##

## Synchronous Session 6

##

#

# ggplot Basics

#

# histogram

#

ggplot(mtcars,aes(x=mpg))+geom\_histogram(bins=10,color='black',fill="white")

ggplot(mtcars,aes(x=mpg))+geom\_histogram(bins=5,color='black',fill="pink")

#

# boxplot

#

ggplot(mtcars,aes(x=cyl,y=mpg))+geom\_boxplot()

ggplot(mtcars,aes(x=as.factor(cyl),y=mpg))+geom\_boxplot()

ggplot(mtcars,aes(x=as.factor(cyl),y=mpg))+geom\_boxplot(aes(fill=factor(cyl)))

#

# bar

#

ggplot(mtcars,aes(x=cyl,y=mpg))+geom\_bar(stat="identity")

ggplot(mtcars,aes(x=cyl,y=mpg))+geom\_bar(stat="identity", color="black",fill="blue")

#

# what if wanted avg mgp per cyl

# use tapply,,, df, group, function

#

ave.mpg<-tapply(mtcars$mpg,mtcars$cyl,mean)

cylName<-unlist(labels(ave.mpg))

df<-data.frame(ave.mpg,cylName)

df

#

ggplot(df,aes(x=cylName,y=ave.mpg))+geom\_bar(stat="identity")

ggplot(df,aes(x=cylName,y=ave.mpg))+geom\_bar(stat="identity", color="black",fill="blue")

ggplot(df,aes(x=cylName,y=ave.mpg))+geom\_bar(stat="identity", color="black",fill="blue")+

ggtitle("avg mpg per cyl")

ggplot(df,aes(x=cylName,y=ave.mpg))+geom\_bar(stat="identity", color="black",fill="blue")+

ggtitle("avg mpg per cyl")+theme(plot.title=element\_text(hjust=0.5))

#

#

#

#############################

#

# line charts

#

ggplot(df,aes(x=cylName,y=ave.mpg, group=1)) +

geom\_line()

ggplot(df,aes(x=cylName,y=ave.mpg, group=1)) +

geom\_line(color="blue", size=2)

ggplot(df,aes(x=cylName,y=ave.mpg, group=1)) +

geom\_line(color="red", size=2) +

geom\_point(color="blue")

#

# layer approach

#

g<-ggplot(df,aes(x=cylName,y=ave.mpg, group=1))

g+geom\_point(color="blue")

g+geom\_line(color="red", size=2)

g+geom\_point(color="blue")+geom\_line(color="red", size=2)

g+geom\_point(color="blue",size=3)+geom\_line(color="red", size=2) # line layer on top of point layer

g+geom\_line(color="red", size=2)+geom\_point(color="blue",size=3) # point layer on top of line layer

g+geom\_point(color="blue",size=3)+geom\_line(color="red", size=2)+ggtitle("test")

#######################################

#

# multiple lines - need to create a new dataframe

# show one line for am==1, another for am==0

# automatic 1 = yes, 0 = no

# create 2 df's

#

amYes <- mtcars[mtcars$am==1, ] # df w/ am=1

amNo <- mtcars[mtcars$am==0, ] # df w/ am=0

# lets look at both dfs

amYes

amNo

#

# calc avg mpg for amYes and amNo similar to earlier tapply example

#

amyes.ave.mpg <- tapply(amYes$mpg, amYes$cyl, mean) # col, group, function

amno.ave.mpg <- tapply(amNo$mpg, amNo$cyl, mean)

#

# let's look at both

#

amyes.ave.mpg

amno.ave.mpg

#

#make one dataframe, but with all the mpg in one column

#

ave.mpg <- c(amno.ave.mpg, amyes.ave.mpg)

ave.mpg

cylNames <- c(cylName, cylName )

cylNames

am <- c(0,0,0,1,1,1)

df <- data.frame(ave.mpg, cylNames, am)

df

# now we can plot some lines

#

ggplot(df,aes(x=cylNames,y=ave.mpg, group=am)) +

geom\_line(color="red", size=2)

ggplot(df,aes(x=cylNames,y=ave.mpg, group=am, color=am)) +

geom\_line(size=2)

ggplot(df,aes(x=cylNames,y=ave.mpg, group=am, color=factor(am))) +

geom\_line(size=2)

ggplot(df, aes(x = cylNames, y=ave.mpg, group=am)) + # side by side

geom\_bar(stat="identity", position="dodge", color="black", aes(fill=am))

ggplot(df, aes(x = cylNames, y=ave.mpg, group=am)) + # side by side

geom\_bar(stat="identity", position="dodge", color="black", aes(fill=factor(am)))

ggplot(df, aes(x = cylNames, y=ave.mpg, group=am)) + # stacked

geom\_bar(stat="identity", color="black", aes(fill=am))

ggplot(df, aes(x = cylNames, y=ave.mpg, group=am)) + # stacked

geom\_bar(stat="identity", color="black", aes(fill=factor(am)))

ggplot(df,aes(x=cylNames,y=ave.mpg, group=am)) +

geom\_bar(stat="identity", position="dodge", color="black", aes(fill=am)) +

geom\_line(color="red", size=2) +

geom\_point(color="blue")

#################

#

# scatter plots

#

#

ggplot(mtcars, aes(x=hp, y=mpg)) + geom\_point()

#add size of bubble -

ggplot(mtcars, aes(x=hp, y=mpg)) + geom\_point(aes(size=hp))

ggplot(mtcars, aes(x=hp, y=mpg)) + geom\_point(aes(size=wt)) # 3 variables,

# as hp incr, wt incr, mpg decr

#add color of bubble

ggplot(mtcars, aes(x=hp, y=mpg)) + geom\_point(aes(size=wt,color=am)) # 4 variables

ggplot(mtcars, aes(x=hp, y=mpg)) + geom\_point(aes(size=wt,color=factor(am))) # 4 variables

ggplot(mtcars, aes(x=hp, y=mpg)) + geom\_point(aes(size=wt,color=am, shape=as.factor(cyl))) # 5 variables

ggplot(mtcars, aes(x=hp, y=mpg)) + geom\_point(aes(size=wt,color=factor(am), shape=as.factor(cyl))) # 5 variables

ggplot(mtcars, aes(x=hp, y=mpg)) + geom\_point(aes(size=wt,color=factor(am), shape=as.factor(cyl))) + # 5 variables

geom\_jitter()

# name the dots

#

mtcars # note car name is a row name

mt <- mtcars # assign mtcars to another df

mt$name <- unlist(rownames(mt)) # take rowname and create a col $name in df mt

mt # look at mt & verify name has been added

ggplot(mt, aes(x=hp, y=mpg)) +

geom\_point(aes(size=wt,color=am, shape=as.factor(cyl))) +

geom\_text(aes(label=name), size=3)

#look at a subset of dots

subMt <- mt[mt$mpg > 18 & mt$disp > 250,] # all the rows that fit this condition

ggplot(mt, aes(x=hp, y=mpg)) + # specify new df subMt

geom\_point(aes(size=wt,color=am, shape=as.factor(cyl))) +

geom\_text(data= subMt, aes(label=name), size=3)

mt[which(mt$name=="Hornet 4 Drive"), ] # all the attributes which gives the index

################

#

# heatmaps - use geom\_tile

#

ggplot(mtcars, aes(x=mtcars$mpg, y=mt$cyl)) +

geom\_tile(aes(fill=mtcars$hp))

ggplot(mtcars, aes(x=mtcars$mpg, y=mt$cyl)) +

geom\_tile(aes(fill=mtcars$hp)) +

scale\_fill\_gradient(low="white", high="blue")

#explore mpg rounded

mt <- mtcars

mt$roundMPG <- round(mt$mpg)

mt

ggplot(mt, aes(x=mt$roundMPG, y=mt$cyl)) +

geom\_tile(aes(fill=mt$hp)) +

scale\_fill\_gradient(low="white", high="blue")

#more rounding

mt$roundMPG <- round(mt$mpg,-1)

ggplot(mt, aes(x=mt$roundMPG, y=mt$cyl)) +

geom\_tile(aes(fill=mt$hp)) + scale\_fill\_gradient(low="white", high="blue")