Overview

Dataset Statistics

Number of Variables	3
Number of Rows	960
Missing Cells	0
Missing Cells (%)	0.0%
Duplicate Rows	0
Duplicate Rows (%)	0.0%
Total Size in Memory	851.0 KB
Average Row Size in Memory	907.7 B
Variable Types	Categorical: 2 Numerical: 1

Dataset Insights

`seqs` has a high cardinality: 960 distinct values	High Cardinality
`labels` has constant length 1	Constant Length
`seqs` has all distinct values	Unique

Variables

seqs categorical Show Details	**Approximate Distinct Count**	960	
	Approximate Unique (%)	**100.0%**	
	Missing	0	
	Missing (%)	0.0%	
	Memory Size	843.4 KB	

seqs — Count vs seqs (ATTGCSC..., GCIPKHK..., GLWIKGN..., IIECFFS..., KKKCIAK..., MAAAAAA..., MAAAAAV..., MAAACGP..., MAADNTG..., MAADSEP...)

Top 10 of 960 seqs

Approximate	593	**Distinct**	

seqs_length numerical	Count		Mean	420.5552
	Approximate Unique (%)	61.8%	Minimum	40
	Missing	0	Maximum	1021
	Missing (%)	0.0%	Zeros	0
	Infinite	0	Zeros (%)	0.0%
	Infinite (%)	0.0%	Negatives	0
	Memory Size	15.0 KB	Negatives (%)	0.0%

Show Details

labels categorical		
	Approximate Distinct Count	2
	Approximate Unique (%)	0.2%
	Missing	0
	Missing (%)	0.0%
	Memory Size	61.9 KB

Show Details

Interactions

Correlations

Pearson	Spearman	KendallTau

Missing Values

Bar Chart	Spectrum	Heat Map	Dendogram

Report generated with DataPrep