

Public part

This dataset has been treated to protect confidentiality. Several methods have been applied to protect the confidentiality: removing variables from the original dataset, reducing detail in variables by recoding and top-coding, and perturbing values of certain variables.

Removing variables

The released dataset has only a selected number of variables contained in the initial survey. Not all variables could be released without breaching confidentiality rules.

Reducing detail in variables by recoding and top-coding

The variable *citizenship* was recoded to a binary value of Danish / Non-danish.

Perturbing values

The *dob* variable was perturbed by converting to age, to which noise was added by using Laplacian mechanism.

$$f(x \mid 0, 1.33) = 1/(2 \cdot 1.33)e^{\left(\frac{-|x-0|}{1.33}\right)}$$

Laplace mechanism used for adding random noise to the age. Minimum and maximum noise is -5.28 and 7.03

Shuffling values

The rows (individuals) in the dataset were shuffled, so that a leakage of additional information with the same structure would not cause a reidentification of the individuals in the dataset.

Auxiliary dataset

Additionally, an auxiliary dataset is provided, containing a table summarizing counts of individuals having a specific marital status, divided by zip and the method of vote, as per survey.

