

Data we will work with

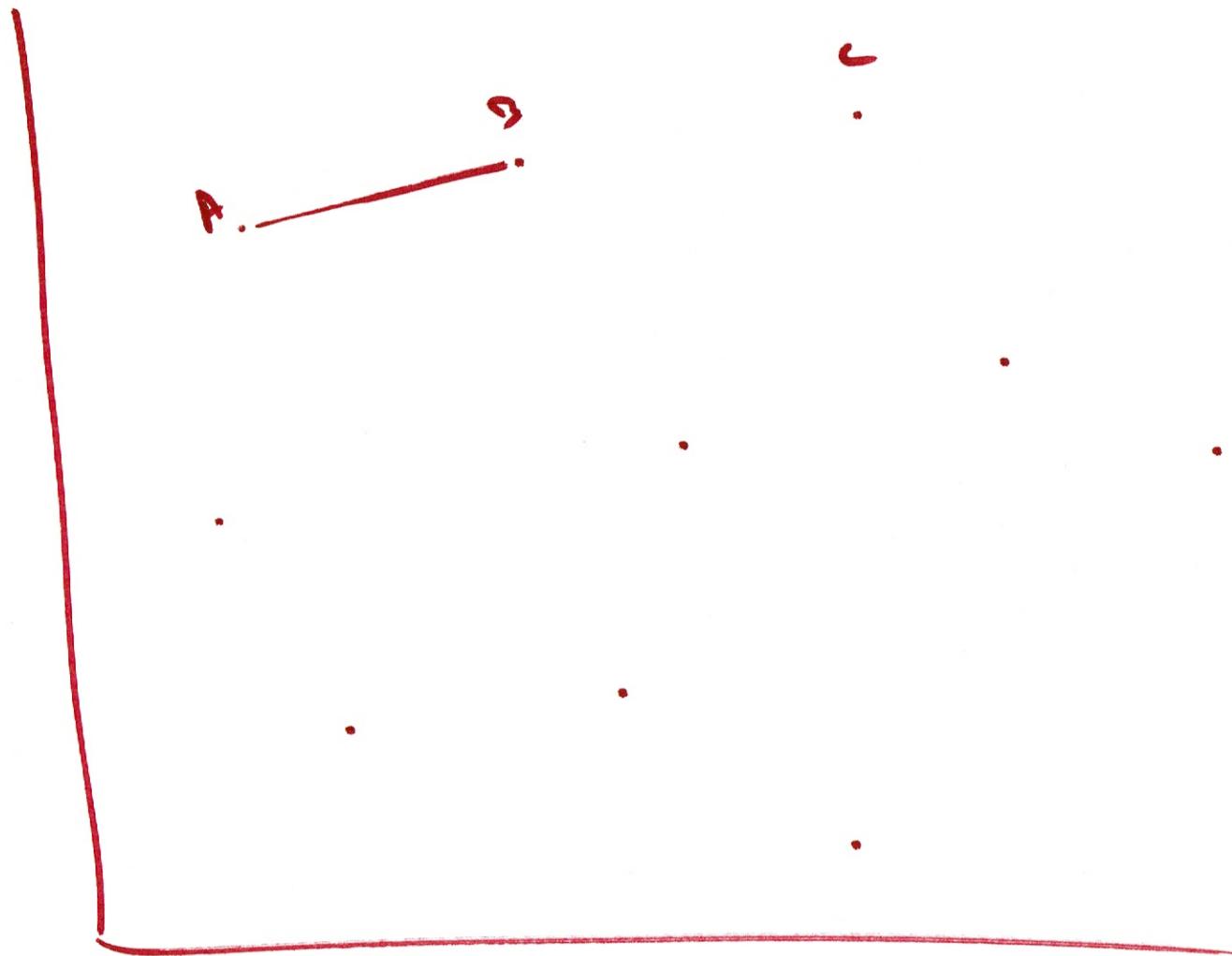
- Customer Spend Data

- AVG_Mthly_Spend: The average monthly amount spent by customer
- No_of_Visits: The number of times a customer visited in a month
- Item Counts: Count of Apparel, Fruits and Vegetable, Staple Items purchased

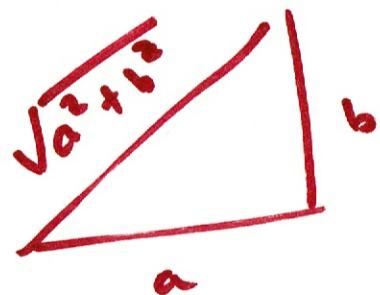


	Cust_ID	Name	Avg_Mthly_Spend	No_Of_Visits	Apparel_Items	FnV_Items	Staples_Items
1	1	A	10000	2	1	1	0
2	2	B	7000	3	0	10	9
3	3	C	7000	7	1	3	4
4	4	D	6500	5	1	1	4
5	5	E	6000	6	0	12	3
6	6	F	4000	3	0	1	8
7	7	G	2500	5	0	11	2
8	8	H	2500	3	0	1	1
9	9	I	2000	2	0	2	2
10	10	J	1000	4	0	1	7

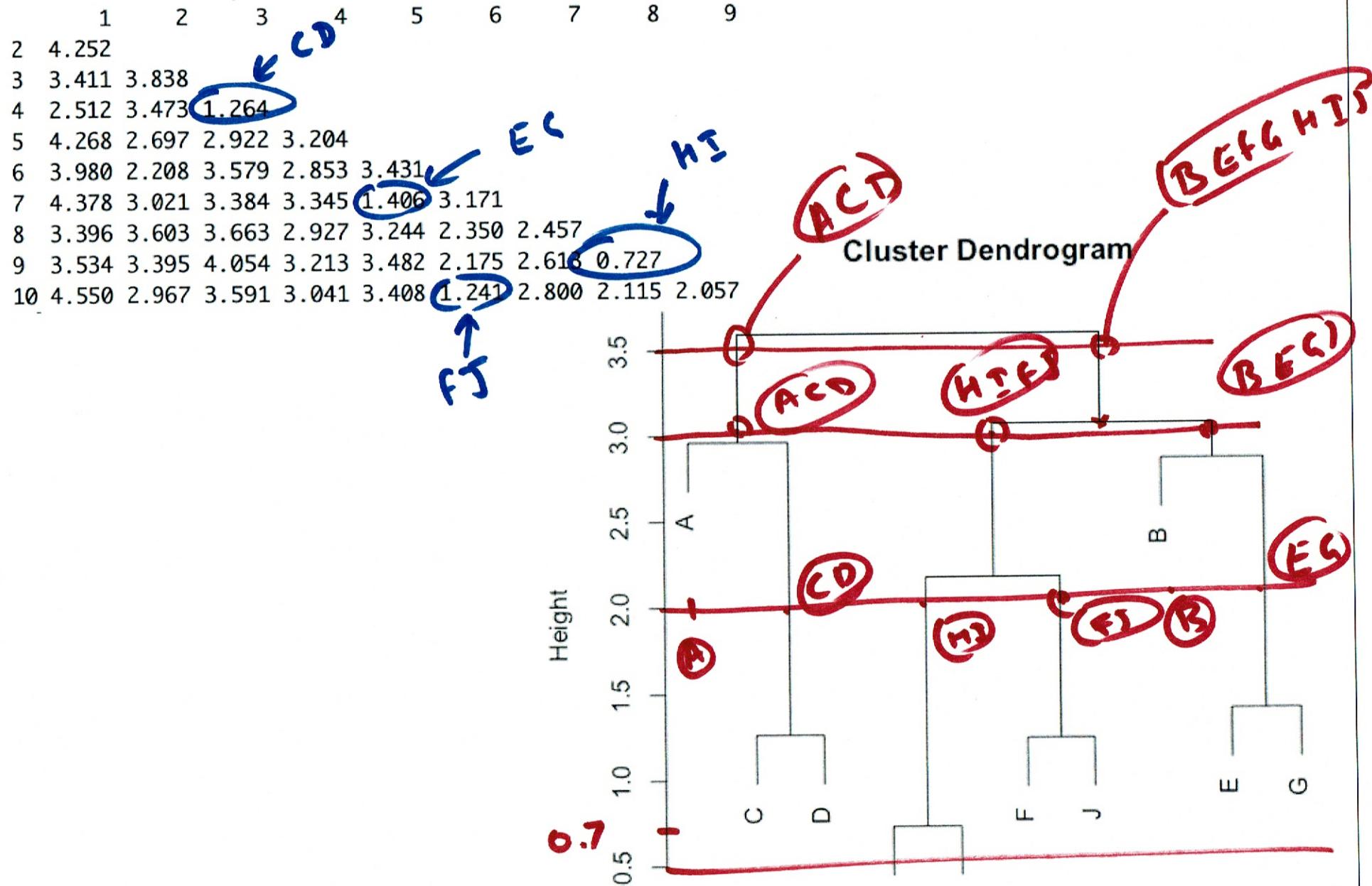
- Can we cluster similar customers together?

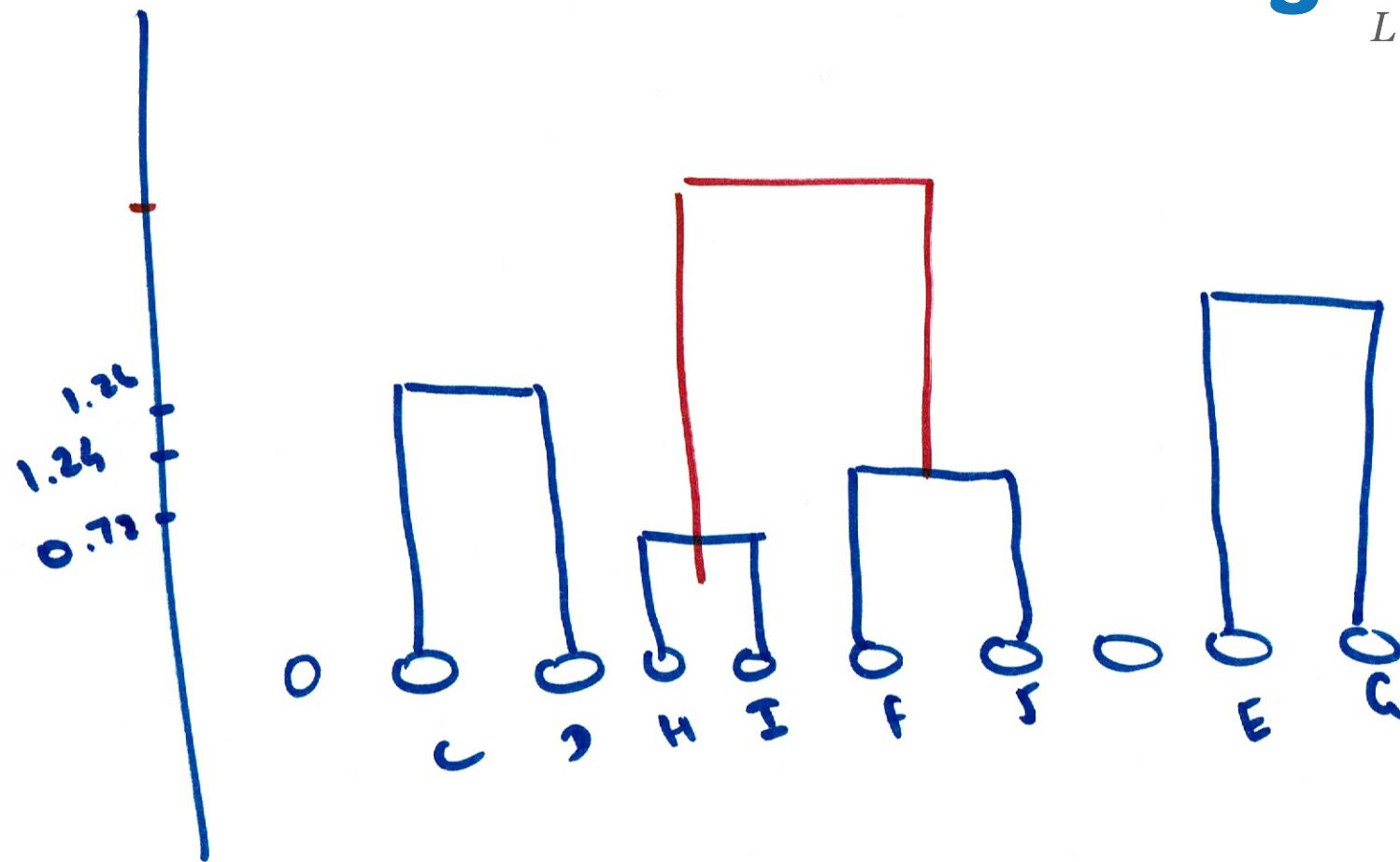


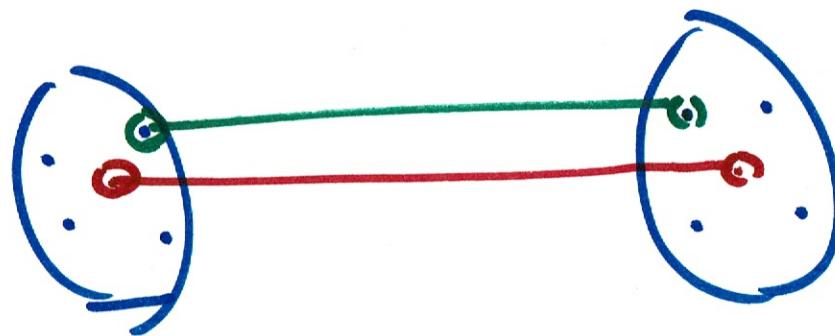
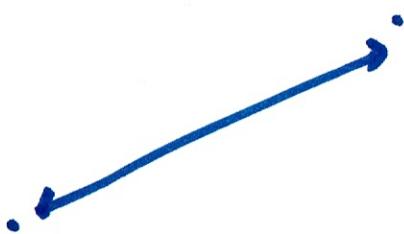
$$\text{dist } A + B = \sqrt{\underbrace{(10000 - 7000)^2}_{+ (2-3)^2 + (1-0)^2} + \underbrace{(2-3)^2 + (1-0)^2}_{+ (1-10)^2 + (0-9)^2}}$$



Distance between objects







Distance between clusters

- Single linkage – Minimum distance or Nearest neighbor
- Complete linkage – Maximum distance or Farthest distance
- Average linkage – Average of the distances between all pairs
- Centroid method – combine cluster with minimum distance between the centroids of the two clusters
- Ward's method – Combine clusters with which the increase in within cluster variance is to the smallest degree

