

HDBSCAN cluster sizes of tokens  $x_i$   
(wikitext\_albert-xlarge-v2\_params-trained\_dense-on\_layers-48, sample 4)

