Interactive inspection of large data
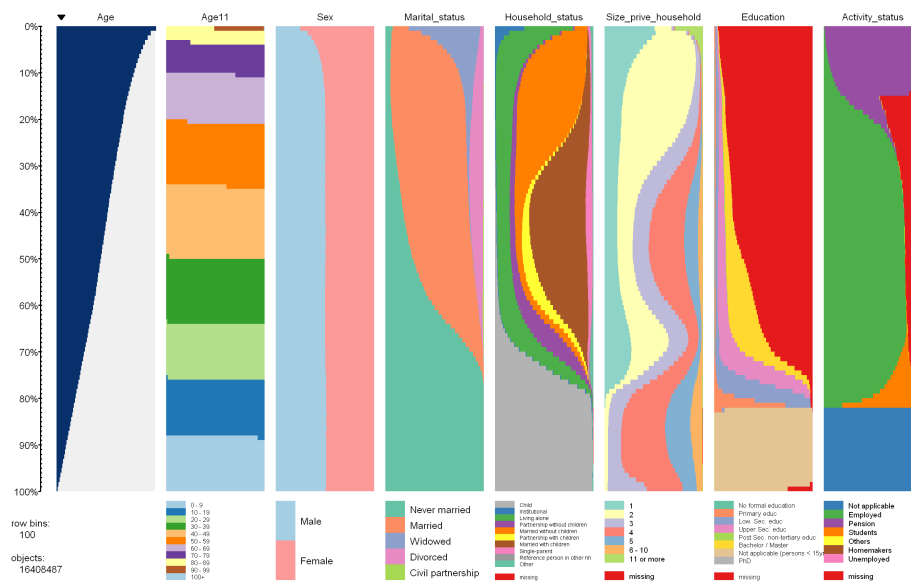
Edwin de Jonge[1⋆]**, Martijn Tennekes**[1]

1. Statistics Netherlands
⋆Contact author: e.dejonge@cbs.nl

Keywords: multivariate, visualization, large data, interactive

We present **tabplotd3** an interactive tool in *R* for inspecting large multivariate data consisting of million rows. **tabplotd3** allows to visually check for anomalies, patterns and missing values in bivariate relations in multivariate data, for categorical as well as numerical variables. The interface creates tableplots [1] from multi million `data.frames` and allows to zoom in to individual observations. We applied the tool to the Dutch (virtual) census [2], creating a tableplot that depict the 17 million inhabitants of the Netherlands.

The presentation will describe tableplots, their use and implemention in **tabplot** and **tabplotd3**. The tool is an interactive web application. To allow interactive actions on multi million data.frames we implemented some aggregations methods in *C*. The user interface is in *javascript* and *SVG*. The tool builds on several other packages including **Rook**, **RJSONIO**, **ff** and our non interactive **tabplot**.

References

[1] Malik, W., A. Unwin, and A. Gribov (2010). An interactive graphical system for visualizing data quality - tableplot graphics. In H. Loracek-Junge and C. Weihs (Eds.), *Classification as a Tool for Research, Proceedings of the 11th IFCS Conference*, pp. 331–339. Berlin: Springer.

[2] Nordholt, E. S. (2005). The dutch virtual census 2001: A new approach by combining different sources. *Statistical Journal of the United Nations Economic Commission for Europe 22(1)*, 25–37.