3 Data Structures - 26 Nov 2020

Dhruv Aggarwal

12/13/2020

# Data Frames(DF) - Most useful features of R & also cited reason for R's ease of use.  
# In dataframe, each column is actually a vector, each of which has same length.  
# Each column can hold different type of data.  
# Also within each column, each element must be of same type, like vectors.  
  
# Creating a Dataframe from vectors  
  
x = 10:1  
y = -4:5  
q = c("Hockey","Football","Baseball","Curlin","Rugby","Lacrosse",  
 "Basketball","Tennis","Cricket","Soccer")  
theDF = data.frame(x,y,q) # this would create a 10x3 data.frame with x, y and q as variable names  
theDF

## x y q  
## 1 10 -4 Hockey  
## 2 9 -3 Football  
## 3 8 -2 Baseball  
## 4 7 -1 Curlin  
## 5 6 0 Rugby  
## 6 5 1 Lacrosse  
## 7 4 2 Basketball  
## 8 3 3 Tennis  
## 9 2 4 Cricket  
## 10 1 5 Soccer

# Assigning Names  
theDF = data.frame (First=x, Second =y, Sport = q)  
theDF

## First Second Sport  
## 1 10 -4 Hockey  
## 2 9 -3 Football  
## 3 8 -2 Baseball  
## 4 7 -1 Curlin  
## 5 6 0 Rugby  
## 6 5 1 Lacrosse  
## 7 4 2 Basketball  
## 8 3 3 Tennis  
## 9 2 4 Cricket  
## 10 1 5 Soccer

# Checking the dimensions of the DF.  
nrow(theDF)

## [1] 10

ncol(theDF)

## [1] 3

dim(theDF)

## [1] 10 3

names (theDF)

## [1] "First" "Second" "Sport"

names(theDF)[3]

## [1] "Sport"

rownames(theDF)

## [1] "1" "2" "3" "4" "5" "6" "7" "8" "9" "10"

# Head and Tail  
head(theDF)

## First Second Sport  
## 1 10 -4 Hockey  
## 2 9 -3 Football  
## 3 8 -2 Baseball  
## 4 7 -1 Curlin  
## 5 6 0 Rugby  
## 6 5 1 Lacrosse

head(theDF, n=7)

## First Second Sport  
## 1 10 -4 Hockey  
## 2 9 -3 Football  
## 3 8 -2 Baseball  
## 4 7 -1 Curlin  
## 5 6 0 Rugby  
## 6 5 1 Lacrosse  
## 7 4 2 Basketball

tail(theDF)

## First Second Sport  
## 5 6 0 Rugby  
## 6 5 1 Lacrosse  
## 7 4 2 Basketball  
## 8 3 3 Tennis  
## 9 2 4 Cricket  
## 10 1 5 Soccer

class(theDF)

## [1] "data.frame"

# Accessing Individual Column using $  
theDF$Sport # gives the third column named Sport

## [1] "Hockey" "Football" "Baseball" "Curlin" "Rugby"   
## [6] "Lacrosse" "Basketball" "Tennis" "Cricket" "Soccer"

# Accessing Specific row and column  
theDF[3,2] # 3rd row and 2nd Column

## [1] -2

theDF[3,2:3] # 3rd Row and column 2 thru 3

## Second Sport  
## 3 -2 Baseball

theDF[c(3,5), 2]# Row 3&5 from Column 2;

## [1] -2 0

# since only one column was selected, it was returned as vector and hence no column names in output.  
  
# Rows 3&5 and Columns 2 through 3  
theDF[c(3,5), 2:3]

## Second Sport  
## 3 -2 Baseball  
## 5 0 Rugby

theDF[ ,3] # Access all Rows for column 3

## [1] "Hockey" "Football" "Baseball" "Curlin" "Rugby"   
## [6] "Lacrosse" "Basketball" "Tennis" "Cricket" "Soccer"

theDF[ , 2:3]

## Second Sport  
## 1 -4 Hockey  
## 2 -3 Football  
## 3 -2 Baseball  
## 4 -1 Curlin  
## 5 0 Rugby  
## 6 1 Lacrosse  
## 7 2 Basketball  
## 8 3 Tennis  
## 9 4 Cricket  
## 10 5 Soccer

theDF[2,]# Access all columns for Row 2

## First Second Sport  
## 2 9 -3 Football

theDF[2:4,]

## First Second Sport  
## 2 9 -3 Football  
## 3 8 -2 Baseball  
## 4 7 -1 Curlin

theDF[ , c("First", "Sport")]# access using Column Names

## First Sport  
## 1 10 Hockey  
## 2 9 Football  
## 3 8 Baseball  
## 4 7 Curlin  
## 5 6 Rugby  
## 6 5 Lacrosse  
## 7 4 Basketball  
## 8 3 Tennis  
## 9 2 Cricket  
## 10 1 Soccer