Use PCA to find the principal components and associated eigenvalues. (8 pts)

What is the minimum number of eigenvalues to keep (> 1.0 rule)? (1 pt)

What is the minimum number of eigenvalues to keep (scree plot)? (3 pts)

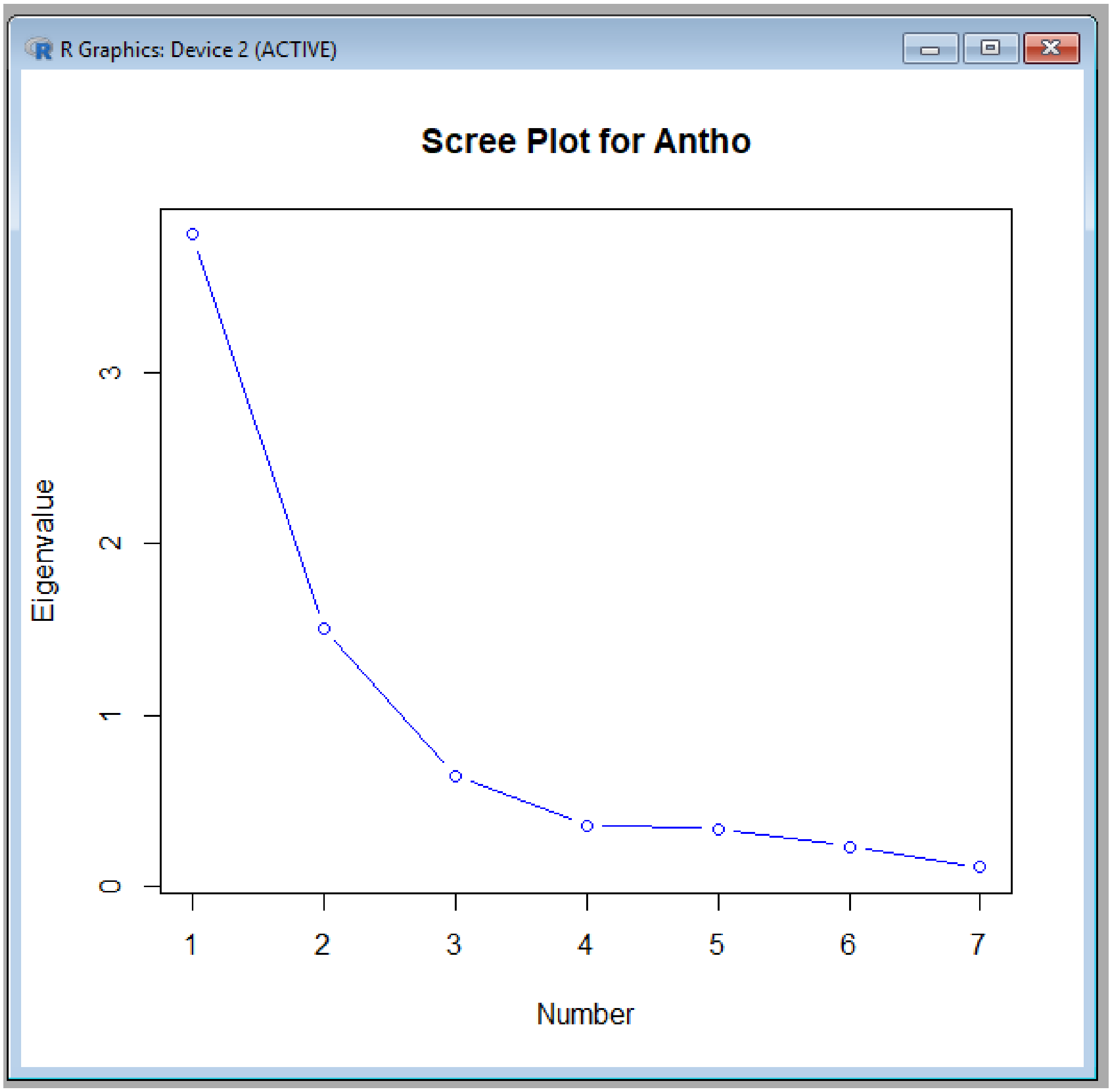
Interpret the first two principal components. (3 pts)

```
> antho<-read.csv('anthro.csv', header=T)

> rownames(antho)<-colnames(antho)

> antho

        HeadL HeadB FaceB LeftFL LeftFAL LeftFTL Height

HeadL   1.000 0.402 0.396  0.301   0.305   0.339  0.340

HeadB   0.402 1.000 0.618  0.150   0.135   0.206  0.183

FaceB   0.396 0.618 1.000  0.321   0.289   0.363  0.345

LeftFL  0.301 0.150 0.321  1.000   0.846   0.759  0.661

LeftFAL 0.305 0.135 0.289  0.846   1.000   0.797  0.800

LeftFTL 0.339 0.206 0.363  0.759   0.797   1.000  0.736

Height  0.340 0.183 0.345  0.661   0.800   0.736  1.000

> eigenAntho<-eigen(antho)

> std<-sqrt(eigenAntho$values)

> proportionVar<-eigenAntho$values/sum(eigenAntho$values)

> cumulativeProp<-cumsum(eigenAntho$values)/sum(eigenAntho$values)

> Loadings<-eigenAntho$vectors

>> #Importance of components:

> #Standard deviation

> std

[1] 1.9492241 1.2256950 0.8061063 0.6000474 0.5823766 0.4850290 0.3375164

> #Proportion of Variance
```

```
> proportionVar
[1] 0.54278208 0.21461832 0.09282963 0.05143670 0.04845178 0.03360759 0.01627391
> #Cumulative Proportion
> cumulativeProp
[1] 0.5427821 0.7574004 0.8502300 0.9016667 0.9501185 0.9837261 1.0000000
> #Loadings:
> Loadings
          [,1]       [,2]         [,3]        [,4]        [,5]         [,6]        [,7]
[1,] -0.2763037 -0.3647677  0.882274766 -0.08573946 -0.06740350  0.005384671 -0.01638732
[2,] -0.2118636 -0.6392041 -0.257527788  0.68707351  0.08129399  0.034955657  0.01762744
[3,] -0.2951449 -0.5123928 -0.381447691 -0.69856220 -0.10071831  0.033740772 -0.07462604
[4,] -0.4375581  0.2349399 -0.069924234  0.10160027 -0.61923662  0.318242311  0.50339046
[5,] -0.4557045  0.2766674 -0.036669136  0.11311530 -0.03907675  0.290305975 -0.78475748
[6,] -0.4502341  0.1784374 -0.059124621  0.05299938 -0.03440885 -0.870489463  0.01445146
[7,] -0.4356893  0.1795404 -0.006212105 -0.08162701  0.76976550  0.233030117  0.35269900
>> eigenAntho$values
[1] 3.7994745 1.5023283 0.6498074 0.3600569 0.3391625 0.2352531 0.1139173
```

#The first two eigenvalues account for 0.757 of the total correlation and are the only eigenvalues greater than 1.0

```
> layout(1)
> plot(1:7, eigenAntho$values, main='Scree Plot for Antho',
+ xlab='Number', ylab='Eigenvalue', col='blue', typ='b')
```



The usable number of principal components can be chosen by the screening criteria.The 'scree' plot is very sharp in slope for the first two components. The distinction is not due to random effects.

Use MDS to analyze this matrix. Show the output. (10 pts)

How many coordinates are need to account for at least 80% of the sum of absolute eigenvalues? (1 pt)

Using the first two coordinates, plot the location of the flowers relative to each other. (4 pts)

```
> flowers<-read.csv('flowers.csv', header=T)

> rownames(flowers)<-colnames(flowers)

> flowersMDS<- cmdscale(flowers, k=3, eig=TRUE)

> flowersMDS
```

```
$`points`
           [,1]        [,2]         [,3]
Bgn  0.441626810 -0.13910199 -0.024866489
Brm -0.411002097 -0.10646858  0.112134733
Cml  0.135998713 -0.05878727 -0.300985873
Dhl -0.117861814 -0.32989431 -0.146573607
Fmn  0.267792041  0.12142878 -0.000862644
Fch  0.295445879 -0.10140928 -0.284987056
Grn  0.167256180 -0.34946203 -0.053768686
Gld -0.061150521 -0.25197021  0.182807089
Hth  0.131374192  0.27948118  0.188840138
Hyd -0.173171040  0.35681051 -0.212378459
Irs  0.267620864  0.18687636  0.084102966
Lly -0.001753092  0.22650611  0.071416733
Lvy  0.051226880  0.34595178  0.113183286
Pny  0.015761013  0.11587345 -0.058215025
Pnc -0.178928108  0.09234363  0.155484871
Rdr -0.427237687 -0.04816267 -0.218569105
Scr -0.491784734 -0.01604355  0.006881662
Tlp  0.088786521 -0.32397190  0.386355466
$eig
 [1]  1.173085e+00  8.944353e-01  5.732009e-01  4.843006e-01  2.638516e-01  2.298165e-01
8.383417e-02  6.645861e-02  2.984017e-02 -2.775558e-17
[11] -2.147959e-02 -4.094094e-02 -4.366468e-02 -8.867401e-02 -1.045334e-01 -1.275889e-01 -
1.714295e-01 -2.045124e-01
$x
NULL
$ac
```

[1] 0

$GOF

[1] 0.5738645 0.6951420

>> cumulativeProp<-cumsum(abs(flowersMDS$eig))/sum(abs(flowersMDS$eig))

> cumulativeProp

[1] 0.2549273 0.4493002 0.5738645 0.6791096 0.7364481 0.7863903 0.8046086 0.8190510 0.8255356 0.8255356 0.8302034 0.8391005 0.8485894 0.8678594

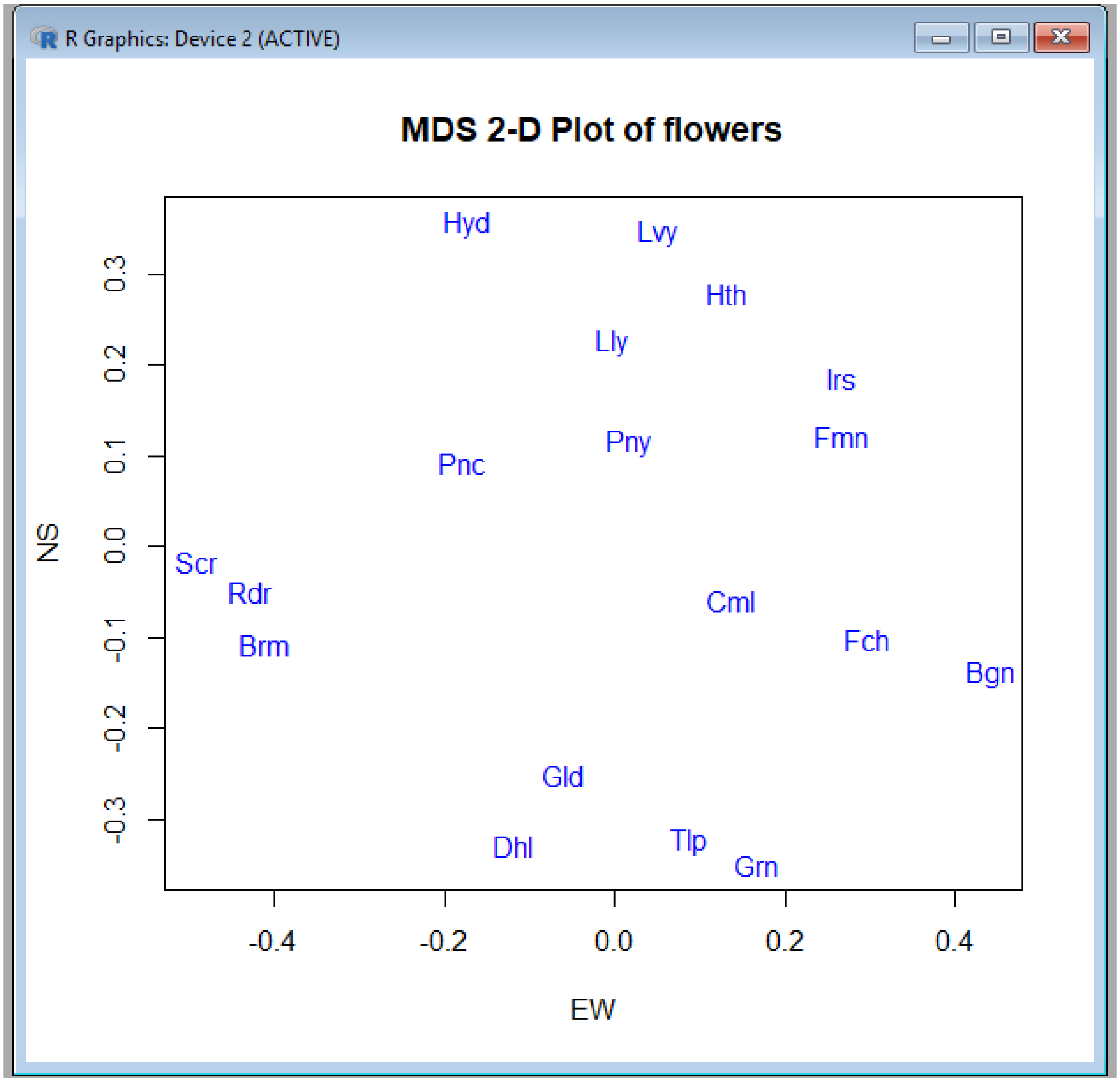
[15] 0.8905760 0.9183028 0.9555567 1.0000000

7 coordinates are need to account for at least 80% of the sum of absolute eigenvalues.

> layout(1)

> plot(flowersMDS$points[,1], flowersMDS$points[,2], type='n', main='MDS 2-D Plot of flowers', xlab='EW', ylab='NS')

> text(flowersMDS$points[,1], flowersMDS$points[,2], labels=colnames(flowers), col='blue')


R Graphics: Device 2 (ACTIVE)
MDS 2-D Plot of flowers
Hyd
Lvy
Hth
Lly
Irs
Pny
Fmn
Pnc
Scr
Rdr
Brm
Cml
Fch
Bgn
Gld
Dhl
Tlp
Grn
NS
0.3
0.2
0.1
0.0
-0.1
-0.2
-0.3
-0.4
-0.2
0.0
0.2
0.4
EW