R masterclass

Hadley Wickham

Chief Scientist

RStudio

May 2015

CC BY NC

HELLO

my name is

Hadley

Winston Chang
Software Engineer
RStudio

Your turn

Introduce yourself to your neighbours. Who are you and what are you using R for?

Day 1

- ~~Introductions~~ & warm ups
- What makes a good function?
- Functional programming
- Object oriented programming (S3)
- Metaprogramming

Advanced R

- This day covers the most important parts of the “Advanced R” book.
- Online for free at http://adv-r.had.co.nz/
- Buy physical/kindle version from http://amzn.com/1466586966

Warmup

Your turn

What are the four common types of atomic vectors? (Bonus points for the two uncommon types)

Brainstorm with your neighbours for 1 minute.

character

double
integer

logical

raw

complex

```
as.character(c(T, F))
as.character(seq_len(5))
as.logical(c(0, 1, 100))
as.logical(c("T", "F", "a"))
as.double(c("A", "100"))
as.double(c(T, F))
```

When vectors of different types occur in an expression, they will be automatically coerced to the same type: character > numeric > logical

`typeof()`

`length()` A scalar is a vector of length 1

Technically, these are all **atomic** vectors

```
typeof(1:10)
typeof("a")
typeof(1.5)
```

```
# We'll talk about this later:
typeof(mtcars)
typeof(factor(1:10))
```

~~mode()~~

~~storage.mode()~~

Designed for S compatibility!

```r
# BEWARE!
# What does is.numeric() do?
x <- 1
y <- 1L
is.integer(x)
is.integer(y)
is.numeric(x)
is.numeric(x)
```

```r
# What does is.vector() do?
is.vector(x)
comment(x) <- "This is a comment!"
is.vector(x)

# Assume a is.*() does something surprising
# until proven otherwise (by reading the docs)
```

```
# Most useful coercion is from logical
# to numeric
x <- c(FALSE, TRUE, FALSE, TRUE, TRUE)
sum(x)
mean(x)

x <- sample(1e4, 1e3)
sum(x > 500)
mean(x > 500)
```

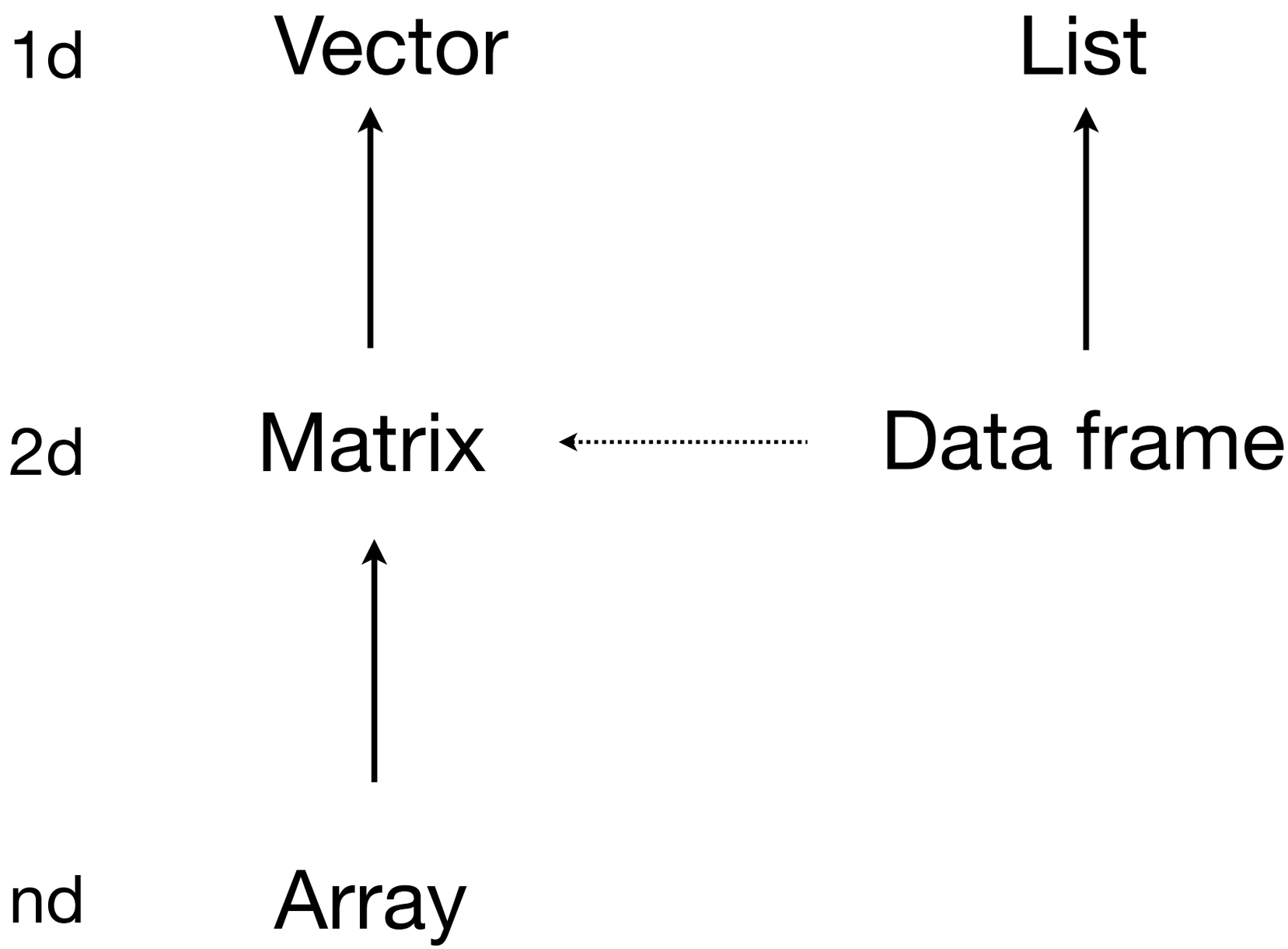
Your turn

How is a list different from an atomic vector?

How is a data frame different from a matrix?

How do you examine the structure of an object?

Brainstorm with your neighbours for 1 minute.


1d
Vector
List
2d
Matrix
Data frame
nd
Array
Same types
Different types

str()

Your turn

What are the six types of thing that you can put inside []?

What's the difference between `[`, `[[` and `$`? What does `drop = FALSE` do?

Brainstorm with your neighbour for 1 minutes.

blank	include all
integer	**+ve**: include **0**: drop all **-ve**: exclude
logical	keep TRUEs
character	lookup by name

	Simplifying	Preserving
Vectors	`x[[1]]`	`x[1:4]`
Matrices/ Data frame	`x[1:4, ]`	`x[1:4, , drop = F]`
Lists	`x[[1]]` `x$name`	`x[1]`

If list `x` is a train carrying objects, then `x[[5]]` is the object in car 5; `x[4:6]` is a train of cars 4-6.

http://twitter.com/#!/RLangTip/status/118339256388304896

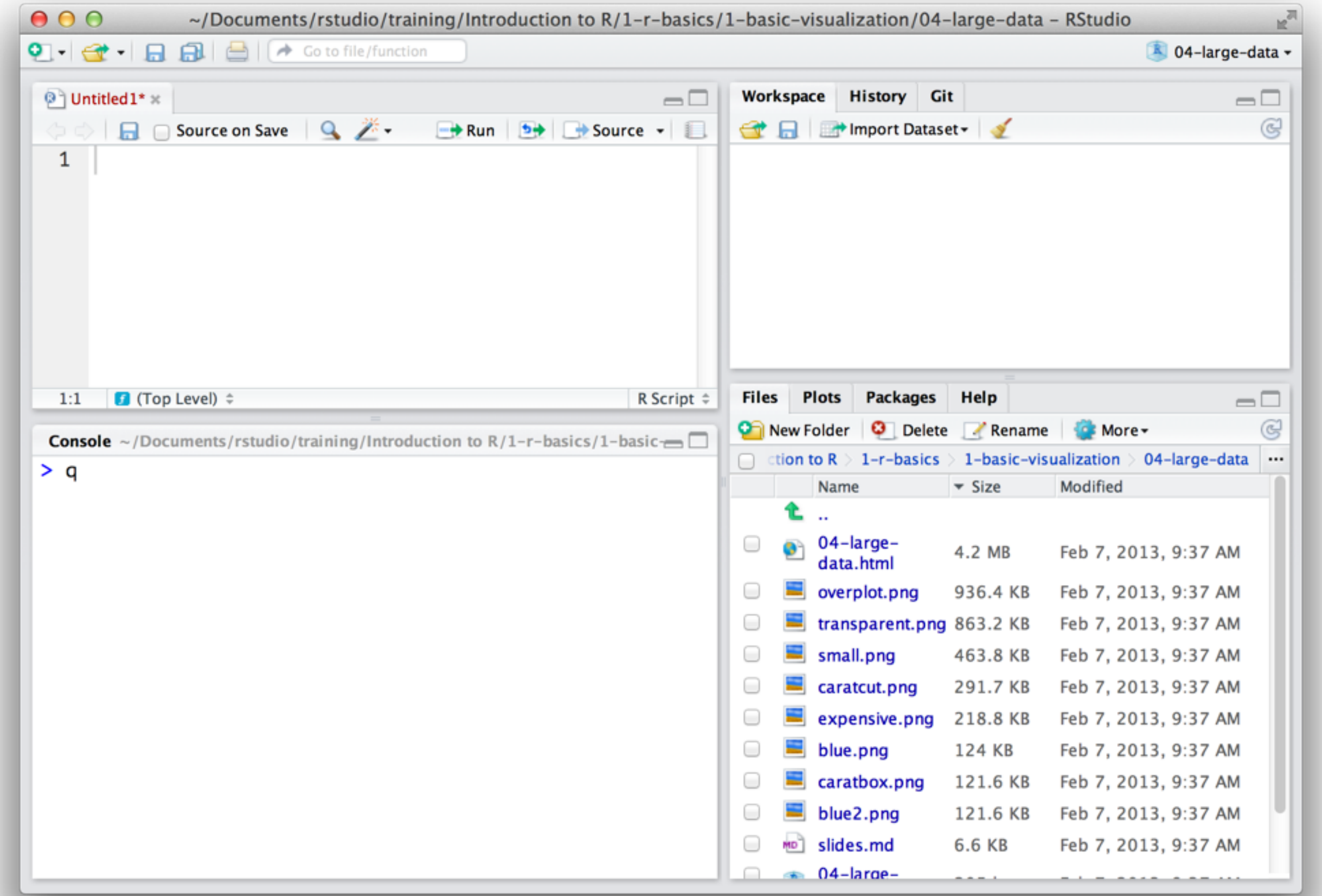
Your turn

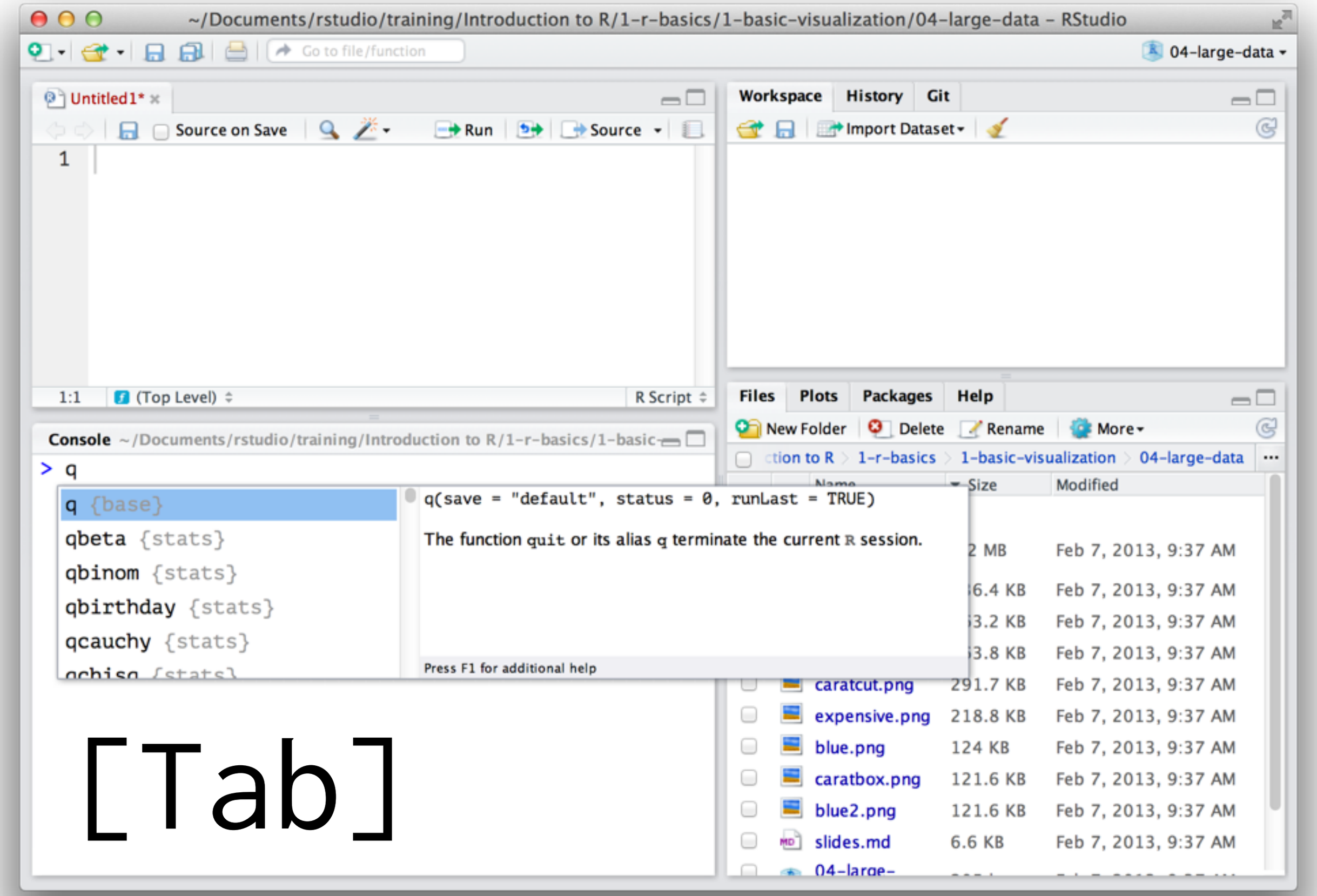
What's the difference between `x[] <- y` and `x <- y` ?

```
x <- factor(c("a", "b", "c"))
x <- c("c", "b", "a")
```

```
x <- factor(c("a", "b", "c"))
x[] <- c("c", "b", "a")
```

Rstudio


~/Documents/rstudio/training/Introduction to R/1-r-basics/1-basic-visualization/04-large-data - RStudio
Go to file/function
04-large-data
Untitled1*
Source on Save
Run
Source
1
1:1
(Top Level)
R Script
Console ~/Documents/rstudio/training/Introduction to R/1-r-basics/1-basic-
> q
Workspace
History
Git
Import Dataset
Files
Plots
Packages
Help
New Folder
Delete
Rename
More
tion to R
1-r-basics
1-basic-visualization
04-large-data
Name
Size
Modified
..
04-large-data.html 4.2 MB Feb 7, 2013, 9:37 AM
overplot.png 936.4 KB Feb 7, 2013, 9:37 AM
transparent.png 863.2 KB Feb 7, 2013, 9:37 AM
small.png 463.8 KB Feb 7, 2013, 9:37 AM
caratcut.png 291.7 KB Feb 7, 2013, 9:37 AM
expensive.png 218.8 KB Feb 7, 2013, 9:37 AM
blue.png 124 KB Feb 7, 2013, 9:37 AM
caratbox.png 121.6 KB Feb 7, 2013, 9:37 AM
blue2.png 121.6 KB Feb 7, 2013, 9:37 AM
slides.md 6.6 KB Feb 7, 2013, 9:37 AM
04-large-


~/Documents/rstudio/training/Introduction to R/1-r-basics/1-basic-visualization/04-large-data - RStudio
Go to file/function
04-large-data
Untitled1*
Source on Save
Run
Source
1
1:1
(Top Level)
R Script
Workspace
History
Git
Import Dataset
Files
Plots
Packages
Help
New Folder
Delete
Rename
More
1-r-basics
1-basic-visualization
04-large-data
Name
Size
Modified
Console ~/Documents/rstudio/training/Introduction to R/1-r-basics/1-basic-
> q
q {base}
qbeta {stats}
qbinom {stats}
qbirthday {stats}
qcauchy {stats}
q(save = "default", status = 0, runLast = TRUE)
The function quit or its alias q terminate the current R session.
Press F1 for additional help
2 MB
Feb 7, 2013, 9:37 AM
6.4 KB
Feb 7, 2013, 9:37 AM
3.2 KB
Feb 7, 2013, 9:37 AM
3.8 KB
Feb 7, 2013, 9:37 AM
caratcut.png
291.7 KB
Feb 7, 2013, 9:37 AM
expensive.png
218.8 KB
Feb 7, 2013, 9:37 AM
blue.png
124 KB
Feb 7, 2013, 9:37 AM
caratbox.png
121.6 KB
Feb 7, 2013, 9:37 AM
blue2.png
121.6 KB
Feb 7, 2013, 9:37 AM
slides.md
6.6 KB
Feb 7, 2013, 9:37 AM
[Tab]

~/Documents/rstudio/training/Introduction to R/1-r-basics/1-basic-visualization/04-large-data - RStudio

Go to file/function

04-large-data

Untitled1*

Source on Save | Run | Source

```
qplot(table - depth, data = diamonds,
qplot(day, data = email)
qplot(day, mails, data = daily, geom = "line", colo
qplot(day, mails, data = daily, geom = "smooth", co
qplot(day, variants, data = daily, geom = "line", c
qplot(wday, hour, data = wh, size = freq)
qplot(mpg, wt, data = mtcars)
qplot(mpg, wt, data = mtcars, colour = cyl)
```

```
> q
```

Workspace | History | Git

Import Dataset

Files | Plots | Packages | Help

New Folder | Delete | Rename | More

ction to R > 1-r-basics > 1-basic-visualization > 04-large-data

Name	Size	Modified
..		
04-large-data.html	4.2 MB	Feb 7, 2013, 9:37 AM
overplot.png	936.4 KB	Feb 7, 2013, 9:37 AM
transparent.png	863.2 KB	Feb 7, 2013, 9:37 AM
small.png	463.8 KB	Feb 7, 2013, 9:37 AM
caratcut.png	291.7 KB	Feb 7, 2013, 9:37 AM
	B	Feb 7, 2013, 9:37 AM
		Feb 7, 2013, 9:37 AM
	B	Feb 7, 2013, 9:37 AM
	B	Feb 7, 2013, 9:37 AM
		Feb 7, 2013, 9:37 AM
04-large-		

[Cmd/Ctrl + ↑]

~/Documents/rstudio/training/Introduction to R/1-r-basics/1-basic-visualization/04-large-data - RStudio

Go to file/function

04-large-data

Untitled1*

Source on Save

Run

Source

```
1  library(ggplot2)
```

1:17 (Top Level) R Script

Console ~/Documents/rstudio/training/Introduction to R/1-r-basics/1-basic-

```
>
```

Workspace History Git

Import Dataset

Files Plots Packages Help

New Folder Delete Rename More

tion to R 1-r-basics 1-basic-visualization 04-large-data

	Name	Size	Modified
	..		
	04-large-data.html	4.2 MB	Feb 7, 2013, 9:37 AM
	overplot.png	936.4 KB	Feb 7, 2013, 9:37 AM
	transparent.png	863.2 KB	Feb 7, 2013, 9:37 AM
	small.png	463.8 KB	Feb 7, 2013, 9:37 AM
	caratcut.png	291.7 KB	Feb 7, 2013, 9:37 AM
	expensive.png	218.8 KB	Feb 7, 2013, 9:37 AM
	blue.png	124 KB	Feb 7, 2013, 9:37 AM
	caratbox.png	121.6 KB	Feb 7, 2013, 9:37 AM
	blue2.png	121.6 KB	Feb 7, 2013, 9:37 AM
	slides.md	6.6 KB	Feb 7, 2013, 9:37 AM
	04-large-		

~/Documents/rstudio/training/Introduction to R/1-r-basics/1-basic-visualization/04-large-data - RStudio

Go to file/function

04-large-data

Untitled1*

Source on Save | Run | Source

```
1  library(ggplot2)
```

[Cmd/ctrl + enter]

1:17 (Top Level) R Script

Workspace | History | Git

Import Dataset

Files | Plots | Packages | Help

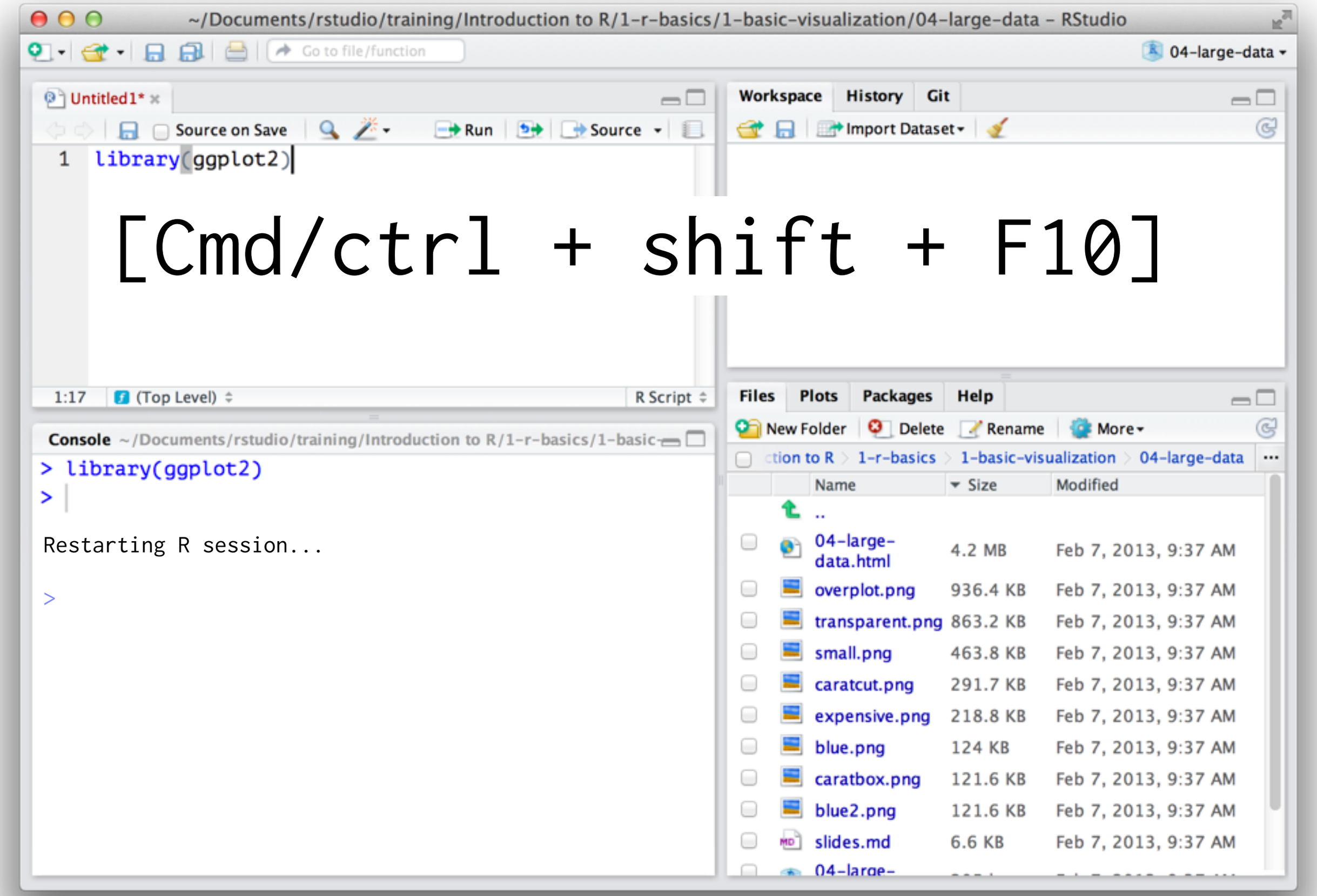
New Folder | Delete | Rename | More

tion to R › 1-r-basics › 1-basic-visualization › 04-large-data

Console ~/Documents/rstudio/training/Introduction to R/1-r-basics/1-basic-

```
> library(ggplot2)
>
```

Name	Size	Modified
..		
04-large-data.html	4.2 MB	Feb 7, 2013, 9:37 AM
overplot.png	936.4 KB	Feb 7, 2013, 9:37 AM
transparent.png	863.2 KB	Feb 7, 2013, 9:37 AM
small.png	463.8 KB	Feb 7, 2013, 9:37 AM
caratcut.png	291.7 KB	Feb 7, 2013, 9:37 AM
expensive.png	218.8 KB	Feb 7, 2013, 9:37 AM
blue.png	124 KB	Feb 7, 2013, 9:37 AM
caratbox.png	121.6 KB	Feb 7, 2013, 9:37 AM
blue2.png	121.6 KB	Feb 7, 2013, 9:37 AM
slides.md	6.6 KB	Feb 7, 2013, 9:37 AM
04-large-		


~/Documents/rstudio/training/Introduction to R/1-r-basics/1-basic-visualization/04-large-data - RStudio
Go to file/function
04-large-data
Untitled1*
Source on Save
Run
Source
1 library(ggplot2)
[Cmd/ctrl + shift + F10]
1:17
(Top Level)
R Script
Workspace
History
Git
Import Dataset
Files
Plots
Packages
Help
New Folder
Delete
Rename
More
1-r-basics
1-basic-visualization
04-large-data
Name
Size
Modified
Console ~/Documents/rstudio/training/Introduction to R/1-r-basics/1-basic-
> library(ggplot2)
>
Restarting R session...
>
04-large-data.html 4.2 MB Feb 7, 2013, 9:37 AM
overplot.png 936.4 KB Feb 7, 2013, 9:37 AM
transparent.png 863.2 KB Feb 7, 2013, 9:37 AM
small.png 463.8 KB Feb 7, 2013, 9:37 AM
caratcut.png 291.7 KB Feb 7, 2013, 9:37 AM
expensive.png 218.8 KB Feb 7, 2013, 9:37 AM
blue.png 124 KB Feb 7, 2013, 9:37 AM
caratbox.png 121.6 KB Feb 7, 2013, 9:37 AM
blue2.png 121.6 KB Feb 7, 2013, 9:37 AM
slides.md 6.6 KB Feb 7, 2013, 9:37 AM

Your turn

Skim https://twitter.com/rstudiotips – which one do you think is most useful?

This work is licensed under the Creative Commons Attribution-Noncommercial 3.0 United States License. To view a copy of this license, visit http://creativecommons.org/licenses/by-nc/3.0/us/ or send a letter to Creative Commons, 171 Second Street, Suite 300, San Francisco, California, 94105, USA.