An increasing number of socio-economic studies are carried out on the basis of households' or individual respondents' questionnaires collected repeatedly over a period of time. The growing relevance of longitudinal data[1] is caused by the widespread use of this data in many research fields. An important example is the data collected as part of the Social Diagnosis[2], Eurobarometer[3], the World Value Survey[4] or the Panel Study of Income Dynamics[5], which are freely accessible. The questions of interest are usually measured on qualitative scales of measurement (nominal or ordinal).

One of the first, but still convincing methods of describing the structures in data is visualization. Graphical data presentation is not only data depiction but it is frequently considered as the basic method of data analysis. This approach is very popular in analyses of bi-dimensional data sets, which can be presented on the surface and submit to the initial visual inspection of the whole data set. The dynamic development of information technologies has contributed to the growing popularity of multivariate statistical methods of data analysis. However, the assumption of many of them is the homogeneity of the data sets of observations. The problem of heterogeneity of the analysed data is very important, especially when responses are observed on each subject repeatedly over time. Then it is often simply assumed (without prior verification) that the analyzed data set is homogeneous and statistical analyses are carried out for each of the periods of time separately [Czapiński 2013, 2015; Szumlicz 2013; Panek et al. 2015].

In light of the above, and based on the results of the critical analysis of literature conducted by the author, the **main objective** of the research project is to investigate the opportunities to use latent variable models in the homogeneous structures identification in longitudinal socio-economic data sets. The nature of the research problem, project' objectives and hypotheses have determined a **methodological approach.** The project methodology assumes integrated approach to the structures recognition in socio-economic longitudinal data sets.

Most of all we will consider the identification of homogenous structures in data set based on the mixture models. In this approach, the number of components in the mixtures is not known, as it is a latent variable of the model. We will focus especially on different variants of latent classes (with covariates and combined with item response theory) and latent Markov models (basic, constrained and extended) which are an important tool for structures in longitudinal data sets recognition. Those model are based on discrete latent variable, the different values of which correspond to different homogenous structures (named latent classes or states) having a common distribution of the response variables. The information criterion AIC [Akaike 1974] [6], BIC [Schwarz 1978] [7], as well as *S* index [Bartolucci et al. 2013, p. 68] [8] will be used for hypothesis testing and models assessment which among others allow the author to compare the results and draw conclusions for future research. All the computation and graphics will be performed in the professional program for statistical computing R, using different libraries as well as our own copyright procedures.

[1] A longitudinal (panel) dataset tracks the same type of information on the same subjects at multiple points in time.
[2] http://www.diagnoza.com/
[3] http://www.gesis.org/eurobarometer-data-service/home/
[4] http://www.worldvaluessurvey.org/WVSContents.jsp
[5] http://psidonline.isr.umich.edu/
[6] Akaike H., 1974, *A new look at statistical model identification*, IEEE Transactions on Automatic Control, 19, 716-723.
[7] Schwarz G., 1978, *Estimating the dimension of a model*, Annals of Statistics, 6,461-464.
[8] Bartolucci F., Farcomeni A., Pennoni F., 2013, *Latent Markov Models for Longitudinal Data*, Chapman and Hall/CRC press, Boca Raton.