**Lec01**

1. Data: symbols

2. Information: data that are processed to be useful; provides answers to "who", "what", "where", and "when" questions

3. Knowledge: application of data and information; answers "how" questions

4. Understanding: appreciation of "why"

5. Wisdom: evaluated understanding

**What is big data?**

“Big data usually includes data sets with sizes beyond the ability of commonly used software tools to capture, manage, and process the data within a tolerable elapsed time.”

**Storage - Analytics - Presentation of results**

**Why hard?**

The amount of data collected grows exponentially with time.

Conventional techniques for analyzing data have a hard time catching up.

**Lec02**

1. Data have swept into every industry and business function and are now an important factor of production.
2. Big data creates value in several ways.
3. Use of big data will become a key basis of competition and growth for individual firms.
4. The use of big data will underpin new waves of productivity growth and consumer surplus.
5. While the use of big data will matter across sectors, some sectors are poised for greater gains.
6. There will be a shortage of talent necessary for organizations to take advantage of big data.
7. Several issues will have to be addressed to capture the full potential of big data/
8. Data policies.
9. Technology and techniques.
10. Organizational change and talent.
11. Access to data.
12. Industry structure

**R**

**Read csv:**

data <- read.csv ("fileName.csv", header = TRUE/FALSE)

**Read table:**

data <- read.table ("fileName.txt", header = TRUE/FALSE, quote = “分隔符”)

**Matrix and array:**

v = c(2,4,6,1,3,5,9,8,7,5,6,7)

cn = c("col1","col2","col3","col4")

rn = c("row1","row2","row3")

mn = c("mat1","mat2","mat3")

a = array(v, dim = c(3,4,3), dimnames = list(rn,cn,mn)):

, , mat1

col1 col2 col3 col4

row1 2 1 9 5

row2 4 3 8 6

row3 6 5 7 7

, , mat2

col1 col2 col3 col4

row1 2 1 9 5

row2 4 3 8 6

row3 6 5 7 7

, , mat3

col1 col2 col3 col4

row1 2 1 9 5

row2 4 3 8 6

row3 6 5 7 7

**Data frame:**

f = data.frame(

name = c("a","b","c"),

age = c(12,14,16),

gender = c("male","female","female"),

stringsAsFactors = FALSE

)

name age gender

1 a 12 male

2 b 14 female

3 c 16 female

**Summery:**

Summery(dataframe)

**Rows and cols:**

abRows <- data[a:b,]

abCols <- data[,a:b]

maxRows <- apply(data,2,max) 2是对比列

minRows <- apply(data,2,min)

averageData<- mean(data[,col])

mode, *class*, **typeOf**

**Empty values and sort:**

unValues <- data[!complete.cases(data),] 输出带empty的

unValues <- data[is.na(data),]

indices <- which(!complete.cases(data)) 输出行的号码

sortByOneCol <- data[order(data["ColName"]),] 从大到小

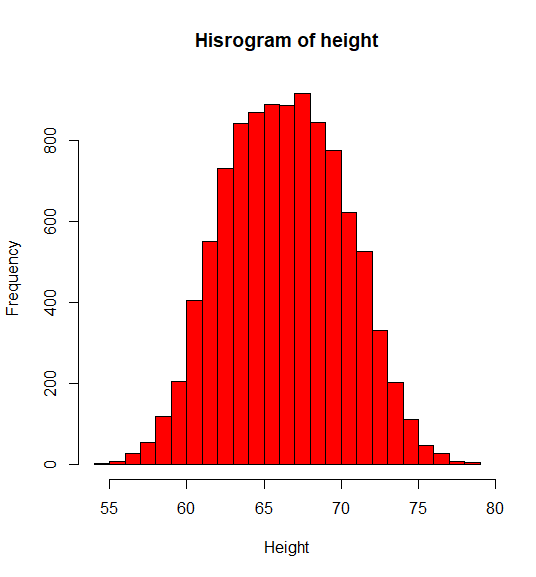
**Add and remove cols:**

newData <- cbind(data, addedData)

selectedData <- subset(data, colName == "value")

**Hist and plot:**

Hist <- hist(data, main = "MainName", xlab = "Height", col = "red", breaks = 20)



plot <- plot(data1, data2, main = "Weight-Height", xlab = "Weight", ylab = "Height", col = "red")

