Homework 3

sat3.txt contains the educational data for each state: Part03 is the % of students taking the SAT in 2003; Verbal03, Math03, and Total03 are the state averages for the respective sections; and Expend01 is the per pupil educational spending in 2001.

1) Construct a graphical display to support the argument that more money is not associated with higher test scores.

2) Construct a graphical display to counter the argument in 1 by also taking the percentage of student in the state taking the exam into account.
3) Considering the three variables Part03, Total03, and Expend01 choose a method of detecting outliers that seems reasonable to you to identify any states you think are outliers and briefly explain your results.

Eigenvalues and Eigenvectors

The eigenvectors of the covariance matrix are exactly the coefficients for principal components…

And the eigenvalues are the variances of the new variables!

Eigenvalues and Eigenvectors?!?

Consider the equation:

\[(\Sigma - \lambda I)x = 0\]

\(\lambda\) is an eigenvalue

\(x\) is an eigenvector

We typically make the length of \(x\) be 1. We typically order them from largest \(\lambda\) to smallest.
Example – Bivariate Normal

Last time we saw that a bivariate normal with correlation matrix
\[
\begin{pmatrix}
1 & r \\
r & 1
\end{pmatrix}
\]

had eigen values \((1+r)\) and \((1-r)\) and principal components
\[
\frac{1}{\sqrt{2}} (x_1 + x_2) \quad \text{and} \quad \frac{1}{\sqrt{2}} (x_1 - x_2)
\]

Example Cont.

```r
library(MASS)
source("http://biostatistics.iop.kcl.ac.uk/publications/everitt/RSPCMA/functions.txt")
mu<-c(0,0)
sigma1<-matrix(c(1,0.5,0.5,1),ncol=2)
ex1<-mvrnorm(n=1000,mu,sigma1)
sigma2<-matrix(c(1,0.9,0.9,1),ncol=2)
ex2<-mvrnorm(n=1000,mu,sigma2)
sigma3<-matrix(c(1,-0.9,-0.9,1),ncol=2)
ex3<-mvrnorm(n=1000,mu,sigma3)
```

Example Cont.

```r
bvbox(ex1)
var(ex1)
```

1.033 0.503
0.503 1.033
Example Cont.

```
pcl<-cbind((1/sqrt(2))*(ex1[,1]+ex1[,2]),
            (1/sqrt(2))*(-ex1[,1]+ex1[,2]))
bvbox(pcl)
var(pcl)
  1.533  -0.003
  -0.003   0.527
```
Example Cont.

Could also use these functions...

\[
\begin{align*}
eigen(\text{cov}(\text{ex1}))\text{\$vectors} \\
\text{princomp}(\text{ex1,cor=F})\text{\$loadings} \\
eigen(\text{cov}(\text{ex1}))\text{\$values} \\
\text{round(\text{var(princomp(ex1,cor=F)\$scores)},2)} \\
\text{princomp}(\text{ex1,cor=F})\text{\$scores}
\end{align*}
\]

Now, Let's try the oildata...

\[
\text{oildata<-read.table(}\text{"http://www.stat.sc.edu/~habing/courses/data/oild ata.txt"},\text{header=TRUE)}
\]
\[
\text{sect3<-as.matrix(oildata[,12:31])}
\]
\[
\text{oilpc<-princomp(sect3,cor=F)}
\]

Now, Let's try the oildata...

\[
> \text{oilpc}\text{\$loadings[,]\[1]} \\
\text{\begin{array}{cccccc}
Q1 & Q2 & Q3 & Q4 & Q5 & Q6 \\
0.14749703 & -0.18642164 & -0.24432644 & 0.09980284 & 0.14749703 & -0.18642164 \\
Q7 & Q8 & Q9 & Q10 & Q11 & Q12 \\
0.35633250 & -0.17468260 & -0.04253903 & 0.01630308 & 0.35633250 & -0.17468260 \\
Q13 & Q14 & Q15 & Q16 & Q17 & Q18 \\
-0.24074574 & 0.07216813 & 0.38062551 & 0.16292494 & -0.24074574 & 0.07216813 \\
Q19 & Q20 & Q21 & Q22 & Q23 & Q24 \\
-0.23470978 & -0.23005868 & -0.19195023 & -0.11866190 & -0.23470978 & -0.23005868
\end{array}}
\]