R

Yi-Ju Tseng

2017-02-06

Contents

M R Hadoop EcoSystems

Chapter 1

R 101

R

1.1 R

R R **Packages** 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 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
```

1.3

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

1.3.1 numeric

( )

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

2^53 bit64 package (?) 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
```

1.3.2 character

"

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

1.3.3 logic

TRUE T **FALSE F**

```
boolT<-TRUE
boolT1<-T
```

```
boolF<-FALSE
boolF1<-F
```

1.3.4 (Date)

Sys.Date()

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

```
## [1] "2017-02-06"
```

lubridate(?) 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"
```

The Yhat Blog

1.4

1.4.1

R

- +
- -
- *
- /

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

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

```
num1-num2
```

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

```
num1*num2
```

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

```
num1/num2
```

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

1.4.2

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
```

1.5 rbookdown

You can label chapter and section titles using `{#label}` after them, e.g., we can reference Chapter **??**. 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
)
```

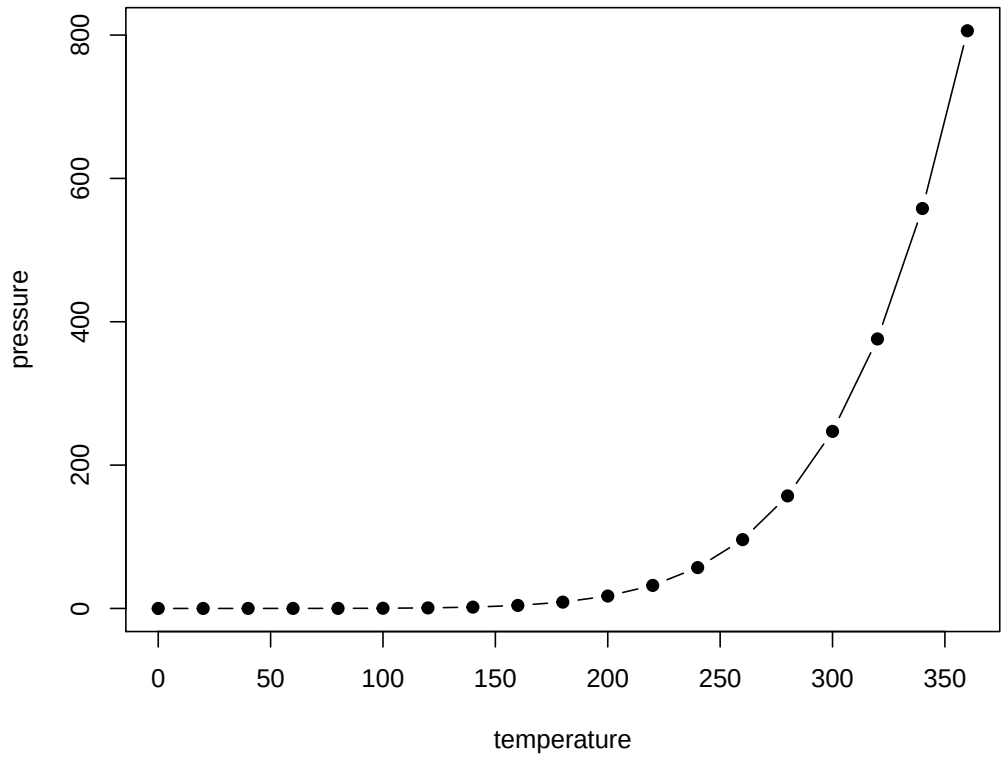

Figure 1.1: Here is a nice figure!

Table 1.1: 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

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

Chapter 2

R

Here is a review of existing methods. **(factor)**

2.1 vector

c()

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

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

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

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

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

vec

2.1.1

2.1.2

2.2 factor

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

2.3 list

2.4 matrix

2.5 data.frame

2.6 data.table ( )

Chapter 3

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

3.1

3.1.1 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
## 1    NBA    Quincy Acy  NYK       SF          68
## 2    NBA  Jordan Adams  MEM       SG          30
## 3    NBA  Steven Adams  OKL        C          70
## 4    NBA   Jeff Adrien  MIN       PF          17
## 5    NBA Arron Afflalo  POR       SG          78
## 6    NBA Alexis Ajinca  NOR        C          68
```

```
##   TotalMinutesPlayed FieldGoalsMade
## 1               1288            152
## 2                249             35
## 3               1776            217
## 4                215             19
## 5               2502            375
## 6                956            181
##   FieldGoalsAttempted ThreesMade ThreesAttempted
## 1                 331         18              60
## 2                  86         10              25
## 3                 399          0               2
## 4                  44          0               0
## 5                 884        118             333
## 6                 328          0               0
##   FreeThrowsMade FreeThrowsAttempted OffensiveRebounds
## 1             76                  97                79
## 2             14                  23                 9
## 3            103                 205               199
## 4             22                  38                23
## 5            167                 198                27
## 6             81                  99               104
##   TotalRebounds Assists Steals Turnovers Blocks
## 1           301      68     27        60     22
## 2            28      16     16        14      7
## 3           522      65     38        99     85
## 4            77      15      4         9      9
## 5           247     129     41       116      7
## 6           315      47     21        69     51
##   PersonalFouls Disqualifications TotalPoints
## 1           147                 1         398
## 2            24                 0          94
## 3           222                 3         537
## 4            30                 0          60
## 5           167                 1        1035
## 6           151                 0         443
##   Technicals Ejections FlagrantFouls GamesStarted
## 1          5         0             0           22
## 2          0         0             0            0
## 3          3         0             0           67
## 4          0         0             0            0
## 5          0         0             0           72
## 6          1         0             0            8
```

3.1.2 Raw Data

3.1.2.1 Html

3.1.2.2 Facebook

3.1.2.3 !?

3.2

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

3.2.1 Open Data

http://bit.ly/1U3pVoI

3.3

3.3.1 (.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)
```

3.4 read.csv

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

3.4.1

readLines,

3.4.2 R

load, R Ex: iris

3.4.3 R

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

3.5

3.5.1 (.csv / Tab)

write.table

3.5.2

writeLines

3.5.3 R

save

3.5.4 R

dump

3.6 R

- # Row
-
- Column
-

Column

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

3.7 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")
```

3.8 Open Data

3.9

Chapter 4

Some *significant* applications are demonstrated in this chapter.

4.1 Example one

4.2 Example two