

Survey Data Cleaning

Method: Survey Data Cleaning

Method type: data_cleaning_workflow

Design: Deterministic audit of duplicate rows, missing cells, categorical domains, numeric ranges, and final retained rows.

Variables: school, sex, age, address, famsize, Pstatus, Medu, Fedu, Mjob, Fjob, reason, guardian, traveltime, studytime, failures, schoolsup, famsup, paid, activities, nursery, higher, internet, romantic, famrel, freetime, goout, Dalc, Walc, health, absences, G1, G2, G3

Inference scope: No hypothesis test is required; data-quality rules determine whether rows or cells require correction.

Formula: duplicate_i=I(row_i appeared earlier); range violation=I(x<min or x>max); clean rows=distinct valid rows.

Source rows: 649

Calculated metrics:

input_rows: 649

duplicate_rows: 0

missing_cells: 0

category_violations: 0

range_violations: 0

clean_rows: 649

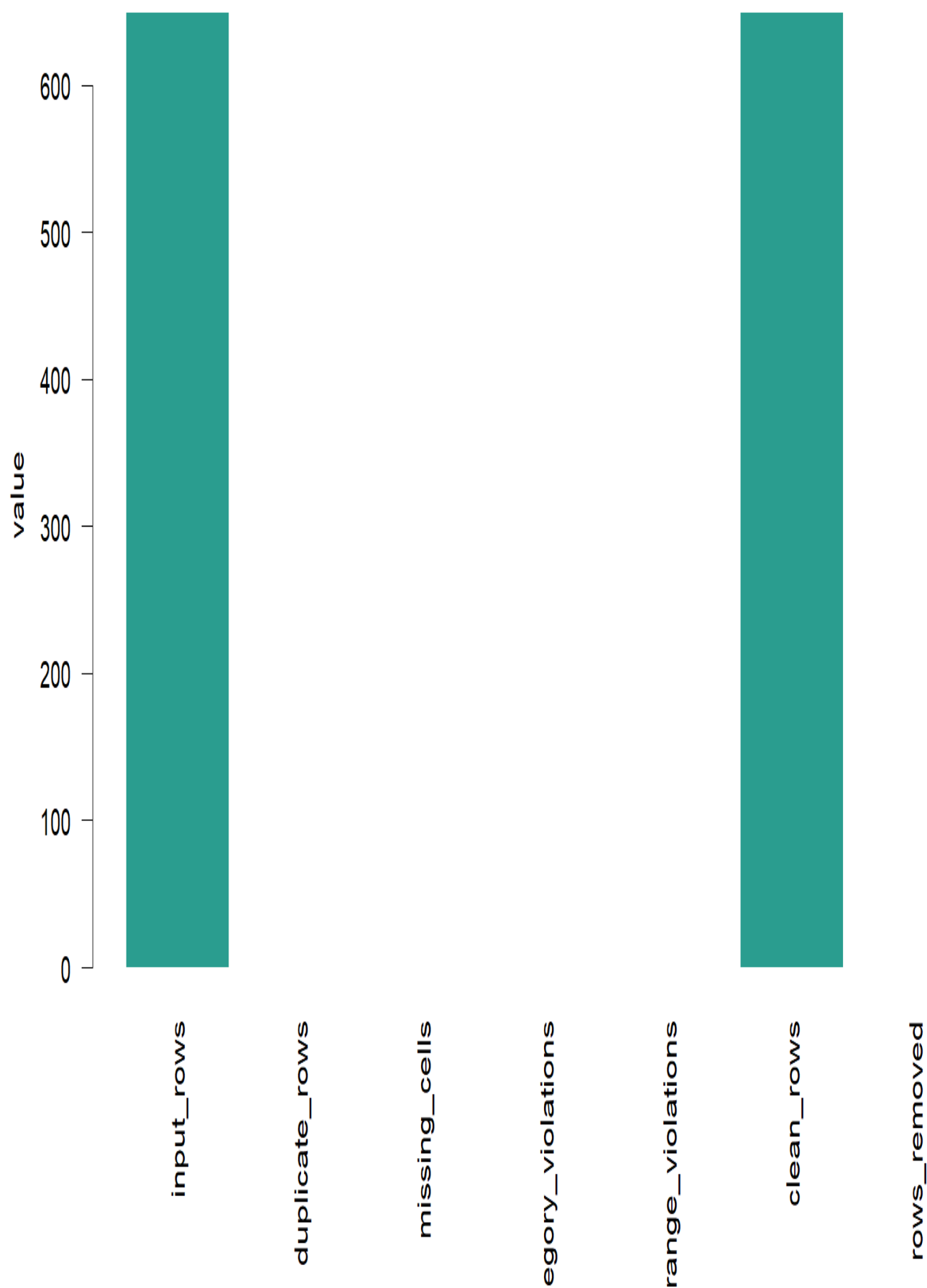
rows_removed: 0

Interpretation:

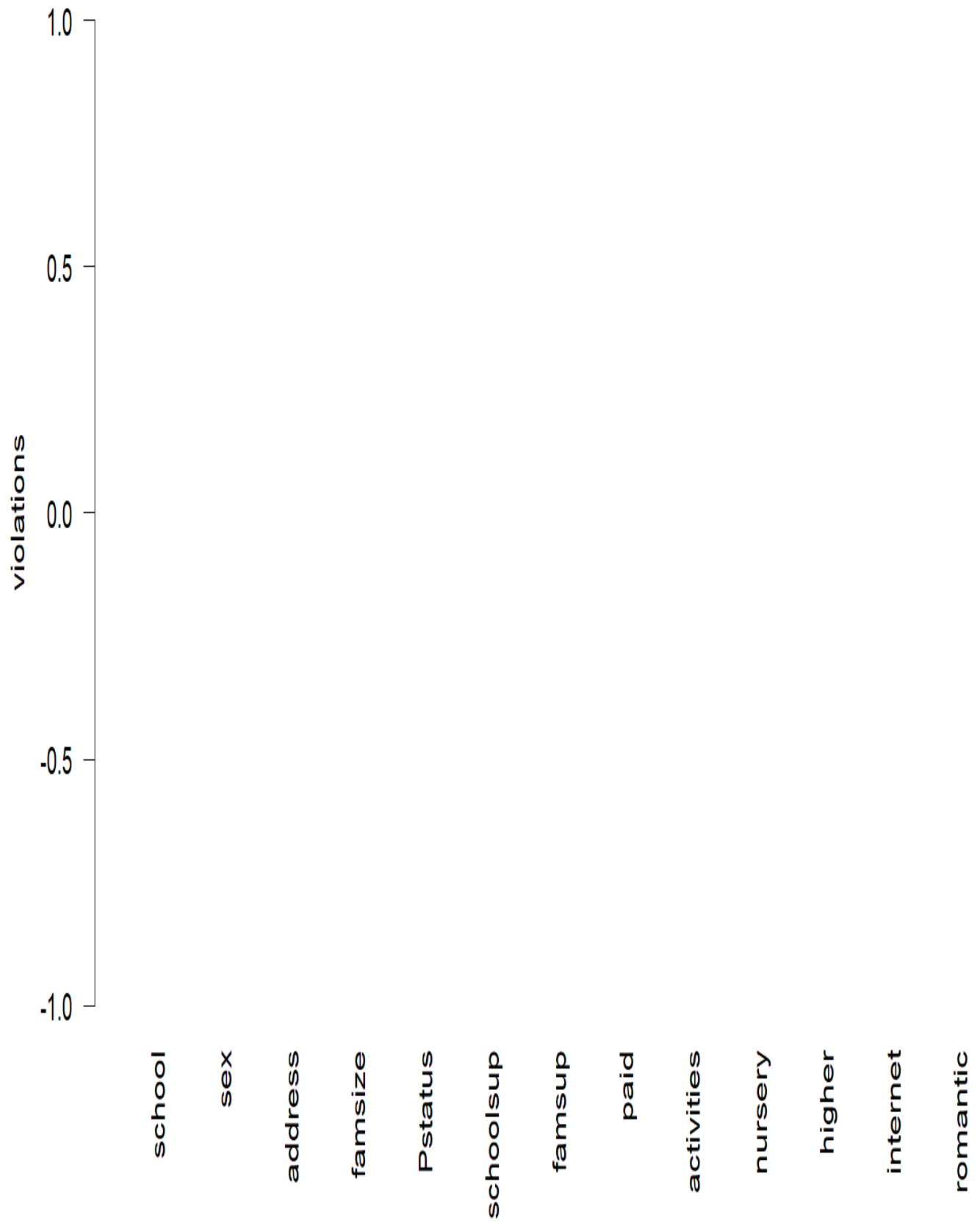
Duplicates, missing cells, domains, and ranges are audited separately.

Input rows 649 clean rows 649

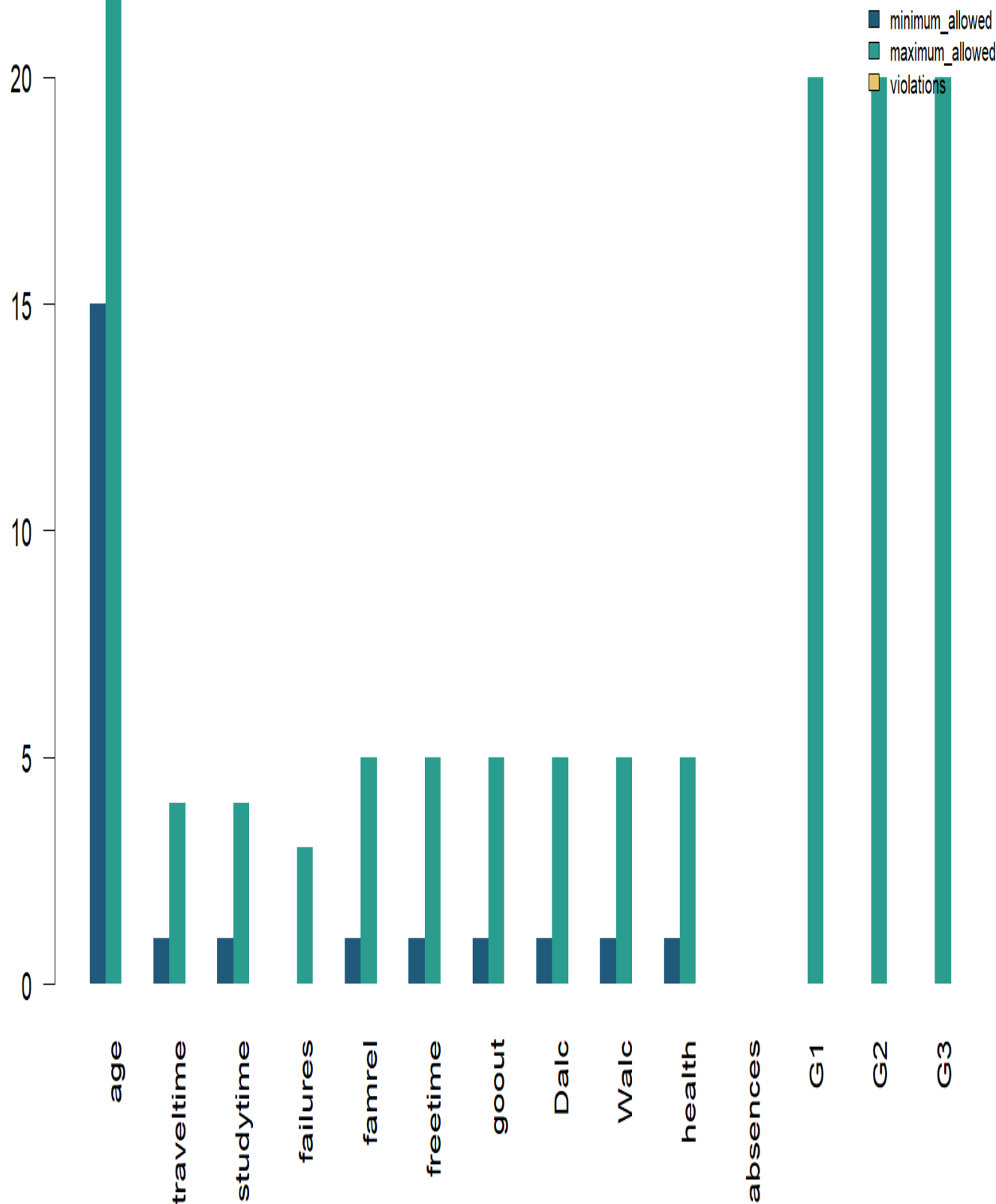
Survey Data Cleaning - primary metrics



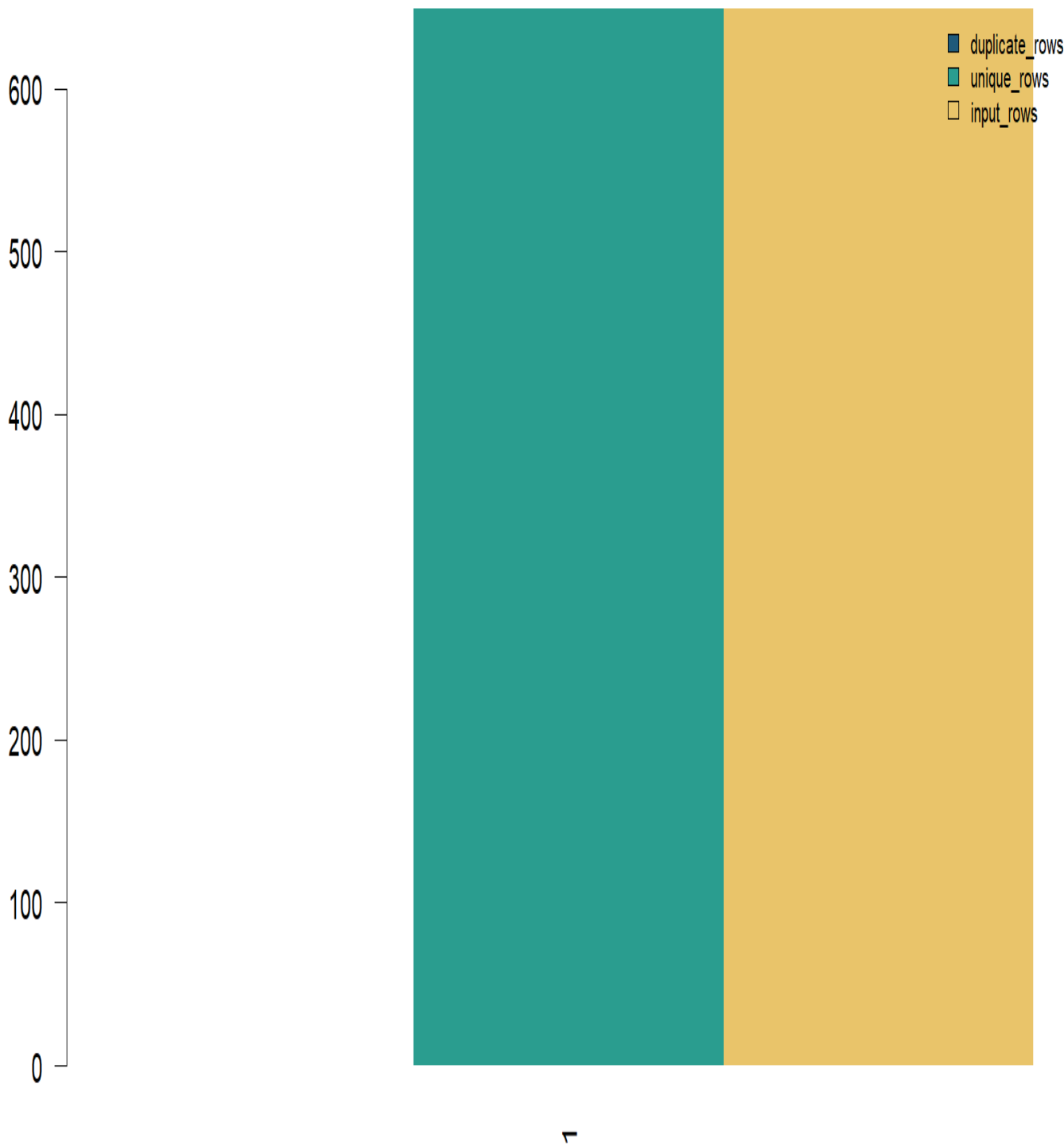
Survey Data Cleaning - categorical domain checks



Survey Data Cleaning - numeric range checks



Survey Data Cleaning - duplicate row summary



Survey Data Cleaning verified results

