**Question 1**

Investigate the air quality of two areas in California. Use the base plotting functions from the graphics package. Show the code you used to generate the graphs. Use the data in AirQualitySanFranciscoData.csv and in AirQualitySanDiegoData.csv. Create four graphs with the year on the horizontal axis. Two of the graphs have number of days ppm exceeded the standard on the vertical axis. Two of the graphs have maximum concentration on the vertical axis. Notice the spacing between the columns of graphs. Make an exact replica of the graph shown below:

**Answer 1**

SanDiegoAir=read.csv("C:/ITU/RClass/AirQualitySanDiegoData.csv",header=TRUE,sep=",",na.strings="NA", strip.white=T)

SFAir=read.csv("C:/ITU/RClass/AirQualitySanFranciscoData.csv",header=TRUE,sep=",",na.strings="NA", strip.white=T)

par(mfrow = c(2,2), mar = c(2, 3, 2, 3), oma = c(2, 2, 2, 2))

layout(rbind(c(1,2), c(3,4)),heights = c(2,2), widths = c(2,2))

yrange1 <- range(SanDiegoAir$ExceedStateLevelDays8.hr)

xrange1 <- range(SFAir$Year)

plot(xrange1, yrange1, xlab ="Year", ylab="Number of Days over Standard",type="n")

with(SFAir,{

lines(Year, ExceedStateLevelDays1.hr, col = "Red",lty = 2)

lines(Year, ExceedStateLevelDays8.hr, col = "Yellow",lty = 3)

lines(Year, ExceedNationalLevelDays8.hr, col = "Green",lty = 2)

points(Year, ExceedStateLevelDays1.hr, col = "Red",bg = "Red")

points(Year, ExceedStateLevelDays8.hr, col = "Yellow",bg = "Yellow")

points(Year, ExceedNationalLevelDays8.hr, col = "Green", ,bg = "Green")

title(main = "San Francisco Air Quality", ylab = "Number of Days over Standard")

legend("topright",title = "Over Standard",c("State Days 1 hour","State Days 8 hour", "National level 8 hours"),col = c("Red","Yellow","Green"),pch = 1,inset = .05,,cex = 0.6)

mtext("Air Quality", cex = 2, side = 3, line = 0, outer = T)

})

plot(xrange1, yrange1,xlab="Year", ylab="Number of Days over Standard",type="n")

with(SanDiegoAir,{

lines(Year, ExceedStateLevelDays1.hr, col = "Red",lty = 2)

lines(Year, ExceedStateLevelDays8.hr, col = "Yellow",lty = 4)

lines(Year, ExceedNationalLevelDays8.hr, col = "Green",lty = 2)

points(Year, ExceedStateLevelDays1.hr, col = "Red",bg = "Red")

points(Year, ExceedStateLevelDays8.hr, col = "Yellow",bg = "Yellow")

points(Year, ExceedNationalLevelDays8.hr, col = "Green",bg = "Green")

title(main = "San Diege Air Quality", ylab = "Number of Days over Standard")

legend("topright",title = "Over Standard",c("State Days 1 hour","State Days 8 hour", "National level 8 hours"),col = c("Red","Yellow","Green"),pch = 1,inset = .05,cex = 0.6)

})

yrange2 <- range(SDair$MaxConcentration1.hr)

plot(xrange1, yrange2, xlab="Year", ylab="Maximum Concentration",type="n")

with(SFAir,{

lines(Year, MaxConcentration1.hr, col = 5,lty = 2)

lines(Year, MaxConcentration8.hr, col = 4,lty = 3)

points(Year, MaxConcentration1.hr, col = 5,bg = 5)

points(Year, MaxConcentration8.hr, col = 4,bg = 4)

title(ylab = "Number of Days over Standard")

legend("topright",title = "ppm",c("Max 1 hour","Max 8 hour"),col = c(4,5),pch = 1,inset = .05,,cex = 0.6)

})

plot(xrange1, yrange2, xlab="Year", ylab="Maximum Concentration",type="n")

with(SanDiegoAir,{

lines(Year, MaxConcentration1.hr, col = 5,lty = 2)

lines(Year, MaxConcentration8.hr, col = 4,lty = 3)

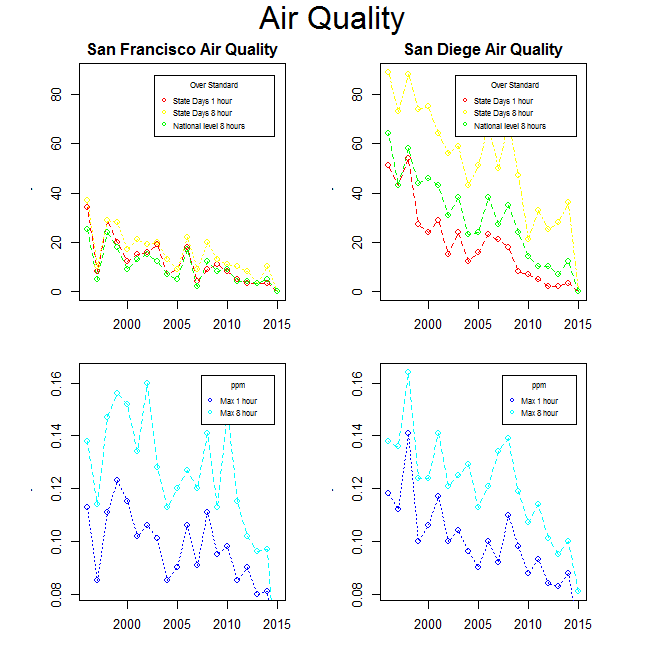
points(Year, MaxConcentration1.hr, col = 5,bg = 5)

points(Year, MaxConcentration8.hr, col = 4,bg = 4)

title(ylab = "Number of Days over Standard")

legend("topright",title = "ppm",c("Max 1 hour","Max 8 hour"),col = c(4,5),pch = 1,inset = .05,,cex = 0.6)

})

****

**Question 2**

Investigate the land areas of USA states. Use the base plotting functions from the graphics package. Show the code you used to generate the graphs. Use the state.x77 data set (recall typing ?state.x77 into the console will provide a discription of this data set). Notice the axis for the charts. Create the graphs and charts exactly the same as those below:

**Answer 2**

d <-as.data.frame(cbind(state.x77,region=state.region))

density <- density(d$Area)

xcor <- as.integer(c(min(d$Area),mean(d$Area),d[rownames(d)== "Texas",colnames(d

)=="Area"],max(d$Area)))

par(mfrow = c(2,1), mar = c(4, 4, 3, 5), oma = c(0, 0, 0, 0))

plot(density, col="red", lwd=2, main="Probability Density Plot \n State Land Area in Sq

uare Miles",xlim=c(min(d$Area),max(d$Area)),xlab="Land Area in Square Mile"

,xaxt="n")

axis(1,at=xcor,labels=xcor)

abline(v=mean(d$Area),col = "black", lty = 2)

text(130000,10e-06,paste("mean = ",round(mean(d$Area),digits=0)))

text(d$Area[rownames(d)=="Alaska"]-85000,8e-06,paste("Alaska Area = \n",d$Area[ro

wnames(d)=="Alaska"]))

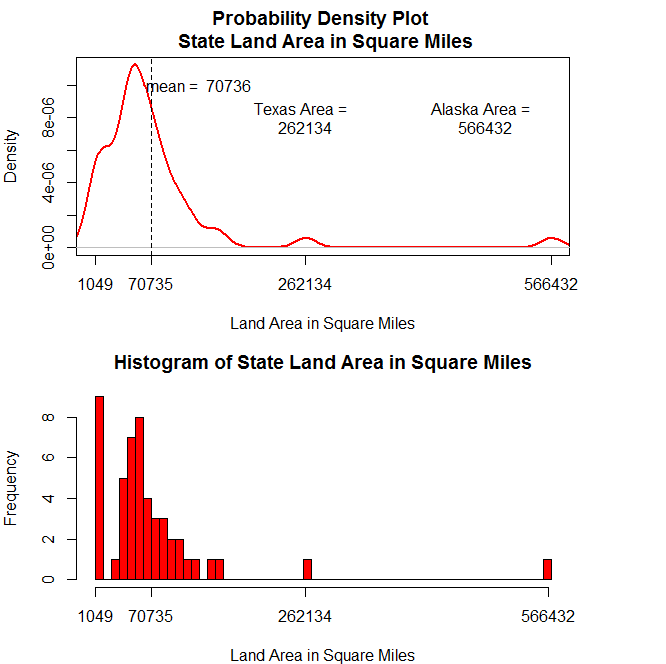
text(d$Area[rownames(d)=="Texas"]-3000,8e-06,paste("Texas Area = \n",d$Area[row

names(d)=="Texas"]))

hist(d$Area, col = "red", main = "Histogram of State Land Area in Square Miles",xlab =

"Land Area in Square Miles",xaxt="n", breaks=50)

axis(1,at=xcor,labels=xcor)

****

**Question 3**

Investigate the land areas of USA regions. State regions can be found in the state.region data set. Use the base plotting functions from the graphics package. Show the code you used to generate the graphs. Use the data sets state.x77 and state.region. Notice the axis for the charts. Watch out; Color code the graphs correctly. Create the graphs and charts exactly the same as those below:

**Answer 3**

par(mfrow = c(3,1))

d <- cbind(d,regionName= as.factor(state.region))

areaByRegion <- 1:4

for(index in 1:4){

indx<-which(d$region==index)

areaByRegion[index] <- sum(d$Area[indx])

}

areasum <- cbind(Region = c("Northeast","South","North Central","West"),as.data.frame(areaByRegion))

par(mfrow = c(3,1), mar = c(3, 2, 3, 2), oma = c(1, 2, 1, 2))

barplot(areasum$areaByRegion, main = "Area By Region",xlab = "Region",ylab = "Area in Square Miles", col = c(2,7,3,4))

x <- d[order(d$Area),]

x$regionName <- factor(x$regionName)

x$color[x$regionName=="Northeast"] <- 2

x$color[x$regionName=="South"] <- 7

x$color[x$regionName=="North Central"] <- 3

x$color[x$regionName=="West"] <- 4

names(x)[4]<-"lifeexp"

names(x)[6]<- "hsgrad"

dotchart(x$Area,

labels = row.names(x),

cex = .5,

groups = x$regionName,

gcolor = "black",

color = x$color,

pch = 19,

main = "Area of State in Square Miles\ngrouped by region",

xlab = "Area in Square Miles",

xlim = c(0,150000))

oldpar <- par(no.readonly=TRUE)

par(lwd=4)

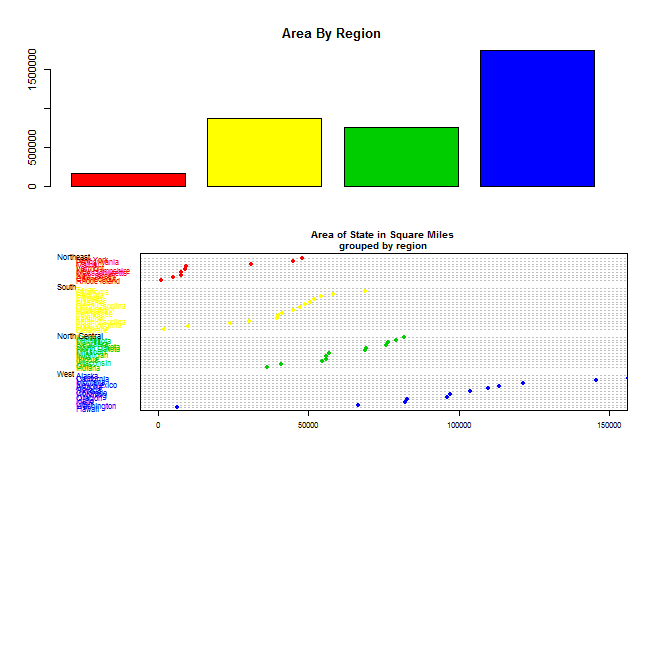
sm.density.compare(x$Area, x$regionName,col = c(2,7,3,4), lty = c(2,3,1,4) )

legend("topright", c("Northeast","North Central","South","West"),

title = "Region Names", inset = .02,lwd = 1,

col=c(2,3,7,4),cex = 0.7,lty = c(2,3,1,4))

title(main = "Probability Density Comparison of Areas by Regions")

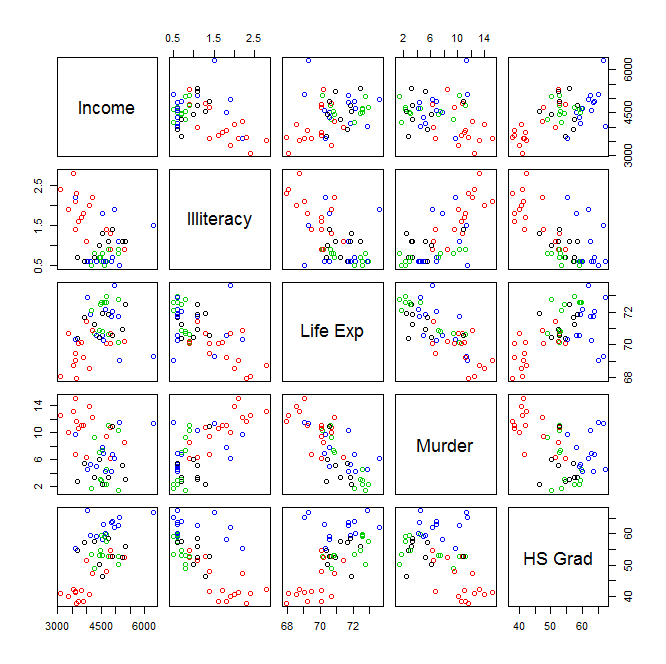
****

**Question 4**

Some interesting observations can be seen by looking at pairs plots of the state.x77 data. Use the base plotting functions from the graphics package. Show the code you used to generate the graphs. Create the graphs and charts exactly the same as those below:

**Answer 4**

pairs(state.x77[,2:6], col = unclass(state.region))

****

Illiteracy <- state.x77[ ,3]

LifeExpectancy<-state.x77[,4]

plot(Illiteracy,LifeExpectancy, main="Illiteracy vs Life Expectancy",col = unclass(state.region))

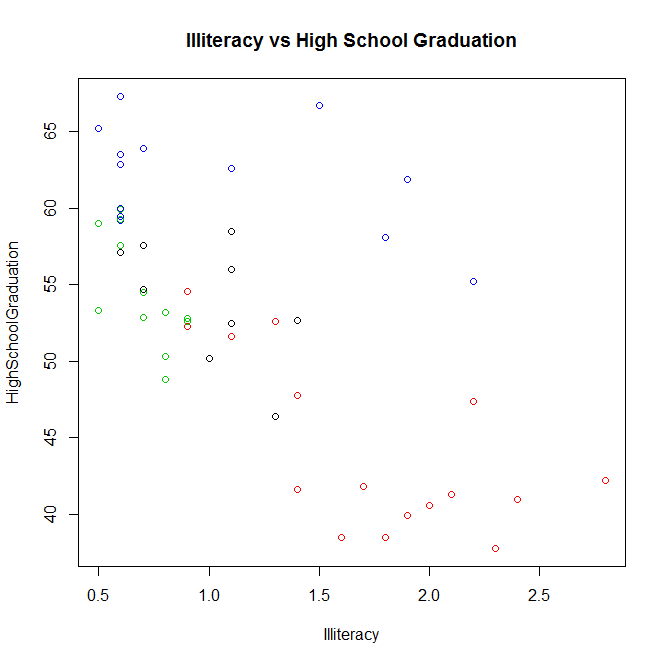
legend("topright",inset=.05,cex = 1,title="Region Names",c("North East","North Central","South","West"),horiz=FALSE,lty=c(1,1),lwd=c(2,2),col=c("red","blue"),text.font=2)



Illiteracy <- state.x77[ ,3]

HighSchoolGraduation<-state.x77[, 6]

plot(Illiteracy,HighSchoolGraduation, main="Illiteracy vs High School Graduation",col = unclass(state.region))



Income<-state.x77[,2]

LifeExpectancy<-state.x77[,4]

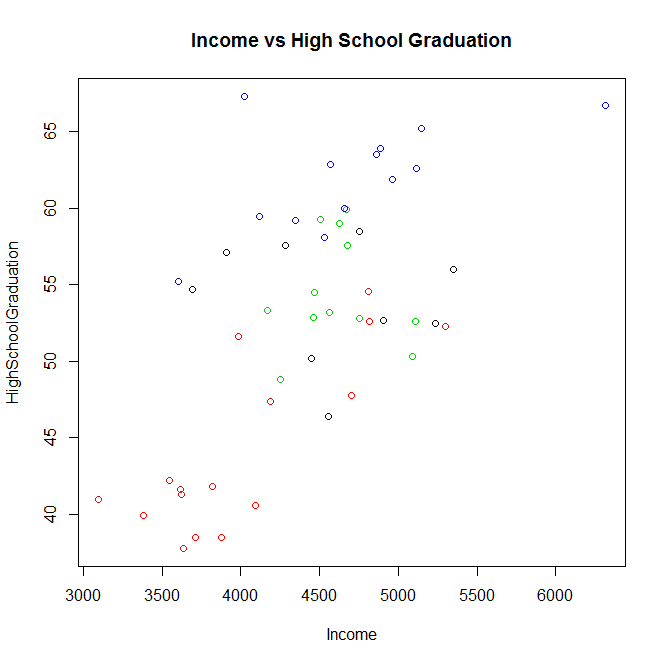
plot(Income,LifeExpectancy, main="Income vs Life Expectancy",col = unclass(state.region))



Income<-state.x77[,2]

HighSchoolGraduation<-state.x77[, 6]

plot(Income,HighSchoolGraduation, main="Income vs High School Graduation",col = unclass(state.region))



**Question 5**

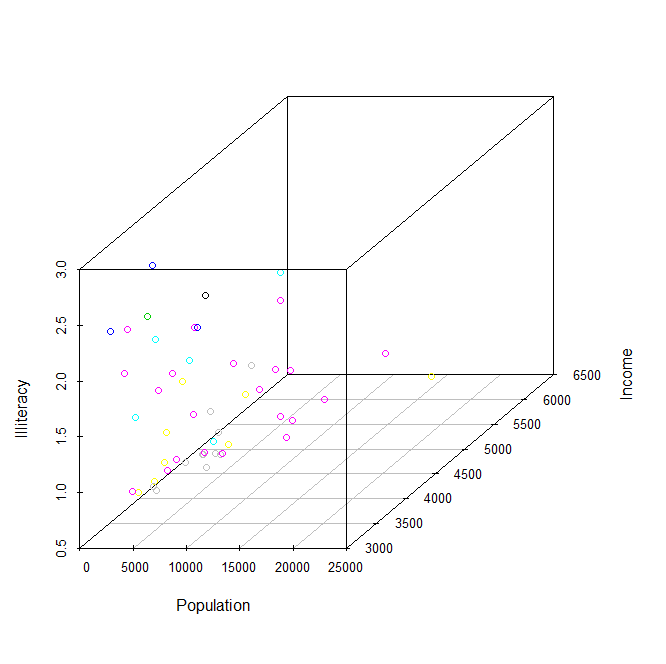
Use the base plotting functions from the graphics package. Show the code you used to generate the graphs. The precip data contains the average amount of precipitation (rainfall) in inches for each of 70 United States (and Puerto Rico) cities. Be creative to generate some graphs and charts using either the precip data or another data set of your choice. What about using the scatterplot3d package?

**Answer 5**

using state.x77 data

state<-state.x77

scatterplot3d(state)

****

**Using data (trees)**

data(trees)

trees[1:2,]

s3d <- scatterplot3d(trees, type="h", highlight.3d=TRUE,

angle=55, scale.y=0.7, pch=16,

main="Example of scatterplot3d plot: Tree Data")

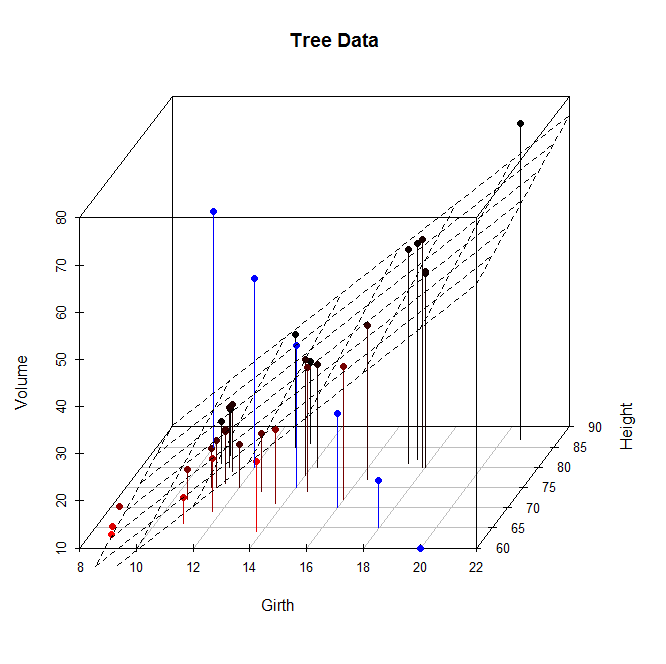
s3d$points3d(seq(10,20,2), seq(85,60,-5),

seq(60,10,-10), col="blue",

type="h", pch=16)

my.lm <- lm(Volume ~ Girth + Height)

s3d$plane3d(my.lm)

****

**Using mt cars data**

with(mtcars, {

   scatterplot3d(disp, wt, mpg,

                 color="blue", pch=19,

                 type="h",

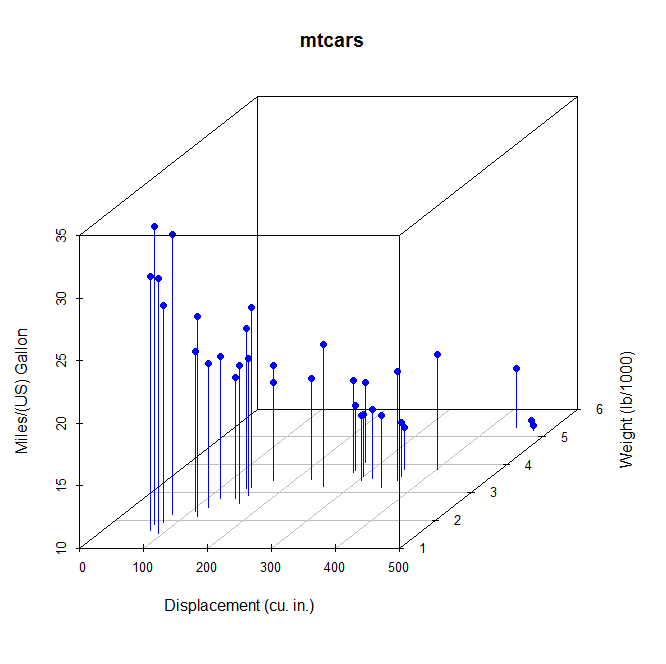
                 main="mtcars",

                 xlab="Displacement (cu. in.)",

                 ylab="Weight (lb/1000)",

                 zlab="Miles/(US) Gallon")

})



with(mtcars, {

s3d <- scatterplot3d(disp, wt, mpg,

color="blue", pch=19,

type="h",

main="mtcars",

xlab="Displacement (cu. in.)",

ylab="Weight (lb/1000)",

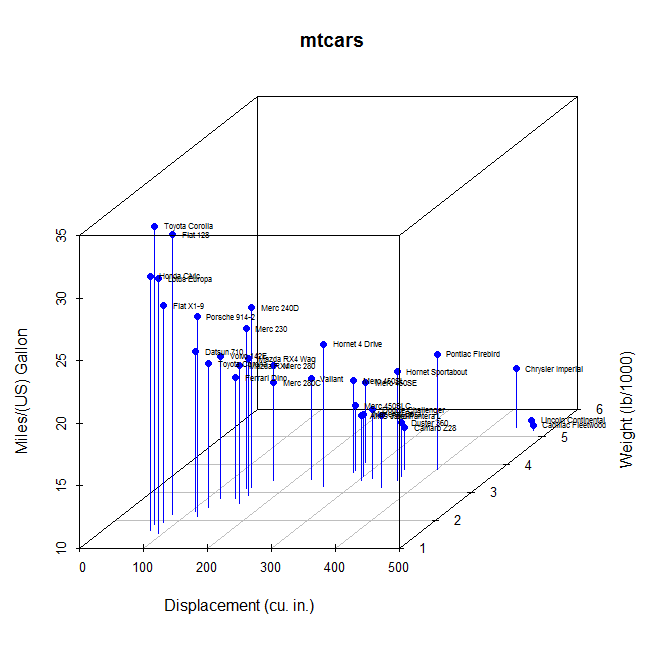
zlab="Miles/(US) Gallon")

s3d.coords <- s3d$xyz.convert(disp, wt, mpg)

labels=row.names(mtcars),

cex=.5, pos=4) # shrink text 50% and place to right of points)

})



mtcars$pcolor[mtcars$cyl==4] <- "red"

mtcars$pcolor[mtcars$cyl==6] <- "blue"

mtcars$pcolor[mtcars$cyl==8] <- "darkgreen"

with(mtcars, {

s3d <- scatterplot3d(disp, wt, mpg,

color=pcolor, pch=19,

type="h", lty.hplot=2,

scale.y=.75,

main="mtcars",

xlab="Displacement (cu. in.)",

ylab="Weight (lb/1000)",

zlab="Miles/(US) Gallon")

s3d.coords <- s3d$xyz.convert(disp, wt, mpg)

text(s3d.coords$x, s3d.coords$y,

labels=row.names(mtcars),

pos=4, cex=.5)

legend("topleft", inset=.05,

bty="n", cex=.5,

title="Number of Cylinders",

c("4", "6", "8"), fill=c("red", "blue", "darkgreen"))

})

