*# \*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*#*

*# \*\*\*Name: Class 1 Presentation Code\*\*\*\*\*\*\*\*\*#*

*# \*\*\*Author: R Tutor\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*#*

*# \*\*\*Date: March 10, 2016\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*#*

*# \*\*\*Require: baseballFull.csv\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*#*

*# \*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*#*

*#remove all existing objects from workspace*

[**rm**](http://inside-r.org/r-doc/base/rm)([**list**](http://inside-r.org/r-doc/base/list)=[**ls**](http://inside-r.org/r-doc/base/ls)())

*####Introduction#####*

*#Example Script*

1+1

*#create a very simple object, a, with numeric elements 1 and 2*

a <- [**c**](http://inside-r.org/r-doc/base/c)(1,2)

a

A

*#Setting Working Directory*

input <- "//AGBOSNAS1/DATA/SHARE/R/2016/Class 1/Code & Examples/Input"

output <- "//AGBOSNAS1/DATA/SHARE/R/2016/Class 1/Code & Examples/Output"

temp <- "//AGBOSNAS1/DATA/SHARE/R/2016/Class 1/Code & Examples/Temp"

[**setwd**](http://inside-r.org/r-doc/base/setwd)(input)

*#assuming the "Code & Examples" folder has the*

*#following subfolders: Input, Output, and Temp*

baseball <- [**read.csv**](http://inside-r.org/r-doc/utils/read.csv)("baseballFULL.csv", header=**TRUE**)

*#Example graphics*

*#installing and loading packages*

[**install.packages**](http://inside-r.org/r-doc/utils/install.packages)("rgl", dependencies=**TRUE**)

[**library**](http://inside-r.org/r-doc/base/library)([rgl](http://inside-r.org/packages/cran/rgl))

[**demo**](http://inside-r.org/r-doc/utils/demo)(bivar)

*#Help Example*

?[**lm**](http://inside-r.org/r-doc/stats/lm)

*#Tab example*

l

*####Data Structure####*

*####Summary of Data Structures####*

*#### A. Scalars & Vectors ####*

*#Vector Creation#*

**pi**

[**is.vector**](http://inside-r.org/r-doc/base/is.vector)(**pi**)

*#make object x with a vector containing elements 1,2,3,and 4*

x <- [**c**](http://inside-r.org/r-doc/base/c)(1,2,3,4)

x

*#using the combine function to create a vector from other vectors*

x1 <- [**c**](http://inside-r.org/r-doc/base/c)(3,2,1)

x2 <- [**c**](http://inside-r.org/r-doc/base/c)(4,5,6)

x3 <- [**c**](http://inside-r.org/r-doc/base/c)(7,8,9)

x4 <- [**c**](http://inside-r.org/r-doc/base/c)(x1,x2,x3)

x4

*#examples of using vectors with special properties*

*#a vector of consecutive integers*

y <- 1:4

y

*#a vector with repeating values, using the rep() function*

x <- [**rep**](http://inside-r.org/r-doc/base/rep)(1,10)

x

*#vector with elements that are characters*

AG\_Office <- [**c**](http://inside-r.org/r-doc/base/c)("Boston", "Chicago", "Dallas",

"Denver", "Los Angeles", "Menlo Park",

"Montreal", "New York", "San Francisco", "Washington")

AG\_Office

*#Indexing Elements of a vector; subvector <- Vector[index]#*

*#a: Index vector of positive integers*

AG\_Office[3]

*#gives us the second to last element of the vector AG\_Office*

second\_to\_last <- AG\_Office[**[length](http://inside-r.org/r-doc/base/length)**(AG\_Office)-1]

second\_to\_last

*#adding an office to the AG\_office vector*

AG\_Office <- [**c**](http://inside-r.org/r-doc/base/c)("Boston", "Chicago", "Dallas",

"Denver", "Los Angeles", "Menlo Park",

"Montreal", "New York", "San Francisco", "Washington",

"Beijing")

second\_to\_last <- AG\_Office[**[length](http://inside-r.org/r-doc/base/length)**(AG\_Office)-1]

second\_to\_last

*#trying to use an index value greater than length(x)*

AG\_Office[15]

*#b: index vector of negative integers; Subvector Returns all except 3rd element*

*#all of AG\_Office except for the third element*

AG\_Office\_Ex\_Dallas <- AG\_Office[-3]

AG\_Office\_Ex\_Dallas

*#c: vector of character string/names*

*#first, change names of positons in the vector;*

*#attach alphanumeric names to vector*

*#elements*

fruit\_prices <- [**c**](http://inside-r.org/r-doc/base/c)(1.1,3.5,5.4,2.3)

fruit\_prices

[**names**](http://inside-r.org/r-doc/base/names)(fruit\_prices) <- [**c**](http://inside-r.org/r-doc/base/c)("orange", "banana", "apple", "peach")

fruit\_prices

lunch <- fruit\_prices[[**c**](http://inside-r.org/r-doc/base/c)("apple", "orange")]

lunch

*#d: logical Vector -- conditional indexing*

*#vector that evaluates to boolean; Boston = #1 "AG\_Office" so 1st position evaluates to True*

is\_Boston\_Office <- AG\_Office=="Boston"

is\_Boston\_Office

AG\_Office[is\_Boston\_Office]

*#### Vector Functions ####*

*#functions are performed on vectors element by element -- like an internal loop*

a <- 1:10

a

a+1

*#summary statistics*

*#Create a vector of 100 random normally distributed values with mean 0 and standard deviation 1*

x <- [**rnorm**](http://inside-r.org/r-doc/stats/rnorm)(n=100,[**mean**](http://inside-r.org/r-doc/base/mean)=0,[**sd**](http://inside-r.org/r-doc/stats/sd)=1)

[**length**](http://inside-r.org/r-doc/base/length)(x)

[**summary**](http://inside-r.org/r-doc/base/summary)(x)

[**mean**](http://inside-r.org/r-doc/base/mean)(x)

[**sd**](http://inside-r.org/r-doc/stats/sd)(x)

*#arithmetic operations*

x <- [**c**](http://inside-r.org/r-doc/base/c)(1,2,3)

y <- [**c**](http://inside-r.org/r-doc/base/c)(4,5,6)

x+1

x+y

*#sample error*

x <- [**c**](http://inside-r.org/r-doc/base/c)(1,2,3,4)

y <- [**c**](http://inside-r.org/r-doc/base/c)(6,0,9,20,22,23)

x+y

*#using NA*

*#try to take the mean of a vector which contains a single NA value*

x <- [**c**](http://inside-r.org/r-doc/base/c)(100, **NA**, 200, 300, 400)

x

[**mean**](http://inside-r.org/r-doc/base/mean)(x)

[**mean**](http://inside-r.org/r-doc/base/mean)(x, na.rm=**TRUE**)

*#### B. Matrices ####*

*#combine vector x of numeric values with vector y of character values*

*#into a matrix called M*

x <- [**c**](http://inside-r.org/r-doc/base/c)(1,2,3)

y <- [**c**](http://inside-r.org/r-doc/base/c)("one", "two", "three")

M <- [**cbind**](http://inside-r.org/r-doc/base/cbind)(x,y)

M

*#### C. Lists ####*

*#redefine y for example*

x1 <- [**c**](http://inside-r.org/r-doc/base/c)(1,2,3)

x2 <- x1+3

x3 <- x1+4

y <- [**cbind**](http://inside-r.org/r-doc/base/cbind)(x1,x2,x3)

y

*#creating a list*

x <- [**c**](http://inside-r.org/r-doc/base/c)("one","two","three")

y

L <- [**list**](http://inside-r.org/r-doc/base/list)(x,y)

L

*#we can refer to list components by using one of the*

*#four types of index vectors*

*#requires double brackets instead of single ones*

L <- [**list**](http://inside-r.org/r-doc/base/list)([**vector**](http://inside-r.org/r-doc/base/vector)=x,[**matrix**](http://inside-r.org/r-doc/base/matrix)=y)

L

L[[1]]

L[["vector"]]

*#refer to list components by their names using the $ sign*

L$vector

L$vector[1]

*#R saves regression results into a list with many components*

*#regress salary on years, runs*

reg1 <- [**lm**](http://inside-r.org/r-doc/stats/lm)(salary~years+runs, [**data**](http://inside-r.org/r-doc/utils/data) = baseball)

reg1\_summary <- [**summary**](http://inside-r.org/r-doc/base/summary)(reg1)

*#seeing if reg1 and reg1\_summary is a list*

[**is.list**](http://inside-r.org/r-doc/base/is.list)(reg1)

[**is.list**](http://inside-r.org/r-doc/base/is.list)(reg1\_summary)

*#use the names() function to get a list of reg1's components*

*#list of reg1\_summary components using the names() function*

[**names**](http://inside-r.org/r-doc/base/names)(reg1\_summary)

*#get the matrix of coefficients*

reg1\_coef <- reg1\_summary$coefficients

reg1\_coef

*#### D. Data Frames ####*

*#Basic Data Frame*

x1 <- [**c**](http://inside-r.org/r-doc/base/c)(1,2,3)

x2 <- [**c**](http://inside-r.org/r-doc/base/c)("AB","BC","CD")

[**D**](http://inside-r.org/r-doc/stats/D) <- [**data.frame**](http://inside-r.org/r-doc/base/data.frame)(x1,x2)

[**D**](http://inside-r.org/r-doc/stats/D)

*#Reading in Dataframes*

*#1. General command: read.table()*

input <- "//AGBOSNAS1/DATA/SHARE/R/2016/Class 1/Code & Examples/Input"

[**setwd**](http://inside-r.org/r-doc/base/setwd)(input)

mytable <- [**read.table**](http://inside-r.org/r-doc/utils/read.table)("baseballFull.csv", header=**TRUE**,sep=",")

*#2. Specialized command: read.csv()*

baseball <- [**read.csv**](http://inside-r.org/r-doc/utils/read.csv)("baseballFull.csv")

*#Saving / Reading an R data frame*

*#save permanent R dataset called BaseballDataset*

temp <- "//agbosnas1/data/Share/R/2016/Class 1/Code & Examples/Temp"

[**setwd**](http://inside-r.org/r-doc/base/setwd)(temp)

[**save**](http://inside-r.org/r-doc/base/save)( baseball, [**file**](http://inside-r.org/r-doc/base/file) = "BaseballDataset.Rda" )

*#read in BaseballDataset*

[**load**](http://inside-r.org/r-doc/base/load)("BaseballDataset.Rda")

*#Accessing Data within Dataframes*

*#Method A: Similarly to a vector/matrix*

*#Access row 1, column 5*

baseball[1,5]

*#Access columns 2 and 3*

baseball[ , [**c**](http://inside-r.org/r-doc/base/c)(2,3)]

*#Access all rows except for row 2*

baseball[-[**c**](http://inside-r.org/r-doc/base/c)(2),]

*#Method B: Using dataframe properties*

*#Find the names of the dataframe's variables*

[**names**](http://inside-r.org/r-doc/base/names)(baseball)

*#Once we know the names, we can index datasets using the name*

baseball[,"team"]

*#...or we can use the $ syntax*

baseball$team

*#Tab example*

*#if multiple possible variables, will provide a list*

baseball$

*#if only one possible variable, will auto fill*

baseball$n

*#Creating variables in a Dataframe*

baseball$combined\_runs <- baseball$runs + baseball$homer

baseball$combined\_runs

*#creating new variables with conditions based on existing variables*

*#will be covered in Class 2*

*#Renaming variables*

*#create a small example data frame*

Var1 <- [**c**](http://inside-r.org/r-doc/base/c)(1,2,3,4)

Var2 <- [**c**](http://inside-r.org/r-doc/base/c)("A","B","C","D")

Example\_Data <- [**data.frame**](http://inside-r.org/r-doc/base/data.frame)(Var1,Var2)

*#rename variables in data frame*

[**names**](http://inside-r.org/r-doc/base/names)(Example\_Data) <- [**c**](http://inside-r.org/r-doc/base/c)("Numbers","Letters")

Example\_Data

*#install and load reshape package*

[**install.packages**](http://inside-r.org/r-doc/utils/install.packages)("reshape")

[**library**](http://inside-r.org/r-doc/base/library)("reshape")

*#view syntax for rename*

rename

*#rename the "Numbers" variable*

Example\_Data <- rename(Example\_Data, [**c**](http://inside-r.org/r-doc/base/c)(Numbers="Numbers\_new"))

Example\_Data

*#Subsetting a Dataframe*

*#Method A: Index Syntax*

*#Select all obs whose team is BOS*

baseballBOS <- baseball[baseball$team == "BOS", ]

[**which**](http://inside-r.org/r-doc/base/which)(baseball$team == "BOS" )

baseballBOS <- baseball[[**which**](http://inside-r.org/r-doc/base/which)( baseball$team == "BOS" ), ]

*#Select all obs whose team is not BOS*

baseballNoBOS <- baseball[-[**which**](http://inside-r.org/r-doc/base/which)( baseball$team == "BOS" ), ]

*#Method B: Subset function*

*#subset to observations whose team is BOS*

baseballBOS <- [**subset**](http://inside-r.org/r-doc/base/subset)( baseball, baseball$team == "BOS" )

*#subset to observations whose team is BOS and only keep name, team, position, and hits variables*

baseballBOS2 <- [**subset**](http://inside-r.org/r-doc/base/subset)( baseball, baseball$team == "BOS",

[**c**](http://inside-r.org/r-doc/base/c)(name, team, position, hits) )

*#variable dropping using subset*

baseball\_sub <- [**subset**](http://inside-r.org/r-doc/base/subset)(baseball, select=-[**c**](http://inside-r.org/r-doc/base/c)(name, salary))

*#### Factors ####*

*#Creating a factor from a numeric vector*

x <- [**c**](http://inside-r.org/r-doc/base/c)(1,2,3,3,4,5,5,6)

xf <- [**factor**](http://inside-r.org/r-doc/base/factor)(x)

x

xf

[**levels**](http://inside-r.org/r-doc/base/levels)(xf)

*#importance of factors: used very often in categorical analysis*

*#summary statistics by group: aggregate(), by(), tapply() covered in Class 2*

*#categorical analyses: ANOVA, Chi-Square Test*

*#Factors v Character Variables*

*#create vectors - one character, and one factor*

character\_vector <- [**c**](http://inside-r.org/r-doc/base/c)("BIRD","MOLE","FISH","MYNOCK")

factor\_vector <- [**as.factor**](http://inside-r.org/r-doc/base/as.factor)([**c**](http://inside-r.org/r-doc/base/c)("SKY","EARTH","SEA","SPACE"))

character\_vector

factor\_vector

*#concatenating these vectors returns the expected results*

[**paste**](http://inside-r.org/r-doc/base/paste)(character\_vector,factor\_vector)

*#attempt to create matrix with these two vectors*

[**matrix**](http://inside-r.org/r-doc/base/matrix) <- [**cbind**](http://inside-r.org/r-doc/base/cbind)(character\_vector,factor\_vector)

[**matrix**](http://inside-r.org/r-doc/base/matrix)

*#to combine as a matrix while keeping the desired strings*

matrix2 <- [**cbind**](http://inside-r.org/r-doc/base/cbind)(character\_vector,[**as.character**](http://inside-r.org/r-doc/base/as.character)(factor\_vector))

matrix2