

Bernhard Meindl¹
Matthias Tempel^{1,2}

¹Statistics Austria

²Vienna Univ. of Technology

WORKSESSION, SDC
Tarragona, October 2011

Protecting statistical tables
R-Package sdcTable

Overview, concepts and practical application

Overview of R-package `sdcTable`

- ▶ `sdcTable` is on CRAN since 2009, major rewrite in 2011
- ▶ S4-classes are extensively used
 - ▶ leads to *robustification*
 - ▶ makes it *easier to add* features
- ▶ Constant use of S4-methods
 - ▶ easier debugging
 - ▶ modularization
- ▶ `sdcTable` is easily to expand, adjust or tweak
 - ▶ open source code
 - ▶ modular design of R

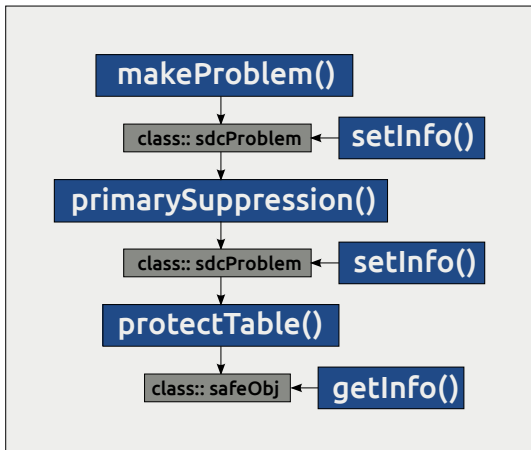
Main features of `sdcTable`

- ▶ Allows to protect tables having
 - ▶ multiple dimensions
 - ▶ arbitrarily complex structure of dimensions
- ▶ Standardized problem generation
 - ▶ Specification of dimensions
 - ▶ Data: micro-data, pre-aggregated data, ...
- ▶ Algorithms/Methods available:
 - ▶ primary cell suppression (different rules, custom)
 - ▶ secondary cell suppr. using a cut and branch algorithm (OPT, HITAS)
 - ▶ secondary cell suppr. using a (simple) variation of GHMITER
 - ▶ protection of 2 tables having common table cells

Things missing in `sdcTable`

- ▶ No GUI (graphical user interface)
- ▶ Some algorithms:
 - ▶ Perturbation
 - ▶ Rounding
- ▶ Comparison of different protected datasets
- ▶ (probably) many other things not yet required (within STAT)

Application of sdcTable



Possible future developments

- ▶ Plans:
 - ▶ additional algorithms (perturbation, rounding, ...)
 - ▶ more flexible output generation (eg. 'templating')
 - ▶ performance improvements
- ▶ Constraints: time and needs
- ▶ Feedback and contributions to `sdcTable` are (very much) desired
 - ▶ code review / contributions
 - ▶ bug hunting
 - ▶ feature requests / general feedback