

Data cleaning documentation

4.5.1. Codes for missing in survey and in dataset.

1. Survey:

Variable label	Code
Don't know	-99
Refused to answer	-66

2. Data:

Variable label	Code
Don't know	.d
Refused to answer	.r
Other	.o
Missing	.

4.5.2. Record of any corrections.

1. Removed variables generated by SurveyCTO.
2. Converted PII variables to missing.

4.5.3. Deidentification for masked variables

1. Encoded the cluster codes and used the encoded values to mask labels using labmask in Stata.

4.5.4. Link observations across different datafiles.

1. Use the hh_id or key variable to merge the rosters with the master data for both experiments.