```
{\rtf1\ansi\ansicpg1252\cocoartf1038\cocoasubrtf360
{\fonttbl\f0\fmodern\fcharset0 Courier;}
{\colortbl;\red255\green255\blue255;}
\margl1440\margr1440\vieww12700\viewh11720\viewkind0
\deftab720
\pard\pardeftab720\ql\qnatural

\f0\fs28 \cf0 \
Chapter 8.  Multivariate Analysis\
\
# Overview of globular cluster properties\
\
GC = read.table("http://astrostatistics.psu.edu/MSMA/datasets/GlobClus_prop.dat", \
  header=T, fill=T)\
summary(GC)\
manyhist <- function(x) \{ \
    par(mfrow=n2mfrow(dim(x)[2]))  \
    for (i in 1:dim(x)[2]) \{ name=names(x)[i]              \
        hist(x[,i], main='', breaks='FD', ylab='', xlab=name) \}\}\
    par(mfrow=c(1,1))\
manyhist(GC[,2:13])\
\
# Univariate boxplots and normal quantile-quantile plots for two variables\
\
par(mfrow=c(2,1))\
boxplot(GC[,8], main='', ylab=expression(Core~r~~ (pc)), pars=list(xlab='',
cex.lab=1.3, cex.axis=1.3, pch=20))\
qqnorm(GC[,8], main='', xlab='', ylab='', cex.lab=1.3, cex.axis=1.3, pch=20)\
boxplot(GC[,20], main='', ylab=expression(CSB ~~(mag/arcmin^2)), \
  cex.lab=1.3, cex.axis=1.3, pars=list(xlab='', pch=20))\
qqnorm(GC[,20], main='', xlab='', ylab='', cex.lab=1.3, \
  cex.axis=1.3, pch=20)\
qqline(GC[,20])\
par(mfrow=c(1,1))\
\
#  Prepare the data\
# 1. Remove objects with NA entries, remove labels\
\
dim(GC) ; GC1 <- na.omit(GC[,-1]) ; dim(GC1)\
\
# 2. Standardize variables \
\
GC2 <- scale(GC1)\
\
# 3. Separate locational and dynamical variables\
\
```

```
GCloc <- GC2[,+c(1:4)]\
GCdyn <- GC2[,-c(1:4)]\
\
# 4. Remove two bad outliers\
\
GCdyn1 <- GCdyn[-c(which.max(GCdyn[,4]), which.max(GCdyn[,11])),]\
GCloc1 <- GCloc[-c(which.max(GCdyn[,4]), which.max(GCdyn[,11])),]\
\
# Bivariate relationships\
\
cor(GCdyn1, method='kendall')  \
var(GCdyn1)\
pairs(GCdyn1[,3:8],pch=20,cex=0.4)\
\
# Elaborated color pairs plot\
\
library(MASS)\
pairs(GCdyn1[,5:8],main='',labels=names(GCdyn1[5:8]), panel=function(x,y) \{\
          abline(lsfit(x,y)$coef,lwd=2,col='deeppink2')\
     lines(lowess(x,y),lwd=3,col='blue3',lty=2)\
          points(x,y,pch=21,bg = c("red", "green3", "blue"))\
          rug(jitter(x,factor=3),side=1,col='lightcoral',ticksize=-.05)\
          rug(jitter(y,factor=3),side=2,col='cornflowerblue',ticksize=-.05)\
          contour(kde2d(x,y)$x, kde2d(x,y)$y, kde2d(x,y)$z, drawlabels=F,add=T,\
   col='darkblue',nlevel=4)\
          \})\
\
# PCA for dynamical variables.\
\
PCdyn <- princomp(GCdyn1)\
plot(PCdyn, main='')\
summary(PCdyn) \
loadings(PCdyn)\
biplot(PCdyn, col='black', cex=c(0.6,1))\
\
# Add principal component values into the data frame\
\
PCdat <- data.frame(names=row.names(GCdyn1), GCdyn1, PCdyn$scores[,1:4])\
\
# Multiple regression to predict globular cluster central surface brightnesses\
\
GCdyn2 <- as.data.frame(GCdyn1)\
attach(GCdyn2)\
CSB_fit1 <- lm(CSBt~.-CSBt,data=GCdyn2) ; CSB_fit1\
CSB_fit2 <- lm(CSBt~.,data=GCdyn2[,c(7:11,13)]) ; CSB_fit2\
str(CSB_fit2)\
```

```
summary(CSB_fit2) \
sd(CSB_fit2$residuals)\
par(mfrow=c(2,1))\
plot(CSB_fit2$fitted.values, CSBt, pch=20, main='Globular cluster central surface
brightness') \
lines(c(-2, 3), c(-2, 3))\
qqnorm(CSB_fit2$residuals, pch=20, main='Residuals') \
qqline(CSB_fit2$residuals)\
par(mfrow=c(1,1))\
\
# MARS nonlinear regression\
\
install.packages('earth') ; library(earth)\
CSB_fit3 <- earth(CSBt~.-CSBt, data=GCdyn2) ; CSB_fit3\
sd(CSB_fit3$residuals) \
qqnorm(CSB_fit2$residuals) ; qqline(CSB_fit2$residuals)\
\
# Some multivariate display techniques: \
# interactive 3-dim scatter plot; 4-dim bubble plot; parallel coordinates plot\
\
install.packages('rgl') ; library(rgl)\
open3d() ; plot3d(GCdyn1[,5:7]) \
snapshot3d(file='GlobClus3D.png') \
\
library(lattice)\
cloud(GCdyn1[,5]~GCdyn1[,6]*GCdyn[,7], screen=list(z=60,x=45,y=20),\
   xlab='log.t.rad', ylab='log.rho.cen', zlab='conc', col=1, pch=1, cex=GCdyn1[,8]+1)\
\
parallel(~GCdyn1, col=c('darkred','darkgreen','orange','blue','black'))\
\
}
```