New developments in SDC tools in R

Bernhard Meindl (bernhard.meindl@statistik.gv.at)

Bukarest, November 2017
What's the plan?

We're going to have a look at two R-packages that can be used for SDC-purposes.
What's the plan?

We're going to have a look at two R-packages that can be used for SDC-purposes

**sdcMicro**

R-package for anonymization of micro data
What's the plan?

We're going to have a look at two R-packages that can be used for SDC-purposes

**sdcMicro**

R-package for anonymization of micro data

**sdcTable**

R-package for anonymization of tabular data
sdcMicro

micro data anonymization
Some History

The beginning

- development started around 2006 at Statistics Austria and TU Vienna (thx Matthias!)
- first version on CRAN
Some History

The beginning

- development started around 2006 at Statistics Austria and TU Vienna (thx Matthias!)
- first version on CRAN

In-Between

- working our way from simple functions to class-oriented implementation
- some "roadblocks" (eg. a not-very usable GUI)
Some History

The beginning

- development started around 2006 at Statistics Austria and TU Vienna (thx Matthias!)
- first version on CRAN

In-Between

- working our way from simple functions to class-oriented implementation
- some "roadblocks" (eg. a not-very usable GUI)

Now

- This year: 10 year anniversary :)
- new shiny interface
General information

Installation

- current stable version (v5.0.4) from CRAN

```r
install.packages("sdcMicro")
```
General information

Installation

- current stable version (v5.0.4) from CRAN
  ```
  install.packages("sdcMicro")
  ```

- development version from github
  ```
  devtools::install_github("sdcTools/sdcMicro", ref="next")
  ```
General information

Installation

- current stable version (v5.0.4) from CRAN
  ```r
  install.packages("sdcMicro")
  ```
- development version from github
  ```r
  devtools::install_github("sdcTools/sdcMicro", ref="next")
  ```
- `next`-branch is eventually merged to `master`
- new stable version is built from `master`-branch
Development

'Home' of the development

- European Project SGA on SDC tools (https://github.com/sdcTools)
- started to improve the development, support of tools for statistical disclosure control
- "umbrella" for many tools, also tau-argus and mu-argus
- sdcMicro: https://github.com/sdcTools/sdcMicro
Development

'Home' of the development

- European Project SGA on SDC tools ([https://github.com/sdcTools](https://github.com/sdcTools))
- started to improve the development, support of tools for statistical disclosure control
- "umbrella" for many tools, also tau-argus and mu-argus

Get in touch or contribute

- tell us about bugs in the issue tracker
- ask for help/improvements
- contributions are very welcome, just file a pull-request!
Current developments (1)

Implementation of methods from mu-argus
Current developments (1)

Implementation of methods from mu-argus

argus-algorithm for rank swapping

?argus_rankswap
Current developments (1)

Implementation of methods from mu-argus

argus-algorithm for rank swapping

?argus_rankswap

argus-algorithm for microaggregation

?argus_microaggregation
Current developments (1)

Implementation of methods from mu-argus

argus-algorithm for rank swapping

?argus_rankswap

argus-algorithm for microaggregation

?argus_microaggregation

detect cells with small counts

?riskyCells
Current developments (2)

Harmonization

- task of SGA on sdc tools: functionalities were identified that should yield the same results independent from the software tool used
- tests have been done to check where harmonization was still needed
- development is underway (e.g. the via the functions shown before)
Current developments (3)

new graphical user interface

- the GUI is based on **shiny**
- goal: non-experts should be able to anonymize micro data
- many neat features
- actively developed, new default theme was contributed by word bank
Current developments (3)

new graphical user interface

- the GUI is based on shiny
- goal: non-experts should be able to anonymize micro data
- many neat features
- actively developed, new default theme was contributed by word bank

Live demonstration (bear with me :))

```r
## start the GUI
sdcApp()
```
sdcTable

tabular data protection
General information

Installation

- current stable version (v0.22.8) from CRAN

```r
install.packages("sdcTable")
```
General information

Installation

- current stable version (v0.22.8) from CRAN

```r
install.packages("sdcTable")
```

'Home' of the development

- Development on Github: https://github.com/bernhard-da/sdcTable
Current developments (1)

Methods to interface with tau-argus

create batch files suitable for tau-argus

?createArgusInput
Current developments (1)

Methods to interface with tau-argus

create batch files suitable for tau-argus

?createArgusInput

call tau-argus and run a batch-file

?runArgusBatchFile

- you can of course run "any" batch-file
- but: tau-argus must be installed
Current developments (2)

Code.cleanup

- methods that are better implemented/tested in tau-argus will likely be removed
- concentration on greedy heuristic protection algorithm
Current developments (2)

Code-cleanup

- methods that are better implemented/tested in tau-argus will likely be removed
- concentration on greedy heuristic protection algorithm

New functionality

- possible implementation of the ABS cell key method (future Eurostat project)
Current developments (3)

easySdcTable

- "Add-On" package developed from Statistics Norway (Øyvind Langsrud)
- includes a graphical user interface based on shiny
- also allows to call tau-Argus
Thanks for your attention

Any questions/comments?

Ask now or

Contact me: bernhard.meindl@statistik.gv.at