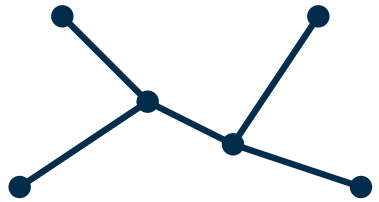
Computational Geometry

EMST & Clustering

Thomas Bläsius

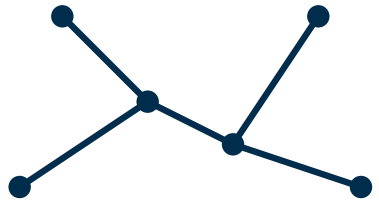
Euclidean MST

Problem: EMST
Let $P \subset \mathbb{R}^2$ be a set of points. Compute a tree $T = (P, E)$ with vertices P such that its total edge length $\sum_{pq \in E} \|p - q\|$ is minimized.



Euclidean MST

Problem: EMST
Let $P \subset \mathbb{R}^2$ be a set of points. Compute a tree $T = (P, E)$ with vertices P such that its total edge length $\sum_{pq \in E} \|p - q\|$ is minimized.



How fast can you compute the EMST?
(assuming a real-RAM)

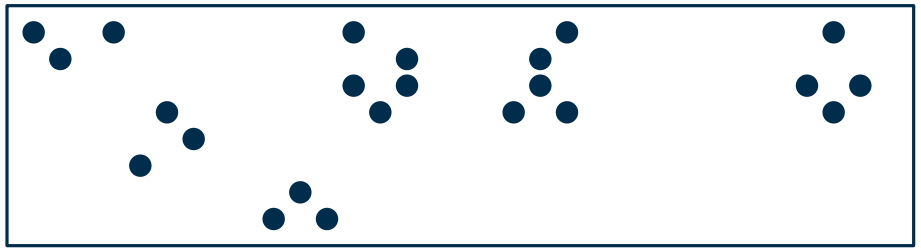
How fast can you compute the EMST?
(assuming integer coordinates and a word-RAM)

Is EMST $\in$ P?

Clustering

Goal

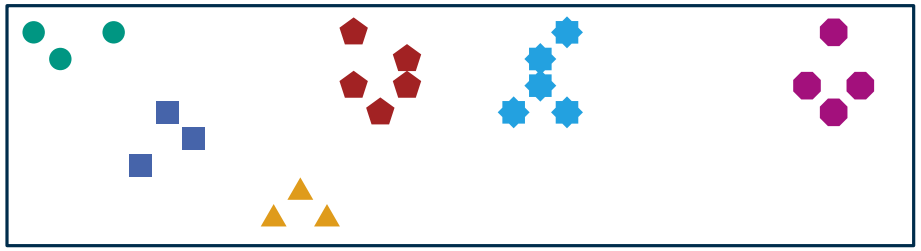
- given a set of points, find a reasonable clustering



Clustering

Goal

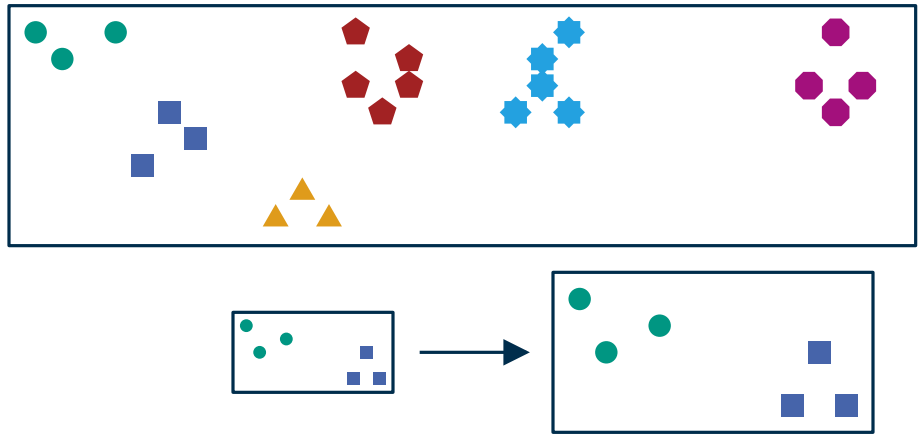
- given a set of points, find a reasonable clustering



Clustering

Goal

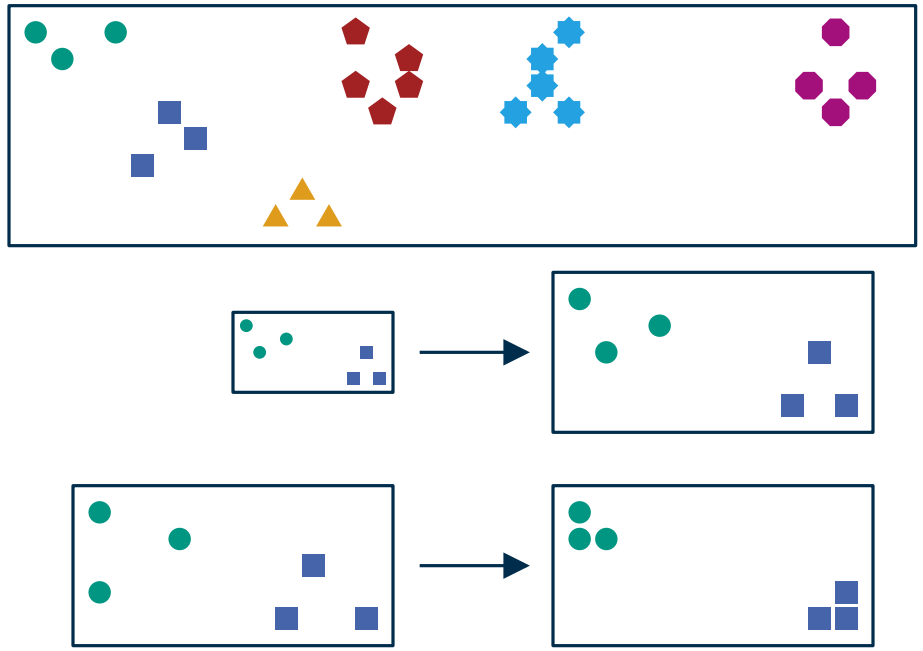
- given a set of points, find a reasonable clustering
- requirements for the algorithm:
  - **scaling invariance:** scaling all distances with the same factor do not change the clustering



Clustering

Goal

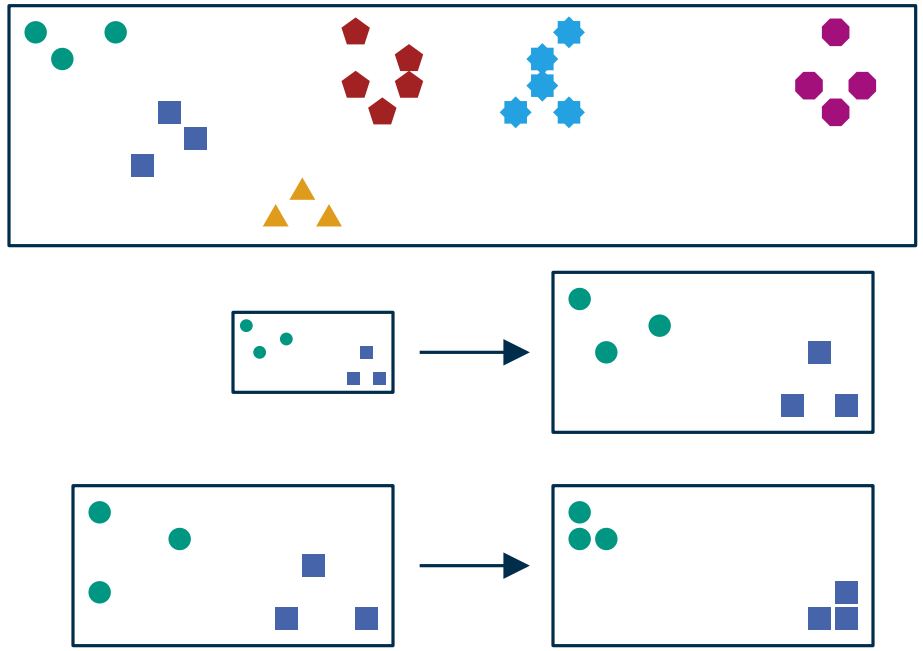
- given a set of points, find a reasonable clustering
- requirements for the algorithm:
  - **scaling invariance:** scaling all distances with the same factor do not change the clustering
  - **consistency:** decreasing distances within clusters and increasing distances between clusters does not change the clustering



Clustering

Goal

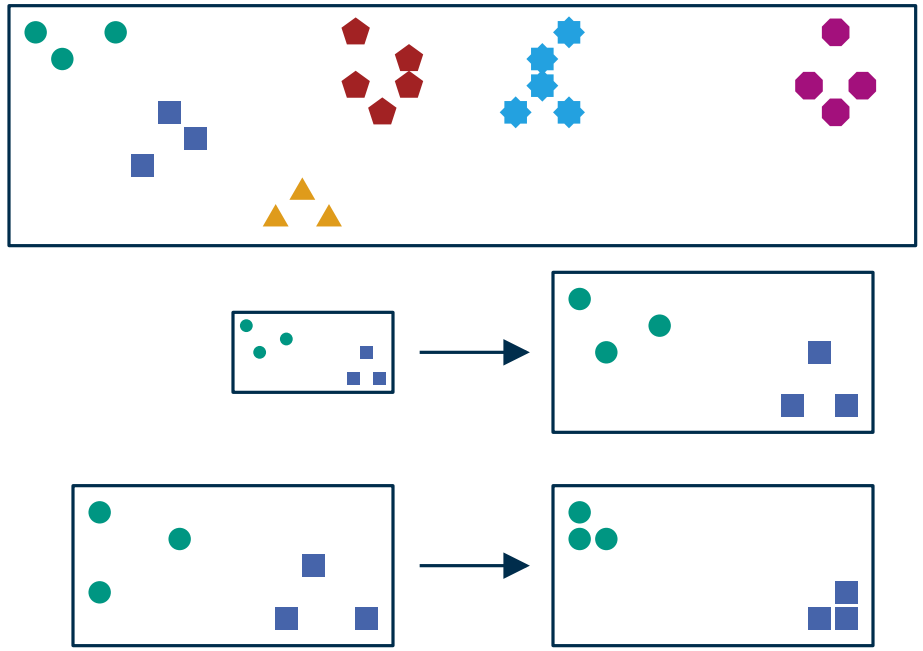
- given a set of points, find a reasonable clustering
- requirements for the algorithm:
  - **scaling invariance:** scaling all distances with the same factor do not change the clustering
  - **consistency:** decreasing distances within clusters and increasing distances between clusters does not change the clustering
  - **richness:** every clustering is in principle possible



Clustering

Goal

- given a set of points, find a reasonable clustering
- requirements for the algorithm:
  - **scaling invariance:** scaling all distances with the same factor do not change the clustering
  - **consistency:** decreasing distances within clusters and increasing distances between clusters does not change the clustering
  - **richness:** every clustering is in principle possible

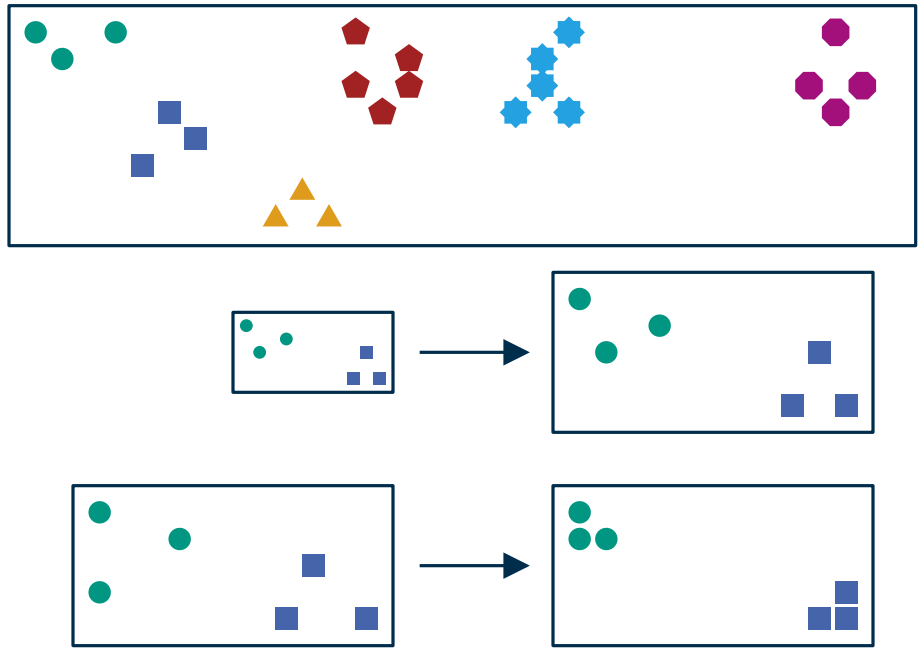


Why can there be no such algorithm?

Clustering

Goal

- given a set of points, find a reasonable clustering
- requirements for the algorithm:
  - **scaling invariance:** scaling all distances with the same factor do not change the clustering
  - **consistency:** decreasing distances within clusters and increasing distances between clusters does not change the clustering
  - **richness:** every clustering is in principle possible



Why can there be no such algorithm?

Which of the three requirements would you drop or relax?