



Visualisation and Imputation of Missing Values

Alexander Kowarik*

Statistics Austria, Vienna, Austria - alexander.kowarik@statistik.gv.at

Matthias Templ

ZHAW, Winterthur, Switzerland - matthias.templ@gmail.com

In recent years many shared tools have been developed in and for Statistical Offices. Our developments are shown exemplarily on basis of shared tools for imputation of missing values. To show a use-case, we will take the package VIM in the focus of our presentation. This package is developed to explore and analyze the structure of missing values in data using visualization methods, to impute these missing values with the built-in imputation methods and to verify the imputation process using visualization tools. The package is regularly used by many organizations and users. We also give an outline how these tools are deployed and how users are supported.

Four different imputation methods are currently implemented in VIM, namely hot-deck imputation, k-nearest neighbor imputation, regression imputation and iterative robust model-based imputation. These methods are implemented in a flexible and efficient manner with many options for customization. New features and applications will be shown.

References:

M. Templ, A. Alfons, P. Filzmoser. Exploring Incomplete Data Using Visualization Techniques. *Advances in Data Analysis and Classification*, 6(1), 29-47, 2012.

A. Kowarik, M. Templ. Imputation with the R-package VIM. *Journal of Statistical Software*, 1-16, 2016

M. Templ, V. Todorov. The Software Environment R for Official Statistics and Survey Methodology. *Austrian Journal of Statistics*, 45(1), 97-124, 2016.

Keywords: imputation, missing values, visualisation.