Make Your Data Confidential with the sdcMicro and sdcMicroGUI packages

Alexander Kowarik, Matthias Templ

Keywords: Statistical disclosure methods, disclosure risk and data utility, S4 class package, graphical user interface

The demand of data from surveys, registers or other data sets containing sensible information on people or enterprises have increased significantly over the last years. However, before providing data to the public or to researchers, confidentiality has to be respected for any data set containing sensible individual information. Confidentiality can be achieved by applying statistical disclosure methods on the data.

We present a methodological and object-oriented approach to anonymize data - the sdcMicro package (Templ and Meindl, 2010; Templ et al., 2013), which includes all popular methods on statistical disclosure control. After specifying an S4-class “sdcMicro” object, all methods are directly applied to this object, whereas all necessary slots are filled in and updated automatically. For example, the disclosure risk and data utility estimates update whenever a method is applied on this object. This allows comparisons (how much a disclosure method influences the data utility/risk) and gives high usability into the hand of the user. Moreover, a reporting system is provided, whereas the whole process of anonymisation is summarized.

In addition, an highly interactive point and click graphical userinterface is implemented and available in the sdcMicroGUI package (Kowarik and Templ, 2013). All essential information is always updated and made visible to the user.

To allow for fast computations, all essential methods in the sdcMicro package are based on efficient C++ code.

References

