

Categorical Data Visualization Reordered

Alexander Pilhöfer^{1,*}, Antony Unwin¹

1. Department of Computer Oriented Statistics and Data Analysis, Institute of Mathematics, University of Augsburg, Germany

*Contact author: alexander.pilhoefer@math.uni-augsburg.de

Keywords: Categorical Data, Reordering, Visualization

The analysis of categorical data can often be challenging, especially when multiple nominal variables are involved. The major problem is the lack of naturally defined distances and orders. Visualisations reveal information in datasets, especially when they are backed up with flexible ordering options.

This talk presents *joint reordering*, an approach for emphasising associations between categorical variables: Each variable is reordered in a way that is optimal given the orderings of the other variables. The optimality criterion used is the *Bertin Classification Criterion* (BCC) [1]. Two algorithms are presented, the general BCC reordering algorithm and a stepwise procedure tailored for the optimization of CPCP plots [2], an extension of Parallel Coordinates for categorical data.

The main example discussed makes use of the big US airport dataset that was the subject of the 2009 JSM Data Expo. Results are visualised using both fluctuation diagrams and CPCP plots.

The reordering and visualisation techniques described are available in the R package **extracat** [2].

References

- [1] Pilhöfer, A., A. Gribov, and A. Unwin (2012). Comparing clusterings using Bertin’s idea. *Visualization and Computer Graphics, IEEE Transactions on* 18(12), 2506–2515.
- [2] Pilhöfer, A. and A. Unwin (2013, 5). New approaches in visualization of categorical data: R package extracat. *Journal of Statistical Software* 53(7), 1–25.