$\label{eq:hdbscan} \begin{aligned} &\text{HDBSCAN cluster sizes of tokens } x_i \\ (\texttt{imdb_albert-xlarge-v2_params-random_dense-on_layers-48}, \text{ sample } 2) \end{aligned}$

