Column	Meaning	What It Shows When Plotted	Use Case
matches	Absolute count of keyword matches for a category in an article.	Which categories or articles have the most keyword mentions (raw volume).	Raw content analysis; not ideal for comparing across texts of different lengths.
percentage / prop_module	Percentage of that category within its module in the same article.	Which category dominates its module within each article.	Internal focus comparison within modules; visualizes theme balance.
prop_article	Percentage of the article's total matches that belongs to this category.	Which theme dominates the article's overall narrative.	Identifying the main thematic focus of each article.
total_matches_article	Total number of keyword matches in the article.	How thematically 'rich' or dense the article is in terms of recognized terms.	Evaluating overall thematic content per document.
module_total	Sum of matches for all categories within a module in that article.	How much a module was discussed in an article.	Tracking which broader themes are emphasized per article.
total_words	Total word count of the article (cleaned text).	Length of the processed article – used for normalization.	Reference measure for normalization; supports density metrics.
matches_per_1000w	Number of matches per 1,000 words – a normalized density measure.	Topic density adjusted for article length. Great for fair comparison.	Fair thematic intensity comparison across articles of different size.

Analytical Question	Recommended Metric	Best Chart Type
Which themes are most frequently mentioned (in raw numbers)?	matches	Barplot, boxplot
Which category dominates each module within articles?	prop_module	Stacked barplot, facet plots
Which topic dominates the overall content of each article?	prop_article	Barplot, heatmap
Which articles are the most thematically rich overall?	total_matches_article	Histogram, barplot by article
Which broader module is most discussed in each article?	module_total	Facet barplot per article or module
Which themes are most intensely discussed (adjusted for article length)?	matches_per_1000w	Heatmap, ranking, barplot
Which articles have similar thematic structures or profiles?	All combined metrics	Clustering, multidimensional scaling
How balanced or diverse is an article's coverage across categories or modules?	prop_article, module_total	Entropy plots, diversity indices
Are there clusters of articles based on thematic focus?	prop_article, matches_per_1000w	PCA, t-SNE, dendrograms