

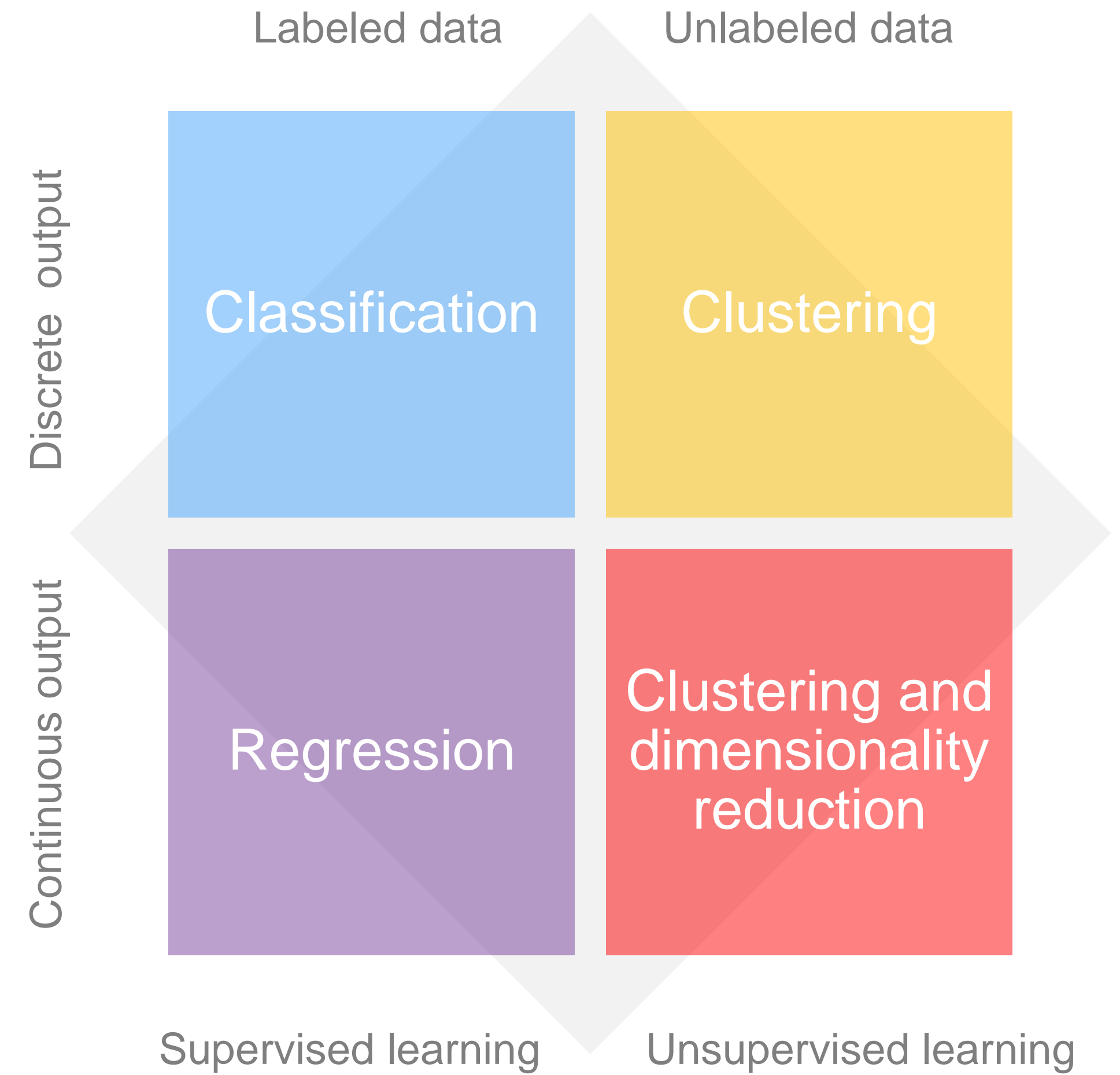
# CREATIVE DATA MINING

Introduction to Unsupervised Machine Learning

09.04.2018

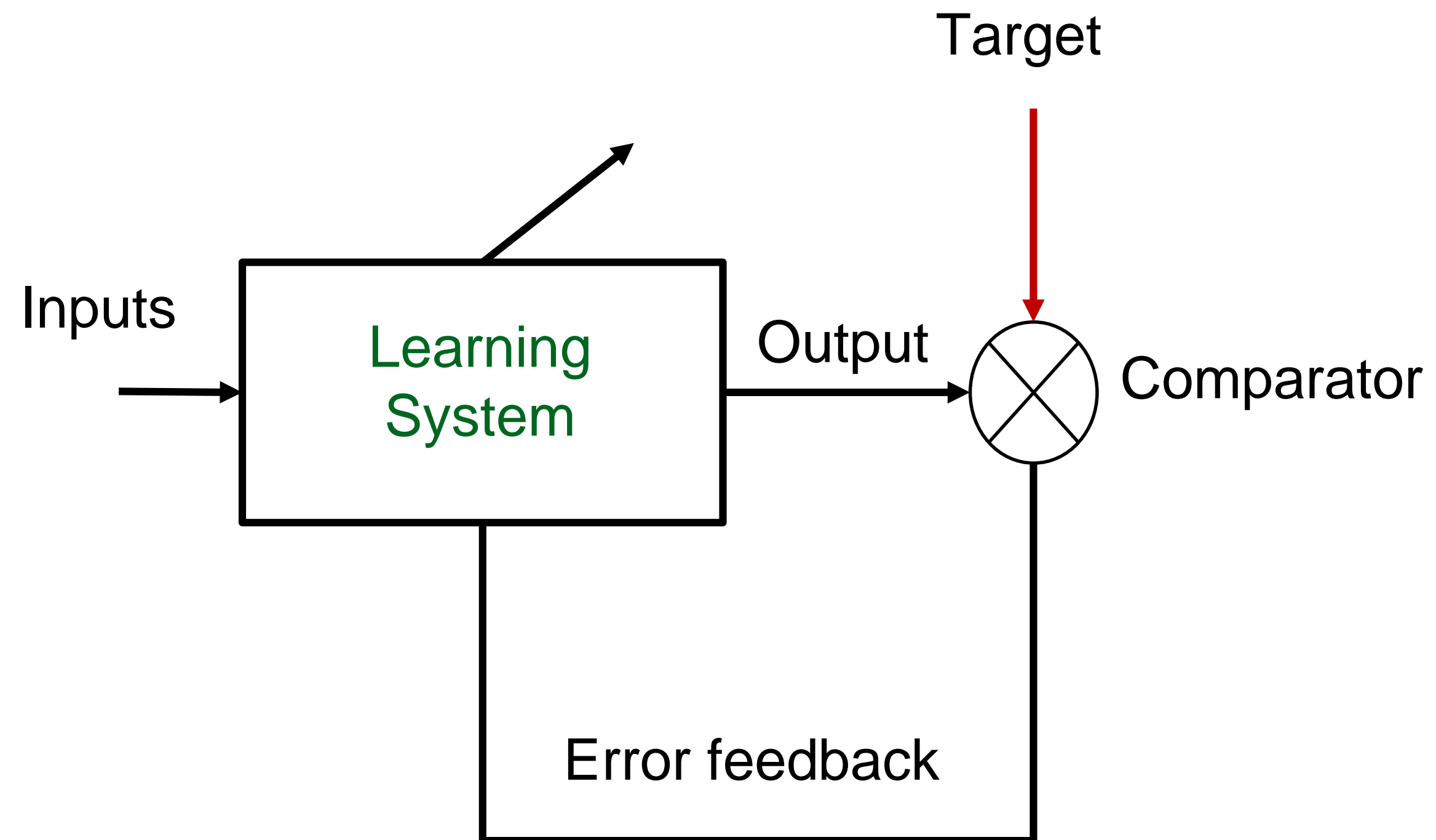
Dr. Varun OJHA

Danielle GRIEGO

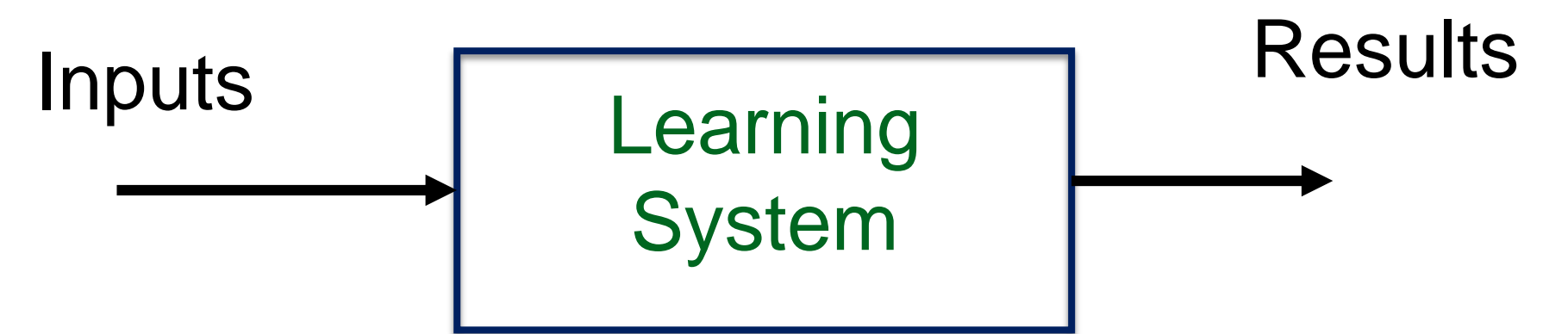


# Learning Systems

## Supervised



## Unsupervised



# Data (unlabeled)

#	Inputs			
	F1	F2	F3	F4
1	76.85	17.27	0.22	34.63
2	76.97	19.54	0.22	34.64
3	77.10	18.51	0.22	34.64
4	85.28	46.09	0.22	34.61
5	85.42	35.83	0.22	34.61
6	88.02	2.59	0.22	34.63
7	77.25	6.34	0.22	34.65
8	77.49	6.98	0.22	34.63
9	85.81	12.18	0.22	34.61

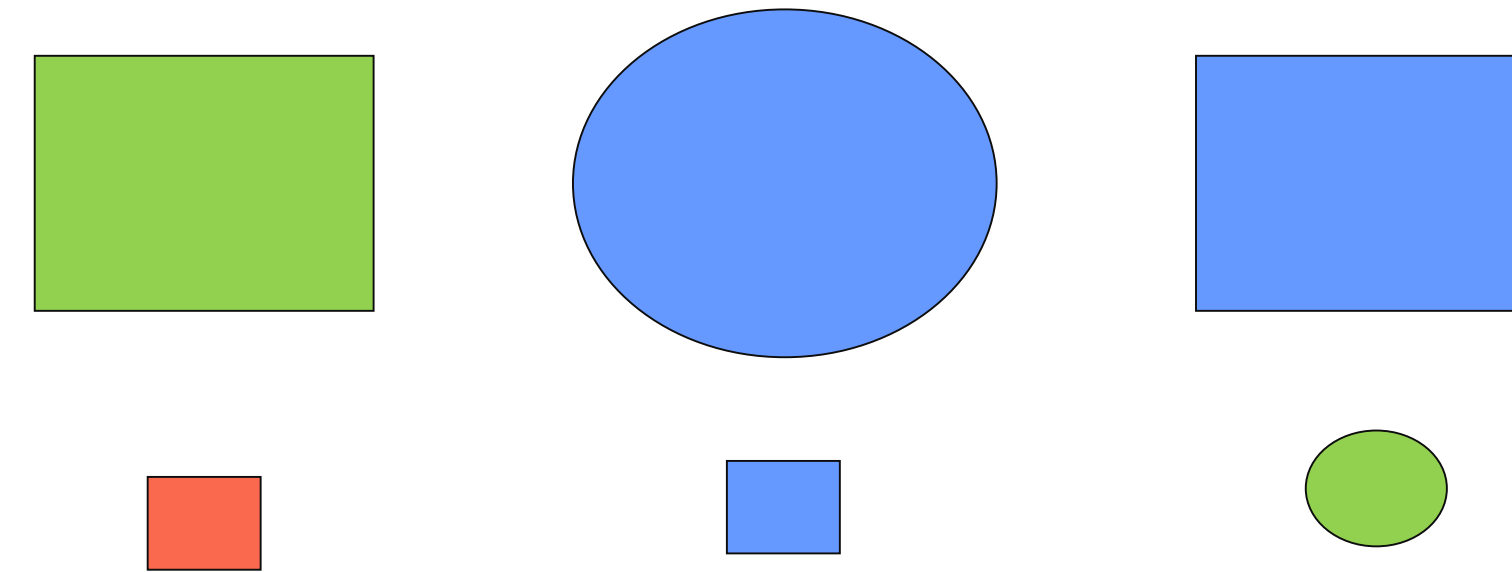
We do not know what is target/output.

We just want to understand some pattern or some sort of grouping of data points.

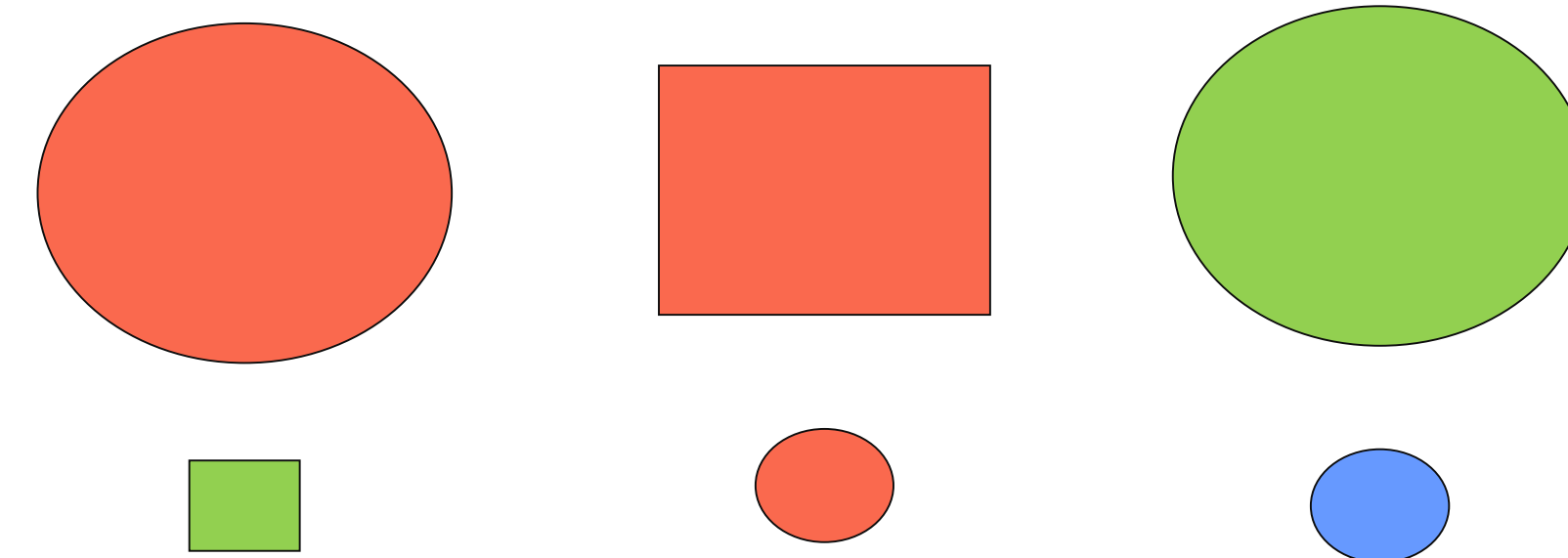
# Data (unlabeled)

## Example: natural concept

Can we form groups of these objects?

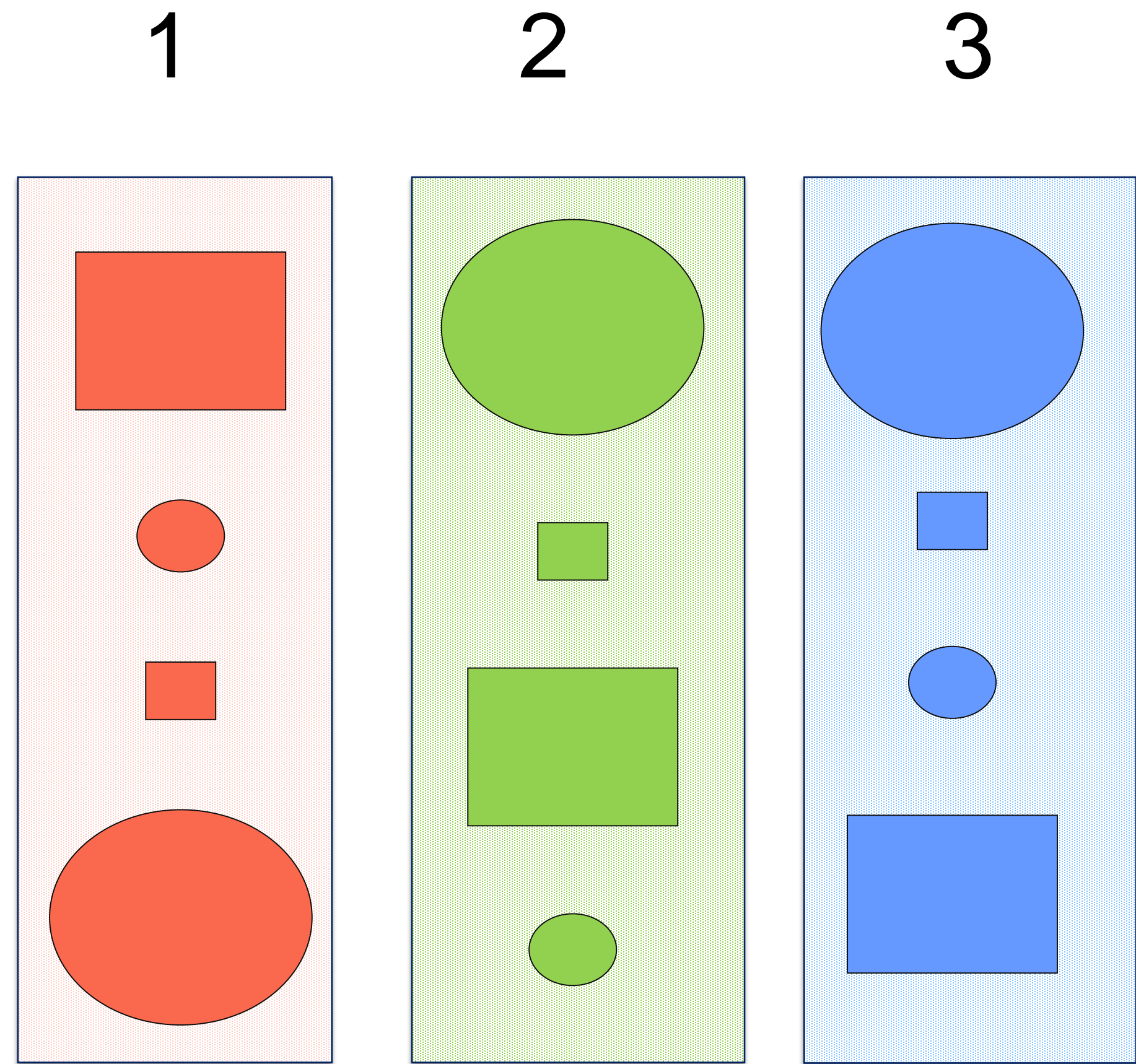


What features these objects has?



# Clustering

Criteria: Color

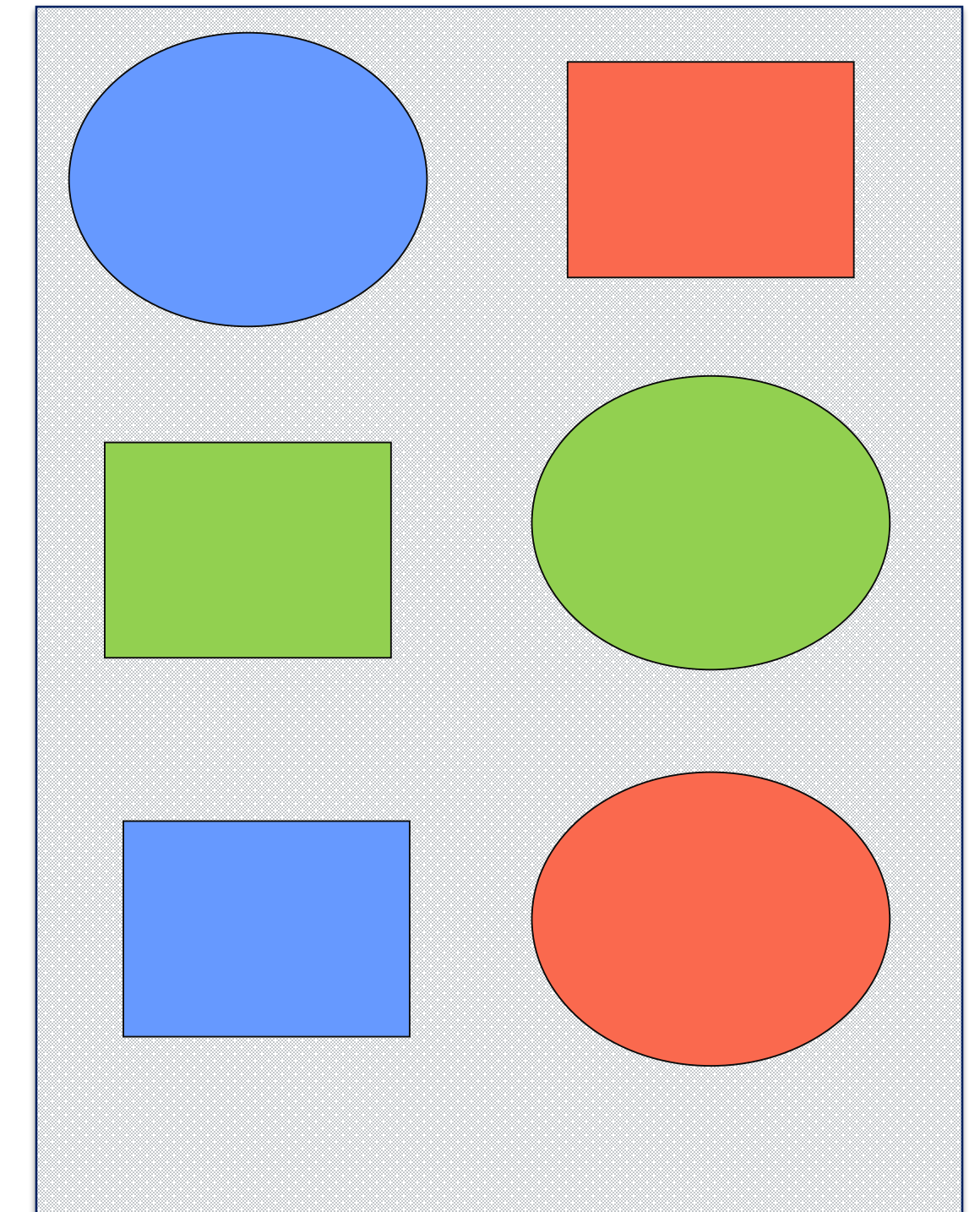
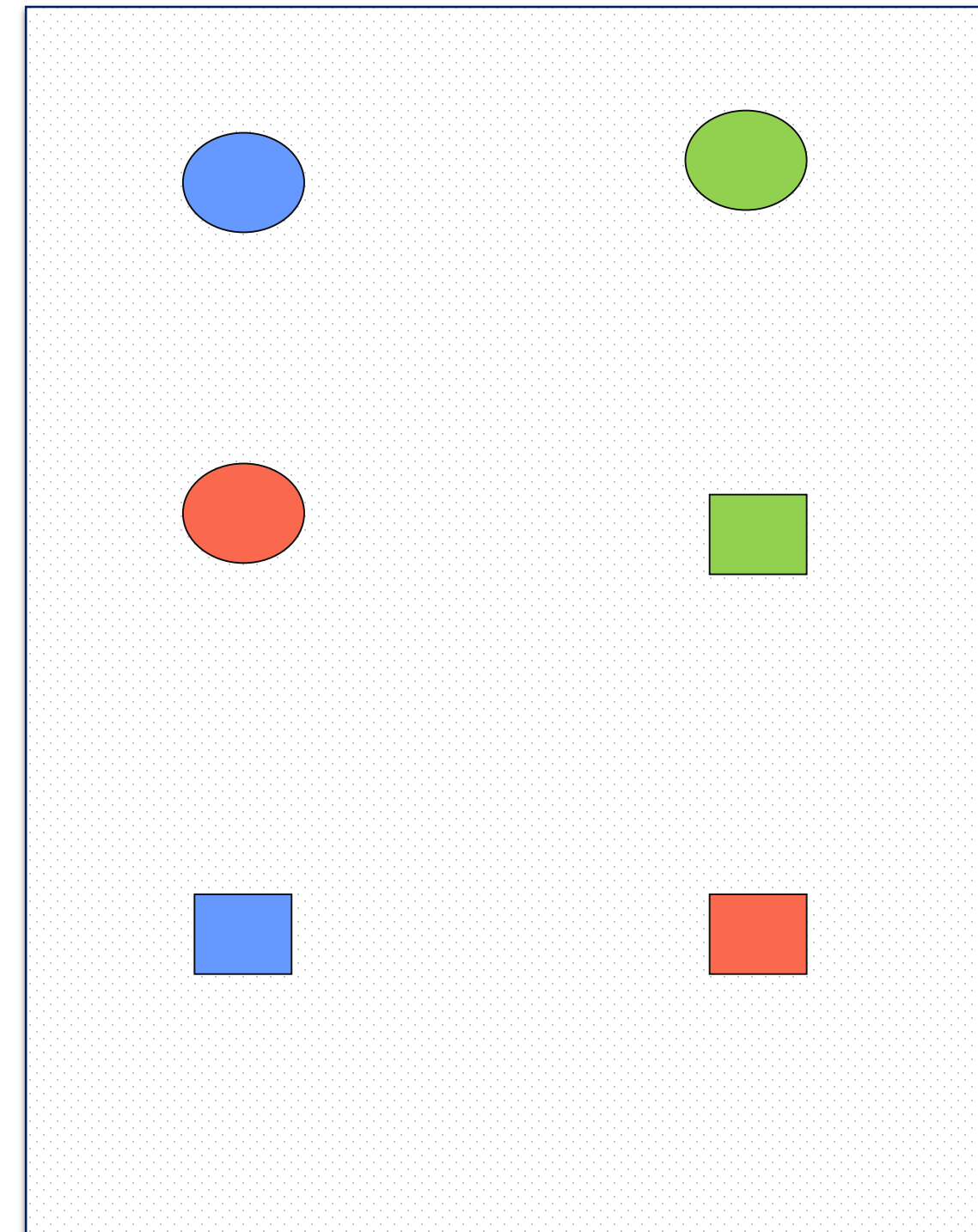


# Clustering

Criteria: Size

1

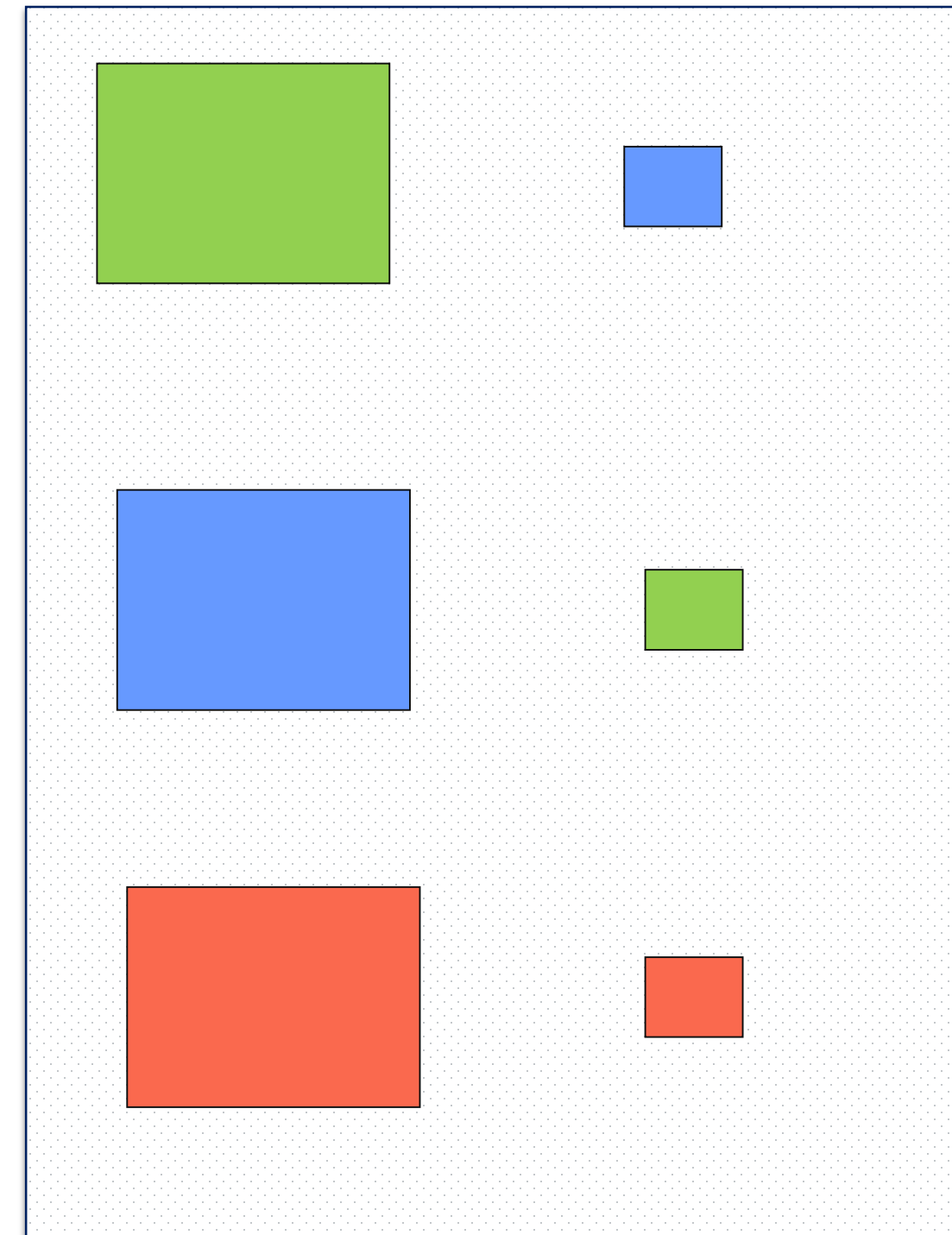
2



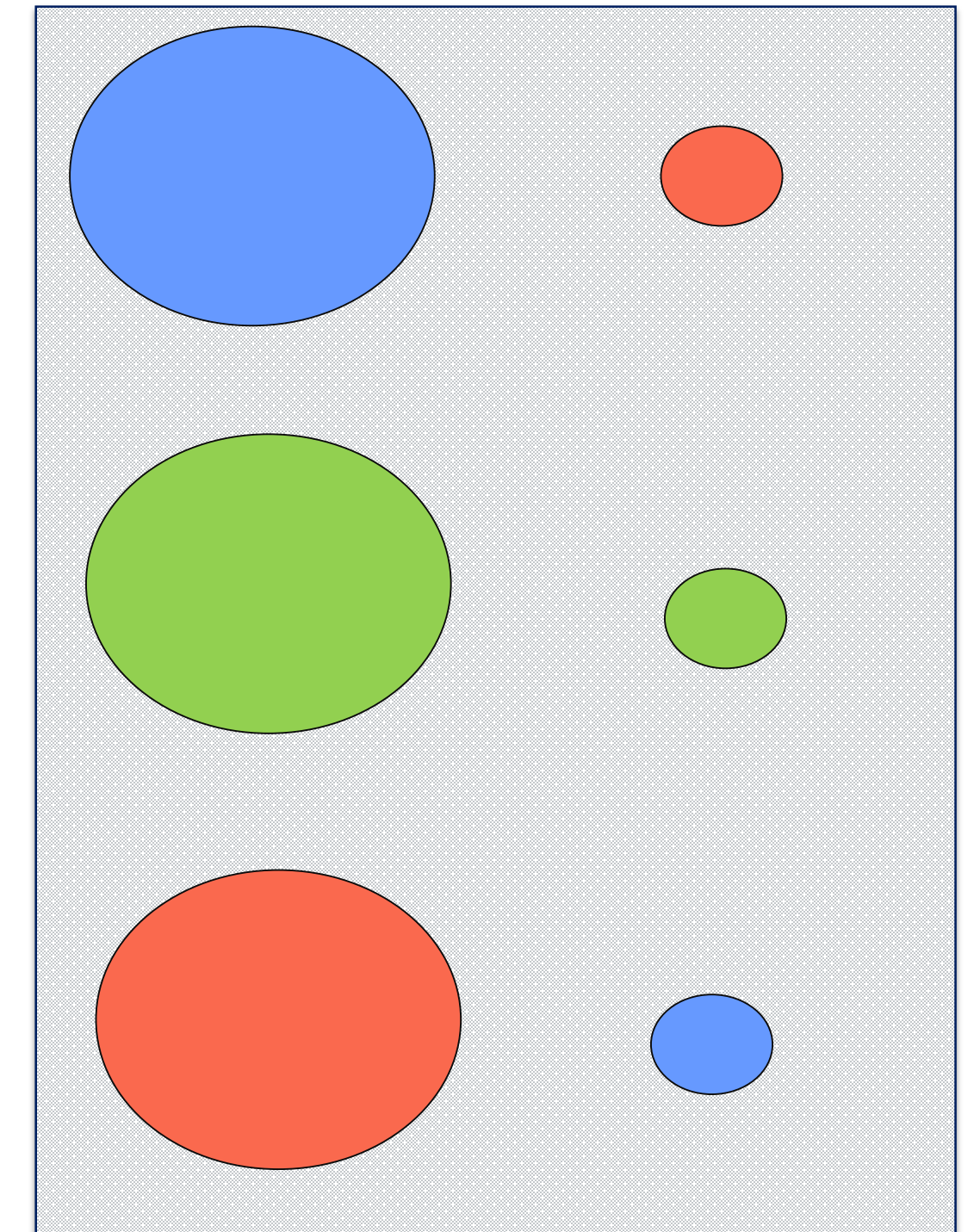
# Clustering

Criteria: Shape

1



2



# Data (unlabeled)

Can we form groups of these samples?

What are the criteria for grouping?

#	Inputs			
	HR	BVP	EDA	TempBF
1	76.85	17.27	0.22	34.63
2	76.97	19.54	0.22	34.64
3	77.10	18.51	0.22	34.64
4	85.28	46.09	0.22	34.61
5	85.42	35.83	0.22	34.61
6	88.02	2.59	0.22	34.63
7	77.25	6.34	0.22	34.65
8	77.49	6.98	0.22	34.63
9	85.81	12.18	0.22	34.61

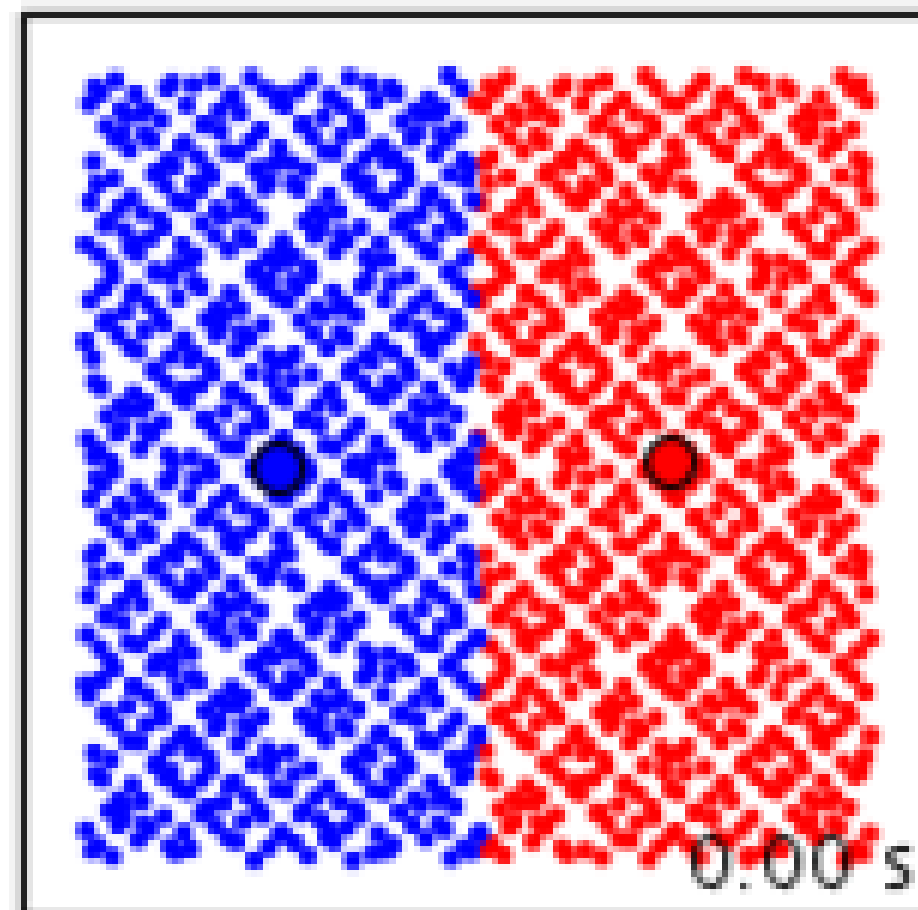
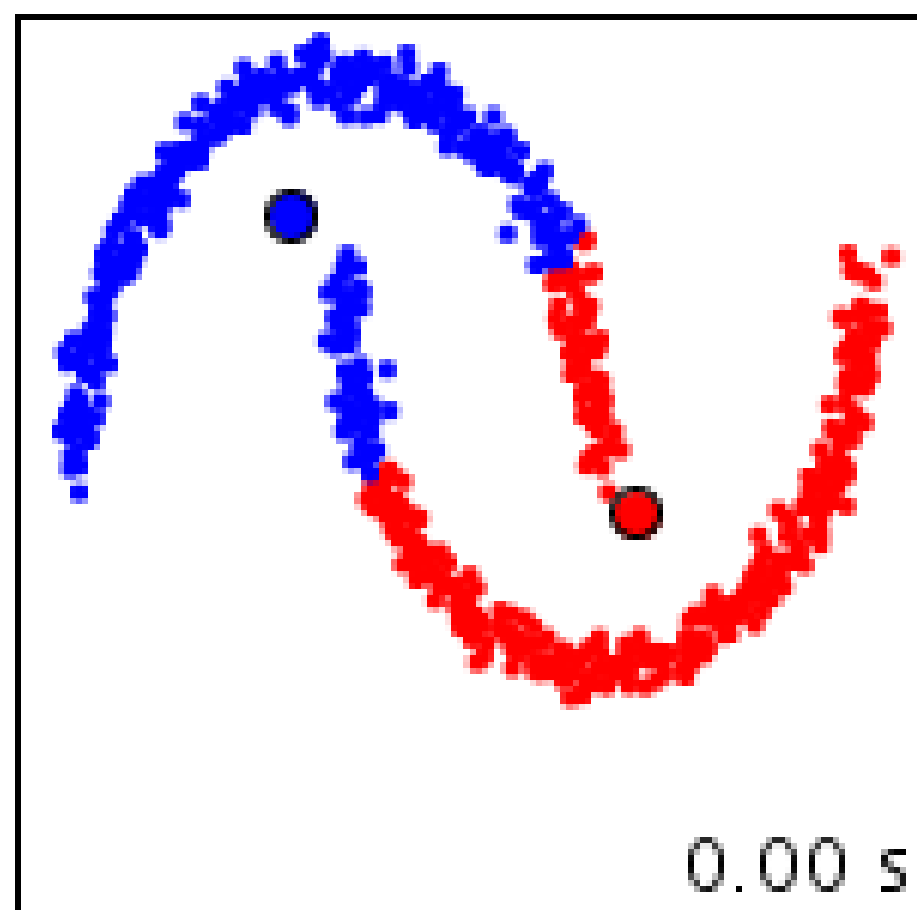
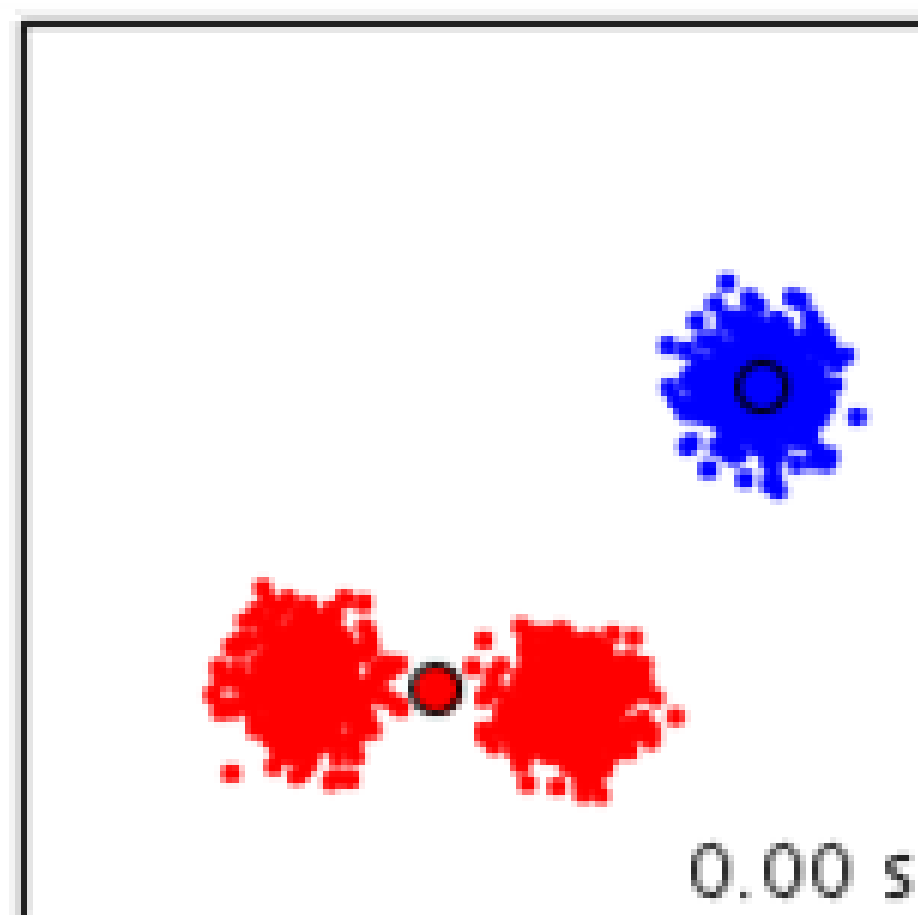
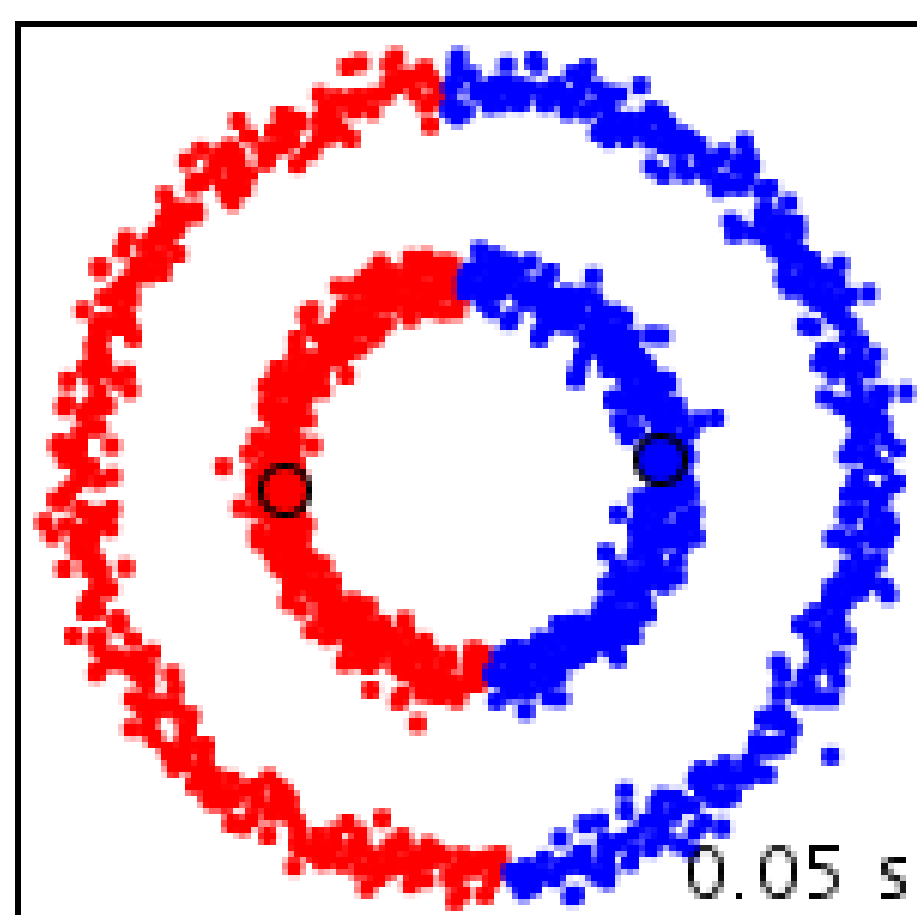


# Clustering

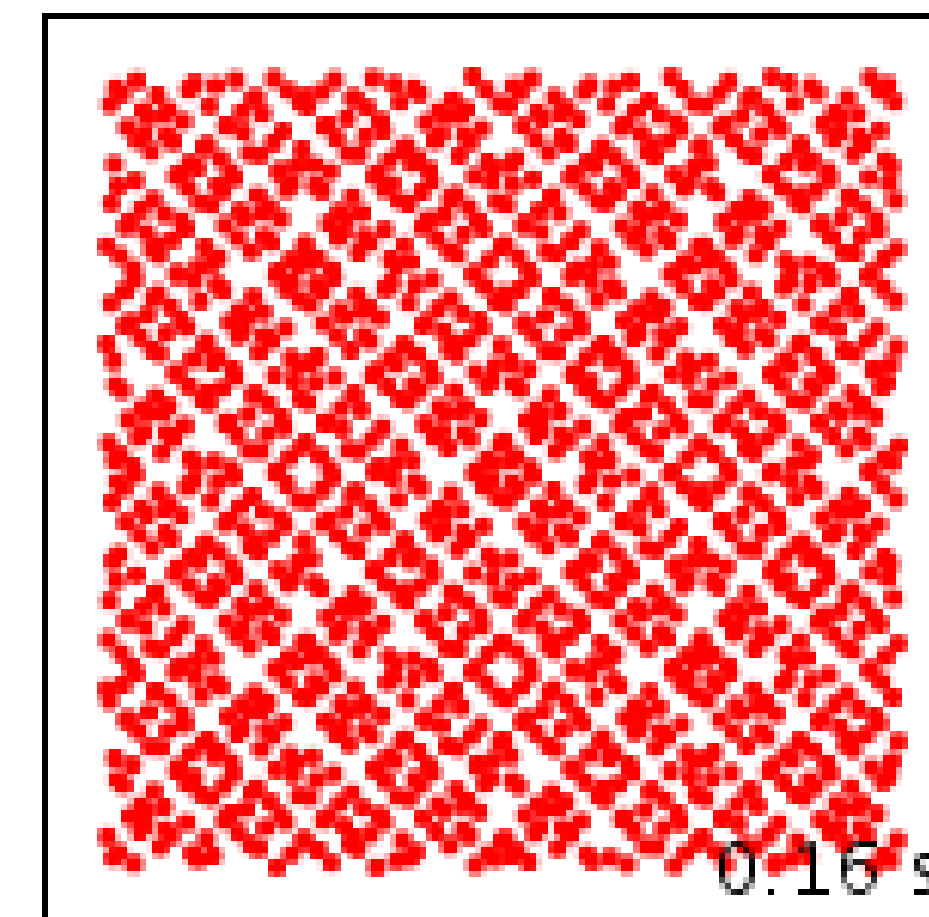
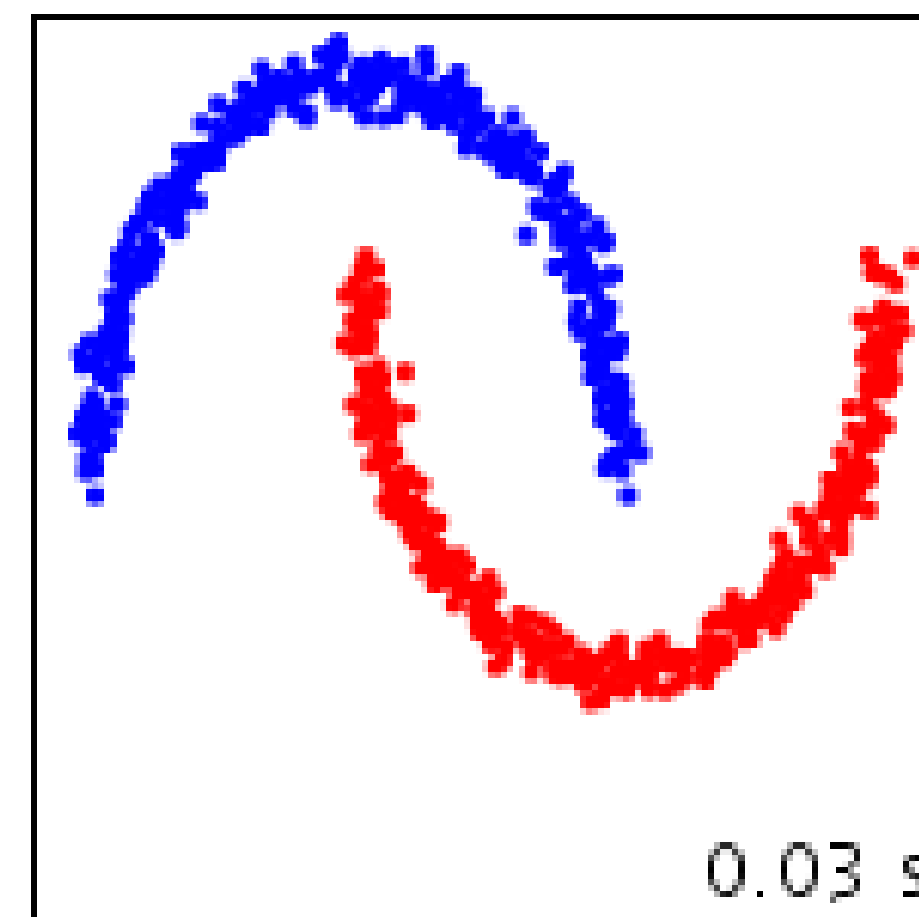
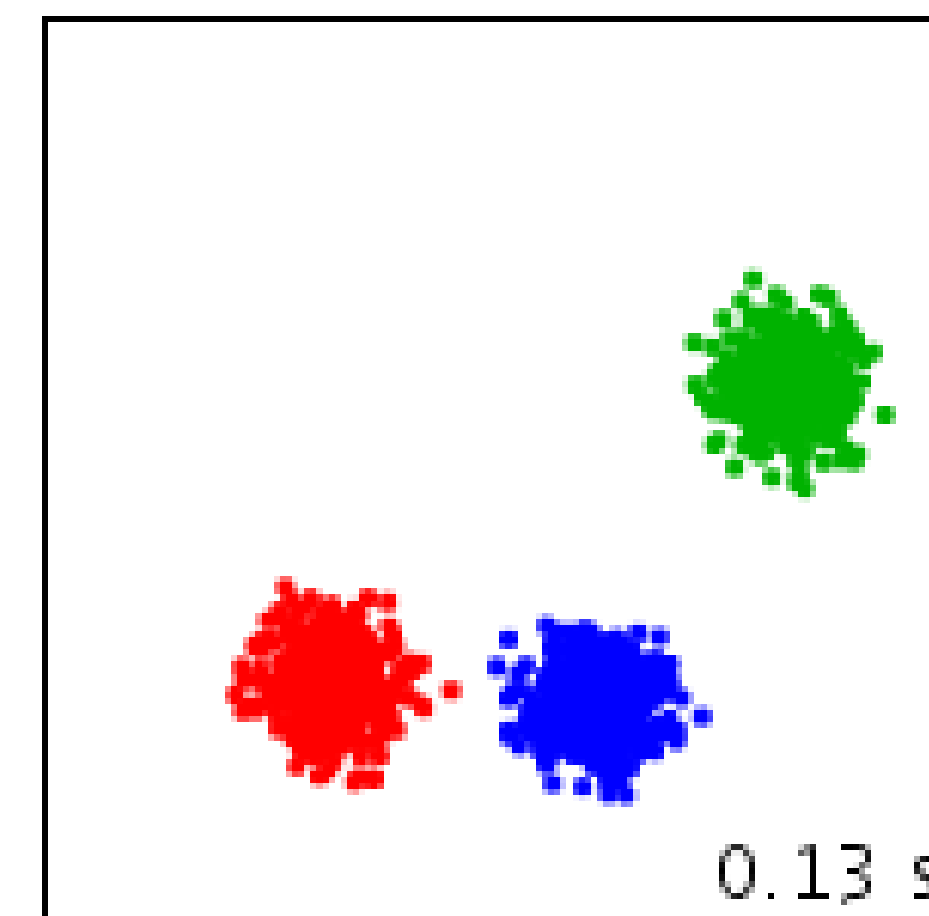
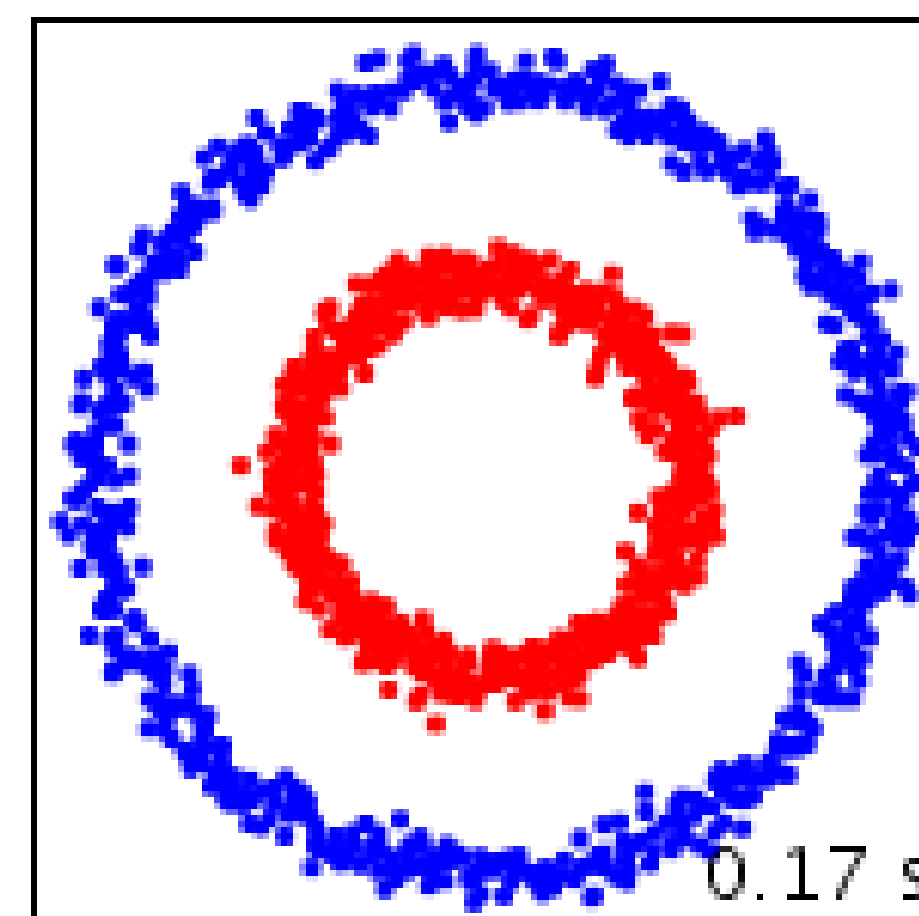
Reference: <http://commons.apache.org/proper/commons-math/userguide/ml.html>

## Distance-based Vs Density-Based

$K = 2$

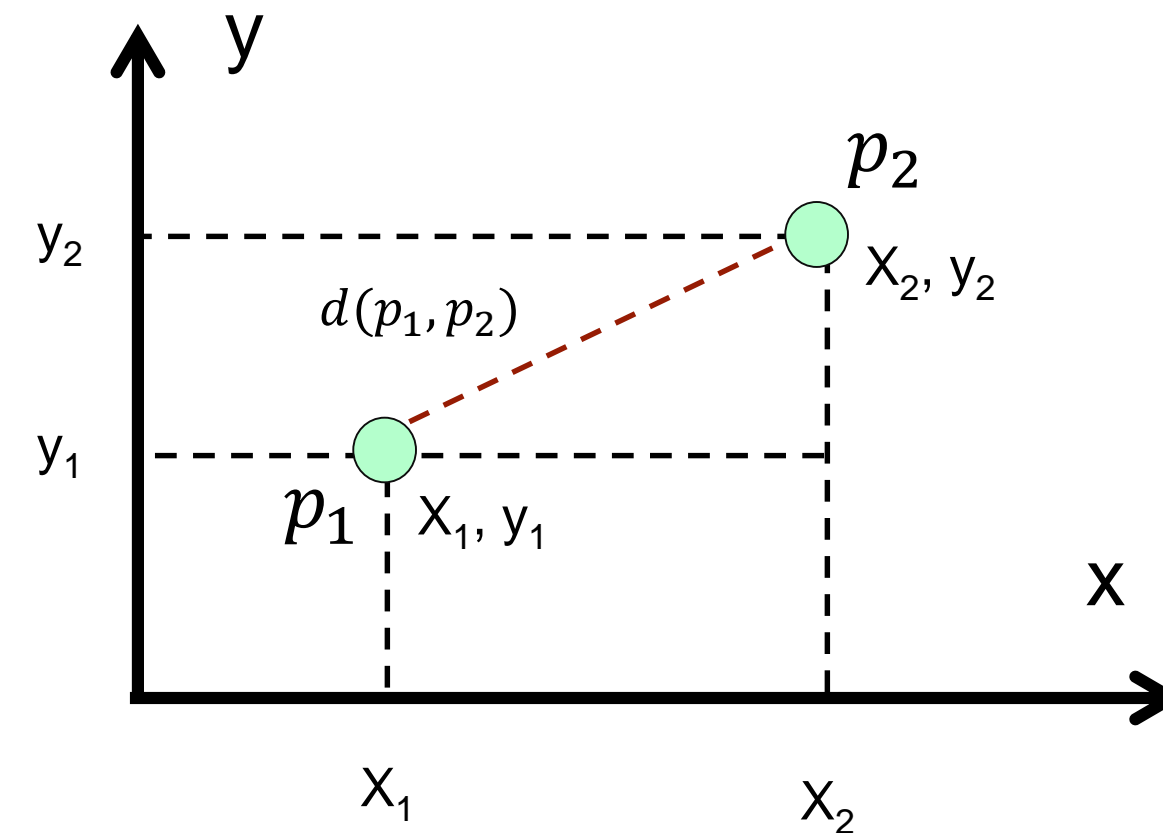


$\epsilon = 2, m = 2$



# Clustering

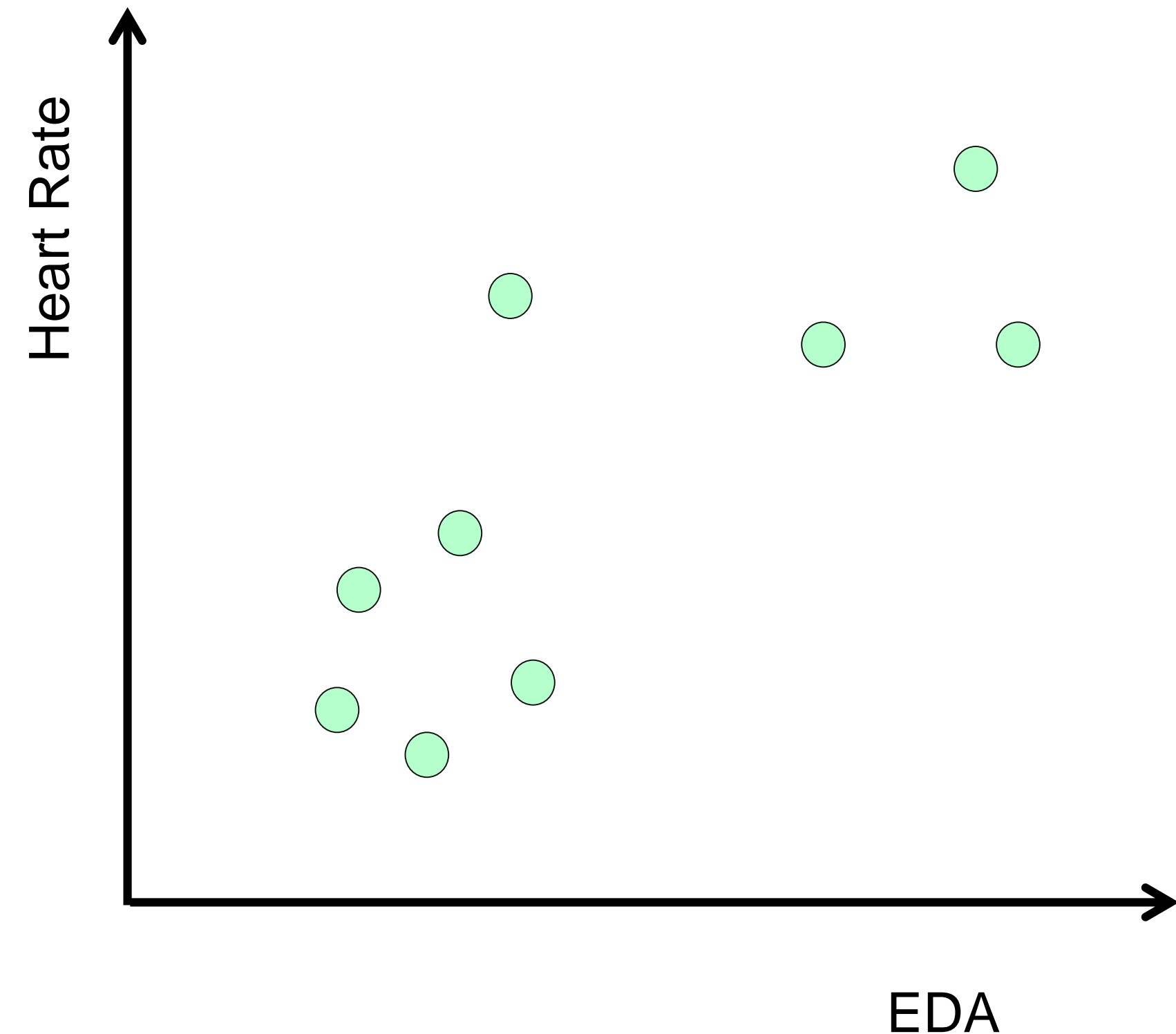
## Distance based: K-Means



Euclidean distance:

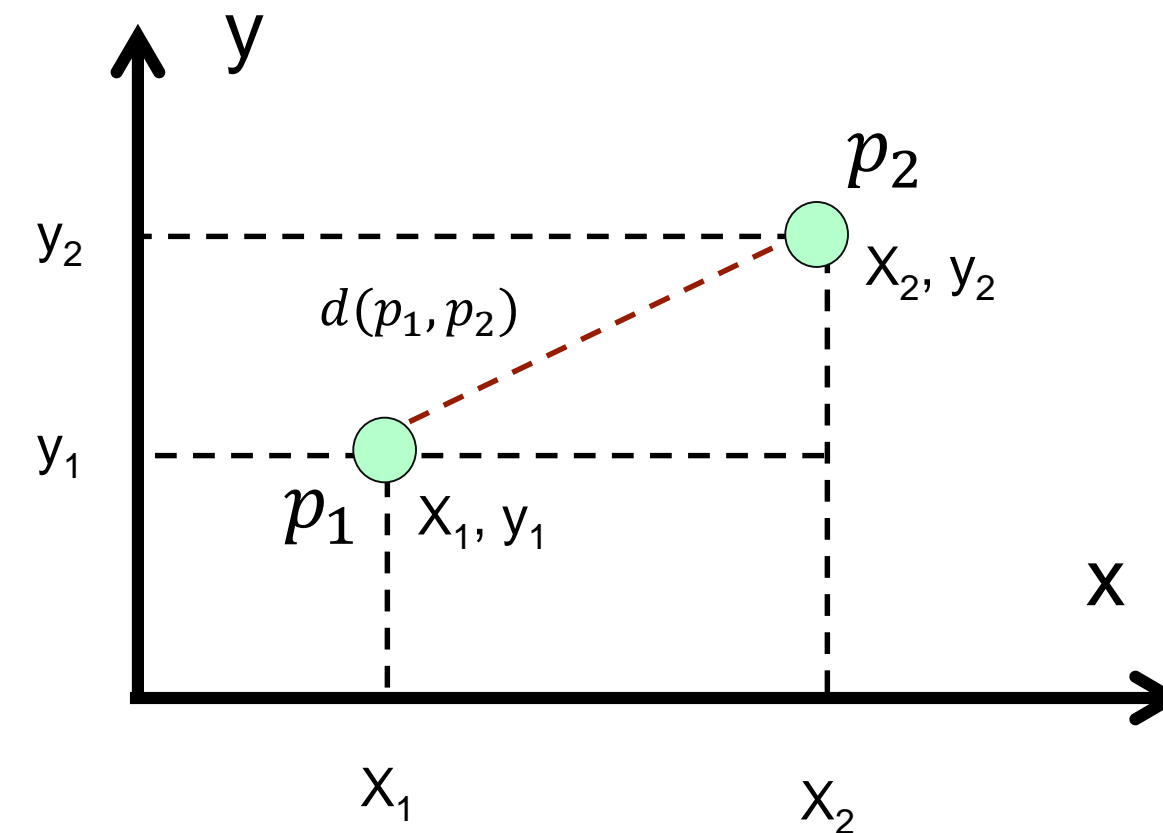
$$d(p_1, p_2) = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2}$$

## ESUM data



# Clustering

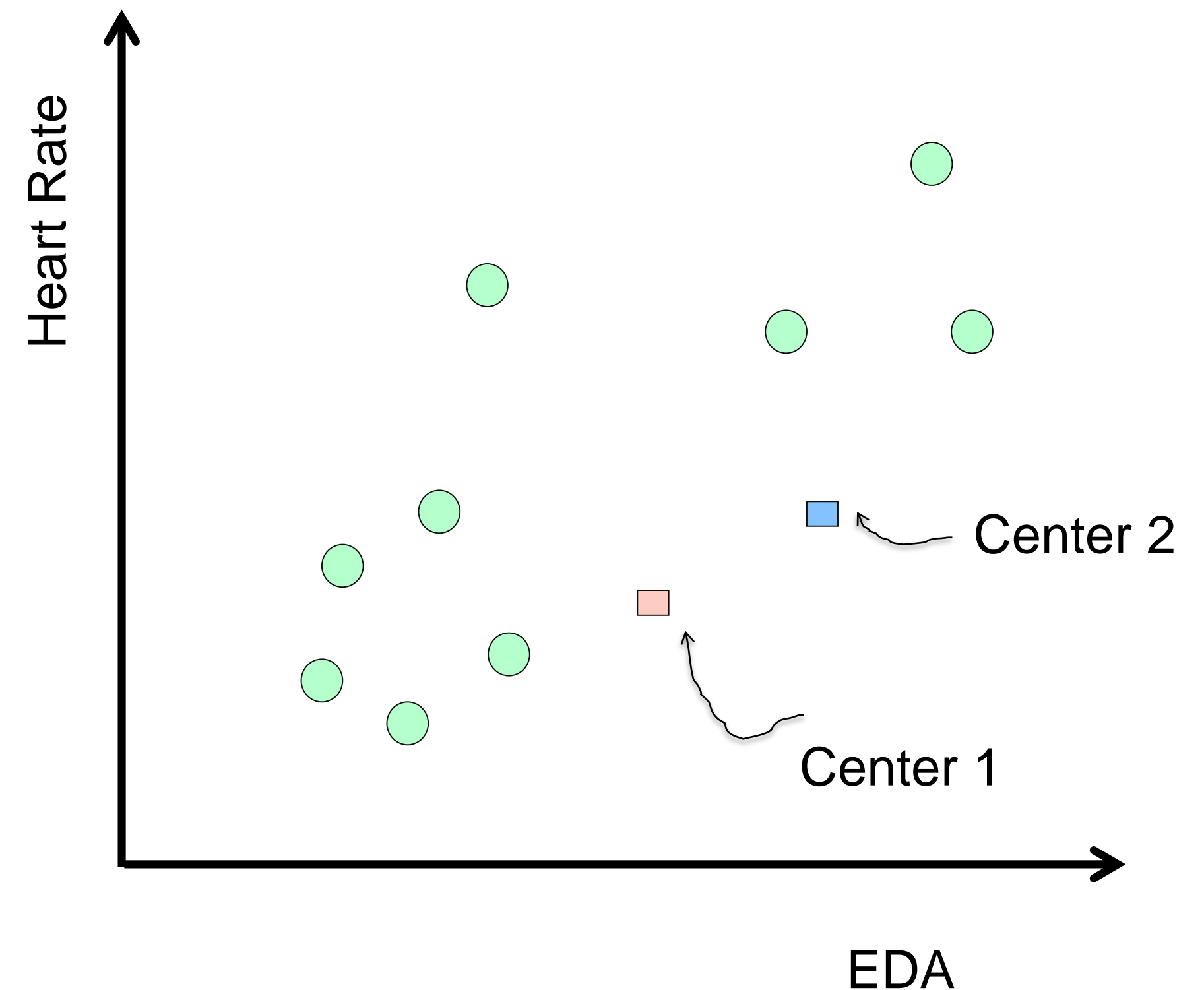
## Distance based: K-Means



Euclidean distance:

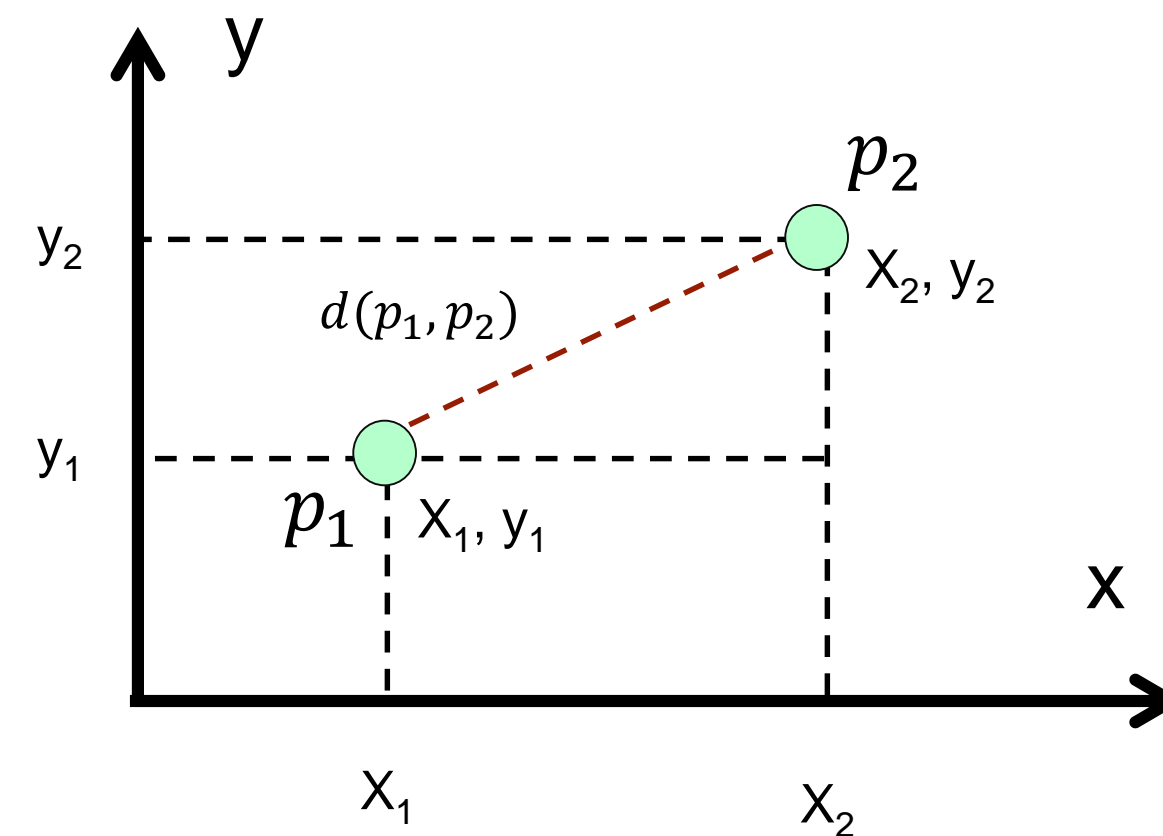
$$d(p_1, p_2) = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2}$$

## Defining centers



# Clustering

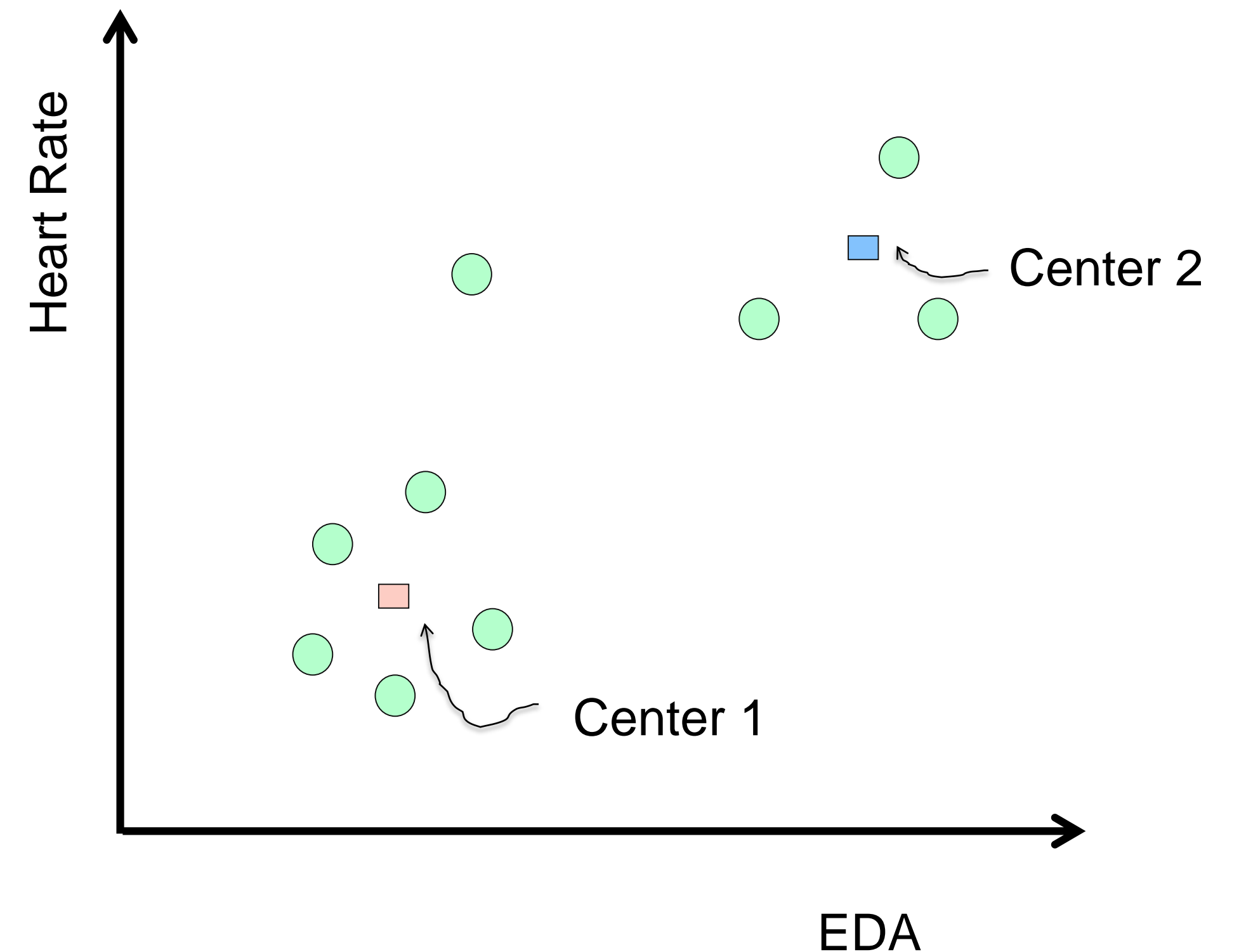
## Distance based: K-Means



Euclidean distance:

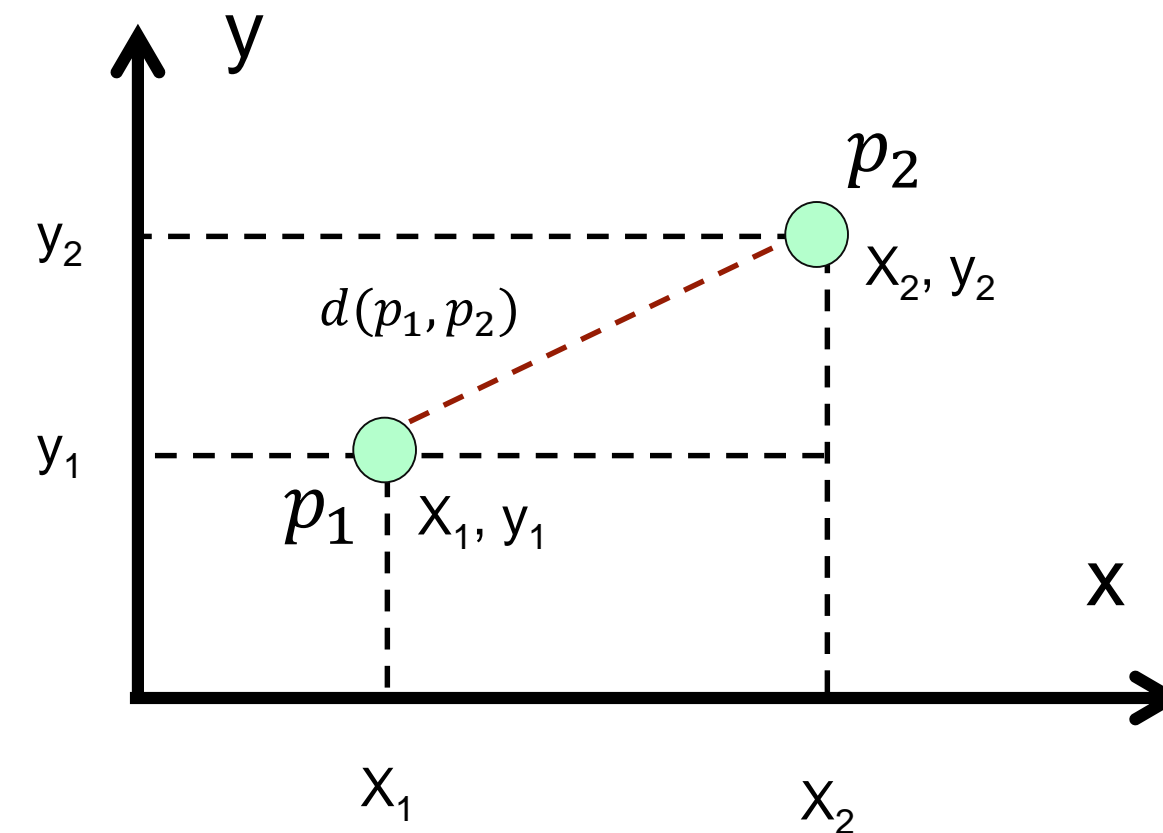
$$d(p_1, p_2) = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2}$$

## Results



# Clustering

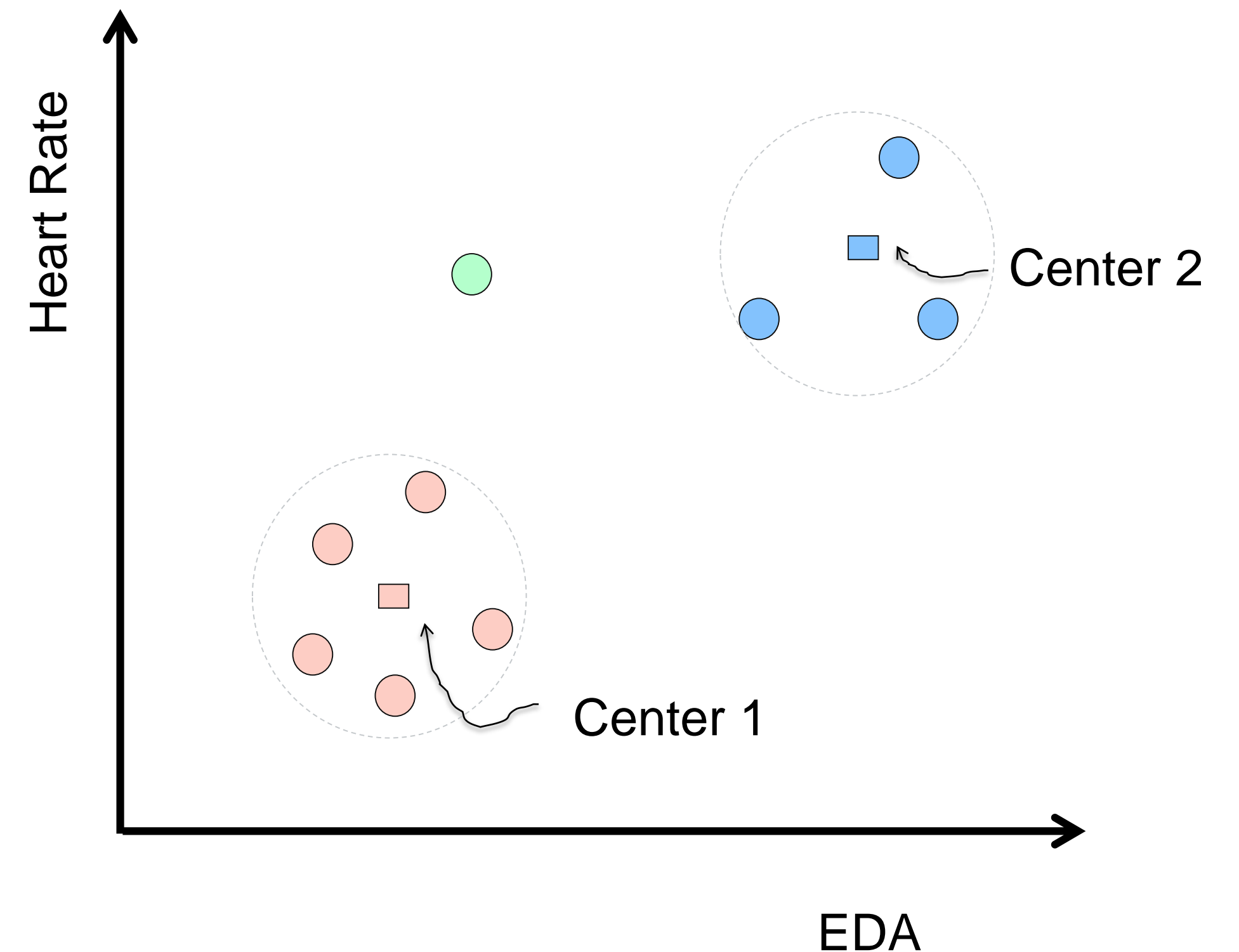
## Distance based: K-Means



Euclidean distance:

$$d(p_1, p_2) = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2}$$

## Results

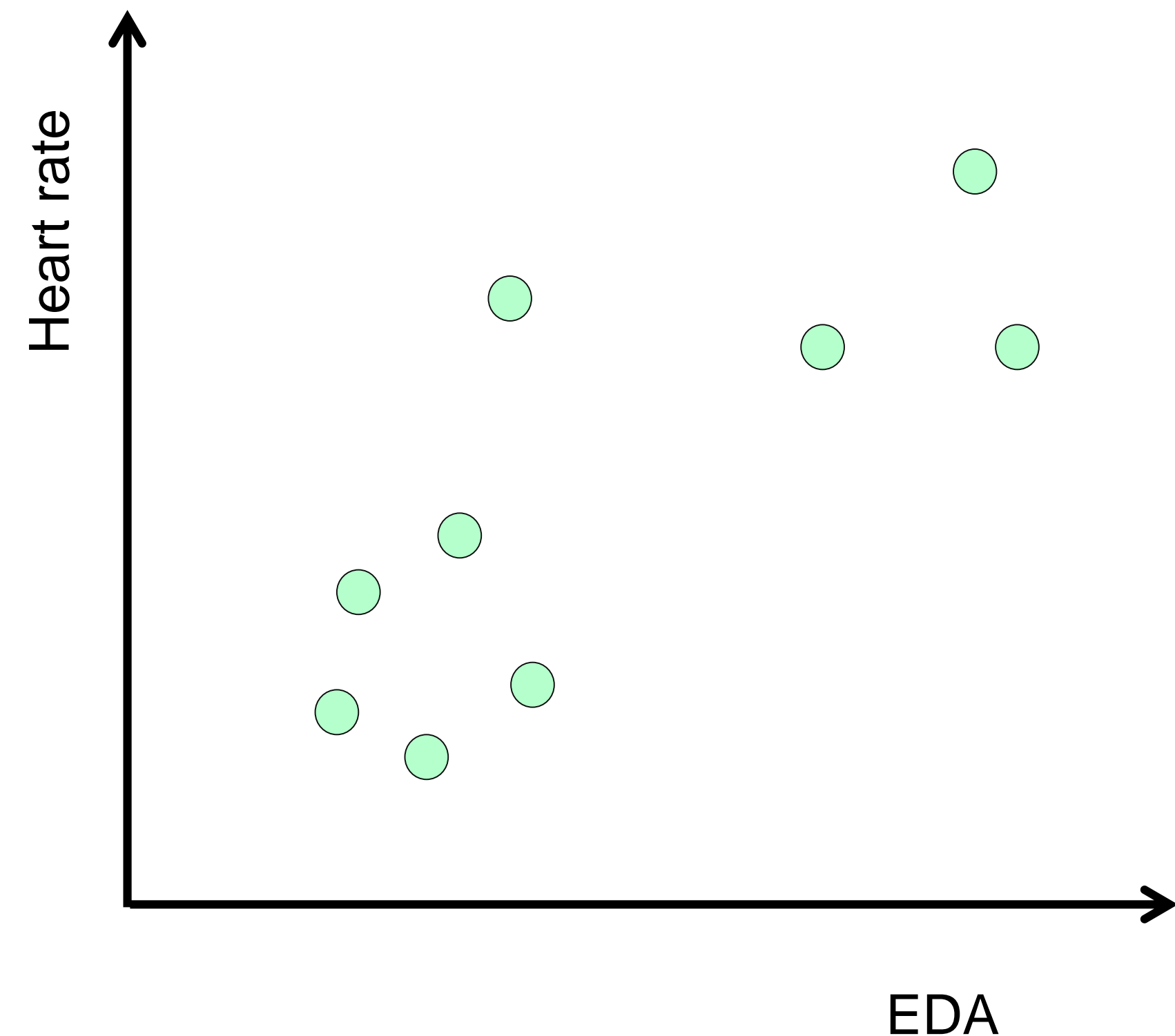


# Clustering

## Density based : DBSCAN

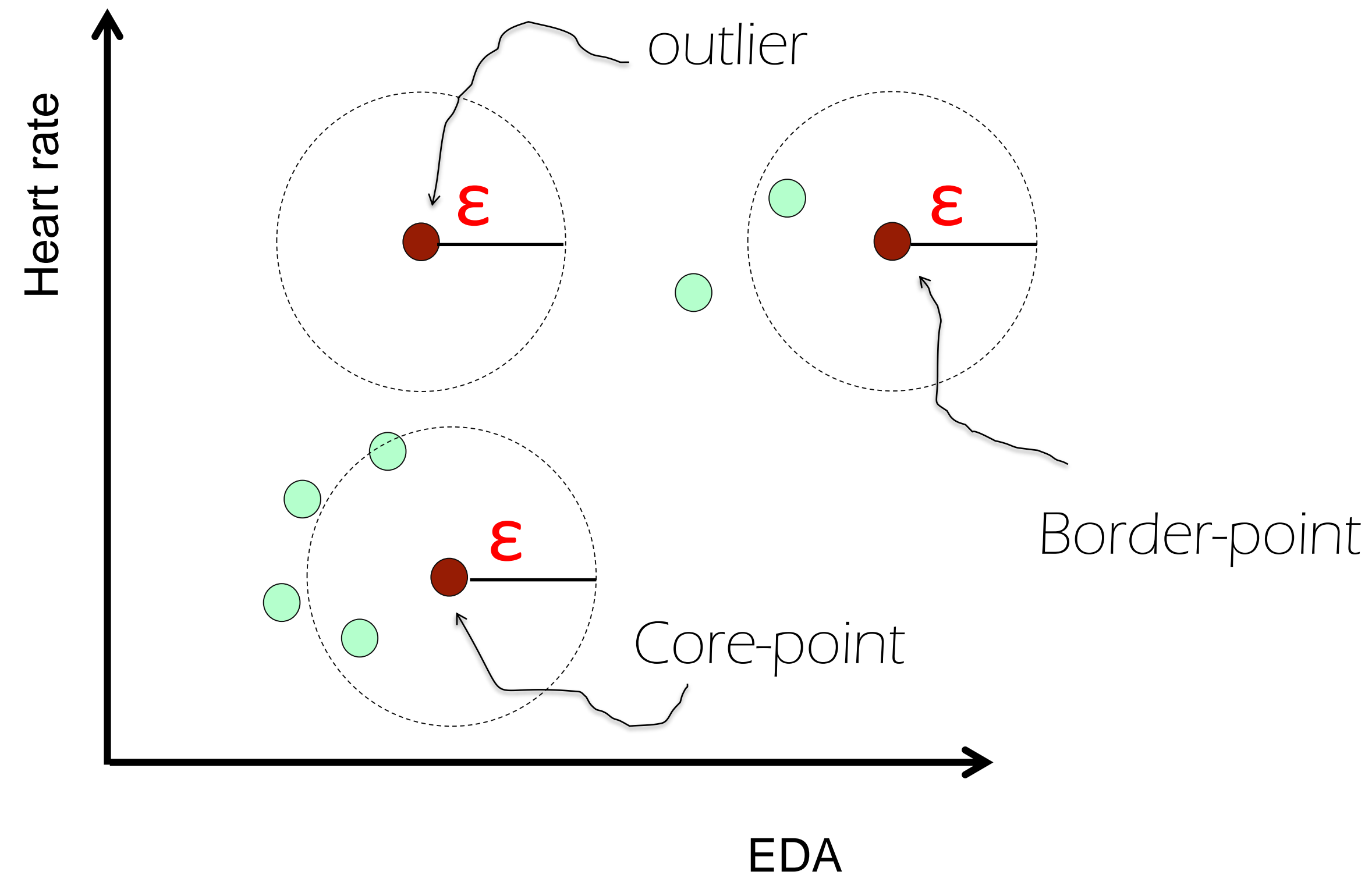
- Select two parameters:
  - $\epsilon$ -Neighborhood
  - $m$ -minimum number of points
- Select a point  $p$  randomly
- Find all points reachable from  $p$  based on  $\epsilon$  and  $m$
- If  $p$  is a *core point*, mark points as a cluster.
- If  $p$  is a *border point*, no points are reachable from  $p$ .
- Select a new point  $p$ .
- Continue until all of the points are visited.

## ESUM data

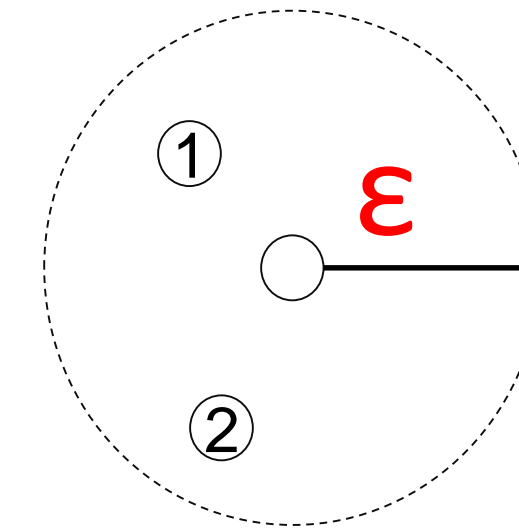


# Clustering

## Density based : DBSCAN



## Define a center

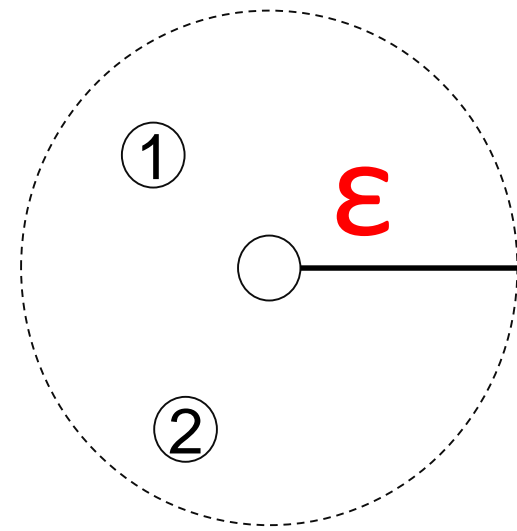


$$\epsilon = 1 \text{ cm}$$

$$m = 2$$

# Clustering (DBSCAN)

For a center



$$\epsilon = 1 \text{ cm}$$

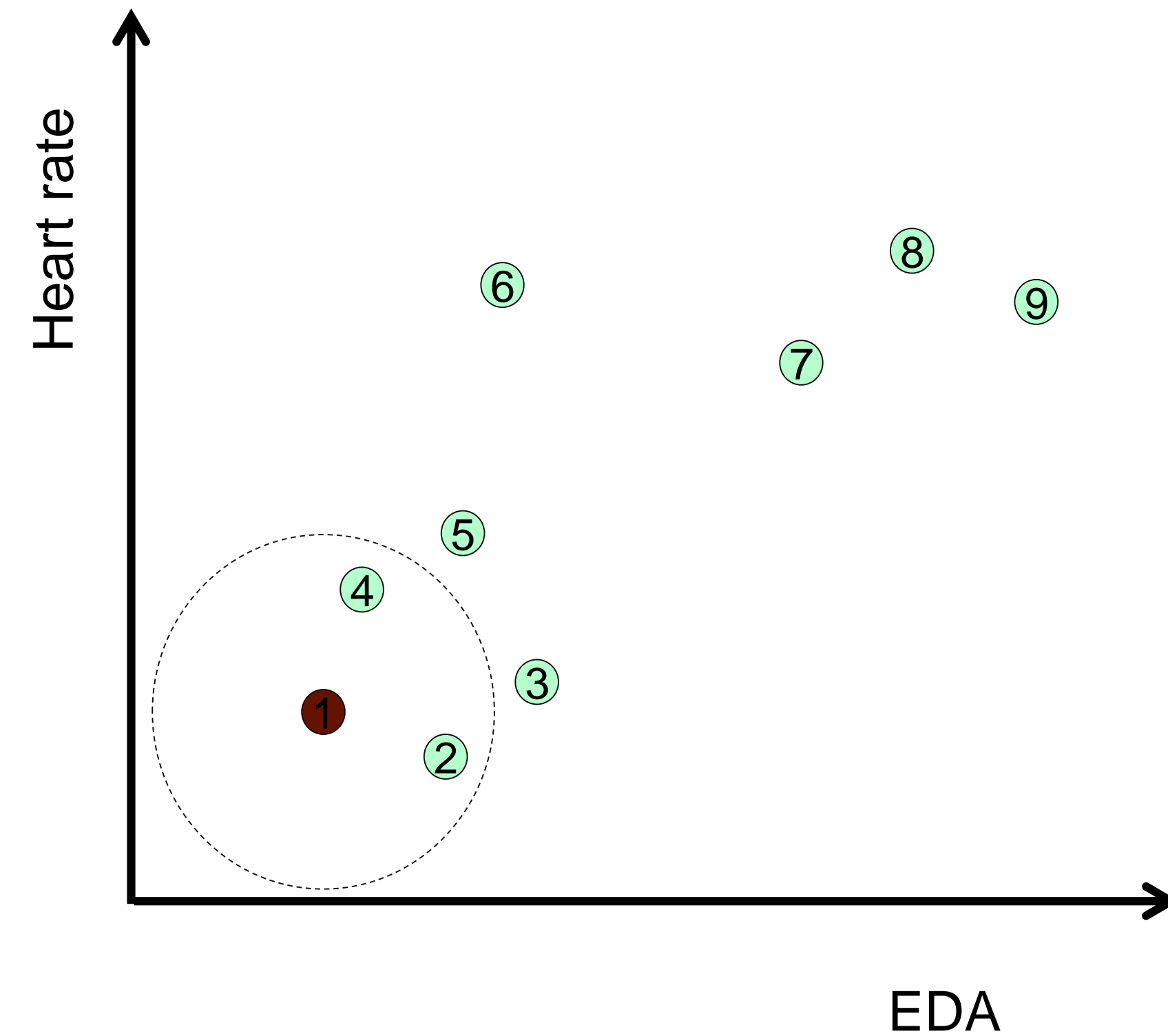
$$m = 2$$

Scanning point

# 1

Points in cluster:

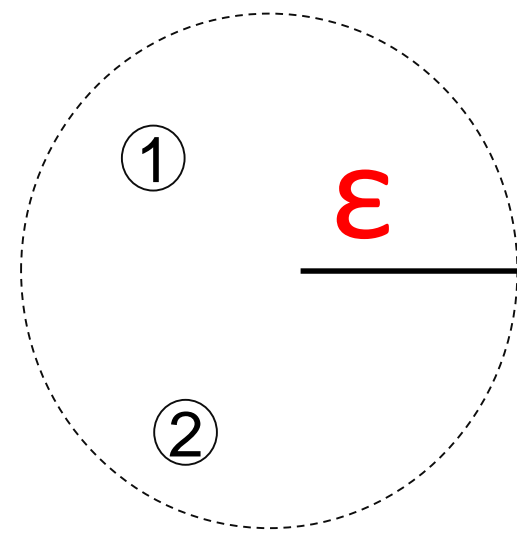
1  $\rightarrow$  2,4





# Clustering (DBSCAN)

## For a center



$$\epsilon = 1 \text{ cm}$$

$$m = 2$$

Scanning point

# 1,

# 2,

#3,

#5

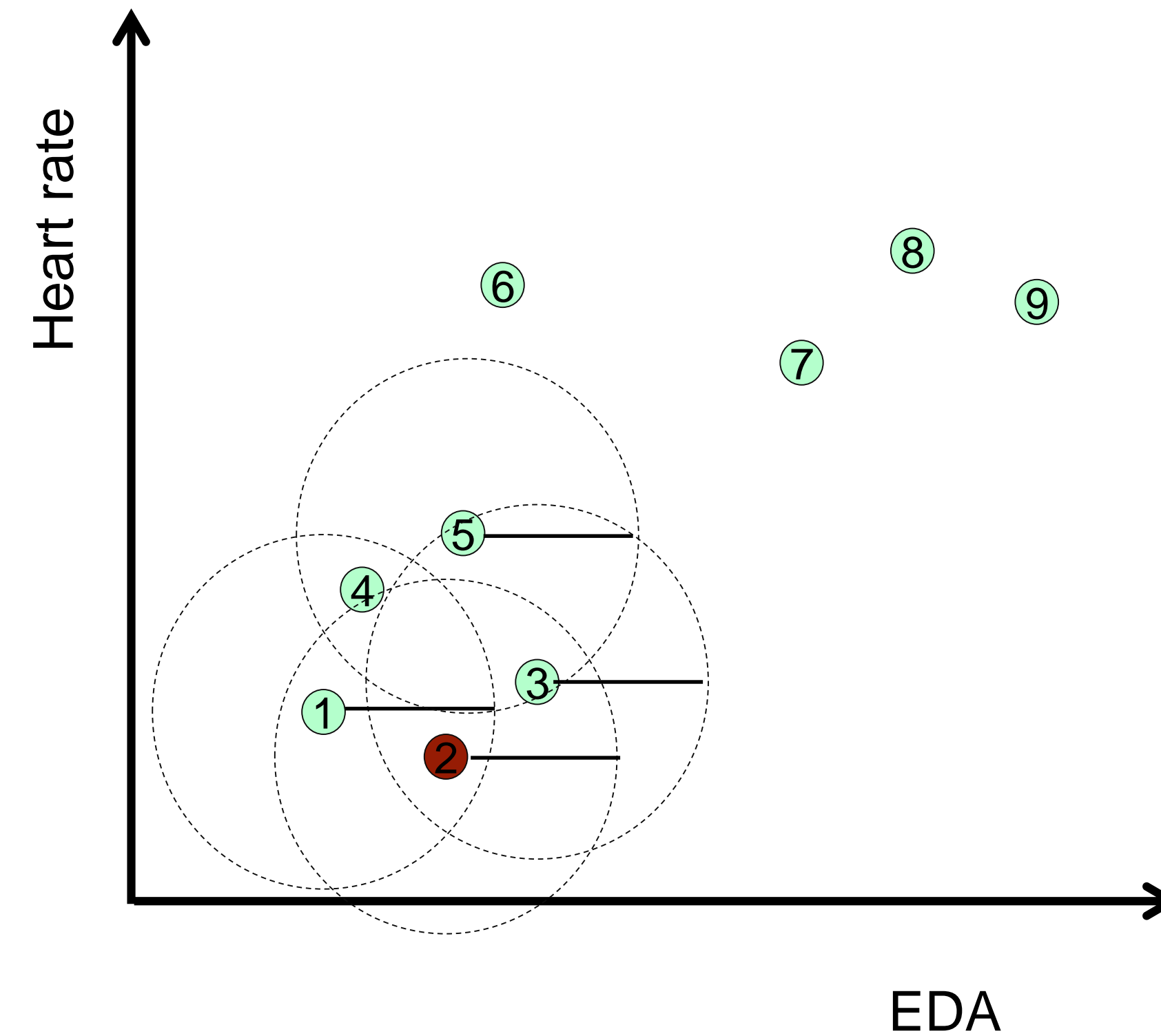
Points in cluster:

1 -> 2,4

2 -> 1, 3,4

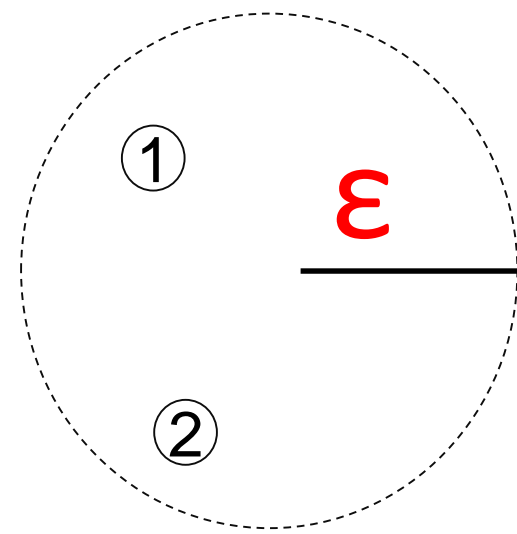
3 -> 2, 5

5 -> 4,3



# Clustering (DBSCAN)

## For a center



$$\epsilon = 1 \text{ cm}$$

$$m = 2$$

Scanning point

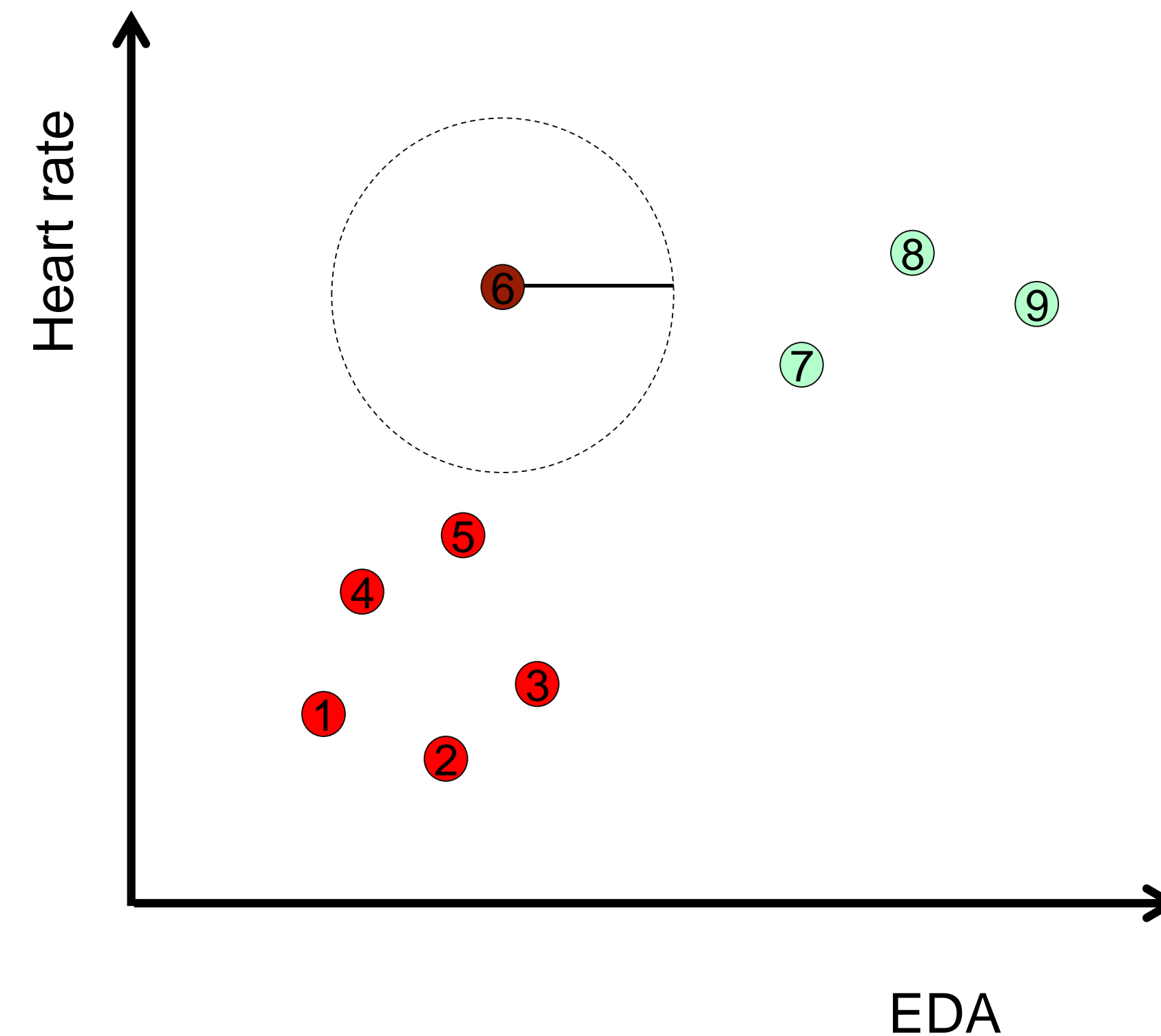
# 6,

Points in cluster:

6  $\rightarrow$  ?

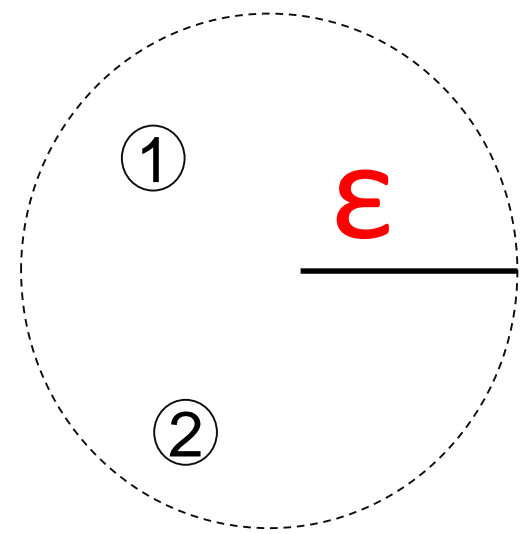
Cluster found:

red



# Clustering

## For a center



$$\epsilon = 1 \text{ cm}$$

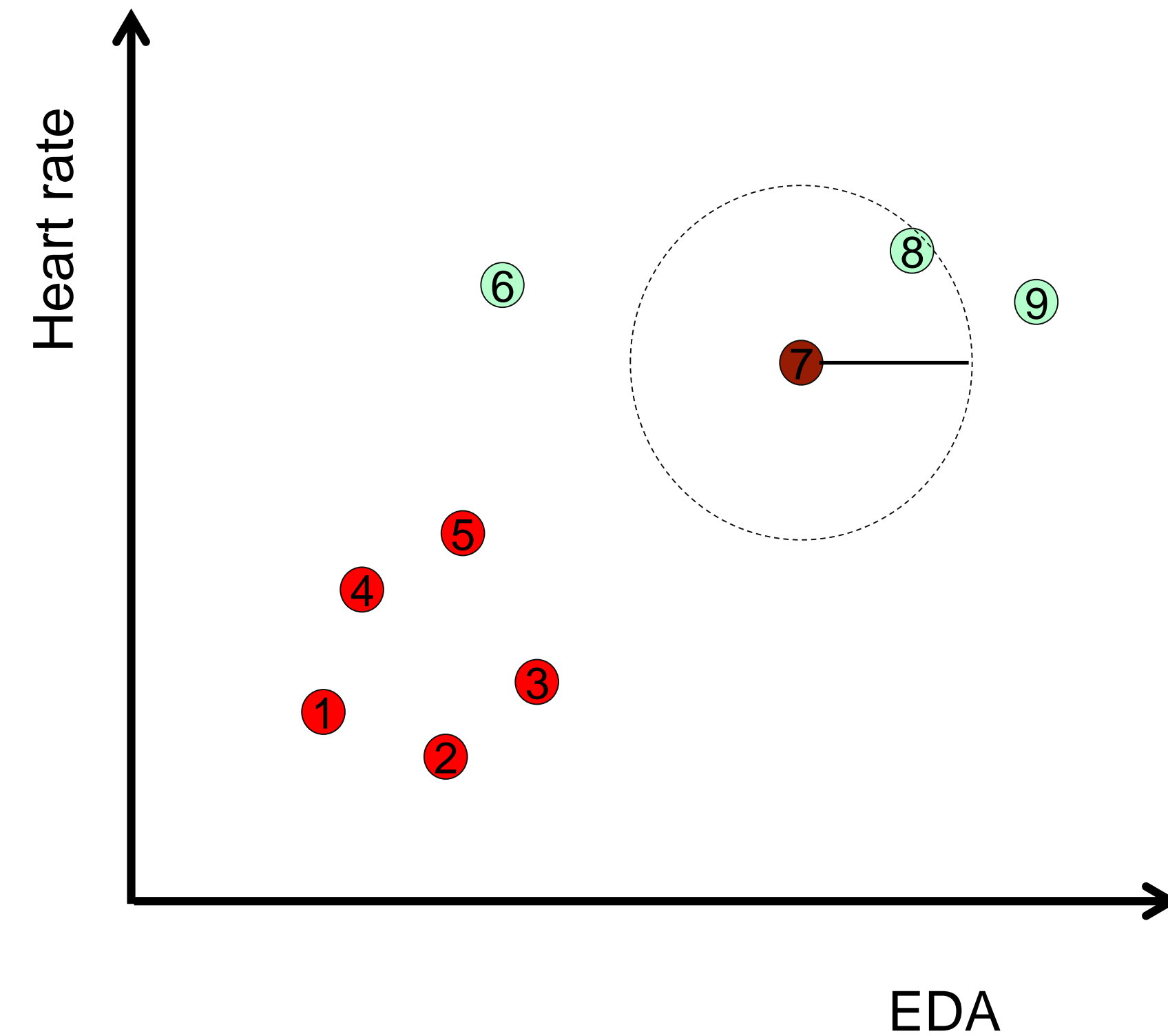
$$m = 2$$

Scanning point

# 7,

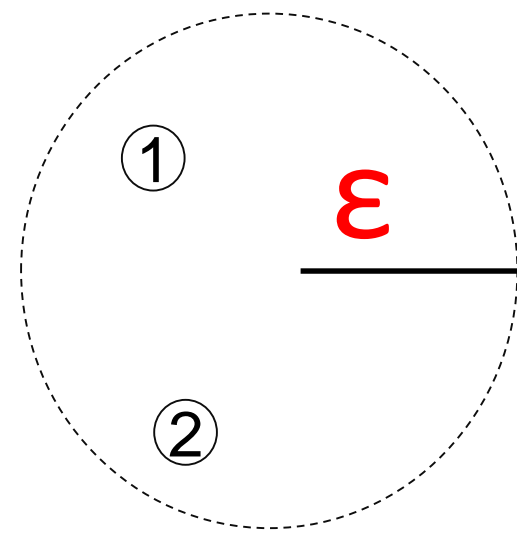
Points in cluster:

7  $\rightarrow$  8



# Clustering (DBSCAN)

For a center



$$\epsilon = 1 \text{ cm}$$

$$m = 2$$

Scanning point

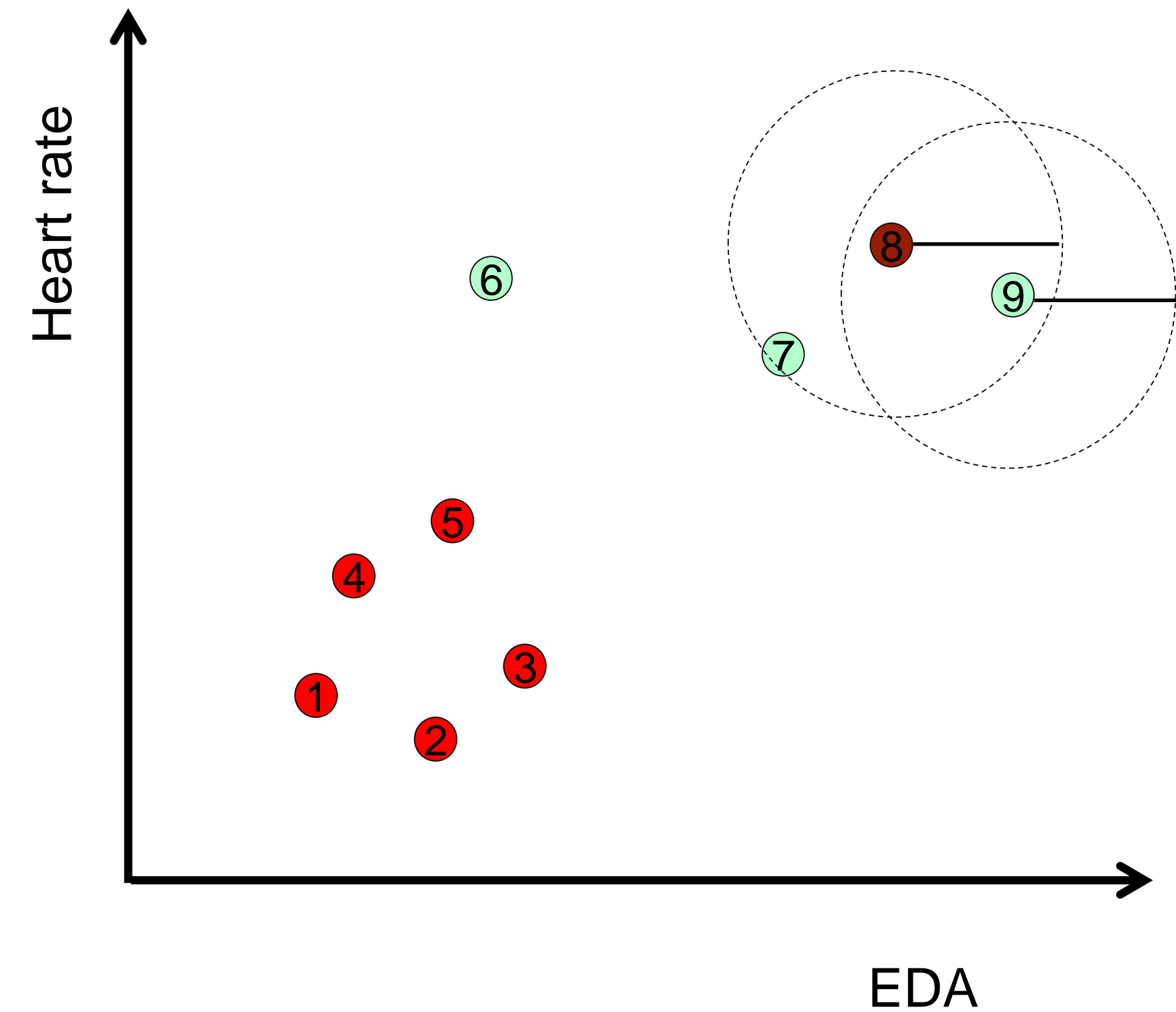
# 8,

#9

Points in cluster:

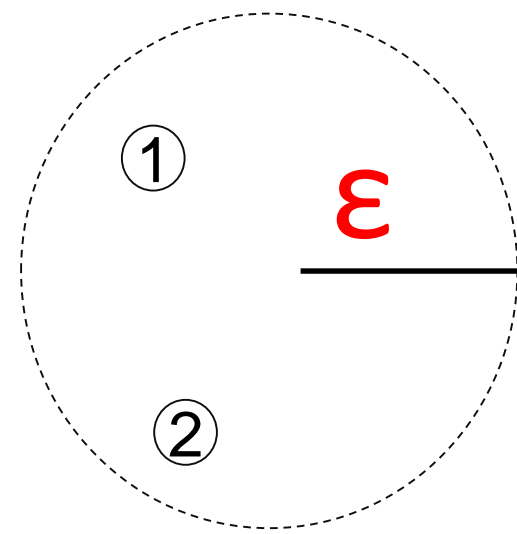
8  $\rightarrow$  7, 9

9  $\rightarrow$  8



# Clustering (DBSCAN)

For a center



$$\epsilon = 1 \text{ cm}$$

$$m = 2$$

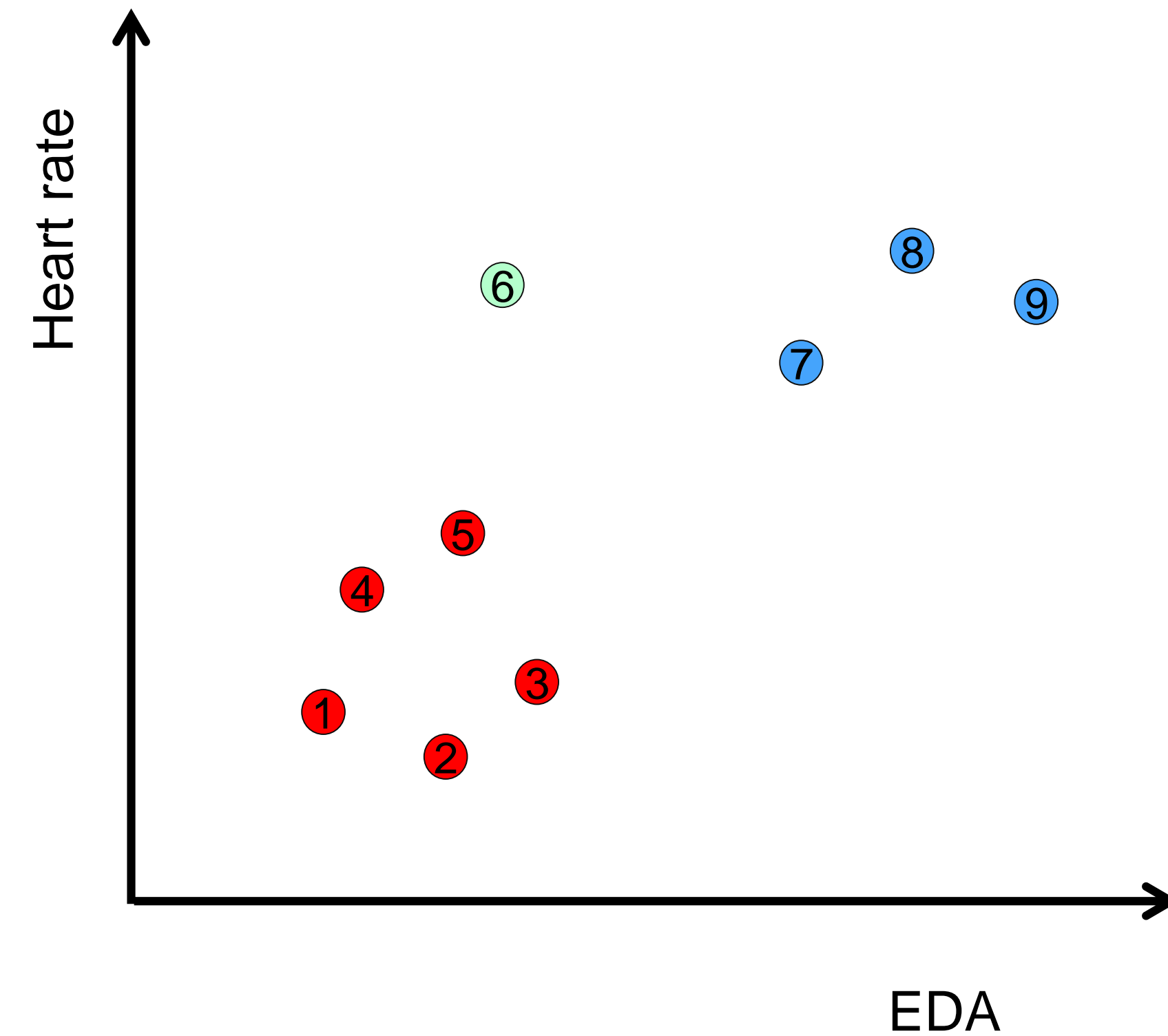
Scanning point

Total cluster 2

Finish

red

blue



Thank you!

Varun Ojha [ojha@arch.ethz.ch](mailto:ojha@arch.ethz.ch)

Danielle Griego [griego@arch.ethz.ch](mailto:griego@arch.ethz.ch)