

» Using census data to assist in sampling of survey on ethnic discrimination

Joao Lopes
Afshin Ashofteh

DMSI / ME
 April 2024

0



Aim



» Create groups of parishes with target population

- Use socio-demographic data from population census;
- Group the 3092 parishes of Portugal;
- Select for marginal population and ethnic minorities.

» Assist in survey on ethnic discrimination

- Survey on Living Conditions, Origins and Trajectories of Population Resident in Portugal (ICOT);
- Sampling scheme accounting for selected groups of parishes.

1



Introduction > procedure



» Procedure

1. Collect socio-demographic data from population census
2. Collect geospatial data of parishes
3. Create hierarchical cluster tree;
4. Create geospatial maps of clusters;
5. Select important clusters of parishes.

» Software

- Use R (main packages “ROracle”, “NbClust”, “sf”, “dendextend”).



INSTITUTO NACIONAL DE ESTATÍSTICA
STATISTICS PORTUGAL

Assist survey on ethnic discrimination

April 2024

» 02/15

2



Data > data source



» 37 variables from ‘Censos 21’ (Statistics Portugal)

- Sex-ratio (2 categories);
- Literacy (1 category);
- Education (11 categories);
- Place-of-birth (12 categories);
- Religion (11 categories);

[at individual-level]

» Geospatial data from <http://dados.gov.pt>

- Shapefiles of Portuguese parishes (“Freguesias”).



INSTITUTO NACIONAL DE ESTATÍSTICA
STATISTICS PORTUGAL

Assist survey on ethnic discrimination

April 2024

» 03/15

3



Data > data merging



» Create final dataset

- Percentages for each category of variables at parish-level;
- Selection of categories (25) related to presence of discrimination;
- Merge geographical information.

ID	Female	illiterate	Nenhum		BRA	...	Orthodox	...
10103	52%	7%	7%	...	2%	...	1%	...
10109	52%	5%	6%	...	1%	...	1%	...
10112	51%	6%	7%	...	1%	...	0%	...
10119	51%	7%	6%	...	1%	...	0%	...
10121	52%	7%	7%	...	4%	...	1%	...
...	...	...	...	...	...	...	...	...



INSTITUTO NACIONAL DE ESTATÍSTICA
STATISTICS PORTUGAL

Assist survey on ethnic discrimination

April 2024

» 04/15

4



Methods > HCA



» Hierarchical cluster tree

- Euclidean distance;
- Ward's minimum variance.

» Optimal number of clusters

- Minimization criteria (McClain, SD index);
- 'Index > 1' criteria (Frey);
- Maximization criteria (Point biserial, Silhouette);

[Flexible approach depending on branch of tree]



INSTITUTO NACIONAL DE ESTATÍSTICA
STATISTICS PORTUGAL

Assist survey on ethnic discrimination

April 2024

» 05/15

5



Methods > Significance



» Set significance thresholds for each category

- 3 negative levels ($< P_{10}$, '-'; $< P_{05}$, '- -'; $< P_{01}$, '- - -');
- 3 positive levels ($> P_{90}$, '+'; $< P_{95}$, '++'; $< P_{99}$, '+++').

» Assess significance of clusters for each category

- Average value within parishes from each cluster.

» Select important clusters

- Significance of each category;
- Geospatial location of parishes.



INSTITUTO NACIONAL DE ESTATÍSTICA
STATISTICS PORTUGAL

Assist survey on ethnic discrimination

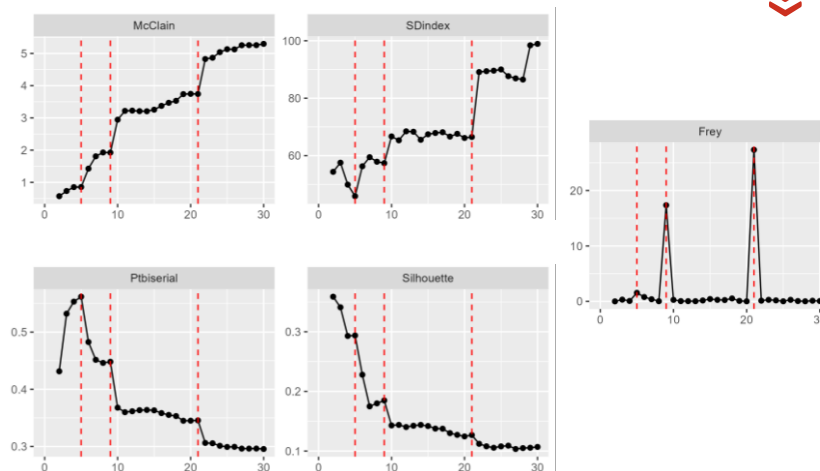
April 2024

» 06/15

6



Results > HCA



INSTITUTO NACIONAL DE ESTATÍSTICA
STATISTICS PORTUGAL

Assist survey on ethnic discrimination

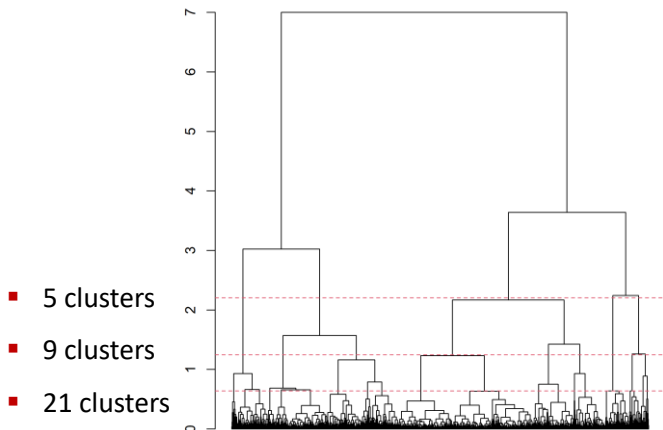
April 2024

» 07/15

7



Results > HCA



INSTITUTO NACIONAL DE ESTATÍSTICA
STATISTICS PORTUGAL

Assist survey on ethnic discrimination

April 2024

» 08/15

8



Results > Significance



» Characterization of 6 selected clusters (A to F)

Variable	A	B	C	D	E	F
n	270	11	26	153	18	16

Variable	A	B	C	D	E	F
Female		- - -				-

Variable	A	B	C	D	E	F
Illiterate						

Variable	A	B	C	D	E	F
“Nenhum”						
“Básico 1”	++		-	--		



INSTITUTO NACIONAL DE ESTATÍSTICA
STATISTICS PORTUGAL

Assist survey on ethnic discrimination

April 2024

» 09/15

9



Results > Significance



» Characterization of 6 selected clusters (A to F)

Variable	A	B	C	D	E	F
BR			+	++		
PALOP			+	++		
E27 + GB			++		+++	+++
VE	--					
East-Eur	--	++	++			
CH						
South-Asia	-	+++				
CN	---				---	
ZA	--				+	
North-Amer						
Other		+++	+			++



INSTITUTO NACIONAL DE ESTATÍSTICA
STATISTICS PORTUGAL

Assist survey on ethnic discrimination

April 2024

» 10/15

10



Results > Significance



» Characterization of 6 selected clusters (A to F)

Variable	A	B	C	D	E	F
Orthodox	--	++	++			+
Protestant			++	+		++
Jehovah's						
Other Christ.			++	+	+	++
Buddhist	--	+++				+
Hindu	---	+++				
Jewish						+
Muslin	--	+++		+		
Other	--	+++				++
No religion			+	++		+++



INSTITUTO NACIONAL DE ESTATÍSTICA
STATISTICS PORTUGAL

Assist survey on ethnic discrimination

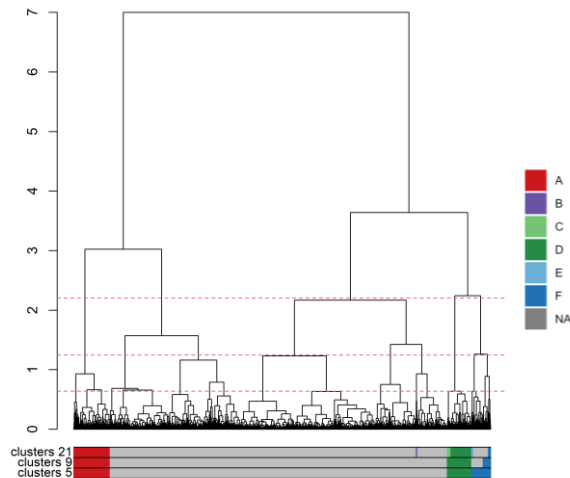
April 2024

» 11/15

11



Results > Final



INSTITUTO NACIONAL DE ESTATÍSTICA
STATISTICS PORTUGAL

Assist survey on ethnic discrimination

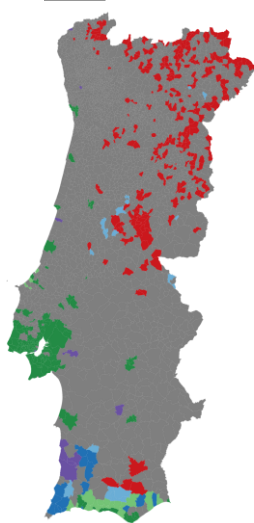
April 2024

» 12/15

12



Results > Final



	Description
A	Low-education Portuguese
B	Southern-Asian in Primary Sector
C	High-class Touristic Areas
D	Afrodescendents in Urban Areas
E	Emigration-related Europeans
F	Europeans in Rural Areas



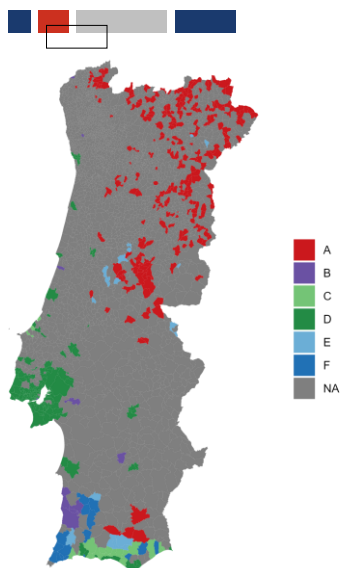
INSTITUTO NACIONAL DE ESTATÍSTICA
STATISTICS PORTUGAL

Assist survey on ethnic discrimination

April 2024

» 13/15

13



Recommendation



» Oversampling priority:

- B and D (164 in 3092);
- **B, D and F (180 in 3092);**
- B, C, D and F (206 in 3092);
- B, C, D, E and F (224 in 3092);
- A, B, C, D, E and F (494 in 3092).



INSTITUTO NACIONAL DE ESTATÍSTICA
STATISTICS PORTUGAL

Assist survey on ethnic discrimination

April 2024

» 14/15

14



Conclusion



- Devised strategy to use **population census data** to assist in efficient targeting of **marginal population** and **ethnic minorities**;
- Considered **3092 parishes** using **25 socio-demographic variables**;
- Used **hierarchical clustering** and **geospatial analysis**;
- Selected 6 potentially important clusters (**494 parishes**);
- Recommend prioritize the use of 3 clusters (**180 parishes**).



INSTITUTO NACIONAL DE ESTATÍSTICA
STATISTICS PORTUGAL

Assist survey on ethnic discrimination

April 2024

» 15/15

15

Thank You



16



Bibliography



Illiteracy and Racism:

Cavaco, 2018, Revista Internacional de Educação de Jovens e Adultos.
Santos, 2019, Boletim Estatístico (Observatório das Migrações).

Municipal Plans for Immigration Integration and Immigrants in Primary Sector:

Pires, 2016, Revista Migrações - Planos de Integração para Migrantes.
Carvalho, 2021, Revista Migrações.
Agostinho, 2022, Master thesis (Universidade Nova de Lisboa).

Migration from EU- countries:

Santos & Godinho, 2018, Boletim Estatístico (Observatório das Migrações).
Oliveira, 2023, Relatório Estatístico Anual (Observatório das Migrações).

Racism towards Afrodescendants:

Santos, 2019, Revista Migrações - Afrodescendentes em Portugal.

Historical Emigration:

Oliveira & Nunes, 2017, Estudos (Statistics Portugal).

17



Motivation



» European Union (EU)

- “A Union of Equality: EU Anti-Racism Action Plan 2020–2025”;
- Assess impact of discrimination by collecting racial and ethnic data.

» Portugal

- “National Plan to Combat Racism and Discrimination 2021-2025”;
- Parliamentary Resolution no. 11/2021 and 16/2021;
- Combat racism and ethnic discrimination in Portugal.



INSTITUTO NACIONAL DE ESTATÍSTICA
STATISTICS PORTUGAL

Assist survey on ethnic discrimination

April 2024

» 18/15

18



Methods > explore data



» Main variables [at parish-level]

- Sex-ratio (1 categories);
- Literacy (1 category);
- Education (2 categories);
- Place-of-birth (11 categories);
- Religion (10 categories)

» Methods

- Calculate summary statistics: n , $\hat{m}$, sd , P_{50} , P_{25} , P_{75} ;
- Create geospatial maps.



INSTITUTO NACIONAL DE ESTATÍSTICA
STATISTICS PORTUGAL

Assist survey on ethnic discrimination

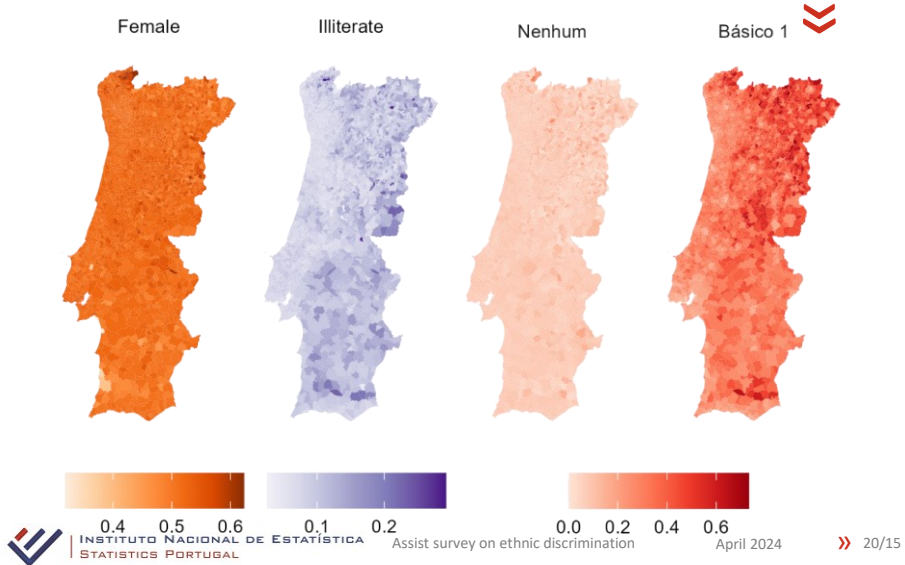
April 2024

» 19/15

19



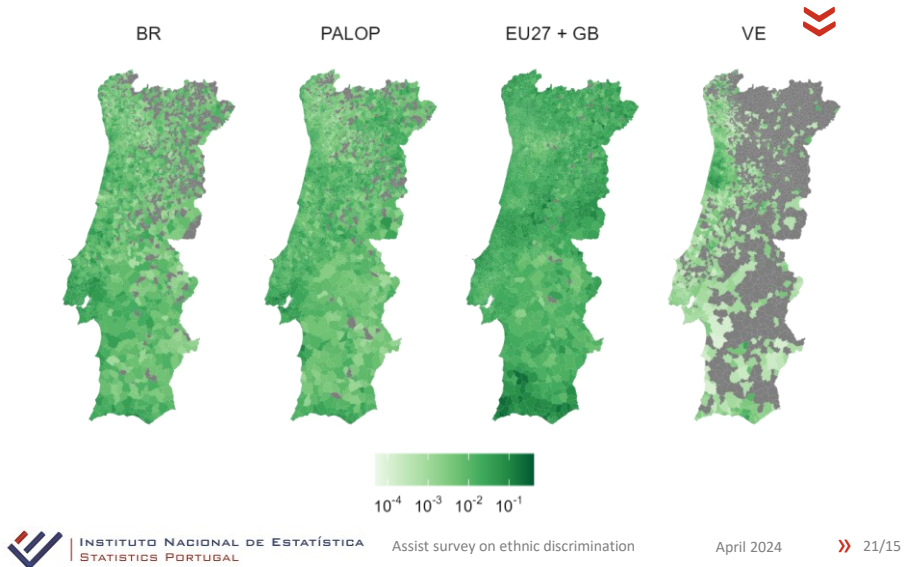
Exploratory results > maps



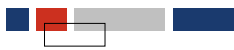
20



Exploratory results > maps



21



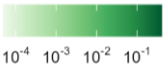
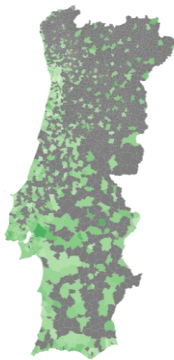
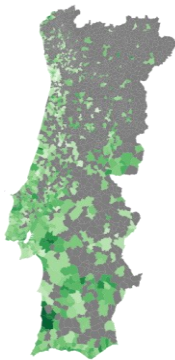
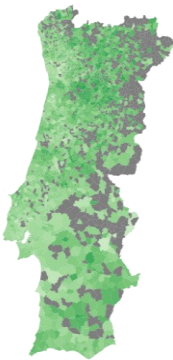
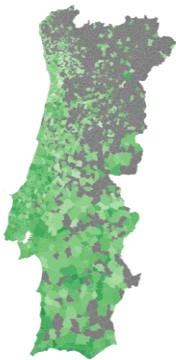
Exploratory results > maps

East-Eur

CH

South-Asia

CN



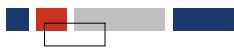
INSTITUTO NACIONAL DE ESTATÍSTICA
STATISTICS PORTUGAL

Assist survey on ethnic discrimination

April 2024

>> 22/15

22

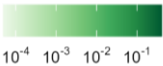
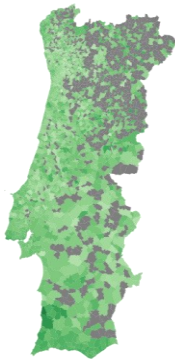
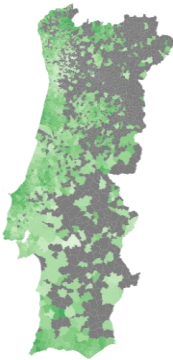
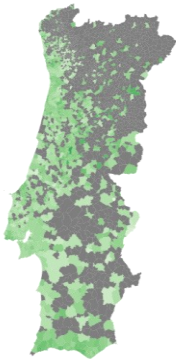


Exploratory results > maps

ZA

North-Amer

Other



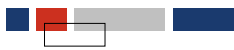
INSTITUTO NACIONAL DE ESTATÍSTICA
STATISTICS PORTUGAL

Assist survey on ethnic discrimination

April 2024

>> 23/15

23



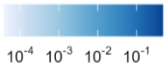
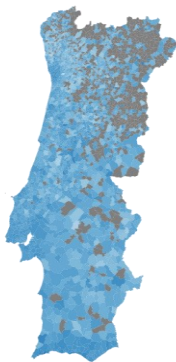
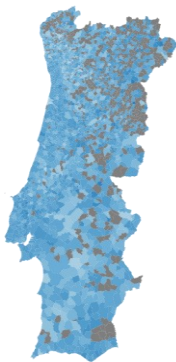
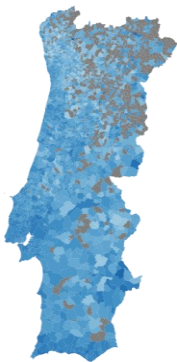
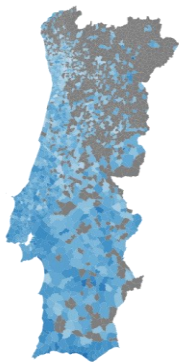
Exploratory results > maps

Orthodox

Protestant/Evangelic

Jehovah's witness

Other Christian



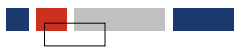
INSTITUTO NACIONAL DE ESTATÍSTICA
STATISTICS PORTUGAL

Assist survey on ethnic discrimination

April 2024

>> 24/15

24



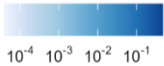
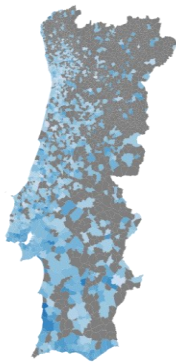
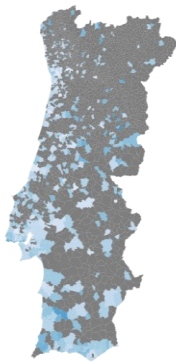
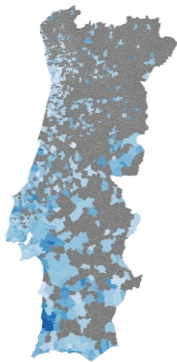
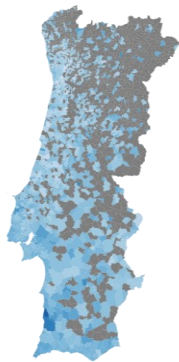
Exploratory results > maps

Buddhist

Hindu

Jewish

Muslim



INSTITUTO NACIONAL DE ESTATÍSTICA
STATISTICS PORTUGAL

Assist survey on ethnic discrimination

April 2024

>> 25/15

25

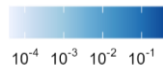
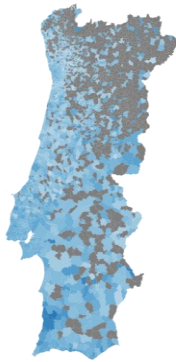


Exploratory results > maps



Other religion

Non-religious



INSTITUTO NACIONAL DE ESTATÍSTICA
STATISTICS PORTUGAL

Assist survey on ethnic discrimination

April 2024

» 26/15

26



Results > Significance



» Significance thresholds

Variable	P01	P05	P10	P90	P95	P99
Female	46.4%	48.9%	49.8%	54.2%	55.1%	57.4%

Variable	P01	P05	P10	P90	P95	P99
Illiterate	4.8%	5.6%	6.1%	12.8%	14.6%	18.9%

Variable	P01	P05	P10	P90	P95	P99
“Nenhum”	1.9%	3.6%	4.7%	10.1%	11.7%	16.2%
“Básico 1”	13.8%	18.5%	21.8%	48.6%	53.4%	61.8%



INSTITUTO NACIONAL DE ESTATÍSTICA
STATISTICS PORTUGAL

Assist survey on ethnic discrimination

April 2024

» 27/15

27



Results > Significance



» Significance thresholds

Variable	P01	P05	P10	P90	P95	P99
BR	0.1%	0.1%	0.2%	2.5%	3.8%	6.6%
PALOP	0.1%	0.2%	0.2%	2.6%	3.5%	8.0%
E27 + GB	0.2%	0.5%	0.7%	4.8%	6.1%	11.6%
VE	0.0%	0.0%	0.0%	1.3%	2.8%	7.9%
East-Eur	0.0%	0.0%	0.1%	1.0%	1.5%	2.9%
CH	0.0%	0.1%	0.1%	1.0%	1.5%	2.8%
South-Asia	0.0%	0.0%	0.0%	1.2%	2.8%	9.9%
CN	0.0%	0.0%	0.0%	0.3%	0.4%	1.0%
ZA	0.0%	0.0%	0.0%	0.4%	0.6%	1.6%
North-Amer	0.0%	0.0%	0.0%	0.8%	1.3%	2.4%
Other	0.0%	0.1%	0.1%	0.8%	1.1%	2.0%



INSTITUTO NACIONAL DE ESTATÍSTICA
STATISTICS PORTUGAL

Assist survey on ethnic discrimination

April 2024

» 28/15

28



Results > Significance



» Significance thresholds

Variable	P01	P05	P10	P90	P95	P99
Orthodox	0.0%	0.1%	0.1%	1.3%	2.0%	4.3%
Protestant	0.1%	0.1%	0.2%	2.9%	3.9%	6.3%
Jehovah's	0.1%	0.1%	0.2%	1.3%	1.6%	2.9%
Other Christ.	0.1%	0.1%	0.2%	1.4%	1.7%	2.5%
Buddhist	0.0%	0.0%	0.0%	0.4%	0.6%	1.3%
Hindu	0.0%	0.0%	0.0%	0.6%	1.4%	3.7%
Jewish	0.0%	0.0%	0.0%	0.2%	0.3%	0.7%
Muslin	0.0%	0.0%	0.0%	0.7%	1.1%	3.1%
Other	0.0%	0.0%	0.1%	0.5%	0.7%	1.7%
No religion	0.6%	1.4%	2.0%	17.3%	22.3%	30.3%



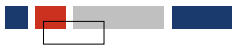
INSTITUTO NACIONAL DE ESTATÍSTICA
STATISTICS PORTUGAL

Assist survey on ethnic discrimination

April 2024

» 29/15

29



	Description
A	<u>Education-level</u> : Very low. <u>Female ratio</u> : Very low.
B	<u>Naturalty</u> : Eastern Europe, Southern Asia, and Other. <u>Religion</u> : Orthodox, Buddhist, Hindu, Muslin, and Other.
C	<u>Naturalty</u> : Brazil, PALOP, EU and UK, and Other. <u>Education-level</u> : Moderately high. <u>Religion</u> : Orthodox, Protestant and Other Christians, and Non-religious.
D	<u>Naturalty</u> : Brazil, PALOP. <u>Education-level</u> : High. <u>Religion</u> : Protestant and Other Christian, Muslin, and Non-religious.
E	<u>Naturalty</u> : EU and UK, and South Africa. <u>Religion</u> : Other Christians. <u>Female ratio</u> : Moderately low.
F	<u>Naturalty</u> : EU and UK, and Other. <u>Religion</u> : Orthodox, Protestant and Other Christian, Buddhist, Jewish, Other, and Non-religious.

