

Q 1)

	a	b	c	d	e	f	g	h
A	5	4		5	2		3	2
B		3	4	4	2	2	1	
C	3		1	4		4	5	3

	a	b	c	d	e	f	g	h
A	1	1	0	1	0	0	1	0
B	0	1	1	1	0	0	0	0
C	1	0	0	1	0	1	1	1

a)

$$(A, B) = 3/5$$

$$(A, C) = 1/2$$

$$(B, C) = 6/7$$

b)

$$(A, B)$$

$$\frac{4 * 3 + 5 * 4 + 2 * 2 + 3 * 1}{\sqrt{5^2 + 4^2 + 5^2 + 2^2 + 3^2 + 2^2} \sqrt{3^2 + 4^2 + 4^2 + 2^2 + 2^2 + 1^2}}$$

$$= 39/\text{root}(83)*\text{root}(50) = 0.605$$

$$(A, C)$$

$$\frac{5 * 3 + 5 * 4 + 3 * 5 + 2 * 3}{\sqrt{5^2 + 4^2 + 5^2 + 2^2 + 3^2 + 2^2} \sqrt{3^2 + 1^2 + 4^2 + 4^2 + 5^2 + 3^2}}$$

$$= 56/\text{root}(83)*\text{root}(76) = 0.705$$

$$(B, C)$$

$$\frac{4 * 1 + 4 * 4 + 2 * 4 + 1 * 5}{\sqrt{3^2 + 4^2 + 4^2 + 2^2 + 2^2 + 1^2} \sqrt{3^2 + 1^2 + 4^2 + 4^2 + 5^2 + 3^2}}$$

$$= 33 / \sqrt{50} \cdot \sqrt{76} = 0.535$$

c)

$$A \cap B = 2$$

$$A \cup B = 5$$

$$(A, B) = 1 - 2/5 = \underline{3/5}$$

$$A \cap C = 3$$

$$A \cup C = 6$$

$$(A, C) = 1 - 3/6 = \underline{1/2}$$

$$B \cap C = 1$$

$$B \cup C = 7$$

$$(B, C) = 1 - 1/7 = \underline{6/7}$$

d)

$$(A, B)$$

$$\frac{2}{\sqrt{1^2 + 1^2 + 1^2 + 1^2} \sqrt{1^2 + 1^2 + 1^2}}$$

$$= (2 / \sqrt{4}) \cdot \sqrt{3} = \underline{0.577}$$

$$(A, C)$$

$$\frac{3}{\sqrt{1^2 + 1^2 + 1^2 + 1^2} \sqrt{1^2 + 1^2 + 1^2 + 1^2 + 1^2}}$$

$$= (3 / \sqrt{4}) \cdot \sqrt{5} = \underline{0.671}$$

$$(B, C)$$

$$\frac{1}{\sqrt{1^2 + 1^2 + 1^2} \sqrt{1^2 + 1^2 + 1^2 + 1^2 + 1^2}}$$

$$= (1/\sqrt{3})\sqrt{5} = \underline{0.258}$$

e)

Avg of A = 3.5 / Avg of B = 2.67 / Avg of C = 3.33

	a	b	c	d	e	f	g	h
A	1.5	0.5		1.5	-1.5		-0.5	-1.5
B		0.33	1.33	1.33	-0.67	-0.67	-1.67	
C	-0.33		-2.33	0.67		0.67	1.67	-0.33

f)

(A, B)

$$\frac{0.5 * 0.33 + 1.5 * 1.33 + -1.5 * -1.33 + -0.5 * -1.67}{\sqrt{1.5^2 + 0.5^2 + 1.5^2 + (-1.5)^2 + (-0.5)^2 + (-1.5)^2} \sqrt{0.33^2 + 1.33^2 + 1.33^2 + (-0.67)^2 + (-0.67)^2 + (-1.67)^2}}$$

$$= 4.99/\sqrt{9.5} \sqrt{7.3334} = 0.5978$$

(A, C)

$$\frac{1.5 * -0.33 + 1.5 * 0.67 + -0.5 * 1.67 + -1.5 * -0.33}{\sqrt{1.5^2 + 0.5^2 + 1.5^2 + (-1.5)^2 + (-0.5)^2 + (-1.5)^2} \sqrt{(-0.33)^2 + (-2.33)^2 + 0.67^2 + 0.67^2 + (1.67)^2 + (-0.33)^2}}$$

$$= 0.17/\sqrt{9.5} \sqrt{9.3334} = 0.0181$$

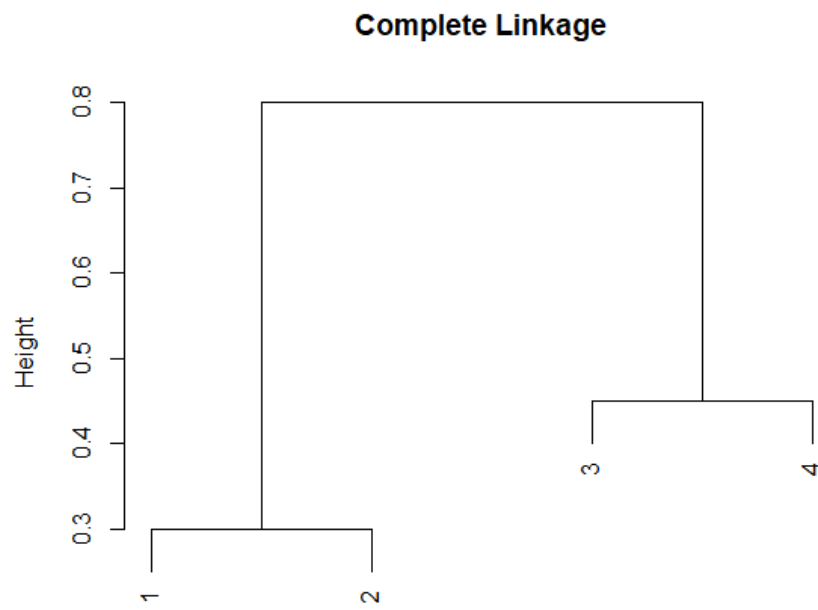
(B, C)

$$\frac{1.33 * -2.33 + 1.33 * 0.67 + -0.67 * 0.67 + -1.67 * 1.67}{\sqrt{0.33^2 + 1.33^2 + 1.33^2 + (-0.67)^2 + (-0.67)^2 + (-1.67)^2} \sqrt{(-0.33)^2 + (-2.33)^2 + 0.67^2 + 0.67^2 + (1.67)^2 + (-0.33)^2}}$$

$$= -5.4456/\sqrt{7.3334} \sqrt{9.3334} = -0.6582$$

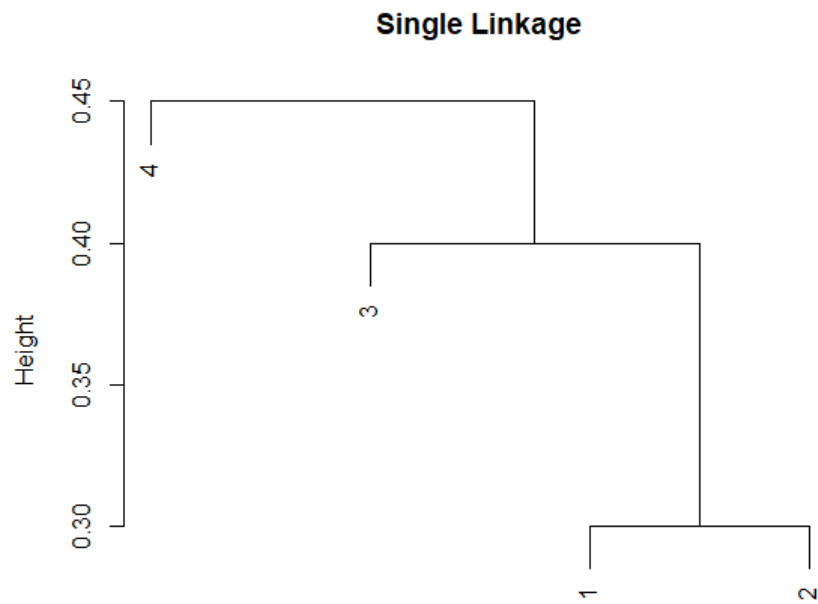
Q 2)

a)



m  
hclust (\*, "complete")

b)



m  
hclust (\*, "single")

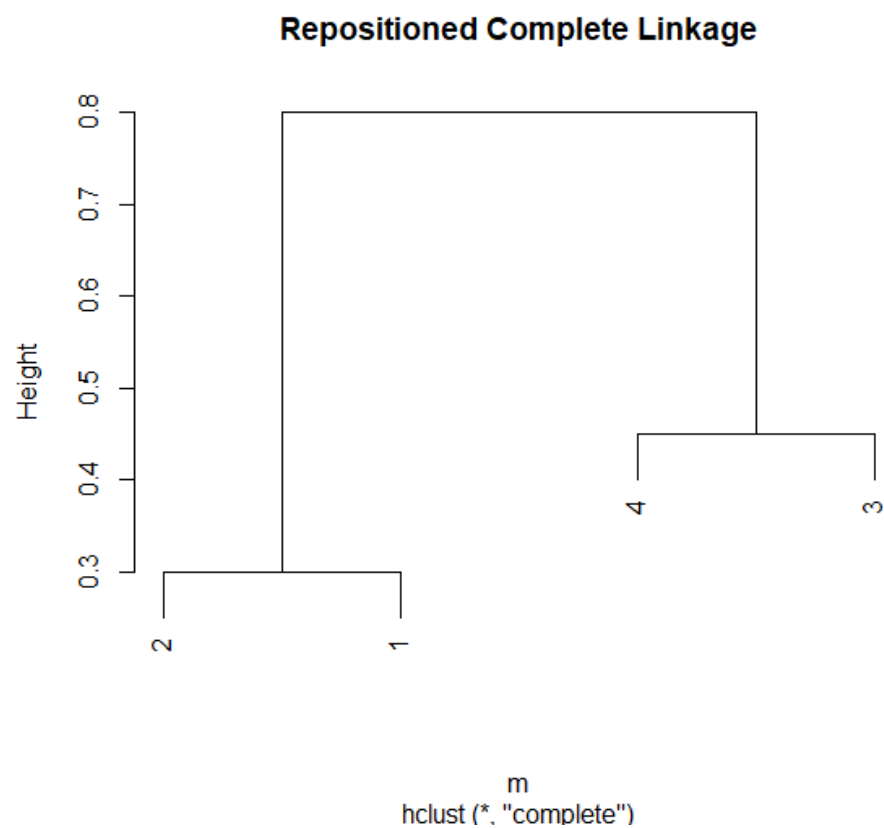
c)

For the first cluster dendrogram at the height of 0.8 when  $i=4$ , I obtained complete linkage of (1,2) and (3,4).

d)

For the second cluster at the height of 0.45 when  $i=4$ , I obtained single linkage of (4), (3,(1,2))

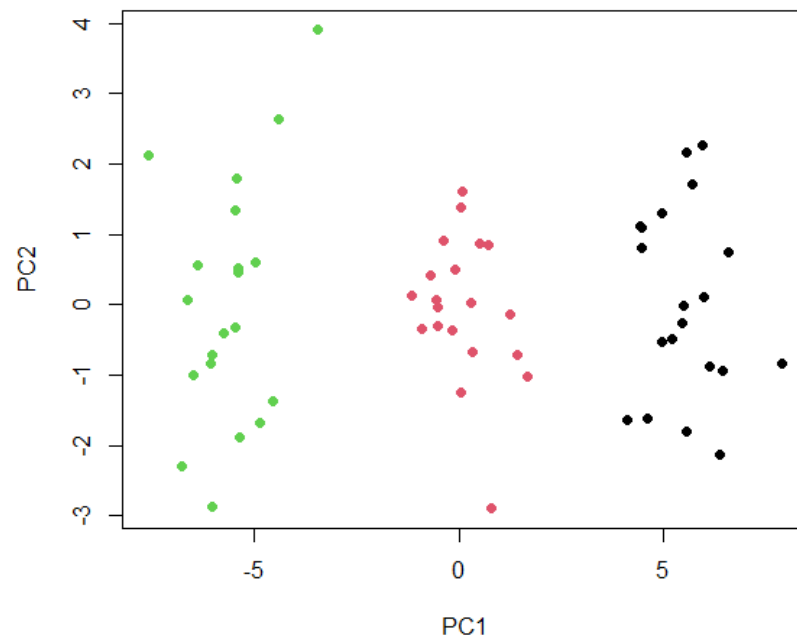
e)



Keeping the dendrogram same, I flipped the numbers in the complete linkage.

### Question 3

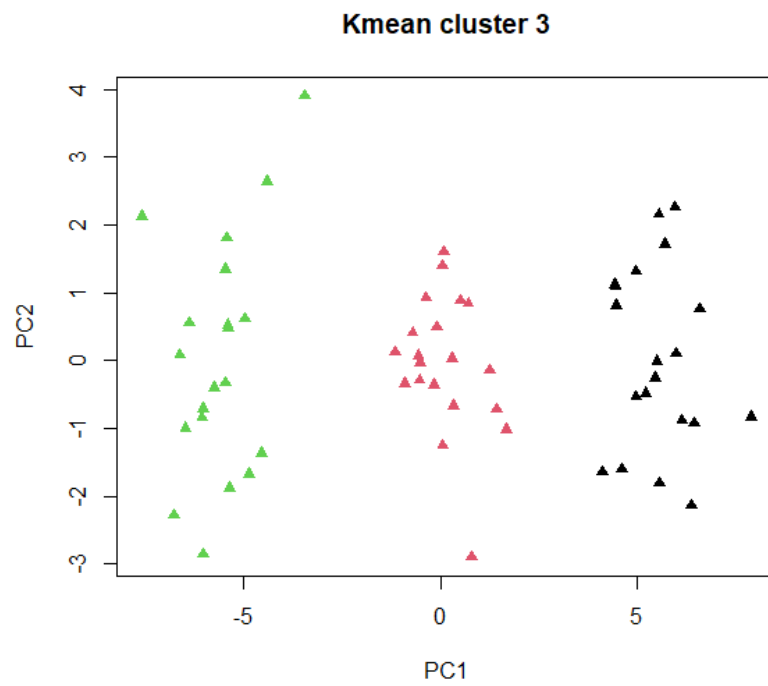
a)



b)

```
> table(datak$cluster, a)
a
  1  2  3
1 20  0  0
2  0 20  0
3  0  0 20
```

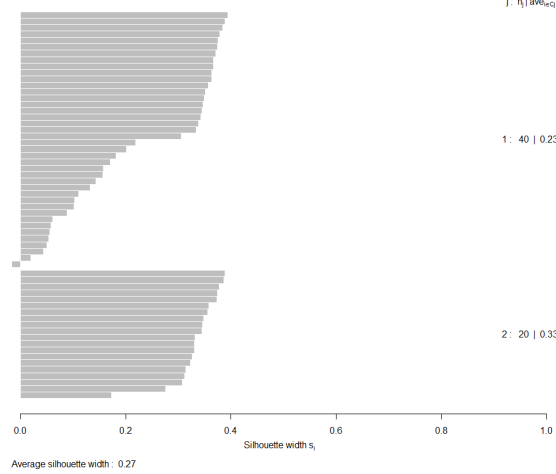
You can see that the data is evenly clustered into 20.



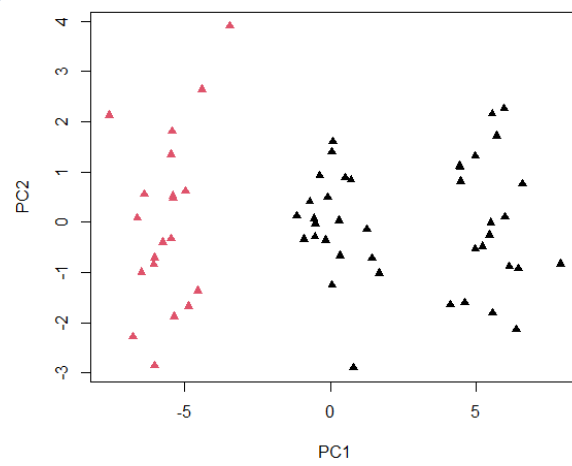
c)

K = 2

Silhouette plot of (x = dataK2\$cluster, dist = dist(data), col = a)  
n = 60

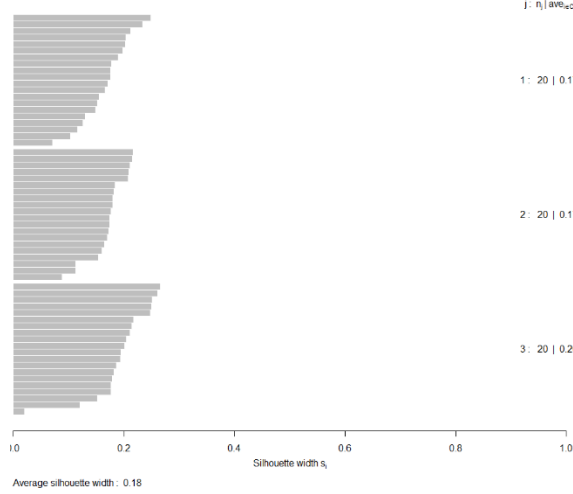


### Kmean cluster 2



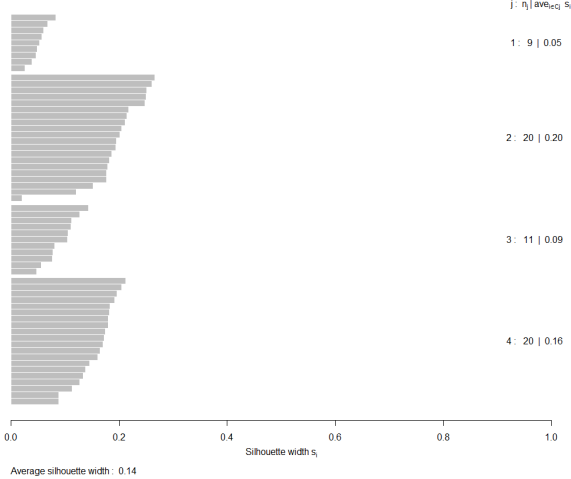
K = 3

Silhouette plot of (x = datak\$cluster, dist = dist(data), col = a)  
n = 60

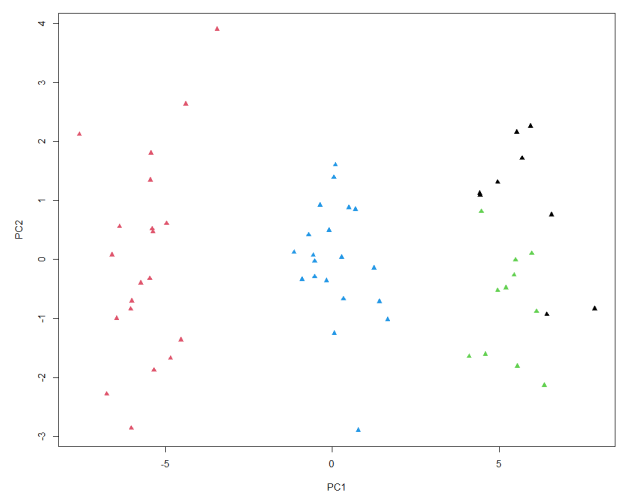


K = 4

Silhouette plot of (x = datak4\$cluster, dist = dist(data), col = a)  
n = 60

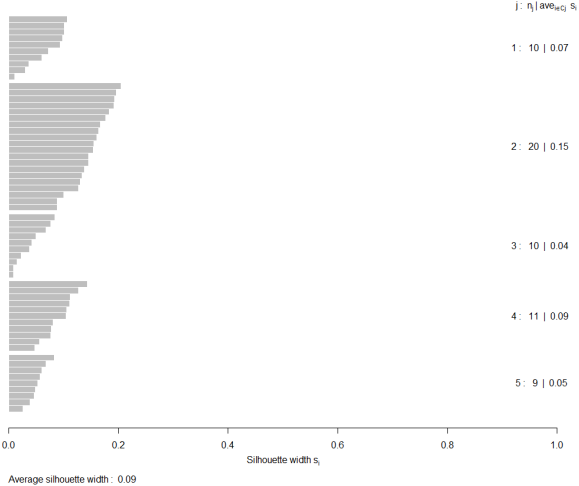


Kmean cluster 4

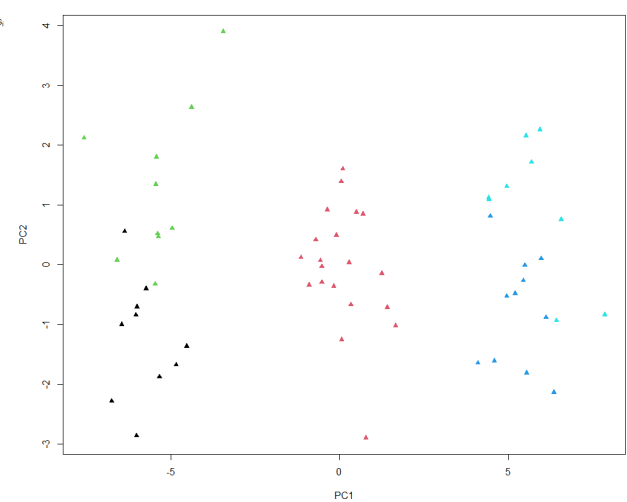


K = 5

Silhouette plot of (x = datak5\$cluster, dist = dist(data), col = a)  
n = 60



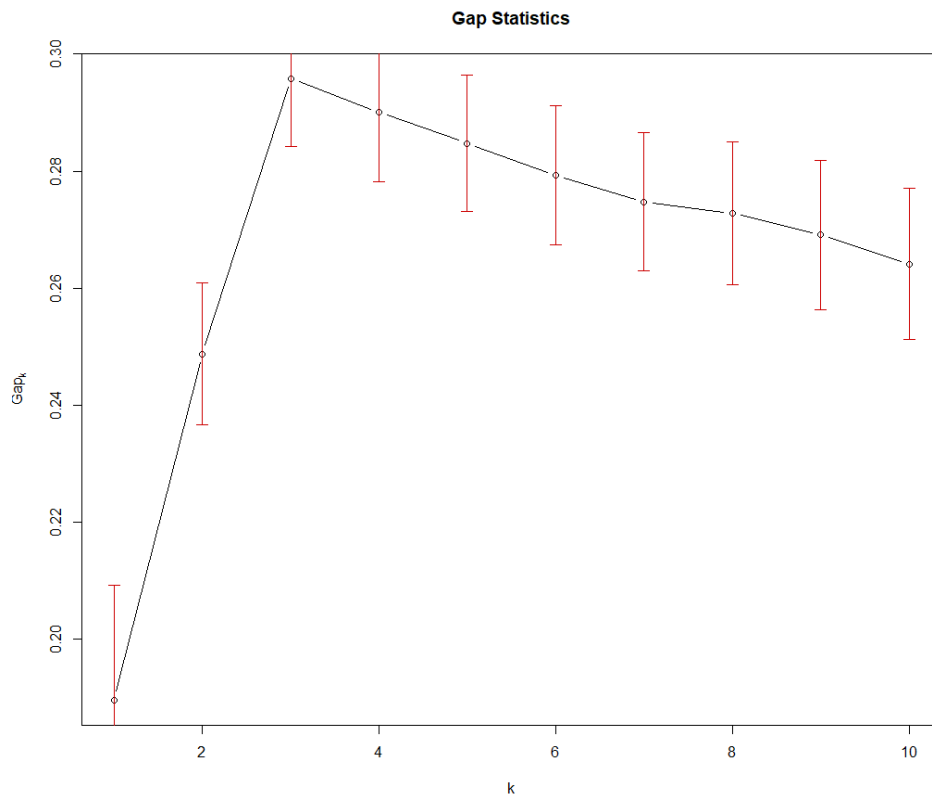
Kmean cluster 5





The graphs shown are performed in order to compare and find the optimal K value for clustering. First of all, K = 2 has one that is pointing to minus, K = 4, 5 have components that are clustered below average. Therefore, K = 3 is the optimal K for silhouette plots.

d)



For gap statistics, K is maximized when K = 3 and immediately decreases after. Therefore, it is the optimal K