

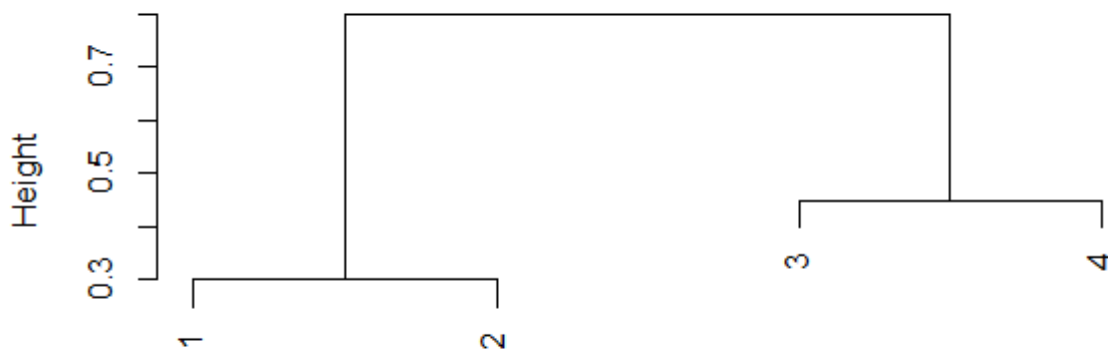
HW5 --Jiayuan Guo -- R code

1. Chapter 10, Problem 2

(a)

```
1 #Problem 1(a)
2 d = as.dist(matrix(c(0, 0.3, 0.4, 0.7,
3                     0.3, 0, 0.5, 0.8,
4                     0.4, 0.5, 0.0, 0.45,
5                     0.7, 0.8, 0.45, 0.0), nrow = 4))
6 plot(hclust(d, method = "complete"))
```

Cluster Dendrogram

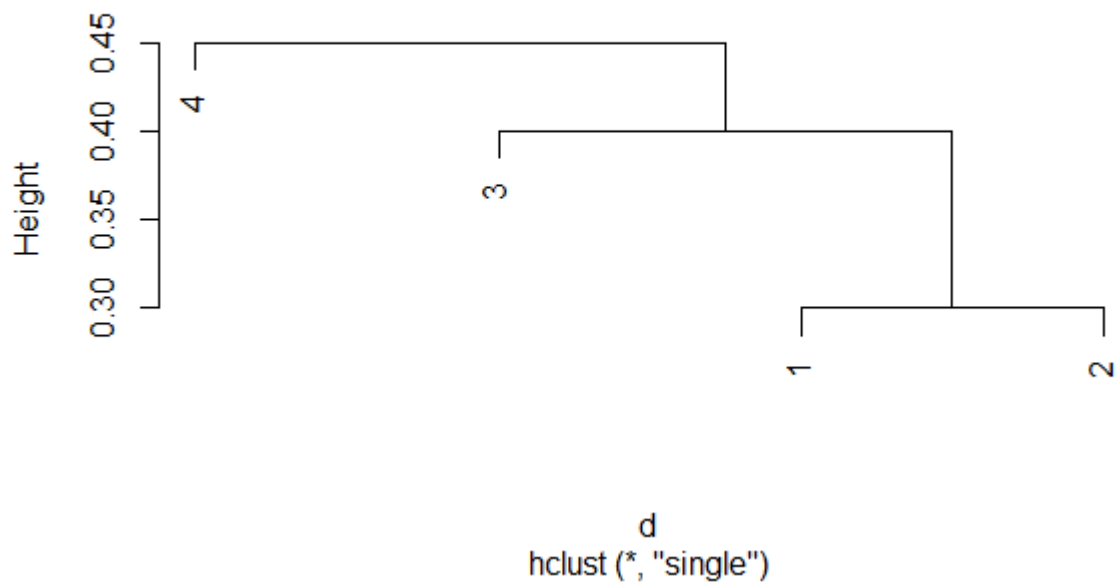


d
hclust (*, "complete")

(b)

```
1 #1(b)
2 plot(hclust(d, method = "single"))
```

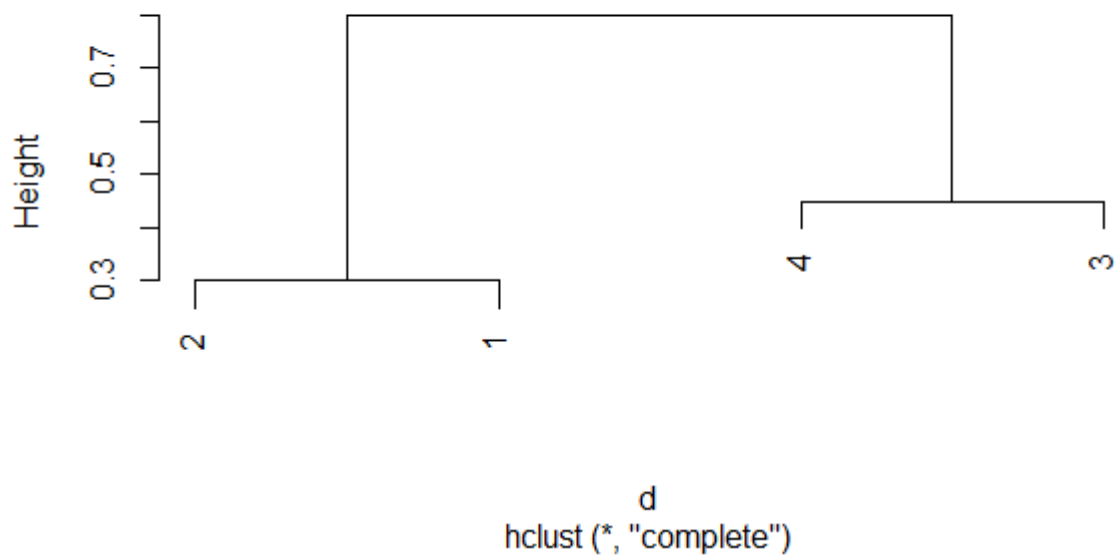
Cluster Dendrogram



(e)

```
1 #1(e)
2 plot(hclust(d, method = "complete"), labels = c(2,1,4,3))
```

Cluster Dendrogram



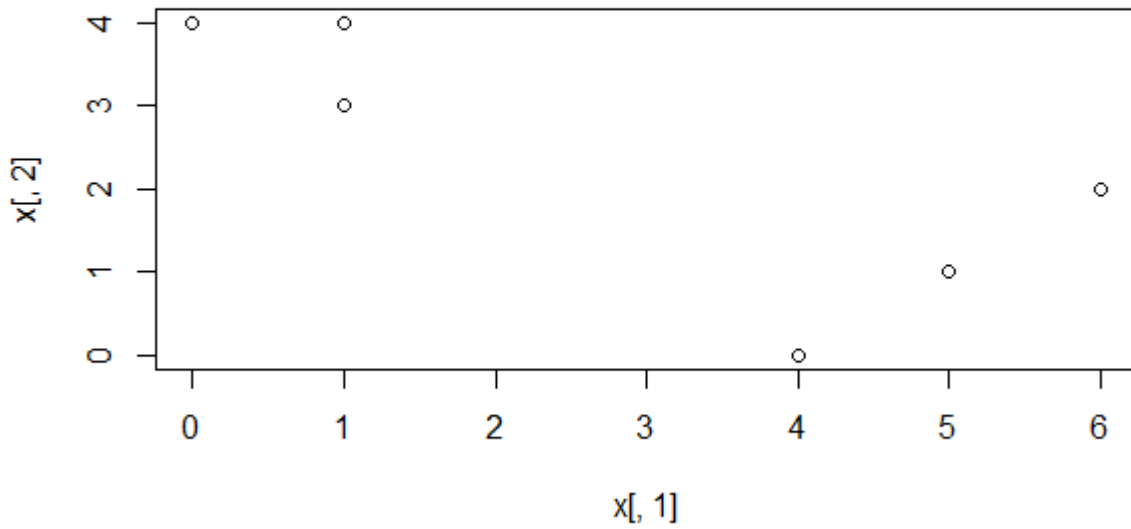
2. Chapter 10, Problem 3

(a)

```

1 #problem2(a)
2 set.seed(1)
3 x = cbind(c(1, 1, 0, 5, 6, 4), c(4, 3, 4, 1, 2, 0))
4 plot(x[,1], x[,2])

```



(b)

```

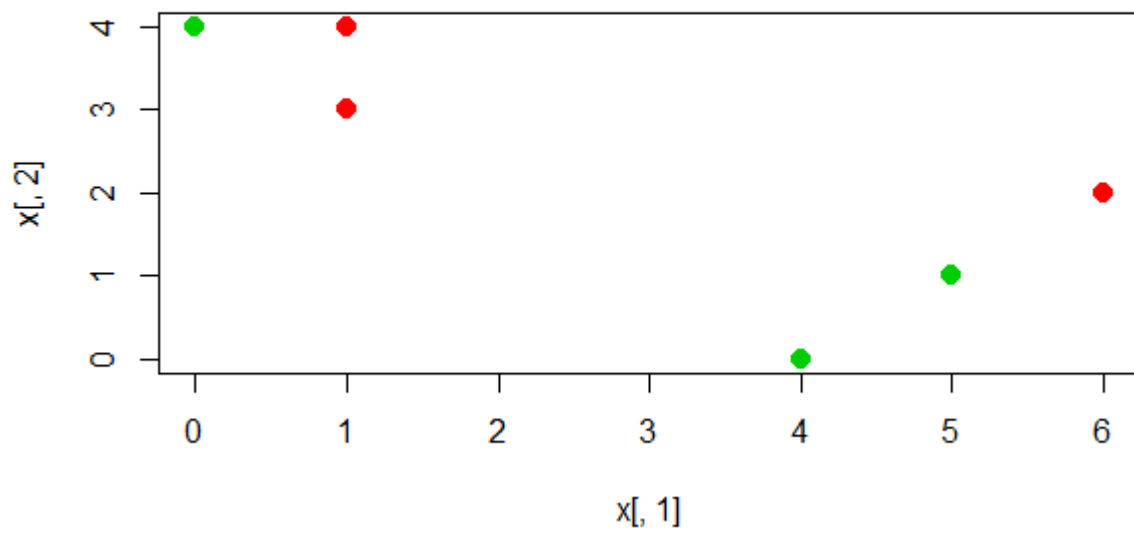
1 #2(b)
2 labels = sample(2, nrow(x), replace = T)
3 labels
4 plot(x[, 1], x[, 2], col = (labels + 1), pch = 20, cex = 2)

```

```

1 Output:
2 > labels
3 [1] 1 1 2 2 1 2

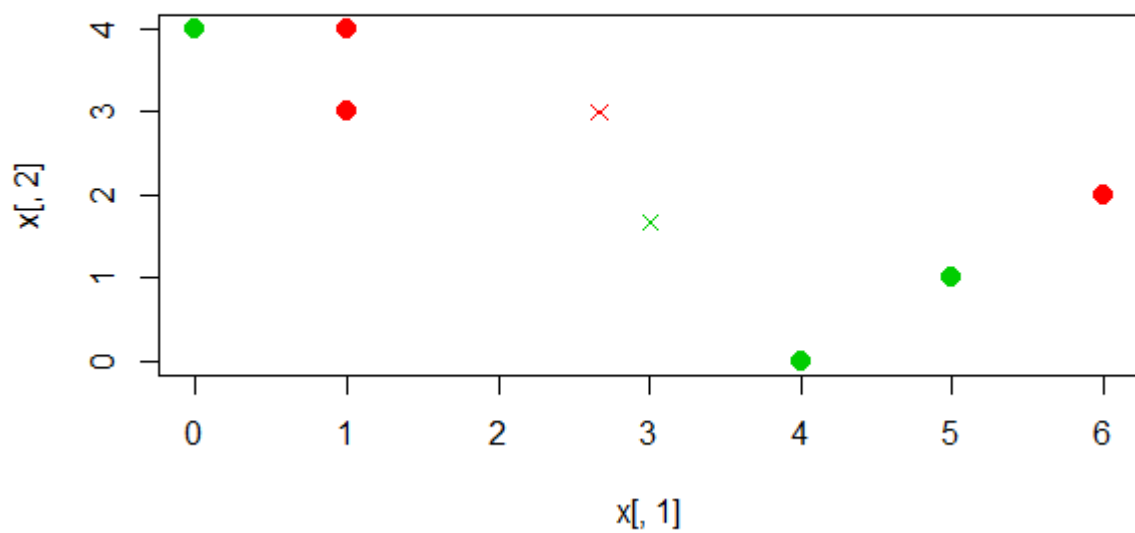
```



(c)

```
1 #2(c)
2 centroid1 = c(mean(x[labels == 1, 1]), mean(x[labels == 1, 2]))
3 centroid2 = c(mean(x[labels == 2, 1]), mean(x[labels == 2, 2]))
4 plot(x[,1], x[,2], col=(labels + 1), pch = 20, cex = 2)
5 points(centroid1[1], centroid1[2], col = 2, pch = 4)
6 points(centroid2[1], centroid2[2], col = 3, pch = 4)
```

```
1 Output:
2 > centroid1
3 [1] 2.666667 3.000000
4 > centroid2
5 [1] 3.000000 1.666667
```

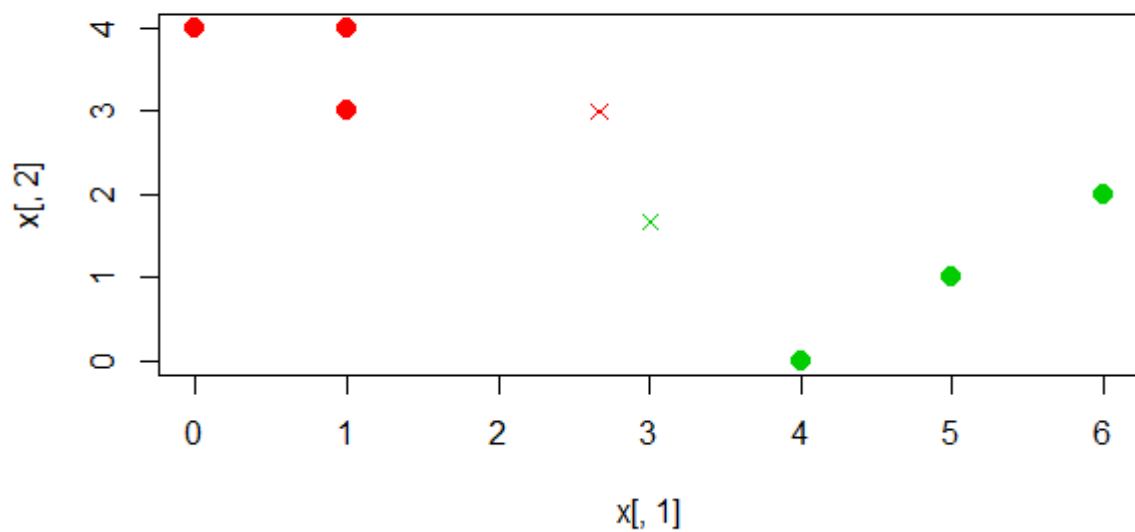


(d)

```

1 #2(d)
2 labels = c(1, 1, 1, 2, 2, 2)
3 plot(x[, 1], x[, 2], col = (labels + 1), pch = 20, cex = 2)
4 points(centroid1[1], centroid1[2], col = 2, pch = 4)
5 points(centroid2[1], centroid2[2], col = 3, pch = 4)

```



(e)

```

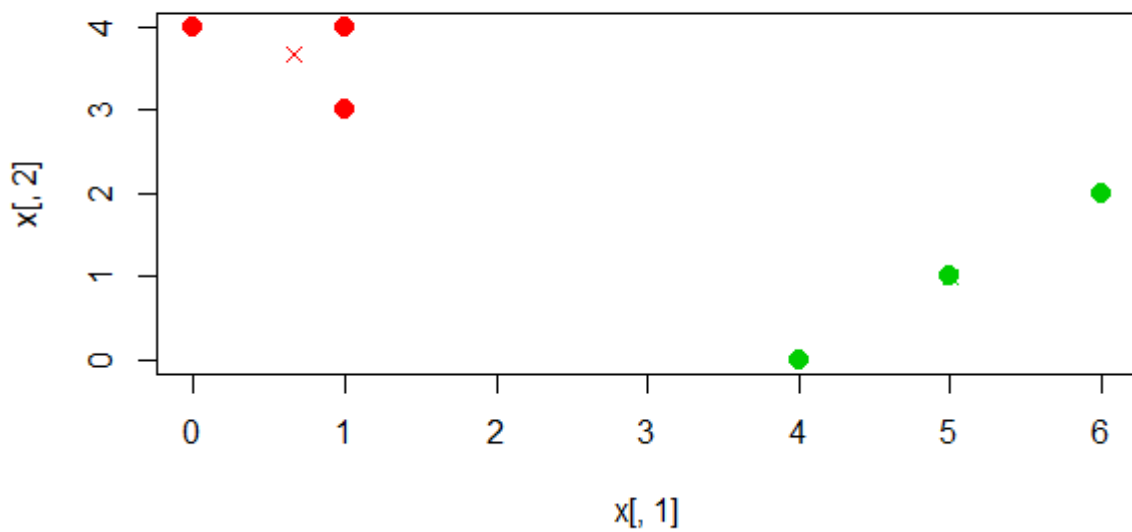
1 #2(e)
2 centroid1 <- c(mean(x[labels == 1, 1]), mean(x[labels == 1, 2]))
3 centroid2 <- c(mean(x[labels == 2, 1]), mean(x[labels == 2, 2]))
4 plot(x[,1], x[,2], col=(labels + 1), pch = 20, cex = 2)
5 points(centroid1[1], centroid1[2], col = 2, pch = 4)
6 points(centroid2[1], centroid2[2], col = 3, pch = 4)

```

```

1 Output:
2 > centroid1
3 [1] 0.6666667 3.6666667
4 > centroid2
5 [1] 5 1

```

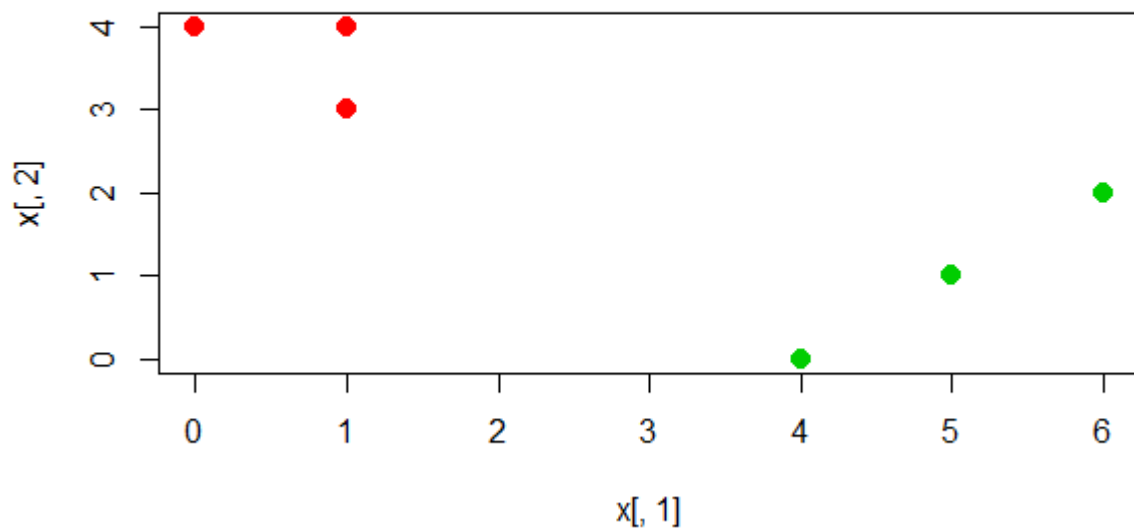


(f)

```

1 #2(f)
2 plot(x[, 1], x[, 2], col=(labels + 1), pch = 20, cex = 2)

```



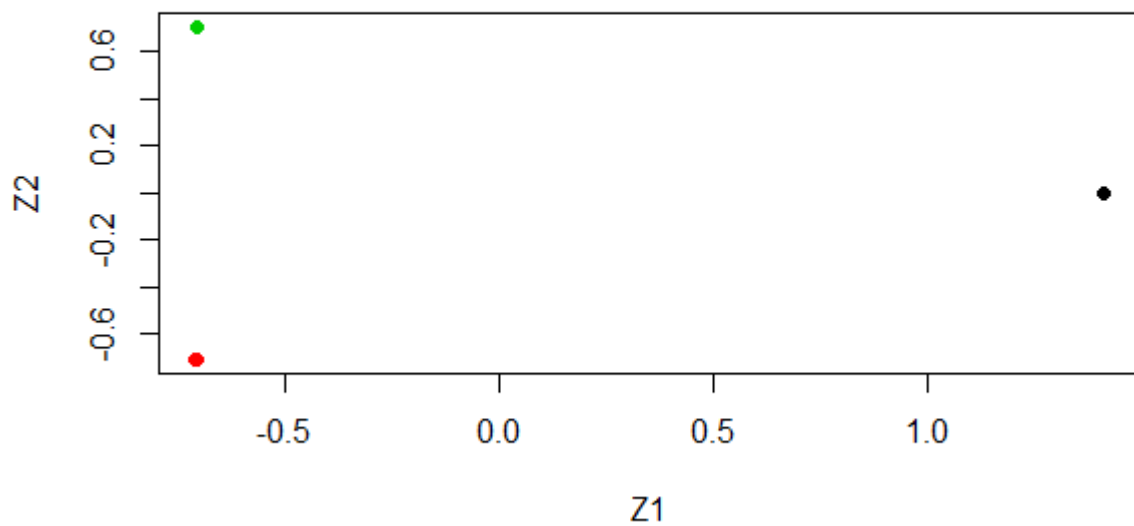
3. Chapter 10, Problem 10

(a)

```
1 #Problem3(a)
2 set.seed(1)
3 x = matrix(rnorm(20*3*50, mean=0, sd=0.001), ncol=50)
4 x[1:20, 2] = 1
5 x[21:40, 1] = 2
6 x[21:40, 2] = 2
7 x[41:60, 1] = 1
8 true.labels = c(rep(1, 20), rep(2, 20), rep(3, 20))
```

(b)

```
1 #3(b)
2 pr.out = prcomp(x)
3 plot(pr.out$x[, 1:2], col = 1:3, xlab = "Z1", ylab = "Z2", pch = 19)
```



(c)

```
1 #3(c)
2 km.out = kmeans(x, 3, nstart=20)
3 table(km.out$cluster, c(rep(1,20), rep(2,20), rep(3,20)))
```

```
1 Output:
2 > table(km.out$cluster, c(rep(1,20), rep(2,20), rep(3,20)))
3      1  2  3
4  1 20  0  0
5  2  0  0 20
6  3  0 20  0
```

(d)

```
1 #3(d)
2 km.out = kmeans(x, 2, nstart = 20)
3 table(true.labels, km.out$cluster)
```

```
1 Output:
2 > table(true.labels, km.out$cluster)
3 true.labels  1  2
4           1 20  0
5           2  0 20
6           3 20  0
```

(e)


```

1 #3(e)
2 km.out = kmeans(x, 4, nstart = 20)
3 table(true.labels, km.out$cluster)

```

```

1 Output:
2 > table(true.labels, km.out$cluster)
3 true.labels  1  2  3  4
4             1 10  0  0 10
5             2  0 20  0  0
6             3  0  0 20  0

```

(f)

```

1 #3(f)
2 km.out = kmeans(pr.out$x[, 1:2], 3, nstart = 20)
3 table(true.labels, km.out$cluster)

```

```

1 Output:
2 > table(true.labels, km.out$cluster)
3 true.labels  1  2  3
4             1  0 20  0
5             2  0  0 20
6             3 20  0  0

```

(g)

```

1 #3(g)
2 km.out = kmeans(scale(x), 3, nstart = 20)
3 table(true.labels, km.out$cluster)

```

```

1 Output:
2 > table(true.labels, km.out$cluster)
3 true.labels  1  2  3
4             1  3  8  9
5             2  8  2 10
6             3  8  8  4

```

4. Chapter 10, Problem 11

(a)

```

1 #Problem4(a)
2 genes = read.csv("Ch10Ex11.csv", header = FALSE)

```

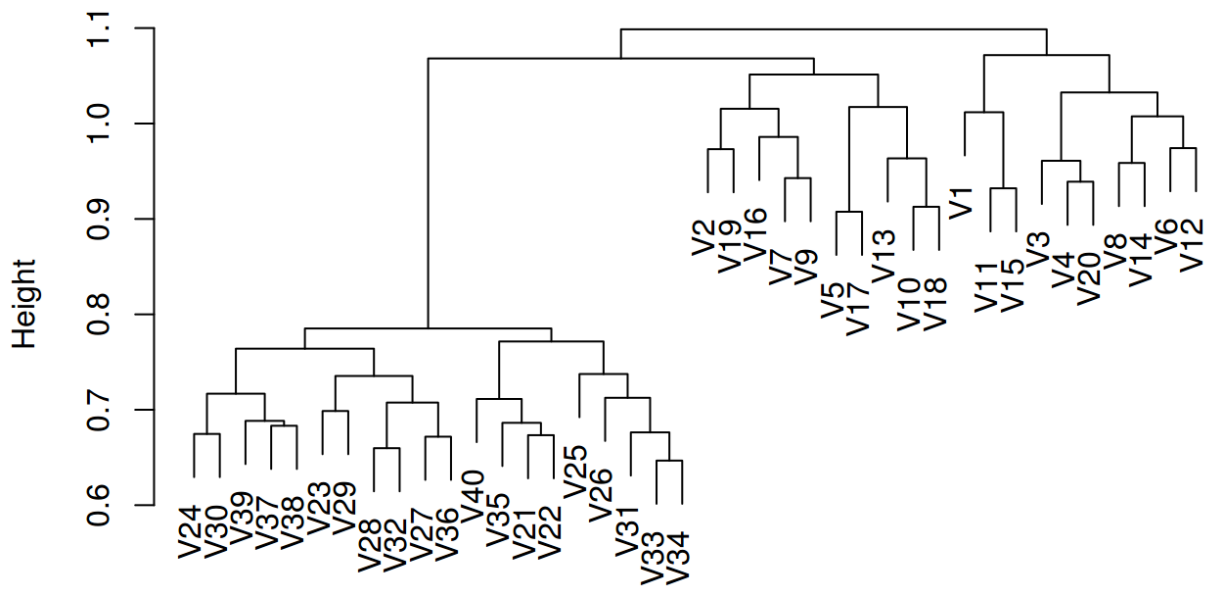
(b)

```

1 #4(b)
2 hc.complete = hclust(as.dist(1 - cor(genes)), method = "complete")
3 plot(hc.complete)

```

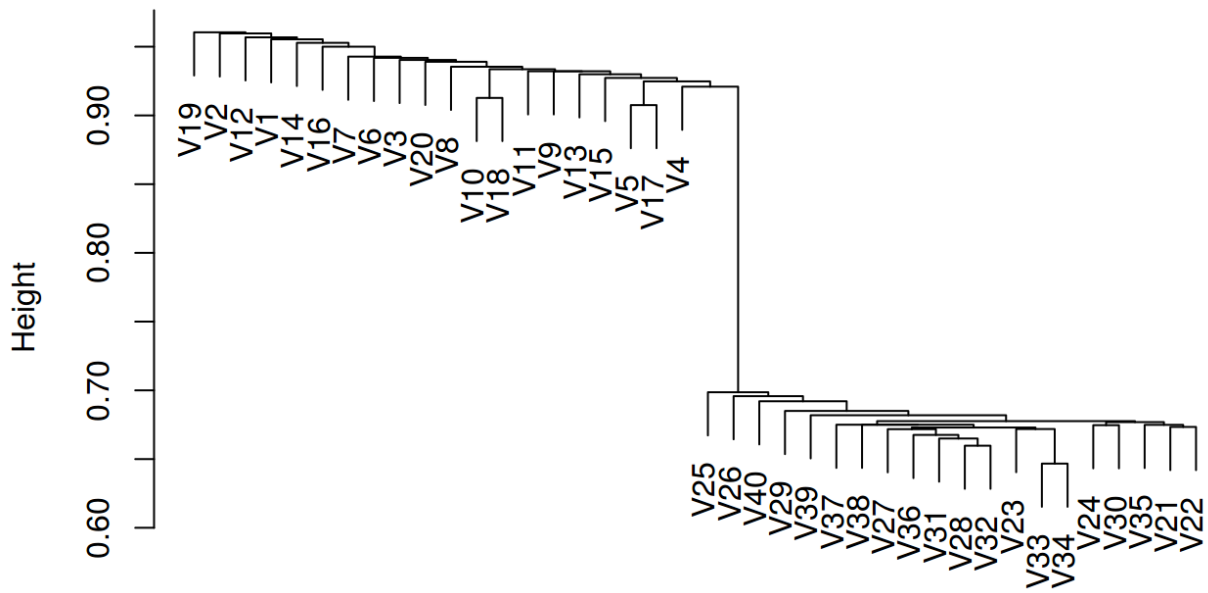
Cluster Dendrogram



```
as.dist(1 - cor(genes))  
hclust (*, "complete")
```

```
1 hc.single = hclust(as.dist(1 - cor(genes)), method = "single")  
2 plot(hc.single)
```

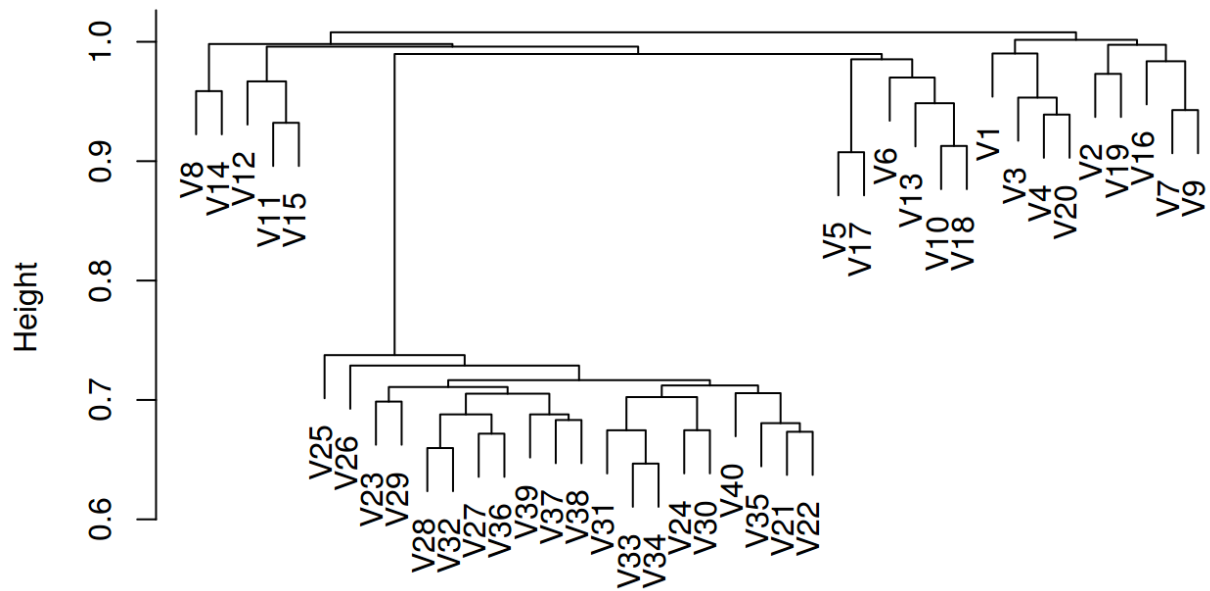
Cluster Dendrogram



```
as.dist(1 - cor(genes))  
hclust (*, "single")
```

```
1 | hc.average <- hclust(as.dist(1 - cor(genes)), method = "average")  
2 | plot(hc.average)
```

Cluster Dendrogram



```
as.dist(1 - cor(genes))
hclust (*, "average")
```

(c)

```
1 #4(c)
2 pr.out = prcomp(t(genes))
3 head(pr.out$rotation)
```

```

1 Output:
2 ##          PC1          PC2          PC3          PC4          PC5
3 ## [1,] -0.002434664 -0.030745799  0.009359932  0.009699551 -0.012847866
4 ## [2,] -0.002016598 -0.025927592  0.050300983 -0.026082885  0.003488293
5 ## [3,]  0.011233842 -0.003937802  0.014564920  0.054373032 -0.020411836
6 ## [4,]  0.013912855  0.025625408  0.033998676 -0.011530298 -0.009364524
7 ## [5,]  0.007293322  0.013590353 -0.008229702 -0.001343010  0.030002978
8 ## [6,]  0.017928318 -0.026302699 -0.020728401 -0.024069152 -0.018619253
9 ##          PC6          PC7          PC8          PC9          PC10
10 ## [1,]  0.023439995  0.010152261 -0.024602570 -0.021925557 -0.035003076
11 ## [2,]  0.001605492 -0.037364376 -0.017332292  0.011319311  0.007802611
12 ## [3,]  0.025337127  0.070772412  0.047340581 -0.013963868  0.023624407
13 ## [4,]  0.029529539  0.002885764 -0.093667774 -0.008391226 -0.019226470
14 ## [5,] -0.017042934  0.003555111 -0.053227214 -0.010479774  0.008446406
15 ## [6,] -0.049103273 -0.040473304 -0.005455454 -0.003882692  0.028472950
16 ##          PC11         PC12         PC13         PC14         PC15
17 ## [1,]  0.068133070  0.002322824 -0.050042837 -0.043957087  0.007542896
18 ## [2,] -0.092523227  0.036265781  0.002951734  0.021272662 -0.040075267
19 ## [3,]  0.017649621  0.021512568  0.013587072  0.005264628 -0.002918920
20 ## [4,]  0.006695624  0.025918069 -0.081179098  0.017689681  0.045951951
21 ## [5,]  0.053250618 -0.076682641 -0.049516326 -0.003282028  0.060755699
22 ## [6,] -0.018103035  0.015433035  0.015967833 -0.006985293 -0.025237500
23 ##          PC16         PC17         PC18         PC19         PC20
24 ## [1,] -0.04567334 -0.019899716  0.02946561 -0.009362957 -0.029855408
25 ## [2,]  0.03433259  0.003735211 -0.01218600 -0.023466062 -0.005495696
26 ## [3,]  0.01881913  0.003284517  0.02597233  0.021581732  0.016808524
27 ## [4,] -0.01062858  0.018342677 -0.03334608 -0.052262385 -0.030868339
28 ## [5,] -0.02562691  0.049934804 -0.04221058 -0.012279815  0.018004932
29 ## [6,] -0.00394582  0.037319024 -0.02541592 -0.029423771 -0.012043007
30 ##          PC21         PC22         PC23         PC24         PC25
31 ## [1,] -0.009190761  0.0230209664 -0.028970518  0.033060132  0.021453017
32 ## [2,] -0.002808309  0.0079065160 -0.007921167 -0.034424716  0.011932971
33 ## [3,]  0.010683143 -0.0392265342  0.004592080  0.026463736 -0.038085712
34 ## [4,]  0.079419742 -0.0001627164  0.070396594 -0.002015954  0.006459925
35 ## [5,] -0.038364004 -0.0230993500 -0.047439556 -0.001129421 -0.001285153
36 ## [6,] -0.004522525  0.0304001071  0.016062043 -0.019329595 -0.034486284
37 ##          PC26         PC27         PC28         PC29         PC30
38 ## [1,]  0.034447853  0.017729906  0.034708970 -0.028136309 -0.009873440
39 ## [2,]  0.051079165  0.032435028 -0.006934708 -0.026307151 -0.008143422
40 ## [3,] -0.064720318 -0.004616608  0.038015189  0.006455198  0.004570640
41 ## [4,]  0.022138389 -0.017120199  0.074901678  0.015812685  0.016391804
42 ## [5,] -0.010772594  0.010889806 -0.005305488  0.015248277  0.029303828
43 ## [6,]  0.001489549  0.028082907 -0.036617970 -0.054760935  0.023337598
44 ##          PC31         PC32         PC33         PC34         PC35
45 ## [1,] -0.03576788  0.016708304 -0.01823350  0.0007957941 -0.01443692
46 ## [2,] -0.04439239  0.011968530  0.04168309  0.0123210140  0.02739196
47 ## [3,]  0.02932866  0.026066011  0.02055204 -0.0716448783  0.02726941
48 ## [4,] -0.03954720  0.014714963  0.02846397  0.0316775643  0.01866774
49 ## [5,]  0.05494446 -0.005416152  0.03476606  0.0245476439 -0.04037835
50 ## [6,]  0.01132569  0.006320203 -0.00237484  0.0061140832  0.01402898
51 ##          PC36         PC37         PC38         PC39         PC40
52 ## [1,]  0.010652118 -0.009366629 -0.012754402  0.0020214363  0.07000786
53 ## [2,] -0.002733484 -0.001318693  0.031410461 -0.0108377476 -0.06326465

```

```
54 ## [3,]  0.020891497 -0.001380233 -0.025857254  0.0008800921 -0.32824953
55 ## [4,] -0.027363133 -0.006080650 -0.025316130 -0.0235404170 -0.01675446
56 ## [5,] -0.046869227 -0.017973802  0.002917167  0.0342753219  0.04896111
57 ## [6,]  0.042083325  0.055817170 -0.010080327  0.0029965594  0.05407104
```

```
1 total.load = apply(pr.out$rotation, 1, sum)
2 index = order(abs(total.load), decreasing = TRUE)
3 index[1:10]
```

```
1 Output:
2 ## [1] 865  68 911 428 624  11 524 803 980 822
```