###

### Dataframe Stuff

#

age<-c(40,50,60)

wt<-c(100,130,150)

myDf <-data.frame(age,wt)

myDf # note: Observations (rows), Variables (columns)

age<-age+3

age

myDf # note: age in df did not change

myDf[1,2]

myDf[2,2]

myDf[1,]

myDf[,2]

myNewData<-myDf[,2]

myNewData

mean(myDf$age) # dataframe variable age

age

myDf$newAge<-age # add age+3 vector to myDf, creates a new column (variable) newAge

myDf

myDfNew<-rbind(myDf,c(10,12,13)) # add a new row

myDfNew

myDfNew[order(myDfNew$age),] # sort ascending

order(myDfNew$age) # just positional values,, didn't change the Df

myDfNewSorted<-myDfNew[order(myDfNew$age),] # save sorted Df

myDfNewSorted

row.names(myDfNewSorted) <- 1:nrow(myDfNewSorted) # rename row names

# or rownames(df) <- NULL

myDfNewSorted

#

# Helpful R Commands ppt

# List all the datasets that come with R package "datasets"

#

data()

#

#

# http://www.r-tutor.com/r-introduction/data-frame

# https://www.r-bloggers.com/working-with-the-data-frame-in-r/

# https://www.r-bloggers.com/5-ways-to-subset-a-data-frame-in-r/

#

mtcars

str(mtcars)

# ?mtcars

myCars<-mtcars

myCars[1,]

myCars[1,3]

myCars[,3]

mean(myCars$mpg)

# how nany different ways can we retrieve the content of the 9th col or "am" variable ?

# [, 9] am Transmission (0 = automatic, 1 = manual)

#

# to retrieve the ninth column vector of the built-in data set mtcars, we write mtcars[[9]].

myCars[[9]]

# to retrieve the ninth column

myCars[,9]

# We can retrieve the same column vector by its name.

myCars[["am"]]

# We can also retrieve with the "$" operator in lieu of the double square bracket operator.

myCars$am

#Yet another way to retrieve the same column vector is to use the single square bracket "[]" operator. We prepend the column name with a comma character, which signals a wildcard match for the row position.

myCars[,"am"]

# and finally

myCars[9]

head(myCars)

tail(myCars)

tail(myCars,10)

meanMpg<-mean(myCars$mpg)

meanMpg

myCars[myCars$mpg>meanMpg,]

myCars$mpg>meanMpg # TRUE TRUE TRUE TRUE FALSE FALSE FALSE TRUE

myCars

# print on 1 line

#

cat("mean mpg :", meanMpg)

# or

cat("max mpg :", mean(myCars$mpg))

#

# ---------------- Chapter 5: Rows and Columns -----------------

myFamilyNames <- c("Dad","Mom","Sis","Bro","Dog")

myFamilyNames

myFamilyAges <- c(43, 42, 12, 8, 5)

myFamilyAges

myFamilyAges[2]

myFamilyGenders <- c("Male","Female","Female","Male","Female")

myFamilyWeights <- c(188,136,83,61,44)

myFamily <- data.frame(myFamilyNames, myFamilyAges, myFamilyGenders, myFamilyWeights)

myFamily

str(myFamily)

summary(myFamily)

myFamily$myFamilyAges

myFamilyAges <- c(myFamilyAges, 11)

myFamilyAges

head(myFamilyAges, 2)

tail(myFamilyAges, 2)

myFamily$myFamilyAges

myFamily <- data.frame(myFamilyNames, myFamilyAges, myFamilyGenders, myFamilyWeights)

#

# max

# which.max

# row.names

max(mtcars$hp)

which.max(mtcars$hp)

row.names(mtcars)

row.names(mtcars[1,])

row.names(mtcars[4,])