

HDBSCAN cluster sizes of tokens  $x_i$   
(imdb\_bert-large-uncased\_params-random\_dense-off\_layers-24, sample 3)

