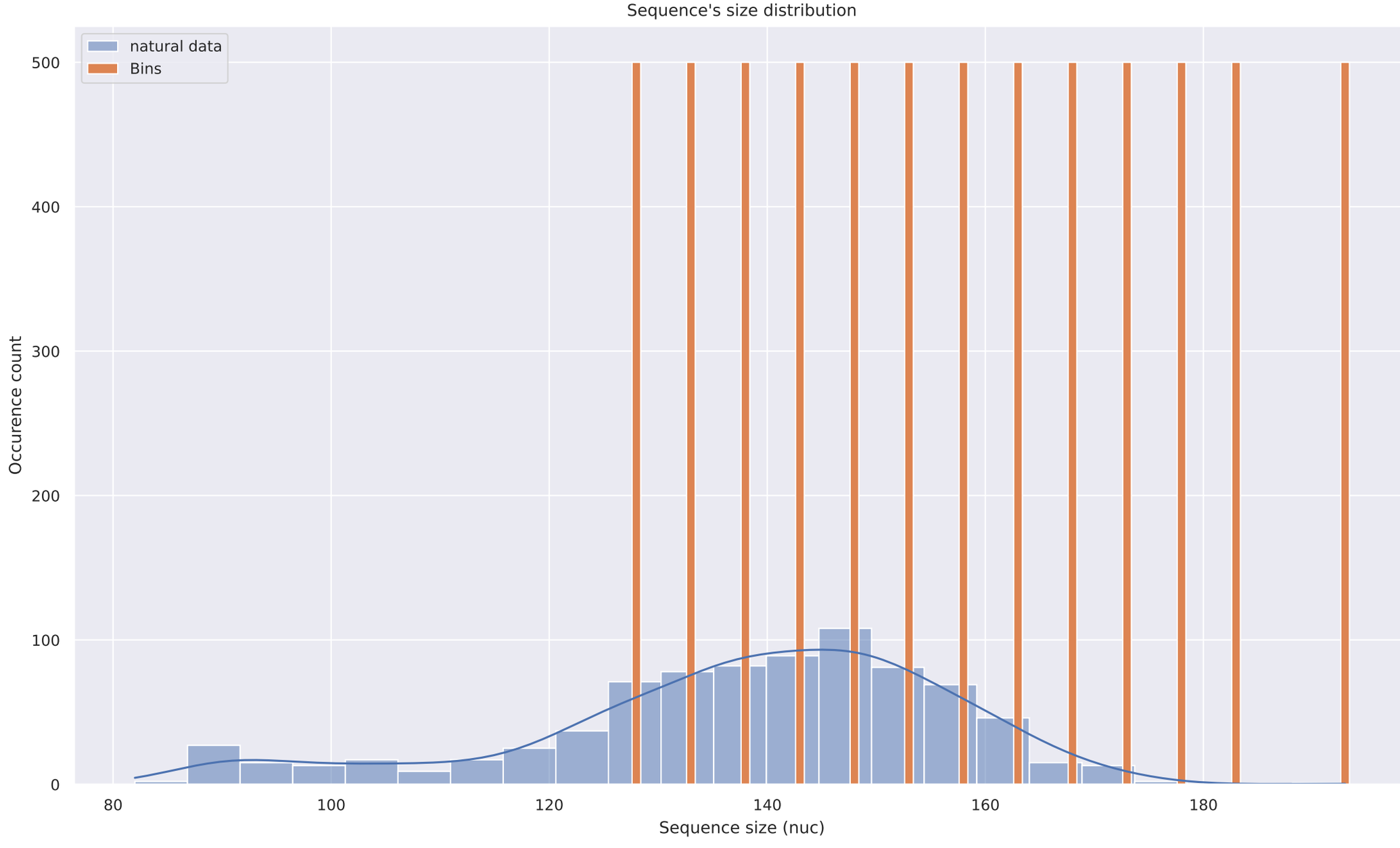
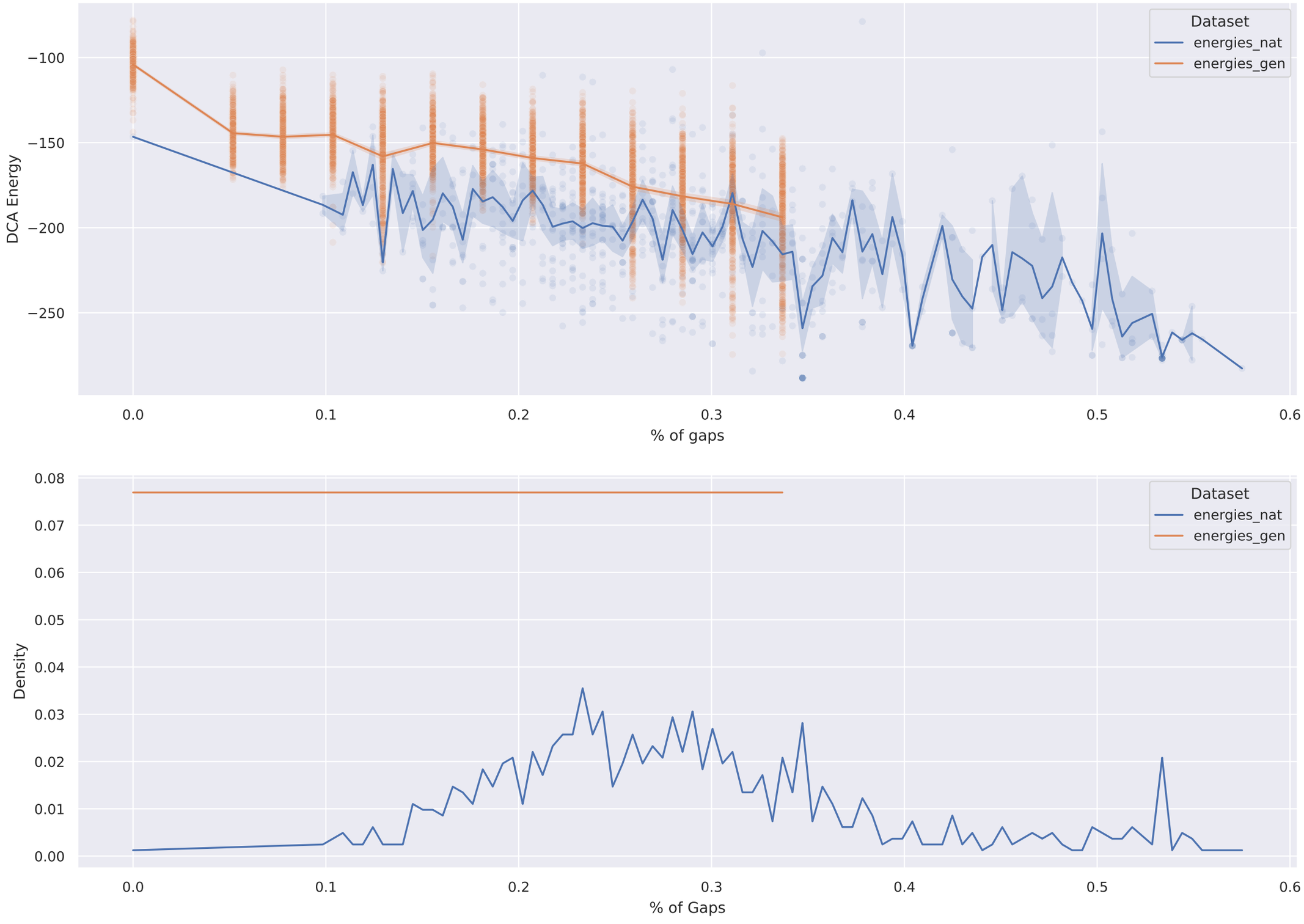
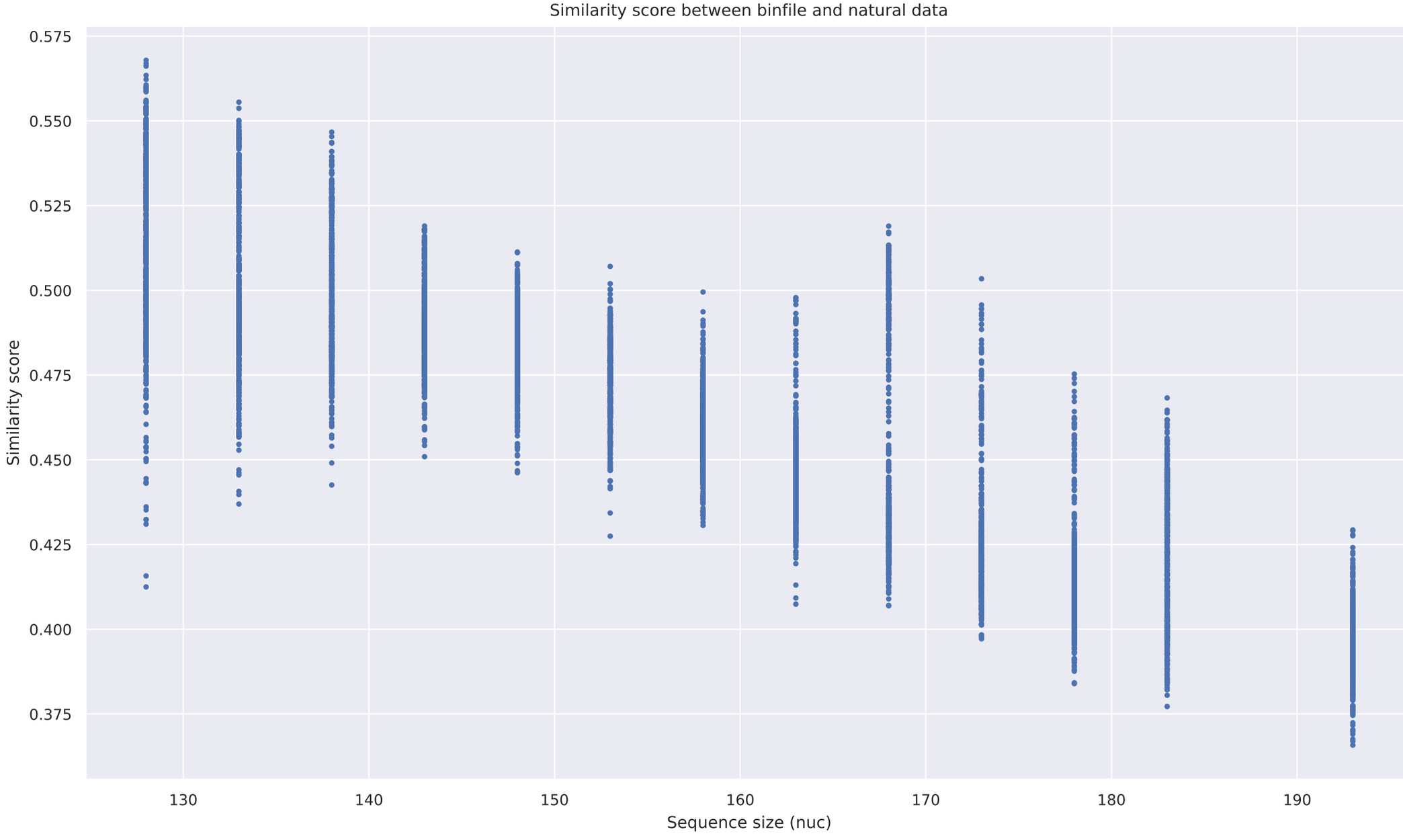

Sequence's size distribution
natural data
Bins
Occurence count
0
100
200
300
400
500
80
100
120
140
160
180
Sequence size (nuc)

Jaccard distribution score depending on size with k=5
Jaccard distance score
0.22
0.21
0.20
0.19
0.18
0.17
0.16
130
140
150
160
170
180
190
Sequence size (nuc)


DCA Energy
−100
−150
−200
−250
0.0
0.1
0.2
0.3
0.4
0.5
0.6
% of gaps
Dataset
energies_nat
energies_gen
Density
0.08
0.07
0.06
0.05
0.04
0.03
0.02
0.01
0.00
0.0
0.1
0.2
0.3
0.4
0.5
0.6
% of Gaps
Dataset
energies_nat
energies_gen


Similarity score between binfile and natural data
Similarity score
0.575
0.550
0.525
0.500
0.475
0.450
0.425
0.400
0.375
130
140
150
160
170
180
190
Sequence size (nuc)