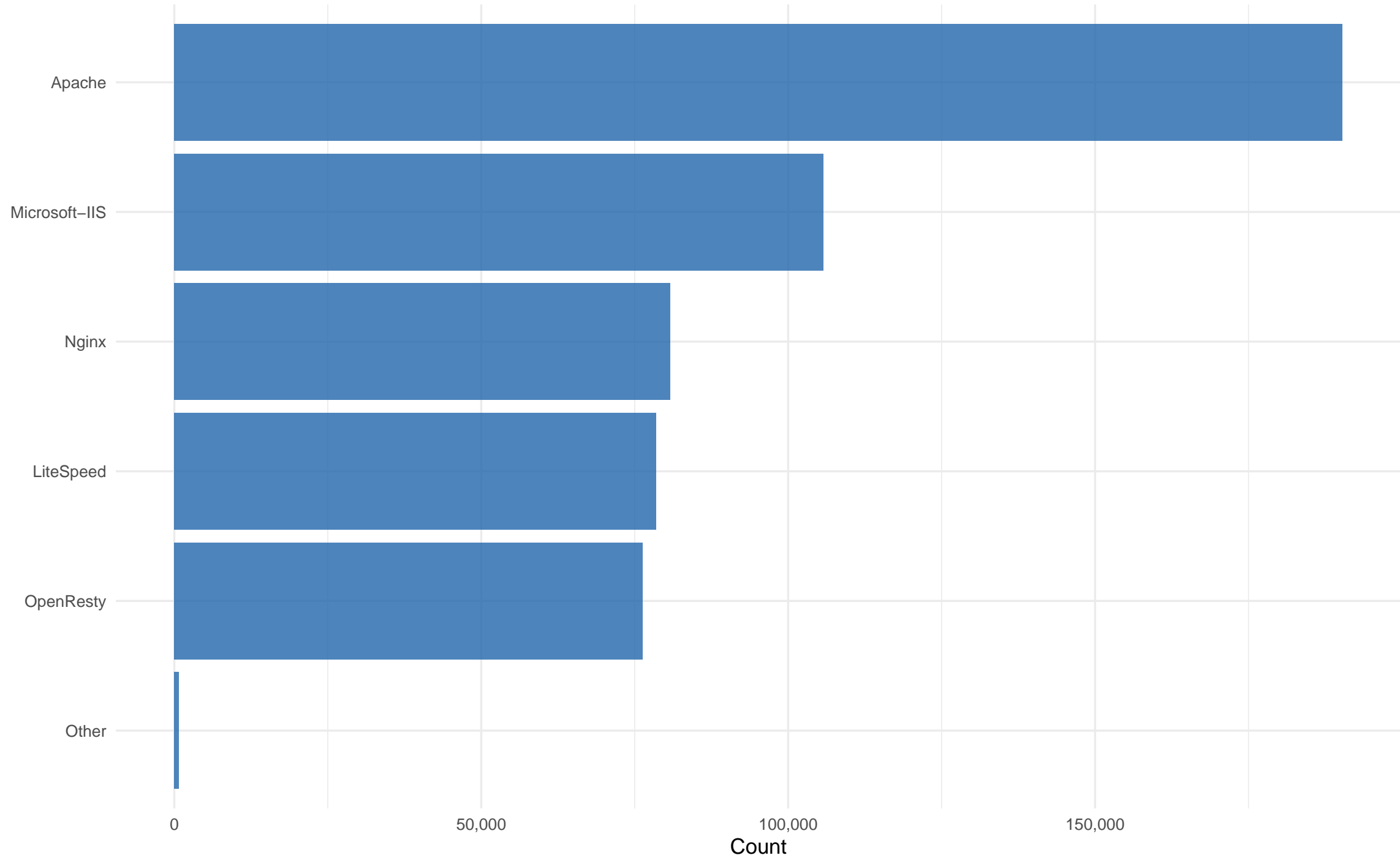


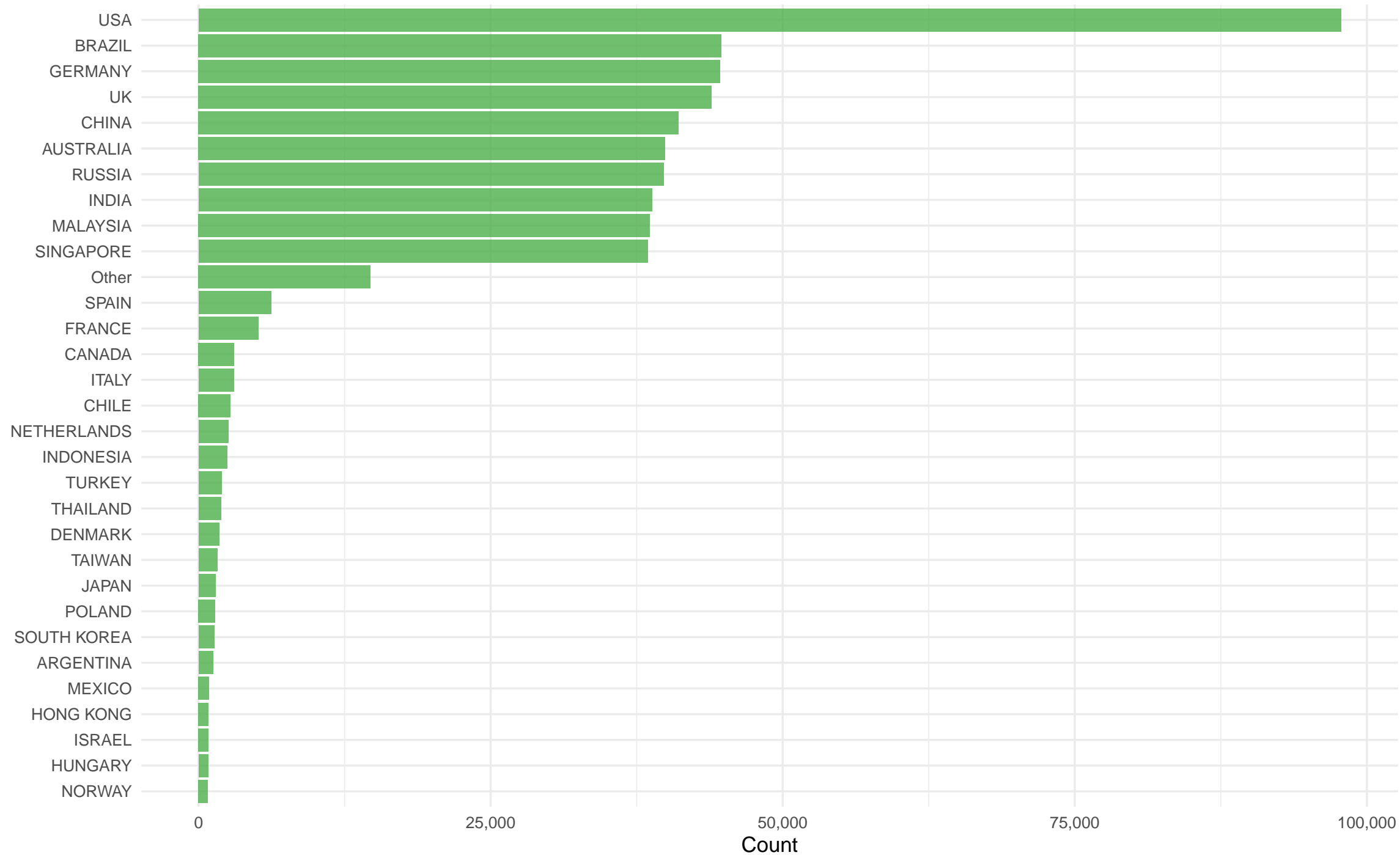
WebServer After Cleaning (6 categories)

Reduced from 176 raw values; 60,295 NA



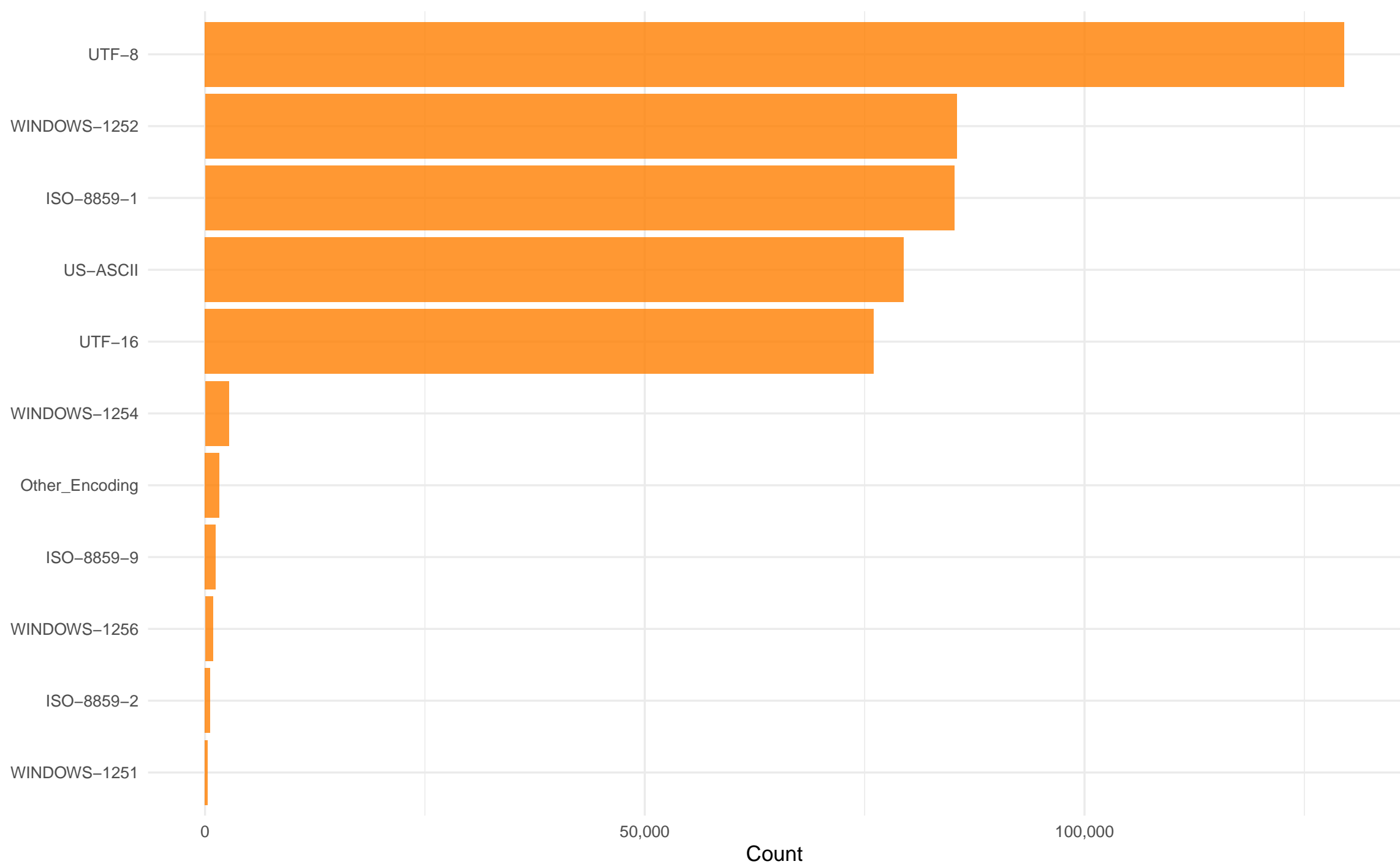
Country After Cleaning (31 categories)

Standardized from ~243 raw variations; 68,202 NA



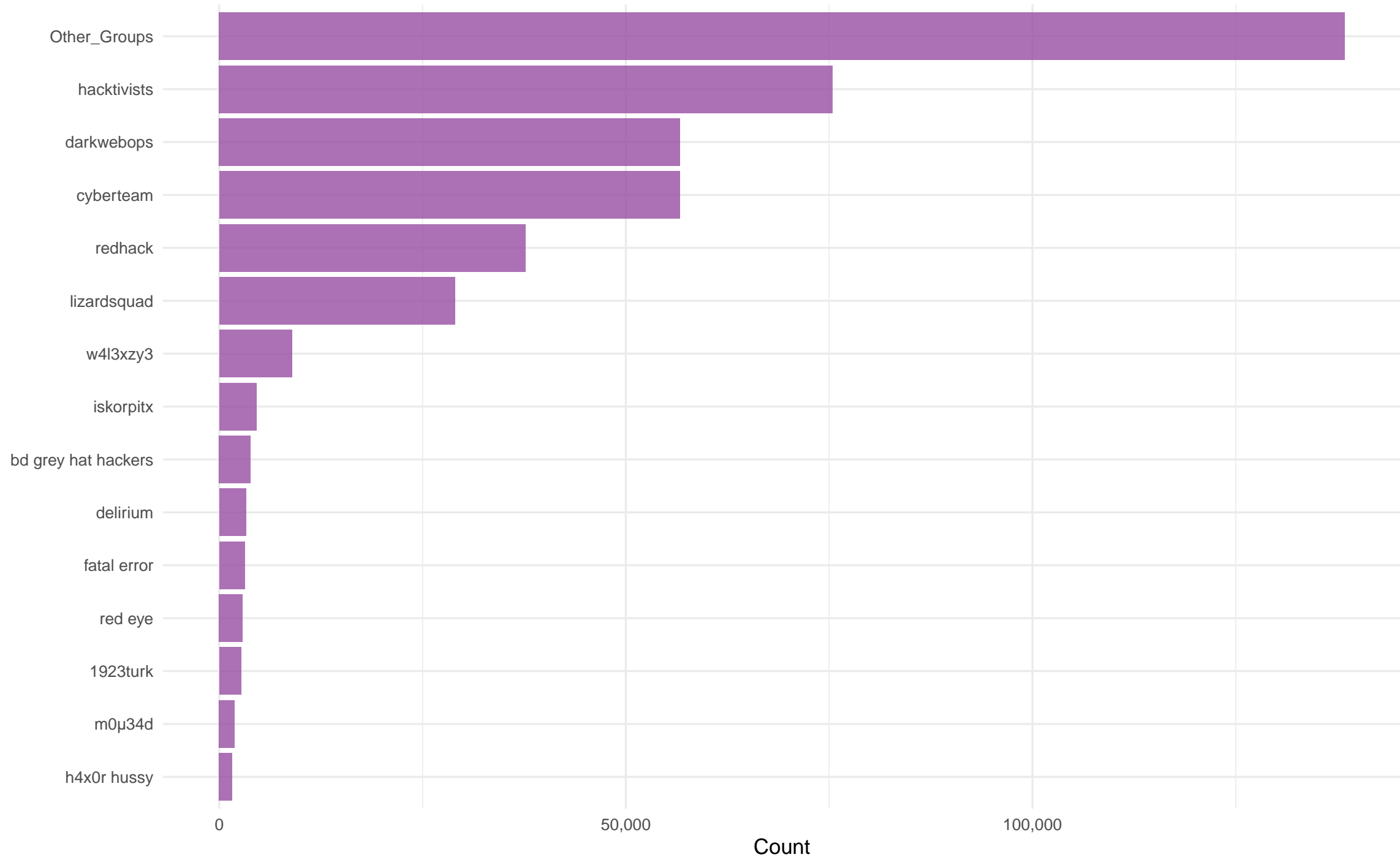
Encoding After Cleaning (11 categories)

Standardized from ~37 raw variations; 129,864 NA



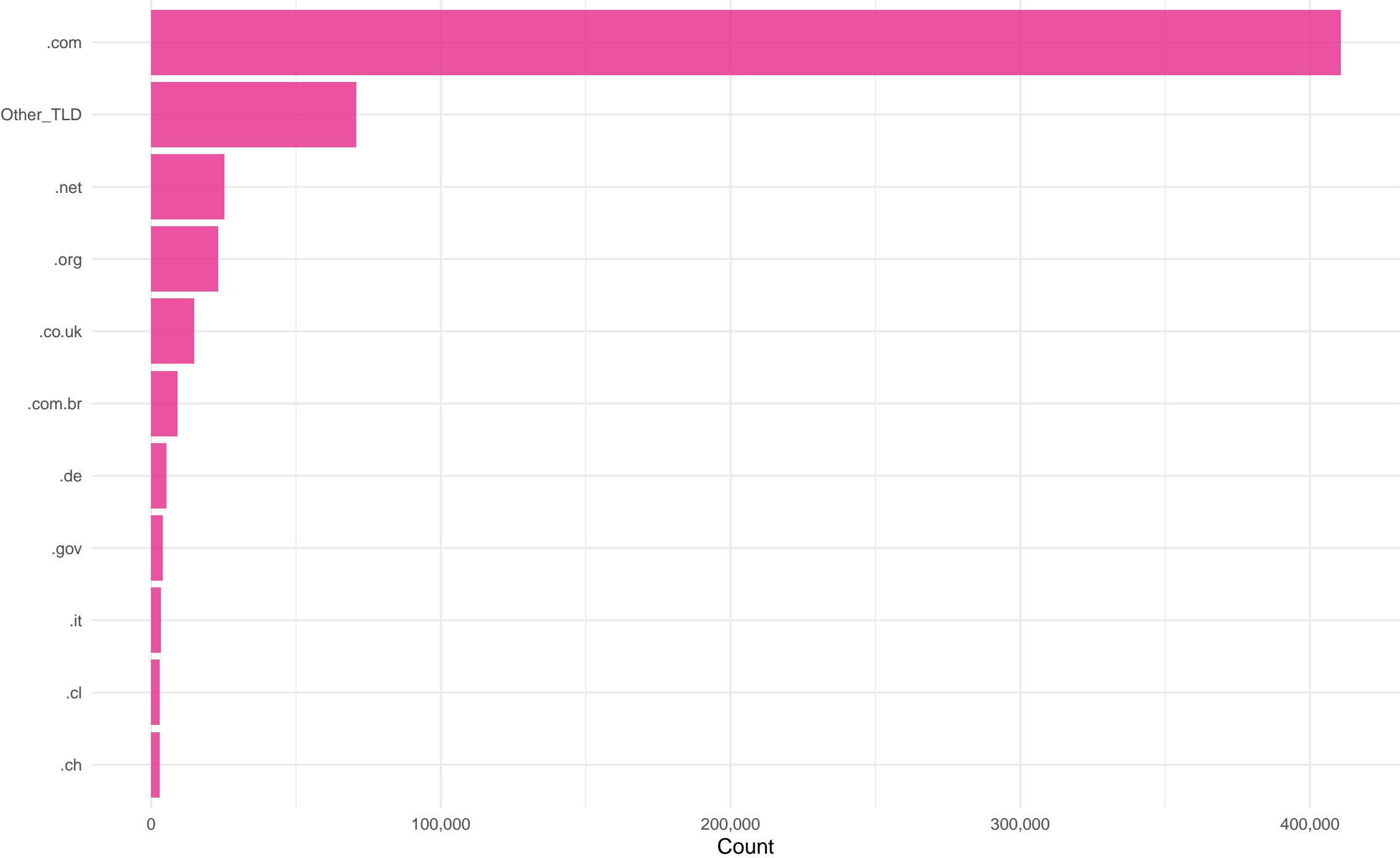
Notify After Cleaning (Top 15 of 51 categories)

Reduced from 9,905 raw values; Anonymous/Unknown -> NA (125,595)



URL Suffix (TLD) Distribution (11 categories)

Extracted from domain; NA: 21,195



Categorical Variable Cardinality: Before vs After Cleaning

Y-axis on log scale; cleaning reduces high-cardinality to manageable levels

