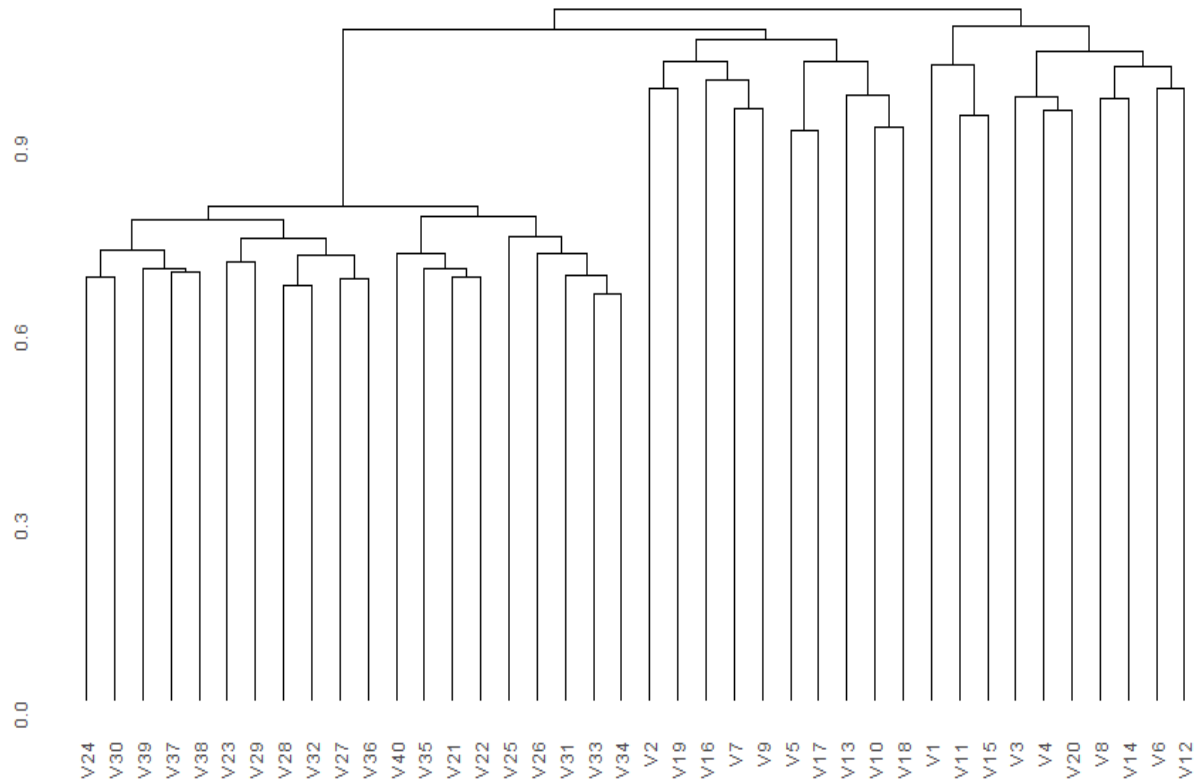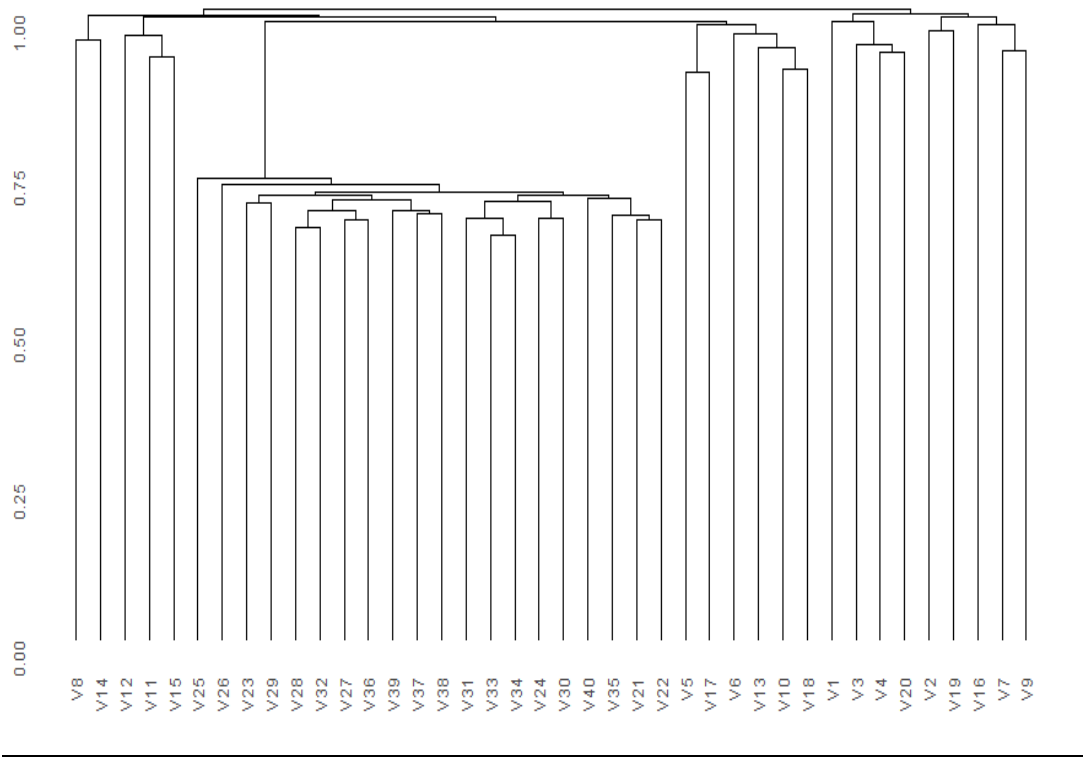3.)

a.) loaded the data of gene expression data set.

b.) Apply hierarchical clustering to the samples using correlation based distance:
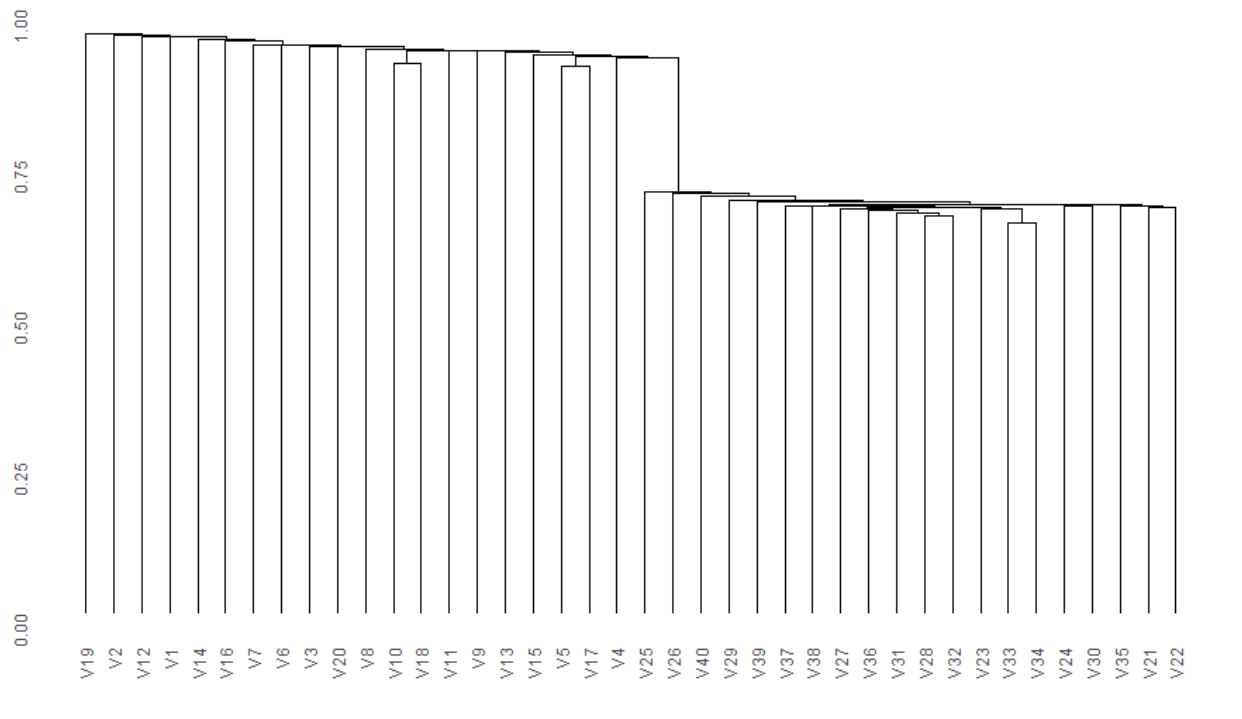
Complete Linkage:

In complete linkage, we define the distance between two clusters to be the maximum distance between any single data point in the first cluster and any single data point in the second cluster. For single linkage it is the minimum distance between any single data point in the first cluster and any single data point in the second cluster. Whereas for Average linkage it is the average distance between the data points.

Average Linkage:

Single Linkage:

The results obtained are different and they are based on the type of linkage that is used.

For,

Complete Linkage: Number of clusters = 2

Average Linkage: Number of clusters = 3

Single Linkage: Number of clusters = 2

c.)

To look at which genes differ the most across the healthy patients and diseased patients, we use principal component analysis. When the weights assigned to each gene are arranged in decreasing order the first ten of them will be the genes which differ a lot.

```
> head(PCA$rotation)
              PC1          PC2          PC3          PC4          PC5          PC6          PC7          PC8
[1,] -0.002434664 -0.030745799  0.009359932  0.009699551 -0.012847866  0.023439995  0.010152261 -0.024602570
[2,] -0.002016598 -0.025927592  0.050300983 -0.026082885  0.003488293  0.001605492 -0.037364376 -0.017332292
[3,]  0.011233842 -0.003937802  0.014564920  0.054373032 -0.020411836  0.025337127  0.070772412  0.047340581
[4,]  0.013912855  0.025625408  0.033998676 -0.011530298 -0.009364524  0.029529539  0.002885764 -0.093667774
[5,]  0.007293322  0.013590353 -0.008229702 -0.001343010  0.030002978 -0.017042934  0.003555111 -0.053227214
[6,]  0.017928318 -0.026302699 -0.020728401 -0.024069152 -0.018619253 -0.049103273 -0.040473304 -0.005455454
              PC9         PC10         PC11         PC12         PC13         PC14         PC15         PC16
[1,] -0.021925557 -0.035003076  0.068133070  0.002322824 -0.050042837 -0.043957087  0.007542896 -0.04567334
[2,]  0.011319311  0.007802611 -0.092523227  0.036265781  0.002951734  0.021272662 -0.040075267  0.03433259
[3,] -0.013963868  0.023624407  0.017649621  0.021512568  0.013587072  0.005264628 -0.002918920  0.01881913
[4,] -0.008391226 -0.019226470  0.006695624  0.025918069 -0.081179098  0.017689681  0.045951951 -0.01062858
[5,] -0.010479774  0.008446406  0.053250618 -0.076682641 -0.049516326 -0.003282028  0.060755699 -0.02562691
[6,] -0.003882692  0.028472950 -0.018103035  0.015433035  0.015967833 -0.006985293 -0.025237500 -0.00394582
             PC17         PC18         PC19         PC20         PC21         PC22         PC23         PC24
[1,] -0.019899716  0.02946561 -0.009362957 -0.029855408 -0.009190761  0.0230209664 -0.028970518  0.033060132
[2,]  0.003735211 -0.01218600 -0.023466062 -0.005495696 -0.002808309  0.0079065160 -0.007921167 -0.034424716
[3,]  0.003284517  0.02597233  0.021581732  0.016808524  0.010683143 -0.0392265342  0.004592080  0.026463736
[4,]  0.018342677 -0.03334608 -0.052262385 -0.030868339  0.079419742 -0.0001627164  0.070396594 -0.002015954
[5,]  0.049934804 -0.04221058 -0.012279815  0.018004932 -0.038364004 -0.0230993500 -0.047439556 -0.001129421
[6,]  0.037319024 -0.02541592 -0.029423771 -0.012043007 -0.004522525  0.0304001071  0.016062043 -0.019329595
             PC25         PC26         PC27         PC28         PC29         PC30         PC31         PC32
[1,]  0.021453017  0.034447853  0.017729906  0.034708970 -0.028136309 -0.009873440 -0.03576788  0.016708304
[2,]  0.011932971  0.051079165  0.032435028 -0.006934708 -0.026307151 -0.008143422 -0.04439239  0.011968530
[3,] -0.038085712 -0.064720318 -0.004616608  0.038015189  0.006455198  0.004570640  0.02932866  0.026066011
[4,]  0.006459925  0.022138389 -0.017120199  0.074901678  0.015812685  0.016391804 -0.03954720  0.014714963
[5,] -0.001285153 -0.010772594  0.010889806 -0.005305488  0.015248277  0.029303828  0.05494446 -0.005416152
[6,] -0.034486284  0.001489549  0.028082907 -0.036617970 -0.054760935  0.023337598  0.01132569  0.006320203
             PC33         PC34         PC35         PC36         PC37         PC38         PC39         PC40
[1,] -0.01823350  0.0007957941 -0.01443692  0.010652118 -0.009366629 -0.012754402  0.0020214363  0.21757848
[2,]  0.04168309  0.0123210140  0.02739196 -0.002733484 -0.001318693  0.031410461 -0.0108377476  0.03272198
[3,]  0.02055204 -0.0716448783  0.02726941  0.020891497 -0.001380233 -0.025857254  0.0008800921 -0.19644706
[4,]  0.02846397  0.0316775643  0.01866774 -0.027363133 -0.006080650 -0.025316130 -0.0235404170 -0.19390100
[5,]  0.03476606  0.0245476439 -0.04037835 -0.046869227 -0.017973802  0.002917167  0.0342753219  0.03476877
[6,] -0.00237484  0.0061140832  0.01402898  0.042083325  0.055817170 -0.010080327  0.0029965594  0.29420482
```

Genes that differ the most across the two groups:

```
> total_load <- apply(PCA$rotation,1,sum)
> indices <- order(abs(total_load),decreasing = TRUE)
> indices[1:10]
 [1] 865 911  68 428 624 980 524 803 822 570
> total_load[indices[1:10]]
 [1]  0.7723461 -0.7283080  0.7224073 -0.6354402 -0.5968822 -0.5558179  0.5510418  0.5295094  0.4942181  0.4825781
>
>
```

These are the ten most different genes across the group.