

HDBSCAN cluster sizes of tokens  $x_i$   
(uniform\_albert-xlarge-v2\_params-trained\_dense-off\_layers-48, sample 8)

