P2 Part A
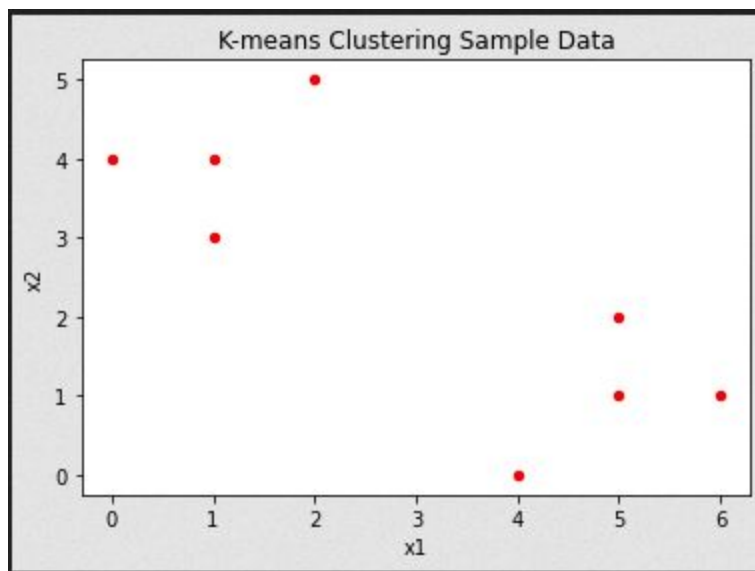
1. Team Jarlsberg
   Erin Dolson
   Siddhesh Mahadeshwar
   Zachary Noel

2.

a.

b. $x1_{group\ 1}$ = (1 + 1 + 4 + 5)/4 = 2.75
   $x2_{group\ 1}$ = (3 + 4 + 0 + 1)/4 = 2
   $x1_{group\ 2}$ = (0 + 2 + 5 + 6)/4 = 3.25
   $x2_{group\ 2}$ = (4 + 5 + 2 + 1)/4 = 3

Centers	X1	X2
Group 1	2.75	2
Group 2	3.25	3

c. For initial group 1, (2.75, 2) is the initial centroid. (c_1)
   For initial group 2, (3.25, 3) is the second initial centroid. (c_2)

   Now, we have to utilize the euclidean distance formula.
   We need to find the distance from c_1 to all of our other samples,
   and then find the distance from c_2 to all of our other samples.

   Distance from c_1 to Sample 1:

$$\sqrt{(2.75-0)^2 + (2-4)^2} = \sqrt{2.75^2 + 2^2} = \sqrt{7.5625 + 2} = \sqrt{11.5625} = 3.4004$$

The same formula is utilized for the below distances.

Distance from c_1 to Sample 2
2.016
Distance from c_1 to Sample 3
2.658
Distance from c_1 to Sample 4
3.092
Distance from c_1 to Sample 5
2.358
Distance from c_1 to Sample 6
2.462
Distance from c_1 to Sample 7:
2.250
Distance from c_1 to Sample 8:
3.400

Distance from c_2 to Sample 1
3.400
Distance from c_2 to Sample 2
2.250
Distance from c_2 to Sample 3
2.462
Distance from c_2 to Sample 4
2.358
Distance from c_2 to Sample 5
3.092
Distance from c_2 to Sample 6
6.658
Distance from c_2 to Sample 7
2.016
Distance from c_2 to Sample 8
3.400

Sample	d2c1 (distance from sample i to center of cluster 1)	d2c2 (distance from sample i to center of cluster 2)	Group
1	3.400	3.400	1
2	2.016	2.250	1

3	2.658	2.462	2
4	3.092	2.358	2
5	2.358	3.092	1
6	2.462	2.658	1
7	2.250	2.016	2
8	3.400	3.400	1

d.

$x1_{group\ 1} = (3.400 + 2.016 + 2.358 + 2.462 + 3.400)/5 = 2.727$
$x2_{group\ 1} = (3.400 + 2.250 + 3.092 + 2.658 + 3.400)/5 = 2.960$
$x1_{group\ 2} = (2.658 + 3.092 + 2.250)/3 = 2.667$
$x2_{group\ 2} = (2.462 + 2.358 + 2.016)/3 = 2.278$

Centers	X1	X2
Group 1	2.727	2.960
Group 2	2.667	2.278

Sample	d2c1 (distance from sample i to center of cluster 1)	d2c2 (distance from sample i to center of cluster 2)	Group
1	2.919	3.175	1
2	1.727	1.817	1
3	2.016	2.397	1
4	2.166	2.803	1
5	3.222	2.639	2
6	3.001	2.660	1
7	2.467	2.35	2
8	3.815	3.57	2

$x1_{group\ 1}$ = (2.919 + 1.727 + 2.016 +2.166 + 3.001)/5 = 2.366
$x2_{group\ 1}$ = (3.175 + 1.817 + 2.397 + 2.803 + 2.660)/5 = 2.570
$x1_{group\ 2}$ = (3.222 + 2.467 + 3.815)/3 = 3.168
$x2_{group\ 2}$ = (2.639 + 2.35 + 3.57)/3 = 2.853

Centers	X1	X2
Group 1	2.366	2.570
Group 2	3.168	2.853

Sample	d2c1 (distance from sample *i* to center of cluster 1)	d2c2 (distance from sample *i* to center of cluster 2)	Group
1	2.765	3.369	1
2	1.432	2.173	1
3	1.978	2.453	1
4	2.457	2.444	1
5	3.045	2.972	2
6	3.066	2.606	2
7	2.695	2.021	2
8	3.959	3.384	2

$x1_{group\ 1}$ = (2.765 + 1.432 + 1.978 +2.457)/4 = 2.158
$x2_{group\ 1}$ = (3.369 + 2.173 + 2.453 + 2.444)/4 = 2.610
$x1_{group\ 2}$ = (3.045 + 3.066 + 2.695 + 3.959)/4 = 3.191
$x2_{group\ 2}$ = (2.972 + 2.606 + 2.021 + 3.384)/4 = 2.746

Centers	X1	X2
Group 1	2.158	2.610
Group 2	3.191	2.746

Sample	d2c1 (distance from sample *i* to center of cluster 1)	d2c2 (distance from sample *i* to center of cluster 2)	Group
1	2.567	3.429	1
2	1.222	2.206	1
3	1.809	2.524	1
4	2.395	2.549	1
5	3.195	2.863	2
6	3.266	2.514	2
7	2.907	1.957	2
8	4.166	3.307	2

Since the cluster groupings remained the same, the centroid values will remain the same throughout when performing further calculations.
Therefore, no further calculations are necessary.

KMeans Clustering of Sample Data
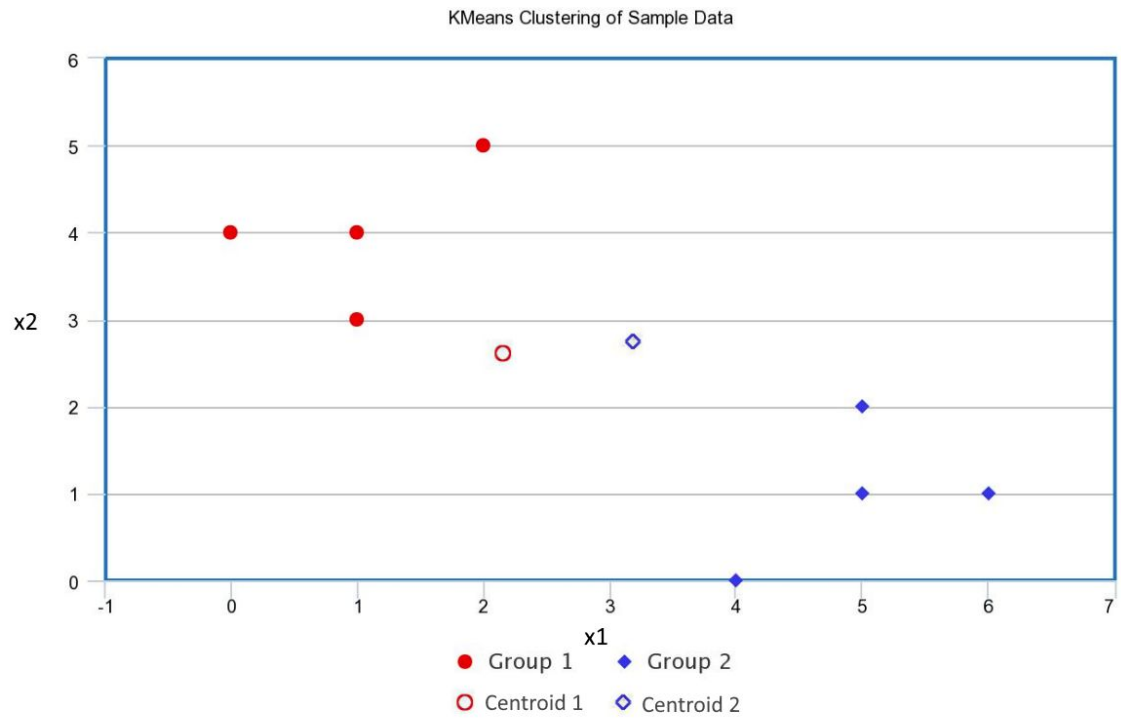
e)

3.

$$D = \begin{bmatrix} -- & 0.3 & 0.4 & 0.7 & 0.6 \\ 0.3 & -- & 0.5 & 0.8 & 0.2 \\ 0.4 & 0.5 & -- & 0.45 & 0.4 \\ 0.7 & 0.8 & 0.45 & -- & 0.35 \\ 0.6 & 0.2 & 0.4 & 0.35 & -- \end{bmatrix}$$

Step 1 - Combine Samples 2 and 5

	1	(2,5)	3	4
1	0	-	-	-
(2,5)	0.6	0	-	-
3	0.4	0.5	0	-
4	0.7	0.8	.45	0

Step 2 - Combine Samples 1 and 3

	(1,3)	(2,5)	4

	0	-	-
(1,3)	0	-	-
(2,5)	0.6	0	-
4	0.7	0.8	0

Step 3 - Combine (1,3) and (2,5)

	((1,3), (2,5))	**4**
((1,3), (2,5))	0	-
4	0.8	0
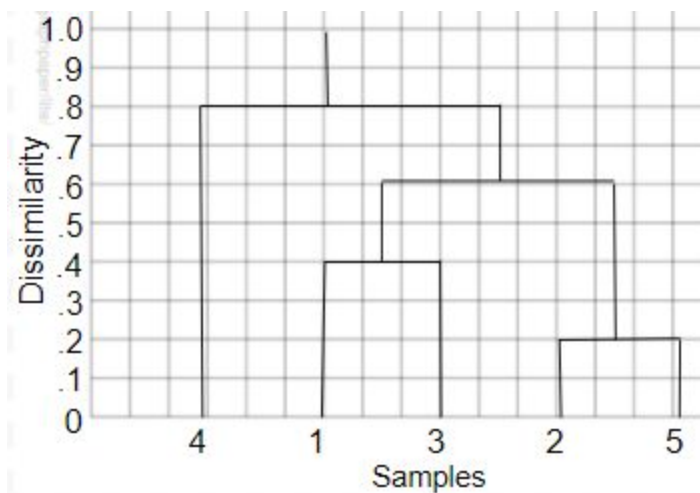
b)

Step 1 - Combine Samples 2 and 5

	1	**(2,5)**	**3**	**4**
1	0	-	-	-
(2,5)	0.3	0	-	-
3	0.4	0.4	0	-
4	0.7	0.35	0.45	0

Step 2 - Combine Samples (2,5) and 1

	(1, (2,5))	**3**	**4**
(1, (2,5))	0	-	-
3	0.4	0	-

4	0.35	0.45	0

Step 3 - Combine Samples ((2,5),1) and 4

	(((1, (2,5)), 4)	3
(((1,(2,5)), 4)	0	-
3	0.4	0