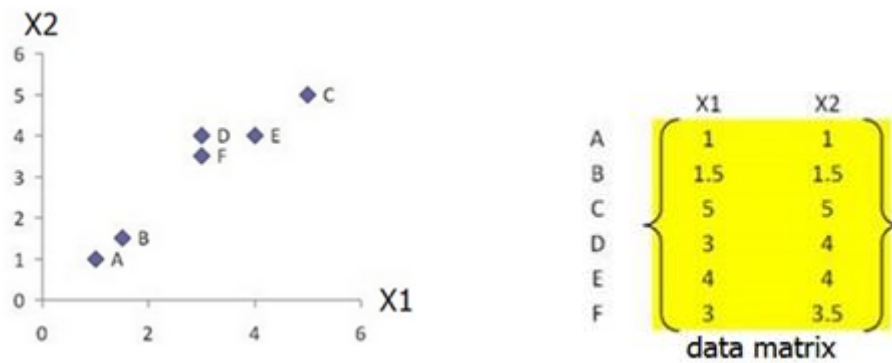


## Clustering analysis with agglomerative clustering



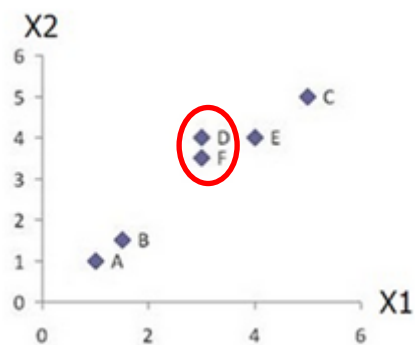
Manhattan distance:  $|x_1 - x_2| + |y_1 - y_2|$

Distance Matrix

Dist	A	B	C	D	E	F
A	0.00	1.00	8.00	7.00	6.00	4.50
B	1.00	0.00	7.00	4.00	5.00	3.50
C	8.00	7.00	0.00	3.00	2.00	3.50
D	7.00	4.00	3.00	0.00	1.00	0.50
E	6.00	5.00	2.00	1.00	0.00	1.50
F	4.50	3.50	3.50	0.50	1.50	0.00

### ITERATION 1

Merge two closest clusters (iteration 1)



Update distance matrix (iteration 1)

Dist	A	B	C	D,F	E
A	0.00	1.00	8.00	4.50	6.00
B	1.00	0.00	7.00	3.50	5.00
C	8.00	7.00	0.00	3.00	2.00
D,F	4.50	3.50	3.00	0.00	1.00
E	6.00	5.00	2.00	1.00	0.00

$$d(D,F) - A = \min (d(D,A), d(F,A)) = \min (7.00, 4.50) = 4.50$$

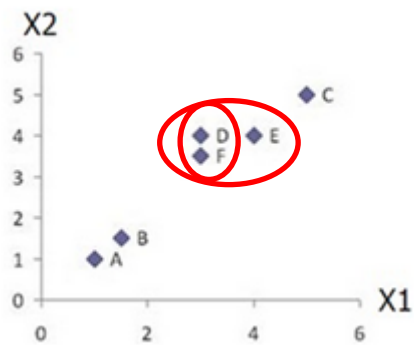
$$d(D,F) - B = \min (d(D,B), d(F,B)) = \min (4.00, 3.50) = 3.50$$

$$d(D,F) - C = \min (d(D,C), d(F,C)) = \min (3.00, 3.50) = 3.00$$

$$d(D,F) - E = \min (d(D,E), d(F,E)) = \min (1.00, 1.50) = 1.00$$

## ITERATION 2

Merge two closest clusters (iteration 2)



Update distance matrix (iteration 2)

Dist	A	B	C	(D,F),E
A	0.00	1.00	8.00	4.50
B	1.00	0.00	7.00	3.50
C	8.00	7.00	0.00	2.00
(D,F),E	4.50	3.50	2.00	0.00

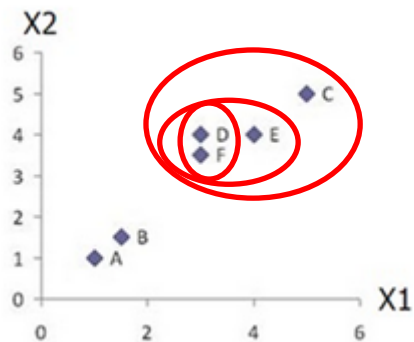
$$d((D,F),E) - A = \min (d((D,F),A), d(E,A)) = \min (4.50, 6.00) = 4.50$$

$$d((D,F),E) - B = \min (d((D,F),B), d(E,B)) = \min (3.50, 5.00) = 3.50$$

$$d((D,F),E) - C = \min (d((D,F),C), d(E,C)) = \min (3.00, 2.00) = 2.00$$

### ITERATION 3

Merge two closest clusters (iteration 3)



Update distance matrix (iteration 2)

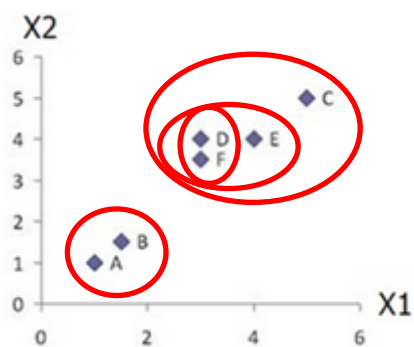
Dist	A	B	((D,F),E),C
A	0.00	1.00	4.50
B	1.00	0.00	3.50
((D,F),E),C	4.50	3.50	0.00

$$d((D,F),E),C - A = \min(d(((D,F),E),A), d(C,A)) = \min(4.50, 8.00) = 4.50$$

$$d((D,F),E),C - B = \min(d(((D,F),E),B), d(C,B)) = \min(3.50, 7.00) = 3.50$$

### ITERATION 4

Merge two closest clusters (iteration 4)

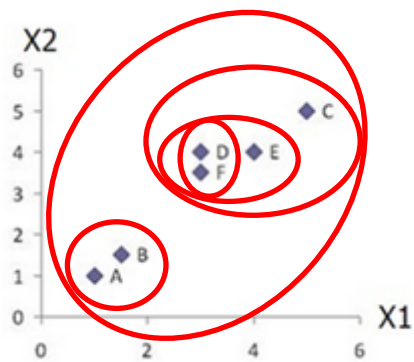


Update distance matrix (iteration 4)

Dist	A,B	((D,F),E),C
A,B	0.00	3.50
((D,F),E),C	3.50	0.00

$$d((A,B) - ((D,F),E),C) = \min(d(((D,F),E),C),A), d(((D,F),E),C),B)) = \min(4.50, 3.50) = 3.50$$

## FINAL RESULT



### When you use Euclidean distance, are the results different?

When using Euclidean distance, the results are the same. However, with Manhattan distance, we had two possible options for selecting the cluster, whereas Euclidean distance only allowed us to choose the cluster [A,B]. Naturally, the distances differ depending on the type of distance used.