



## Visualizing aggregated symbolic data with continuous and categorical variables

Junji Nakano\*

The Institute of Statistical Mathematics, Tokyo, Japan – [nakanoj@ism.ac.jp](mailto:nakanoj@ism.ac.jp)

Nobuo Shimizu

The Institute of Statistical Mathematics, Tokyo, Japan – [nobuo@ism.ac.jp](mailto:nobuo@ism.ac.jp)

Yoshikazu Yamamoto

Tokushima Bunri University, Kagawa, Japan – [yamamoto@fe.bunri-u.ac.jp](mailto:yamamoto@fe.bunri-u.ac.jp)

Traditional symbolic data analysis uses information about marginal distribution of each variable mainly for simplicity. In recent “Big data” era, huge amount of individual data are available and are often divided into some naturally defined groups. Such groups can be summarized by several descriptive statistics calculated by using information about the joint distribution. We call such statistics as “aggregated symbolic data”. We consider that individual data consists of continuous and categorical variables, and use up to second moments related statistics to express each group. We visualize all aggregated symbolic data together by an extended parallel coordinate plot. Several real data examples are illustrated.

**Keywords:** aggregated symbolic data; descriptive statistics; parallel coordinate plot.