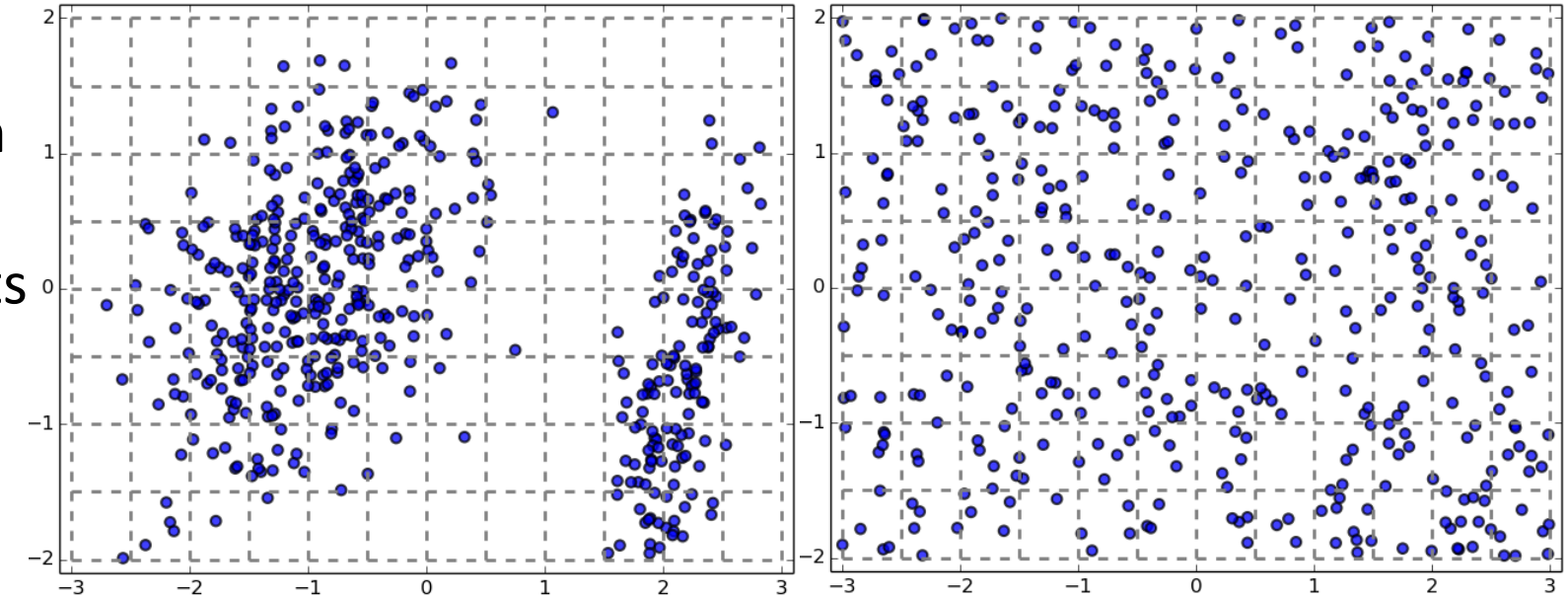
Testing Clustering Tendency: A Spatial Histogram Approach

- **Spatial Histogram Approach:** Contrast the *d*-dimensional histogram of the input dataset ***D*** with the histogram generated from random samples
  - Dataset D is clusterable if the distributions of two histograms are rather different
- Method outline
  - Divide each dimension into equi-width bins, count how many points lie in each cells, and obtain the empirical joint probability mass function (EPMF)



  - Do the same for the randomly sampled data
  - Compute how much they differ using the *Kullback-Leibler* (*KL*) *divergence* value