R

Yi-Ju Tseng

2017-02-07

Contents

R R Hadoop EcoSystems

Chapter 1

R 101

R

1.1 R

R 2000 R 1.0.0 R R **Packages** R (Standing on the shoulders of giants (Hal R. Varian, Google)) 2017 1 R 10,000 (R Studio) R Studio Quick list of useful R packages

Package

```
install.packages("  ")
```

ggplot2 R Console

```
install.packages("ggplot2")
```

library( )

```
library(ggplot2)
```

1.2

R (1, 2,....) mean() :

```
mean(c(1,2,3,4,5,6)) ## 1~6
```

```
## [1] 3.5
```

?

```
?mean
```

seq() from, to, by

```
seq(from=1,to=9,by=2)#1~9  2
```

```
## [1] 1 3 5 7 9
```

```
seq(1,9,2)#
```

```
## [1] 1 3 5 7 9
```

```
seq(by=2,to=9,from=1)#
```

```
## [1] 1 3 5 7 9
```

1.3

R R R <- <- ( ) <- ->

```
a<-1
2->b
a
```

```
## [1] 1
```

```
b
```

```
## [1] 2
```

R = =

```
c=1
c
```

```
## [1] 1
```

str() str()

```
d<-3
str(d)
```

```
##  num 3
```

:

- break, else, FALSE, for, function, if, Inf, NA, NaN, next, repeat, return, TRUE, while
- .
-

1.4

R (Console) > + R

1.5

R **(numeric)** **(character)** **(logic)** **(Date)**

1.5.1 numeric

```
num1<-100
num2<-1000.001
```

2^53 bit64 package (Oehlschlägel, 2015) 2^63

```
print(2^53, digits=20)
```

```
## [1] 9007199254740992
```

```
print(2^53+1, digits=20) # +1   2^53
```

```
## [1] 9007199254740992
```

```
library(bit64) #  bit64 package
print(as.integer64(2)^53, digits=20)
```

```
## integer64
## [1] 9007199254740992
```

```
print(as.integer64(2)^53+1, digits=20)#  bit64
```

```
## integer64
## [1] 9007199254740993
```

1.5.2 character

"

```
char1<-"abcTest"
char2<-"100"
char3<-"200"
#char2+char3 #  Error message: non-numeric argument to binary operator
```

1.5.3 logic

TRUE T FALSE F

```
boolT<-TRUE
boolT1<-T
boolF<-FALSE
boolF1<-F
```

1.5.4 (Date)

Sys.Date()

```
dateBook<-Sys.Date()
dateBook
```

```
## [1] "2017-02-07"
```

lubridate(Grolemund et al., 2016) package / / ymd() y year m month d day / /

```
library(lubridate)
ymd('2012/3/3')
```

```
## [1] "2012-03-03"
```

```
mdy('3/3/2012')
```

```
## [1] "2012-03-03"
```

The Yhat Blog

1.6

1.6.1

R

- +
- -
- *
- /
- %%
- ^

```
num1<-1
num2<-100
num1+num2
```

```
## [1] 101
```

```
num1-num2
```

```
## [1] -99
```

```
num1*num2
```

```
## [1] 100
```

```
num1/num2
```

```
## [1] 0.01
```

```
100%%3 ##100 3
```

```
## [1] 1
```

```
2^3 ##2 3
```

```
## [1] 8
```

1.6.2

- round()
- floor()
- ceiling()

```
num1<-1.568
num2<-2.121
round(num1,digits = 2) #
```

```
## [1] 1.57
```

```
round(num2,digits = 1) #
```

```
## [1] 2.1
```

```
floor(num1) ##1.568
```

```
## [1] 1
```

```
ceiling(num2) ##2.121
```

```
## [1] 3
```

1.6.3

R

- >
- <
- ==
- >=
- <=

```
num1<-1
num2<-100
num1>num2
```

```
## [1] FALSE
```

```
num1<num2
```

```
## [1] TRUE
```

```
char1<-"abcTest"
char2<-"defTest"
char1>char2
```

```
## [1] FALSE
```

JAVA R & |

- &
- |

```
TRUE & TRUE
```

```
## [1] TRUE
```

```
TRUE & FALSE
```

```
## [1] FALSE
```

```
TRUE | TRUE
```

```
## [1] TRUE
```

```
TRUE | FALSE
```

```
## [1] TRUE
```

!

```
!TRUE
```

```
## [1] FALSE
```

```
!FALSE
```

```
## [1] TRUE
```

1.7

- Message
- Warning
- Error
- Condition

```
log(-1)
```

```
## Warning in log(-1): NaNs produced
```

```
## [1] NaN
```

```
mena(NA)
```

```
## Error in eval(expr, envir, enclos): could not find function "mena"
```

1:

```
# Error: could not find function "fetch_NBAPlayerStatistics"
#   "fetch_NBAPlayerStatistics" function
```

SportsAnalytics package

2:

```
# Error in library(knitr): there is no package called 'knitr'
#   "knitr" package
```

knitr package

1.8 Help

R R ? ?

```
?ggplot2
?ymd
```

Stack Overflow

Chapter 2

R

2.1 vector

c()

```
vec<-c('a','b','c','d','e')
```

a~e vec **(element)** a vec **1** b **2** vec **4**

```
vec[4] ## 4
```

```
## [1] "d"
```

```
vec[c(2,3)] ## 2 3
```

```
## [1] "b" "c"
```

vec

<-

```
vec[3]
```

```
## [1] "c"
```

```
vec[3]<-'z' ##     "z"
vec[3]
```

```
## [1] "z"
```

2.1.1

1~20 :

```
1:20 ## c(1,2,...,19,20)
```

```
##  [1]  1  2  3  4  5  6  7  8  9 10 11 12 13 14 15 16 17 18 19 20
```

seq()

```
seq(from=1,to=20,by=1) ##1~20  1
```

```
##  [1]  1  2  3  4  5  6  7  8  9 10 11 12 13 14 15 16 17 18 19 20
```

```
seq(from=1,to=50,by=2) ##1~50   2
```

```
##  [1]  1  3  5  7  9 11 13 15 17 19 21 23 25 27 29 31 33 35 37 39 41 43 45 47 49
```

2.1.2

```
numvec<-1:10 ## c(1,2,3,4,5,6,7,8,9,10)
numvec+3 ##   +3
```

```
##  [1]  4  5  6  7  8  9 10 11 12 13
```

```
numvec*2 ##   *2
```

```
##  [1]  2  4  6  8 10 12 14 16 18 20
```

```
numvec1<-1:3 ## c(1,2,3)
numvec2<-4:6 ## c(4,5,6)
numvec1+numvec2
```

```
## [1] 5 7 9
```

```
numvec1*numvec2
```

```
## [1]  4 10 18
```

2.2 factor

factor(,levels=) levels

```
factor(c("  ","   ","   "),
       levels = c("  ","   ","   "))
```

```
## [1]
## Levels:
```

2.3 list

R **list** list()

```
listSample<-list(Students=c("Tom","Kobe","Emma","Amy"),Year=2017,
                 Score=c(60,50,80,40),School="CGU")
listSample
```

```
## $Students
## [1] "Tom"  "Kobe" "Emma" "Amy"
##
## $Year
## [1] 2017
##
## $Score
## [1] 60 50 80 40
```

```
##
## $School
## [1] "CGU"
```

2.3.1

$

```
listSample$Students ##   Students
```

```
## [1] "Tom"  "Kobe" "Emma" "Amy"
```

[[]]

```
listSample[[1]] ##
```

```
## [1] "Tom"  "Kobe" "Emma" "Amy"
```

list

```
listSample[1] ##
```

```
## $Students
## [1] "Tom"  "Kobe" "Emma" "Amy"
```

2.3.2

```
listSample[[1]]
```

```
## [1] "Tom"  "Kobe" "Emma" "Amy"
```

```
listSample[[1]]<-c(" "," "," "," "," ") ## Students
listSample[[1]]
```

```
## [1] " " " " " " " " " "
```

$ <-

```
listSample$Gender<-c("M","F","M","F","M") ## Gender
```

2.4 matrix

```
a <- matrix(c(1:6), nrow=3, ncol=2) ## 3x2    1~6
a
```

```
##      [,1] [,2]
## [1,]    1    4
## [2,]    2    5
## [3,]    3    6
```

2.5 data.frame

Column Row Excel `data.frame()`

```
StuDF <- data.frame(StuID=c(1,2,3,4,5), ##  =
                    name=c(" "," "," "," "," "),
                    score=c(80,60,90,70,50))
StuDF
```

```
##   StuID name score
## 1     1      80
## 2     2      60
## 3     3      90
## 4     4      70
## 5     5      50
```

StuID, name, score R V1Vn R StuID score name R 1n colnames() rownames

```
colnames(StuDF) ##
```

```
## [1] "StuID" "name"  "score"
```

```
rownames(StuDF) ##
```

```
## [1] "1" "2" "3" "4" "5"
```

str()

```
str(StuDF)
```

```
## 'data.frame':    5 obs. of  3 variables:
##  $ StuID: num  1 2 3 4 5
##  $ name : Factor w/ 5 levels " "," ",..: 4 2 5 3 1
##  $ score: num  80 60 90 70 50
```

2.6 data.table

data.table data.frame data.table (Dowle et al., 2016) package `data.table` Chapter 6 A data.table R tutorial by DataCamp DataCamp

2.7

:

- names()
- dimnames()
- length()
- dim()
- class()
- table()
- str()

names()

```
head(islands) ##R
```

```
##       Africa   Antarctica         Asia    Australia Axel Heiberg       Baffin
##        11506         5500        16988         2968           16          184
```

```
head(names(islands)) ##
```

```
## [1] "Africa"       "Antarctica"   "Asia"         "Australia"    "Axel Heiberg"
## [6] "Baffin"
```

```
head(USArrests) ##R
```

```
##            Murder Assault UrbanPop Rape
## Alabama      13.2     236       58 21.2
## Alaska       10.0     263       48 44.5
## Arizona       8.1     294       80 31.0
## Arkansas      8.8     190       50 19.5
## California    9.0     276       91 40.6
## Colorado      7.9     204       78 38.7
```

```
head(names(USArrests)) ##
```

```
## [1] "Murder"   "Assault"  "UrbanPop" "Rape"
```

```
 dimnames()
```

```
dimnames(USArrests)
```

```
## [[1]]
##  [1] "Alabama"        "Alaska"         "Arizona"        "Arkansas"
##  [5] "California"     "Colorado"       "Connecticut"    "Delaware"
##  [9] "Florida"        "Georgia"        "Hawaii"         "Idaho"
## [13] "Illinois"       "Indiana"        "Iowa"           "Kansas"
## [17] "Kentucky"       "Louisiana"      "Maine"          "Maryland"
## [21] "Massachusetts"  "Michigan"       "Minnesota"      "Mississippi"
## [25] "Missouri"       "Montana"        "Nebraska"       "Nevada"
## [29] "New Hampshire"  "New Jersey"     "New Mexico"     "New York"
## [33] "North Carolina" "North Dakota"   "Ohio"           "Oklahoma"
## [37] "Oregon"         "Pennsylvania"   "Rhode Island"   "South Carolina"
## [41] "South Dakota"   "Tennessee"      "Texas"          "Utah"
## [45] "Vermont"        "Virginia"       "Washington"     "West Virginia"
## [49] "Wisconsin"      "Wyoming"
##
## [[2]]
## [1] "Murder"   "Assault"  "UrbanPop" "Rape"
```

```
 length()
```

```
length(islands)
```

```
## [1] 48
```

```
length(USArrests)
```

```
## [1] 4
```

```
 dim()         dimnames()
```

```
dim(USArrests)
```

```
## [1] 50  4
```

```
 class()
```

```
class(1)
```

```
## [1] "numeric"
```

```
class("Test")
```

```
## [1] "character"
```

```
class(Sys.Date())
```

```
## [1] "Date"
```

```
 table()
iris$Species ##
```

```
##   [1] setosa     setosa     setosa     setosa     setosa     setosa
##   [7] setosa     setosa     setosa     setosa     setosa     setosa
##  [13] setosa     setosa     setosa     setosa     setosa     setosa
##  [19] setosa     setosa     setosa     setosa     setosa     setosa
##  [25] setosa     setosa     setosa     setosa     setosa     setosa
##  [31] setosa     setosa     setosa     setosa     setosa     setosa
##  [37] setosa     setosa     setosa     setosa     setosa     setosa
##  [43] setosa     setosa     setosa     setosa     setosa     setosa
##  [49] setosa     setosa     versicolor versicolor versicolor versicolor
##  [55] versicolor versicolor versicolor versicolor versicolor versicolor
##  [61] versicolor versicolor versicolor versicolor versicolor versicolor
##  [67] versicolor versicolor versicolor versicolor versicolor versicolor
##  [73] versicolor versicolor versicolor versicolor versicolor versicolor
##  [79] versicolor versicolor versicolor versicolor versicolor versicolor
##  [85] versicolor versicolor versicolor versicolor versicolor versicolor
##  [91] versicolor versicolor versicolor versicolor versicolor versicolor
##  [97] versicolor versicolor versicolor versicolor virginica  virginica
## [103] virginica  virginica  virginica  virginica  virginica  virginica
## [109] virginica  virginica  virginica  virginica  virginica  virginica
## [115] virginica  virginica  virginica  virginica  virginica  virginica
## [121] virginica  virginica  virginica  virginica  virginica  virginica
## [127] virginica  virginica  virginica  virginica  virginica  virginica
## [133] virginica  virginica  virginica  virginica  virginica  virginica
## [139] virginica  virginica  virginica  virginica  virginica  virginica
## [145] virginica  virginica  virginica  virginica  virginica  virginica
## Levels: setosa versicolor virginica
```

```
table(iris$Species) ##
```

```
##
##     setosa versicolor  virginica
##         50         50         50
```

```
 str()
str(iris)
```

```
## 'data.frame':    150 obs. of  5 variables:
##  $ Sepal.Length: num  5.1 4.9 4.7 4.6 5 5.4 4.6 5 4.4 4.9 ...
##  $ Sepal.Width : num  3.5 3 3.2 3.1 3.6 3.9 3.4 3.4 2.9 3.1 ...
##  $ Petal.Length: num  1.4 1.4 1.3 1.5 1.4 1.7 1.4 1.5 1.4 1.5 ...
##  $ Petal.Width : num  0.2 0.2 0.2 0.2 0.2 0.4 0.3 0.2 0.2 0.1 ...
##  $ Species     : Factor w/ 3 levels "setosa","versicolor",..: 1 1 1 1 1 1 1 1 1 1 ...
```

```
str(listSample)
```

```
## List of 5
```

```
##  $ Students: chr [1:5] " " " " " " " " " ...
##  $ Year    : num 2017
##  $ Score   : num [1:4] 60 50 80 40
##  $ School  : chr "CGU"
##  $ Gender  : chr [1:5] "M" "F" "M" "F" ...
```

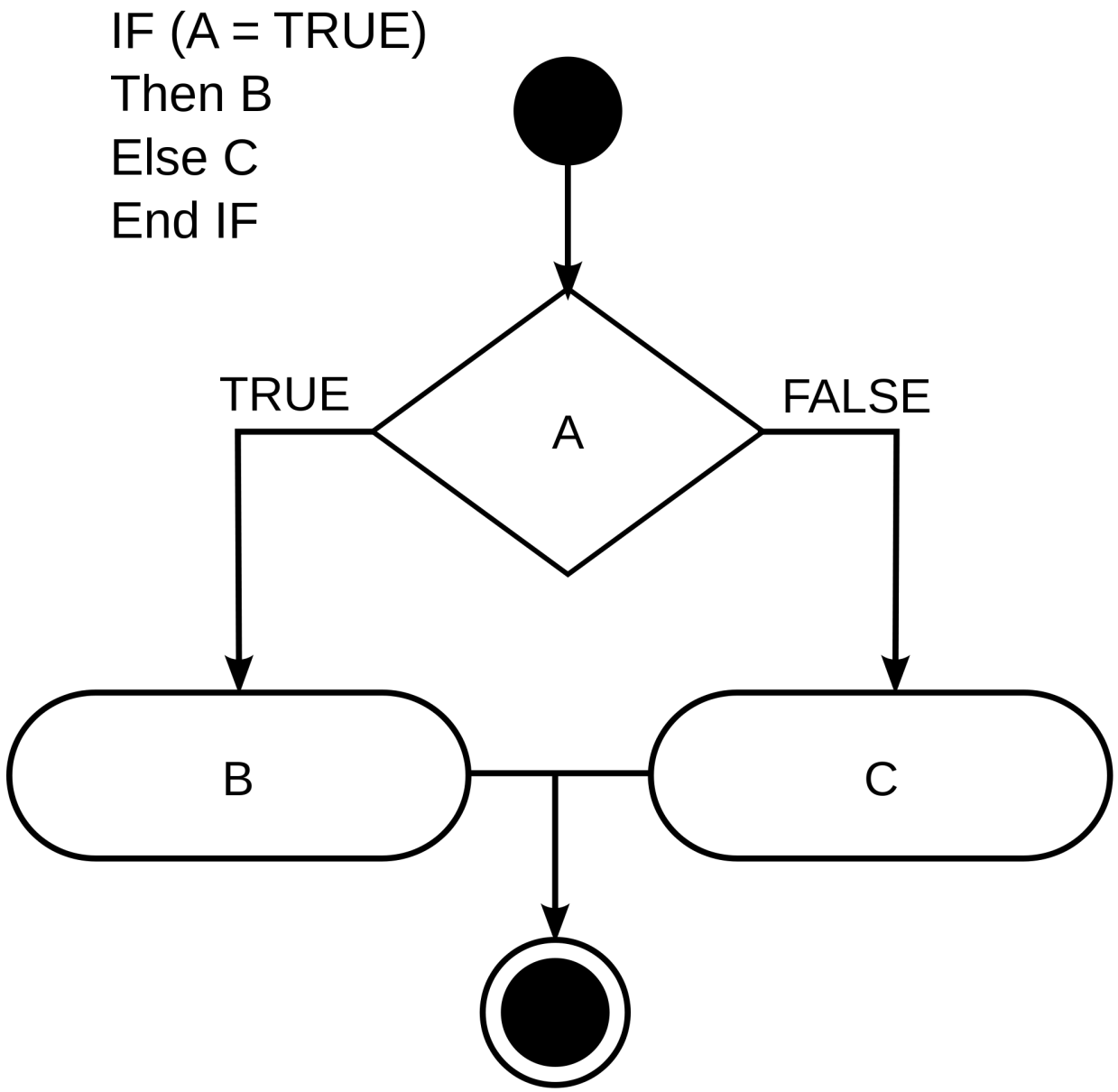
Chapter 3

3.1

3.1.1 if-else

if-else **if-else** **if** **(TRUE)** if **(FALSE)** **else** **else**

```
knitr::include_graphics("figure/ifelse.png")
```



if else {} {} {}

60 60 :

```
score<-59
if(score>=60){
  print(" ")
}else{
  print("  ")
}
```

```
## [1] "  "
```

```
score<-80
if(score>=60){
  print(" ")
}else{
```

```
  print("  ")
}
```

```
## [1] " "
```

3.1.2 if-else if-else

90 60 90 60 if else **else if** :

```
score<-95
if(score>=90){
  print(" ")
}else if(score>=60){
  print(" ")
}else{
  print("  ")
}
```

```
## [1] " "
```

if-else if-else if else if if 95 90 (if) 60 (else if)

3.1.3 if

if if if-else 60 60 60 :

```
CHscore<-95 ##
ENscore<-55 ##
if(CHscore>=60){
  if(ENscore>=60){
    print("   ")
  }else{
    print("      ")
  }
}else{
  if(ENscore>=60){
    print("      ")
  }else{
    print("   ")
  }
}
```

```
## [1] "      "
```

3.1.4 ifelse()

ifelse() if-else ifelse(, ,) :

```
score<-80
ifelse(score>=60," "," ")
```

```
## [1] " "
```

ifelse()

```
scoreVector<-c(30,90,50,60,80)
ifelse(scoreVector>=60," ", "  ")
```

```
## [1] "  " " "  "   "  " " "  "   " "
```

3.2

3.2.1 for

R for for for (in){ } :

```
for (n in 1:10){ #n   1:10
  print(n)
}
```

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

for if-else :

```
for (n in 1:10){
  if(n%%2==0){ #
    print(n)
  }else{
    print(" ") #   " "
  }
}
```

```
## [1] " "
## [1] 2
## [1] " "
## [1] 4
## [1] " "
## [1] 6
## [1] " "
## [1] 8
## [1] " "
## [1] 10
```

3.2.2 while

while while

```
x<-0
while(x<=5){
  print(x)
```

```
  x<-x+1
}
```

```
## [1] 0
## [1] 1
## [1] 2
## [1] 3
## [1] 4
## [1] 5
```

3.2.3 break

break

```
for(n in 1:10){
  if(n==5){
    break ##  5
  }
  print(n)
}
```

```
## [1] 1
## [1] 2
## [1] 3
## [1] 4
```

3.2.4 next

next

```
for(n in 1:10){
  if(n==5){
    next ## 5
  }
  print(n)
}
```

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

Chapter 4

What is 'Data'?

http://en.wikipedia.org/wiki/Data

`Data are values of qualitative or quantitative variables, belonging to a set of items.`

-
- tidy
- R

`Raw data` -> **Processing script** -> `Tidy data` -> Data analysis -> Data communication

4.1

4.1.1 Tidy Data

- Column Column Name
- Raw
-
- index
- One file, one table

```
if (!require('SportsAnalytics')){
    install.packages("SportsAnalytics")
    library(SportsAnalytics)
}
NBA1415<-fetch_NBAPlayerStatistics("14-15")
```

```
head(NBA1415)
```

```
##   League          Name Team Position GamesPlayed TotalMinutesPlayed
## 1    NBA    Quincy Acy  NYK       SF          68                1288
## 2    NBA  Jordan Adams  MEM       SG          30                 249
## 3    NBA  Steven Adams  OKL        C          70                1776
## 4    NBA   Jeff Adrien  MIN       PF          17                 215
## 5    NBA Arron Afflalo  POR       SG          78                2502
## 6    NBA Alexis Ajinca  NOR        C          68                 956
##   FieldGoalsMade FieldGoalsAttempted ThreesMade ThreesAttempted FreeThrowsMade
```

```
## 1            152                 331         18              60             76
## 2             35                  86         10              25             14
## 3            217                 399          0               2            103
## 4             19                  44          0               0             22
## 5            375                 884        118             333            167
## 6            181                 328          0               0             81
##   FreeThrowsAttempted OffensiveRebounds TotalRebounds Assists Steals Turnovers
## 1                  97                79           301      68     27        60
## 2                  23                 9            28      16     16        14
## 3                 205               199           522      65     38        99
## 4                  38                23            77      15      4         9
## 5                 198                27           247     129     41       116
## 6                  99               104           315      47     21        69
##   Blocks PersonalFouls Disqualifications TotalPoints Technicals Ejections
## 1     22           147                 1         398          5         0
## 2      7            24                 0          94          0         0
## 3     85           222                 3         537          3         0
## 4      9            30                 0          60          0         0
## 5      7           167                 1        1035          0         0
## 6     51           151                 0         443          1         0
##   FlagrantFouls GamesStarted
## 1             0           22
## 2             0            0
## 3             0           67
## 4             0            0
## 5             0           72
## 6             0            8
```

4.1.2 Raw Data

4.1.2.1 Html

4.1.2.2 Facebook

4.1.2.3 !?

4.2

-
-
- Open Data
  - http://data.taipei/
  - http://data.tycg.gov.tw/
  - http://data.moi.gov.tw/
-
-

4.2.1 Open Data

4.3

4.3.1 (.csv / Tab / Excel)

`read.table`, `read.csv`,

The `read.table`, `read.csv`

- `file`,
- `header`, T/F
- `sep`,
- `colClasses`,
- `comment.char`,
- `skip`,
- `stringsAsFactors`, 'Factor'

xlsx

```
if (!require('xlsx')){
    install.packages("xlsx")
    library(xlsx)
}
ExcelData <- read.xlsx("data.xlsx",sheetIndex=1,header=TRUE)
head(ExcelData)
```

4.4 read.csv

```
data <- read.csv('open.csv')
data
```

4.4.1

`readLines`,

4.4.2 R

`load`, R Ex: iris

4.4.3 R

`source`, R Obejct or script, , ASCII (`dump`)

4.5

4.5.1 (.csv / Tab)

`write.table`

4.5.2

`writeLines`

4.5.3 R

`save`

4.5.4 R

`dump`

4.6 R

- # Row
-
- Column
-

Column

```
initial <- read.csv("open.csv", nrows = 100)
classes <- sapply(initial, class)
tabAll <- read.csv("open.csv", colClasses = classes)
```

4.7 download.file

RCurl Package

download.file(URL, destfile= , method=?)

method = "curl"–>For **https**

```
if (!require('RCurl')){
    install.packages("RCurl")
    library(RCurl)
}
download.file("https://raw.githubusercontent.com/yijutseng/BigDataCGUIM/master/files/opendata10401.csv"
              destfile = "open.csv", method = "curl")
```

4.8 Open Data

4.9

Chapter 5

5.1

Chapter 1.5 **(numeric)** **(character)** **(logic)** **(Date)**

5.1.1

`is.` TRUE

- `is.numeric( )`
- `is.character( )`
- `is.logical( )`

```
num<-100
cha<-'200'
boo<-T
is.numeric(num)
```

```
## [1] TRUE
```

```
is.numeric(cha)
```

```
## [1] FALSE
```

```
is.character(num)
```

```
## [1] FALSE
```

```
is.character(cha)
```

```
## [1] TRUE
```

```
is.logical(boo)
```

```
## [1] TRUE
```

`class( )`

```
class(num)
```

```
## [1] "numeric"
```

```
class(cha)
```

```
## [1] "character"
```

```
class(boo)
```

```
## [1] "logical"
```

```
class(Sys.Date())
```

```
## [1] "Date"
```

5.1.2

as.

- as.numeric( )
- as.character( )
- as.logical( )

```
as.numeric(cha)
```

```
## [1] 200
```

```
as.numeric(boo)
```

```
## [1] 1
```

```
as.character(num)
```

```
## [1] "100"
```

```
as.character(boo)
```

```
## [1] "TRUE"
```

NA Warning: NAs introduced by coercion Warning: NA

```
as.numeric("abc")
```

```
## Warning: NAs introduced by coercion
```

```
## [1] NA
```

lubridate(Grolemund et al., 2016) package / / ymd() y year m month d day / / mdy()

```
library(lubridate)
ymd('2012/3/3')
```

```
## [1] "2012-03-03"
```

```
mdy('3/3/2012')
```

```
## [1] "2012-03-03"
```

5.2

5.2.1

- strsplit()
- substr()
- toupper() tolower()
- paste() paste0()
- gsub()
- str_trim() stringr(Wickham, 2016b) package

```
strsplit ("Hello World"," ")
```

```
## [[1]]
## [1] "Hello" "World"
```

```
toupper("Hello World")
```

```
## [1] "HELLO WORLD"
```

```
tolower("Hello World")
```

```
## [1] "hello world"
```

```
paste("Hello", "World", sep='')
```

```
## [1] "HelloWorld"
```

```
substr("Hello World", start=2,stop=4)
```

```
## [1] "ell"
```

```
gsub("o","0","Hello World")
```

```
## [1] "Hell0 W0rld"
```

```
library(stringr)
str_trim(" Hello World ")
```

```
## [1] "Hello World"
```

5.2.2

grep() grepl():

- (index) grep(,)
- (TRUE or FALSE) grepl(,)

```
grep("A",c("Alex","Tom","Amy","Joy","Emma")) ##      A   "A"
```

```
## [1] 1 3
```

```
grepl("A",c("Alex","Tom","Amy","Joy","Emma")) ##      A      "A"
```

```
## [1]  TRUE FALSE  TRUE FALSE FALSE
```

```
grepl("a",c("Alex","Tom","Amy","Joy","Emma")) ##      a      "a"
```

```
## [1] FALSE FALSE FALSE FALSE  TRUE
```

5.3 Subset

5.3.1 ( )

{#vector} []

```
letters ##R
```

```
##  [1] "a" "b" "c" "d" "e" "f" "g" "h" "i" "j" "k" "l" "m" "n" "o" "p" "q" "r" "s"
## [20] "t" "u" "v" "w" "x" "y" "z"
```

```
letters[1] ## letters
```

```
## [1] "a"
```

```
letters[1:10] ## letters
```

```
##  [1] "a" "b" "c" "d" "e" "f" "g" "h" "i" "j"
```

```
letters[c(1,3,5)] ## letters  1,3,5
```

```
## [1] "a" "c" "e"
```

```
letters[c(-1,-3,-5)] ## letters   1,3,5
```

```
##  [1] "b" "d" "f" "g" "h" "i" "j" "k" "l" "m" "n" "o" "p" "q" "r" "s" "t" "u" "v"
## [20] "w" "x" "y" "z"
```

head() tail()

```
head(letters,5) ## letters
```

```
## [1] "a" "b" "c" "d" "e"
```

```
tail(letters,3) ## letters
```

```
## [1] "x" "y" "z"
```

5.3.2

data.frame (Row) (Column) [] , **Row, Column** , , (index) (TRUE/FALSE)

- : dataFrame[row index,column index]
- : dataFrame[c(T,F,T),c(T,F,T)]
- : dataFrame[row name,column name]

```
iris[1,2] ##  Row  Column
```

```
## [1] 3.5
```

```
iris[1:3,] ## 1~3 Row   Column
```

```
##   Sepal.Length Sepal.Width Petal.Length Petal.Width Species
## 1          5.1         3.5          1.4         0.2  setosa
## 2          4.9         3.0          1.4         0.2  setosa
## 3          4.7         3.2          1.3         0.2  setosa
```

```
iris[,"Species"] ##  Row  Species Column
```

```
##   [1] setosa     setosa     setosa     setosa     setosa     setosa
##   [7] setosa     setosa     setosa     setosa     setosa     setosa
##  [13] setosa     setosa     setosa     setosa     setosa     setosa
##  [19] setosa     setosa     setosa     setosa     setosa     setosa
##  [25] setosa     setosa     setosa     setosa     setosa     setosa
##  [31] setosa     setosa     setosa     setosa     setosa     setosa
##  [37] setosa     setosa     setosa     setosa     setosa     setosa
##  [43] setosa     setosa     setosa     setosa     setosa     setosa
##  [49] setosa     setosa     versicolor versicolor versicolor versicolor
##  [55] versicolor versicolor versicolor versicolor versicolor versicolor
##  [61] versicolor versicolor versicolor versicolor versicolor versicolor
##  [67] versicolor versicolor versicolor versicolor versicolor versicolor
```

```
##  [73] versicolor versicolor versicolor versicolor versicolor versicolor
##  [79] versicolor versicolor versicolor versicolor versicolor versicolor
##  [85] versicolor versicolor versicolor versicolor versicolor versicolor
##  [91] versicolor versicolor versicolor versicolor versicolor versicolor
##  [97] versicolor versicolor versicolor versicolor virginica  virginica
## [103] virginica  virginica  virginica  virginica  virginica  virginica
## [109] virginica  virginica  virginica  virginica  virginica  virginica
## [115] virginica  virginica  virginica  virginica  virginica  virginica
## [121] virginica  virginica  virginica  virginica  virginica  virginica
## [127] virginica  virginica  virginica  virginica  virginica  virginica
## [133] virginica  virginica  virginica  virginica  virginica  virginica
## [139] virginica  virginica  virginica  virginica  virginica  virginica
## [145] virginica  virginica  virginica  virginica  virginica  virginica
## Levels: setosa versicolor virginica
```

```
iris[1:10,c(T,F,T,F,T)] ## 1~10 Row 1,3,5 Column (TRUE)
```

```
##    Sepal.Length Petal.Length Species
## 1           5.1          1.4  setosa
## 2           4.9          1.4  setosa
## 3           4.7          1.3  setosa
## 4           4.6          1.5  setosa
## 5           5.0          1.4  setosa
## 6           5.4          1.7  setosa
## 7           4.6          1.4  setosa
## 8           5.0          1.5  setosa
## 9           4.4          1.4  setosa
## 10          4.9          1.5  setosa
```

$ Column

```
iris$Species ##  Row  Species Column
```

```
##   [1] setosa     setosa     setosa     setosa     setosa     setosa
##   [7] setosa     setosa     setosa     setosa     setosa     setosa
##  [13] setosa     setosa     setosa     setosa     setosa     setosa
##  [19] setosa     setosa     setosa     setosa     setosa     setosa
##  [25] setosa     setosa     setosa     setosa     setosa     setosa
##  [31] setosa     setosa     setosa     setosa     setosa     setosa
##  [37] setosa     setosa     setosa     setosa     setosa     setosa
##  [43] setosa     setosa     setosa     setosa     setosa     setosa
##  [49] setosa     setosa     versicolor versicolor versicolor versicolor
##  [55] versicolor versicolor versicolor versicolor versicolor versicolor
##  [61] versicolor versicolor versicolor versicolor versicolor versicolor
##  [67] versicolor versicolor versicolor versicolor versicolor versicolor
##  [73] versicolor versicolor versicolor versicolor versicolor versicolor
##  [79] versicolor versicolor versicolor versicolor versicolor versicolor
##  [85] versicolor versicolor versicolor versicolor versicolor versicolor
##  [91] versicolor versicolor versicolor versicolor versicolor versicolor
##  [97] versicolor versicolor versicolor versicolor virginica  virginica
## [103] virginica  virginica  virginica  virginica  virginica  virginica
## [109] virginica  virginica  virginica  virginica  virginica  virginica
## [115] virginica  virginica  virginica  virginica  virginica  virginica
## [121] virginica  virginica  virginica  virginica  virginica  virginica
## [127] virginica  virginica  virginica  virginica  virginica  virginica
## [133] virginica  virginica  virginica  virginica  virginica  virginica
```

```
## [139] virginica  virginica  virginica  virginica  virginica  virginica
## [145] virginica  virginica  virginica  virginica  virginica  virginica
## Levels: setosa versicolor virginica
```

Row `subset()` `subset( , )`

```
subset(iris,Species=="virginica") ##Species "virginica" Row   Column
```

```
##     Sepal.Length Sepal.Width Petal.Length Petal.Width   Species
## 101          6.3         3.3          6.0         2.5 virginica
## 102          5.8         2.7          5.1         1.9 virginica
## 103          7.1         3.0          5.9         2.1 virginica
## 104          6.3         2.9          5.6         1.8 virginica
## 105          6.5         3.0          5.8         2.2 virginica
## 106          7.6         3.0          6.6         2.1 virginica
## 107          4.9         2.5          4.5         1.7 virginica
## 108          7.3         2.9          6.3         1.8 virginica
## 109          6.7         2.5          5.8         1.8 virginica
## 110          7.2         3.6          6.1         2.5 virginica
## 111          6.5         3.2          5.1         2.0 virginica
## 112          6.4         2.7          5.3         1.9 virginica
## 113          6.8         3.0          5.5         2.1 virginica
## 114          5.7         2.5          5.0         2.0 virginica
## 115          5.8         2.8          5.1         2.4 virginica
## 116          6.4         3.2          5.3         2.3 virginica
## 117          6.5         3.0          5.5         1.8 virginica
## 118          7.7         3.8          6.7         2.2 virginica
## 119          7.7         2.6          6.9         2.3 virginica
## 120          6.0         2.2          5.0         1.5 virginica
## 121          6.9         3.2          5.7         2.3 virginica
## 122          5.6         2.8          4.9         2.0 virginica
## 123          7.7         2.8          6.7         2.0 virginica
## 124          6.3         2.7          4.9         1.8 virginica
## 125          6.7         3.3          5.7         2.1 virginica
## 126          7.2         3.2          6.0         1.8 virginica
## 127          6.2         2.8          4.8         1.8 virginica
## 128          6.1         3.0          4.9         1.8 virginica
## 129          6.4         2.8          5.6         2.1 virginica
## 130          7.2         3.0          5.8         1.6 virginica
## 131          7.4         2.8          6.1         1.9 virginica
## 132          7.9         3.8          6.4         2.0 virginica
## 133          6.4         2.8          5.6         2.2 virginica
## 134          6.3         2.8          5.1         1.5 virginica
## 135          6.1         2.6          5.6         1.4 virginica
## 136          7.7         3.0          6.1         2.3 virginica
## 137          6.3         3.4          5.6         2.4 virginica
## 138          6.4         3.1          5.5         1.8 virginica
## 139          6.0         3.0          4.8         1.8 virginica
## 140          6.9         3.1          5.4         2.1 virginica
## 141          6.7         3.1          5.6         2.4 virginica
## 142          6.9         3.1          5.1         2.3 virginica
## 143          5.8         2.7          5.1         1.9 virginica
## 144          6.8         3.2          5.9         2.3 virginica
## 145          6.7         3.3          5.7         2.5 virginica
## 146          6.7         3.0          5.2         2.3 virginica
```

```
## 147          6.3         2.5          5.0         1.9 virginica
## 148          6.5         3.0          5.2         2.0 virginica
## 149          6.2         3.4          5.4         2.3 virginica
## 150          5.9         3.0          5.1         1.8 virginica
```

Row `grepl()`

```
knitr::kable(iris[grepl("color",iris$Species),]) ##Species "color"
```

	Sepal.Length	Sepal.Width	Petal.Length	Petal.Width	Species
51	7.0	3.2	4.7	1.4	versicolor
52	6.4	3.2	4.5	1.5	versicolor
53	6.9	3.1	4.9	1.5	versicolor
54	5.5	2.3	4.0	1.3	versicolor
55	6.5	2.8	4.6	1.5	versicolor
56	5.7	2.8	4.5	1.3	versicolor
57	6.3	3.3	4.7	1.6	versicolor
58	4.9	2.4	3.3	1.0	versicolor
59	6.6	2.9	4.6	1.3	versicolor
60	5.2	2.7	3.9	1.4	versicolor
61	5.0	2.0	3.5	1.0	versicolor
62	5.9	3.0	4.2	1.5	versicolor
63	6.0	2.2	4.0	1.0	versicolor
64	6.1	2.9	4.7	1.4	versicolor
65	5.6	2.9	3.6	1.3	versicolor
66	6.7	3.1	4.4	1.4	versicolor
67	5.6	3.0	4.5	1.5	versicolor
68	5.8	2.7	4.1	1.0	versicolor
69	6.2	2.2	4.5	1.5	versicolor
70	5.6	2.5	3.9	1.1	versicolor
71	5.9	3.2	4.8	1.8	versicolor
72	6.1	2.8	4.0	1.3	versicolor
73	6.3	2.5	4.9	1.5	versicolor
74	6.1	2.8	4.7	1.2	versicolor
75	6.4	2.9	4.3	1.3	versicolor
76	6.6	3.0	4.4	1.4	versicolor
77	6.8	2.8	4.8	1.4	versicolor
78	6.7	3.0	5.0	1.7	versicolor
79	6.0	2.9	4.5	1.5	versicolor
80	5.7	2.6	3.5	1.0	versicolor
81	5.5	2.4	3.8	1.1	versicolor
82	5.5	2.4	3.7	1.0	versicolor
83	5.8	2.7	3.9	1.2	versicolor
84	6.0	2.7	5.1	1.6	versicolor
85	5.4	3.0	4.5	1.5	versicolor
86	6.0	3.4	4.5	1.6	versicolor
87	6.7	3.1	4.7	1.5	versicolor
88	6.3	2.3	4.4	1.3	versicolor
89	5.6	3.0	4.1	1.3	versicolor
90	5.5	2.5	4.0	1.3	versicolor
91	5.5	2.6	4.4	1.2	versicolor
92	6.1	3.0	4.6	1.4	versicolor
93	5.8	2.6	4.0	1.2	versicolor
94	5.0	2.3	3.3	1.0	versicolor
95	5.6	2.7	4.2	1.3	versicolor
96	5.7	3.0	4.2	1.2	versicolor
97	5.7	2.9	4.2	1.3	versicolor
98	6.2	2.9	4.3	1.3	versicolor
99	5.1	2.5	3.0	1.1	versicolor
100	5.7	2.8	4.1	1.3	versicolor

(Raw) `head()` `tail()`

```
head(iris,5) ## iris
```

```
##   Sepal.Length Sepal.Width Petal.Length Petal.Width Species
## 1          5.1         3.5          1.4         0.2  setosa
## 2          4.9         3.0          1.4         0.2  setosa
## 3          4.7         3.2          1.3         0.2  setosa
## 4          4.6         3.1          1.5         0.2  setosa
## 5          5.0         3.6          1.4         0.2  setosa
```

```
tail(iris,3) ## iris
```

```
##     Sepal.Length Sepal.Width Petal.Length Petal.Width   Species
## 148          6.5         3.0          5.2         2.0 virginica
## 149          6.2         3.4          5.4         2.3 virginica
## 150          5.9         3.0          5.1         1.8 virginica
```

5.4

5.4.1 sort

sort()

```
head(islands) ##
```

```
##       Africa   Antarctica         Asia    Australia Axel Heiberg       Baffin
##        11506         5500        16988         2968           16          184
```

```
head(sort(islands)) ##
```

```
##       Vancouver          Hainan Prince of Wales           Timor          Kyushu
##              12              13              13              13              14
##          Taiwan
##              14
```

decreasing TRUE

```
head(sort(islands,decreasing = T)) ##
```

```
##          Asia        Africa North America South America    Antarctica
##         16988         11506          9390          6795          5500
##        Europe
##          3745
```

5.4.2 order

order() order() iris$Sepal.Length 14 iris$Sepal.Length 14

```
order(iris$Sepal.Length)
```

```
##   [1]  14   9  39  43  42   4   7  23  48   3  30  12  13  25  31  46   2  10
##  [19]  35  38  58 107   5   8  26  27  36  41  44  50  61  94   1  18  20  22
##  [37]  24  40  45  47  99  28  29  33  60  49   6  11  17  21  32  85  34  37
##  [55]  54  81  82  90  91  65  67  70  89  95 122  16  19  56  80  96  97 100
##  [73] 114  15  68  83  93 102 115 143  62  71 150  63  79  84  86 120 139  64
##  [91]  72  74  92 128 135  69  98 127 149  57  73  88 101 104 124 134 137 147
## [109]  52  75 112 116 129 133 138  55 105 111 117 148  59  76  66  78  87 109
## [127] 125 141 145 146  77 113 144  53 121 140 142  51 103 110 126 130 108 131
```

```
## [145] 106 118 119 123 136 132
```

```
iris$Sepal.Length[14]
```

```
## [1] 4.3
```

decreasing TRUE iris$Sepal.Length 132 iris$Sepal.Length 132

```
order(iris$Sepal.Length,decreasing = T)
```

```
##   [1] 132 118 119 123 136 106 131 108 110 126 130 103  51  53 121 140 142  77
##  [19] 113 144  66  78  87 109 125 141 145 146  59  76  55 105 111 117 148  52
##  [37]  75 112 116 129 133 138  57  73  88 101 104 124 134 137 147  69  98 127
##  [55] 149  64  72  74  92 128 135  63  79  84  86 120 139  62  71 150  15  68
##  [73]  83  93 102 115 143  16  19  56  80  96  97 100 114  65  67  70  89  95
##  [91] 122  34  37  54  81  82  90  91   6  11  17  21  32  85  49  28  29  33
## [109]  60   1  18  20  22  24  40  45  47  99   5   8  26  27  36  41  44  50
## [127]  61  94   2  10  35  38  58 107  12  13  25  31  46   3  30   4   7  23
## [145]  48  42   9  39  43  14
```

```
iris$Sepal.Length[132]
```

```
## [1] 7.9
```

order iris

```
head(iris) ##
```

```
##   Sepal.Length Sepal.Width Petal.Length Petal.Width Species
## 1          5.1         3.5          1.4         0.2  setosa
## 2          4.9         3.0          1.4         0.2  setosa
## 3          4.7         3.2          1.3         0.2  setosa
## 4          4.6         3.1          1.5         0.2  setosa
## 5          5.0         3.6          1.4         0.2  setosa
## 6          5.4         3.9          1.7         0.4  setosa
```

```
head(iris[order(iris$Sepal.Length),]) ## Sepal.Length
```

```
##    Sepal.Length Sepal.Width Petal.Length Petal.Width Species
## 14          4.3         3.0          1.1         0.1  setosa
## 9           4.4         2.9          1.4         0.2  setosa
## 39          4.4         3.0          1.3         0.2  setosa
## 43          4.4         3.2          1.3         0.2  setosa
## 42          4.5         2.3          1.3         0.3  setosa
## 4           4.6         3.1          1.5         0.2  setosa
```

```
head(iris[order(iris$Sepal.Length,decreasing = T),]) ##
```

```
##     Sepal.Length Sepal.Width Petal.Length Petal.Width   Species
## 132          7.9         3.8          6.4         2.0 virginica
## 118          7.7         3.8          6.7         2.2 virginica
## 119          7.7         2.6          6.9         2.3 virginica
## 123          7.7         2.8          6.7         2.0 virginica
## 136          7.7         3.0          6.1         2.3 virginica
## 106          7.6         3.0          6.6         2.1 virginica
```

5.5

- Row rbind()
- Column cbind()

rbind() cbind() :

```
rbind(c(1,2,3), #
      c(4,5,6)  #
      )
```

```
##      [,1] [,2] [,3]
## [1,]    1    2    3
## [2,]    4    5    6
```

:

```
irisAdd<-rbind(iris, #
      c(1,1,1,1,"versicolor")  #
      )
tail(irisAdd)
```

```
##     Sepal.Length Sepal.Width Petal.Length Petal.Width    Species
## 146          6.7           3          5.2         2.3  virginica
## 147          6.3         2.5            5         1.9  virginica
## 148          6.5           3          5.2           2  virginica
## 149          6.2         3.4          5.4         2.3  virginica
## 150          5.9           3          5.1         1.8  virginica
## 151            1           1            1           1 versicolor
```

:

```
cbind(c(1,2,3), #
      c(4,5,6)  #
      )
```

```
##      [,1] [,2]
## [1,]    1    4
## [2,]    2    5
## [3,]    3    6
```

:

```
irisAdd<-cbind(iris, #
      rep("Add",nrow(iris))  #
      )
tail(irisAdd)
```

```
##     Sepal.Length Sepal.Width Petal.Length Petal.Width   Species
## 145          6.7         3.3          5.7         2.5 virginica
## 146          6.7         3.0          5.2         2.3 virginica
## 147          6.3         2.5          5.0         1.9 virginica
## 148          6.5         3.0          5.2         2.0 virginica
## 149          6.2         3.4          5.4         2.3 virginica
## 150          5.9         3.0          5.1         1.8 virginica
##     rep("Add", nrow(iris))
## 145                    Add
## 146                    Add
## 147                    Add
## 148                    Add
## 149                    Add
## 150                    Add
```

5.6

R reshape2(Wickham, 2016a) package

- melt(/ ,id.vars=)
- dcast(/ , ~)

airquality Ozone, Solar.R, Wind, Temp, Month, Day (Column) Month Day variable value

```
library(reshape2)
head(airquality)
```

```
##   Ozone Solar.R Wind Temp Month Day
## 1    41     190  7.4   67     5   1
## 2    36     118  8.0   72     5   2
## 3    12     149 12.6   74     5   3
## 4    18     313 11.5   62     5   4
## 5    NA      NA 14.3   56     5   5
## 6    28      NA 14.9   66     5   6
```

```
airqualityM<-melt(airquality,id.vars = c("Month","Day")) ##   "Month","Day"
head(airqualityM)
```

```
##   Month Day variable value
## 1     5   1    Ozone    41
## 2     5   2    Ozone    36
## 3     5   3    Ozone    12
## 4     5   4    Ozone    18
## 5     5   5    Ozone    NA
## 6     5   6    Ozone    28
```

airqualityM Month, Day, variable, value (Column) variable value :

```
library(reshape2)
##  "Month","Day"     variable
airqualityCast<-dcast(airqualityM, Month +Day~variable)
head(airqualityCast)
```

```
##   Month Day Ozone Solar.R Wind Temp
## 1     5   1    41     190  7.4   67
## 2     5   2    36     118  8.0   72
## 3     5   3    12     149 12.6   74
## 4     5   4    18     313 11.5   62
## 5     5   5    NA      NA 14.3   56
## 6     5   6    28      NA 14.9   66
```

5.7

(Missing Value) is.na() NA TRUE

```
naVec<-c("a","b",NA,"d","e")
is.na(naVec)
```

```
## [1] FALSE FALSE  TRUE FALSE FALSE
```

```
naVec[!is.na(naVec)] ##   is.na()  FALSE
```

```
## [1] "a" "b" "d" "e"
```

complete.cases TRUE

```
head(airquality)
```

```
##   Ozone Solar.R Wind Temp Month Day
## 1    41     190  7.4   67     5   1
## 2    36     118  8.0   72     5   2
## 3    12     149 12.6   74     5   3
## 4    18     313 11.5   62     5   4
## 5    NA      NA 14.3   56     5   5
## 6    28      NA 14.9   66     5   6
```

```
complete.cases(airquality)
```

```
##   [1]  TRUE  TRUE  TRUE  TRUE FALSE FALSE  TRUE  TRUE  TRUE FALSE FALSE  TRUE
##  [13]  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE
##  [25] FALSE FALSE FALSE  TRUE  TRUE  TRUE  TRUE FALSE FALSE FALSE FALSE FALSE
##  [37] FALSE  TRUE FALSE  TRUE  TRUE FALSE FALSE  TRUE FALSE FALSE  TRUE  TRUE
##  [49]  TRUE  TRUE  TRUE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
##  [61] FALSE  TRUE  TRUE  TRUE FALSE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE FALSE
##  [73]  TRUE  TRUE FALSE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE FALSE FALSE
##  [85]  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE FALSE
##  [97] FALSE FALSE  TRUE  TRUE  TRUE FALSE FALSE  TRUE  TRUE  TRUE FALSE  TRUE
## [109]  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE FALSE  TRUE  TRUE  TRUE FALSE  TRUE
## [121]  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE
## [133]  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE
## [145]  TRUE  TRUE  TRUE  TRUE  TRUE FALSE  TRUE  TRUE  TRUE
```

```
head(airquality[complete.cases(airquality),]) ##   complete.cases()  TRUE
```

```
##   Ozone Solar.R Wind Temp Month Day
## 1    41     190  7.4   67     5   1
## 2    36     118  8.0   72     5   2
## 3    12     149 12.6   74     5   3
## 4    18     313 11.5   62     5   4
## 7    23     299  8.6   65     5   7
## 8    19      99 13.8   59     5   8
```

skydome20 R –(10) (Impute Missing Value)

Chapter 6

6.1

6.2 data.table

You can label chapter and section titles using `{#label}` after them, e.g., we can reference Chapter 1. If you do not manually label them, there will be automatic labels anyway, e.g., Chapter **??**.

Figures and tables with captions will be placed in `figure` and `table` environments, respectively.

```
par(mar = c(4, 4, .1, .1))
plot(pressure, type = 'b', pch = 19)
```

Reference a figure by its code chunk label with the `fig:` prefix, e.g., see Figure **??**. Similarly, you can reference tables generated from `knitr::kable()`, e.g., see Table **??**.

```
knitr::kable(
  head(iris, 20), caption = 'Here is a nice table!',
  booktabs = TRUE
)
```

You can write citations, too. For example, we are using the **bookdown** package (Xie, 2016) in this sample book, which was built on top of R Markdown and **knitr** (Xie, 2015).

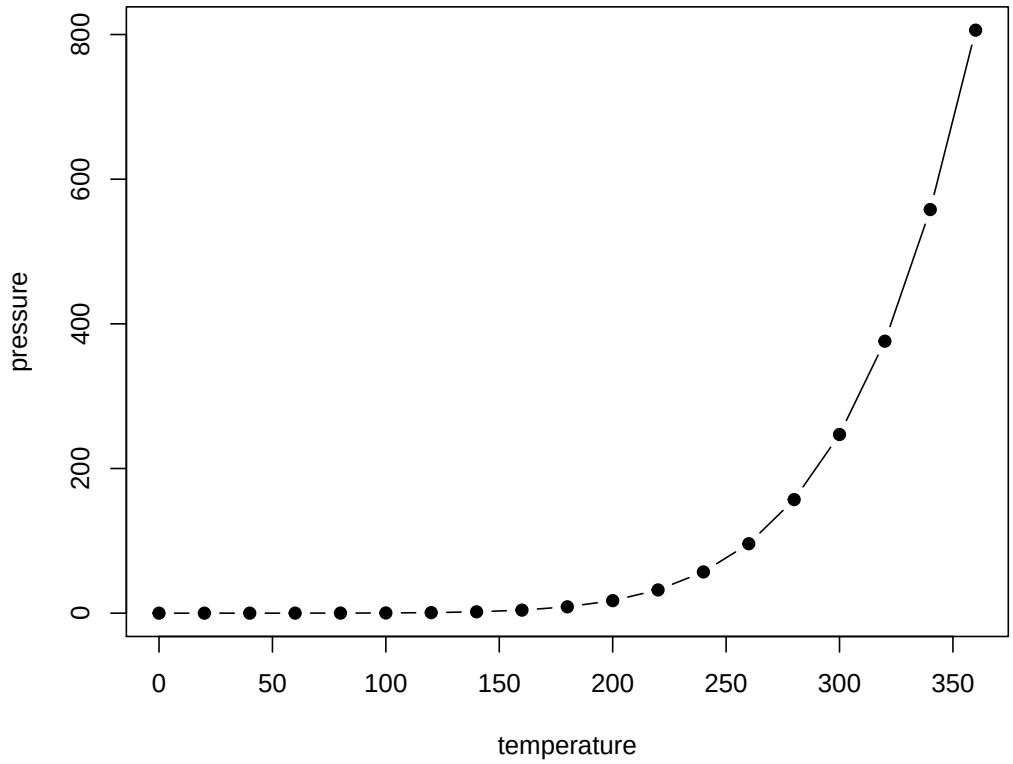


Figure 6.1: Here is a nice figure!

Table 6.1: Here is a nice table!

Sepal.Length	Sepal.Width	Petal.Length	Petal.Width	Species
5.1	3.5	1.4	0.2	setosa
4.9	3.0	1.4	0.2	setosa
4.7	3.2	1.3	0.2	setosa
4.6	3.1	1.5	0.2	setosa
5.0	3.6	1.4	0.2	setosa
5.4	3.9	1.7	0.4	setosa
4.6	3.4	1.4	0.3	setosa
5.0	3.4	1.5	0.2	setosa
4.4	2.9	1.4	0.2	setosa
4.9	3.1	1.5	0.1	setosa
5.4	3.7	1.5	0.2	setosa
4.8	3.4	1.6	0.2	setosa
4.8	3.0	1.4	0.1	setosa
4.3	3.0	1.1	0.1	setosa
5.8	4.0	1.2	0.2	setosa
5.7	4.4	1.5	0.4	setosa
5.4	3.9	1.3	0.4	setosa
5.1	3.5	1.4	0.3	setosa
5.7	3.8	1.7	0.3	setosa
5.1	3.8	1.5	0.3	setosa

Chapter 7

You can label chapter and section titles using `{#label}` after them, e.g., we can reference Chapter 1. If you do not manually label them, there will be automatic labels anyway, e.g., Chapter **??**.

Figures and tables with captions will be placed in `figure` and `table` environments, respectively.

```
par(mar = c(4, 4, .1, .1))
plot(pressure, type = 'b', pch = 19)
```

Reference a figure by its code chunk label with the `fig:` prefix, e.g., see Figure **??**. Similarly, you can reference tables generated from `knitr::kable()`, e.g., see Table **??**.

```
knitr::kable(
  head(iris, 20), caption = 'Here is a nice table!',
  booktabs = TRUE
)
```

You can write citations, too. For example, we are using the **bookdown** package (Xie, 2016) in this sample book, which was built on top of R Markdown and **knitr** (Xie, 2015).

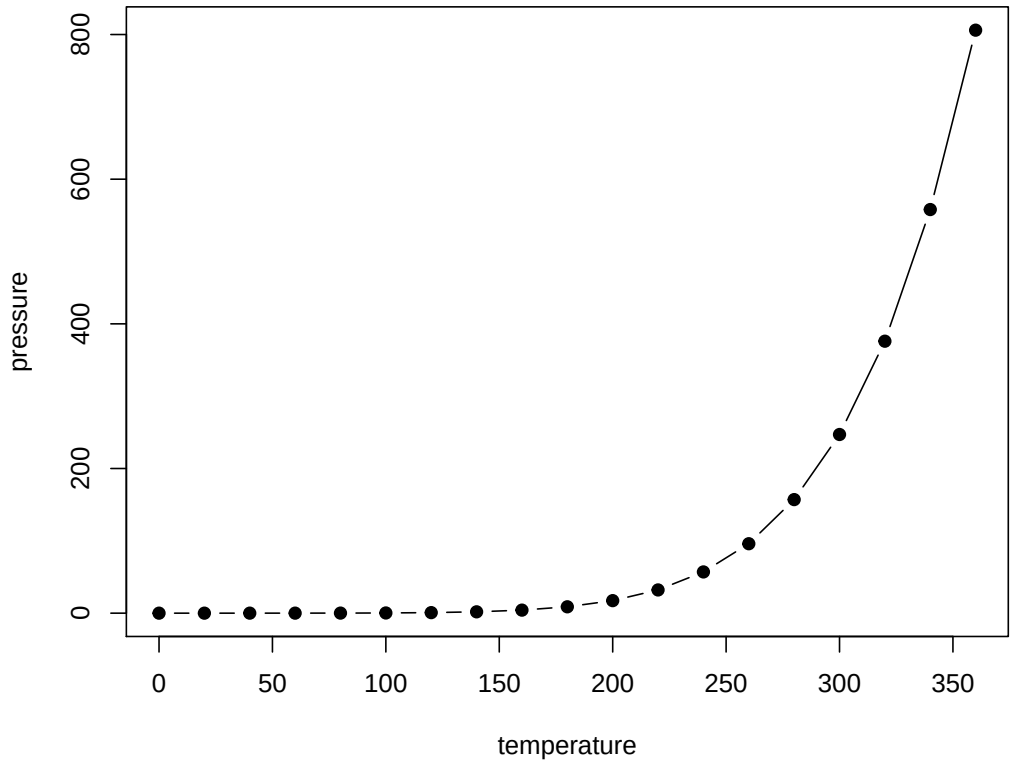


Figure 7.1: Here is a nice figure!

Table 7.1: Here is a nice table!

Sepal.Length	Sepal.Width	Petal.Length	Petal.Width	Species
5.1	3.5	1.4	0.2	setosa
4.9	3.0	1.4	0.2	setosa
4.7	3.2	1.3	0.2	setosa
4.6	3.1	1.5	0.2	setosa
5.0	3.6	1.4	0.2	setosa
5.4	3.9	1.7	0.4	setosa
4.6	3.4	1.4	0.3	setosa
5.0	3.4	1.5	0.2	setosa
4.4	2.9	1.4	0.2	setosa
4.9	3.1	1.5	0.1	setosa
5.4	3.7	1.5	0.2	setosa
4.8	3.4	1.6	0.2	setosa
4.8	3.0	1.4	0.1	setosa
4.3	3.0	1.1	0.1	setosa
5.8	4.0	1.2	0.2	setosa
5.7	4.4	1.5	0.4	setosa
5.4	3.9	1.3	0.4	setosa
5.1	3.5	1.4	0.3	setosa
5.7	3.8	1.7	0.3	setosa
5.1	3.8	1.5	0.3	setosa

Chapter 8

8.1 R + Hadoop

8.2 RHadoop (Cloudera)

2016/05/12

8.2.1 /

- Cloudera Hadoop Platform: CDH-5.4.5
- R for Linux 3.3.0
- RStudio Server
- RHadoop (latest version on May 12, 2016)
  - ravro-1.0.3
  - plyrmr-0.6.0
  - rmr-3.3.1
  - rhdfs-1.0.8
  - rhbase-1.2.1

8.2.2

- RHadoop
- RHadoop
- Setting persistent environment variable in CentOS 7 issue
- How to resolve “Permission denied” errors in CDH

8.2.3

1. Cloudera CDH QuickStart VM Cloudera VM
2. R
3. RHadoop RHadoop
4. RStudio Server

8.2.3.1 Cloudera CDH QuickStart VM

Cloudera CDH QuickStart VM Cloudera Linux Hadoop Hadoop

VM Virtural Box

- Cloudera CDH QuickStart VM
- Virtural Box

Cloudera CDH QuickStart VM

8.2.3.2 R

- Cloudera CDH Linux CentOS
- Extra Packages for Enterprise Linux (EPEL) `sudo yum install epel-release`
- EPRL yum R Terminal

```
sudo yum install epel-release
sudo yum update
sudo yum install R
```

8.2.3.3 RHadoop-1

`HADOOP_CMD` `HADOOP_STREAMING` `HADOOP_STREAMING`

1. `HADOOP_STREAMING`

   ```
   find / -name hadoop-streaming-*.jar
   ```

2. `HADOOP_CMD` `HADOOP_STREAMING`

   ```
   echo export HADOOP_CMD="/usr/bin/hadoop">/etc/profile.d/hadoopenv.sh
   echo export HADOOP_STREAMING=
   "/opt/cloudera/parcels/CDH-5.4.5-1.cdh5.4.5.p0.7/lib/hadoop-mapreduce/
       hadoop-streaming-2.6.0-cdh5.4.5.jar" > /etc/profile.d/hadoopenv.sh
   chmod 0755 /etc/profile.d/hadoopenv.sh
   ```

8.2.3.4 RHadoop-2 rmr2

- Node
- packages Depends Imports packages
- packages R Terminal `R` R

```
install.packages(c("methods","Rcpp", "RJSONIO", "digest", "functional",
                   "reshape2","stringr", "plyr", "caTools","quickcheck","testthat"),
                 dependencies=TRUE, repos='http://cran.rstudio.com/')
```

- `q()` R
- rmr2
- `rmr2_2.3.0.tar.gz`

```
sudo R CMD INSTALL rmr2_2.3.0.tar.gz
```

8.2.3.5 RHadoop-3 rhdfs

- R Node
- Check JDK JDK 1.8.0_91
- Check JAVA_HOME

```
echo $JAVA_HOME
```

/usr/java/jdk1.8.0_91

```
echo export JAVA_HOME="/usr/java/jdk1.8.0_91">/etc/profile.d/jdkenv.sh
```

R JAVA Terminal

```
R CMD javareconf
```

R Terminal R R rJava package

```
install.packages("rJava",dependencies=TRUE, repos='http://cran.rstudio.com/')
```

R rhdfs rhdfs

- /usr/bin/hadoop HADOOP_CMD
- rhdfs_1.0.8.tar.gz

```
sudo HADOOP_CMD=/usr/bin/hadoop R CMD INSTALL rhdfs_1.0.8.tar.gz
```

8.2.4

- hdfs
- user01

```
sudo -u hdfs hadoop fs -mkdir /user/user01
sudo -u hdfs hadoop fs -chown user01 /user/user01
```

8.2.5

R

```
Sys.setenv(HADOOP_CMD="/usr/bin/hadoop")
Sys.setenv(HADOOP_STREAMING="/opt/cloudera/parcels/CDH-5.4.5-1.cdh5.4.5.p0.7/lib/hadoop-mapreduce/hadoo
library(rmr2)
#test mapreduce
small.ints = to.dfs(1:100)
out<-mapreduce(
    input = small.ints,
    map = function(., v) cbind(v, v^2))
head(from.dfs(out))
```

8.2.6 RStudio Server

Terminal

- https://download2.rstudio.org/rstudio-server-rhel-0.99.896-x86_64.rpm Check

```
wget https://download2.rstudio.org/rstudio-server-rhel-0.99.896-x86_64.rpm
sudo yum install --nogpgcheck rstudio-server-rhel-0.99.896-x86_64.rpm
```

http://localhost:8787/ RStudio Server

8.3 RHadoop MapReduce: easy word count

```
Debate<-readLines("https://raw.githubusercontent.com/yijutseng/BigDataCGUIM/master/RepDebateMiami.txt")
DebateSplit<-unlist(strsplit(tolower(Debate),split = ' |\\.|\\,|\\?'))
#table(DebateSplit)
```

```
DebateSplitDFS = to.dfs(DebateSplit)
result = mapreduce(
    input = DebateSplitDFS,
    map = function(.,v) keyval(v, 1),
    reduce = function(k,vv) keyval(k, sum(vv)))
head(result)
```

8.4 R + Spark

Chapter 9

9.1 R

9.2 RStudio

Yi-Ju Tseng

Lab:

Bibliography

Dowle, M., Srinivasan, A., Short, T., with contributions from R Saporta, S. L., and Antonyan, E. (2016). *data.table: Extension of Data.frame*. R package version 1.9.8.

Grolemund, G., Spinu, V., and Wickham, H. (2016). *lubridate: Make Dealing with Dates a Little Easier*. R package version 1.6.0.

Oehlschlägel, J. (2015). *bit64: A S3 Class for Vectors of 64bit Integers*. R package version 0.9-5.

Wickham, H. (2016a). *reshape2: Flexibly Reshape Data: A Reboot of the Reshape Package*. R package version 1.4.2.

Wickham, H. (2016b). *stringr: Simple, Consistent Wrappers for Common String Operations*. R package version 1.1.0.

Xie, Y. (2015). *Dynamic Documents with R and knitr*. Chapman and Hall/CRC, Boca Raton, Florida, 2nd edition. ISBN 978-1498716963.

Xie, Y. (2016). *bookdown: Authoring Books and Technical Documents with R Markdown*. R package version 0.3.9.