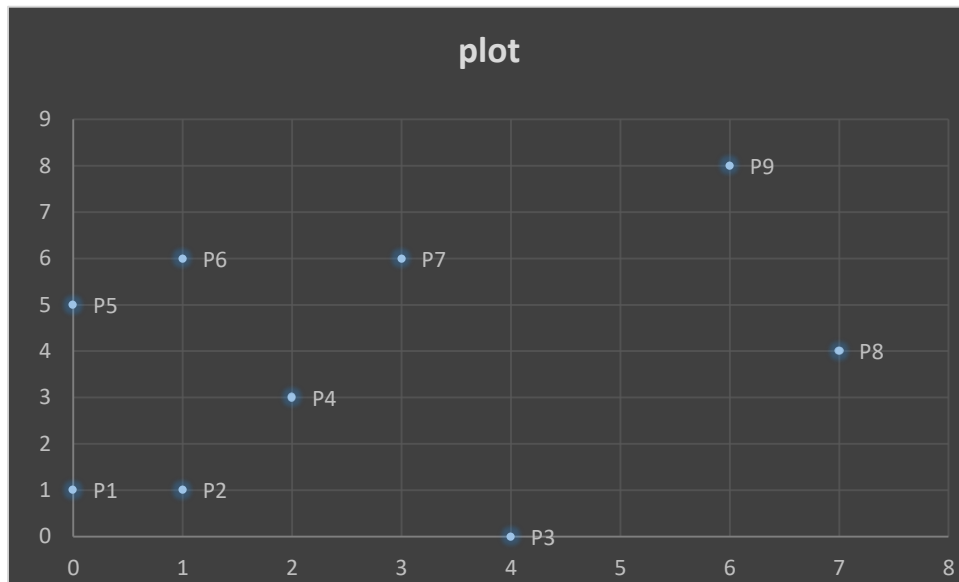


Exercise1



- Calculate the Euclidean distance ie

$$\text{Distance} [(x,y), (a, b)] = \sqrt{(x-a)^2 + (y-b)^2}$$

- I create a euclidean distance matrix for all the values

Distance Matrix

	P1	P2	P3	P4	P5	P6	P7		
P1	0								
P2	1	0							
P3	4.123106	3.162278	0						
P4	2.828427	2.236068	3.605551	0					
P5	5	4.123106	6.403124	2.828427	0				
P6	5.09902	5	6.708204	3.162278	1.414214	0			
P7	5.830952	5.385165	6.082763	3.162278	3.162278	2	0		
P8	7.615773	6.708204	5	5.09902	7.071068	6.324555	4.472136	0	
P9	9.219544	8.602325	8.246211	6.403124	6.708204	5.385165	3.605551	4.123106	0

To find the first cluster we find the classifier with the minimum distance

Min Distance = **1** and the corresponds to **P1** and **P2**

Therefore the first cluster is (P1, P2)

We again update the distance Matrix $\text{MIN}[\text{dist}(\text{P1}, \text{P2}), \text{P3}]$

$$= \text{MIN}[\text{dist}(\text{P1}, \text{P3}), \text{dist}(\text{P2}, \text{P3})]$$

$$= \min(4.123106, 3.162278) = 3.162278$$

We repeat the procedure until we update the distance matrix

Update distance matrix for cluster P1, p2

	P1,P2	P3	P4	P5	P6	P7	P8	P9
P1,P2	0							
P3	3.162278	0						
P4	2.236068	3.605551	0					
P5	4.123106	6.403124	2.828427	0				
P6	5	6.708204	3.162278	1.414214	0			
P7	5.385165	6.082763	3.162278	3.162278	2	0		
P8	6.708204	5	5.09902	7.071068	6.324555	4.472136	0	
P9	8.602325	8.246211	6.403124	6.708204	5.385165	3.605551	4.123106	0

Min Distance = **1.4142** and it corresponds to **P6** and **P5**.

Therefore the second cluster is (P5, P6)

We again update the distance Matrix $\text{MIN}[\text{dist}(\text{P5, P6}), \text{P1}]$
 $= \text{MIN}[\text{dist}(\text{P5, P1}), (\text{P6, P1})]$
 $= \min[5, 5.083]$
 $= 5$

We again update the distance Matrix $\text{MIN}[\text{dist}(\text{P5, P6}), \text{P2}]$

$= \text{MIN}[\text{dist}(\text{P5, P2}), (\text{P6, P2})]$
 $= \min[4.12, 5]$
 $= 4.12$

Update distance matrix for cluster P5, P6

	P1,P2	P3	P4	P5, P6	P7	P8	P9
P1,P2	0						
P3	4.123106	0					
P4	2.828427	3.605551	0				
P5,P6	4.123106	6.403124	2.828427	0			
P7	5.385165	6.082763	3.162278	2	0		
P8	6.708204	5	5.09902	6.324555	4.472136	0	
P9	8.602325	8.246211	6.403124	5.385165	3.605551	4.123106	0

We again find the minimum Euclidean distance

Min Distance = **2** and it corresponds to **(P6,P5) and P7**

Update distance matrix for cluster (P5, P6), P7

	P1,P2	P3	P4	P5, P6, P7	P8	P9
P1,P2	0					
P3	4.123106	0				
P4	2.828427	3.605551	0			
P5,P6,P7	4.123106	6.082763	2.828427	0		
P8	6.708204	5	5.09902	4.472136	0	
P9	8.602325	8.246211	6.403124	3.605551	4.123106	0

We again find the minimum Euclidean distance

Min Distance = 2.82847 which corresponds to cluster **(P5, P6),P7)** and P4

Update distance matrix for cluster ((P5, P6),P7)P4

	P1,P2,	P3	P5, P6, P7, P4	P8	P9
P1,P2	0				
P3	4.123106	0			
P5,P6,P7, P4	2.828427	3.605551	0		
P8	6.708204	5	4.472135955	0	
P9	8.602325	8.246211	3.605551275	4.123106	0

We again calculate the min Euclidean distance]

Min Distance = **2.82** corresponding to **((P5, P6),P7)P4)** and **(P1, P2)**

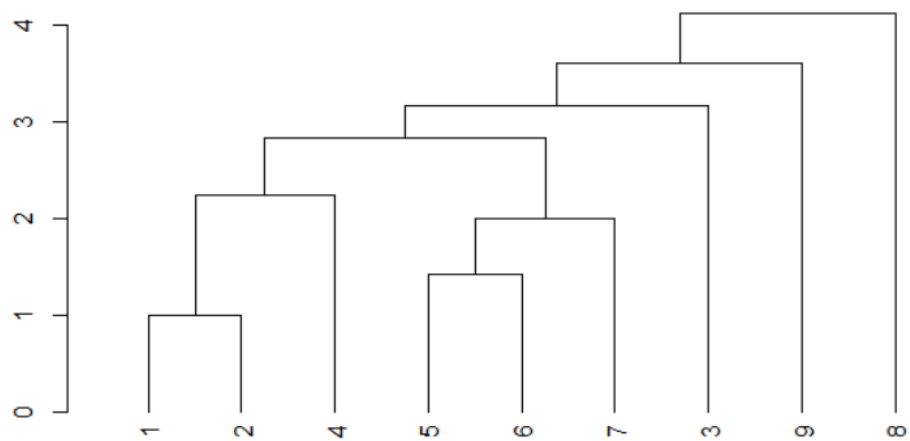
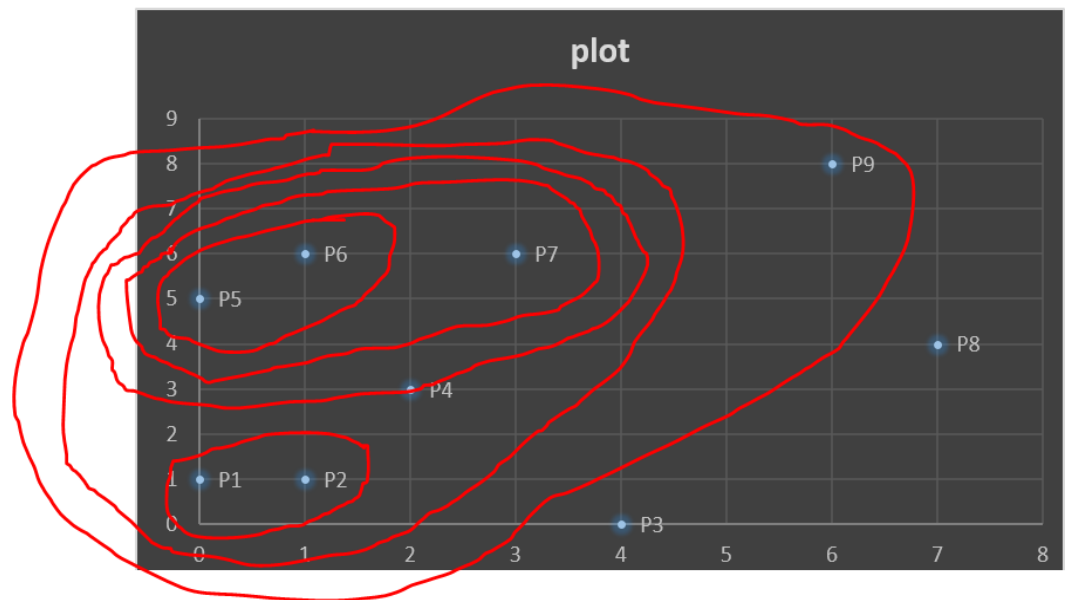
	P1,P2,(P5,P6,P7, P4)	P3	P8	P9
P1,P2,(P5,P6,P7, P4)	0			
P3	3.605551275	0		
P8	4.472135955	5	0	
P9	3.605551275	8.246211	4.123106	0

	P1,P2,(P5,P6,P7, P4)P3	P8
P1,P2,(P5,P6,P7, P4)P3	0	
P8	4.472135955	0
P9	3.605551275	4.123106

Minimum distance = 3.6

We update the eculidean distance again

	P1,P2,(P5,P6,P7, P4)P3)P9	P8, P9
P1,P2,(P5,P6,P7, P4)P3)P9	0	
P8,	4.472135955	0



1b Repeat the same as in (1a) but now calculate distances between clusters with complete linkage

- I create a euclidean distance matrix for all the values

	P1	P2	P3	P4	P5	P6	P7	P8	P9
P1	0								
P2	1	0							
P3	4.123106	3.162278	0						
P4	2.828427	2.236068	3.605551	0					
P5	5	4.123106	6.403124	2.828427	0				
P6	5.09902	5	6.708204	3.162278	1.414214	0			
P7	5.830952	5.385165	6.082763	3.162278	3.162278	2	0		
P8	7.615773	6.708204	5	5.09902	7.071068	6.324555	4.472136	0	
P9	9.219544	8.602325	8.246211	6.403124	6.708204	5.385165	3.605551	4.123106	0

Minimum distance = 1 and corresponds to [P1,P2]

Therefore first cluster contains [P1, P2]

We again update the distance Matrix $\text{MAX}[\text{dist}(P1, P2), P3]$
 $= \text{MAX}[\text{dist}(P1, P3), (P2, P3)]$

$= \text{MAX}[(4.123106, 3.162278)] = 4.123106$

We repeat the above procedure for all the features intersecting with P1 and P2

	P1,P2	P3	P4	P5	P6	P7	P8	P9
P1,P2	0							
P3	4.123106	0						
P4	2.828427	3.605551	0					
P5	5	6.403124	2.828427	0				
P6	5.09902	6.708204	3.162278	1.414214	0			
P7	5.830952	6.082763	3.162278	3.162278	2	0		
P8	7.615773	5	5.09902	7.071068	6.324555	4.472136	0	
P9	9.219544	8.246211	6.403124	6.708204	5.385165	3.605551	4.123106	

We again find the minimum distance from the updated Euclidean distance matrix

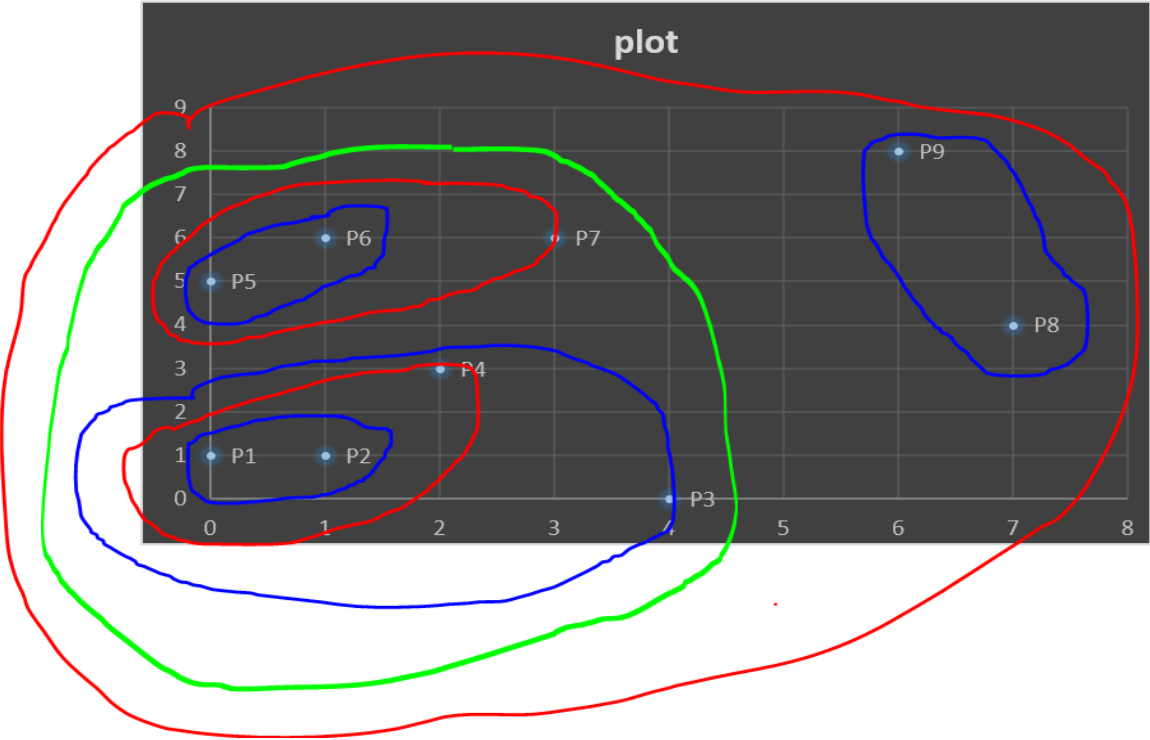
	P1,P2	P3	P4	P5, P6	P7	P8	P9
P1,P2	0						
P3	4.123106	0					
P4	2.828427	3.605551	0				
P5, P6	5.09902	6.708204	3.162278	0			
P7	5.830952	6.082763	3.162278	3.162278	0		
P8	7.615773	5	5.09902	7.071068	4.472136	0	
P9	9.219544	8.246211	6.403124	6.708204	3.605551	4.123106	

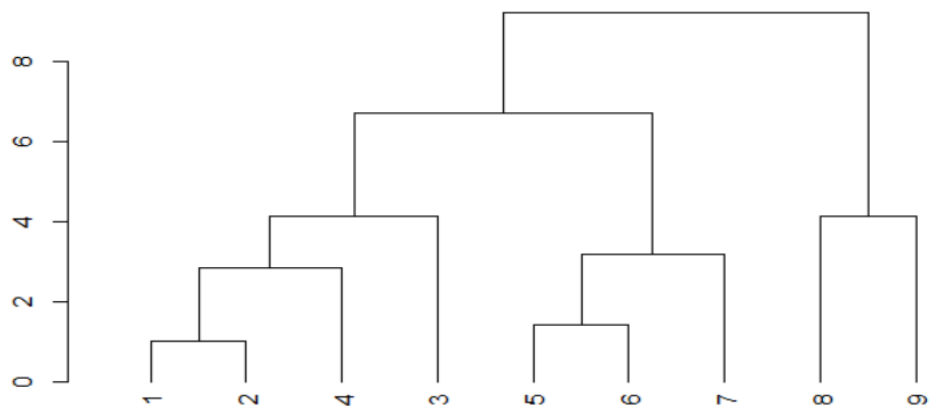
We again update and the Euclidean distance matrix and get the minimum distance

	P1,P2, P4	P3	P5, P6	P7	P8	P9
P1,P2,P4	0					
P3	4.123106	0				
P5, P6	5.09902	6.708204	0			
P7	5.830952	6.082763	3.162278	0		
P8	7.615773	5	7.071068	4.472136	0	
P9	9.219544	8.246211	6.708204	3.605551	4.123106	

	P1,P2, P4, P3	P5, P6	P7	P8	P9
P1,P2,P4, P3	0				
P5, P6	6.708203932	0			
P7	6.08276253	3.162278	0		
P8	7.615773106	7.071068	4.472136	0	
P9	9.219544457	6.708204	3.605551	4.123106	

	P1,P2, P4, P3	P5, P6	P8,P9	P7
P1,P2,P4, P3	0			
P5, P6	6.708203932	0		
P8, P9	9.219544457	6.708204	0	
P7	7.615773106	7.071068	4.472136	0





Exercise 2

ID	X	Y
P1	0	1
P2	1	1
P3	4	0
P4	2	3
P5	0	5
P6	1	6
P7	3	6
P8	7	4
P9	6	8

	EUCLIDEAN DISTANCE				
Dataset	Cluster P1	Cluster P2	Cluster P3	Cluster P4	ASSINGMENT
P5(0, 5)	2	2.3	6.1	2.8	P1

ID	P1
CLUSTERS ASSIGNED	P1, P5

Update the cluster centroid for cluster P1		
cluster	X	Y
Cluster P1	$(0+0)/2 = 0$	$(1+5)/2 = 3$

	EUCLIDEAN DISTANCE				
Dataset	Cluster P1	Cluster P2	Cluster P3	Cluster P4	ASSINGMENT
P6(1,6)	2	5	6.7	3.3	P1

--	--	--	--	--	--

ID	P1
CLUSTERS ASSIGNED	P1, P5, P6

	Update the cluster centroid for cluster P1			
	cluster	X	Y	
	Cluster P1	$(0+0+1)/3 = 0.3$	$(1+5+6)/3 = 4$	

	EUCLIDEAN DISTANCE				
Dataset	Cluster P1	Cluster P2	Cluster P3	Cluster P4	ASSINGMENT
P7(3, 6)	3.36	5.38	6.01	3.3	P4

ID	P4
CLUSTERS ASSIGNED	P4, P7

Update the cluster centroid for cluster P4		
Cluster ID	X	Y
P4	$(2+3)/2 = 2.5$	$(3+6)/2 = 4.5$

	EUCLIDEAN DISTANCE				
Dataset	Cluster P1	Cluster P2	Cluster P3	Cluster P4	ASSINGMENT
P8(7, 4)	6.67	6.7	5	4.5	P4

ID	P4
CLUSTERS ASSIGNED	P4, P7, P8

Update the cluster centroid for cluster P4		
cluster	X	Y
Cluster P4	$(2+3+7)/3 = 3.33$	$(3+6+4)/3 = 4.33$

	EUCLIDEAN DISTANCE				
--	--------------------	--	--	--	--

Dataset	Cluster P1	Cluster P2	Cluster P3	Cluster P4	ASSINGMENT
P9(6, 8)	6.64	8.63	8.24	4.53	P4

ID	CLUSTERS ASSIGNED
P1	P1, P5, P6
P2	P2
P3	P3
P4	P4, P7, P8, P9

cluster centroid for all the clusters				
		MEAN		
	ID	X	Y	
cluster P1	P1, P5, P6	$(0+0+1)/3 = 0.3$	$(1+5+6)/3 = 4$	
Cluster P2	P2	1	1	
Cluster P3	P3	4	0	
Cluster P4	[P4, P7, P8, P9]	$(2+3+7+6)/4 = 4.5$	$3+6+4+8)/4 = 5.25$	

			EUCLIDANT DISTANCE				
ID	X	Y	CLUSTER P1	CLUSTER P2	CLUSTER P3	CLUSTER P4	ASSIGNMENT
P1	0	1	3.0149627	1	4.123105626	6.189709202	P2
P2	1	1	3.0805844	0	3.16227766	5.505678886	P2
P3	4	0	5.4488531	3.16227766	0	5.273755777	P3
P4	2	3	1.9723083	2.23606798	3.605551275	3.363406012	P1
P5	0	5	1.0440307	4.12310563	6.403124237	4.506939094	P1
P6	1	6	2.118962	5	6.708203932	3.579455266	P1
P7	3	6	3.3600595	5.38516481	6.08276253	1.677050983	P4
P8	7	4	6.7	6.70820393	5	2.795084972	P4
P9	6	8	6.9634761	8.60232527	8.246211251	3.132491022	P4

ID	OLD CLUSTERS	NEW CLUSTERS ASSIGNED
P1	P1, P5, P6	P4, P5, P6

P2	P2	P1, P2
P3	P3	P3
P4	P4, P7, P8, P9	P7, P8, P9

I again calculated the centroid for each clusters as shown below

cluster centroid for all the clusters			
		MEAN	
	NEW CLUSTERS ASSIGNED	X	Y
cluster P1	P4, P5, P6	1	4.666666667
Cluster P2	P1, P2	0.5	1
Cluster P3	P3	4	0
Cluster P4	P7, P8, P9	5.333333333	6

EUCLIDANT DISTANCE							
	X	Y	CLUSTER P1	CLUSTER P2	CLUSTER P3	CLUSTER P4	ASSIGNMENT
P1	0	1	3.8009063	0.5	4.123105626	7.308139298	P2
P2	1	1	3.667	0.5	3.16227766	6.614295125	P2
P3	4	0	5.5480527	3.64005494	0	6.145640731	P3
P4	2	3	1.9439365	2.5	3.605551275	4.482064257	P1
P5	0	5	1.0539872	4.03112887	6.403124237	5.422997326	P1
P6	1	6	1.333	5.02493781	6.708203932	4.33	P1
P7	3	6	2.403516	5.59016994	6.08276253	2.33	P4
P8	7	4	6.0369602	7.15891053	5	2.605551765	P4
P9	6	8	6.0090672	8.90224691	8.246211251	2.10924157	P4

ID	OLD CLUSTERS	NEW CLUSTERS ASSIGNED
P1	P4, P5, P6	P4, P5, P6
P2	P2, P1	P1, P2
P3	P3	P3
P4	P7, P8, P9	P7, P8, P9