NATIONAL RESEARCH UNIVERSITY

School of Data Analysis and Artificial Intelligence Department of Computer Science

DATA SCIENCE FOR BUSINESS

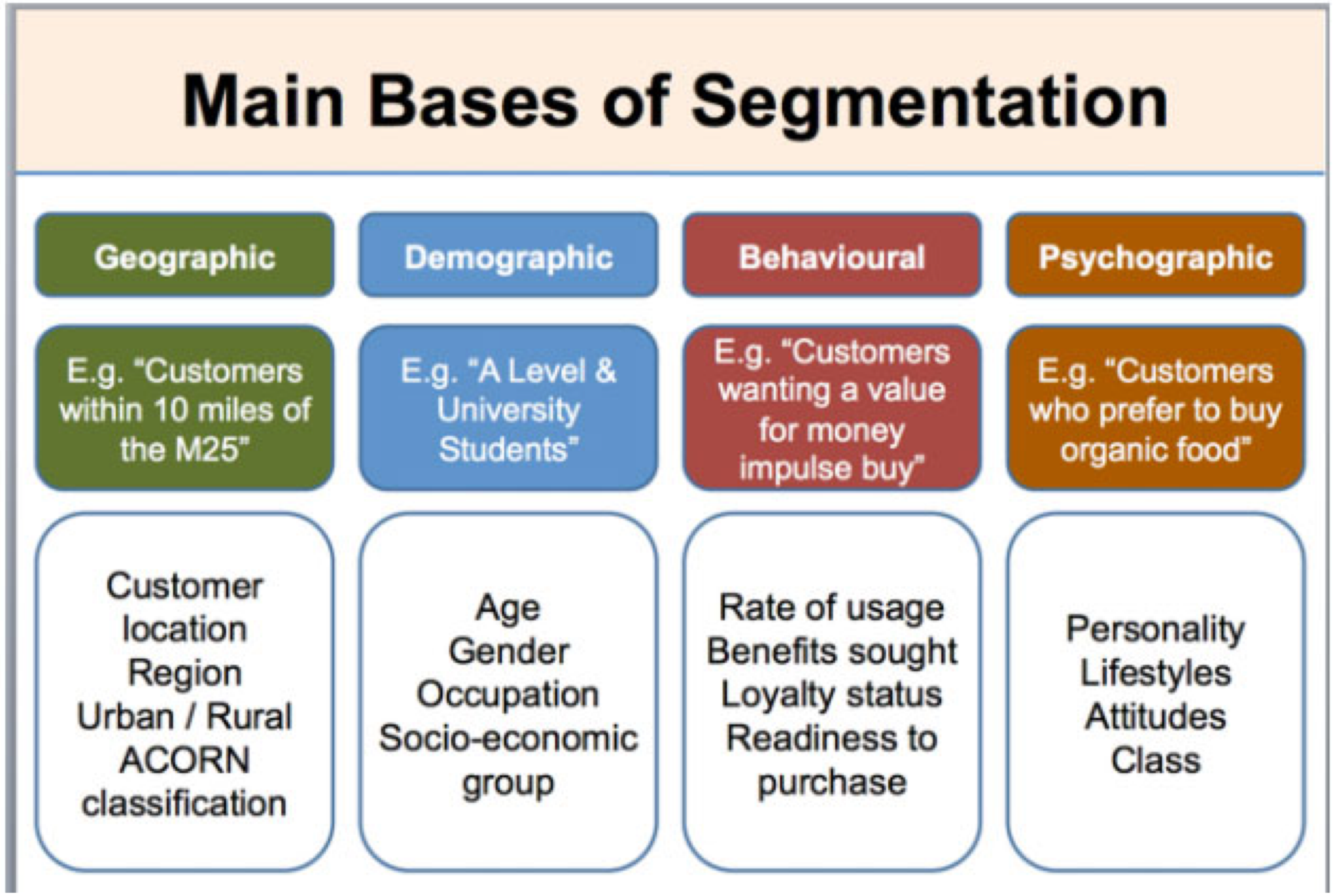
Lecture 5. Customer segmentation. Clustering.

Moscow, May 13nd, 2022.

lzhukov@hse.ru

CUSTOMER SEGMENTATION

Customer segmentation as “the practice of dividing a customer base into groups of individuals that are similar in specific ways relevant to marketing.”



WHY SEGMENT CUSTOMERS?

Marketing and service differentiation

- Improve marketing focus – segments have different interests, values, tastes and reasons to purchase.
- Identify most and least profitable customers
- Build loyal relationships
- Create personas – representative customer

Product & brand

- Brands to appeal to particular segments
- Customize products and services
- Predict future purchasing patterns

Pricing

- Price products differently
- Willingness to pay for optimal value

CUSTOMER SEGMENTATION MODELS

SEGMENTATION MODEL	HOW TO SEGMENT CUSTOMERS
Demographic Segmentation	Age, gender, income, education, and marital status
Geographic Segmentation	Country, state, city, and town
Psychographic Segmentation	Personality, attitude, values, and interests
Technographic Segmentation	Mobile-use, desktop-use, apps, and software
Behavioral Segmentation	Tendencies and frequent actions, feature or product use, and habits
Needs-Based Segmentation	Product/ service must-haves and needs of specific customer groups
Value-Based Segmentation	Economic value of specific customer groups on the business

BEHAVIORAL SEGMENTATION (RFM)

RFM Metrics

RECENCY

The freshness of the customer activity, be it purchases or visits

E.g. Time since last order or last engaged with the product

FREQUENCY

The frequency of the customer transactions or visits

E.g. Total number of transactions or average time between transactions/ engaged visits

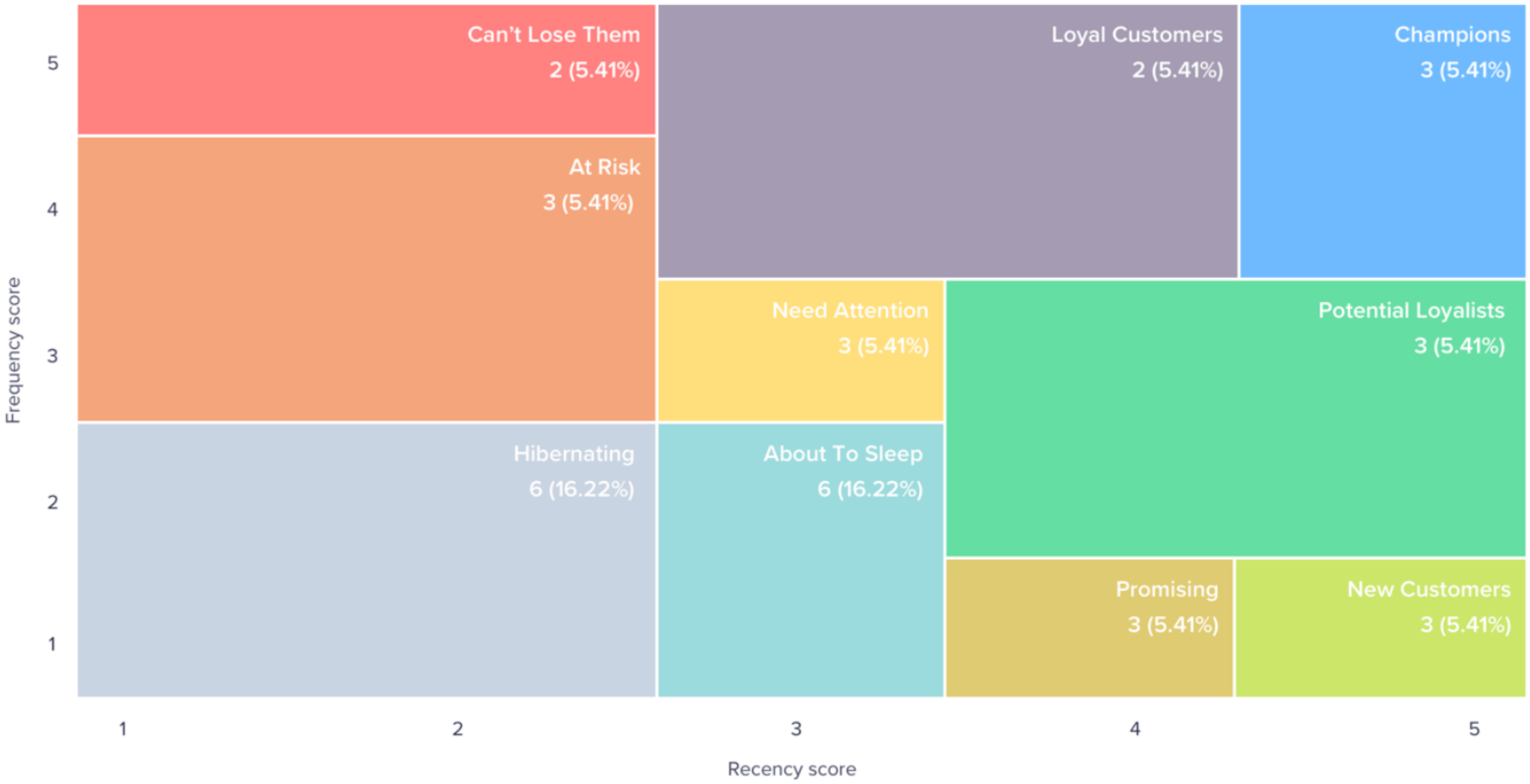
MONETARY

The intention of customer to spend or purchasing power of customer

E.g. Total or average transactions value

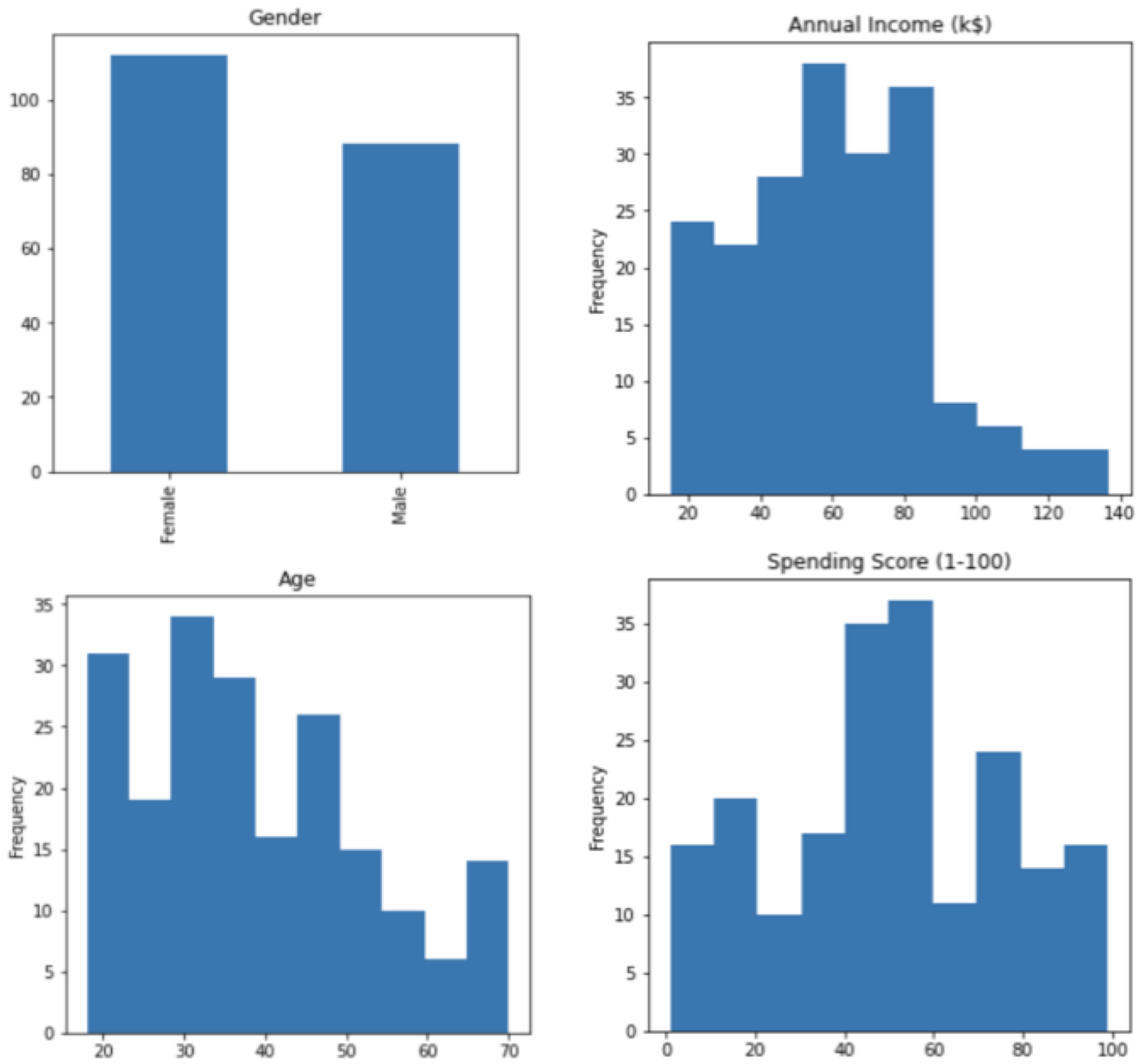
CUSTOMER ID	RECENCY (DAY)	FREQUENCY (NUMBER)	MONETARY (TOTAL)
1	4	6	540
2	6	11	940
3	46	1	35
4	23	3	65
5	15	4	179
6	32	2	56
7	7	3	140
8	50	1	950
9	34	15	2630
10	10	5	191
11	3	8	845
12	1	10	1510
13	27	3	54
14	18	2	40
15	5	1	25

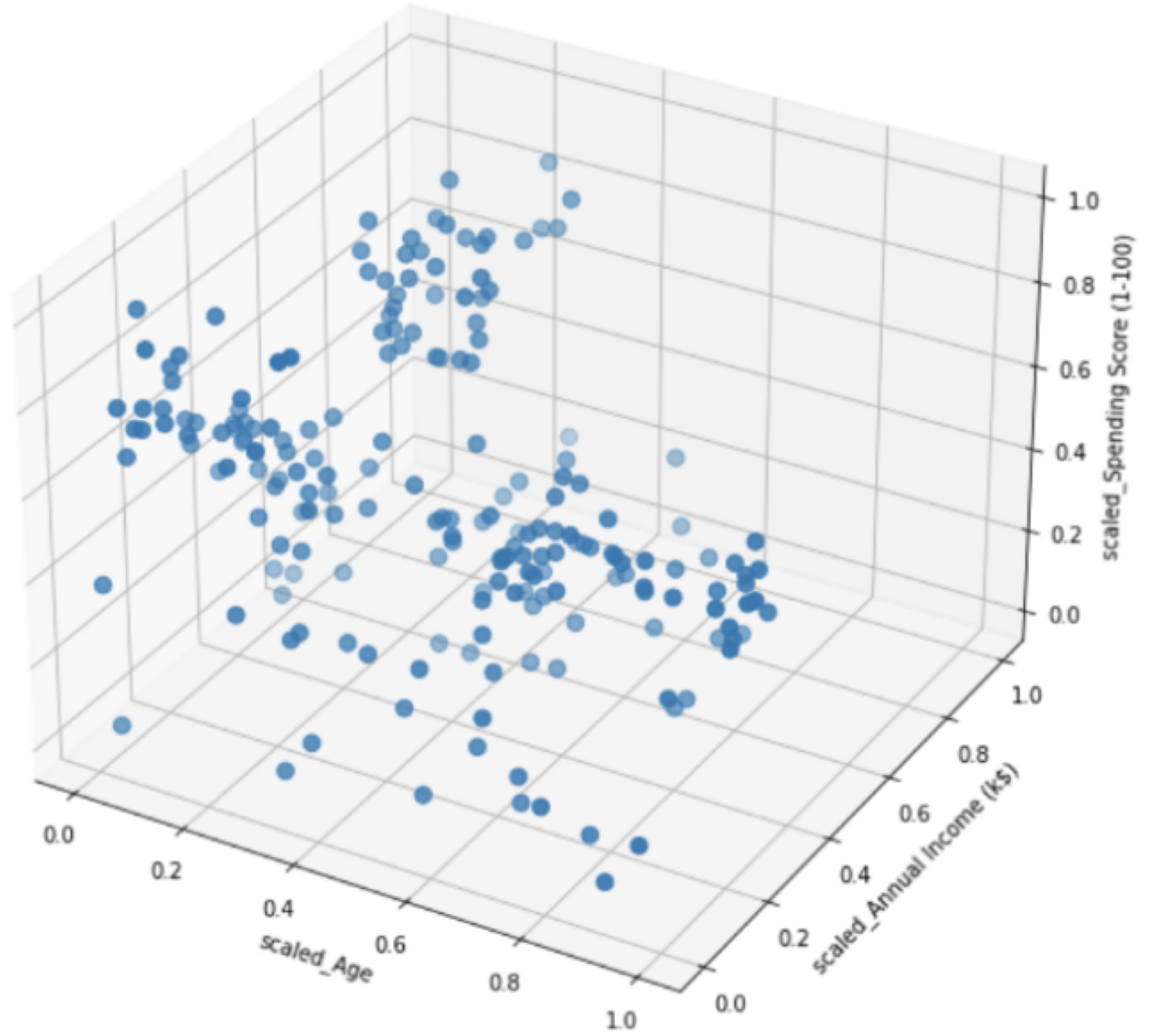
RFM SEGMENTS ANALYSIS



SEGMENTATION BY HISTOGRAM

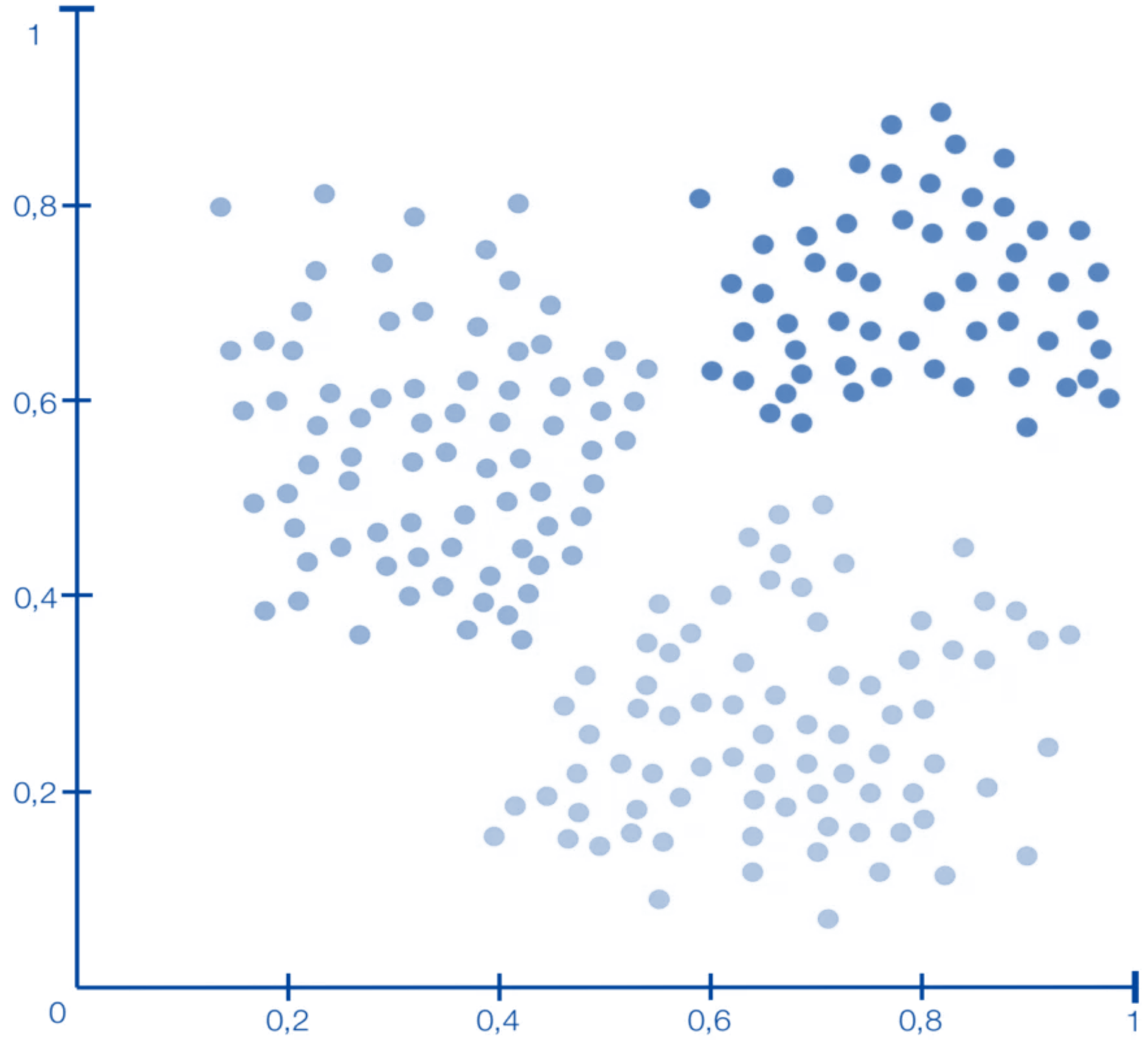
Univariate analysis





CLUSTERING

Unsupervised learning – no target variable, the goal is to discover patterns in data



Problem set-up:

- Result: assignment to clusters
- Predictors: numerical and categorical

Examples of clustering algorithms

- K-means
- Hierarchical clustering
- DBSCAN
- Spectral clustering
- Gaussian

DISTANCE AND SIMILARITY

Euclidean distance:

$$d_{euc}(x,y) = \sqrt{\sum_{i=1}^{n}(x_i - y_i)^2}$$

Manhattan distance:

$$d_{man}(x,y) = \sum_{i=1}^{n}|(x_i - y_i)|$$

Pearson correlation distance:

$$d_{cor}(x,y) = 1 - \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2 \sum_{i=1}^{n}(y_i - \bar{y})^2}}$$

Attribute	Person A	Person B
Sex	Male	Female
Age	23	40
Years at current address	2	10
Residential status (1=Owner, 2=Renter, 3=Other)	2	1
Income	50,000	90,000

K-MEANS CLUSTERING

Problem set-up:

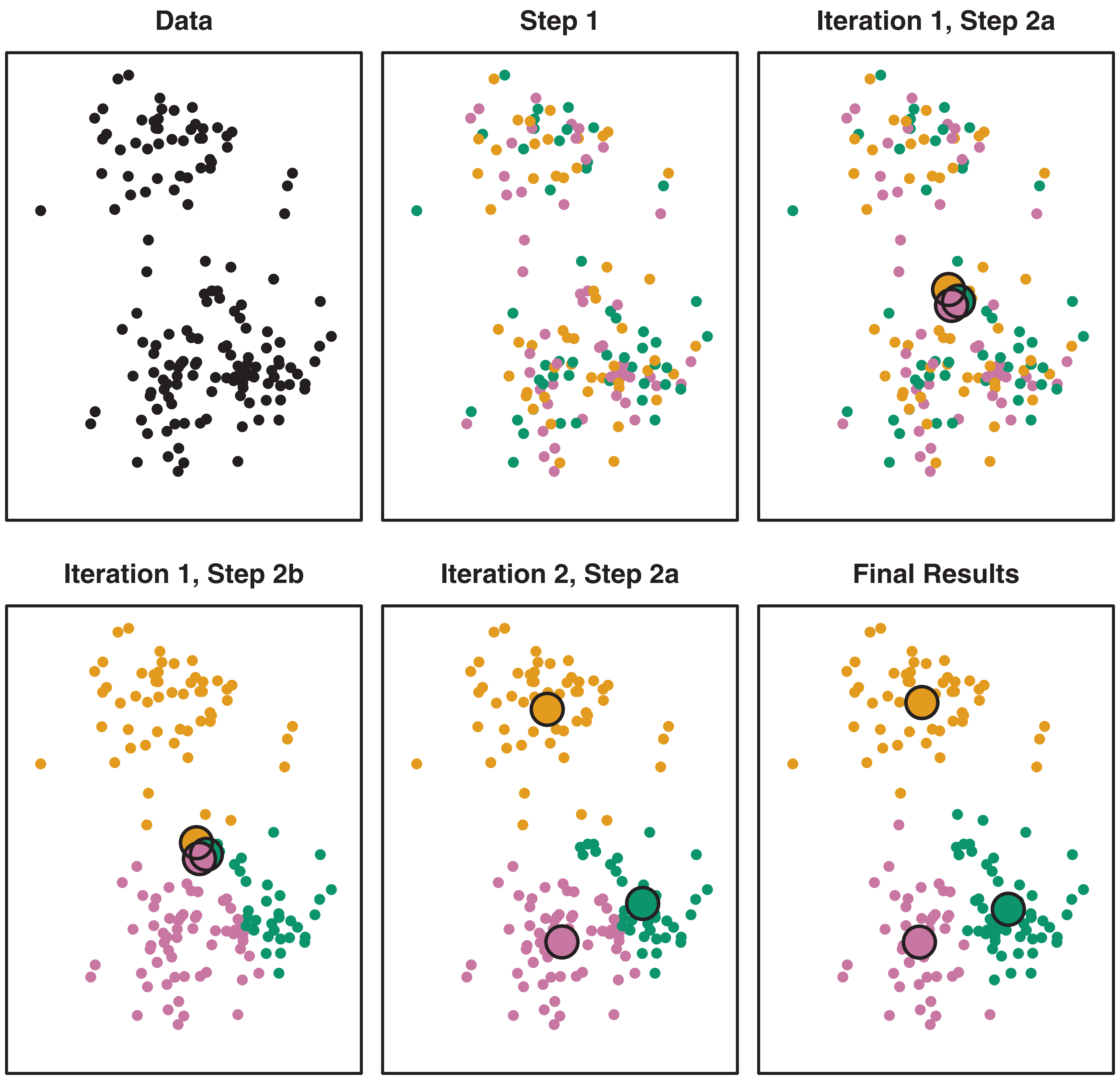
- Result: assignment to clusters (each observation to a single cluster)
- Predictors: numerical
- Pre-specified number of clusters !

Algorithm:

- Minimizing total within-cluster Euclidian distances
- Select k points randomly as initial centroids
- Assign each data point to nearest centroid
- Compute the new centroid location as mean of cluster
- Iterate until converge (no shift in centroid position)

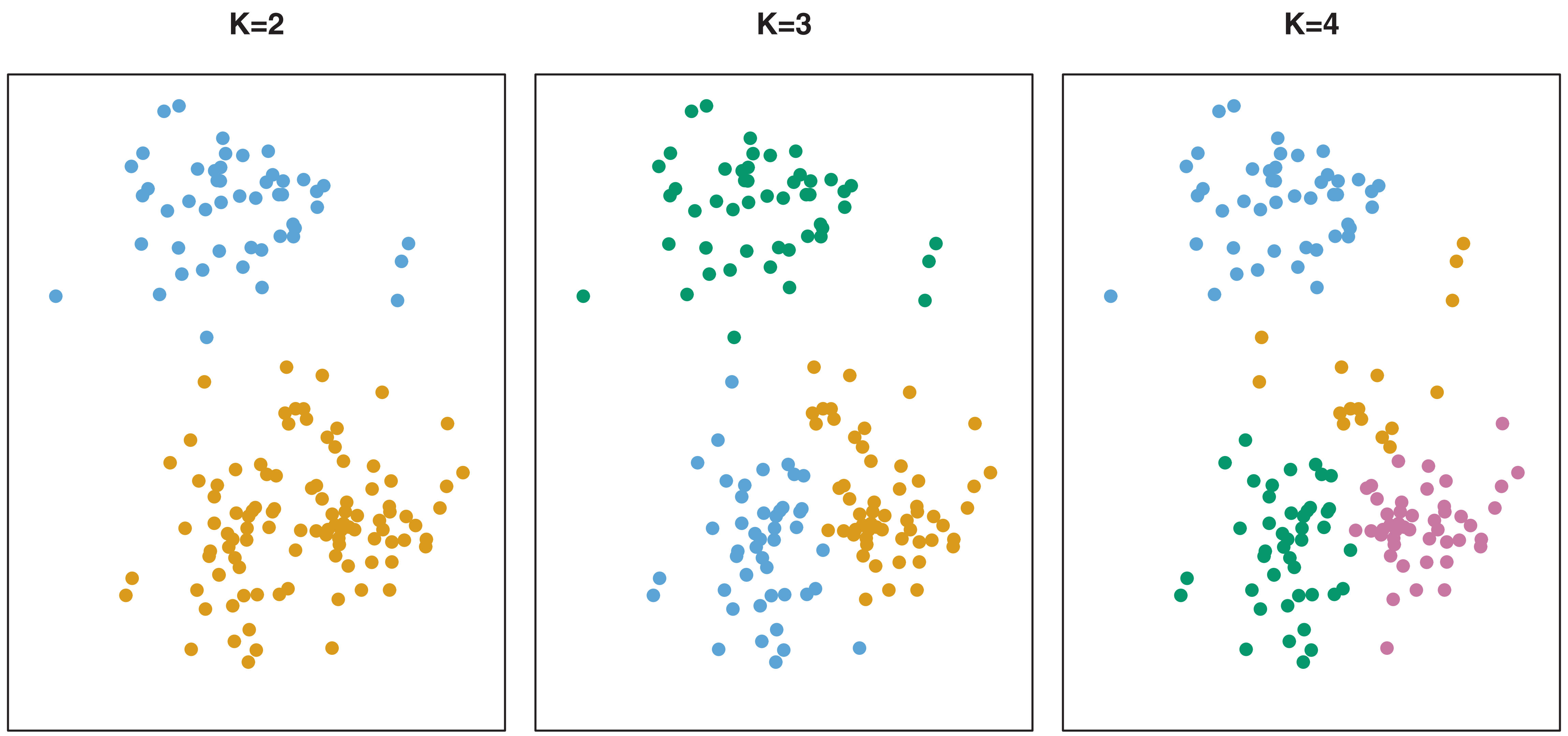
$$W(C_k) = \sum_{x_i \in C_k} (x_i - \mu_k)^2$$

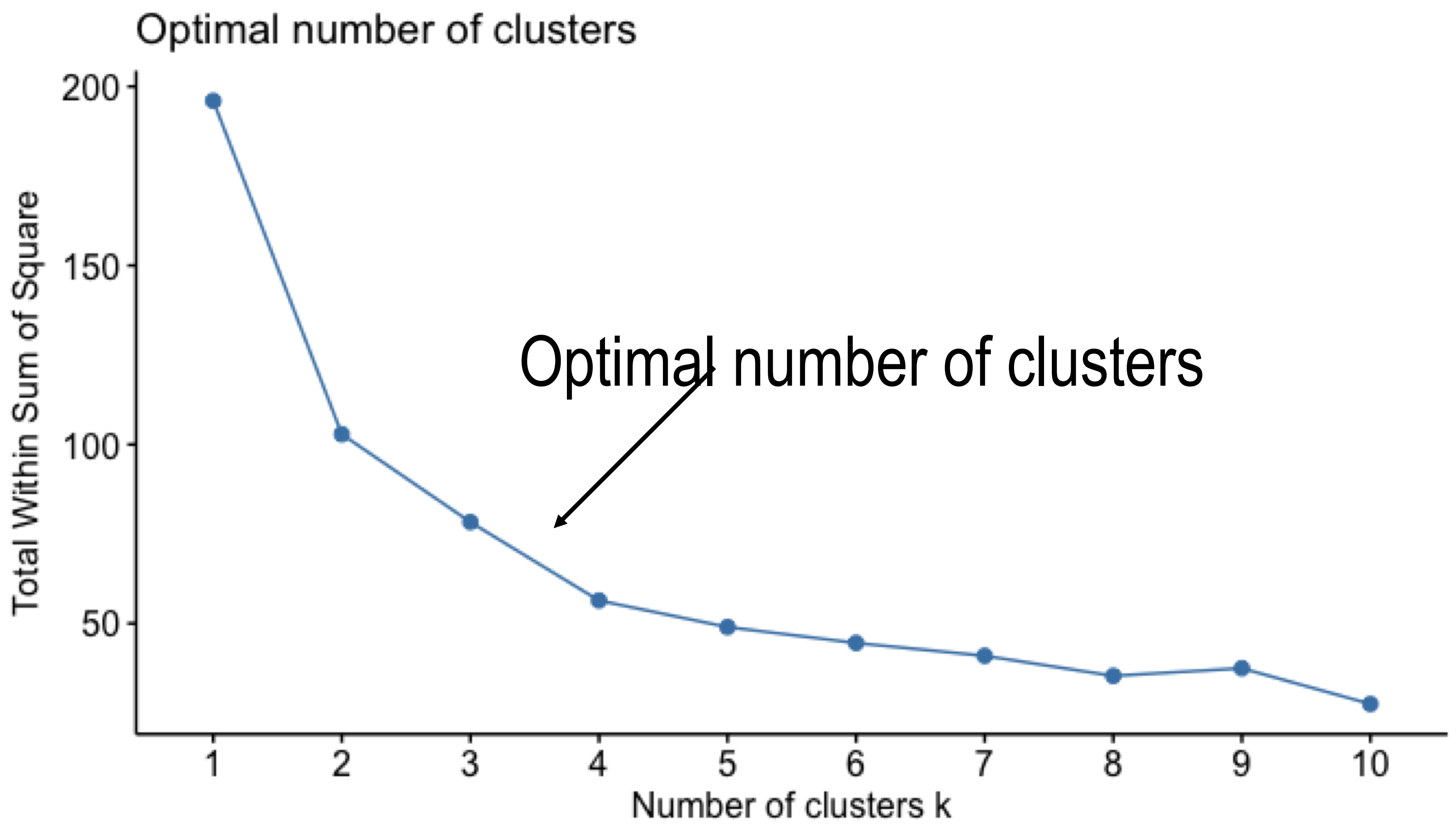
$$minimize\left(\sum_{k=1}^{k} W(C_k)\right)$$



OPTIMAL NUMBER OF CLUSTERS

K-means clustering example



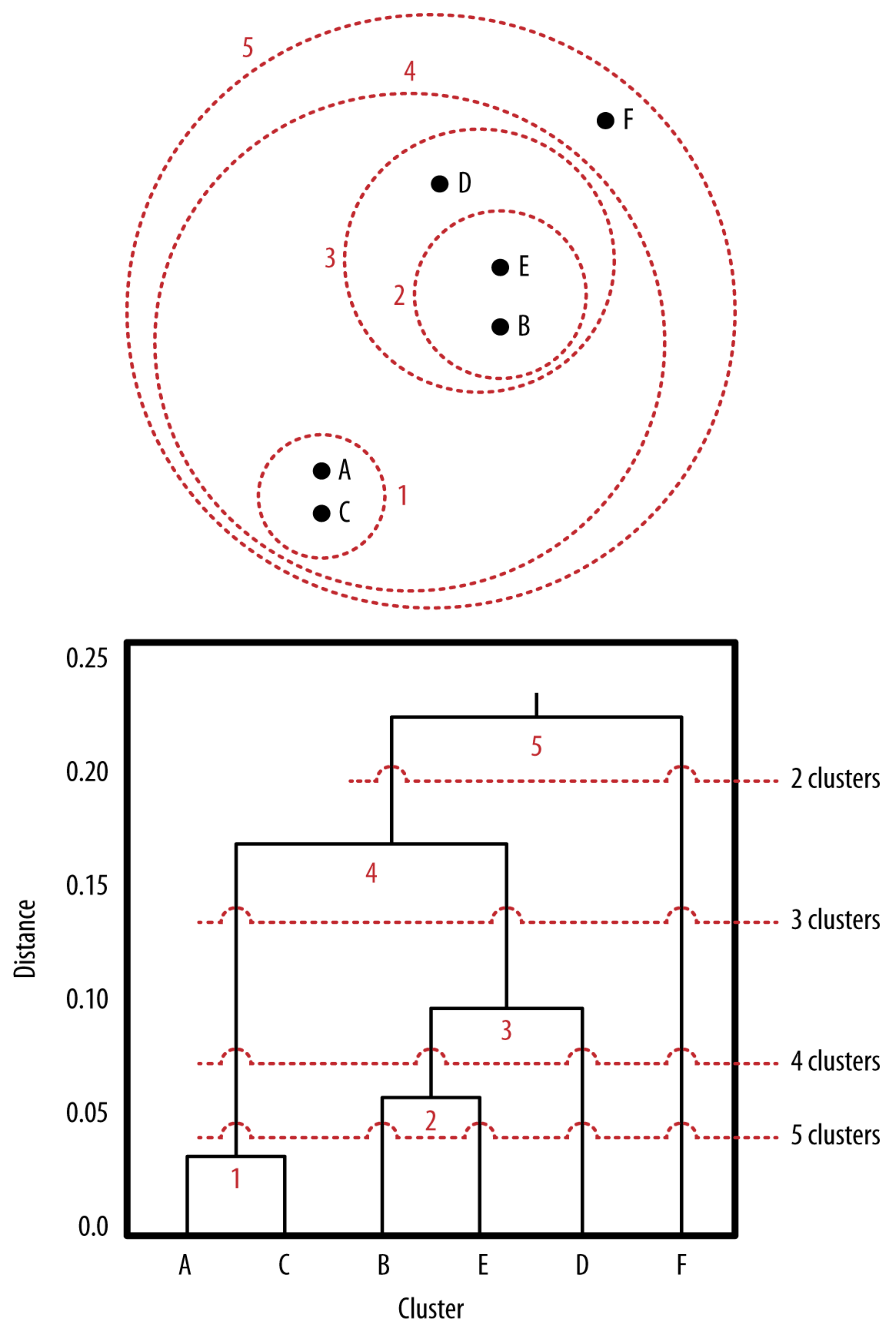


- Elbow method $\left(\sum_{k=1}^{k} W(C_k)\right)$

AGGLOMERATIVE CLUSTERING

Problem set-up:
- Result: dendrogram – tree structure
- Predictors: numerical or categorical – need similarity
- Does not require number of clusters!
- Need post-processing to generate clusters

Algorithm:
- Bottom up, agglomerative construction
- Vertical y-axis – distance/dissimilarity
- Steps:
  - start with each point in its own cluster
  - merge two most similar clusters; iterate until one left
- Linkage: single, complete, average, centroid
- Hierarchical

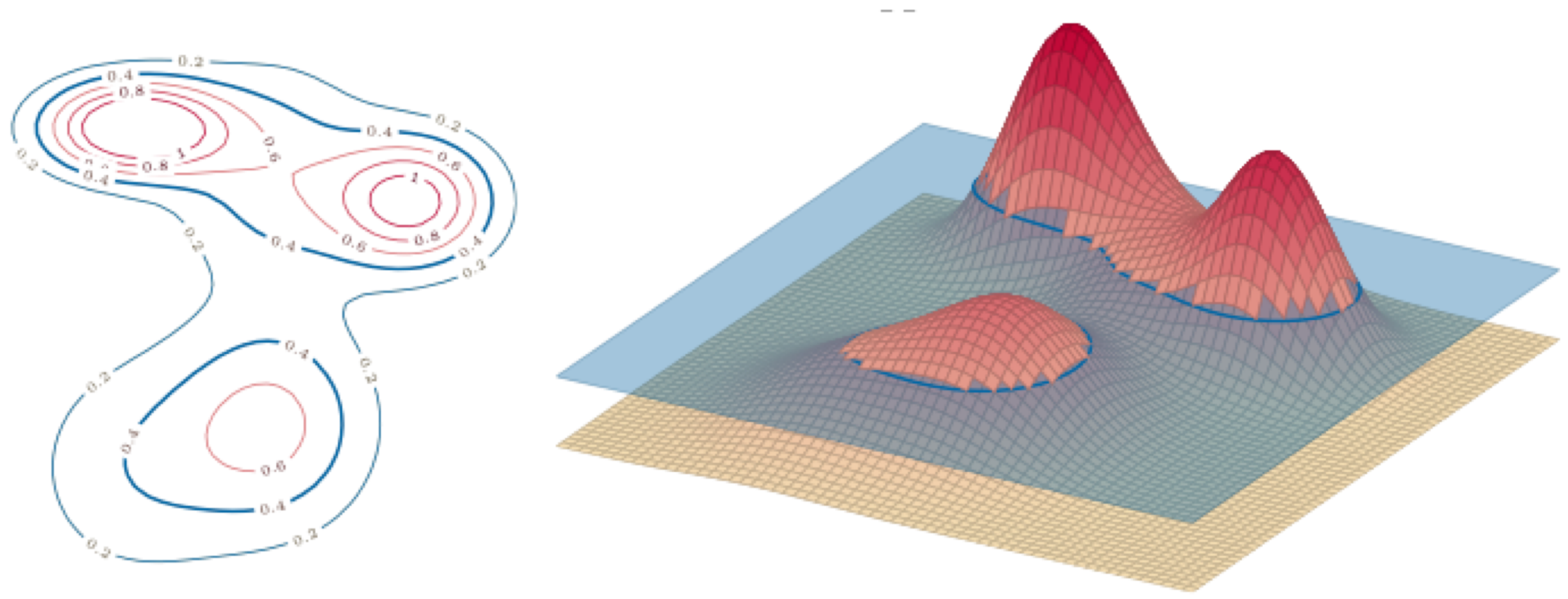


DBSCAN

Density-based spatial clustering and application with noise

Problem set-up:

- Two parameters:
- neighborhood radius eps
- min number of points to form a cluster
- Does not require number of clusters!
- Works with any distance/similarity functions
- All points within a cluster are

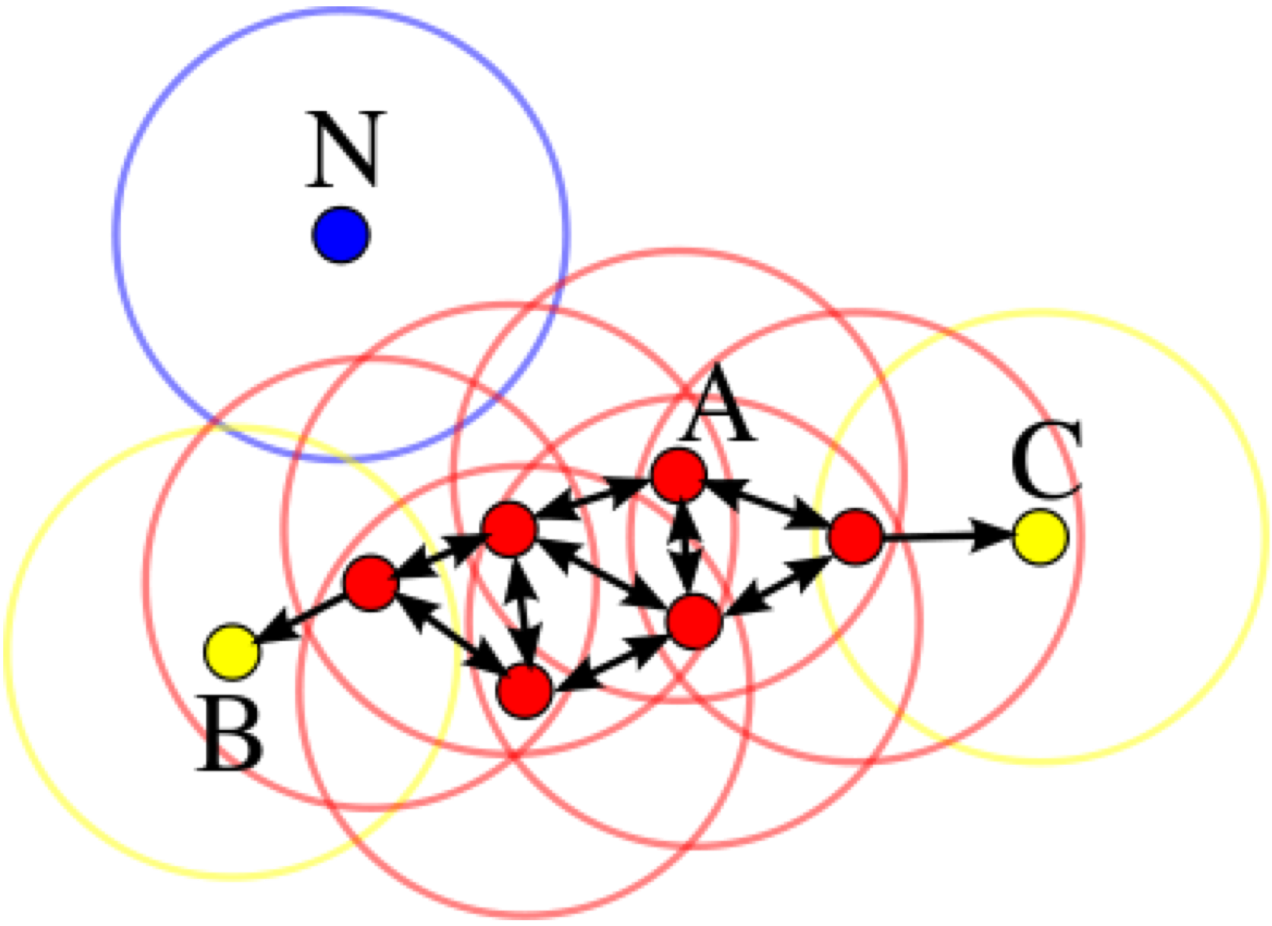


(a) Contour plot

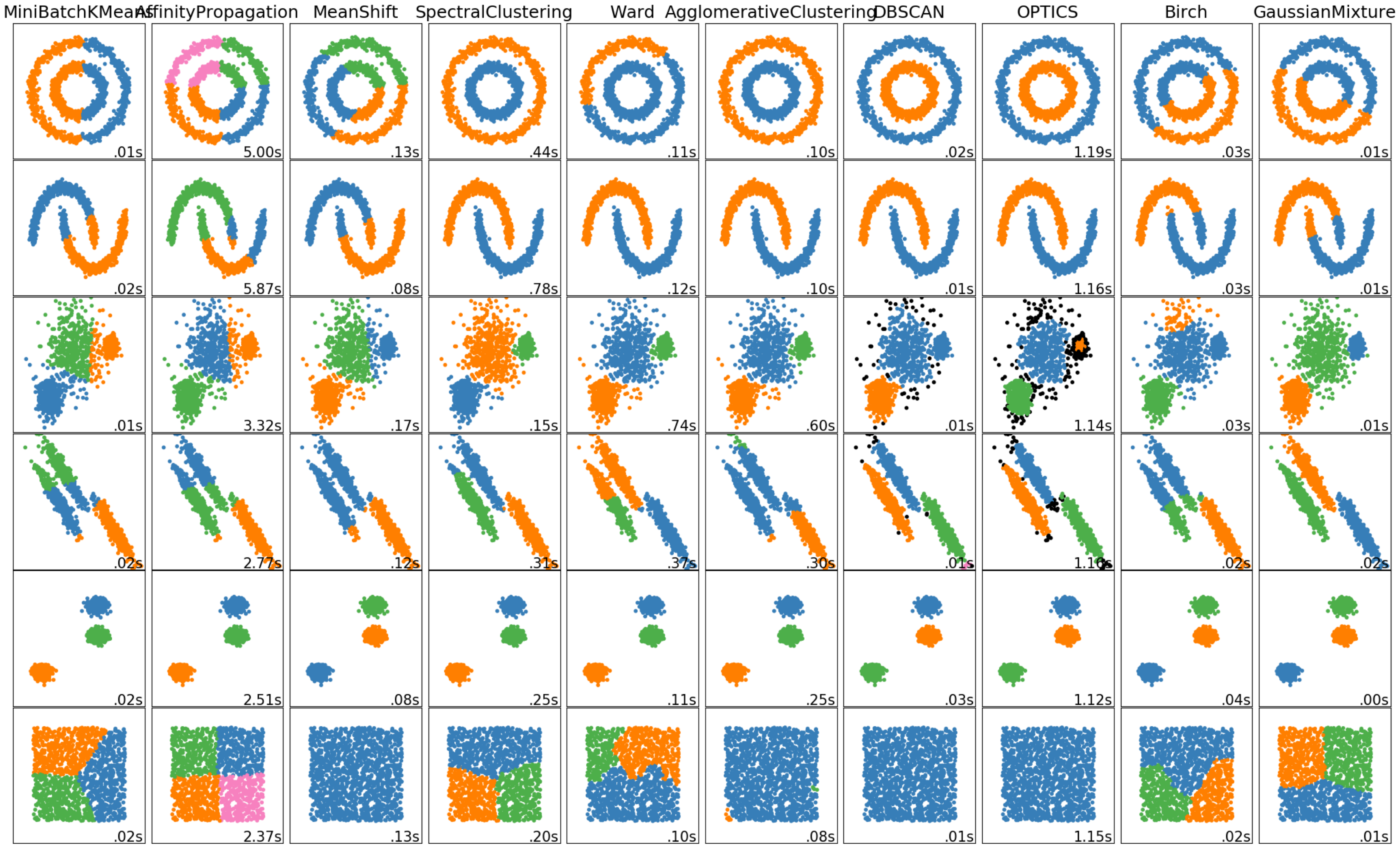
(b) Density model with level set

Algorithm:

- Minimizing total within-cluster Euclidian distances
- Randomly select a point and if enough points in eps start a cluster (core point), if not – noise
- Add all points in eps neighborhood to the cluster, iterate until all reachable points found
- Continue with unvisited pointes



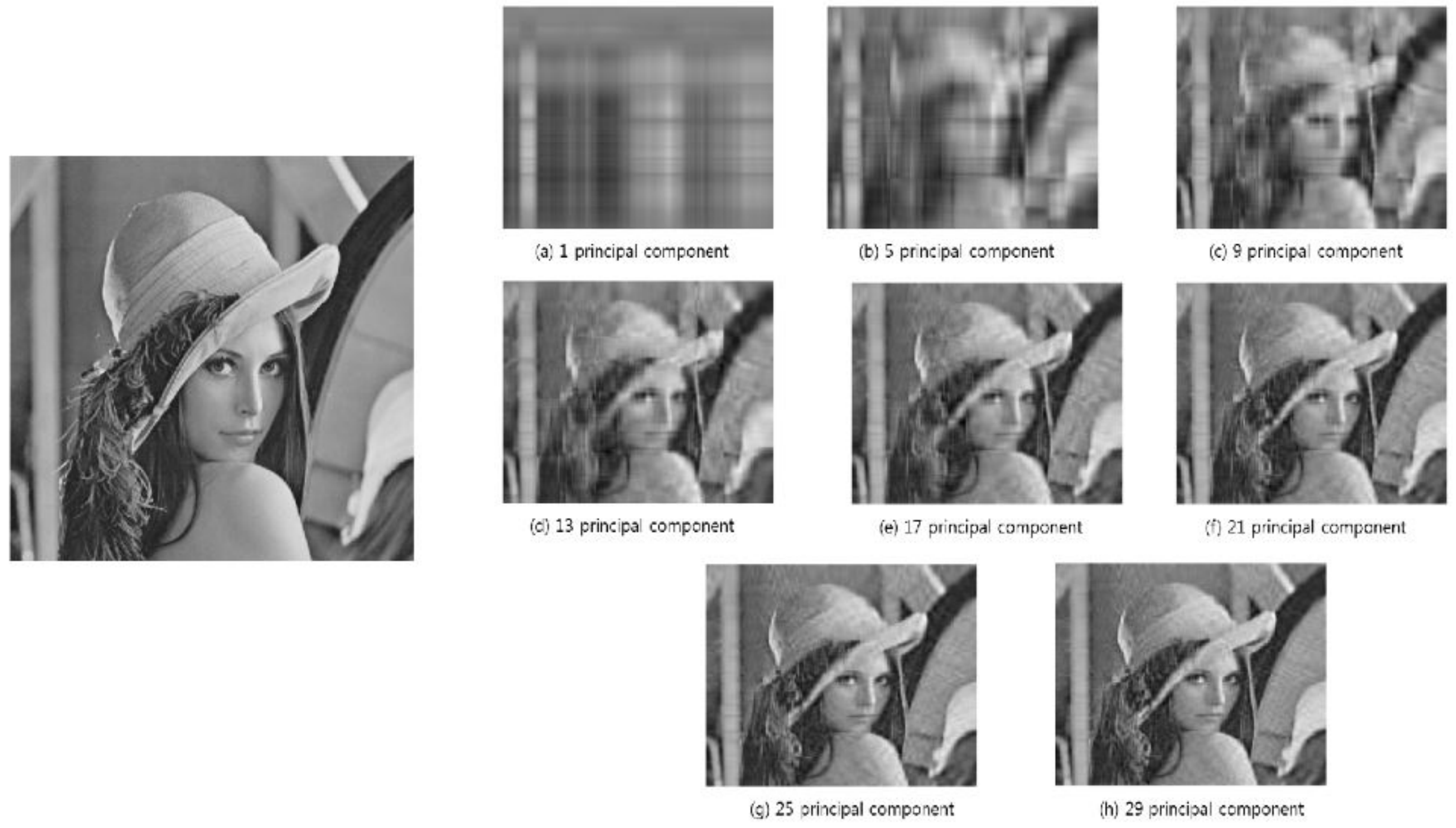
TESTING CLUSTERING METHODS

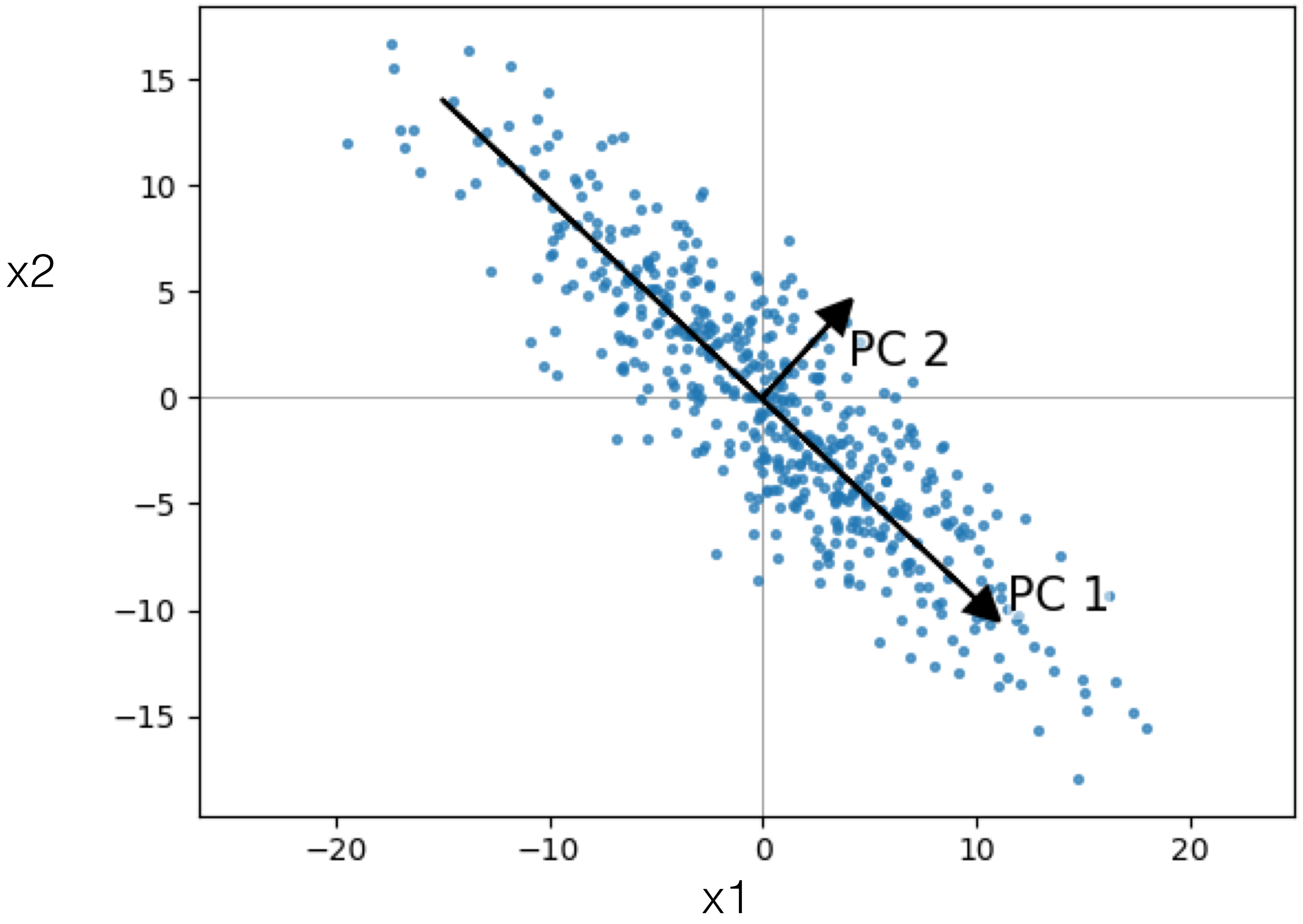


PRINCIPAL COMPONENT ANALYSIS (PCA)

Data dimensionality reduction

- Principal components analysis – construction of orthogonal basis along the directions of largest variance
- Computed by eigenvalue decomposition of covariance matrix
- Dimensionality reduction – reconstruction of data with less components





PRINCIPAL COMPONENT ANALYSIS (PCA)

Maximizing variance

X (n x p), n – points/measurements, p – features
row – single measurements

$$t_{k(i)} = \mathbf{x}_{(i)} \cdot \mathbf{w}_{(k)}$$

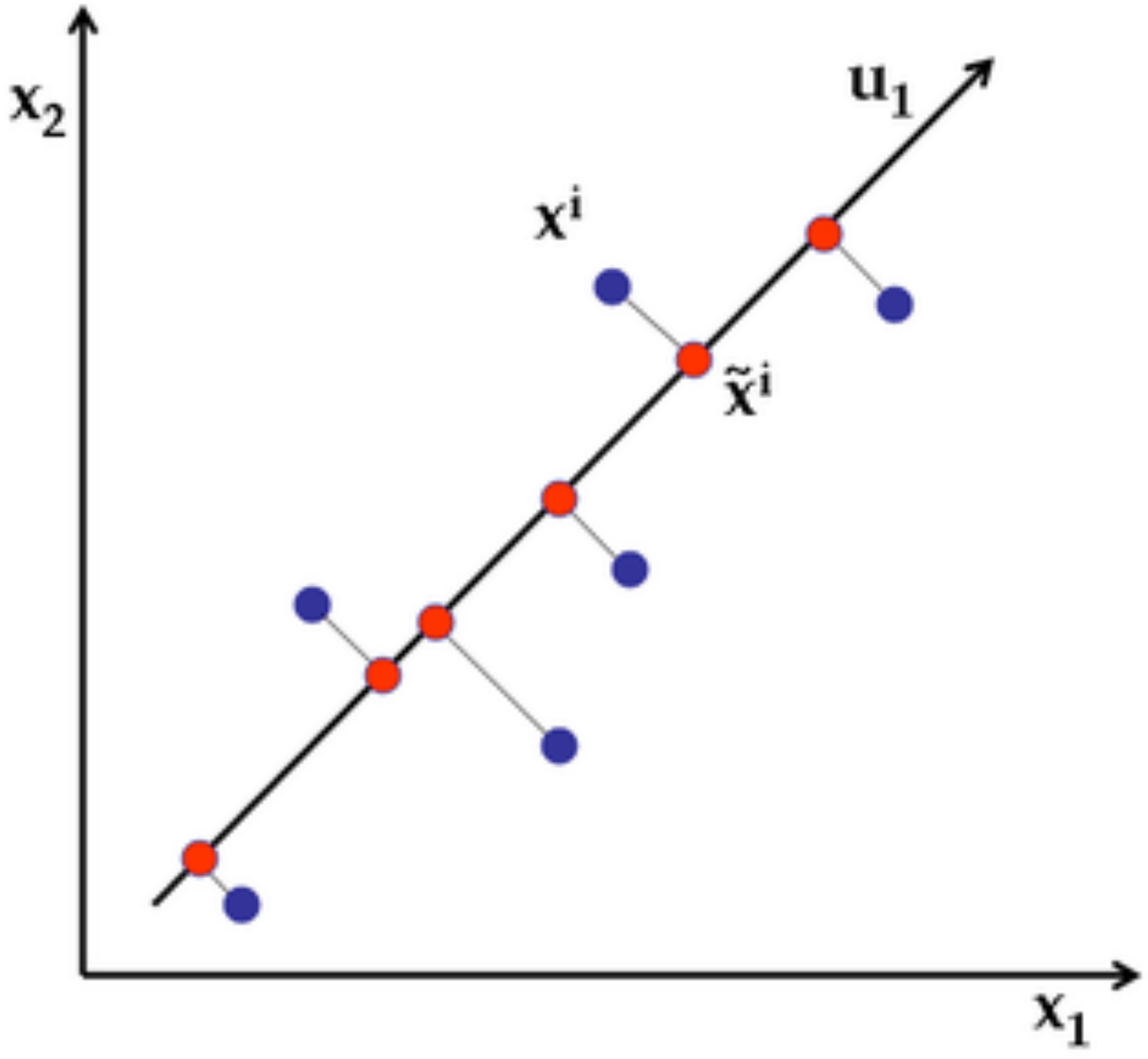
$$\mathbf{w}_{(1)} = \arg\max_{\Vert\mathbf{w}\Vert=1} \left\{ \sum_i (t_1)^2_{(i)} \right\} = \arg\max_{\Vert\mathbf{w}\Vert=1} \left\{ \sum_i \left(\mathbf{x}_{(i)} \cdot \mathbf{w}\right)^2 \right\}$$

$$\mathbf{w}_{(1)} = \arg\max \left\{ \frac{\mathbf{w}^\mathsf{T}\mathbf{X}^\mathsf{T}\mathbf{X}\mathbf{w}}{\mathbf{w}^\mathsf{T}\mathbf{w}} \right\} \qquad R(M,x) = \frac{x^* M x}{x^* x}.$$

Raileigh quotient – maximum possible value is the largest eigenvalue/eigenvector

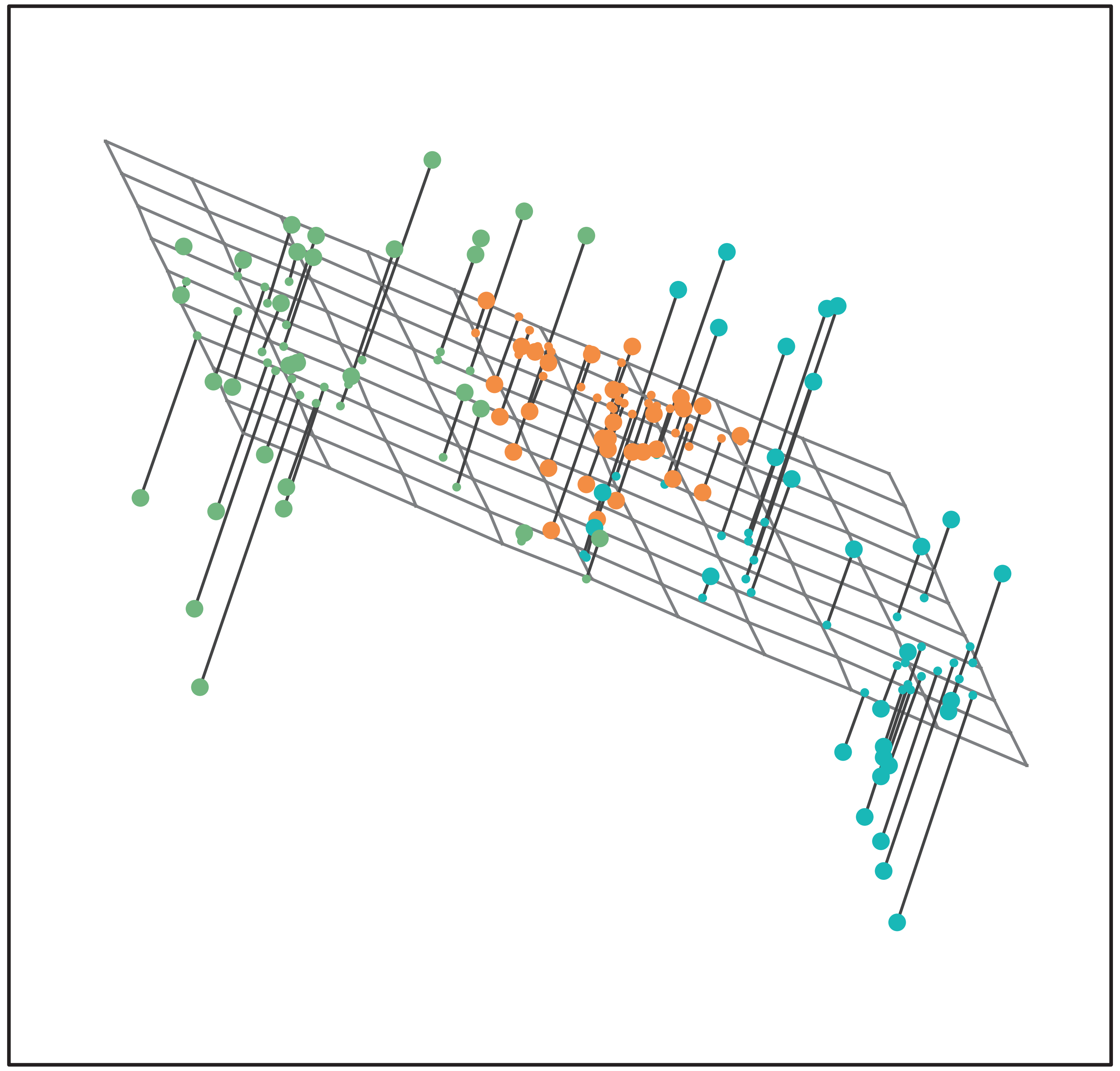
$$\mathbf{X}^\mathsf{T}\mathbf{X} = \mathbf{W}\mathbf{\Lambda}\mathbf{W}^\mathsf{T} \qquad \mathbf{T} = \mathbf{X}\mathbf{W}$$

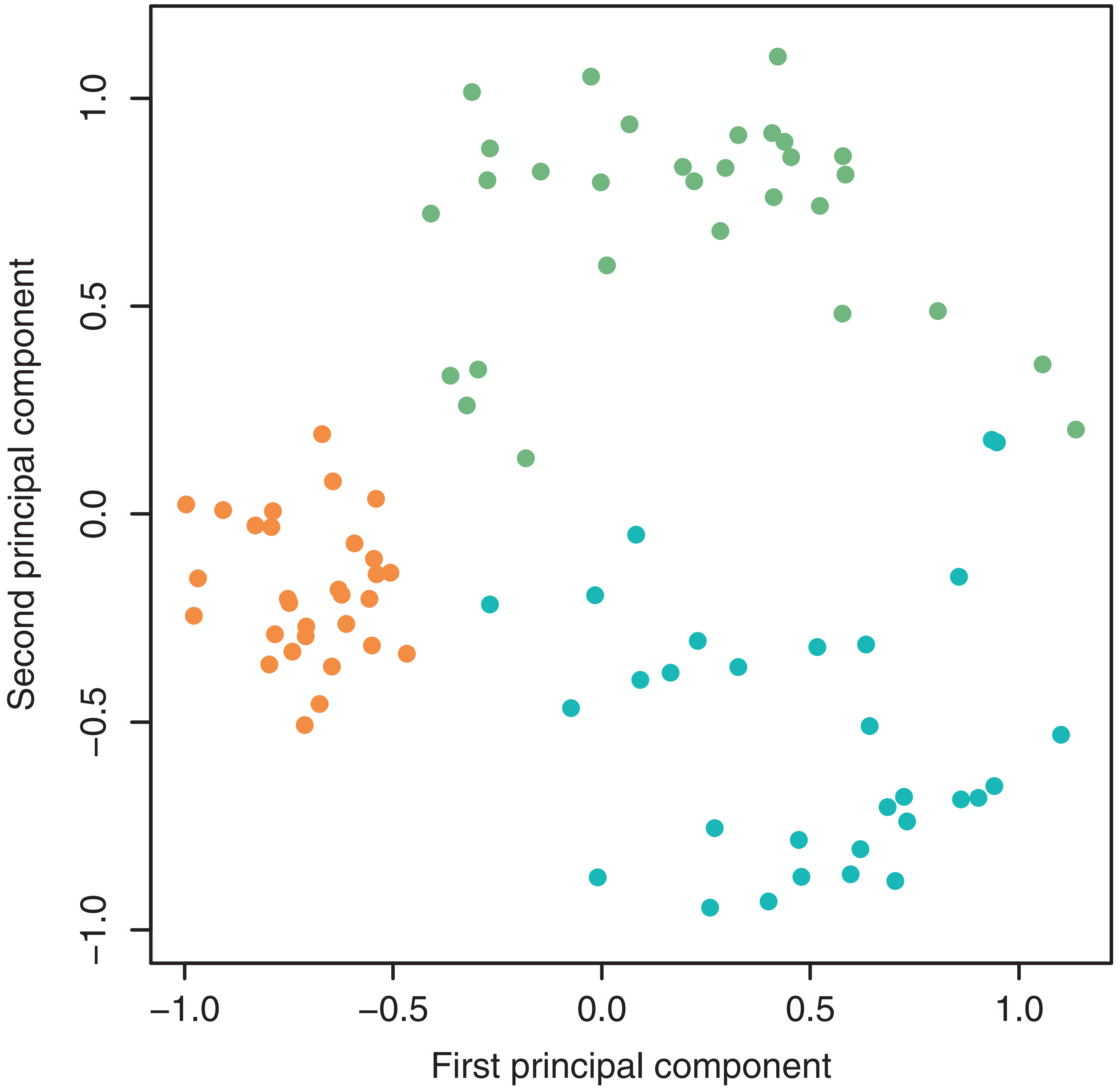
$$\mathbf{T}_L = \mathbf{X}\mathbf{W}_L$$



PRINCIPAL COMPONENT ANALYSIS (PCA)

Dimensionality reduction 3D -> 2D





HIGHER SCHOOL OF ECONOMICS

NATIONAL RESEARCH
UNIVERSITY