Unsupervised Learning: A time perspective analysis of visitors' behaviors

G. Casolla, V. Schiano, S. Cuomo, F. Giampaolo, F. Piccialli
University of Naples Federico II

Mathematical mOdelling and
Data Analysis research group
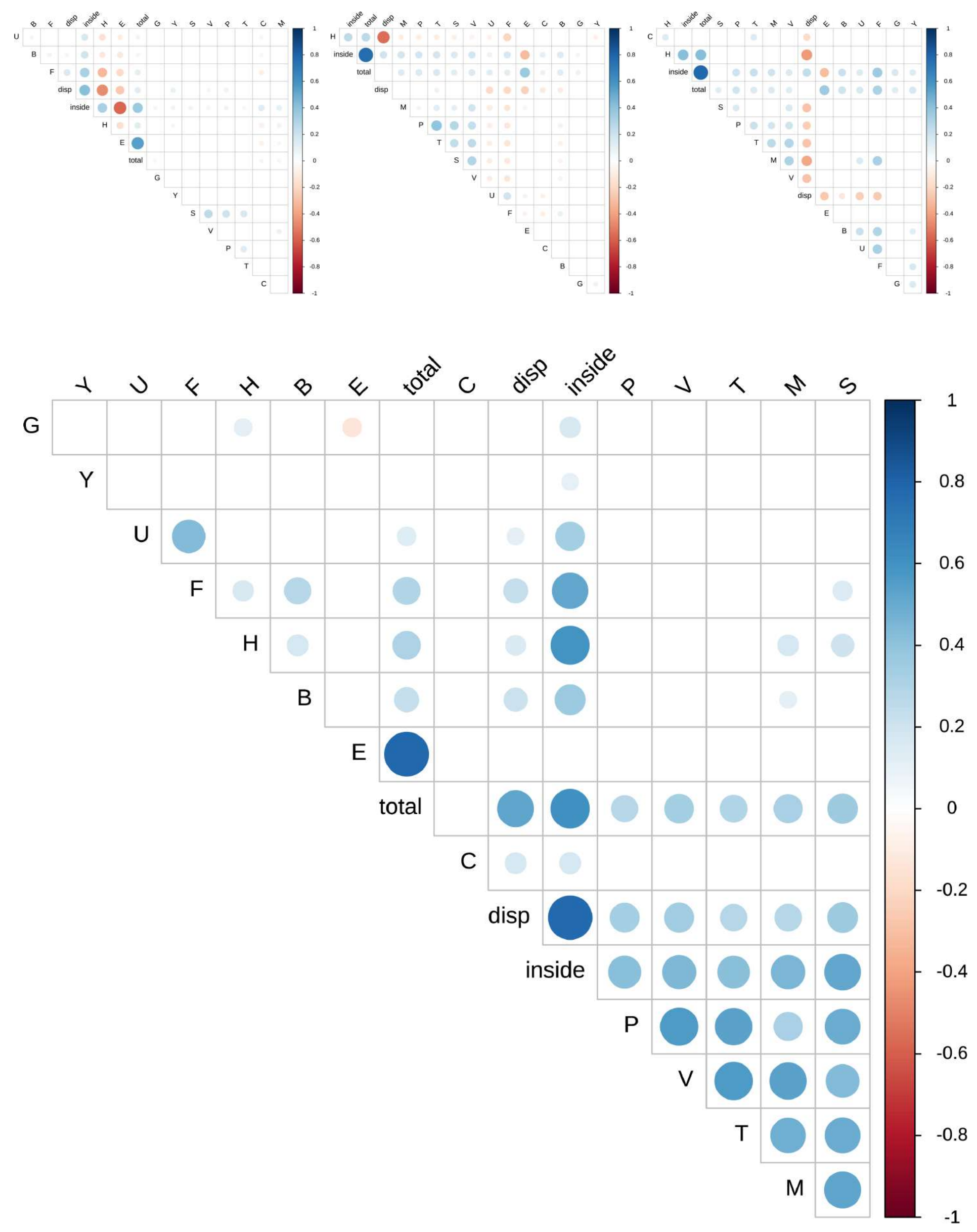http://www.labdma.unina.it
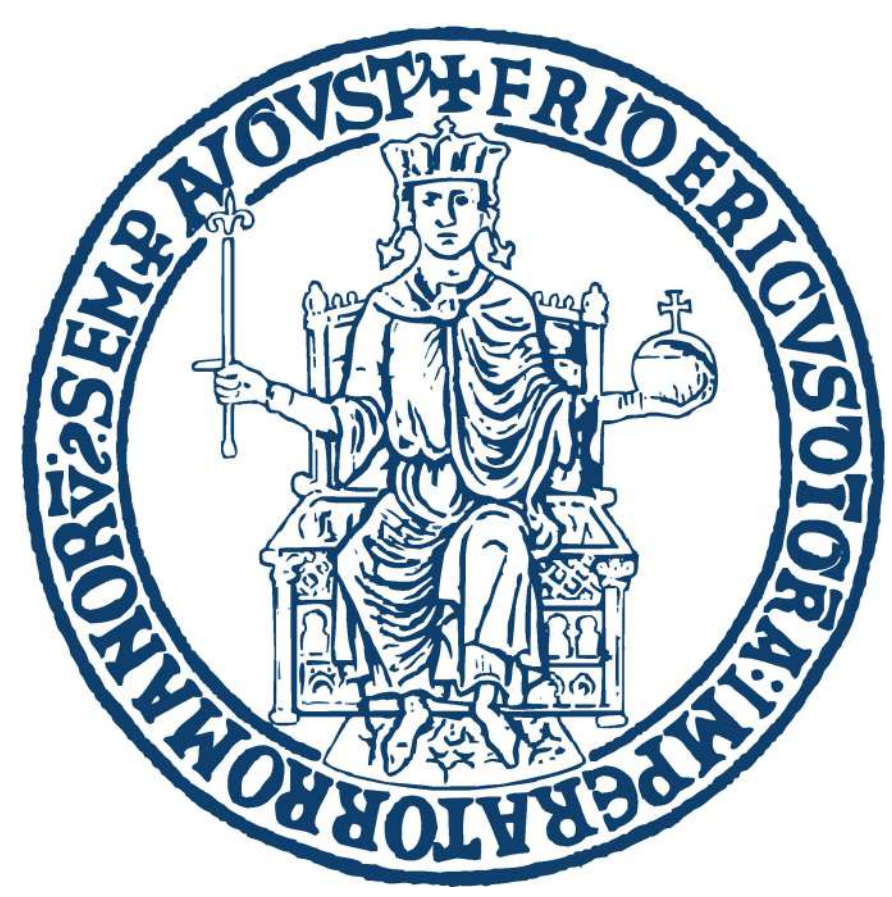
Introduction (1)

Stakeholders of cultural assets need more advanced and comprehensive ICT systems for monitoring and modelling visitor's behaviors.

In order to asses such needs, a set of IoT data about visitors' behaviors in the MANN museum have been used. The dataset is composed of 4258, 4962, and 9634 elements for the months of August, September, and October, respectively. For each visitor the dataset contains:

- time they spent in each of the 13 nodes of the museum

- *total* time of the visit

- *inside* time, i.e. time inside the museum, except the time spent in queue

- *displacement* time, i.e. time period visitor has not been tracked for more than 12 seconds

Time correlation (3)

Find correlations among data can help to highlight relevant features to analyze.

## K value selection and PCA (2)

To choose the number of clusters, k, we used a heuristic approach based on methodologies implemented in the `NbClust` package, [1], and on the ideas reported in [2] as *difference-like* and *optimization-like criteria*.

To choose the final k an *ensemble* approach through a *voting* technique is used.

Clustering results (4)

Cluster profiles results, varying by months, obtained by K-Medoids (PAM) [3] because it is less sensitive to outliers than K-Means.
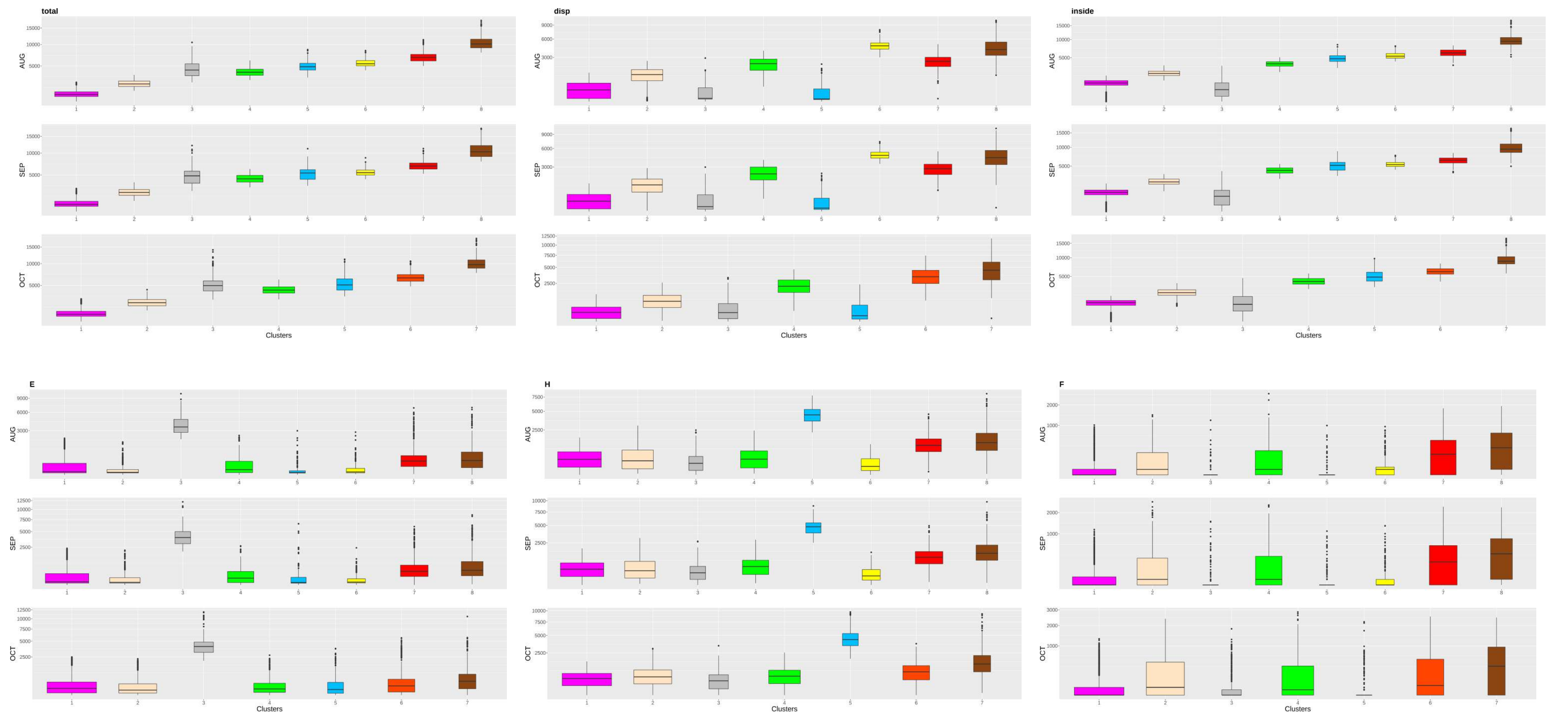
Data visualization: A comparison (5)

Example of a three field plot [4], for the month of October (time quantiles and medoids).

References (6)

[1] M. Charrad, N. Ghazzali, V. Boiteau, A. Niknafs, NbClust: An R Package for Determining the Relevant Number of Clusters in a Data Set. Journal of Statistical Software, Vol. 61(6), pp.136, 2014

[2] L. Vendramin, R. J. G. B. Campello, and E. R. Hruschka, Relative clustering validity criteria: A comparative overview. Statistical Analysis and Data Mining: The ASA Data Science Journal, vol. 3, no. 4, pp. 209-235, 2010

[3] A. P. Reynolds, G. Richards, B. de la Iglesia, and V. J. Rayward-Smith, Clustering rules: A comparison of partitioning and hierarchical clustering algorithms. Journal of Mathematical Modelling and Algorithms, vol. 5, no. 4, pp. 475-504, 2006

[4] M. Aria, C. Cuccurullo, bibliometrix: An R-tool for comprehensive science mapping analysis. Journal of Informetrics 11 (4), pp. 959-975, 2017