Clustering for the crime data

library(readr)  
crime\_data<-read.csv("file:///E:/assignments data/clustering/crime\_data.csv")  
View(crime\_data)  
attach(crime\_data)  
summary(crime\_data)

## X Murder Assault UrbanPop   
## Alabama : 1 Min. : 0.800 Min. : 45.0 Min. :32.00   
## Alaska : 1 1st Qu.: 4.075 1st Qu.:109.0 1st Qu.:54.50   
## Arizona : 1 Median : 7.250 Median :159.0 Median :66.00   
## Arkansas : 1 Mean : 7.788 Mean :170.8 Mean :65.54   
## California: 1 3rd Qu.:11.250 3rd Qu.:249.0 3rd Qu.:77.75   
## Colorado : 1 Max. :17.400 Max. :337.0 Max. :91.00   
## (Other) :44   
## Rape   
## Min. : 7.30   
## 1st Qu.:15.07   
## Median :20.10   
## Mean :21.23   
## 3rd Qu.:26.18   
## Max. :46.00   
##

str(crime\_data)

## 'data.frame': 50 obs. of 5 variables:  
## $ X : Factor w/ 50 levels "Alabama","Alaska",..: 1 2 3 4 5 6 7 8 9 10 ...  
## $ Murder : num 13.2 10 8.1 8.8 9 7.9 3.3 5.9 15.4 17.4 ...  
## $ Assault : int 236 263 294 190 276 204 110 238 335 211 ...  
## $ UrbanPop: int 58 48 80 50 91 78 77 72 80 60 ...  
## $ Rape : num 21.2 44.5 31 19.5 40.6 38.7 11.1 15.8 31.9 25.8 ...

crime\_data1<-crime\_data[, 2:5]  
View(crime\_data1)  
  
#Normalizing the data  
normalized\_data<-scale(crime\_data1)  
normalized\_data

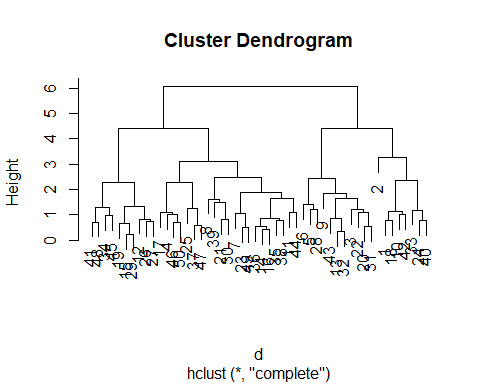
## Murder Assault UrbanPop Rape  
## [1,] 1.24256408 0.78283935 -0.52090661 -0.003416473  
## [2,] 0.50786248 1.10682252 -1.21176419 2.484202941  
## [3,] 0.07163341 1.47880321 0.99898006 1.042878388  
## [4,] 0.23234938 0.23086801 -1.07359268 -0.184916602  
## [5,] 0.27826823 1.26281442 1.75892340 2.067820292  
## [6,] 0.02571456 0.39885929 0.86080854 1.864967207  
## [7,] -1.03041900 -0.72908214 0.79172279 -1.081740768  
## [8,] -0.43347395 0.80683810 0.44629400 -0.579946294  
## [9,] 1.74767144 1.97077766 0.99898006 1.138966691  
## [10,] 2.20685994 0.48285493 -0.38273510 0.487701523  
## [11,] -0.57123050 -1.49704226 1.20623733 -0.110181255  
## [12,] -1.19113497 -0.60908837 -0.79724965 -0.750769945  
## [13,] 0.59970018 0.93883125 1.20623733 0.295524916  
## [14,] -0.13500142 -0.69308401 -0.03730631 -0.024769429  
## [15,] -1.28297267 -1.37704849 -0.58999237 -1.060387812  
## [16,] -0.41051452 -0.66908525 0.03177945 -0.345063775  
## [17,] 0.43898421 -0.74108152 -0.93542116 -0.526563903  
## [18,] 1.74767144 0.93883125 0.03177945 0.103348309  
## [19,] -1.30593210 -1.05306531 -1.00450692 -1.434064548  
## [20,] 0.80633501 1.55079947 0.10086521 0.701231086  
## [21,] -0.77786532 -0.26110644 1.34440885 -0.526563903  
## [22,] 0.99001041 1.01082751 0.58446551 1.480613993  
## [23,] -1.16817555 -1.18505846 0.03177945 -0.676034598  
## [24,] 1.90838741 1.05882502 -1.48810723 -0.441152078  
## [25,] 0.27826823 0.08687549 0.30812248 0.743936999  
## [26,] -0.41051452 -0.74108152 -0.86633540 -0.515887425  
## [27,] -0.80082475 -0.82507715 -0.24456358 -0.505210947  
## [28,] 1.01296983 0.97482938 1.06806582 2.644350114  
## [29,] -1.30593210 -1.36504911 -0.65907813 -1.252564419  
## [30,] -0.08908257 -0.14111267 1.62075188 -0.259651949  
## [31,] 0.82929443 1.37080881 0.30812248 1.160319648  
## [32,] 0.76041616 0.99882813 1.41349461 0.519730957  
## [33,] 1.19664523 1.99477641 -1.41902147 -0.547916860  
## [34,] -1.60440462 -1.50904164 -1.48810723 -1.487446939  
## [35,] -0.11204199 -0.60908837 0.65355127 0.017936483  
## [36,] -0.27275797 -0.23710769 0.16995096 -0.131534211  
## [37,] -0.66306820 -0.14111267 0.10086521 0.861378259  
## [38,] -0.34163624 -0.77707965 0.44629400 -0.676034598  
## [39,] -1.00745957 0.03887798 1.48258036 -1.380682157  
## [40,] 1.51807718 1.29881255 -1.21176419 0.135377743  
## [41,] -0.91562187 -1.01706718 -1.41902147 -0.900240639  
## [42,] 1.24256408 0.20686926 -0.45182086 0.605142783  
## [43,] 1.12776696 0.36286116 0.99898006 0.455672088  
## [44,] -1.05337842 -0.60908837 0.99898006 0.178083656  
## [45,] -1.28297267 -1.47304350 -2.31713632 -1.071064290  
## [46,] 0.16347111 -0.17711080 -0.17547783 -0.056798864  
## [47,] -0.86970302 -0.30910395 0.51537975 0.530407436  
## [48,] -0.47939280 -1.07706407 -1.83353601 -1.273917376  
## [49,] -1.19113497 -1.41304662 0.03177945 -1.113770203  
## [50,] -0.22683912 -0.11711392 -0.38273510 -0.601299251  
## attr(,"scaled:center")  
## Murder Assault UrbanPop Rape   
## 7.788 170.760 65.540 21.232   
## attr(,"scaled:scale")  
## Murder Assault UrbanPop Rape   
## 4.355510 83.337661 14.474763 9.366385

#distance matrix  
d <- dist(normalized\_data, method="euclidean")

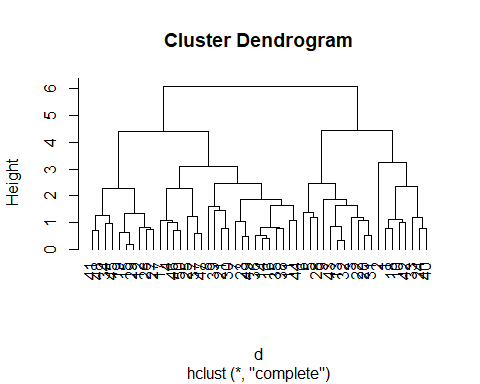
#Heirarchical clustering using complete linkage method  
fit<-hclust(d, method="complete")  
fit

##   
## Call:  
## hclust(d = d, method = "complete")  
##   
## Cluster method : complete   
## Distance : euclidean   
## Number of objects: 50

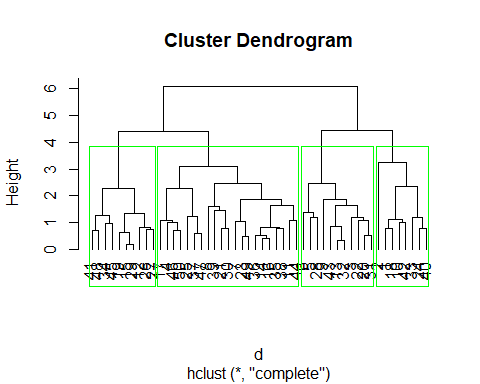
#dendrogram  
plot(fit)



plot(fit, hang=-1)



rect.hclust(fit,plot(fit, hang=-1), k=4, border="green")



#using Cutree  
groups<-cutree(fit, k=4)  
membership<-as.matrix(groups)

final<-data.frame(crime\_data, membership)  
final1<-final[,c(ncol(final),1:(ncol(final)-1))]  
final1

## membership X Murder Assault UrbanPop Rape  
## 1 1 Alabama 13.2 236 58 21.2  
## 2 1 Alaska 10.0 263 48 44.5  
## 3 2 Arizona 8.1 294 80 31.0  
## 4 3 Arkansas 8.8 190 50 19.5  
## 5 2 California 9.0 276 91 40.6  
## 6 2 Colorado 7.9 204 78 38.7  
## 7 3 Connecticut 3.3 110 77 11.1  
## 8 3 Delaware 5.9 238 72 15.8  
## 9 2 Florida 15.4 335 80 31.9  
## 10 1 Georgia 17.4 211 60 25.8  
## 11 3 Hawaii 5.3 46 83 20.2  
## 12 4 Idaho 2.6 120 54 14.2  
## 13 2 Illinois 10.4 249 83 24.0  
## 14 3 Indiana 7.2 113 65 21.0  
## 15 4 Iowa 2.2 56 57 11.3  
## 16 3 Kansas 6.0 115 66 18.0  
## 17 3 Kentucky 9.7 109 52 16.3  
## 18 1 Louisiana 15.4 249 66 22.2  
## 19 4 Maine 2.1 83 51 7.8  
## 20 2 Maryland 11.3 300 67 27.8  
## 21 3 Massachusetts 4.4 149 85 16.3  
## 22 2 Michigan 12.1 255 74 35.1  
## 23 3 Minnesota 2.7 72 66 14.9  
## 24 1 Mississippi 16.1 259 44 17.1  
## 25 3 Missouri 9.0 178 70 28.2  
## 26 4 Montana 6.0 109 53 16.4  
## 27 4 Nebraska 4.3 102 62 16.5  
## 28 2 Nevada 12.2 252 81 46.0  
## 29 4 New Hampshire 2.1 57 56 9.5  
## 30 3 New Jersey 7.4 159 89 18.8  
## 31 2 New Mexico 11.4 285 70 32.1  
## 32 2 New York 11.1 254 86 26.1  
## 33 1 North Carolina 13.0 337 45 16.1  
## 34 4 North Dakota 0.8 45 44 7.3  
## 35 3 Ohio 7.3 120 75 21.4  
## 36 3 Oklahoma 6.6 151 68 20.0  
## 37 3 Oregon 4.9 159 67 29.3  
## 38 3 Pennsylvania 6.3 106 72 14.9  
## 39 3 Rhode Island 3.4 174 87 8.3  
## 40 1 South Carolina 14.4 279 48 22.5  
## 41 4 South Dakota 3.8 86 45 12.8  
## 42 1 Tennessee 13.2 188 59 26.9  
## 43 2 Texas 12.7 201 80 25.5  
## 44 3 Utah 3.2 120 80 22.9  
## 45 4 Vermont 2.2 48 32 11.2  
## 46 3 Virginia 8.5 156 63 20.7  
## 47 3 Washington 4.0 145 73 26.2  
## 48 4 West Virginia 5.7 81 39 9.3  
## 49 3 Wisconsin 2.6 53 66 10.8  
## 50 3 Wyoming 6.8 161 60 15.6

aggregate(crime\_data[,-1],by=list(final$membership),mean)

## Group.1 Murder Assault UrbanPop Rape  
## 1 1 14.087500 252.7500 53.50000 24.53750  
## 2 2 11.054545 264.0909 79.09091 32.61818  
## 3 3 5.871429 134.4762 70.76190 18.58095  
## 4 4 3.180000 78.7000 49.30000 11.63000

# Group 2 have the higher rate of crime