DSSV 2023 Data Science, Statistics & Visualisation
EC DA 2023 European Conference on Data Analysis

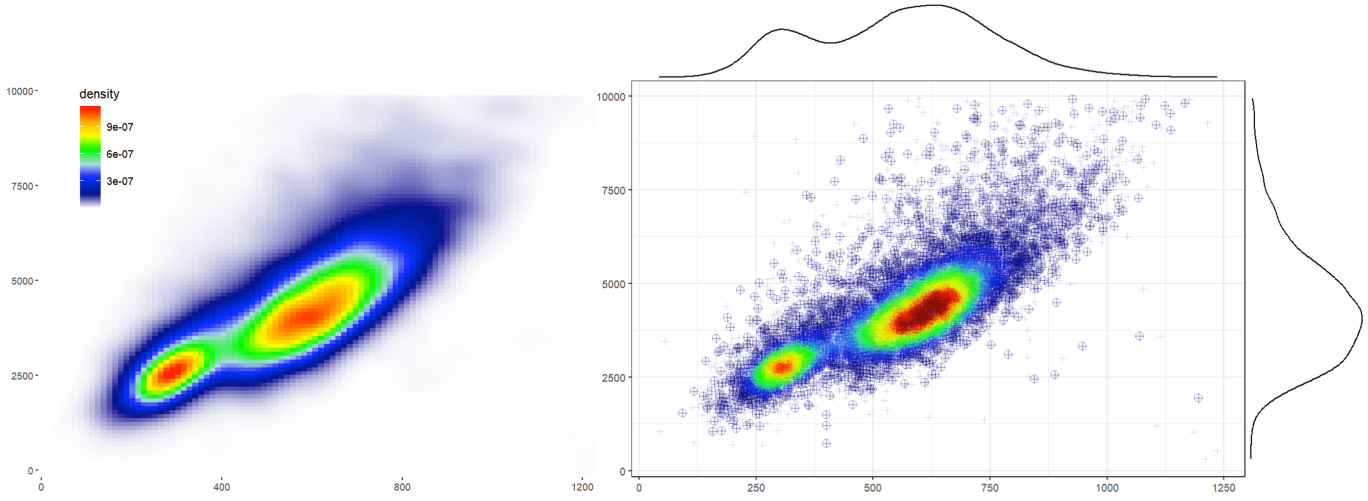
Computing Sensitive Color Transitions for the Identification of Two-Dimensional Structures

Luca Brinkmann[1], Quirin Stier[1], and Michael C. Thrun[1]

[1]Mathematics and Computer Science, Philipps University Marburg, Hans-Meerwein-Strasse 6, D-35032 Marburg, Germany

Keywords: visualization, two-dimensional density estimation

Research data obtained during economics or medical studies experiments often displays a complex distribution even in two-dimensional space. Hence, the statistical identification of subgroups in research data poses an analytical challenge for the interactive identification of meaningful multimodal structures in data [1]. State-of-the-art visualization methods of two-dimensional densities, like ggplot2 (CRAN.R-project.org/package=ggplot2) provide continuous contour plots, as shown in the figure on the left-hand side. These visualization methods are often not sensitive enough to detect multimodal structures. Especially, heights of density peaks are not visualized in a discriminatory manner in a way that height differences of the peaks are represented correctly (see figure on the left side). We propose to employ two-dimensional probability density estimation as one-dimensional information within the 1D distribution cluster algorithm (DDCAL) [2] to find appropriate color transitions. Furthermore, we present the two-dimensional density estimation in a discrete scatter plot and still remain in the framework of ggplot2. As a result, we can visualize multimodal structures with higher sensitivity in comparison to conventional visualization approaches, as presented in the figure on the right. Based on several examples of flow cytometry and economic data, we can show an improvement in distinguishing multimodal structures.



[1] M. C. Thrun, Q. Stier, and A. Ultsch, "Interactive toolbox for two-dimensional gaussian mixture modeling," in *Machine Learning and Knowledge Discovery in Databases: European Conference, ECML PKDD 2022, Grenoble, France, September 19–23, 2022, Proceedings, Part VI*, pp. 658–661, Springer, 2023.

[2] M. Lux and S. Rinderle-Ma, "Ddcal: Evenly distributing data into low variance clusters based on iterative feature scaling," *Journal of Classification*, pp. 1–39, 2023.