NOTES 2

Introduction to R

Acknowledgement: some of the contents are borrowed with or without modification from *An Introduction to Statistical Learning, with applications in R* (Springer, 2013) with permission from the authors: G. James, D. Witten, T. Hastie and R.Tibshirani.

R

- A software
- A programming language
- Download the base package
- Download and install contributed packages as needed

Basic Command

- Can be typed directly in R console (line by line)
- Or can run by creating a new script (File->New Script) and save as a R file
- Single variable
  - 5
  - x<-56/7+1 =
  - y=sqrt(x)
  - z<-y*2
  - a=y^2

- help(rnorm)
- ?rnorm

- ls() Denoted: list of all variables
- rm() remove variables
- rm(list = ls())

- # comments, until the end of the line.
- q() quit…

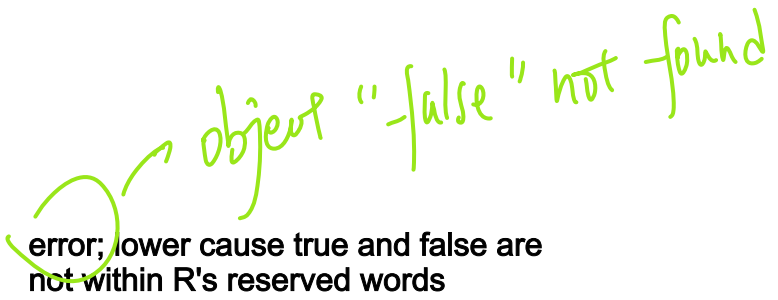
Other Basic functions

- abs(x)
- exp(x)
- ceiling(x)
- floor(x)
- cos(x)
- sin(x)
- log(x)
- log10(x)
- log(x,n)
- sort(x)
- sqrt(x)

- logical operator (<, <=, >, >=, ==, !=)
  - 5==6 F
  - 3<5 T
  - 2!=2 F
- &, I, ! use () when linking expressions
  - (5==6)&(3<5) F

Name a variable, matrix, or vector

- Start with a letter then can include any numbers and letters
- Case-sensitive Y != y
- Do not use underscore "_". avoid names such as "TRUE, FALSE, NA, if, function…." which are already reserved. Check the full list of reserved words by using help in the console
  - ?reserved
- Let's check…
  - a = TRUE
  - b = FALSE
  - how about a =false?



- # logical values T(=TRUE) and F(=FALSE):
- y <- c(T,T,F,T,F)
- y <- ((1:8) > 5)

Vector: One Row

- rm(list = ls())
- x=c(1,3,2,5,7,9)
- y=x[2]
- y=x[-2]
- z=x[2:5]
- a=x[2,5]
- a=x[c(2,5)]
- b=rep(2, 5)
- b = seq(from=1, to=6, by=1)
- sum(b), mean(b), length(b)
- c=b+x
- cor(b,c)

+c before:
combine all numbers in parameters together

Index starts with 1

python starts with 0

everything except the second one

error

index must be continuous; have to put the c before to make work

repeat 2 for 5 times

If the shorter one can be repeated, then it still can work although they are not the same length
Eg:
1,2,3,4,5,6
can only add 1/2/3 vectors--
1: x
2: x,x
3: x,x,x

Matrix: Two dimensional

- x=matrix(data=c(1,2,3,4,5,6,7,8,9,10,11,12), nrow=3,ncol=4)
- x[1,2]
- x[2,]
- x[,4]
- x[1, c(1,3)]
- x[c(1,3), c(1,3)]
- x[1:2,1:2]
- x[,-4]
- sum(x)
- sum(x[2,])
- apply(x, 1, mean)
- apply(x, 2, min)
- y=matrix(data=c(1,2,3,4,5,6,7,8,9,10,11,12), nrow=3,ncol=4, byrow=TRUE)
- dim(x)

number of rows

number of colomns

second row

forth column

sum of numbers in second row

1- this function would be applied per row

2- this function would be applied per colomn

default is bycolomn

dim-dimension
>>outcome:
number of rows, number of columns

Data Frame

What if the number if mean???????

- x <- matrix(c(1,3,5,7,9,11),ncol=2,byrow=T)
- rownames(x)=c("Record 1","Record 2","Record 3")
- colnames(x)=c("Var 1","Var 2")
- dimnames(x)<-list(c("Row 1","Row 2","Row 3"),c("Col1","Col2"))

define the attribute name

- # data frames
- x<-data.frame(x)
- attach(x)
- Col1
- detach(x)

easy to look for the specific number of the columns or rows

Difference between matrix and data frame:

in matrix, you can have only numbers;
However, you can have different types of data in data frame

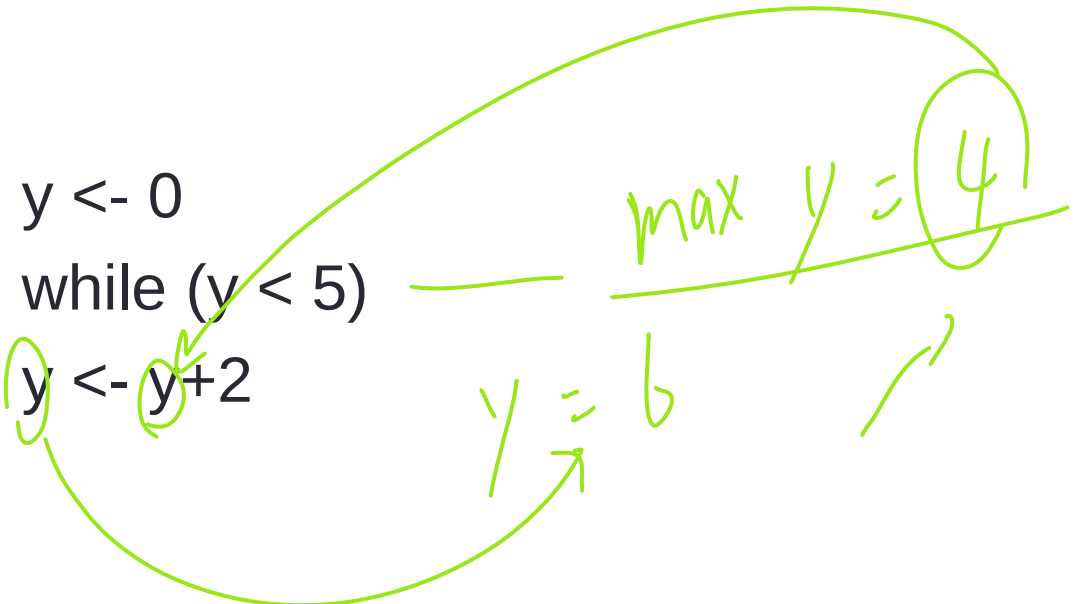
- x<-data.frame(c(1,5,9),c(3,7,11))
- dimnames(x)<-list(c("Row 1","Row 2","Row 3"),c("Col1","Col2"))

- e<- data.frame(
  id = c (1:5),
  name = c("A","B","C","D","E"),
  score = c(80,90.2,85,93,92)
)

LOOP

- x <- rep(1,5)
- y <- NULL
- for (i in 1:length(x))
- y <- c(y, x[i]*i)

- y <- 0
- while (y < 5)
- y <- y+2



Package

- Install package
  - install.packages(“package name”)
  - Click: package → install package→ the target package
  - Click: package →install package(s) from local zip files→ select the package from your local disk

- In each session, you need to first load the package before using it
  - library(“package name”)
  - Click: package → load package→ the target package

Load and Save Data

- read.table("A.data")
- read.csv(“http://www...../abc.csv”)
- read.table(“c:\\temp\\my folder\.....\\abc.txt”)
- A<-read.csv(file.choose(),header=T)

- write.table(A, “C:/FileName.txt", sep="\t")
- write.csv(A, “C:/FileName.csv")

Examine Object

- head(A)
- dim(A)
- str(A) structure function
- summary(A)

- names(A) a list of variable names
- attach(A) Try again!
- attach and names usually go together
  - attach(A)
  - names(A)

If u want to refer the attribute, u have to attach first

- **Please read Chapter 2.3 in the textbook for other basic commands that we do not cover in this notes.**