Statistics 503 Homework 2  
Due: in class on Wednesday, Feb 16 2005

Background: The graphics is the olive oil case study are not suitable for high quality publication. We would like to produce better quality plots that elucidate the class/cluster structure in this data using R. One part of this is to use the lattice package for low-dimensional plots, and the other part is to reproduce the tour plots revealing higher dimensional interactions from ggobi and enhance them.

Purpose: To produce useful presentation graphics for classification data.

Exercises:

1. This code using the lattice package produces the following dotplots for the olive oil data:

```
library(lattice)
d.olive0<-d.olive[,3:10]
d.olive.df<-data.frame(Region=rep(c(rep("South",323),
    rep("Sardinia",98),rep("North",151)),8),
y=as.vector(as.matrix(d.olive0)),
meas=factor(rep(1:8, rep(572,8)), labels=names(d.olive0))
jpeg("olivedotplot.jpg", width=1200, height=600, pointsize=12,
    quality=300, bg="white")
par(pty="s", mar=c(2,1,1,1))
plt.bg<-trellis.par.get("background")
plt.bg$col="grey90"
trellis.par.set("background", plt.bg)
stripplot(Region~y|meas, data=d.olive.df, scales=list(x="free"),
    strip=function(...) strip.default(style=1,...),
    panel=function(x,y)
        panel.grid(h=-1, v=5, col="white")
        panel.stripplot(x,y,col=1, pch=16),
xlab="", pch=16, col=1, layout=c(4,2), aspect=1,
    as.table=T)
dev.off()
```
These plots help to decide which variables are important to separate the three regions. Eicosenoic acid clearly distinguishes southern oils from the others, and linoleic acid distinguishes northern oils from sardinian oils.

Your job is to do the same for the three areas of the northern oils. You’ll need to subset the data to extract just the oils where region is 3, and then write code similar to the above to get a dotplot for each fatty acid. Hand in your plot and your R code.

2. Using ggobi, we found a linear combinations of variables that separates the oils from the three regions. The ggobi screendump is below. We reconstructed the projected data using R and enhanced the plot with the following code:

```R
coefs<-matrix(c(0,0,0,0,0,0,0,0,0,0,0.991/1022,-0.011/1022,0,0, 
                0.135/105,0.013/105,0.009/57,1.000/57),ncol=2,byrow=T) 
proj<-as.matrix(d.olive) %*% coefs 
jpeg("olivetour2.jpg",width=500,height=500,pointsize=12, 
    quality=300,bg="white") 
par(pty="s",bg="gray90") 
plot(proj[,1],proj[,2],xlab="0.97xLinoleic+1.29xArachidic (x1000)", 
    ylab="Eicosenoic",type="n",xlim=c(0.4,1.6),ylim=c(-0.1,1.1)) 
abline(v=c(1.4,1.2,1.0,0.8,0.6),h=c(0.0,0.2,0.4,0.6,0.8,1.0), 
    col="white",lty=1) 
box() 
points(proj[1:323,1],proj[1:323,2],pch="1") 
points(proj[324:421,1],proj[324:421,2],pch="2") 
points(proj[422:572,1],proj[422:572,2],pch="3") 
legend(1.3,1.1,pch=c("1","2","3"),legend=c("South","Sardinia","North")) 
dev.off()
```

Your job is to do something similar for the three areas of the northern oils. Find a linear combination of the fatty acids where the three areas (Umbria, East and West Liguria) are reasonably separated. Take a screendump of the plot window. Reconstruct this projection in R and enhance it similarly to the plot above. Hand in the screendump, the R plot and your R code.