Chapter 6 Homework

Shuowen Wei

1.

(a).

> M<-matrix(0,6,6)

> fix(M)

> M

col1 col2 col3 col4 col5 col6

[1,] 0 8 9 13 18 16

[2,] 8 0 1 5 10 8

[3,] 9 1 0 4 9 7

[4,] 13 5 4 0 5 3

[5,] 18 10 9 5 0 2

[6,] 16 8 7 3 2 0

> P<-round(cmdscale(M,k=1,eig=TRUE)$points)

> P

[,1]

[1,] 11

[2,] 3

[3,] 2

[4,] -2

[5,] -7

[6,] -5

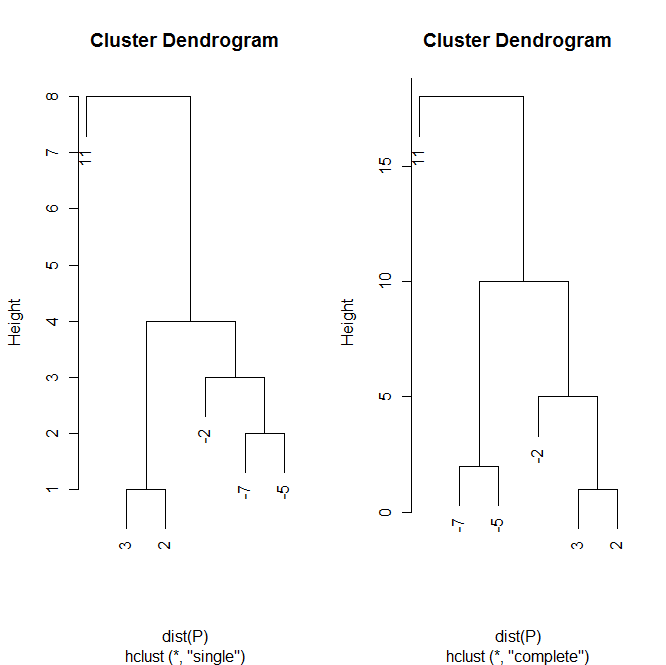
(b).

> rownames(P)<-as.vector(P)

> pm(1,2)

> plot(hclust(dist(P),method="single"))

> plot(hclust(dist(P),method="complete"))



(c).

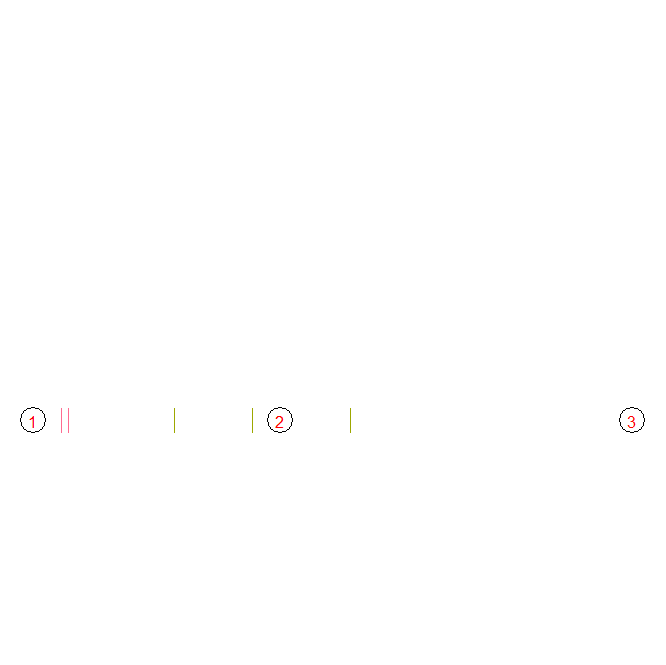
Difference:

In the “single” cluster diagram, (-2, -5, -7) are in the same cluster, while in the “complete” diagram, (-2, 2, 3) are in the same cluster.

(d).

(e).

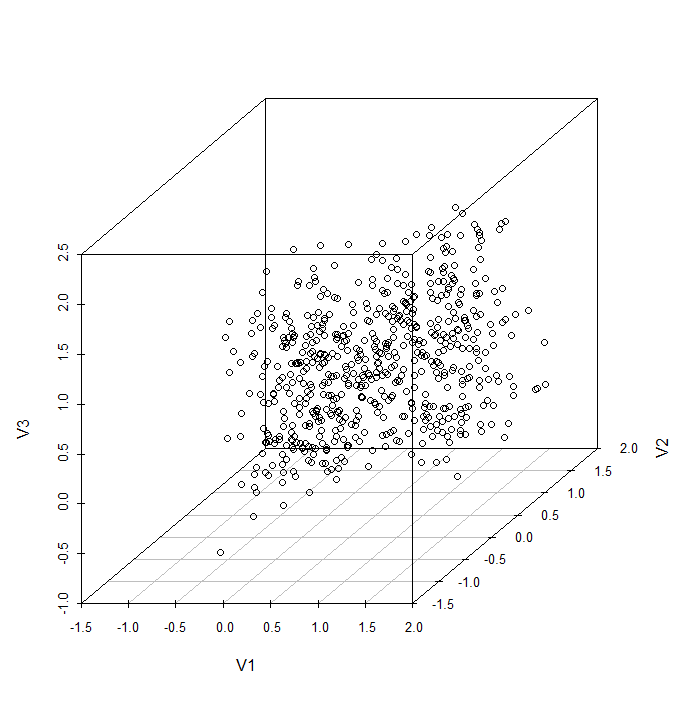
(f).



2.

(a).

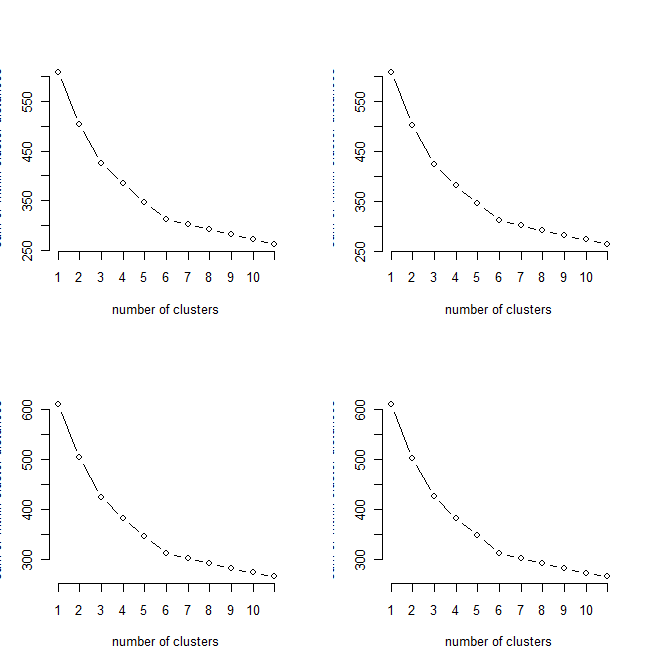
> scatterplot3d(Q2)



(b).

>pm(2,2)

> plot(stepFlexclust(Q2,2:11),type="l")



So we choose k=6.

(c).

> clQ2<-cclust(Q2,k=6,save.data=TRUE);info(clQ2,"distsum")

[1] 312.2132

> clQ2<-cclust(Q2,k=6,save.data=TRUE);info(clQ2,"distsum")

[1] 333.5442

> clQ2<-cclust(Q2,k=6,save.data=TRUE);info(clQ2,"distsum")

[1] 312.0826

> clQ2<-cclust(Q2,k=6,save.data=TRUE);info(clQ2,"distsum")

[1] 312.0386

> clQ2<-cclust(Q2,k=6,save.data=TRUE);info(clQ2,"distsum")

[1] 312.0386

> clQ2<-cclust(Q2,k=6,save.data=TRUE);info(clQ2,"distsum")

[1] 312.0386

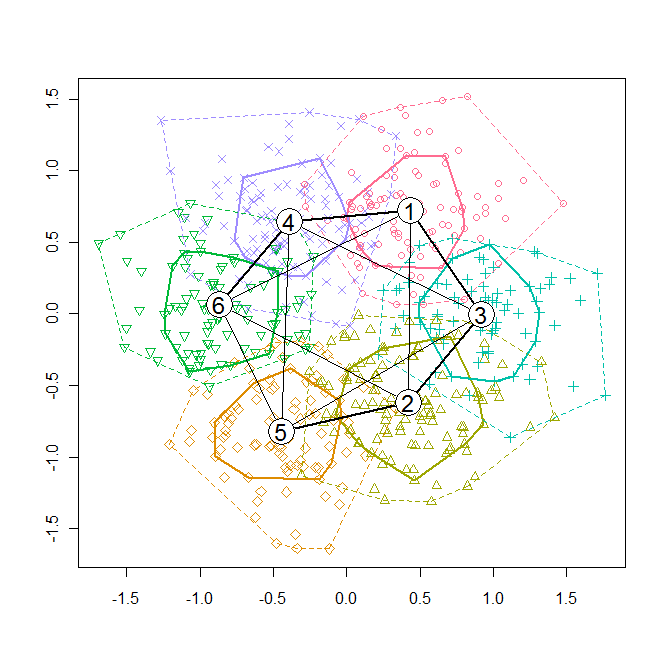
> clQ2<-cclust(Q2,k=6,save.data=TRUE);info(clQ2,"distsum")

[1] 312.1111

> clQ2<-cclust(Q2,k=6,save.data=TRUE);info(clQ2,"distsum")

[1] 312.0236

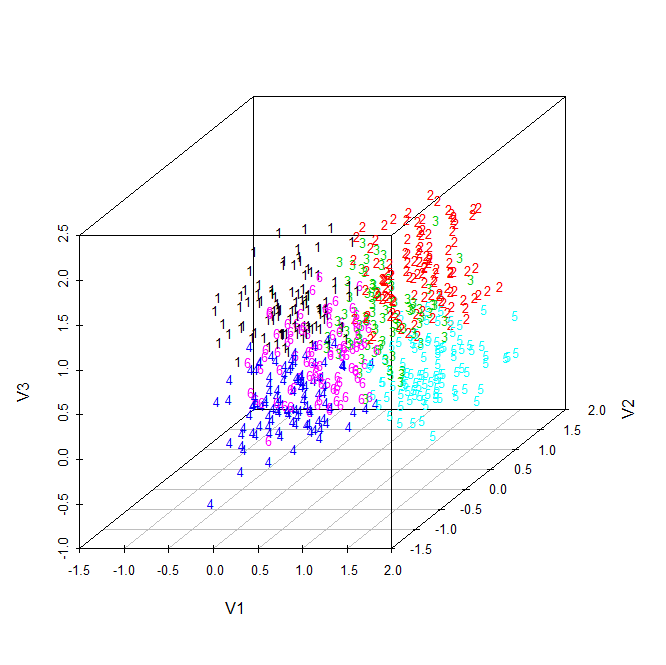
> plot(clQ2,project=prcomp(Q2))



(d).

> U<-scatterplot3d(Q2,color=clusters(clQ2),type="n")

> text(U$xyz.convert(Q2),labels=clusters(clQ2),cex=0.8,col=clusters(clQ2))



(e).

> round(parameters(clQ2))

V1 V2 V3

[1,] 0 0 1

[2,] 1 1 1

[3,] 1 0 1

[4,] 0 0 0

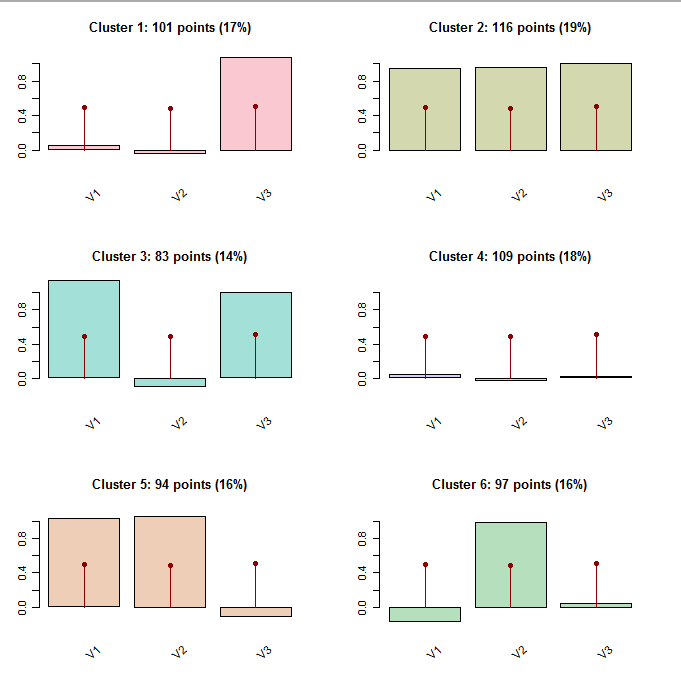
[5,] 1 1 0

[6,] 0 1 0

These are the centroids of each clusters, to which some noise arose.

(f).

> barplot(clQ2)



(g).

> round(clusterSim(clQ2),3)

[,1] [,2] [,3] [,4] [,5] [,6]

[1,] 1.000 0.062 0.238 0.338 0.000 0.024

[2,] 0.060 1.000 0.305 0.000 0.296 0.048

[3,] 0.313 0.297 1.000 0.037 0.011 0.000

[4,] 0.287 0.000 0.025 1.000 0.042 0.339

[5,] 0.000 0.339 0.009 0.061 1.000 0.222

[6,] 0.028 0.085 0.000 0.370 0.191 1.000

No, the matrix of shadow is not symmetric.

(h).

> round(parameters(clQ2),2)

V1 V2 V3

[1,] 0.05 -0.03 1.08

[2,] 0.95 0.96 1.01

[3,] 1.14 -0.09 1.00

**[4,] 0.05 -0.03 0.03**

**[5,] 1.03 1.06 -0.10**

[6,] -0.16 0.99 0.05

> getshads

function(M,clM,i,j)

{

L<-1:dim(M)[1]

DM<-dist2(M,parameters(clM))

DMo<-t(apply(DM,1,order))

DMoi<-DMo[clusters(clM)==i,]

ni<-dim(DMoi)[1]

v<-L[(DMo[,1]==i)&(DMo[,2]==j)]

dii<-DM[(DMo[,1]==i)&(DMo[,2]==j),i]

dij<-DM[(DMo[,1]==i)&(DMo[,2]==j),j]

d<-2\*dii/(dii+dij)

list(v,ni,dii,dij,2\*dii/(dii+dij),sum(d)/ni)

}

> getshads(Q2,clQ2,4,5)

[[1]]

[1] 9 49 91 211 511

[[2]]

[1] 109

[[3]]

[1] 0.8416709 0.7455965 0.6743029 0.7943657 0.9161304

[[4]]

[1] 0.8734882 0.9524135 0.9360710 0.9477534 0.9800263

[[5]]

[1] 0.9814494 0.8782004 0.8374489 0.9119534 0.9663024

[[6]]

[1] **0.04197573**

> Q2[c(9,49,91,211,511),]

V1 V2 V3

9 0.7719156 0.31513490 0.2827063

49 0.5112443 0.33091010 -0.4352284

91 0.6200470 0.22406520 -0.2266260

211 0.8031773 0.13729780 -0.1559020

511 0.9525517 0.07923577 -0.0691824

(i).

> clusterSim(clQ2)

[,1] [,2] [,3] [,4] [,5] [,6]

[1,] 1.00000000 0.06212742 0.237757262 0.33788764 0.00000000 0.0242123

[2,] 0.06035198 1.00000000 0.304703975 0.00000000 0.29587865 0.0475078

[3,] 0.31286469 0.29731752 1.000000000 0.03700704 0.01132577 0.0000000

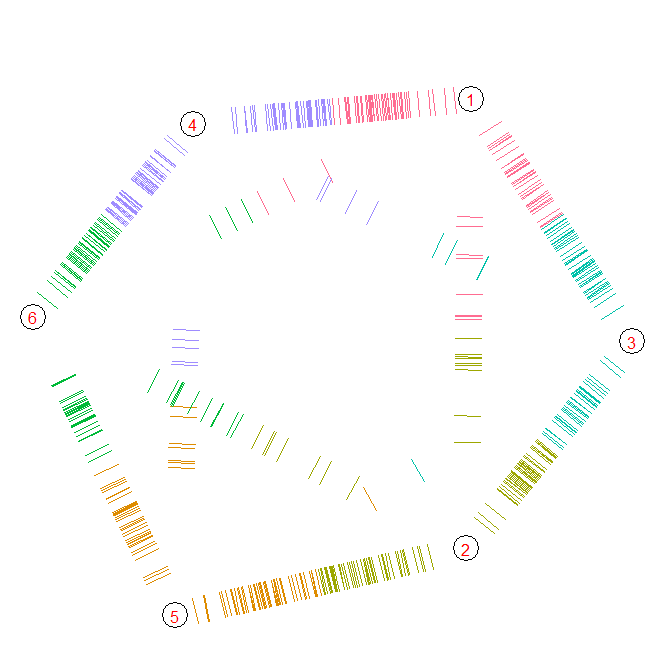
[4,] 0.28664084 0.00000000 0.024994713 1.00000000 **0.04197573** 0.3391284

[5,] 0.00000000 0.33873210 0.009059388 0.06107388 1.00000000 0.2217917

[6,] 0.02792617 0.08531799 0.000000000 0.36982306 0.19060091 1.0000000

(j).

> shadowStars(clQ2,project=prcomp(Q2))



3.

(a).

> demo("Ch-CA")

> newdata=crime[,c("Rape","Burglary","Theft")]

> newdata

Rape Burglary Theft

ME 14.8 803 2347

NH 21.5 755 2208

VT 21.8 949 2697

MA 29.7 1071 2189

RI 21.4 1294 2568

CT 23.8 1198 2758

NY 30.5 1221 2924

NJ 33.2 1071 2822

PA 25.1 735 1654

OH 38.6 988 2574

IN 25.9 887 2333

IL 32.4 1180 2938

MI 67.4 1509 3378

WI 20.1 783 2802

MN 31.8 1004 2785

IA 12.5 956 2801

MO 29.2 1136 2500

ND 11.6 385 2049

SD 17.7 554 1939

NE 24.6 748 2677

KS 32.9 1188 3008

DE 56.9 1042 3090

MD 43.6 1296 2978

DC 52.4 1728 4131

VA 26.5 813 2522

WV 18.9 625 1358

NC 26.4 1225 2423

SC 41.3 1340 2846

GA 43.9 1453 2984

FL 52.7 2221 4373

KY 23.1 824 1740

TN 47.0 1325 2126

AL 28.4 1159 2304

MS 25.8 1076 1845

AR 28.9 1030 2305

LA 40.1 1461 3417

OK 36.4 1787 3142

TX 51.6 2049 3987

MT 17.3 783 3314

ID 20.0 1003 2800

WY 21.9 817 3078

CO 42.3 1792 4231

NM 46.9 1845 3712

AZ 43.0 1908 4337

UT 25.3 915 4074

NV 64.9 1604 3489

WA 53.4 1861 4267

OR 51.1 1967 4163

CA 44.9 1696 3384

AK 72.7 1162 3910

HI 31.0 1339 3759

> newdatasc<-scale(newdata,center=FALSE)

> newdatasc

Rape Burglary Theft

ME 0.3946208 0.6224821 0.7650575

NH 0.5732668 0.5852727 0.7197473

VT 0.5812658 0.7356607 0.8791479

MA 0.7919080 0.8302346 0.7135538

RI 0.5706004 1.0031032 0.8370974

CT 0.6345930 0.9286844 0.8990322

NY 0.8132389 0.9465139 0.9531436

NJ 0.8852305 0.8302346 0.9198944

PA 0.6692556 0.5697688 0.5391585

OH 1.0292138 0.7658933 0.8390533

IN 0.6905865 0.6875986 0.7604939

IL 0.8638997 0.9147309 0.9577073

MI 1.7971247 1.1697703 1.1011352

WI 0.5359378 0.6069782 0.9133750

MN 0.8479015 0.7782965 0.9078335

IA 0.3332946 0.7410871 0.9130490

MO 0.7785763 0.8806223 0.8149313

ND 0.3092974 0.2984503 0.6679177

SD 0.4719452 0.4294584 0.6320607

NE 0.6559238 0.5798464 0.8726284

KS 0.8772315 0.9209325 0.9805253

DE 1.5171572 0.8077539 1.0072551

MD 1.1625317 1.0046536 0.9707462

DC 1.3971711 1.3395381 1.3465925

VA 0.7065846 0.6302341 0.8221027

WV 0.5039415 0.4844973 0.4426707

NC 0.7039183 0.9496147 0.7898314

SC 1.1012055 1.0387622 0.9277178

GA 1.1705307 1.1263593 0.9727020

FL 1.4051702 1.7217096 1.4254778

KY 0.6159285 0.6387612 0.5671922

TN 1.2531878 1.0271343 0.6930176

AL 0.7572454 0.8984518 0.7510407

MS 0.6879201 0.8341105 0.6014193

AR 0.7705772 0.7984515 0.7513667

LA 1.0692092 1.1325609 1.1138481

OK 0.9705540 1.3852747 1.0242057

TX 1.3758402 1.5883759 1.2996524

MT 0.4612798 0.6069782 1.0802729

ID 0.5332714 0.7775213 0.9127231

WY 0.5839322 0.6333349 1.0033434

CO 1.1278690 1.3891507 1.3791897

NM 1.2505215 1.4302360 1.2100100

AZ 1.1465335 1.4790733 1.4137428

UT 0.6745883 0.7093040 1.3280120

NV 1.7304657 1.2434139 1.1373181

WA 1.4238347 1.4426392 1.3909247

OR 1.3625085 1.5248099 1.3570236

CA 1.1971943 1.3147319 1.1030910

AK 1.9384416 0.9007774 1.2745525

HI 0.8265707 1.0379870 1.2253307

attr(,"scaled:scale")

Rape Burglary Theft

37.50435 1289.99688 3067.74327

(b).

> mahalanobis(newdatasc,center=colMeans(newdatasc,),cov=cov(newdatasc))

ME NH VT MA RI CT NY

1.7862522 1.2084534 0.8448333 1.4298655 2.7923225 1.0465452 0.1633077

NJ PA OH IN IL MI WI

0.1686348 3.1781616 1.3232661 0.6907968 0.0251870 7.1760252 2.1141753

MN IA MO ND SD NE KS

0.3428616 3.2171761 0.5530947 4.1393552 2.4386139 1.7501073 0.0518216

DE MD DC VA WV NC SC

7.2774854 0.6479206 2.5486974 0.9640740 4.4540855 1.6090542 0.6605497

GA FL KY TN AL MS AR

0.9305026 6.2653831 2.6977640 5.9101276 1.4191595 3.5801965 0.8135673

LA OK TX MT ID WY CO

0.4343021 4.3931454 4.1430358 5.8199353 1.2666434 3.0095181 3.5801966

NM AZ UT NV WA OR CA

2.4179725 4.5915068 10.7798682 5.2872865 3.1498402 3.4397882 1.5731053

AK HI

17.3237541 2.5706772

(c).

> pairsbg

function(M,bg=TRUE){

if (is.null(rownames(M))){

rownames(M)<-1:(dim(M)[1])

}

pairs(M,panel=function(x,y,...){

text(x,y,abbreviate(row.names(M)),cex=0.6)

if (bg){

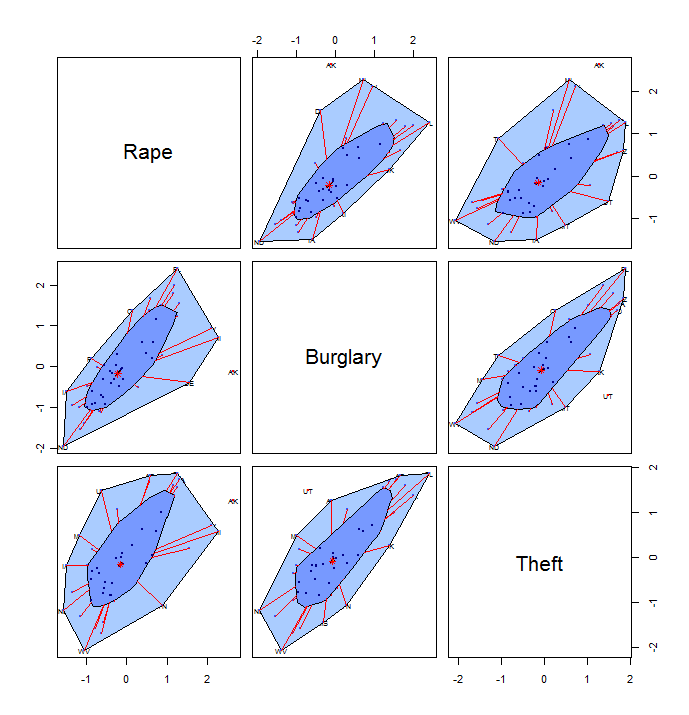
bagplot(cbind(x,y),add=TRUE)

}

})

}

> pairsbg(newdatasc)



(d).

> dim(newdata)

[1] 51 3

> Mcr<-newdata[(rownames(newdata)!="UT"),]

> dim(Mcr)

[1] 50 3

> Mcr<-Mcr[(rownames(Mcr)!="AK"),]

> dim(Mcr)

[1] 49 3

> MQ4<-Mcr

> MQ4s<-matrix(scale(MQ4,scale=TRUE,center=TRUE),ncol=3)

> MQ4s

[,1] [,2] [,3]

[1,] -1.37115819 -0.95946793 -0.73843669

[2,] -0.88287179 -1.07153371 -0.92434924

[3,] -0.86100822 -0.61860119 -0.27031156

[4,] -0.28526754 -0.33376734 -0.94976175

[5,] -0.89015965 0.18687159 -0.44284911

[6,] -0.71525108 -0.03725997 -0.18872403

[7,] -0.22696468 0.01643822 0.03330103

[8,] -0.03019255 -0.33376734 -0.10312401

[9,] -0.62050895 -1.11822778 -1.66532445

[10,] 0.36335171 -0.52754774 -0.43482410

[11,] -0.56220609 -0.76335282 -0.75716169

[12,] -0.08849540 -0.07928463 0.05202604

[13,] 2.46225446 0.68883289 0.64052620

[14,] -0.98490178 -1.00616200 -0.12987402

[15,] -0.13222254 -0.49019248 -0.15261152

[16,] -1.53877890 -0.60225826 -0.13121152

[17,] -0.32170682 -0.18201160 -0.53379913

[18,] -1.60436961 -1.93537408 -1.13701180

[19,] -1.15981035 -1.54080915 -1.28413684

[20,] -0.65694823 -1.08787663 -0.29706156

[21,] -0.05205612 -0.06060700 0.14565106

[22,] 1.69702950 -0.40147374 0.25532609

[23,] 0.72774455 0.19154100 0.10552605

[24,] 1.36907595 1.20013300 1.64766399

[25,] -0.51847895 -0.93612089 -0.50437412

[26,] -1.07235606 -1.37504519 -2.06122457

[27,] -0.52576681 0.02577703 -0.63678666

[28,] 0.56012385 0.29426796 -0.07102400

[29,] 0.74960812 0.55808948 0.11355105

[30,] 1.39093952 2.35114193 1.97133908

[31,] -0.76626608 -0.91043915 -1.55029942

[32,] 0.97553168 0.25924741 -1.03402427

[33,] -0.38000967 -0.12831341 -0.79594921

[34,] -0.56949395 -0.32209382 -1.40986188

[35,] -0.34357039 -0.42949019 -0.79461171

[36,] 0.47266957 0.57676711 0.69268872

[37,] 0.20301887 1.33788052 0.32487611

[38,] 1.31077309 1.94957289 1.45506393

[39,] -1.18896177 -1.00616200 0.55492618

[40,] -0.99218964 -0.49252719 -0.13254902

[41,] -0.85372036 -0.92678208 0.23927609

[42,] 0.63300241 1.34955404 1.78141403

[43,] 0.96824382 1.47329333 1.08725133

[44,] 0.68401741 1.62037967 1.92318907

[45,] 2.28005804 0.91062974 0.78898874

[46,] 1.44195451 1.51064859 1.82956404

[47,] 1.27433381 1.75812719 1.69046400

[48,] 0.82248669 1.12542248 0.64855120

[49,] -0.19052540 0.29193326 1.15011385

(e).

> prcomp(MQ4s)

Standard deviations:

[1] 1.5856241 0.5861205 0.3771721

Rotation:

PC1 PC2 PC3

[1,] 0.5594205 0.75731280 -0.3369362

[2,] 0.6007711 -0.09039511 0.7942939

[3,] 0.5710716 -0.64676584 -0.5055405

(f).

> rownames(MQ4s)<-rownames(MQ4)

> prMQ4s<-prcomp(MQ4s)

> kmeans(MQ4s,centers=2)$cluster

ME NH VT MA RI CT NY NJ PA OH IN IL MI WI MN IA MO ND SD NE KS DE MD DC VA WV NC SC

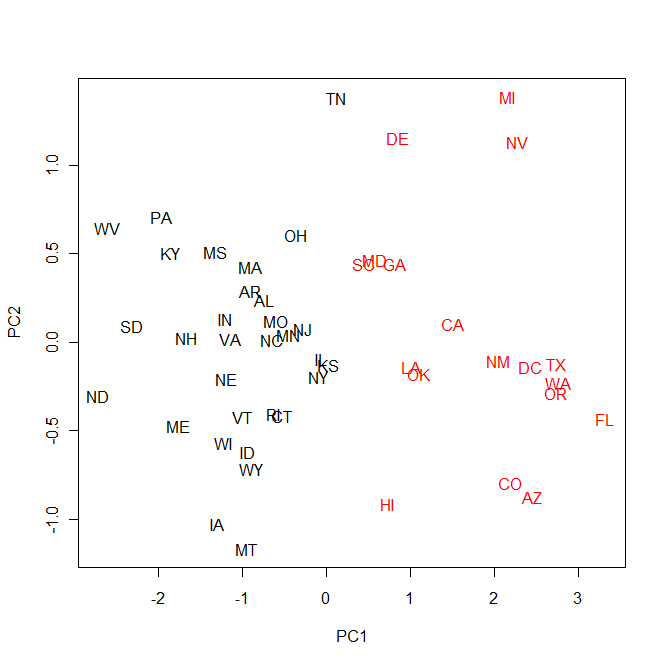
1 1 1 1 1 1 1 1 1 1 1 1 2 1 1 1 1 1 1 1 1 2 2 2 1 1 1 2

GA FL KY TN AL MS AR LA OK TX MT ID WY CO NM AZ NV WA OR CA HI

2 2 1 1 1 1 1 2 2 2 1 1 1 2 2 2 2 2 2 2 2

> plot(prMQ4s$x[,1:2],type="n")

> text(prMQ4s$x[,1:2],labels=rownames(MQ4s),col=kmeans(MQ4s,center=2)$cluster)

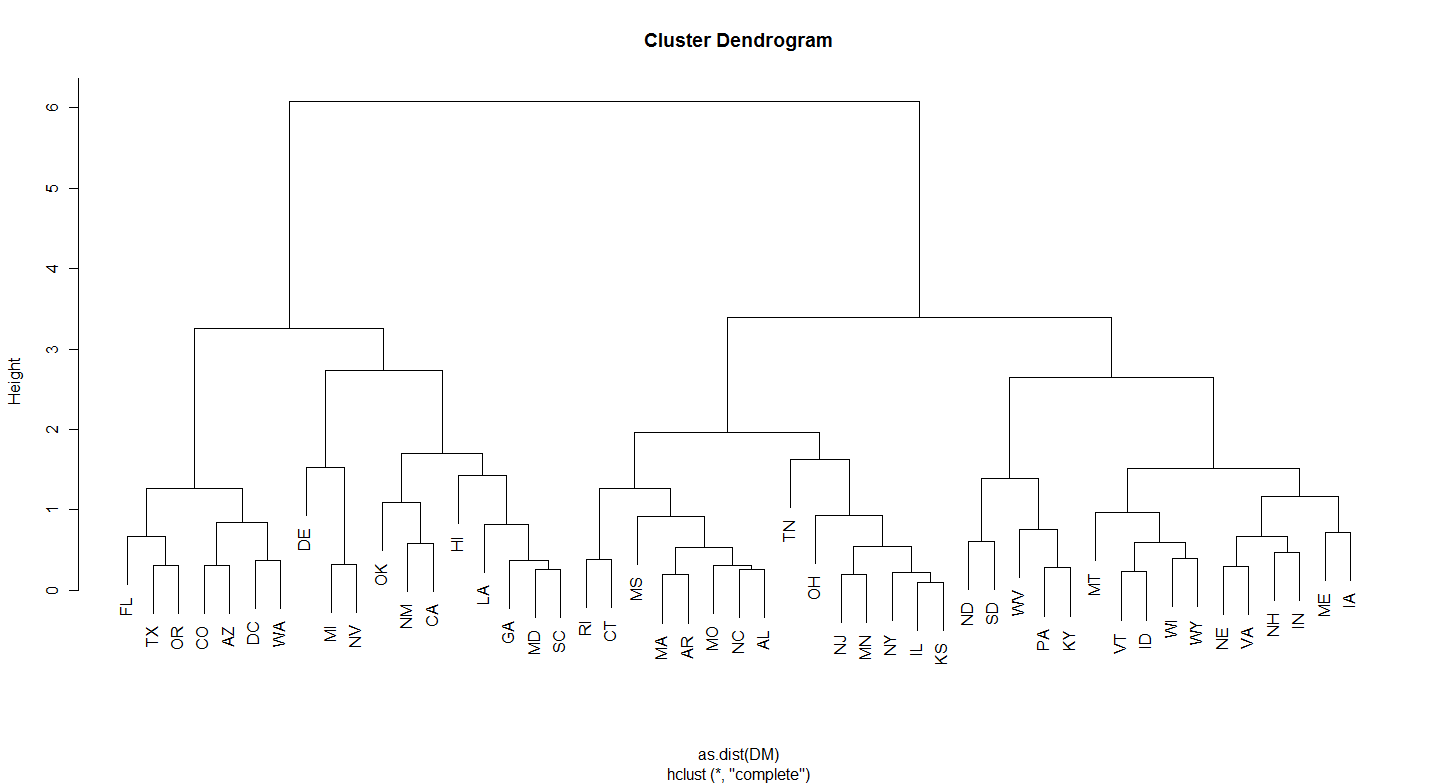


(g).

> DM<-as.matrix(dist(MQ4s))

> DMc<-hclust(as.dist(DM),method="complete")

> plot(DMc)



> cl<-cutree(DMc,4)

> cl

ME NH VT MA RI CT NY NJ PA OH IN IL MI WI MN IA MO ND SD NE KS DE MD DC VA WV NC

1 1 1 2 2 2 2 2 1 2 1 2 3 1 2 1 2 1 1 1 2 3 3 4 1 1 2

SC GA FL KY TN AL MS AR LA OK TX MT ID WY CO NM AZ NV WA OR CA HI

3 3 4 1 2 2 2 2 3 3 4 1 1 1 4 3 4 3 4 4 3 3

> sumcldist

function (M,cl)

{

mv<-NULL

Dt<-0

t<-unique(cl)

for (i in t)

{

U<-cbind(M[cl==i,])

m<-colMeans(U)

mv<-rbind(mv,m)

Dm<-dist2(U,m)

print(Dm)

Dt<-Dt+sum(Dm)

4}

rownames(mv)<-t

list(mv[order(t),],Dt)

}

> sumcldist(MQ4s,cl)

[,1]

ME 0.4040385

NH 0.2717827

VT 0.5825641

PA 1.0576214

IN 0.4958889

WI 0.5447293

IA 0.8868493

ND 1.2007803

SD 0.8212689

NE 0.4992082

VA 0.4966188

WV 1.4343161

KY 0.9078785

MT 1.2474597

ID 0.7576279

WY 0.9267938

[,1]

MA 0.5021327

RI 0.7613763

CT 0.5939096

NY 0.5465920

NJ 0.4558588

OH 0.6855220

IL 0.5560452

MN 0.4727243

MO 0.1199329

KS 0.6573698

NC 0.3946270

TN 1.3777799

AL 0.3554542

MS 1.0051132

AR 0.4300891

[,1]

MI 1.4902784

DE 1.2940502

MD 0.6611460

SC 0.8032888

GA 0.4745243

LA 0.5369023

OK 1.0604874

NM 1.0067091

NV 1.3568533

CA 0.5245773

HI 1.3715333

[,1]

DC 0.5330018

FL 0.7447867

TX 0.4345281

CO 0.6190381

AZ 0.5052226

WA 0.3372769

OR 0.1568021

[[1]]

[,1] [,2] [,3]

1 -0.9771584 -1.0219213 -0.6743203

2 -0.2148183 -0.1624001 -0.4830633

3 0.9775193 0.6406531 0.5214887

4 1.1577281 1.6770796 1.7569569

[[2]]

[1] 35.36096

(h).

> clMQ4s<-cclust(MQ4s,k=4,save.data=TRUE);info(clMQ4s,"distsum")

[1] 35.54995

> clMQ4s<-cclust(MQ4s,k=4,save.data=TRUE);info(clMQ4s,"distsum")

[1] 38.30754

> clMQ4s<-cclust(MQ4s,k=4,save.data=TRUE);info(clMQ4s,"distsum")

[1] 35.75229

> clMQ4s<-cclust(MQ4s,k=4,save.data=TRUE);info(clMQ4s,"distsum")

[1] 34.7572

> clMQ4s<-cclust(MQ4s,k=4,save.data=TRUE);info(clMQ4s,"distsum")

[1] 35.69497

> clMQ4s<-cclust(MQ4s,k=4,save.data=TRUE);info(clMQ4s,"distsum")

[1] 35.69497

> clMQ4s<-cclust(MQ4s,k=4,save.data=TRUE);info(clMQ4s,"distsum")

[1] 38.08004

> clMQ4s<-cclust(MQ4s,k=4,save.data=TRUE);info(clMQ4s,"distsum")

[1] 34.6263

> clMQ4s<-cclust(MQ4s,k=4,save.data=TRUE);info(clMQ4s,"distsum")

[1] 37.49781

> clMQ4s<-cclust(MQ4s,k=4,save.data=TRUE);info(clMQ4s,"distsum")

[1] 35.54995

> clMQ4s<-cclust(MQ4s,k=4,save.data=TRUE);info(clMQ4s,"distsum")

[1] 34.45261

> clMQ4s<-cclust(MQ4s,k=4,save.data=TRUE);info(clMQ4s,"distsum")

[1] 34.57281

> clMQ4s<-cclust(MQ4s,k=4,save.data=TRUE);info(clMQ4s,"distsum")

[1] 34.6263

> clMQ4s<-cclust(MQ4s,k=4,save.data=TRUE);info(clMQ4s,"distsum")

[1] 34.28704

> round(clusterSim(clMQ4s),3)

[,1] [,2] [,3] [,4]

[1,] 1.000 **0.000 0.000** 0.613

[2,] **0.000** 1.000 0.316 0.299

[3,] **0.000** 0.541 1.000 **0.000**

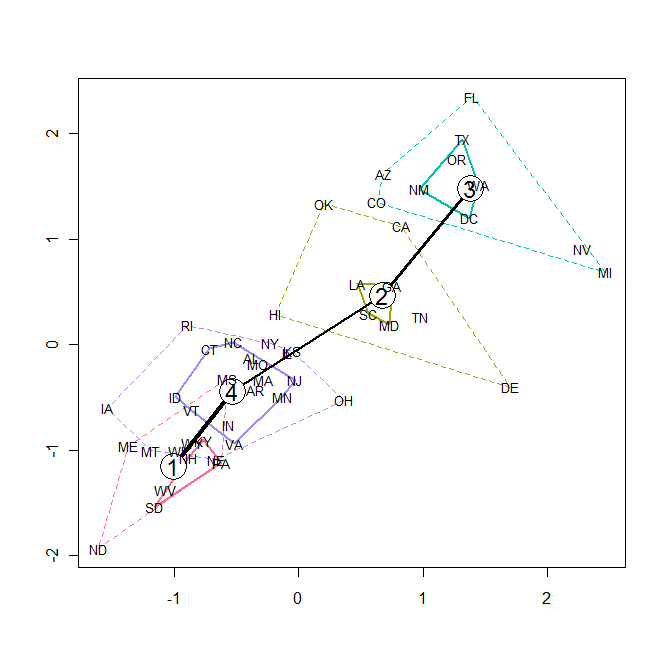
[4,] 0.432 0.246 **0.000** 1.000

(i).

**Without projecting:**

> plot(clMQ4s,type="n")

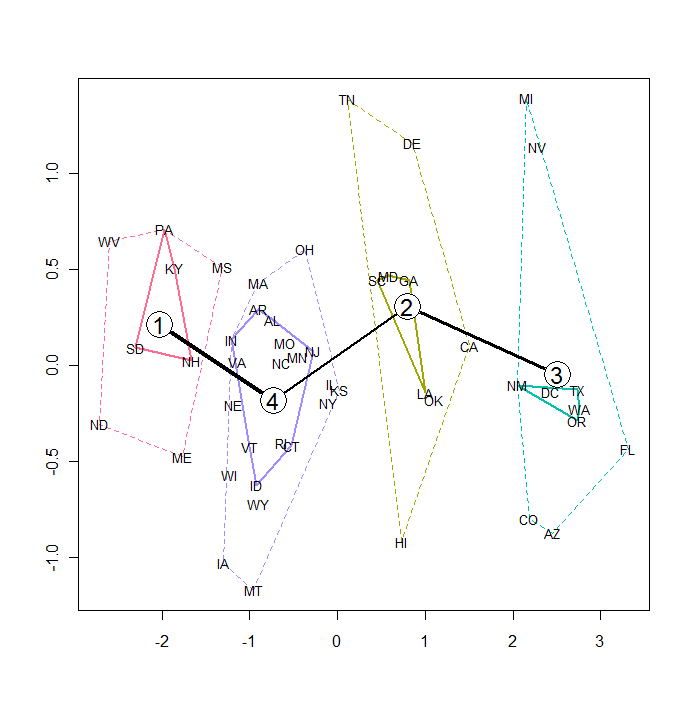
> text(MQ4s,labels=rownames(MQ4s),cex=0.8)



**With projecting:**

> plot(clMQ4s,project=prcomp(MQ4s),type="n")

> text(prcomp(MQ4s)$x,labels=rownames(MQ4s),cex=0.8)



(j).

> clMQ4s\_2<-cclust(MQ4s,k=4,save.data=TRUE);info(clMQ4s\_2,"distsum")

[1] 34.28704

> clMQ4s\_2<-cclust(MQ4s,k=4,save.data=TRUE);info(clMQ4s\_2,"distsum")

[1] 35.54995

> clMQ4s\_2<-cclust(MQ4s,k=4,save.data=TRUE);info(clMQ4s\_2,"distsum")

[1] 39.34811

> clMQ4s\_2<-cclust(MQ4s,k=4,save.data=TRUE);info(clMQ4s\_2,"distsum")

[1] 37.06694

> clMQ4s\_2<-cclust(MQ4s,k=4,save.data=TRUE);info(clMQ4s\_2,"distsum")

[1] 35.75229

> clMQ4s\_2<-cclust(MQ4s,k=4,save.data=TRUE);info(clMQ4s\_2,"distsum")

[1] 36.89089

> clMQ4s\_2<-cclust(MQ4s,k=4,save.data=TRUE);info(clMQ4s\_2,"distsum")

[1] 35.54995

> clMQ4s\_2<-cclust(MQ4s,k=4,save.data=TRUE);info(clMQ4s\_2,"distsum")

[1] 38.27761

> clMQ4s\_2<-cclust(MQ4s,k=4,save.data=TRUE);info(clMQ4s\_2,"distsum")

[1] 38.1727

> clMQ4s\_2<-cclust(MQ4s,k=4,save.data=TRUE);info(clMQ4s\_2,"distsum")

[1] 39.53837

> round(clusterSim(clMQ4s\_2),3)

[,1] [,2] [,3] [,4]

[1,] 1.000 0.222 0.355 0.092

[2,] 0.000 1.000 0.646 0.000

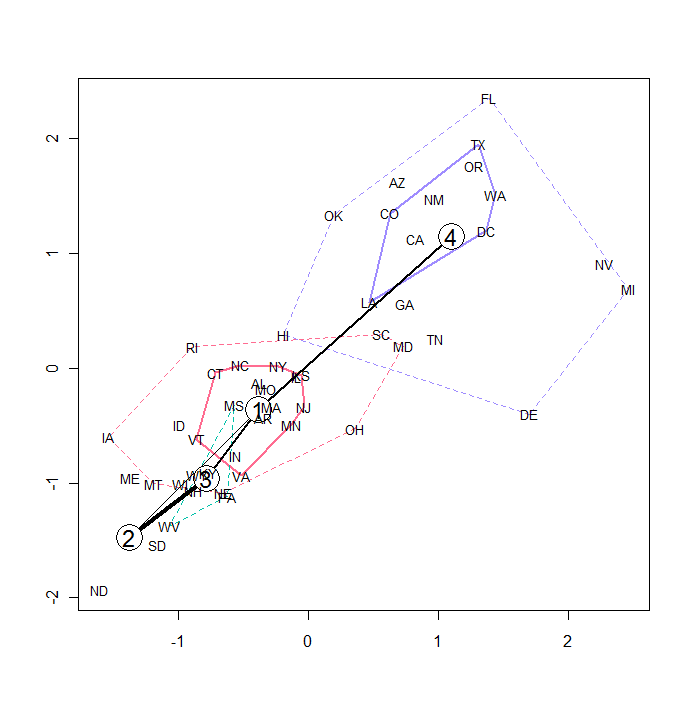
[3,] 0.150 0.463 1.000 0.000

[4,] 0.578 0.000 0.000 1.000

**Without Projecting:**

> plot(clMQ4s\_2,type="n")

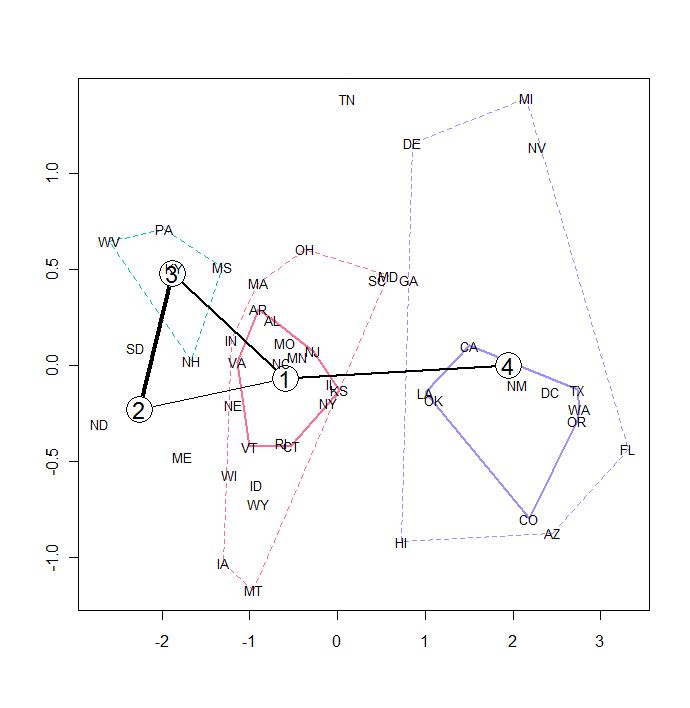
> text(MQ4s,labels=rownames(MQ4s),cex=0.8)



**With Projecting:**

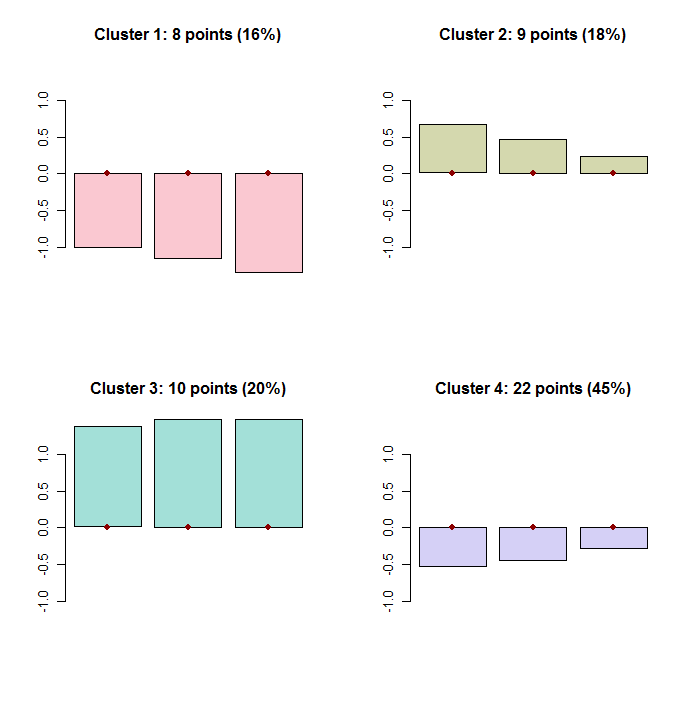
> plot(clMQ4s\_2,project=prcomp(MQ4s),type="n")

> text(prcomp(MQ4s)$x,labels=rownames(MQ4s),cex=0.8)



(k).

> barplot(clMQ4s)



**The centroids:**

> round(parameters(clMQ4s),2)

[,1] [,2] [,3]

[1,] -1.01 -1.15 -1.35

[2,] 0.67 0.47 0.24

[3,] 1.38 1.48 1.48

[4,] -0.54 -0.45 -0.28

(l).

> MQ4s[rownames(MQ4s)=="NC",]

[1] -0.52576681 0.02577703 -0.63678666

> dist2(rbind(MQ4s[rownames(MQ4s)=="NC",]),parameters(clMQ4s))

[,1] [,2] [,3] [,4]

[1,] 1.458116 1.548558 3.200503 0.5896936

stripes(clMQ4s,type="all")

