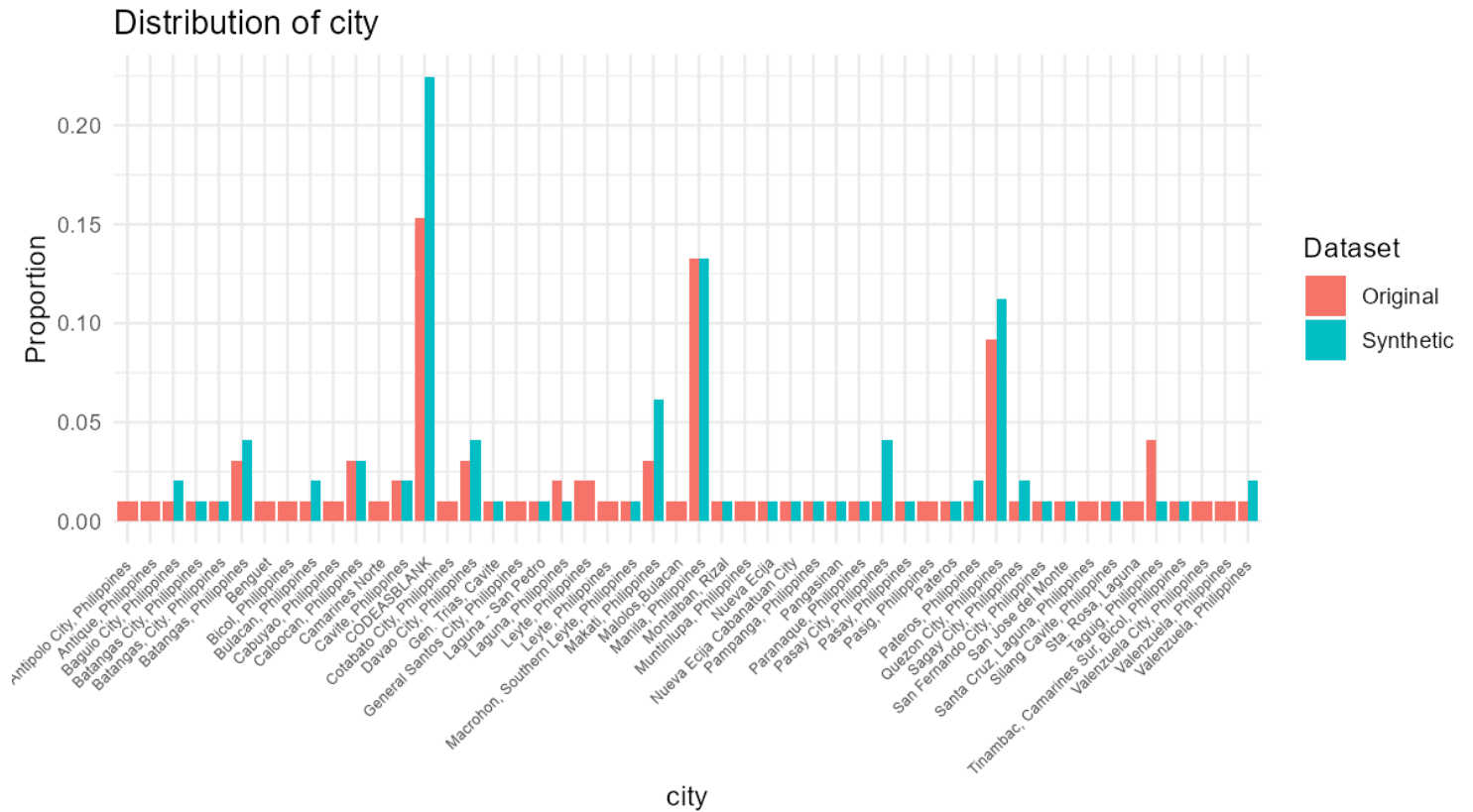


All Variable Distributions

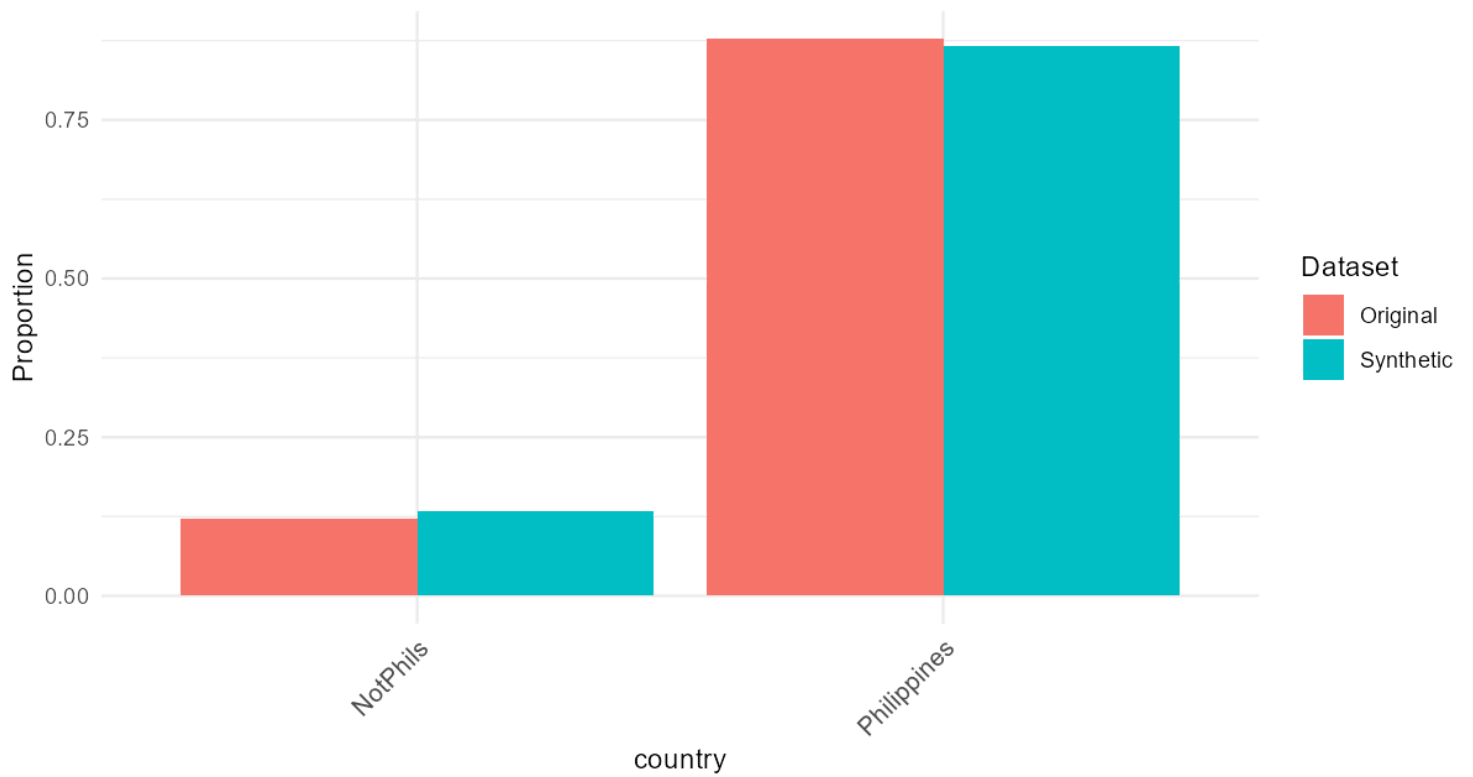
Total variables plotted: 19

city



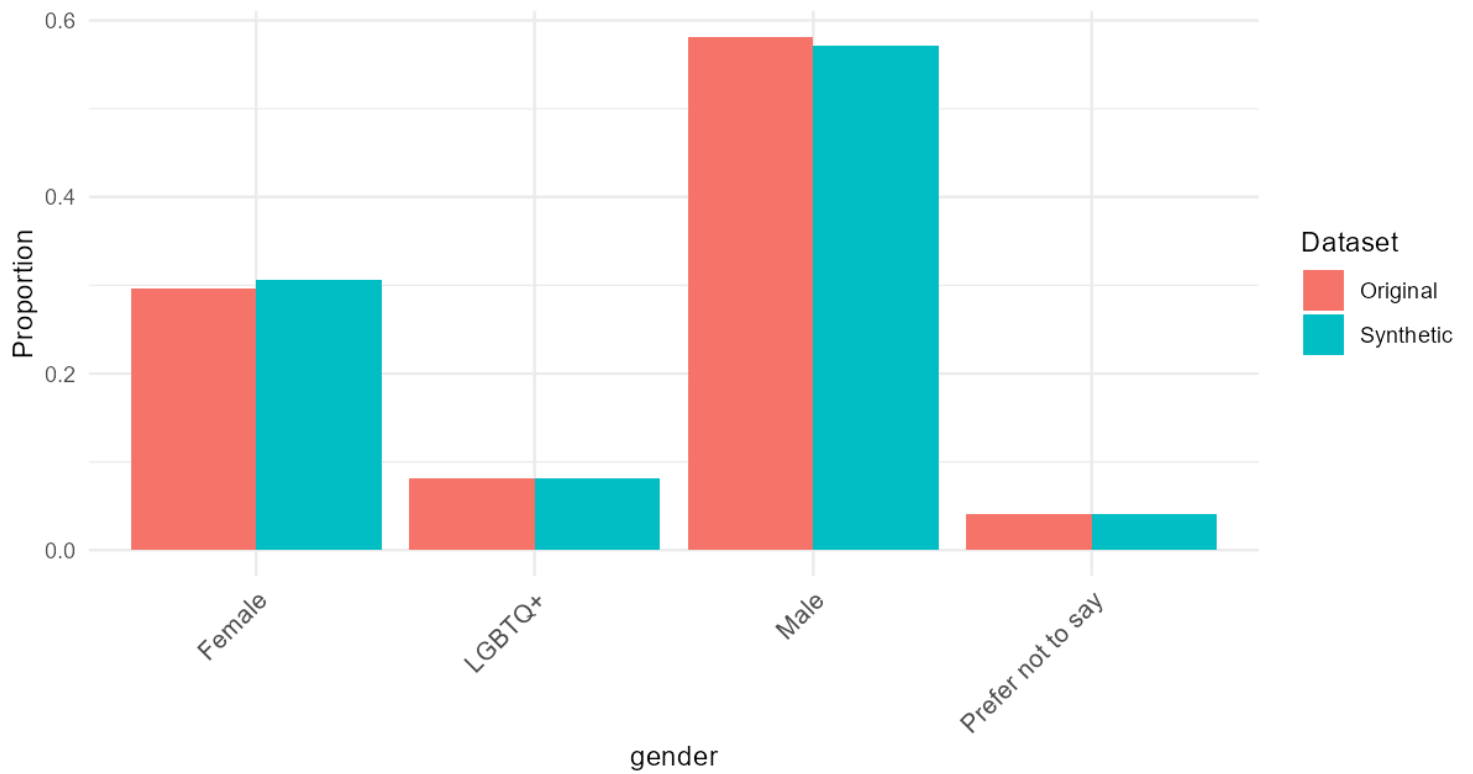
country

Distribution of country



gender

Distribution of gender



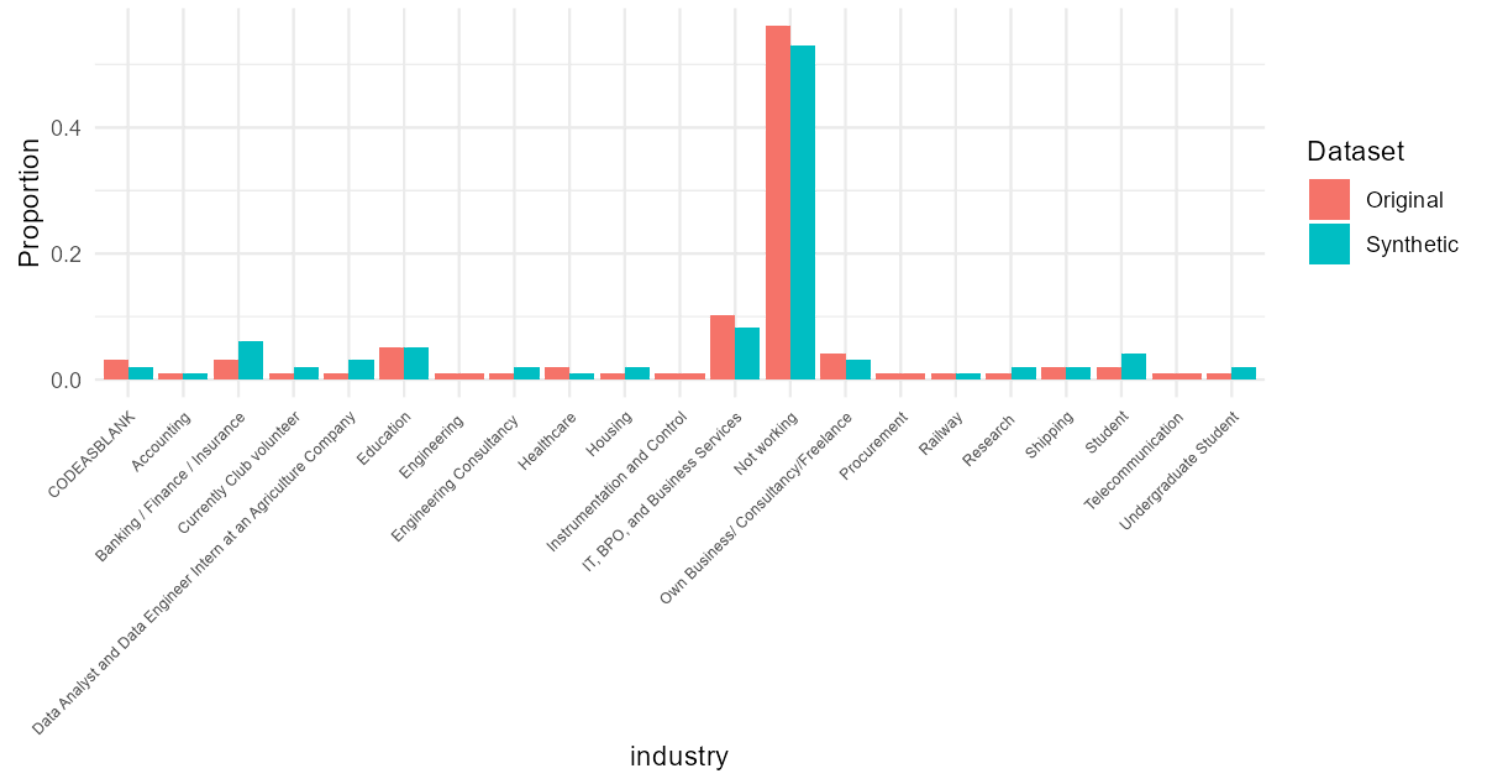
educstat

educstat	Original	Synthetic
College degree	0.83	0.80
Master's degree	0.03	0.05
Other - specify	0.04	0.04
Secondary school student	0.09	0.10

[illegible]

industry

Distribution of industry



generaltools

Distribution of generaltools



whatused

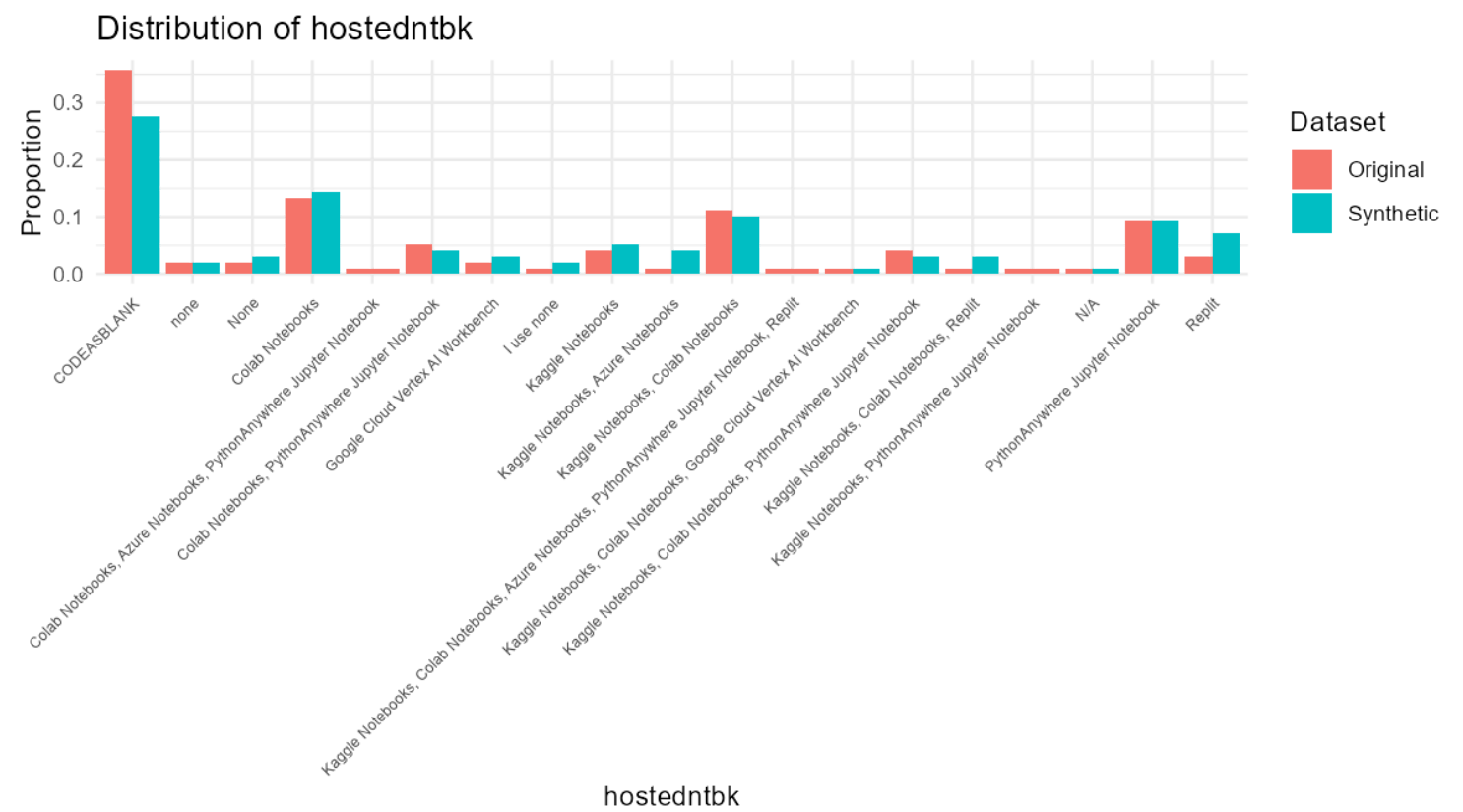
Dataset

- Original
- Synthetic

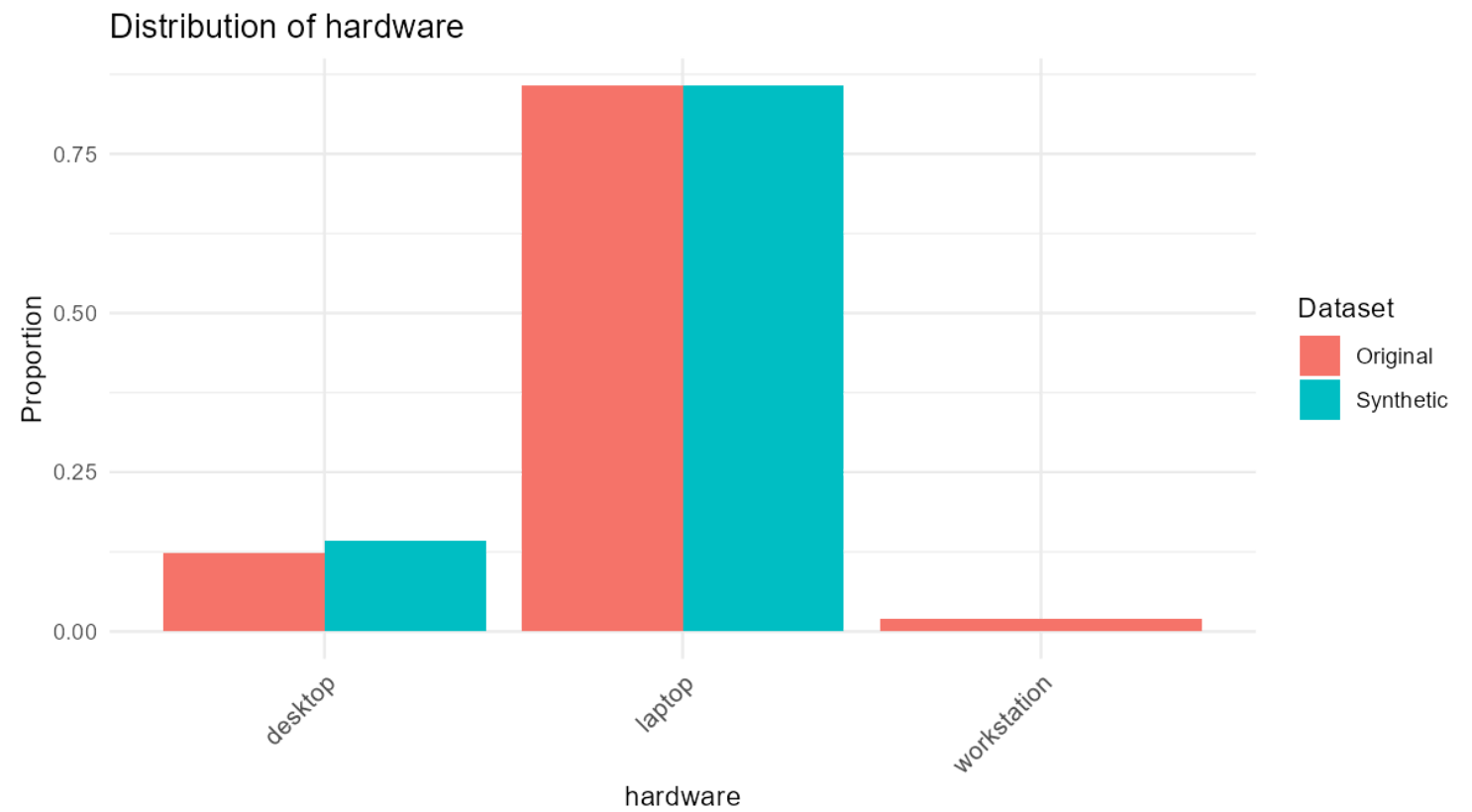
useai

Bar chart showing the proportion of 'useai' for various datasets, categorized by 'Original' (red) and 'Synthetic' (teal) data. The y-axis represents the proportion from 0.0 to 0.2. The x-axis lists 30 different datasets. The chart shows that for many datasets, the proportion of 'useai' is low, but there are notable peaks for 'Original' data in 'CODEASBLANK' and 'Chatgpt, Claude, Gemini' (approx. 0.25), and for 'Synthetic' data in 'Chatgpt, Claude, Gemini' (approx. 0.28) and 'Chatgpt, Copilot' (approx. 0.17).

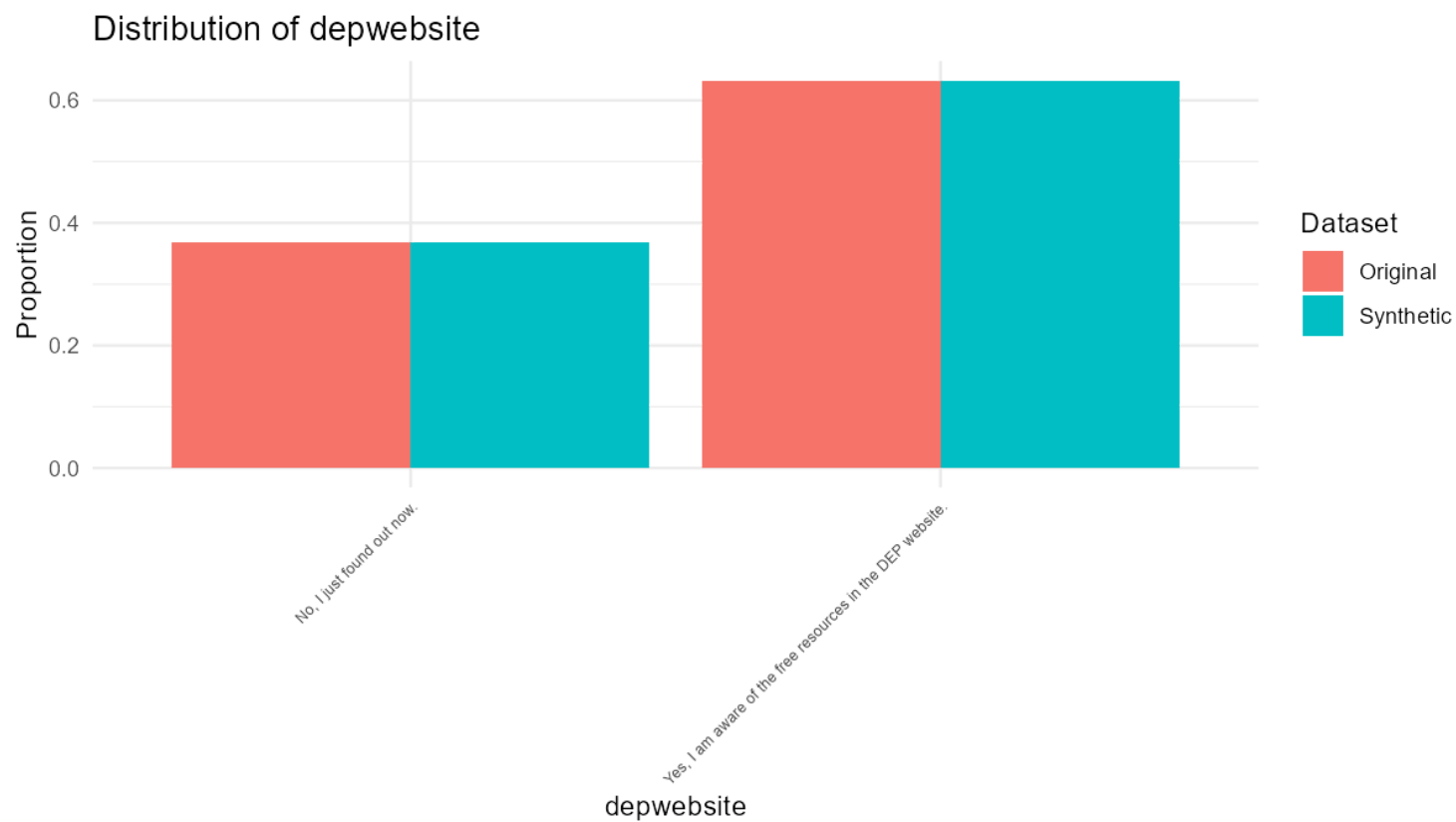
hostedntbk



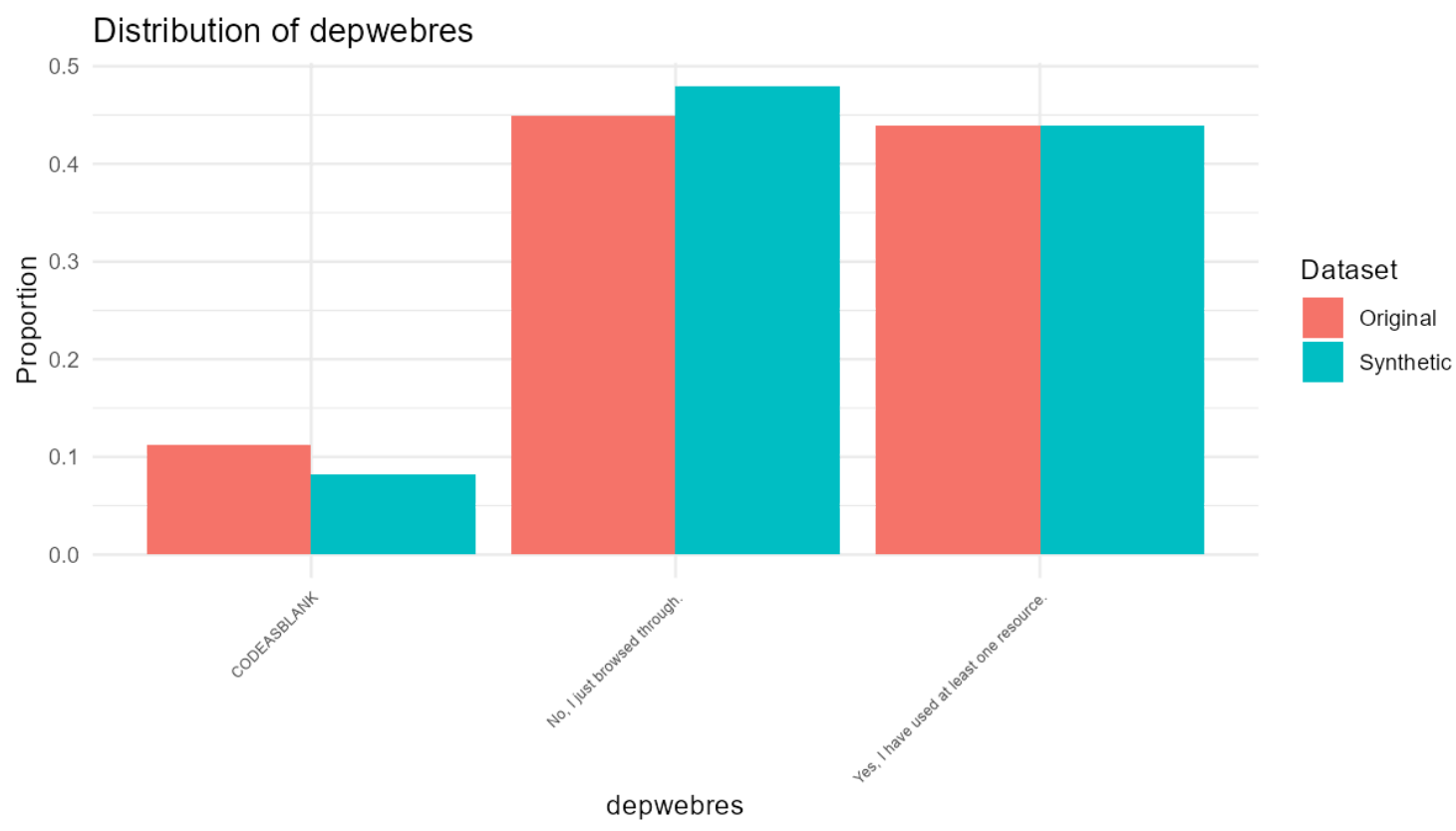
hardware



depwebsite



depwebres

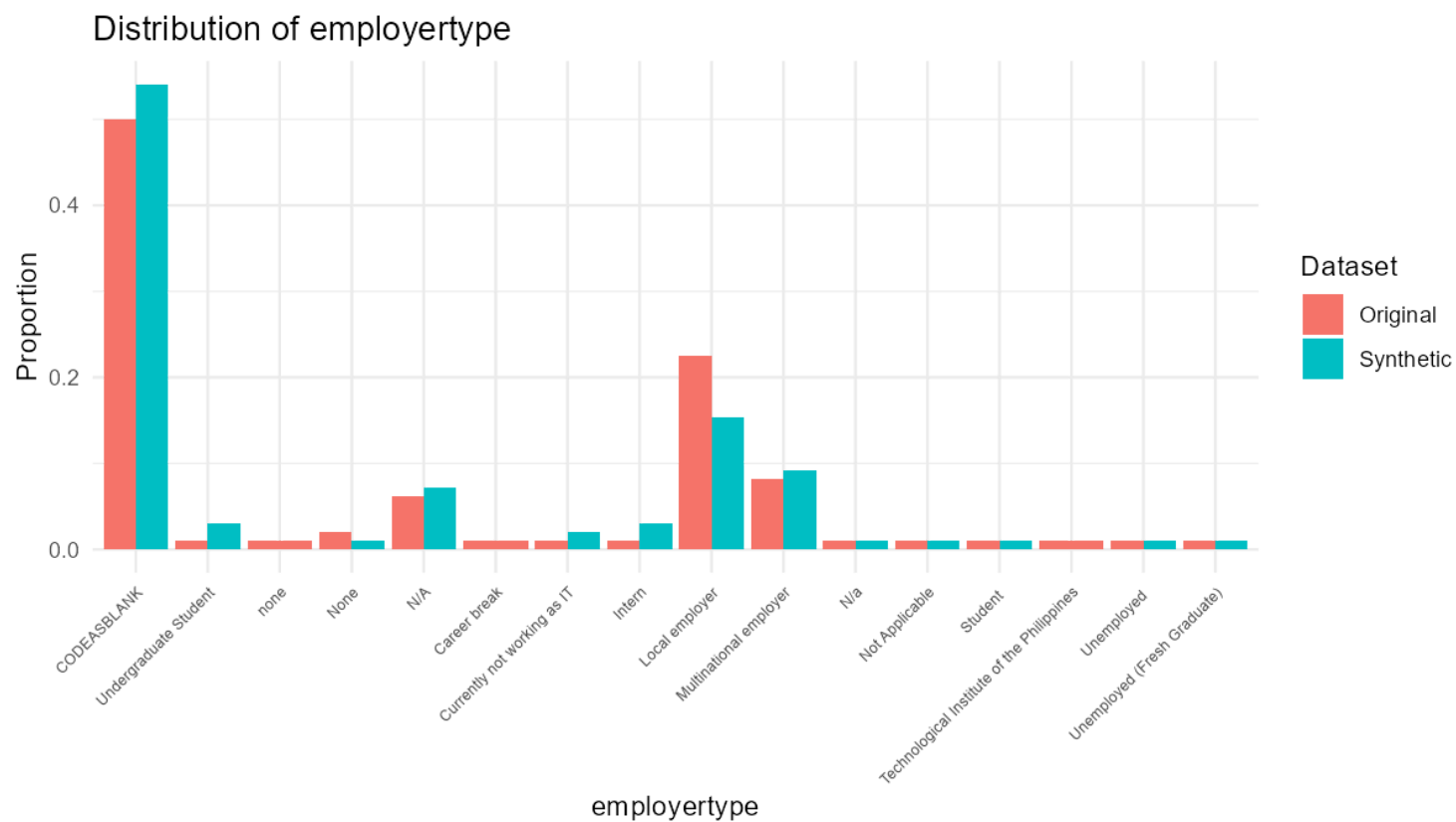


otherfb

volunteer

comms

employertype



age_grp

