Flowchart

Data Exploration Step 1
Original Data
(N = 63614)
assignmt2a_ed_data.sas7bdat

Data Validation Step 1
(N = 63614)

Adjusted
- Type column in ED data dictionary
  - Number -> Numeric
  - Date -> Numeric

Data Validation Step 2
(N = 63614)

Identified Missing values
- 2.95% of the observation in 'health_insurance' variable is missing.
- 3.64% of the observation in 'interpreter' variable is missing.
- 1.04% of the observation in 'sex_ed' variable is missing.
- 4.01% of the observation in 'cob_ed' variable is missing.

Data Cleaning Step 1
(N = 63614)

Removed
- 1047 complete duplicates

Data Cleaning Step 2
(N = 62567)

Recoded
- cob_ed(1, 2, 99) -> cob_ed(1, 2, 3)

Related documents:
- Data dictionary (/Users/annie/Desktop/HDAT9400/HDAT9400_A2a_data/ED Data Dictonary.docx)
- SAS code: (/Users/annie/Desktop/HDAT9400/HDAT9400_A2a_data/tidy.sas7bdat)

Final Dataset
(N = 62567)
tidy.sas7bdat