```
#=====================================================================
#  file:          c:\Projects\CBP\Rcourse\CommonAnalysesRbit003.r
#  function:      common analyses, snook data
#
#  programmer:    Elgin S. Perry, Ph. D.
#
#  date:          6/1/2014
#
#  address:       2000 Kings Landing Rd.
#                 Huntingtown,  Md.  20639
#
#  voice phone:   (410)535-2949
#  email:         EPERRY@chesapeake.net
#=====================================================================

#install.packages()
#library(lattice) #Used for contour plots [contourplot()]
#library(nlme)    #used for gam Mixed model [gamm()]
#library(MASS)    #used for glm Mixed model [glmmPQL()]
#library(mgcv)    #Wood's gam package
#library(chron)   #date functions
#library(doBy)    # Allows "BY processing similar to SAS
#library(FitAR)   #AR package from McLeod and Zhang
#library(Hmisc)   #stat function by Frank Harrell
#library(cluster) #cluster analysis routines
options(stringsAsFactors = FALSE)

source("C:/Projects/Rtp/dfsum.r")
source("C:/Projects/Rtp/DistFunct.r")
#source("C:/Projects/Rtp/RTF.r")

#  be sure to change \ to /
ProjRoot <- 'c:/Projects/CBP/Rcourse/'
setwd(ProjRoot);
RTFout <- paste(ProjRoot,"RTFexample.rtf",sep='')

datafile <- paste(ProjRoot,"snook.tdf",sep='');
snook <- read.table(datafile, header=TRUE, sep="\t", na.strings="NA", dec=".",
strip.white=TRUE,stringsAsFactors = FALSE)
dfsum(snook)
snook[snook$length==40 & snook$water.body=='Atlantic'&snook$season=="May-Oct",'wgt.mean'] <-
NA
#[1] "length"     "water.body" "season"     "wgt.mean"   "wgt.min"    "wgt.max"

# scatter plots and box plots
plot(snook$length,snook$wgt.mean)

snook$wbcol <- ifelse(snook$water.body=='Gulf','green','blue')
plot(snook$length,snook$wgt.mean,xlab='Length',ylab='Mean
Weight',main='Snook',col=snook$wbcol)
#legend("topleft",c('Gulf','Atlantic'), col=c('green','blue'),pch=21)
legend(22,30,c('Gulf','Atlantic'), col=c('green','blue'),pch=21)

points(snook$length,snook$wgt.min,pch=24, col=c('green','blue'))
points(snook$length,snook$wgt.max,pch=25, col=c('green','blue'))


boxplot(wgt.mean ~ water.body,data=snook,outline=TRUE)
boxplot(wgt.mean ~ water.body,data=snook,outline=TRUE,ylab='Mean Weight',main='Snook')

# simple statistics
unique(snook$water.body)
unique(snook$season)
mean(snook$wgt.mean)
```

```
mean(snook$wgt.mean,na.rm=TRUE)
sd(snook$wgt.mean,na.rm=TRUE)
min(snook$wgt.mean,na.rm=TRUE)
max(snook$wgt.mean,na.rm=TRUE)
range(snook$wgt.mean,na.rm=TRUE)

snook <- snook[!is.na(snook$wgt.mean),] # eliminate the missing value record
DistSum(snook$wgt.mean) -> ds.mean.wgt
ds.mean.wgt

# frequency tables
table(snook$water.body)
table(snook$water.body, snook$season)

snook$Season <- ifelse(snook$season=='Apr-Sep' | snook$season=='May-Oct','summer','winter')
table(snook$water.body, snook$Season)
a<-table(snook$water.body, snook$Season)
a
a[1,2]
a[1,]
a[,2]

addmargins(table(snook$water.body, snook$Season))

chisq.test(snook$water.body, snook$Season)

# correlation
cor(snook$length,snook$wgt.mean)
cor.test(snook$length,snook$wgt.mean,method='spearman')

cor(snook[,4:6],use="complete.obs")


#linear regression
lm(wgt.mean~length,data=snook)
lm1 <- lm(wgt.mean~length,data=snook)
# context sensitive functions
summary(lm1)
anova(lm1)
plot(lm1)

layout(matrix(1:4, ncol = 2))# sets up for multiple plots per page
#par(mfrow=c(2,2)) # sets up for multiple plots per page
plot(lm1)
layout(1)

# other model diagnostics
plot(resid(lm1))
shapiro.test(resid(lm1))
boxplot(resid(lm1) ~ snook$water.body)
hist(resid(lm1))
DistPlot(resid(lm1))

# plot of data and predicted values
plot(snook$length,snook$wgt.mean,xlab='Length',ylab='Mean
Weight',main='Snook',col=snook$wbcol)
lines(snook$length,predict(lm1),lwd=2,col='red')


# log-linear regression
lm2 <- lm(log(wgt.mean)~length,data=snook)
summary(lm2)
par(mfrow=c(2,2)) # sets up for multiple plots per page
plot(lm2)
par(mfrow=c(1,1))
```

```
plot(snook$length,snook$wgt.mean,xlab='Length',ylab='Mean
Weight',main='Snook',col=snook$wbcol)
lines(snook$length,exp(predict(lm2)),lwd=2,col='red')

# loess regression
lm3 <- loess(wgt.mean~length,data=snook)
summary(lm3)
plot(snook$length,snook$wgt.mean,xlab='Length',ylab='Mean
Weight',main='Snook',col=snook$wbcol)
lines(snook$length,exp(predict(lm2)),lwd=2,col='red')
lines(snook$length,predict(lm3),lwd=2,col='brown')

lm4 <- loess(wgt.mean~length,data=snook,span=0.5)
anova(lm4,lm3)
```