```
#Some commands for stat seminar series talk about R and RStudio

#Normal curve probabilities using R
par(mfrow = c(1,1))
plotDist("norm", col="blue", mean=2, xlim=c(-4,8), lwd=3)
plotDist("norm", col="green", kind='histogram', add=TRUE, mean=5)
plotDist("norm", col="red", kind='histogram', under=TRUE, mean=0)
#normal probabilities
xpnorm(80, mean=100, sd=15)
xpnorm(c(80,110), mean=100, sd=15)
xqnorm(.9,mean=100,sd=15)
xpnorm(68,mean=65,sd=3,manipulate=TRUE) #gives three sliders
par(mfrow = c(1,1))
par(mar=c(4,4,1,2)+0.1) #control margins in plots

## Overlapping points (continuous)
x1 <- c(rnorm(10000,0,1.25),rnorm(10000,-1,.25)) ; y1 <-
c(rnorm(10000,0,1.25),rnorm(10000,-1,.25));
plot(x1, y1)
plot(x1, y1,
     pch='.', col=rgb(0,0,0,0.2))
smoothScatter(x1, y1, nbin=200)
rm(x1, y1)

##Using R and mosaic package
require(mosaic) #get the mosaic package
require(mosaicData) #get the package of mosaid data sets
data(HELPrct) #get the HELPrct data frame
?HELPrct #What is the HELPrct data frame?
names(HELPrct) #What are the variables?
str(HELPrct) #What is the structure?
head(HELPrct) #see the first 6 rows of data
options(digits = 3) #suggest(!) thta only 3 digits be printed
#goal(y ~ x|z, data=mydata)
tally(~sex, data=HELPrct) #tally the sex variable in the HELPrct data
frame
tally(~substance, data=HELPrct)
tally(~sex | substance, data=HELPrct) #sex breakdown by substance
xtabs(~sex + substance, data=HELPrct) #same thing
mosaicplot(~sex + substance, data=HELPrct) #graph the data
xchisq.test(xtabs(~sex + substance, data=HELPrct)) #test of
independence

library(mosaicData)
require(ggplot2)
data(SaratogaHouses) #This command loads the SaratogaHouses data
SaraH <- SaratogaHouses #This renames the SaratogaHouses data, to save
on future typing
mPlot(SaraH)
```

```
library(openintro) #grab the openintro package
edaPlot(SaraH) #play around with plots using edaPlot from the
openintro package
par(mfrow=c(1,1)) #return to a 1x1 display area for graphs
par(mar=c(5,5,1,2)+0.1) #control margins in plots

NCHS3 <- read.csv("https://blackboard.oberlin.edu/bbcswebdav/
xid-1346993_1")
mPlot(NCHS3)
ggplot(data=NCHS3, aes(x=age, y=height)) + geom_point()  +
aes(colour=sex)  + stat_smooth(method=loess) +
theme(legend.position="top") + labs(title="")

Births1978 <- read.csv("https://blackboard.oberlin.edu/bbcswebdav/
xid-1346994_1")
mPlot(Births1978)

ggplot(data=Births1978, aes(x=dayofyear, y=resids)) + geom_point()   +
stat_smooth(method=loess) + theme(legend.position="none") +
labs(title="")

###testosterone ANOVA in R
par(mfrow = c(1,1))
Dabbs <- read.csv("https://blackboard.oberlin.edu/bbcswebdav/
xid-1354051_1")
str(Dabbs)
par(mar=c(5,5,2,1))
boxplot(Dabbs$testosterone~Dabbs$occupation)
library(lattice)
xyplot(Dabbs$testosterone~Dabbs$occupation,pch=16,cex=1.3)
options(digits=3)
tapply(Dabbs$testosterone,Dabbs$occupation,mean)
tapply(Dabbs$testosterone,Dabbs$occupation,sd)
summary(aov(Dabbs$testosterone~Dabbs$occupation))
mod1=aov(Dabbs$testosterone~Dabbs$occupation)
summary(mod1)
plot(mod1,which=1)
plot(mod1,which=2)
TukeyHSD(mod1)

##New zagat data from Agresti's book
data(zagat,package="smss")
str(zagat)
require(mosaic)
ggplot(zagat,aes(City)) + geom_bar() + ylab("") + xlab("") +
ggtitle("Numbers of restaurants rated")
ggplot(data=zagat, aes(x=City, y=Cost)) + geom_boxplot() +
labs(title="Average cost of a meal")

zagatX <- zagat %>% filter(Food > 2)
```

```
ggplot(zagatX,aes(Cost, Food)) + geom_point() + facet_wrap(~City) +
stat_smooth(method=lm)

#Now using mPlot on my own:
ggplot(data=zagatX, aes(x=Cost, y=Food)) + geom_point()  +
facet_wrap(~City, ncol=4)  + stat_smooth(method=lm) +
theme(legend.position="none") + labs(title="")

library("GGally")
ggpairs(zagatX, columns=c(1,3:6), upper = list(continuous = "cor",
combo = "facetdensity"), lower =
          list(continuous = "points", combo = "box"),
diag=list(continuous = "density", discrete = "bar"))

## rflip and do() example
rflip(4,prob=.7) #make 4 draws from a 70-30 population
do(10)*rflip(4,prob=0.7) #do that 10 times
mysim <- do(1000)*rflip(40,prob=0.7) #do that 1000 times and save the
results
hist(mysim$heads) #look at the results

## another do() example
library(Stat2Data)
data("FirstYearGPA")
FY <- FirstYearGPA
lm(GPA ~ SATM, data=FY) #gives 0.0012 as slope
lm(GPA ~ SATM, data=FY)$coeff[2] #just look at the slope
do(3)*lm(GPA ~ shuffle(SATM), data=FY)$coeff[2] #break link b/w GPA
and     SATM
null.dist <- do(1000)*lm(GPA ~ shuffle(SATM), data=FY)$coeff[2] #1000
random slopes
histogram(null.dist$SATM, v=0.0012) #look at the 1000 slopes
with(null.dist, tally(abs(SATM.)>=0.0012)) #How many are far from
zero?
with(null.dist, tally(abs(SATM.)>=0.0012, format='prop')) #What
proportion are   far from zero?
```