Regional variables in TwinLife

* release of two regional variables:
  + federal state / region: ewi (east-west indicator)
  + community size (political): gkpol, recoded from 7 into 4 groups

The decision of the released depth of the regional variables were based on two rules which are :

1. for each combination of regional variables, sex and birth year there should be at least one “statistical twin pair” for each twin pair in the sample.
2. each cluster (resulting from the combination of the two regional variables) should represent a minimum number of people in the population (strictly accurate: persons that belong to a twin family in Germany; here simplified: persons living in Germany) of 400,000.

Regarding the 1st rule, even the highest simplification into two dimensions in each regional variable would result in two cases without statistical twin in the data set. To leave at least some analysis options to the users we instead decided to offer the political community size in 4 classes and the federal states (or, rather, geographical region) in 2 classes, although this concept leaves six twin pairs with combinations of region, sex and birth year without statistical twin in the data set. Given the strong simplification of all sensitive characteristics in the data, the re-identification of these six twin pairs is considered as ruled out without an excessive extent of (illicit) re-identification activities.

Regarding the 2nd rule there are no official statistics of the number of twin families in German communities. Because of this we applied this rule on the basis of persons living in Germany but were more strict