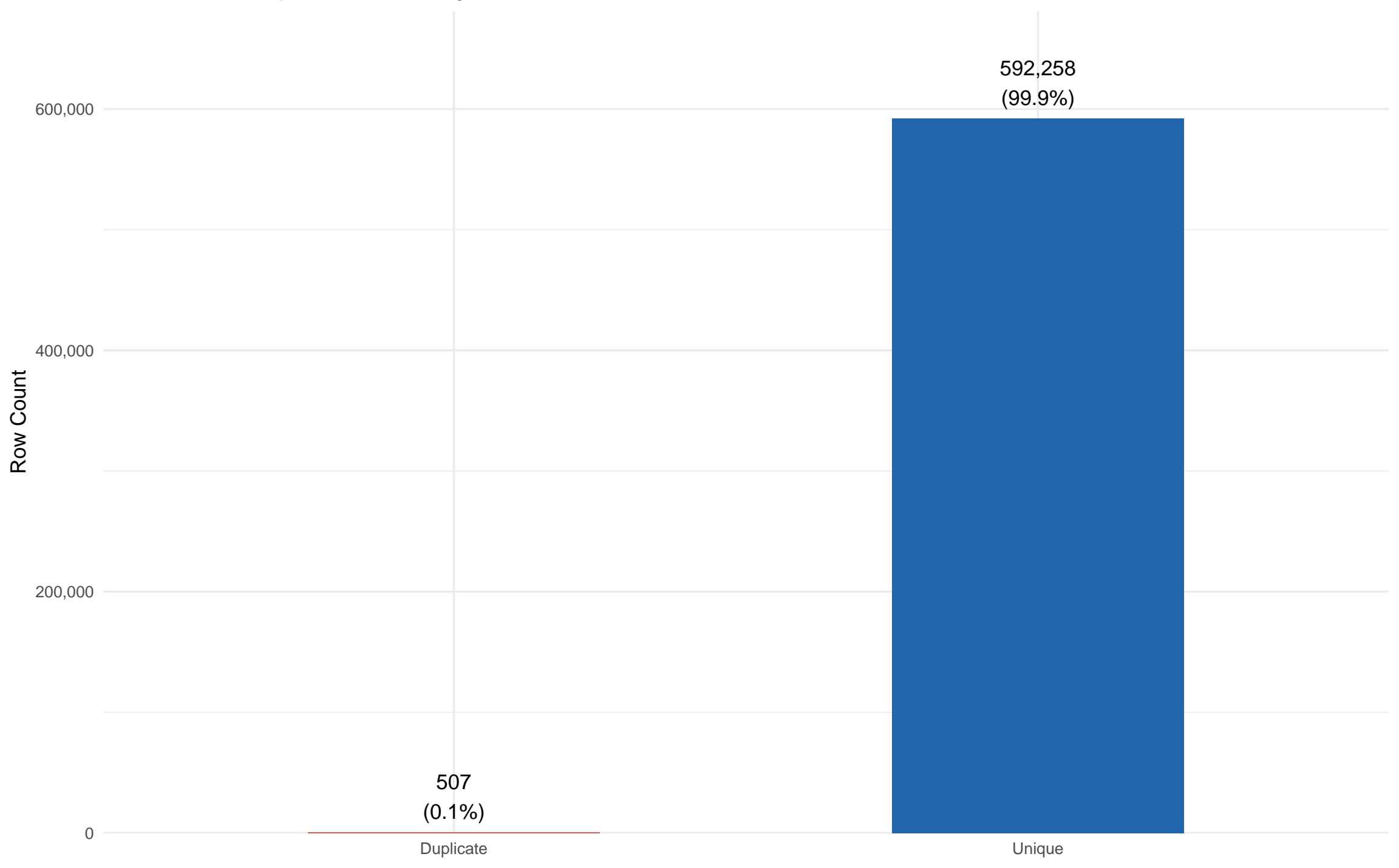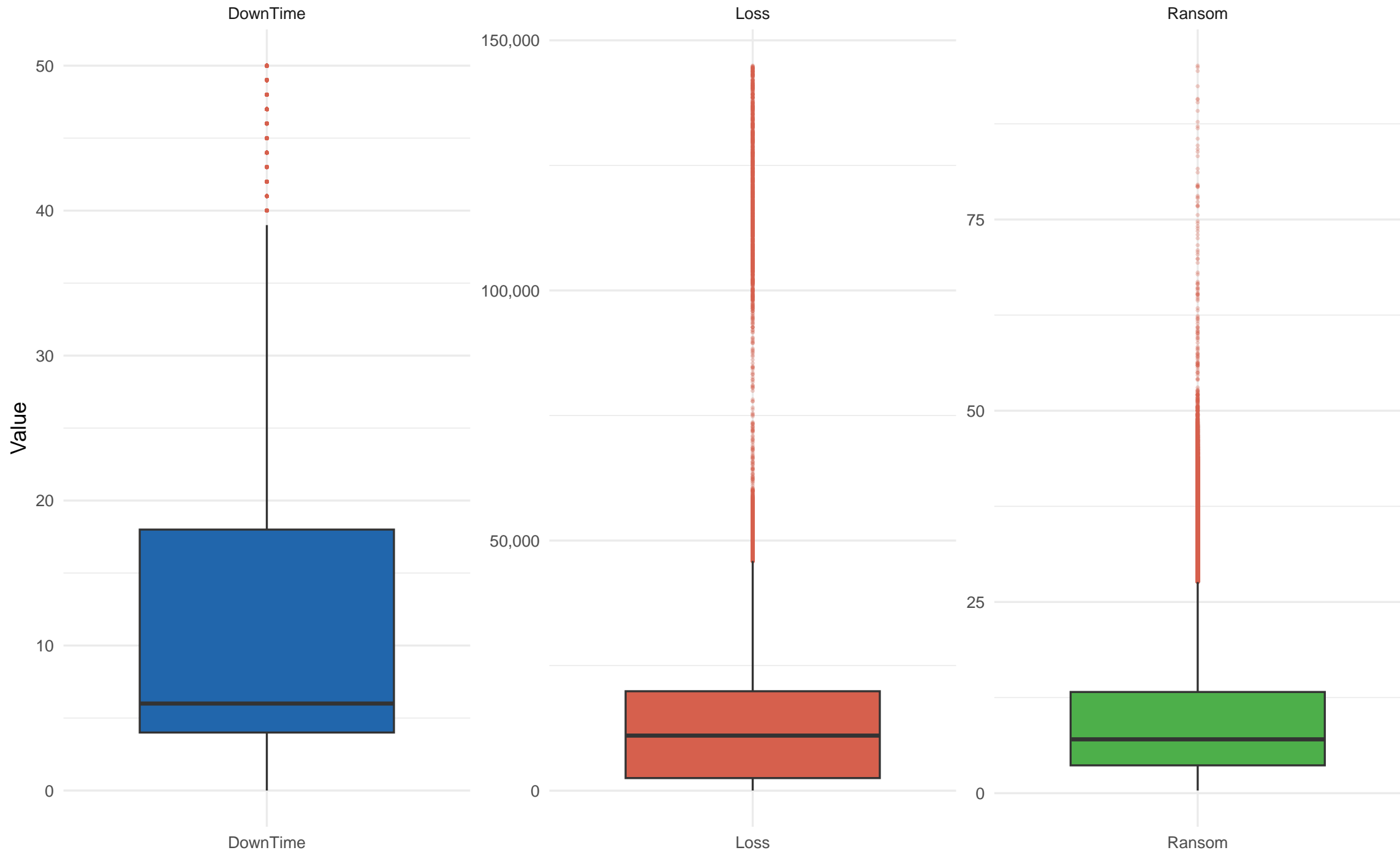# Duplicate Row Detection

Exact row–level duplicates in raw merged data
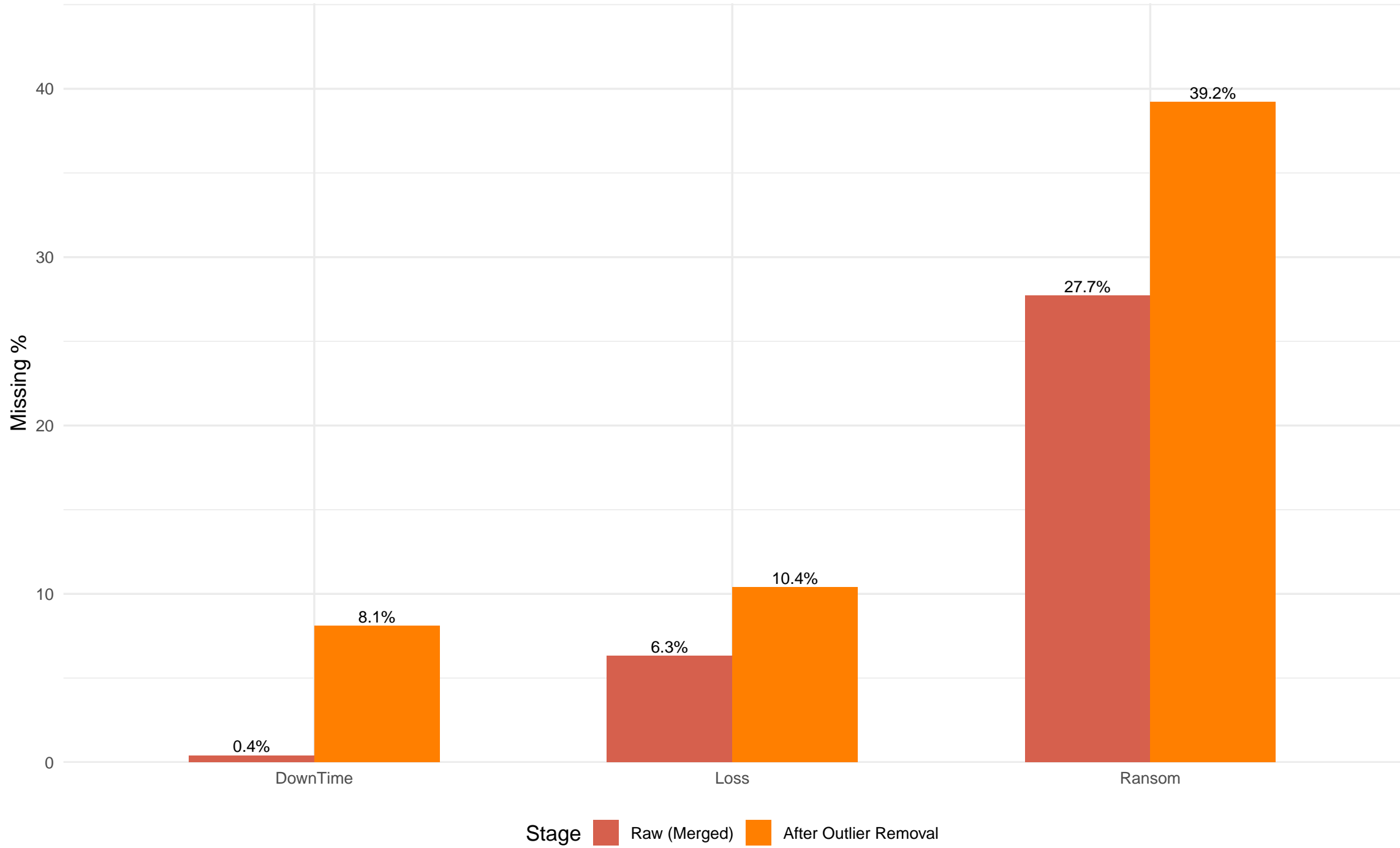
Numeric Variables: Boxplot After Format Conversion
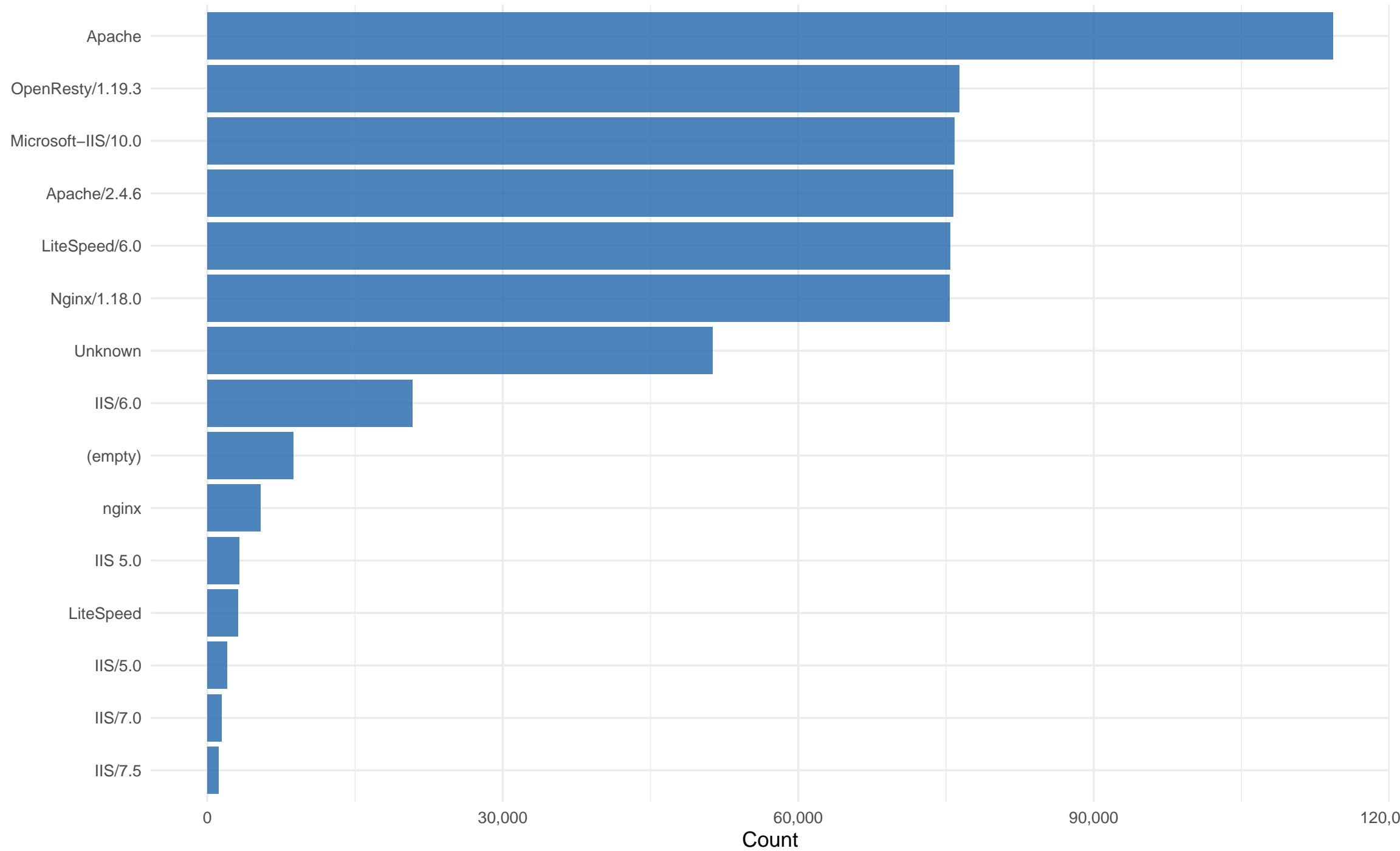Red dots indicate potential outliers (after IQR−based detection)

Missing Values: Raw vs After Outlier Removal

Outlier removal converts extreme values to NA, increasing missing rate

Stage ▮ Raw (Merged) ▮ After Outlier Removal

Raw WebServer: Top 15 Values (Before Cleaning)

High cardinality with inconsistent naming

Count

Country: Inconsistent Naming Examples

Multiple representations of same country need standardization