Visualization and Clustering of Population Pyramids

JELENA GUSIĆ
Faculty of Organization and Informatics, University of Zagreb, Varaždin, Croatia
e-mail: jelena.gusic@foi.hr

DUŠAN MUNĐAR
Faculty of Organization and Informatics, University of Zagreb, Varaždin, Croatia
e-mail: dusan.mundjar@foi.hr

DAMIR HORVAT
Faculty of Organization and Informatics, University of Zagreb, Varaždin, Croatia
e-mail: damir.horvat1@foi.hr

Population pyramids are a type of histograms. They present the age distribution of a population, also taking gender into consideration. Two gender histograms are plotted horizontally back-to-back. The number of males/females or the corresponding proportion is presented on the x-axis and age-groups are presented on the y-axis [1].

Our presentation has two goals. The first goal is to demonstrate a possibility for clustering of data suitable for presentation in the form of a histogram, i.e. to find units (or in our case populations) that are similar by structure to each other. Similarity or distance of the items has to be defined by a distance measure. There are wide range of measures that can be used [2]. Among others, we will use the Euclidean distance. The second goal is to demonstrate a possibility for graphical visualization of the population pyramids and the geo-spatial presentation of the clustered data using R programming language [3].

As an example, we use census data from the Croatian Bureau of Statistics for 20 Croatian counties and the City of Zagreb for the years 2001 and 2011 [4]. Visualization and clustering of counties directs attention on change in populations’ age structure and indicates that pattern of population aging is different among counties in Croatia.

Key words: clustering, data visualization, demographics structure, population aging

MSC 2010: 91C20, 62-07, 62-09
Figure 1: Counties in Croatia clustered by population age structure.

Figure 2: Population proportions pyramids - Croatia 2011 - clustered counties.

References


