

# saeSim: Simulation Tools for Small Area Estimation

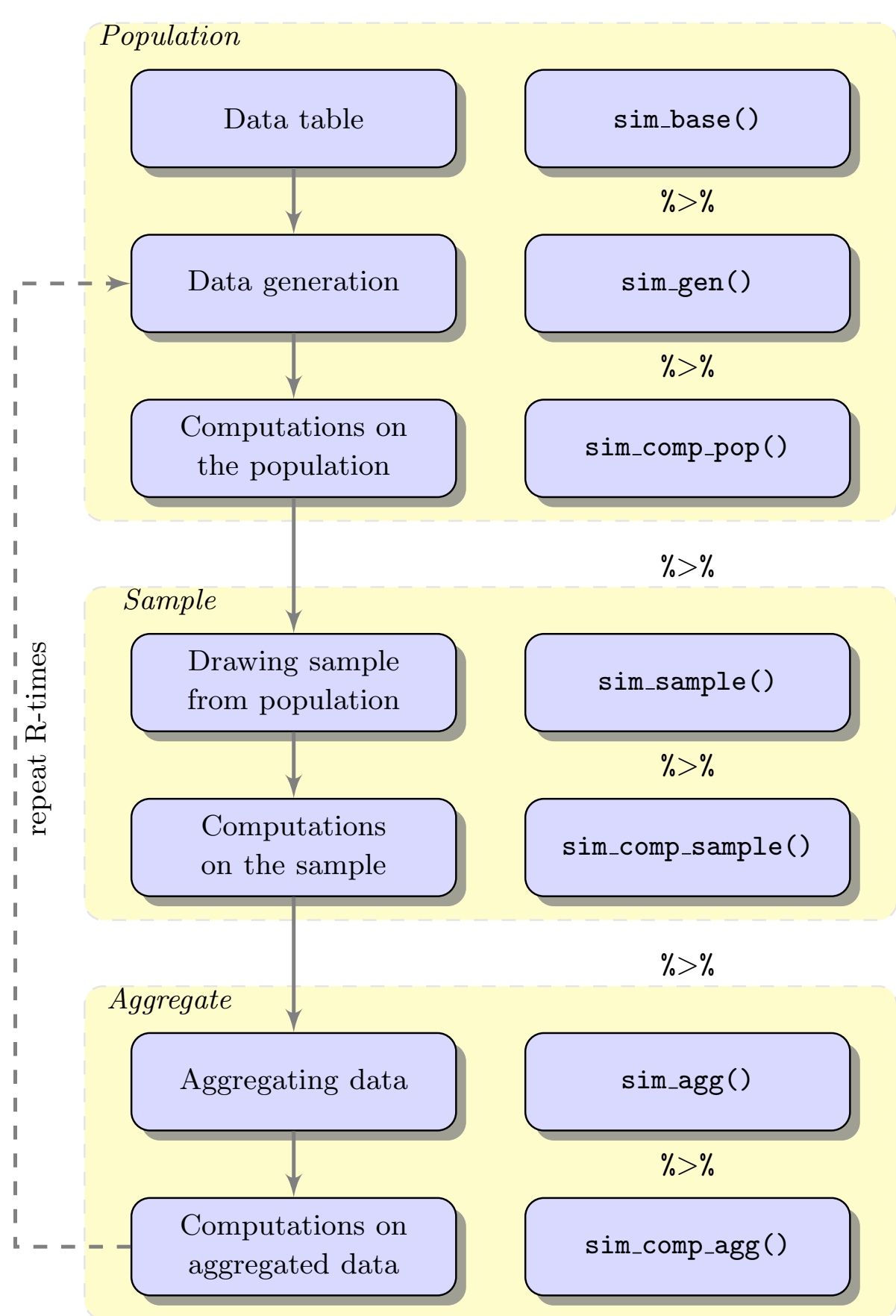
## Sebastian Warnholz & T. Schmid

Sebastian.Warnholz@fu-berlin.de



### Introduction

- Unified markup/tools for the composition of simulation studies in the context of small area estimation.
- Small area estimation summarises the development and application of statistical methods to report statistics for small groups. *Small* refers to the small number of sampled units.
- Model and design based simulation studies have been used to introduce new methods to the field.
- The package highlights a specific way to map a simulation study into R, namely in terms of a pipeline where a data frame is modified in each step. Also frequently used tools are ready to use which make the composition of such studies considerably more easy.
- With this package the composition of a simulation study is reduced to *chaining the steps together*.



### Data Generation

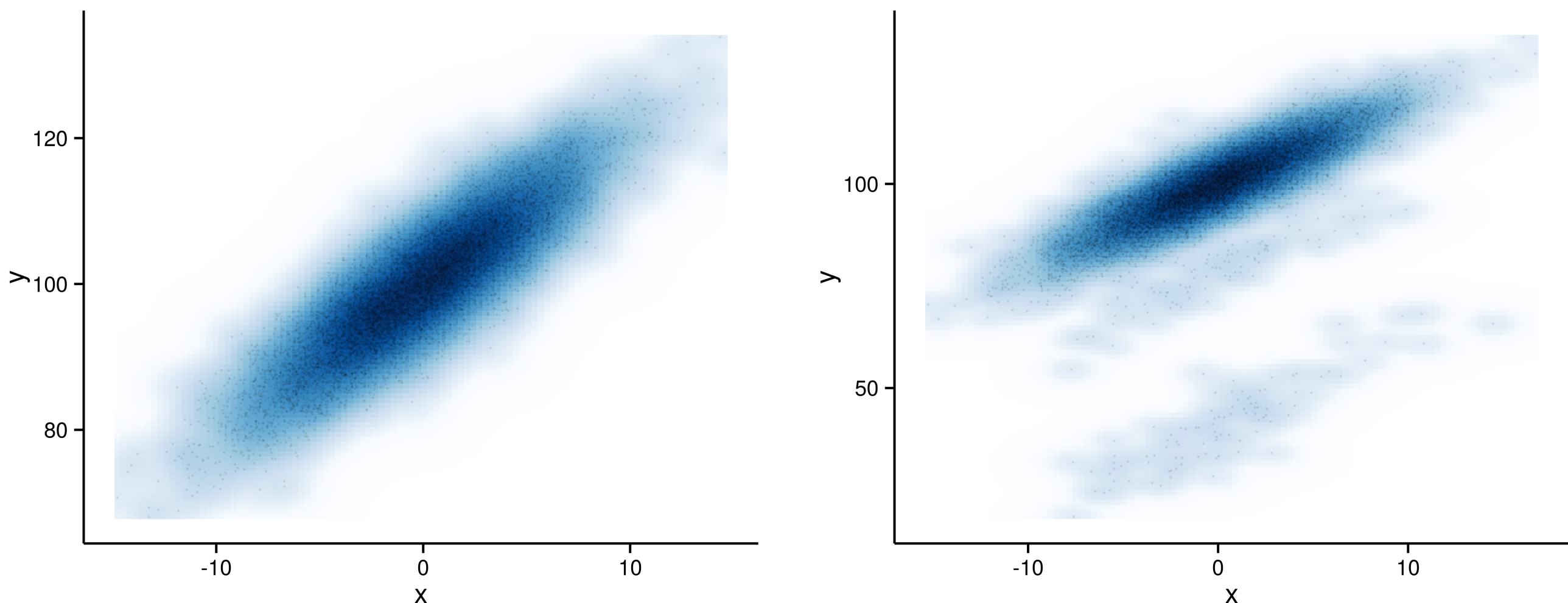
Predefined setups and functions to generate random numbers as random effects including spatial correlation. Also a simple interface to univariate random number generators in R.

```
setup <- sim_base() %>%
  sim_gen_x() %>% # x ~ N(0, 16) iid
  sim_gen_e() %>% # e ~ N(0, 16) iid
  sim_gen_generic(rt, df = 2, groupVars = "idD", name = "v") %>%
  sim_resp_eq(y = 100 + 2 * x + v + e)
setup
## data.frame [10,000 x 6]
##
##   idD idU      x      e      v      y
## 1    1    1  3.225797  0.7579929  1.477616 108.68720
## 2    1    2 -1.440586 -1.7312849  1.477616  96.86516
## 3    1    3  1.262536  3.0967488  1.477616 107.09944
## 4    1    4 -1.793524 -1.1042322  1.477616  96.78633
## 5    1    5  1.144710  0.2524723  1.477616 104.01951
## 6    1    6  0.874534 -3.8347551  1.477616  99.39193
## .. ... ..
```

### Outliers

- Set the frequency or probability for adding contaminated observations
- Specify contamination within domains or across the population
- Change between area and unit level contamination

```
setup %>%
  sim_gen_cont(gen_generic(rnorm, sd = 150, name = "e"), 0.1, "unit")
## data.frame [10,000 x 7]
##
##   idD idU      x      e      v idC      y
## 1    1    1 -2.166447 -1.029188  1.208294 FALSE  95.84621
## 2    1    2 -4.770198  2.172307  1.208294 FALSE  93.84021
## 3    1    3  2.393438 -4.610486  1.208294 FALSE 101.38468
## 4    1    4  2.754315 -5.162324  1.208294 FALSE 101.55460
## 5    1    5  2.735989 -3.642082  1.208294 FALSE 103.03819
## 6    1    6 -2.055526  5.921647  1.208294 FALSE 103.01889
## .. ... ..
autoplot(setup)
autoplot(setup %>% sim_gen_vc()) # contamination on area-level
```



### Timo Schmid

Department of Economics  
Freie Universität Berlin  
D-14195 Berlin, Germany  
Timo.Schmid@fu-berlin.de

### Basc idea

Define scenarios, explain differences!

```
setup1 <- sim_base_lm() %>% sim_sample(sample_number(5))
setup2 <- sim_base_lm() %>% sim_sample(sample_fraction(0.05))
```

- setup1 and setup2 differ in the specific way samples are drawn. sim\_sample is responsible to find the position in the process
- Every sim\_\* function expects a simulation setup or data.frame as first argument
- Every sim\_\* controls at which position in the process a function is called
- For every step in the process tools are named using the corresponding prefix, i.e. gen\_generic or sample\_fraction

### Sampling

Sampling schemes like simple random sampling in domains and cluster sampling.

### Parallel computations

Simulation studies are embarrassingly parallel. For parallel computations we utilize parallelMap which makes it easy to switch between different parallel back ends in R (multicore, socket, mpi, BatchJobs).

### Conclusions

- Remark 1
- Remark 2

### References

Alfons, A., Templ, M. & Filzmoser, P. (2010), ‘An object-oriented framework for statistical simulation: The R package simFrame’, *Journal of Statistical Software* **37**(3), 1–36.  
**URL:** <http://www.jstatsoft.org/v37/i03/>

Bache, S. M. & Wickham, H. (2014), *magrittr: A Forward-Pipe Operator for R*. R package version 1.5.  
**URL:** <http://CRAN.R-project.org/package=magrittr>

Warnholz, S. & Schmid, T. (2015), *saeSim: Simulation Tools for Small Area Estimation*. R package version 0.7.0.  
**URL:** <http://CRAN.R-project.org/package=saeSim>

Wickham, H. & Francois, R. (2015), *dplyr: A Grammar of Data Manipulation*. R package version 0.4.1.  
**URL:** <http://CRAN.R-project.org/package=dplyr>