Introduction to R, Part II

Some useful functions

Tim Frasier

Getting data into, and out of, R

Getting Data Into R

`read.table` function

- Needs at least 3 arguments
  - Name of the file to read in (in quotes)
  - Whether or not the file has a header row
  - How fields are delimited (commas, tabs, space, etc.)

Getting Data Into R

`read.table` function

```
afdata = read.table("agefat.csv", header = TRUE, sep = ",")
```

x	age	fat	sex
1	24	15.5	male
2	37	20.9	male
3	41	18.6	male
4	60	28	male
5	31	34.7	female
6	39	30.2	female
…	…	…	…

Getting Data Out of R

`write.csv` function

```
write.csv(afdata, “myafdata.csv”)
```

Getting Data Out of R

`write.csv` function

- By default, R will put quotes around anything it thinks is text
- Also adds a column of row "names"
    - This can be annoying

```
"","X","age","fat","sex"
"1",1,24,15.5,"male"
"2",2,37,20.9,"male"
"3",3,41,18.6,"male"
"4",4,60,28,"male"
"5",5,31,34.7,"female"
"6",6,39,30.2,"female"
"7",7,58,21.3,"male"
"8",8,23,9.5,"male"
```

Getting Data Out of R

`write.csv` function

```
write.csv(afdata, file = "myafdata.csv", row.names = FALSE, quote = FALSE)
```

```
X,age,fat,sex
1,24,15.5,male
2,37,20.9,male
3,41,18.6,male
4,60,28,male
5,31,34.7,female
6,39,30.2,female
7,58,21.3,male
8,23,9.5,male
```

Some Utility Functions

Some Utility Functions

`summary()`

- Identifies the type, or "class" of the object, and returns appropriate summary information

Some Utility Functions

`summary()`

```
summary(afdata)
```

```
       X                age              fat              sex
 Min.   :1     Min.   :23.00   Min.   : 7.80   female :15
 1st Qu.:7     1st Qu.:37.00   1st Qu.:21.30   male   :10
 Median :13    Median :49.00   Median :29.10
 Mean   :13    Mean   :44.96   Mean   :27.37
 3rd Qu.:19    3rd Qu.:57.00   3rd Qu.:33.00
 Max.   :25    Max.   :61.00   Max.   :42.00
```

Some Utility Functions

`sd()`

- Not sure why `summary` doesn't provide standard deviation
- Can get with `sd`

```
sd(afdata$age)
[1] 13.14813
```

```
sd(afdata$fat)
[1] 8.679272
```

Some Utility Functions

`head()`

- Shows just the first few lines of an object (the "header")
  - Good for identifying the structure/format of large data files

```
head(afdata)
   X   age     fat     sex
1  1    24    15.5    male
2  2    37    20.9    male
3  3    41    18.6    male
4  4    60    28.0    male
5  5    31    34.7  female
6  6    39    30.2  female
```

Some Utility Functions

`tail()`

- Shows just the last few lines of an object (the "header")
  - Good for identifying the structure/format of large data files

```
tail(afdata)
    X   age     fat     sex
20  20  56      32.5    female
21  21  57      30.3    female
22  22  58      33.0    female
23  23  58      33.8    female
24  24  60      41.1    female
25  25  61      34.5    female
```

Some Utility Functions

`str()`

- Summarizes the structure of an object
  - Very useful for complex data sets (*e.g.*, lists and arrays)

```
str(afdata)

'data.frame':  25 obs. of 4 variables:
$ X : int   1 2 3 4 5 6 7 8 9 10 …
$ age   : int 24 37 41 60 31 39 58 23 23 27 …
$ fat   : num 15.5 20.9 18.6 28 34.7 30.2 21.3 9.5 27.9 7.8 …
$ sex   : Factor w/ 2 levels "female", "male": 2 2 2 2 1 1 2 2 1 2
```

Some Utility Functions

`length()`

- Returns the length of a vector
  - Very useful when you want to run the same function on files of different sizes

```
length(afdata[ ,1])
[1] 25
```

Some Utility Functions

`cbind()`

- Combines different data files as different columns in one new data frame

```
x = 1:10
y = c(“A”, “B”)
z = rep(y, times = 5)

new = cbind(x, z)
```

RStudio

Go to file/function

Project: (None)

afdata | Untitled1* | new

10 observations of 2 variables

	x	z
1	1	A
2	2	B
3	3	A
4	4	B
5	5	A
6	6	B
7	7	A
8	8	B
9	9	A
10	10	B

Environment | History

Import Dataset | Clear | List

Global Environment

Data

new — chr [1:10, 1:2] "1" "2" "3" "4" "5" "6" "7" "8" "9" ...

Values

x — int [1:10] 1 2 3 4 5 6 7 8 9 10

y — chr [1:2] "A" "B"

Files | Plots | Packages | Help | Viewer

R: Object Summaries | Find in Topic

summary {base} R Documentation

Object Summaries

Description

summary is a generic function used to produce result summaries of the results of various model fitting functions. The function invokes particular methods which depend on the class of the first argument.

Usage

```
summary(object, ...)

## Default S3 method:
summary(object, ..., digits = max(3, getOption("digits")-3))
## S3 method for class 'data.frame'
summary(object, maxsum = 7,
        digits = max(3, getOption("digits")-3), ...)

## S3 method for class 'factor'
summary(object, maxsum = 100, ...)
```

Console ~/R/

```
'citation()' on how to cite R or R packages in publications.

Type 'demo()' for some demos, 'help()' for on-line help, or
'help.start()' for an HTML browser interface to help.
Type 'q()' to quit R.

> x <- 1:10
> y <- c("A", "B")
> z <- rep(y, times = 5)
> z
 [1] "A" "B" "A" "B" "A" "B" "A" "B" "A" "B"
> new <- cbind(x, z)
> View(new)
>
```

Some Utility Functions

`rbind()`

- Combines different data files as different rows in one new data frame

Some Utility Functions

`rbind()`

- Combines different data files as different rows in one new data frame

```
x = 1:10
y = c("A", "B")
z = rep(y, times = 5)

new2 = rbind(x, z)
```

RStudio

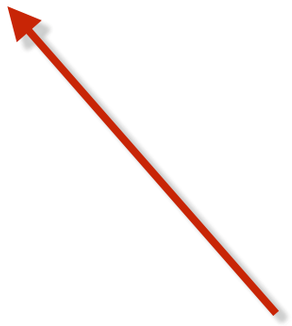
Go to file/function

Project: (None)

afdata | Untitled1* | new | new2

2 observations of 10 variables

	row.names	V1	V2	V3	V4	V5	V6	V7	V8	V9	V10
1	x	1	2	3	4	5	6	7	8	9	10
2	z	A	B	A	B	A	B	A	B	A	B



Environment | History

Import Dataset | Clear | List

Global Environment

Data

new	chr [1:10, 1:2] "1" "2" "3" "4" "5" "6" "7" "8" "9" ...
new2	chr [1:2, 1:10] "1" "A" "2" "B" "3" "A" "4" "B" "5" ...

Values

x	int [1:10] 1 2 3 4 5 6 7 8 9 10

Files | Plots | Packages | Help | Viewer

R: Object Summaries | Find in Topic

summary {base} R Documentation

Object Summaries

Description

summary is a generic function used to produce result summaries of the results of various model fitting functions. The function invokes particular methods which depend on the class of the first argument.

Usage

```
summary(object, ...)

## Default S3 method:
summary(object, ..., digits = max(3, getOption("digits")-3))
## S3 method for class 'data.frame'
summary(object, maxsum = 7,
        digits = max(3, getOption("digits")-3), ...)

## S3 method for class 'factor'
summary(object, maxsum = 100, ...)
```

Console ~/R/

```
Type 'demo()' for some demos, 'help()' for on-line help, or
'help.start()' for an HTML browser interface to help.
Type 'q()' to quit R.

> x <- 1:10
> y <- c("A", "B")
> z <- rep(y, times = 5)
> z
 [1] "A" "B" "A" "B" "A" "B" "A" "B" "A" "B"
> new <- cbind(x, z)
> View(new)
> new2 <- rbind(x, z)
> View(new2)
>
```

Some Utility Functions

`sample()`

- Can be use to randomize a list (key to simulations!)
  - Can sample with or without replacement

Some Utility Functions

`sample()`

- Can be use to randomize a list (key to simulations!)
  - Can sample with or without replacement

```
myList = 1:10

myList
```

```
 1  2  3  4  5  6  7  8  9 10
```

```
sample(myList, size = 10, replace = FALSE)
```

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

```
sample(myList, size = 10, replace = TRUE)
```

```
 [1] 10 10  9  6  3  4  1  8  3  3
```

Some Utility Functions

`print()`

- Prints text to screen
  - May not seem useful now, but key for checking code

```
s = “Bayesian analysis is cool!”
print(s)
[1] “Bayesian analysis is cool!”
```

Some Utility Functions

`?` - Help

- `?` then function name will open the help information for that function

```
?summary
```

RStudio

afdata | Untitled1*

25 observations of 4 variables

	X	age	fat	sex
1	1	24	15.5	male
2	2	37	20.9	male
3	3	41	18.6	male
4	4	60	28.0	male
5	5	31	34.7	female
6	6	39	30.2	female
7	7	58	21.3	male
8	8	23	9.5	male
9	9	23	27.9	female
10	10	27	7.8	male
11	11	27	17.8	male
12	12	39	31.4	female
13	13	41	25.9	male
14	14	45	27.4	male

Console ~/R/

```
Type 'license()' or 'licence()' for distribution details.

  Natural language support but running in an English locale

R is a collaborative project with many contributors.
Type 'contributors()' for more information and
'citation()' on how to cite R or R packages in publications.

Type 'demo()' for some demos, 'help()' for on-line help, or
'help.start()' for an HTML browser interface to help.
Type 'q()' to quit R.

> ?summary
>
```

Environment | History

Import Dataset | Clear | List

Global Environment

Environment is empty

Files | Plots | Packages | Help | Viewer

R: Object Summaries | Find in Topic

summary {base} R Documentation

Object Summaries

Description

summary is a generic function used to produce result summaries of the results of various model fitting functions. The function invokes particular methods which depend on the class of the first argument.

Usage

```
summary(object, ...)

## Default S3 method:
summary(object, ..., digits = max(3, getOption("digits")-3))
## S3 method for class 'data.frame'
summary(object, maxsum = 7,
        digits = max(3, getOption("digits")-3), ...)

## S3 method for class 'factor'
summary(object, maxsum = 100, ...)
```

Google is your friend!

- Can find help for just about anything in R
- Countless number of good online resources for R

Questions?