

HDBSCAN cluster sizes of tokens  $x_i$   
(randomtext\_bert-large-uncased\_params-trained\_dense-on\_layers-24)

