Exploring the multivariate structure of missing values using the R package VIM

Matthias Templ[1,2,*] **, Andreas Alfons**[1]**, Peter Filzmoser**[1]

1. Department of Statistics and Probability Theory, Vienna University of Technology
2. Department of Methodology, Statistics Austria

* Contact author: templ@tuwien.ac.at

Keywords: Missing Values, Visualization, Missing Value Mechanism, R.

In our presentation we will describe and demonstrate the usefulness of exploring missing values in data by using visualization tools in order to get a first impression of the data but also as pre-processing step before imputation.

Before choosing an imputation procedure to impute missing values in data one should be aware of the missing value mechanism. We propose the use of visualization tools for the detection of the missing value mechanism(s). Univariate plots such as histograms and spineplots, multivariate plots such as multiple scatterplots and parallel coordinate plots and maps are adapted in order to visualize missing values (see also Theus et al., 1997), but new plots ("matrixplot" and others) are provided as well (see, e.g., Templ and Filzmoser, 2008). In addition to that, interactivity is provided when using these plots, i.e. the users have the possibility to highlight or re-arrange the plots by clicking on the plots.

VIM can be used for data from essentially any field. If spatial coordinates are available, it is possible to load a background map and present information about missing values on that map.

Our developed R-package VIM (Templ and Alfons, 2009) includes a graphical user interface for interactive, easy to grasp graphics in order to make the tools available also for non-experts in R.

We will illustrate our proposed visualization methods through a short demo using real-life data sets.

This work was partly funded by the European Union (represented by the European Commission) within the 7th framework programme for research (Theme 8, Socio-Economic Sciences and Humanities, Project AMELI (Advanced Methodology for European Laeken Indicators), Grant Agreement No. 217322).

References

Templ, M. and Filzmoser, P. (2008). Visualisation of Missing Values Using the R-package VIM. Research Report CS-2008-1, Department of Statistics and Probability Theory, Vienna University of Technology. http://www.statistik.tuwien.ac.at/forschung/CS/CS-2008-1complete.pdf.

Templ, M. and Alfons, A. (2008). VIM: Visualization and Imputation of Missing Values. R package version 1.2.4. http://cran.r-project.org.

M. Theus, H. Hofmann, B. Siegl, A. Unwin.(1997). MANET: Extensions to Interactive Statistical Graphics for Missing Values. In New Techniques and Technologies for Statistics II, IOS-Press Amsterdam, 247-259. http://home.vrweb.de/~martin.theus/NTTS.pdf