Interactive inspection of large data

Edwin de Jonge¹, Martijn Tennekes¹

1. Statistics Netherlands, The Hague/Heerlen

¹Contact author: e.dejonge@cbs.nl

Keywords: multivariate analysis, visualization, large data, web application

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

The presentation will describe tableplots, their use and implemention in tabplot[3] and tabplotd3. The tool is a R package that spawns an interactive web application. To allow fast 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

