

HDBSCAN cluster sizes of tokens  $x_i$   
(wikitext\_bert-large-uncased\_params-random\_dense-on\_layers-24, sample 7)