliveira, 2023). However, the presence of significant proportion of Jewish population may justify considering these parishes for an efficient sampling of this small religious minority in Portugal.

Acknowledgements

Bibliography

Agostinho MAF (2022) O impacto da migração laboral em Odemira-O desafio da integração. Master thesis. Universidade Nova de Lisboa.
URL: http://hdl.handle.net/10362/147067

Carvalho JM (2021) A imigração e a agricultura no Alentejo no século XXI. in Revista Migrações, Observatório das Migrações, dezembro 2021, n.º 17, Lisboa: ACM, pp. 87-104.
URL: https://www.om.acm.gov.pt/-/numero-nao-temati-5

Cavaco C (2018) Analfabetismo em Portugal – os dados estatísticos, as políticas públicas e os analfabetos. Revista Internacional de Educação de Jovens e Adultos, 1(2), 17-31.
DOI: http://dx.doi.org/10.31892/rieja.v1i2.6156

Charrad M, Ghazzali N, Boiteau V, Niknafs A (2014) NbClust: An R Package for Determining the Relevant Number of Clusters in a Data Set. Journal of Statistical Software, 61(6), 1-36.
URL http://www.jstatsoft.org/v61/i06/

European Commission (2020) A union of equality: EU anti-racism action plan 2020–2025. European Commission: Brussels.
URL:
https://ec.europa.eu/info/sites/info/files/a_union_of_equality_eu_action_plan_against_racism_2020_-2025_en.pdf

Farkas L (2017) Data collection in the field of ethnicity. Publications Office of the European Union, Luxembourg.
URL: https://commission.europa.eu/system/files/2021-09/data_collection_in_the_field_of_ethnicity.pdf

Fengyang Z (2016) Os chineses em Portugal: as razões da vinda e a sua situação atual. PhD thesis. Universidade do Minho.
URL: https://hdl.handle.net/1822/44336

Frey T, Van Groenewoud H (1972) A Cluster Analysis of the D-Squared Matrix of White Spruce Stands in Saskatchewan Based on the Maximum-Minimum Principle. Journal of Ecology, 60(3), 873-886.

Defining clusters of parishes with ethnic minorities at risk of discrimination.

DOI: https://doi.org/10.2307/2258571

Galili T (2015) dendextend: an R package for visualizing, adjusting, and comparing trees of hierarchical clustering. Bioinformatics, 31(22), 3718-3720.
DOI: https://10.1093/bioinformatics/btv428

González-Leonardo M, Newsham N, and Rowe F (2023) Understanding Population Decline Trajectories in Spain Using Sequence Analysis. Geographical Analysis, 55(4), 495-516.
DOI: https://doi.org/10.1111/gean.12357

Halkidi M, Vazirgiannis M, Batistakis I (2000). Quality Scheme Assessment in the Clustering Process. In Principles of Data Mining and Knowledge Discovery, volume 1910 of Lecture Notes in Computer Science, pp. 265–276. Springer-Verlag, Berlin Heidelberg. Proceedings of the 4th European Conference, PKDD 2000, Lyon, France, September 13–16 2000.
DOI: https://doi.org/10.1007/3-540-45372-5_26

McClain JO, Rao VR (1975). CLUSTISZ: A Program to Test for The Quality of Clustering of a Set of Objects. Journal of Marketing Research, 12(4), 456–460.
URL: http://www.jstor.org/stable/3151097.

Milligan GW (1980). An Examination of the Effect of Six Types of Error Perturbation on Fifteen Clustering Algorithms. Psychometrika, 45(3), 325–342.
DOI: https://doi.org/10.1007/BF02293907

Milligan GW (1981). A Monte Carlo Study of Thirty Internal Criterion Measures for Cluster Analysis. Psychometrika, 46(2), 187–199.
DOI: https://doi.org/10.1007/BF02293899

Monteiro R (2023), “Diversidade religiosa em destinos de imigração”, Boletim Estatístico OM N.º 8, Coleção Imigração em Números (coordenação de Catarina Reis Oliveira), Observatório das Migrações.
URL: http://www.om.acm.gov.pt/publicacoes-om/colecao-imigracao-em-numeros/boletins-estatisticos

Defining clusters of parishes with ethnic minorities at risk of discrimination.

Morales L, Saji AK, Prandner D, Bergh J, Bernat A et al (2020) Surveys to Ethnic and Migrant Minorities across Europe: Identifying Knowledge Strengths and Gaps Using Survey Metadata. [Research Report] Centre d'études européennes et de politique comparée. 2020, pp.75. hal-03123373.
URL: https://sciencespo.hal.science/hal-03123373/

Mukhin D, James DA, Luciani J (2021). ROracle: OCI Based Oracle Database Interface for R. R package version 1.3-1.1,
URL: https://CRAN.R-project.org/package=ROracle

Neuwirth E (2022). RColorBrewer: ColorBrewer Palettes. R package version 1.1-3,
URL: https://CRAN.R-project.org/package=RColorBrewer

Ning D (2012) A Comunidade Chinesa em Portugal: acerca de atividades económicas, associativismo, integração e a segunda geração. Master thesis. Universidade do Minho.
URL: https://hdl.handle.net/1822/24071

Oliveira CR (2023). Indicadores de Integração de Imigrantes 2023: Relatório Estatístico Anual. Observatório das Migrações, ACM, IP, 2023.
URL: https://www.om.acm.gov.pt/publicacoes-om/colecao-imigracao-em-numeros/relatorios-anuais

Oliveira C, Neves S (2017) Emigrantes portugueses e seus descendentes no mercado de trabalho europeu. Estudos. Statistics Portugal
URL:
https://www.ine.pt/xportal/xmain?xpid=INE&xpgid=ine_estudos&ESTUDOSest_boui=299950104&ESTUDOSmodo=2

Pebesma E (2018) Simple Features for R: Standardized Support for Spatial Vector Data. The R Journal 10 (1), 439-446.
DOI: https://doi.org/10.32614/RJ-2018-009

Pebesma E, Bivand R (2023). Spatial Data Science: With Applications in R. Chapman and Hall/CRC.
DOI: https://doi.org/10.1201/9780429459016

R Core Team (2023). R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria.
URL: https://www.R-project.org/

Rousseeuw P (1987). Silhouettes: A Graphical Aid to the Interpretation and Validation of Cluster Analysis. Journal of Computational and Applied Mathematics, 20, 53–65.
DOI: https://doi.org/10.1016/0377-0427(87)90125-7

Santos T (2019a), Um retrato social do racismo para com os afrodescendentes em Portugal, in Revista Migrações - Número Temático Afrodescendentes em Portugal, Observatório das Migrações, dezembro 2019, nº 16, Lisboa: ACM, pp.53-77
URL: https://www.om.acm.gov.pt/-/afrodescendentes-em-portugal

Santos T (2019b), "Discriminação em razão da origem racial, étnica ou migratória nos países da União Europeia: perceções e experiências reportadas em inquéritos", Boletim Estatístico OM N.º 5, Coleção Imigração em Números (coordenação de Catarina Reis Oliveira), Observatório das Migrações.
URL: http://www.om.acm.gov.pt/publicacoes-om/colecao-imigracao-em-numeros/boletins-estatisticos

Santos T, Godinho S (2018), Europeus Comunitários em Portugal: Uma Análise Exploratória, Boletim Estatístico OM N.º 4, Coleção Imigração em Números (coordenação de Catarina Reis Oliveira), Observatório das Migrações.
URL: http://www.om.acm.gov.pt/publicacoes-om/colecao-imigracao-em-numeros/boletins-estatisticos

Statistics Canada (2015). Visible Minority of Person. Definitions, Data Sources and Methods.
URL: https://www23.statcan.gc.ca/imdb/p3Var.pl?Function=DEC&Id=45152

Statistics Portugal (2022a) Censos 2021. XVI Recenseamento Geral da População. VI Recenseamento Geral da Habitação : Resultados definitivos. Lisboa : INE.
URL: https://www.ine.pt/xurl/pub/65586079

Statistics Portugal (2022b) Censos 2021. O que nos dizem os Censos sobre a população de nacionalidade estrangeira residente em Portugal. Lisboa : INE.
URL: https://www.ine.pt/xurl/pub/66196836

Statistics Portugal. (2023). Destaque Mais de 1,2 milhões de pessoas já sofreram discriminação em Portugal. Lisboa : INE.
URL:
https://www.ine.pt/xportal/xmain?xpid=INE&xpgid=ine_destaques&DESTAQUESdest_boui=625453018&DESTAQUEStema=5414321&DESTAQUESmodo=2

Ward Jr JH (1963) Hierarchical Grouping to Optimize an Objective Function, Journal of the American Statistical Association, 58, 236–244.
DOI: https://doi.org/10.2307/2282967

Wickham H, Averick M, Bryan J, Chang W, McGowan LD, François R, Grolemund G, Hayes A, Henry L, Hester J, Kuhn M, Pedersen TL, Miller E, Bache SM, Müller K, Ooms J, Robinson D, Seidel DP, Spinu V, Takahashi K, Vaughan D, Wilke C, Woo K, Yutani H (2019). Welcome to the tidyverse. Journal of Open Source Software, 4(43), 1686.
DOI: https://doi.org/10.21105/joss.01686

Wickham H, Seidel D (2022). scales: Scale Functions for Visualization. R package version 1.2.1.
URL: https://CRAN.R-project.org/package=scales