City = read.csv("http://course1.winona.edu/bdeppa/DSCI%20415/Data/City77.csv")

row.names(City) = City$City

City = City[,-1]

names(City)

hist(City$medinc,prob=T,nclass=20,col="blue",main="Histogram of Median Income",xlab="Median Income")

lines(density(City$medinc),lty=2,col="red",lwd=2)

require(ggplot2)

a = ggplot(City,aes(medinc))

a + geom\_histogram(binwidth=1000,aes(y=..density..),fill="blue",color="black") +

geom\_density(linetype=2) +

xlab("Median Income") +

ggtitle("Median Income for Top 77 Cities")

plot(City$poverty,City$infmort,xlab="Percent Below Poverty Level",ylab="Infant Mortality Rate")

lines(lowess(City$poverty,City$infmort),lty=2,col="blue",lwd=2)

abline(lm(infmort~poverty,data=City),lwd=2)

title(main="Infant Morality vs. Poverty Level")

a = ggplot(City,aes(poverty,infmort))

a + geom\_point() + geom\_smooth() +

geom\_text(aes(label=row.names(City)),size=2) +

xlab("Poverty") + ylab("Infant Mortality") +

ggtitle("Infant Mortality vs. Poverty Rate")

Olives = read.csv("http://course1.winona.edu/bdeppa/DSCI%20415/Data/Olives.txt")

names(Olives)

par(mfrow=c(3,2))

boxplot(split(Olives$Oleic,Olives$Area.Name),xlab="Area",ylab="Oleic Acid")

boxplot(split(Olives$Linoleic,Olives$Area.Name),xlab="Area",ylab="Linoleic Acid")

boxplot(split(Olives$Strearic,Olives$Area.Name),xlab="Area",ylab="Strearic Acid")

boxplot(split(Olives$Linolenic,Olives$Area.Name),xlab="Area",ylab="Linolenic Acid")

boxplot(split(Olives$Eicosanoic,Olives$Area.Name),xlab="Area",ylab="Eicosanoic Acid")

boxplot(split(Olives$Eicosenoic,Olives$Area.Name),xlab="Area",ylab="Eicosenoic Acid")

require(s20x)

boxqq(Oleic~Region.Name,data=Olives)

install.packages("s20x")

a = ggplot(Olives,aes(Area.Name,Oleic))

a + geom\_boxplot(fill="lightblue") + xlab("Area Name") +

ylab("Oleic Acid") + ggtitle("Oleic Acid vs. Growing Area")

install.packages("violinmplot")

require(violinmplot)

violinmplot(Area.Name~Oleic,xlab="Oleic Acid",ylab="Area Name",data=Olives,horizontal=T)

library(ggplot2)

a = ggplot(Olives,aes(Area.Name,Oleic))

a + geom\_violin(fill="lightblue") + xlab("Area Name") +

ylab("Oleic Acid") + ggtitle("Oleic Acid vs. Growing Area")

Statplot(Olives$Oleic,xname="Oleic")

install.packages("Statplot")

data("UCBAdmissions")

UCBAdmissions

temp = data.frame(UCBAdmissions)

temp

DeptCount = margin.table(UCBAdmissions,3)

DeptCount

par(mfrow=c(1,2))

pie(DeptCount)

barplot(DeptCount,xlab="Department",ylab="Frequency")

par(mfrow=c(1,2))

DeptAdmit = margin.table(UCBAdmissions,c(1,3))

DeptAdmit

barplot(DeptAdmit,xlab="Department",ylab="Frequency")

barplot(DeptAdmit,xlab="Department",ylab="Frequency",beside=TRUE)

GenderAdmit = margin.table(UCBAdmissions,c(1,2))

GenderAdmit

pie(GenderAdmit[,1],main="Males")

pie(GenderAdmit[,2],main="Females")

barplot(GenderAdmit,xlab="Gender",ylab="Frequency")

barplot(GenderAdmit,xlab="Gender",ylab="Frequency",beside=T)

par(mfrow=c(1,2))

mosaicplot(~Gender+Admit,data=UCBAdmissions,color=T)

mosaicplot(~Admit+Gender,data=UCBAdmissions,color=T)

par(mfrow=c(1,2))

mosaicplot(~Dept+Admit,data=UCBAdmissions,color=T)

mosaicplot(~Admit+Dept,data=UCBAdmissions,color=T)

mosaicplot(~Dept+Gender+Admit,data=UCBAdmissions,color=T)

mosaicplot(~Dept+Admit+Gender,data=UCBAdmissions,color=T)

mosaicplot(~Gender+Dept+Admit,data=UCBAdmissions,color=T)

names(Olives)

olive.mat = Olives[,c(7,6,5,3)]

pairs(olive.mat)

panel.cor = function (x, y, digits = 2, prefix = "", cex.cor)

{

usr <- par("usr")

on.exit(par(usr))

par(usr = c(0, 1, 0, 1))

r <- abs(cor(x, y))

txt <- format(c(r, 0.123456789), digits = digits)[1]

txt <- paste(prefix, txt, sep = "")

if (missing(cex.cor))

cex <- 0.8/strwidth(txt)

text(0.5, 0.5, txt, cex = cex \* r)

}

panel.trendsd = function (x, y, f = 0.5)

{

par(err = -1)

xs <- sort(x, index = T)

x <- xs$x

ix <- xs$ix

y <- y[ix]

trend <- lowess(x, y, f)

e2 <- (y - trend$y)^2

scatter <- lowess(x, e2)

uplim <- trend$y + sqrt(abs(scatter$y))

lowlim <- trend$y - sqrt(abs(scatter$y))

points(x, y, pch = 1)

lines(trend, col = "Blue")

lines(scatter$x, uplim, lty = 2, col = "Red")

lines(scatter$x, lowlim, lty = 2, col = "Red")

invisible()

}

panel.hist = function (x, ...)

{

usr <- par("usr")

on.exit(par(usr))

par(usr = c(usr[1:2], 0, 1.5))

h <- hist(x, plot = FALSE)

breaks <- h$breaks

nB <- length(breaks)

y <- h$counts

y <- y/max(y)

rect(breaks[-nB], 0, breaks[-1], y, col = "cyan", ...)

}

pairs.trendsd = function(data,...) {

pairs(data,lower.panel=panel.cor,upper.panel=panel.trendsd,

diag.panel=panel.hist,...)}

require(lattice)

pairs.trendsd(olive.mat)

Boston = read.csv("http://course1.winona.edu/bdeppa/DSCI%20415/Data/Boston.txt")

names(Boston)

install.packages("corrgram")

require(corrgram)

corrgram(Boston[,-c(1:4)],lower.panel=panel.pts,upper.panel=panel.pie)

corrgram(Boston[,-c(1:4)],lower.panel=panel.shade,upper.panel=panel.pie)

require(corrplot)

Boston.corr = cor(Boston[,-c(1:4)]) # find pairwise correlations between all variables.

options(digits=2) # reduce the number of significant digits shown.

Boston.corr

corrplot(Boston.corr)

corrplot(Boston.corr,method="ellipse")

data(iris)

names(iris)

library(lattice)

xyplot(Sepal.Length~Petal.Length|Species,data=iris)

xyplot(Sepal.Length~Petal.Length|Species,layout=c(2,2),data=iris)

SepWid = equal.count(iris$Sepal.Width)

plot(SepWid)

print(SepWid)

xyplot(Sepal.Length~Petal.Length|SepWid,data=iris)

xyplot(Sepal.Length~Petal.Length|SepWid,groups=Species,layout=c(3,2),

auto.key=list(columns=3),main="Sepal Length \* Petal Length | Sepal

Width",data=iris)

splom(~iris[,1:4],groups=Species,auto.key=list(columns=3),data=iris)

bwplot(Species~Sepal.Width,data=iris,xlab="Sepal Width (mm)",main="Boxplots of Iris Sepal Width")

stripplot(Species ~ jitter(Sepal.Width), data = iris, xlab="Sepal Width (mm)",main="Sepal Width Across Species")

stripplot(Species ~ jitter(Sepal.Width), data = iris, aspect = 1,

jitter=T,xlab="Sepal Width (mm)",main="Sepal Width Across Species")

coplot(Sepal.Width~Sepal.Length|Species,data=iris)

coplot(Sepal.Width~Sepal.Length|Petal.Width,number=4,overlap=.2,data=iris)

Ozdata = read.csv("http://course1.winona.edu/bdeppa/DSCI%20415/Data/Ozdata.csv")

names(Ozdata)

coplot(upoz~safb|inbh\*v500,number=c(4,4),panel=panel.smooth,data=Ozdata)

coplot(upoz~safb|inbh\*v500,number=c(4,4),overlap=.25,panel=function(x,y,...) panel.smooth(x,y,span=.6,...),data=Ozdata)

install.packages("ash")

library(ash)

Swiss = read.csv("http://course1.winona.edu/bdeppa/DSCI%20415/Data/Swiss.csv")

names(Swiss)

d1 = ash1(bin1(Swiss$diagon,nbin=50),3)

hist(Swiss$diagon,nclass=20,prob=T,col="blue",main="Histogram of Bill Diagonal (mm)")

lines(d1)

Statplot(Swiss$diagon)

diagrt.bin <- bin2(cbind(Swiss$diagon,Swiss$right),nbin=c(50,50))

diagrt.1 <- ash2(diagrt.bin,m=c(5,5))

persp(diagrt.1,xlab="Diagonal Length",ylab="Right Height",zlab="",cex=.5,theta=-45,phi=30,shade=1,col="royal blue")

image(diagrt.1,xlab="Bill Diagonal",ylab="Right Height")

contour(diagrt.1,xlab="Bill Diagonal",ylab="Right Height",add=T)

points(Swiss$diagon,Swiss$right,pch=as.character(Swiss$genu),cex=.4)

Genuine <- Swiss$genu

Genuine[Genuine==0] <- "Forged"

Genuine[Genuine==1] <- "Real"

xyplot(Swiss$diagon~Swiss$bottom|Genuine,groups=Genuine)

splom(~Swiss[,1:6],groups=Genuine,auto.key=list(columns=2))

pairs.image <- function(x) {

pairs(x,panel=function(x,y) {

foo <- bin2(cbind(x,y),nbin=c(75,75))

foo <- ash2(foo,m=c(6,6))

image(foo,add=T,xlab="",ylab="",col=topo.colors(1000))

points(x,y,pch=".")

})

}

pairs.image(Swiss[,1:6])

pairs.persp <- function(x) {

par(bg="sky blue")

pairs(x,panel=function(x,y) {

foo <- bin2(cbind(x,y),nbin=c(75,75))

foo <- ash2(foo,m=c(8,8))

par(new=T)

persp(foo,xlab="",ylab="",theta=-45,phi=35,col="royal blue",

shade=.75,box=F,scale=F,border=NA,expand=.9)

})

par(new=F,bg="white")

}

pairs.persp(Swiss[,1:3])

 

 

 

 

 

 

 

 

 

 

 

 

 

 







