

# Divisive Clustering applied to the Finnish, Portuguese and Spanish Data from the European Social Survey 2002/2003

*Soile Mustjärvi*  
[soile.mustjarvi@uta.fi](mailto:soile.mustjarvi@uta.fi)

## OUTLINE

1. Objectives
2. Descriptions of the variables and symbolic objects
3. Some interesting graphics
4. Divisive classification for modal variables
5. Divisive classification for interval variables

## OBJECTIVES

- To present some interesting results of Symbolic Data Analysis and European Social Survey relating to politics  
([www.europeansocialsurvey.com](http://www.europeansocialsurvey.com))
- To compare results of three different countries and combined data
- Methods: *DIV* (SPCA, SGCA)
- Data: Finland, Spain, Portugal and combined data

### European Social Survey:

Scatter Plot of 20 Participating Countries: trust\_gov = average trust to EU parliament, Own Parliament and Politicians; trust\_leg = average trust to legal system and police. Scale from 0 to 10



## European Social Survey:

Scatter Plot of 20 Participating Countries: trust\_gov = average trust to EU parliament, Own Parliament and Politicians (scale from 0 to 10); polact = average participation to political etc. activities (scale from 0 to 7).



## BACKGROUND VARIABLES

- **COUNTRY (country):**
  - 1 = Finland, 2 = Spain, 3 = Portugal
- **SEX (Gender):**
  - 1 = Male, 2 = Female
- **AGE (Age group):**
  - 1 = → 30, 2 = 31 – 60, 3 = 61 →
- **EMP (Employment status):**
  - 1 = Employed, 2 = Not gainfully employed, 3 = No answer
- **AREA (Residential area):**
  - 1 = Urban, 2 = Rural

## SYMBOLIC VARIABLES

- **MODAL VARIABLES:**
  - divided to three or four groups
- **INTERVAL VARIABLES** (min/max)&(99%/1%):
  - are formed by summing two or three different variables together and dividing these sums by the number of summed variables
  - In the case of “99%/1%”-data the biggest / smallest outliers are replaced with a critical points which are 99 % and 1 %.

## DESIGN-WEIGHT

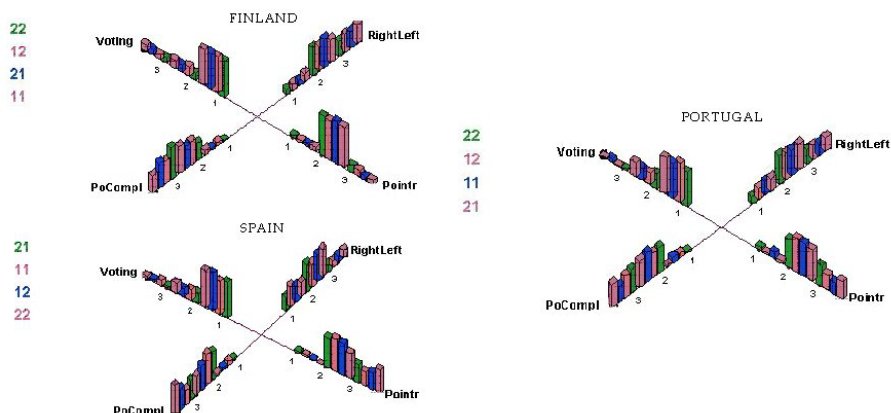
$$w = \frac{1}{\text{PROB1} \cdots \text{PROB}k}$$

- k depends on the number of stages of the sampling design
- The mean of the design weight is 1 for each country

# SYMBOLIC OBJECTS

- **FOR EACH COUNTRY:**
  - DIV: gender, employment status and age ( $2*3*3 = 18$ )
  - GRAPHICS: gender and residential area ( $2*2 = 4$ ).
- **FOR COMBINED DATA:**
  - DIV: country, gender, employment status and age ( $3*2*3*3 = 54$ )

## GRAPHICS: for the whole country on our four main variables relating to politics



## DIV FOR MODAL VARIABLES

CI	FINLAND	SPAIN	PORTUGAL
1	Male, employed, under 30 and Female, employed, under 60.	Female, no answer to employment status, under 30	Female, employed or no answer to employment status, over 60
2	Male/Female, no answer to employment status, under 30	Female, not in paid work, under 30	Male, employed, over 60
3	Male/Female, no answer to employment status, over 60 and Male/Female, not in paid work, 30-60 and Female, not in paid work, under 30	Male, no answer to employment status, over 60 and Female, employed, under 30	Male, not in paid work, over 60
4	Male, not in paid work, under 30	Female, no answer to employment status, over 60 and Male, no answer to employment status, 30-60 and Female, not in paid work, 30-60 and Male, not in paid work, over 60 and Male, employed, over 60	Male/Female, no answer to employment status/not in paid work, under 30
5	Male/Female, no answer to employment status, 30-60 and Male/Female, no answer to employment status/employed, over 60	Female, not in paid work, over 60	Male/Female, employed, under 30/30-60 and Male/Female, no answer to employment status, 30-60 and Male, not in paid work, 30-60
6	Male, employed, 30-60 and Female, not in paid work, over 60	Male, every employment status, under 30	Female, not in paid work, over 60
7	Male, not in paid work, over 60	Male/Female, employed, 30-60 and Female, no answer to employment status, 30-60 and Male, not in paid work, 30-60 and Female, employed, over 60	Male, no answer to employment status, over 60 and Female, not in paid work, 30-60

## DIV FOR MODAL VARIABLES AND COMBINED DATA

CLUSTERS	COMBINED DATA
1	Finnish, Male/Female, employed, under 60.
2	Rest of the objects...
3	Finnish, Male, not in paid work/no answer to employment status, under 30
4	Spanish/Portuguese, Male/Female, all employment status, under 30
5	Portuguese, Male, employed, over 60
6	Spanish/Portuguese, Female, all employment status, over 60 and Portuguese, Male, not in paid work, over 60
7	Finnish, Male/Female, all employment status, all age group

## DIV FOR INTERVAL VARIABLES AND FINNISH DATA

Clust.	SYMBOLIC OBJECTS (99 %/1 %)	SYMBOLIC OBJECTS (min/max)
1	Female, no answer to employment status/employed, under 30 and Male, no answer to employment status, under 30	Male/Female, no answer to employment status, over 30 and Male/Female, employed, 30-60 and Male, no answer to employment status, under 30
2	Male, employed, over 60 and Female, not in paid work, over 60	Male, employed, over 60
3	<b>Female, not in paid work, under 30</b>	<b>Female, not in paid work, under 30</b>
4	Male, not in paid work, over 60 and Female, not in paid work, 30-60 and Female, employed, over 60	Male/Female, employed, under 30 and Male, not in paid work, under 60 and Female, no answer to employment status, under 30
5	Male/Female, no answer to employment status, over 30 and Male, not in paid work, under 60 and Male, employed, under 60 and Female, employed, 30-60	Female, not in paid work, over 30 and Male/Female, employed, over 60

## DIV FOR INTERVAL VARIABLES AND SPANISH DATA

Clust	SYMBOLIC OBJECTS (99 %/1 %)	SYMBOLIC OBJECTS (min/max)
1	Male, no answer to employment status, under 30 and Female, employed, under 30 and Male, not in paid work, under 30	male, not in paid work, under 30/over 60
2	Male, not in paid work, over 60	Male, employed, 30-60 and Male/Female, no answer to employment status, over 30
3	Male, employed, under 60 and Male/Female, no answer to employment status, over 60 and Male/Female, not in paid work, 30-60 and Female, no answer to employment status, under 60 and Female, employed, 30-60 and Male, no answer to employment status, 30-60	<b>Female, employed/not in paid work, over 60 and Female, not in paid work, under 30</b>
4	Male, employed, over 60	Female, employed/not in paid work, 30-60 and Male, employed, over 60
5	<b>Female, Employed/not in paid work, over 60 and Female, not in paid work, under 30</b>	Male/Female, employed/no answer to employment status, under 30 and Female, not in paid work, 30-60


## DIV FOR INTERVAL VARIABLES AND PORTUGUESE DATA

Clust	SYMBOLIC OBJECTS (99 %/1 %)	SYMBOLIC OBJECTS (min/max)
1	Female, employed, over 60 and Female, not in paid work, all age groups	Female, employed, over 60 and Female, not in paid work, over 30 and Male/Female, not in paid work, under 30
2	Rest of the objects...	Female, employed, under 30 and Male/Female, no answer to employment status, over 60 and Male, employed, under 30
3	Male, not in paid work, over 60	Male, not in paid work, 30-60 and Female/Male, no answer to employment status, 30-60 and Male, employed, 30-60 and Female, no answer to employment status, under 30 and Female, employed, under 30 and Male, not in paid work, over 60
4	Male, not in paid work, under 30	Male, employed, over 60
5	Male, employed, over 60	Male, no answer to employment status, under 30

## DIV FOR INTERVAL VARIABLES AND COMBINED DATA

CL	SYMBOLIC OBJECTS (99 %/1 %)	SYMBOLIC OBJECTS (min/max)
1	Finnish/ Spanish/Portuguese, Many different objects	Spanish/Portuguese, employed/not in paid work, all age groups
2	Finnish, Male/Female, employed/not in paid work, over 60 and Spanish, female, not in paid work, over 60/under 30 and Portuguese, male, employed, over 60	Finnish, no answer to employment status, over 30 and Spanish, no answer to employment status, over 30 and Portuguese, employed/ no answer to employment status
3	Spanish, Female, employed, over 60	Finnish, employed/not in paid work and Spanish, female, not in paid work, under 30 or over 60 and Portuguese, male, employed, over 60
4	Finnish, female, not in paid work, under 30	Finnish, employed/not in paid work, all age groups and Spanish, employed/not in paid work, over 30 and Portuguese, Many different objects
5	Finnish/Spanish/Portuguese, Many different objects	Finnish, no answer to employment status, under 60 and Spanish, Many different objects and Portuguese, male, no answer to employment status, under 30





*Thank you!*

