

Yeast Image Analysis Data

Statistical Consulting Center, Department of Statistical Science, Duke University

January 21, 2025

1 Import and Summarize the Data

```
system("du -h ThreeStrainsFourImagesTwoWells.csv")
y<-read.csv("ThreeStrainsFourImagesTwoWells.csv",as.is=TRUE)
dim(y)

## [1] 20117    29

length(unique(y$ObjectNumber))

## [1] 1014

summary(y$ObjectNumber)

##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##      1.0   210.0   420.0   423.7   629.0  1014.0

length(unique(y$FileName_CellPosition))

## [1] 24

y$strain<-substr(y$FileName_CellPosition,36,36)
table(y$strain)

##
##      C      D      E
## 6409 7183 6525

y$well<-substr(y$FileName_CellPosition,37,38)
table(y$well)

##
##      03      04
## 9441 10676

img1<-substr(y$FileName_CellPosition,43,43)
img2<-substr(y$FileName_CellPosition,48,48)
y$image<-paste0(img1,".",img2)
table(y$image)

##
## 1.1 1.2 1.3 2.1
## 4994 5052 5193 4878
```

```

rm(img1,img2)
length(unique(y$Number_Object_Number))

## [1] 1014

dropme<-c("FileName_CellPosition","PathName_CellPosition",
          "Number_Object_Number","Location_Center_Z",
          "AreaShape_EulerNumber")
y<-y[,!(colnames(y) %in% dropme)]
y$uniqueWell<-paste0(y$strain,"-",y$well)
y$uniqueImage<-paste0(y$uniqueWell,"-",y$image)
y$uniqueID<-paste0(y$uniqueImage,"-",y$ObjectNumber)
y<-y[,!(colnames(y) %in% c("image","well","ObjectNumber"))]
length(unique(y$uniqueID))

## [1] 20117

dim(y)

## [1] 20117    30

rownames(y)<-y$uniqueID
head(y)

##           AreaShape_Area AreaShape_BoundingBoxArea
## C-03-1.1-1             865                  1216
## C-03-1.1-2             931                  1190
## C-03-1.1-3             673                   930
## C-03-1.1-4             383                   575
## C-03-1.1-5             462                   576
## C-03-1.1-6            1224                  2016
##           AreaShape_BoundingBoxMaximum_X AreaShape_BoundingBoxMaximum_Y
## C-03-1.1-1                  1177                      41
## C-03-1.1-2                   776                      43
## C-03-1.1-3                   973                      43
## C-03-1.1-4                  1458                      40
## C-03-1.1-5                   489                      49
## C-03-1.1-6                   427                      55
##           AreaShape_BoundingBoxMinimum_X AreaShape_BoundingBoxMinimum_Y
## C-03-1.1-1                  1139                      9
## C-03-1.1-2                   742                      8
## C-03-1.1-3                   943                     12
## C-03-1.1-4                  1433                     17
## C-03-1.1-5                   465                     25
## C-03-1.1-6                   385                      7
##           AreaShape_Center_X AreaShape_Center_Y AreaShape_Compactness
## C-03-1.1-1          1155.9526          24.62312          1.225075
## C-03-1.1-2           758.2385          25.18153          1.067232
## C-03-1.1-3           957.6152          25.72363          1.110523
## C-03-1.1-4          1443.5979          27.90339          1.138146
## C-03-1.1-5           476.5455          36.30952          1.062148
## C-03-1.1-6           403.6201          35.41748          1.817441
##           AreaShape_ConvexArea AreaShape_Eccentricity

```

##	C-03-1.1-1	928	0.5162789	
##	C-03-1.1-2	956	0.2729518	
##	C-03-1.1-3	701	0.2547679	
##	C-03-1.1-4	405	0.3181507	
##	C-03-1.1-5	478	0.3280023	
##	C-03-1.1-6	1454	0.5251552	
##	AreaShape_EquivalentDiameter	AreaShape_Extent	AreaShape_FormFactor	
##	C-03-1.1-1	33.18663	0.7113487	0.8162763
##	C-03-1.1-2	34.42944	0.7823529	0.9370037
##	C-03-1.1-3	29.27269	0.7236559	0.9004770
##	C-03-1.1-4	22.08282	0.6660870	0.8786220
##	C-03-1.1-5	24.25359	0.8020833	0.9414880
##	C-03-1.1-6	39.47715	0.6071429	0.5502242
##	AreaShape_MajorAxisLength	AreaShape_MaxFerretDiameter		
##	C-03-1.1-1	36.16614	37.36308	
##	C-03-1.1-2	35.10777	34.82815	
##	C-03-1.1-3	29.83668	30.41381	
##	C-03-1.1-4	22.85095	24.08319	
##	C-03-1.1-5	24.97477	25.00000	
##	C-03-1.1-6	44.64158	52.47857	
##	AreaShape_MaximumRadius	AreaShape_MeanRadius	AreaShape_MedianRadius	
##	C-03-1.1-1	15.23155	5.485455	5.000000
##	C-03-1.1-2	16.64332	5.998279	5.099020
##	C-03-1.1-3	13.60147	5.072066	4.472136
##	C-03-1.1-4	10.00000	3.846808	3.605551
##	C-03-1.1-5	11.66190	4.288451	4.000000
##	C-03-1.1-6	16.55295	5.766630	5.000000
##	AreaShape_MinFerretDiameter	AreaShape_MinorAxisLength		
##	C-03-1.1-1	30.85774	30.97342	
##	C-03-1.1-2	33.00000	33.77465	
##	C-03-1.1-3	28.17446	28.85213	
##	C-03-1.1-4	20.82649	21.66361	
##	C-03-1.1-5	23.00000	23.59309	
##	C-03-1.1-6	37.00848	37.99027	
##	AreaShape_Orientation	AreaShape_Perimeter	AreaShape_Solidity	
##	C-03-1.1-1	-84.84566	115.39697	0.9321121
##	C-03-1.1-2	-32.85928	111.74012	0.9738494
##	C-03-1.1-3	40.67816	96.91169	0.9600571
##	C-03-1.1-4	-35.46540	74.01219	0.9456790
##	C-03-1.1-5	-38.76810	78.52691	0.9665272
##	C-03-1.1-6	40.37440	167.19596	0.8418157
##	Location_Center_X	Location_Center_Y	strain	uniqueWell
##	C-03-1.1-1	1155.9526	24.62312	C
##	C-03-1.1-2	758.2385	25.18153	C
##	C-03-1.1-3	957.6152	25.72363	C
##	C-03-1.1-4	1443.5979	27.90339	C
##	C-03-1.1-5	476.5455	36.30952	C
##	C-03-1.1-6	403.6201	35.41748	C
##	uniqueID			
##	C-03-1.1-1	C-03-1.1-1		
##	C-03-1.1-2	C-03-1.1-2		

```
## C-03-1.1-3 C-03-1.1-3
## C-03-1.1-4 C-03-1.1-4
## C-03-1.1-5 C-03-1.1-5
## C-03-1.1-6 C-03-1.1-6

apply(is.na(y),2,sum)

##           AreaShape_Area           AreaShape_BoundingBoxArea
##                0                0
## AreaShape_BoundingBoxMaximum_X AreaShape_BoundingBoxMaximum_Y
##                0                0
## AreaShape_BoundingBoxMinimum_X AreaShape_BoundingBoxMinimum_Y
##                0                0
##           AreaShape_Center_X           AreaShape_Center_Y
##                0                0
##           AreaShape_Compactness           AreaShape_ConvexArea
##                0                0
##           AreaShape_Eccentricity AreaShape_EquivalentDiameter
##                0                0
##           AreaShape_Extent           AreaShape_FormFactor
##                0                0
## AreaShape_MajorAxisLength AreaShape_MaxFerretDiameter
##                0                0
##           AreaShape_MaximumRadius           AreaShape_MeanRadius
##                0                0
##           AreaShape_MedianRadius AreaShape_MinFerretDiameter
##                0                0
## AreaShape_MinorAxisLength AreaShape_Orientation
##                0                0
##           AreaShape_Perimeter           AreaShape_Solidity
##                0                0
##           Location_Center_X           Location_Center_Y
##                0                0
##                strain                uniqueWell
##                0                0
##           uniqueImage                uniqueID
##                0                0
```

Add counts of neighbors within 50 units (10 times the median object radius). These computations are (necessarily) made within each image using the function `nNN()`, which is loaded from the input file `RFuncs.R`, but copied here.

```
## Two measures of location are the same:
cor(y$AreaShape_Center_X,y$Location_Center_X)

## [1] 1

cor(y$AreaShape_Center_Y,y$Location_Center_Y)

## [1] 1

## BY-IMAGE computation:
## number of (nearest) neighbors (with centroids) within a ball of radius r
```

```
## around each feature's centroid.
## nNN<-function(img,r){
##   y.i<-y[y$uniqueImage==img,c("Location_Center_X","Location_Center_Y")]
##   ## pairwise Euclidean distances
##   d.x<-(outer(y.i$Location_Center_X,y.i$Location_Center_X,FUN="-")^2)
##   d.y<-(outer(y.i$Location_Center_Y,y.i$Location_Center_Y,FUN="-")^2)
##   dist<-sqrt(d.x + d.y)
##   dist<-(0.5*dist + 0.5*t(dist))
##   ## number of (all) neighbors within a ball of radius=r
##   n.nbrs<-apply(dist<r,1,sum)
##   return(n.nbrs)
## }
```

```
## median median = 5 for all:
summary(y$AreaShape_MedianRadius[y$strain=="C"])

##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##      2.000  4.183   5.000   4.778   5.099   9.137

summary(y$AreaShape_MedianRadius[y$strain=="D"])

##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##      2.000  4.123   5.000   4.721   5.000   9.849

summary(y$AreaShape_MedianRadius[y$strain=="E"])

##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##      2.000  4.123   5.000   4.615   5.000   9.000

## diameter 10 each; in linear arrangement centers occur
## at distance 2r=d from each other
table(nNN(img="C-03-1.1",r=50))

##
##      0      1      2      3      4
## 687    24    16    11     5

table(y$uniqueImage)

##
## C-03-1.1 C-03-1.2 C-03-1.3 C-03-2.1 C-04-1.1 C-04-1.2 C-04-1.3 C-04-2.1
##      743      643      751      734      815      939      954      830
## D-03-1.1 D-03-1.2 D-03-1.3 D-03-2.1 D-04-1.1 D-04-1.2 D-04-1.3 D-04-2.1
##      834      841      873      812      969     1014      944      896
## E-03-1.1 E-03-1.2 E-03-1.3 E-03-2.1 E-04-1.1 E-04-1.2 E-04-1.3 E-04-2.1
##      840      773      806      791      793      842      865      815

uImage<-unique(y$uniqueImage)
y$nNN50<-rep(NA,nrow(y))
for (i in 1:length(uImage)){
  y[y$uniqueImage==uImage[i], "nNN50"]<-nNN(img=uImage[i],r=50)
}
table(y$nNN50,useNA="always")
```

```
##
##      0      1      2      3      4      5      6      7      8 <NA>
## 18676  490  419  296  159   63   12    1    1    0
```

```
table(y$uniqueImage,y$nn50,useNA="always")
```

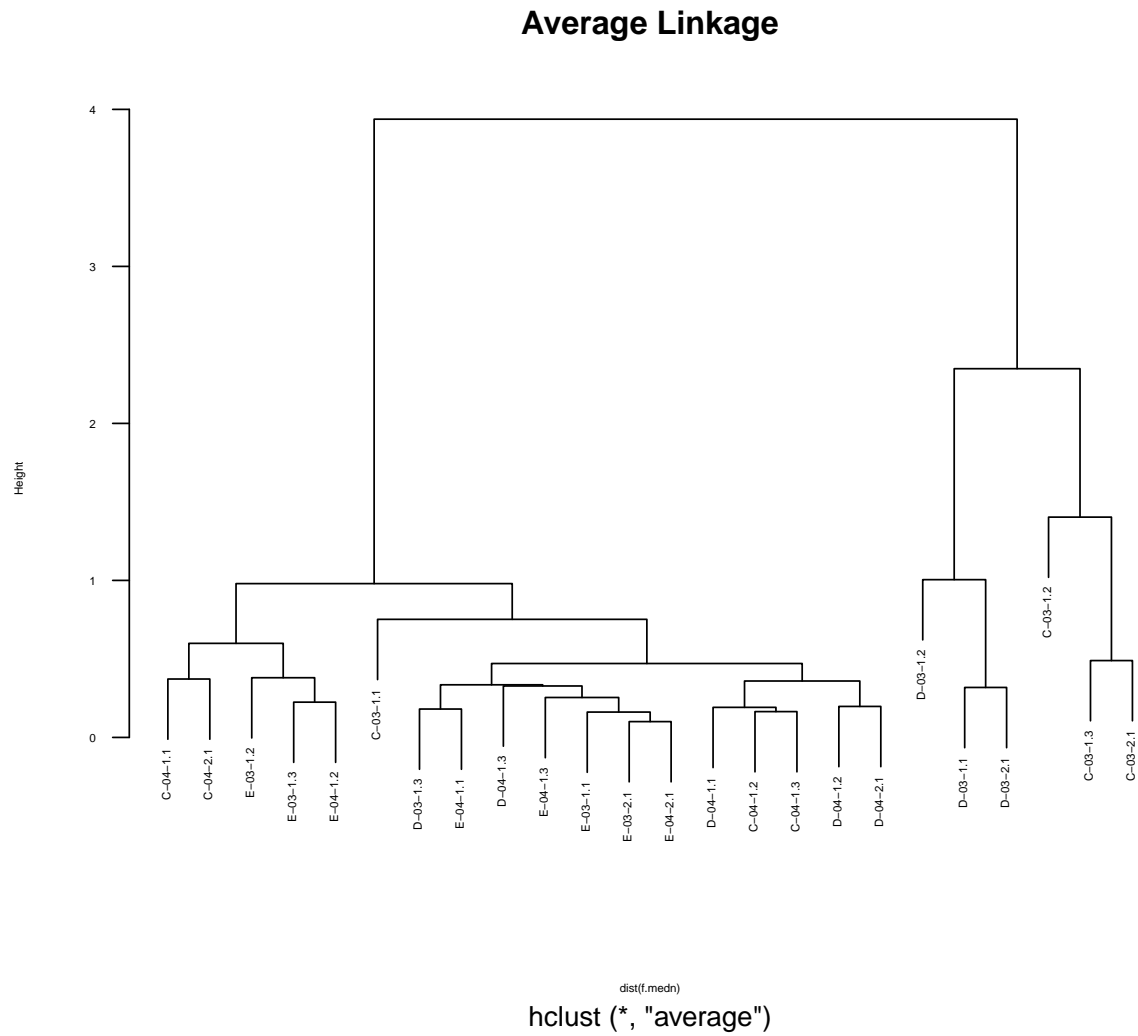
```
##
##      0      1      2      3      4      5      6      7      8 <NA>
## C-03-1.1 687  24  16  11   5   0   0   0   0   0
## C-03-1.2 596  22  12  10   2   1   0   0   0   0
## C-03-1.3 713  14  14   7   3   0   0   0   0   0
## C-03-2.1 689  20  12   9   3   1   0   0   0   0
## C-04-1.1 767  19  14   9   5   1   0   0   0   0
## C-04-1.2 874  15  22  15   8   3   2   0   0   0
## C-04-1.3 879  25  14  15  14   7   0   0   0   0
## C-04-2.1 773  23  10  10  10   4   0   0   0   0
## D-03-1.1 773  19  24  11   6   1   0   0   0   0
## D-03-1.2 774  23  19  15   9   0   1   0   0   0
## D-03-1.3 821  22  12  13   4   1   0   0   0   0
## D-03-2.1 740  30  16  20   4   0   1   0   1   0
## D-04-1.1 898  25  14  17   7   4   3   1   0   0
## D-04-1.2 944  23  26  11   4   5   1   0   0   0
## D-04-1.3 880  14  22  19   6   3   0   0   0   0
## D-04-2.1 826  15  20  15  14   5   1   0   0   0
## E-03-1.1 783  19  16  15   6   1   0   0   0   0
## E-03-1.2 719  20  20  10   2   2   0   0   0   0
## E-03-1.3 746  18  15  16   7   4   0   0   0   0
## E-03-2.1 727  17  28   4   9   6   0   0   0   0
## E-04-1.1 728  19  18  11  10   6   1   0   0   0
## E-04-1.2 782  15  16  12  13   4   0   0   0   0
## E-04-1.3 799  25  23  12   3   2   1   0   0   0
## E-04-2.1 758  24  16   9   5   2   1   0   0   0
## <NA>      0   0   0   0   0   0   0   0   0
```

2 Hierarchical Agglomerative Clustering With `hclust()`

```
drop<-c("strain","uniqueWell","uniqueImage","uniqueID","AreaShape_Center_X","AreaShape_Center_Y","Locat
keep<-(!(colnames(y) %in% drop))
temp<-scale(y[,keep])
## image-level feature summaries
f.medn<-matrix(NA,length(uImage),ncol(temp))
colnames(f.medn)<-paste0(colnames(temp),".medn")
for (i in 1:ncol(temp)){
  l.out<-unlist(lapply(split(temp[,i],y$uniqueImage),median,na.rm=TRUE))
  f.medn[,i]<-l.out
  if (i == 1) rownames(f.medn)<-names(l.out)
}
```

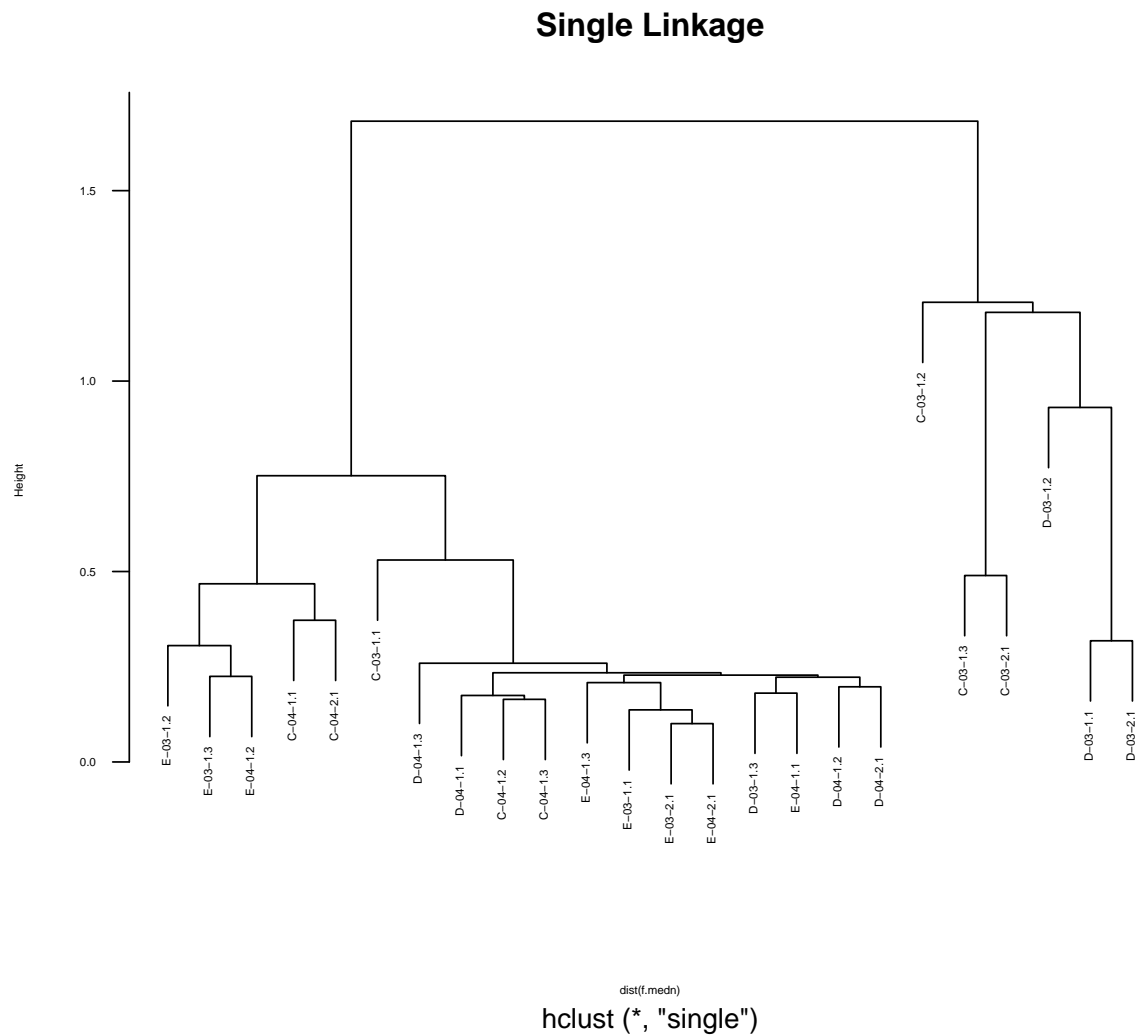
2.1 Average Linkage Clustering

```
h<-hclust(dist(f.medn),method="average")
plot(h,main="Average Linkage",cex=0.4, cex.axis=0.4,cex.lab=0.4,las=1)
```



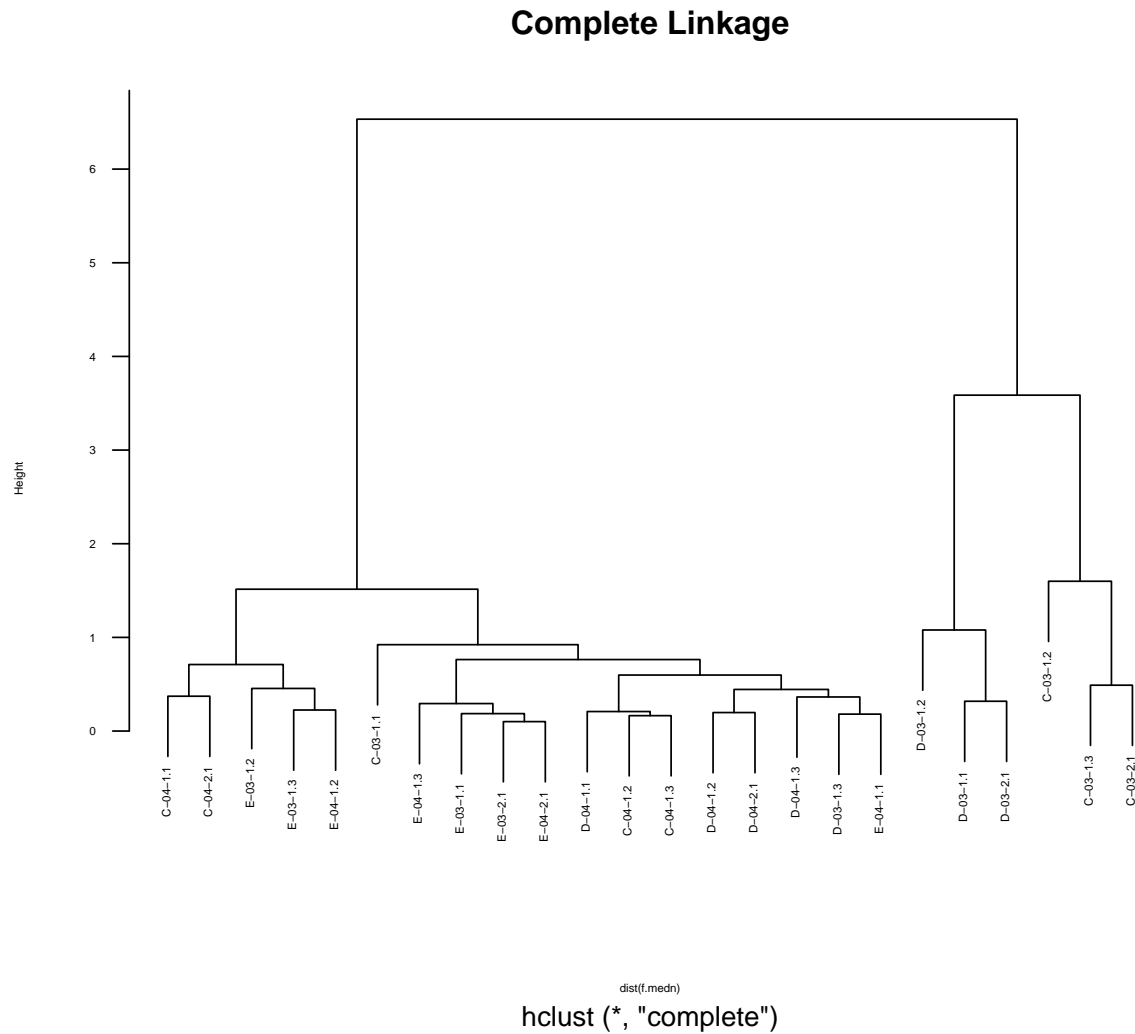
2.2 Single Linkage Clustering

```
h<-hclust(dist(f.medn),method="single")
plot(h,main="Single Linkage",cex=0.4, cex.axis=0.4,cex.lab=0.4,las=1)
```



2.3 Complete Linkage Clustering

```
h<-hclust(dist(f.medn),method="complete")
plot(h,main="Complete Linkage",cex=0.4, cex.axis=0.4,cex.lab=0.4,las=1)
```



2.4 Three Group Classification

```
hc.class<-cutree(h,3)
table(cutree(h,3))
```

```
##
##  1  2  3
## 18  3  3
```

2.5 Heatmap

```
heatmap.2(as.matrix(f.medn),scale="none",trace="none",
          mar=c(15,6),cexCol=0.7)
```

