

Problem 1 and 2:

```
myCCA=function(mat,x,y=-x,num,size){
  if(num!=dim(mat)[2]) {print("wrong input")}
  else{
    leny=ifelse(y==x,num-length(x),length(y))
    r11=mat[x,x];r22=mat[y,y];r12=mat[x,y];r21=mat[y,x]
    E1=solve(r11) %*% r12 %*% solve(r22) %*%r21
    E2=solve(r22) %*% r21 %*% solve(r11) %*%r12
    e1=eigen(E1);e2=eigen(E2);eigseq=1:min(length(x),leny)
    corr=sqrt(e1$values)[eigseq]
    test=-(size-0.5*(length(x)+leny+1))*sum(log(1-e1$values[eigseq]))
    barplot(corr,xlab="dimension",ylab="CC")
    result=ifelse(pchisq(test,length(x)*leny),"reject Ho","not reject Ho")
    p.val=pchisq(test,length(x)*leny)
    return(list(e1=E1,e2=E2,eigenx=e1,eigeny=e2,correlation=corr,pval=p.val[1],test=result[1]))
  }
}
}#test:myCCA(mat=R,x=c(1,2),num=4,size=25)
#test:myCCA(mat=R,x=c(1,2),y=3,num=4,size=25)
```

Input the headsize data, we get:

```
> myCCA(mat=R,x=c(1,2),num=4,size=25)
$e1
      head1 breadth1
head1  0.3225003 0.3168319
breadth1 0.3018705 0.3021324

$e2
      head2 breadth2
head2  0.3013980 0.3002082
breadth2 0.3185347 0.3232347

$eigenx
$eigenx$values
[1] 0.621744734 0.002887956

$eigenx$vectors
      [,1] [,2]
[1,] 0.7269968 -0.7040109
[2,] 0.6866408 0.7101892
```

```
$eigeny
$eigeny$values
[1] 0.621744734 0.002887956
```

```
$eigeny$vectors
      [,1]      [,2]
[1,] -0.6837994 -0.7091095
[2,] -0.7296700  0.7050984
```

```
$correlation
[1] 0.7885079 0.0537397
```

```
$pval
[1] 0.9997939
```

```
$test
[1] "reject Ho"
```

### Problem 3

(a)

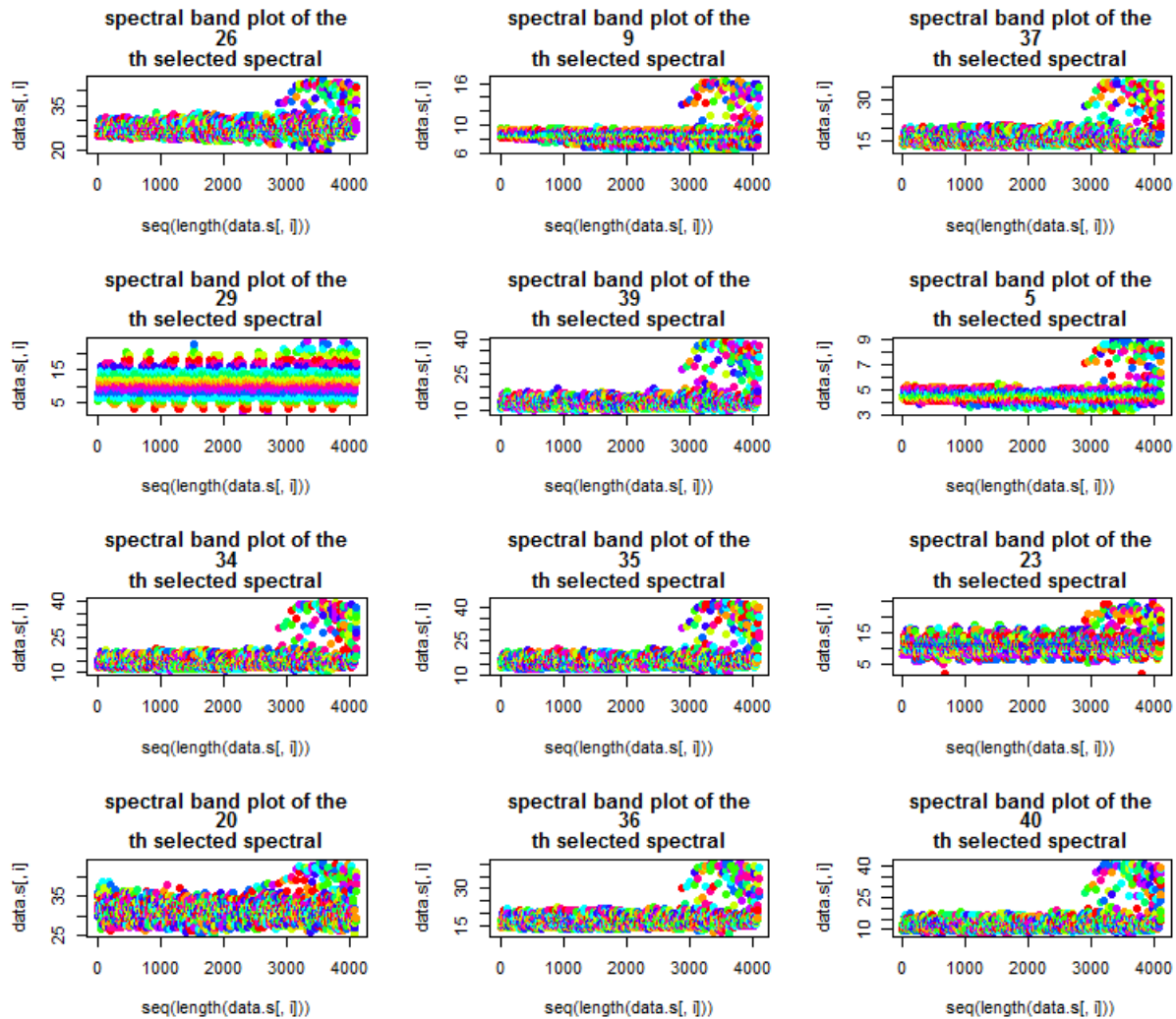
```
data=dat.mat
set.seed(110)
index=sample(1:42,12,replace=F)
set.seed(110)
data.s=data[,sample(1:42,12,replace=F)]
tocontrast=function(x){
  new=(x-min(x)+1)/(max(x)-min(x)+1)
  return(new)
}
data.s.1=tocontrast(data.s)
```

Using the initial values as index of color (10 color base):

```
palette(rainbow(10))
par(mfrow=c(4,3))
for (i in 1:12){
  ith=index[i]
  name=c("spectral band plot of the",ith,"th selected spectral")
```

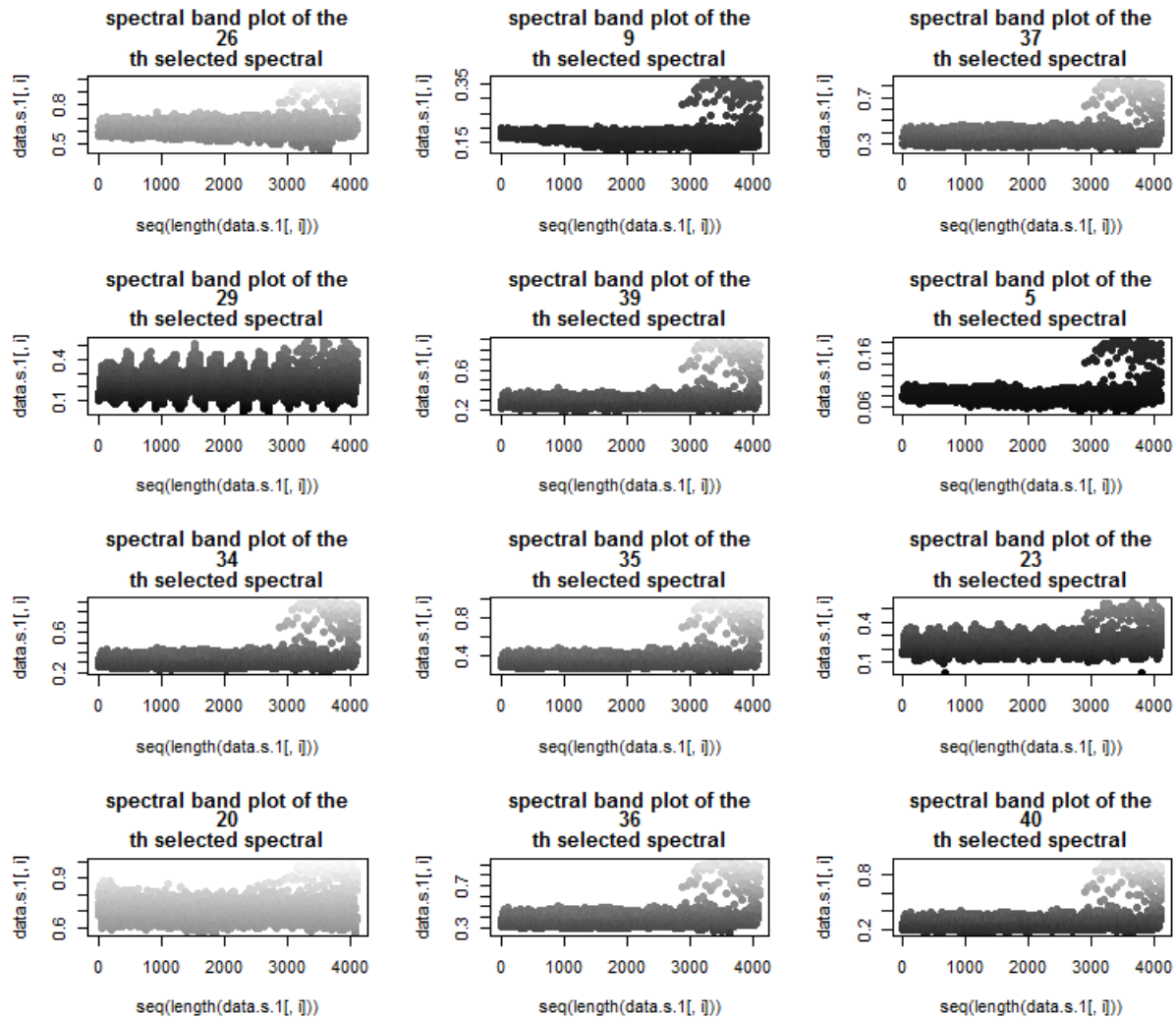
```
plot(data.s.1[,i]~seq(length(data.s.1[,i])),main=name,pch=16,col=data.s.1[,i]*100)
```

```
}
```



Converting the contrast colors (black and white):

```
palette(gray(seq(0,1,0.01)))
par(mfrow=c(4,3))
for (i in 1:12){
  ith=index[i]
  name=c("spectral band plot of the",ith,"th selected spectral")
  plot(data.s.1[,i]~seq(length(data.s.1[,i])),main=name,pch=16,col=data.s.1[,i]*100)
}
```

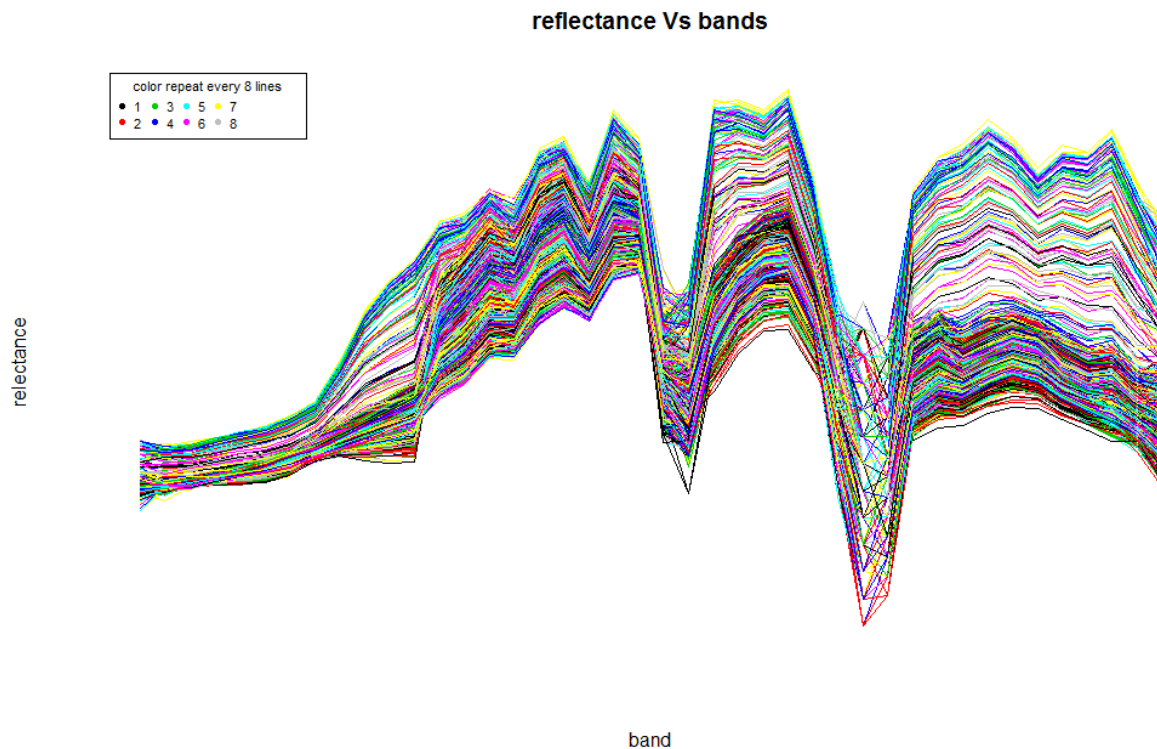


(2)

```
plot(1,ylim=c(0,1),xlim=c(1,42),xlab="band",ylab="relectance",main="reflectance Vs bands",axes=F)
col.index=1

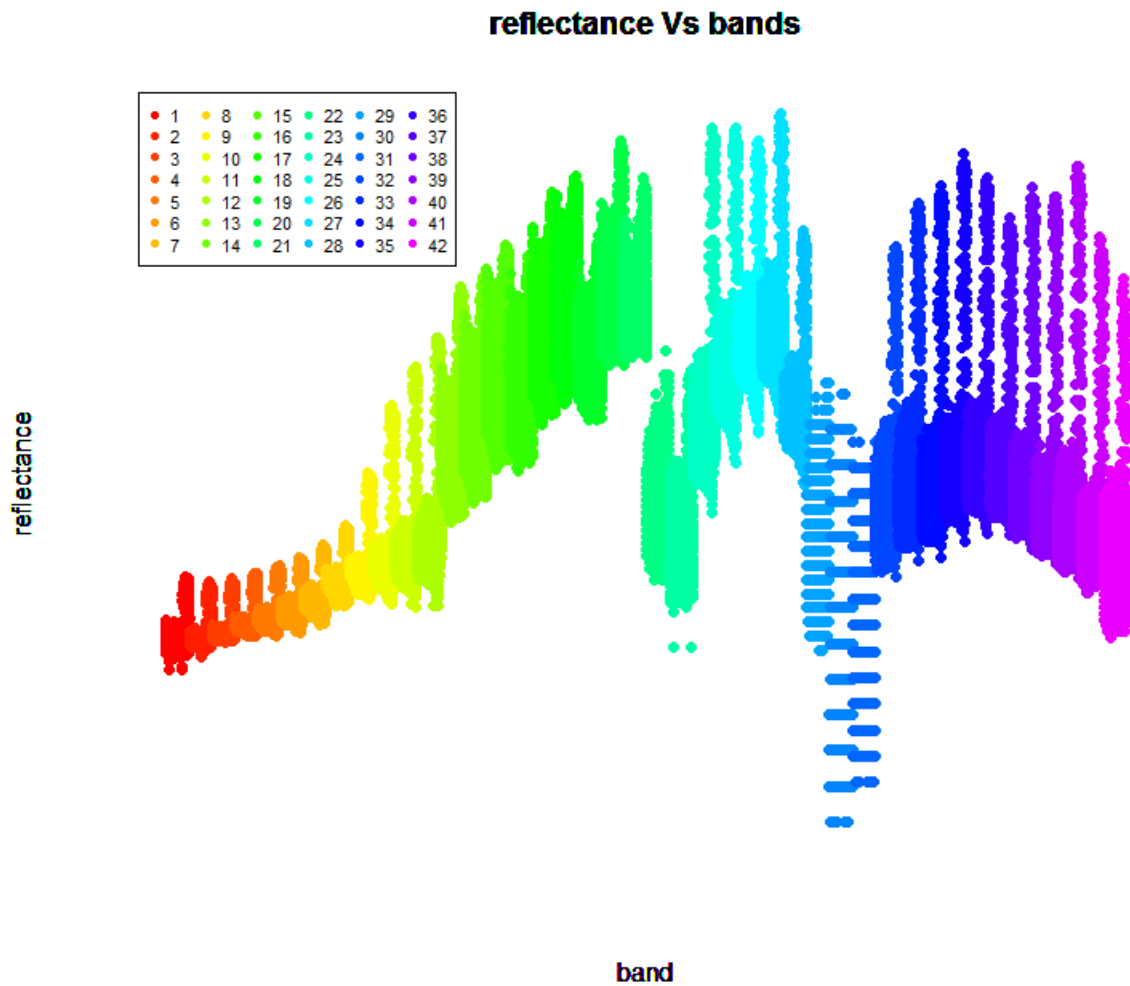
for (i in 1:4096){
  lines(seq(1:42),data.a[i,],col=col.index,type="l")
  col.index=col.index+1
  if(col.index==33){col.index=1}
}
```

```
legend("topleft",col=seq(1,8,1),pch=16,legend=seq(1,8,1),ncol=6,inset=0.01,cex=0.7,title="color repeat  
every 8 lines")
```



When we focus on the reflectance variation, we can find that their tendencies are very similar.

```
palette(rainbow(42))
for(i in 1:42){
  plot(data.a[,i]~seq(i*10-10,i*10,length.out=length(data.a[,i])),axes=F,pch=16,col=i,
    ,xlab="band",ylab="reflectance",main="reflectance Vs bands",xlim=c(0,420),ylim=c(0,1))
  par(new=T)
}
legend("topleft",col=seq(1,42,1),pch=16,legend=seq(1,42,1),ncol=6,inset=0.01,cex=0.7)
```



```
> apply(data.a,2,var)*10^3
col1 col2 col3 col4 col5 col6 col7 col8
0.151 0.095 0.079 0.080 0.097 0.111 0.127 0.174
col9 col10 col11 col12 col13 col14 col15 col16
0.337 0.749 1.110 1.467 2.185 2.913 2.965 2.379
col17 col18 col19 col20 col21 col22 col23 col24
2.564 2.827 1.256 1.821 1.198 1.975 2.163 3.687
col25 col26 col27 col28 col29 col30 col31 col32
2.568 1.937 2.169 1.504 4.080 15.195 9.675 3.827
col33 col34 col35 col36 col37 col38 col39 col40
4.050 3.661 3.790 3.235 2.669 3.268 3.460 4.462
col41 col42
3.275 3.557
```

Combine the plot and sample variance, We can find from the plot that 1:9 are less variable, 9:28 become more variable, and 29:42(especially 30:31) become the most variable group among these 42 bands.

(3)

```

library(corrplot)

r=cor(data)

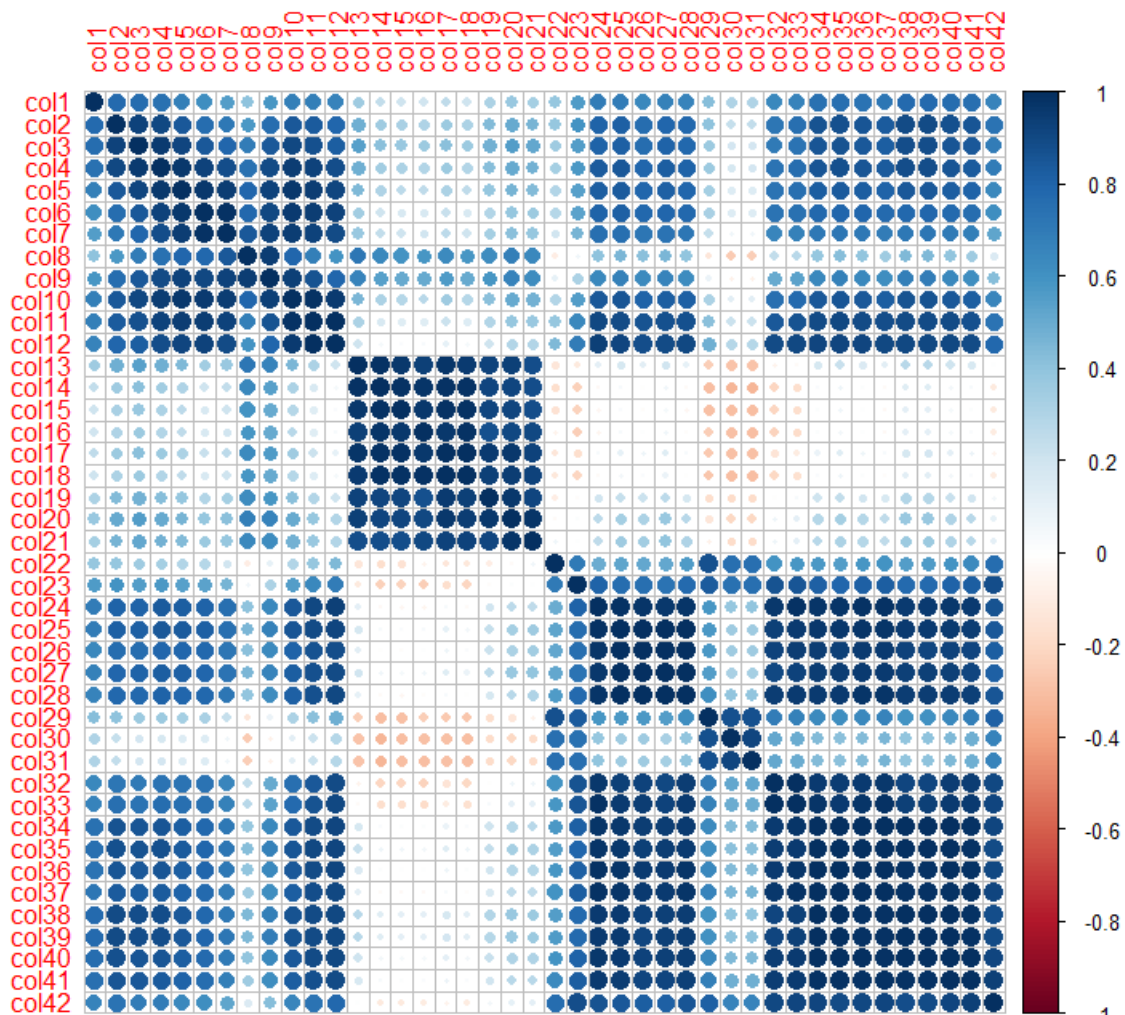
corrplot(r,method="circle") ###1

library(lattice)

levelplot(r) ###2

corrplot.mixed(r,lower="ellipse",upper="number",order="FPC")###3

```

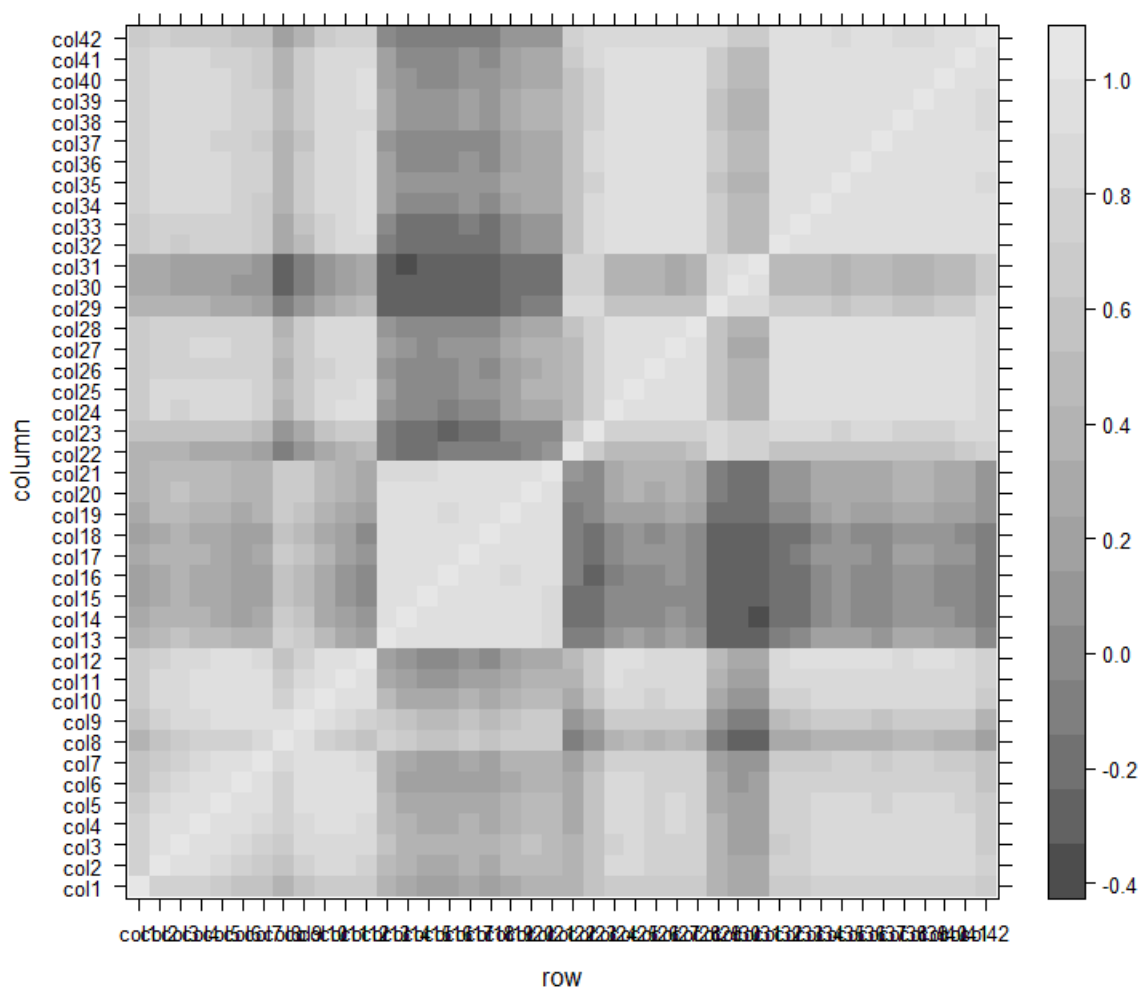


We can see:

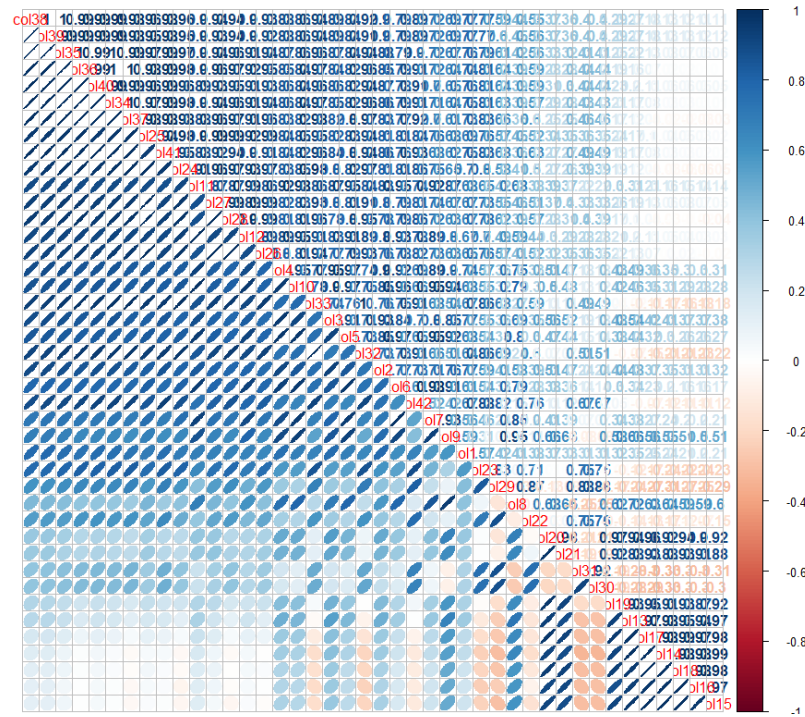
The 22-42 bands are more correlated to each other.

We can separate 42 bands into one group consist of 1:21 bands and the other group consist of 22:42.

Or we can separate 42 bands into two distinct groups, which contains 1:12 and 12:42 respectively.



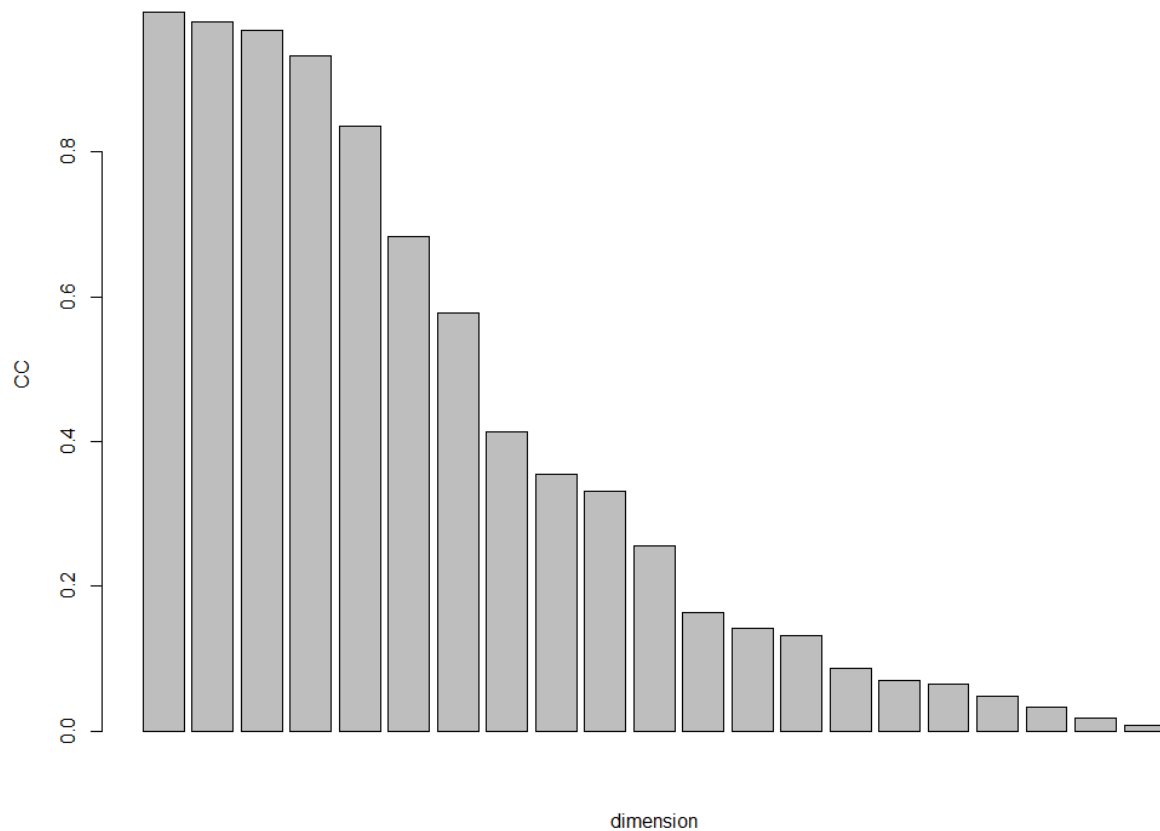




(4)

For the first group(the first 21 bands):

```
myCCA(mat=r,x=seq(1,21),num=42,size=4096)
myCCA(mat=r,x=seq(1,12),num=42,size=4096)
res.cc=cc(data.a[,1:12],data.a[,13:42])
plt.cc(res.cc)
res.cc2=cc(data.a[,1:21],data.a[,22:42])
plt.cc(res.cc2)
```

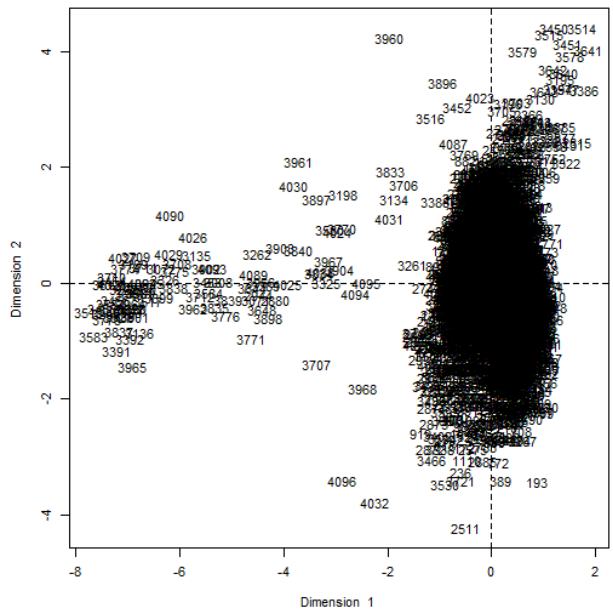


The correlation of each dimension:

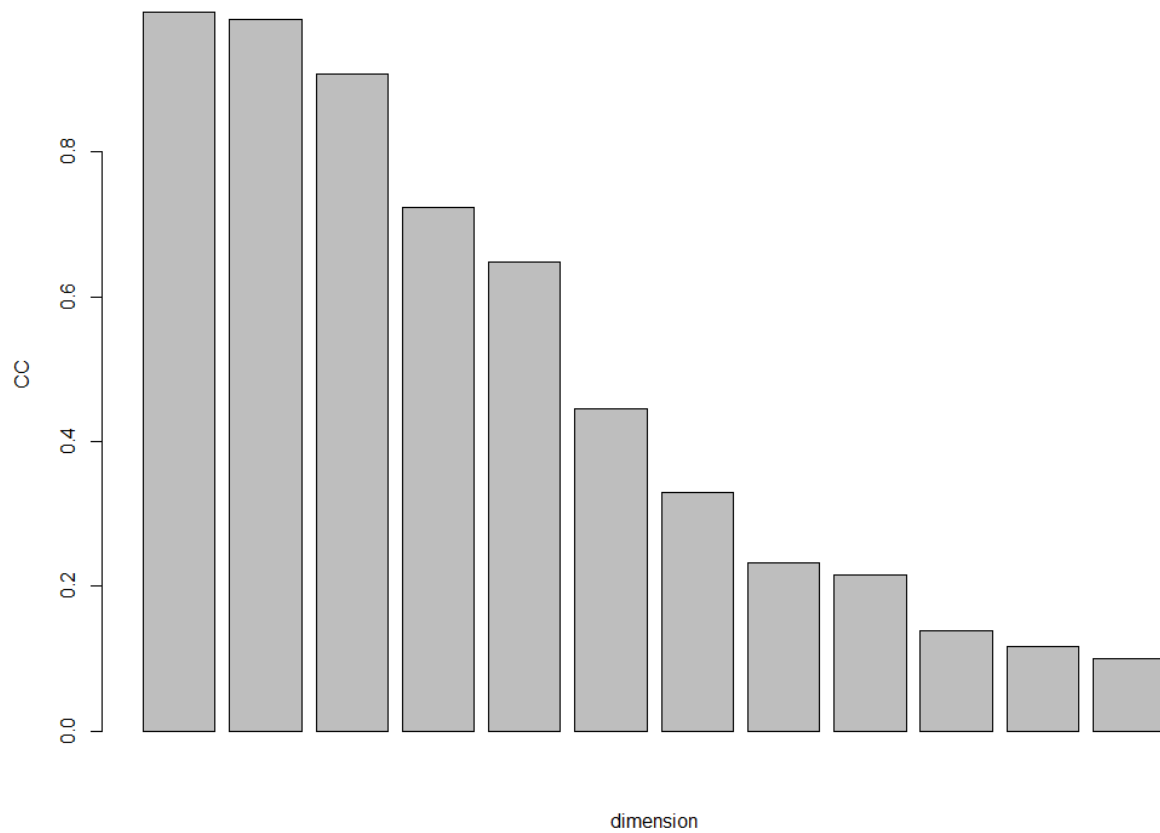
\$correlation

```
[1] 0.9932 0.9798 0.9679 0.9332 0.8367 0.6835 0.5783 0.4136 0.3554 0.3311 0.1645 0.1422
[14] 0.1317 0.0872 0.0699 0.0652 0.0480 0.0326 0.0175 0.007
```

And the hypothesis has the P-value near to 1, which means we should reject the null hypothesis that all correlations are zero.



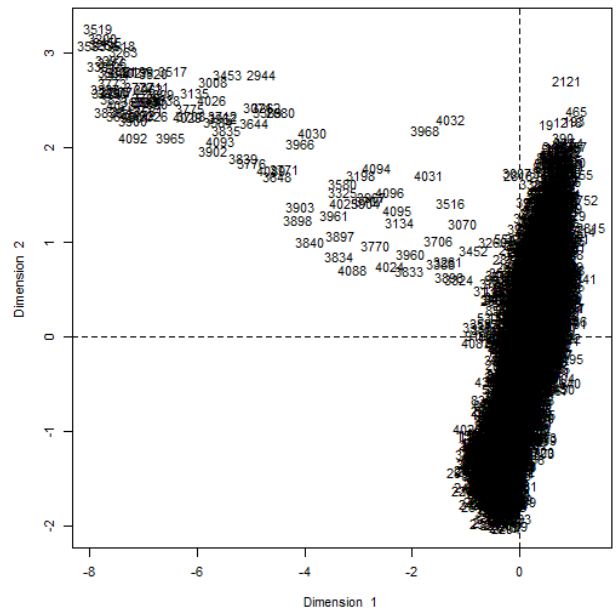
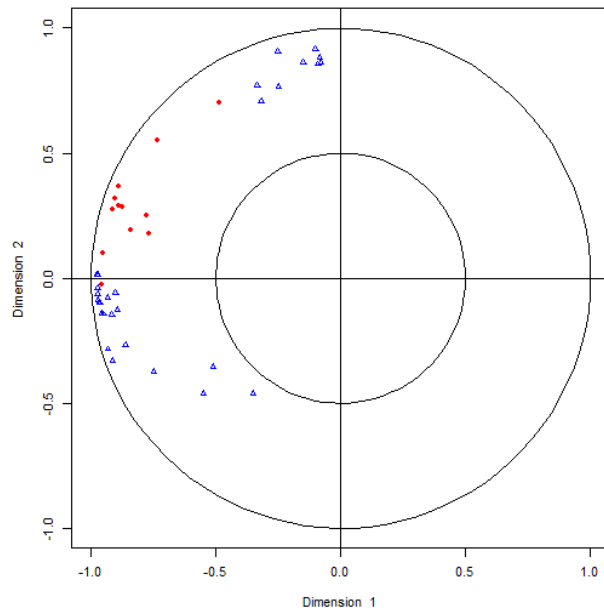
For the second group (the first 12):



```
$correlation
```

```
[1] 0.99 0.98 0.91 0.72 0.65 0.44 0.33 0.23 0.22 0.14 0.12 0.10
```

The p-value of hypothesis test is near 1 too, which means we will draw the same conclusion as above.



When compared these two biplot, the second method seems better than the first one, in some sense that the second one has a more clear relationship between the first and second dimension.