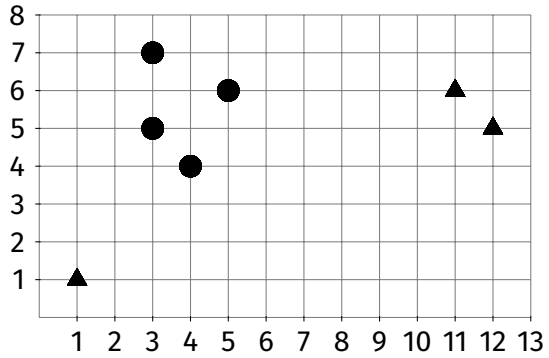
TU Dortmund 13.5.2022
Artificial Intelligence / Data Science and Data Engineering **Machine Learning for Complex Data**
Prof. Dr. Erich Schubert
Prof. Dr. Emmanuel Müller **Summer 2022**
Daniel Wilmes

Assignment 5: k-Medoids Clustering und P3C

Due: Thursday, 19.5.2022

Problem 5-1 *k*-Medians Clustering **9**

Given the following data set with 7 objects (in $\mathbb{R}^2$):



The data set are all points (both triangles and circles) in the above plot. All coordinates are integer.

The three triangles are the data points chosen as *initial cluster centers*.

Cluster this data set using k-Medians using the axis-wise median and Manhattan distance.
As median of an even-sized set, use the average of the two central values ("middle median").

Problem 5-2 P3C **11**

Consider the four-dimensional dataset:

	d1	d2	d3	d4
A	15	12	16	9
B	14	13	18	3
C	12	14	14	15
D	16	13	19	19
E	5	6	9	4
F	4	11	10	18
G	6	17	8	13
H	6	9	14	16
I	14	19	13	15
J	19	3	15	14

a) Determine the relevant dimensions and the relevant bins per dimension.
Use a significance level of $\alpha = 0.35$ with ${\chi^2}_{0.35}(3) = 3.2$; ${\chi^2}_{0.35}(2) = 2.1$; ${\chi^2}_{0.35}(1) = 0.8$;.

b) Merge the relevant cells and determine the p-signatures of dimensionality two.

c) Evaluate which of the p-signatures are cluster cores.