

These materials adapted by Amelia McNamara from
the RStudio [CC BY-SA](#) materials Introduction to R
(2014) and [Master the Tidyverse](#) (2017).

Introduction to R & RStudio: deck 01: Getting started

Amelia McNamara

Visiting Assistant Professor of Statistical and Data Sciences
Smith College

January 2018

HELLO

my name is

Amelia

@AmeliaMN

HELLO

my name is

Katie

Katie Leap

HELLO

my name is

Kelly

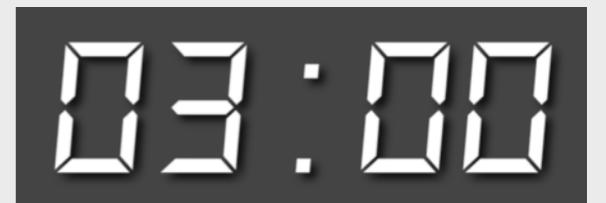
Kelly O'Briant

@b23kelly **@RLadiesDC**

Your turn

Introduce yourself to your nearest 2-4 neighbors. Tell them:

- Who you are
- What you do with data



Schedule

9:00 - 10:30

10:30 - 11:00

Morning break

11:00 - 12:30

12:30 - 1:30

Lunch

1:30 - 3:00

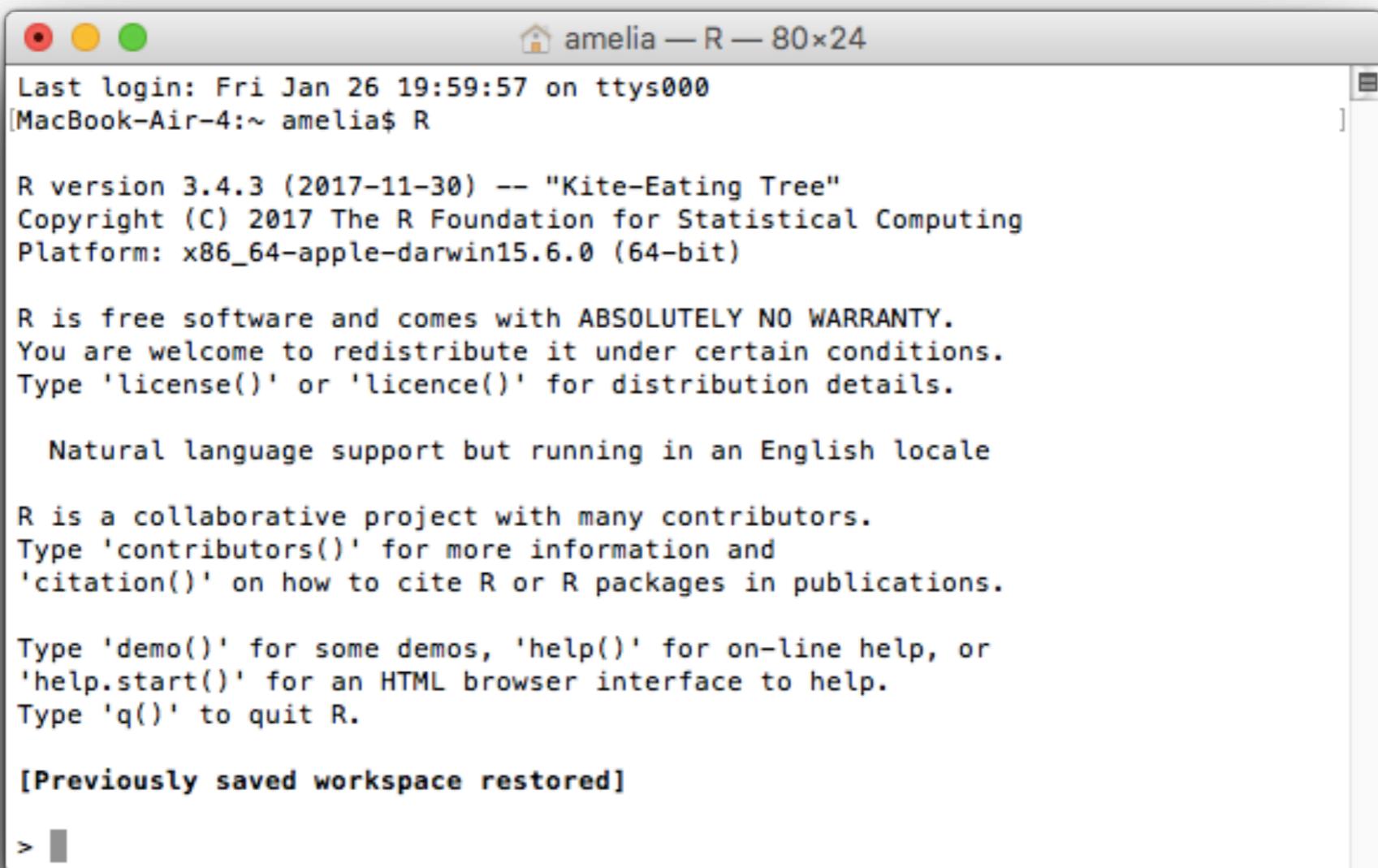
3:00 - 3:30

Afternoon break

3:30 - 5:00



R: a computer programming language



```
Last login: Fri Jan 26 19:59:57 on ttys000
[MacBook-Air-4:~ amelia$ R

R version 3.4.3 (2017-11-30) -- "Kite-Eating Tree"
Copyright (C) 2017 The R Foundation for Statistical Computing
Platform: x86_64-apple-darwin15.6.0 (64-bit)

R is free software and comes with ABSOLUTELY NO WARRANTY.
You are welcome to redistribute it under certain conditions.
Type 'license()' or 'licence()' for distribution details.

Natural language support but running in an English locale

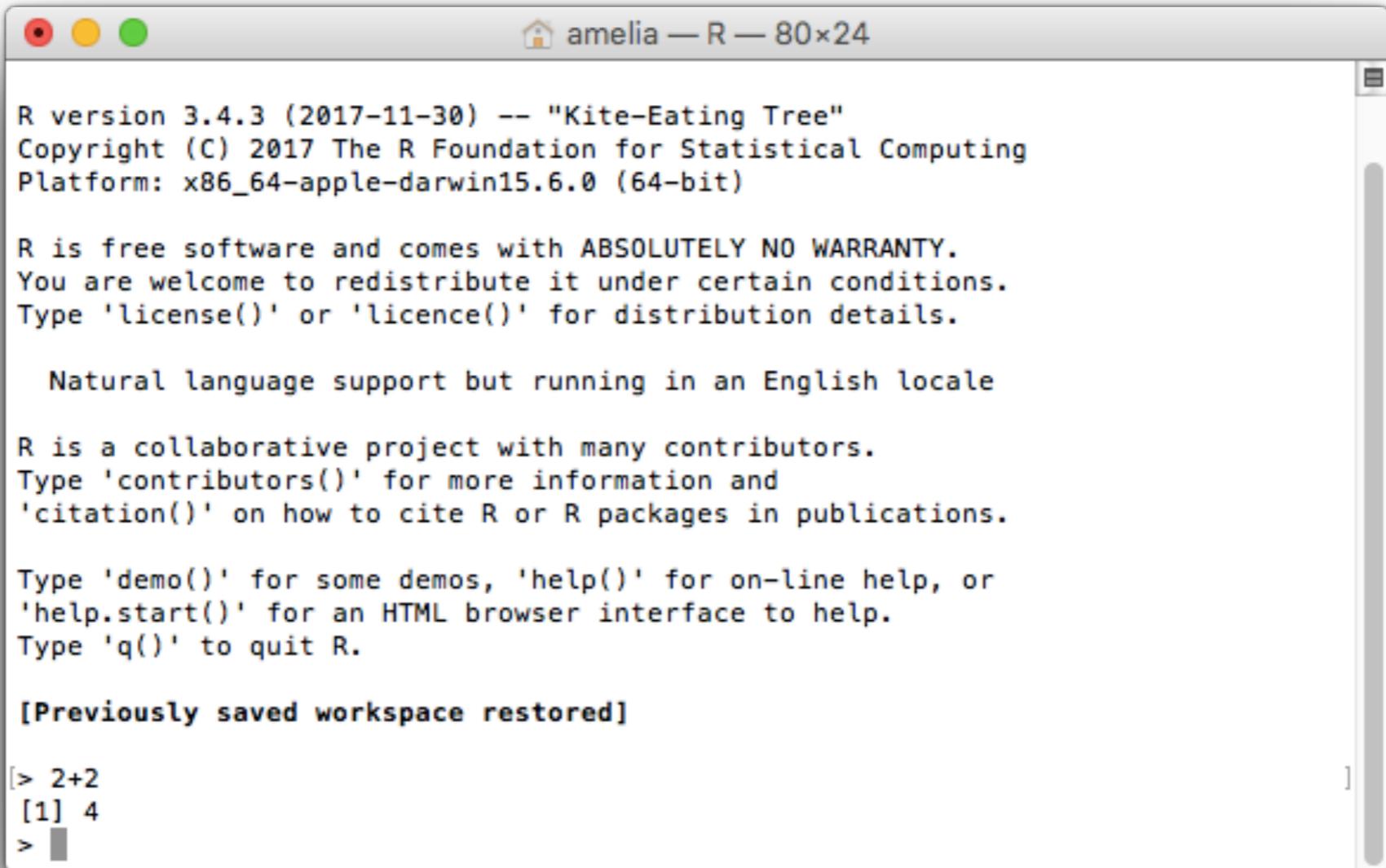
R is a collaborative project with many contributors.
Type 'contributors()' for more information and
'citation()' on how to cite R or R packages in publications.

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.

[Previously saved workspace restored]

> ]
```

R: a computer programming language



```
R version 3.4.3 (2017-11-30) -- "Kite-Eating Tree"
Copyright (C) 2017 The R Foundation for Statistical Computing
Platform: x86_64-apple-darwin15.6.0 (64-bit)

R is free software and comes with ABSOLUTELY NO WARRANTY.
You are welcome to redistribute it under certain conditions.
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.

[Previously saved workspace restored]

[> 2+2
[1] 4
> ]
```

R: a computer programming language

1. Descends from S, Bell Labs
2. Evolved in university environment
3. Full language
4. ...but can be used as a simple application
5. Designed for use with data

```
> bechdel
```

```
# A tibble: 1,794 x 15
```

	year	imdb	title	test	clean_test	binary	budget	domgross	intgross
	<int>	<chr>	<chr>	<chr>	<fctr>	<chr>	<int>	<dbl>	<dbl>
1	2009	tt1003034	Perrier's Bounty	nowomen	nowomen	FAIL	6600000	828	828
2	2008	tt1226681	Pontypool	nowomen-disagree	nowomen	FAIL	1500000	3865	31916
3	2012	tt1874789	Supporting Characters	men	men	FAIL	60000	4917	4917
4	2007	tt0861739	Tropa de Elite	ok-disagree	ok	PASS	6537890	8744	14319195
5	2007	tt0964587	St. Trinian's	ok	ok	PASS	11400000	15000	22446568
6	2011	tt1535616	The Divide	ok	ok	PASS	3000000	18000	18000
7	1996	tt0115591	August	dubious	dubious	FAIL	3400000	12636	12636
8	2006	tt0783238	The Dead Girl	ok	ok	PASS	3300000	19875	19875
9	2005	tt0342272	Dear Wendy	notalk	notalk	FAIL	8000000	23106	446438
10	2011	tt1788391	Kill List	dubious	dubious	FAIL	800000	29063	462206

```
# ... with 1,784 more rows, and 6 more variables: code <chr>, budget_2013 <int>, domgross_2013 <dbl>,
```

```
# intgross_2013 <dbl>, period_code <int>, decade_code <int>
```

```
> bechdel %>% skim(domgross_2013)
```

Skim summary statistics

n obs: 1794

n variables: 15

Variable type: numeric

	variable	missing	complete	n	mean	sd	p25	median	p75	hist
domgross_2013		18	1776	1794	9.5e+07	1.3e+08	2.1e+07	5.6e+07	1.2e+08	█_____

```
> bechdel %>% skim(clean_test)
```

Skim summary statistics

n obs: 1794

n variables: 15

Variable type: factor

	variable	missing	complete	n	n_unique	top_counts	ordered
clean_test		0	1794	1794	5	ok: 803, not: 514, men: 194, dub: 142	FALSE

```
> qplot(budget_2013, domgross_2013, data=bechdel, color = binary)
```



```
> lm(domgross_2013~budget_2013, data=bechdel)
```

Call:

```
lm(formula = domgross_2013 ~ budget_2013, data = bechdel)
```

Residuals:

Min	1Q	Median	3Q	Max
-256686756	-47529500	-27186696	15143559	1690886212

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	3.615e+07	3.782e+06	9.559	<2e-16 ***
budget_2013	1.056e+00	4.823e-02	21.896	<2e-16 ***

Signif. codes: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 11180000 on 1774 degrees of freedom
(18 observations deleted due to missingness)

Multiple R-squared: 0.2128, Adjusted R-squared: 0.2123
F-statistic: 479.4 on 1 and 1774 DF, p-value: < 2.2e-16

Movie explorer

Filter

Minimum number of reviews on Rotten Tomatoes

 10 40 70 100 130 160 190 220 250 280 300

Year released

 1,940 1,948 1,956 1,964 1,972 1,980 1,988 1,996 2,004 2,012 2,014

Minimum number of Oscar wins (all categories)

 0 1 2 3 4

Dollars at Box Office (millions)

 0 80 160 320 400 480 560 640 720 800

Genre (a movie can have multiple genres)

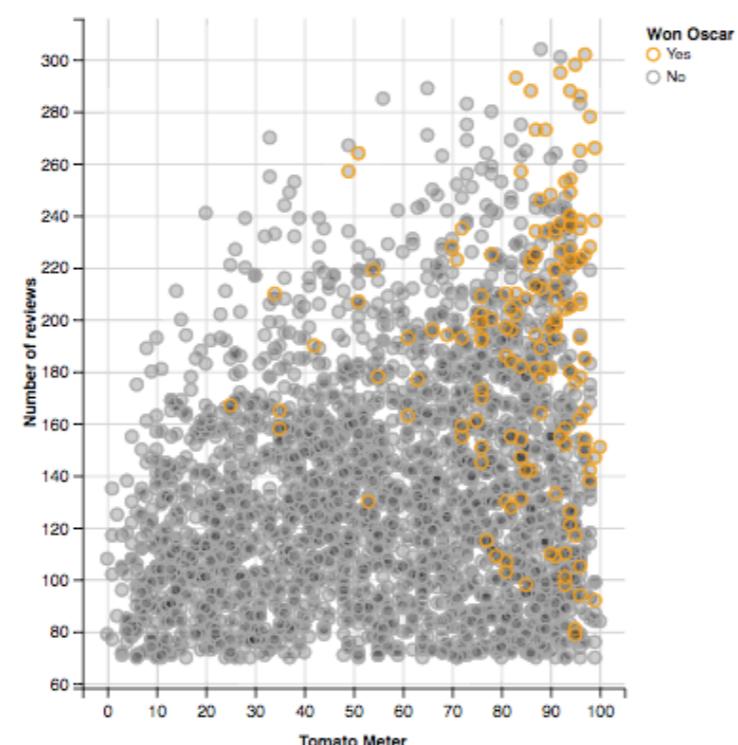
Director name contains (e.g., Miyazaki)

Cast names contains (e.g. Tom Hanks)

X-axis variable

Y-axis variable

Note: The Tomato Meter is the proportion of positive reviews (as judged by the Rotten Tomatoes staff), and the Numeric rating is a normalized 1-10 score of those reviews which have star ratings (for example, 3 out of 4 stars).



Number of movies selected:
2758

intRo X Amelia

www.intro-stats.com

intRo Home ? Github Email </> Download Print

Data Sources Transform Summaries Graphical Numerical Inference Contingency Regression T test

Choose Dataset: MPG

Show 10 entries

Categorical Variable Numeric Variable

manufacturer	model	displ	year	cyl	trans	drv	cty	hwy	fl
audi	a4	1.8	1999	4	auto(l5)	f	18	29	p
audi	a4	1.8	1999	4	manual(m5)	f	21	29	p
audi	a4	2.0	2008	4	manual(m6)	f	20	31	p
audi	a4	2.0	2008	4	auto(av)	f	21	30	p
audi	a4	2.8	1999	6	auto(l5)	f	16	26	p
audi	a4	2.8	1999	6	manual(m5)	f	18	26	p
audi	a4	3.1	2008	6	auto(av)	f	18	27	p
audi	a4 quattro	1.8	1999	4	manual(m5)	4	18	26	p
audi	a4 quattro	1.8	1999	4	auto(l5)	4	16	25	p
audi	a4 quattro	2.0	2008	4	manual(m6)	4	20	28	p

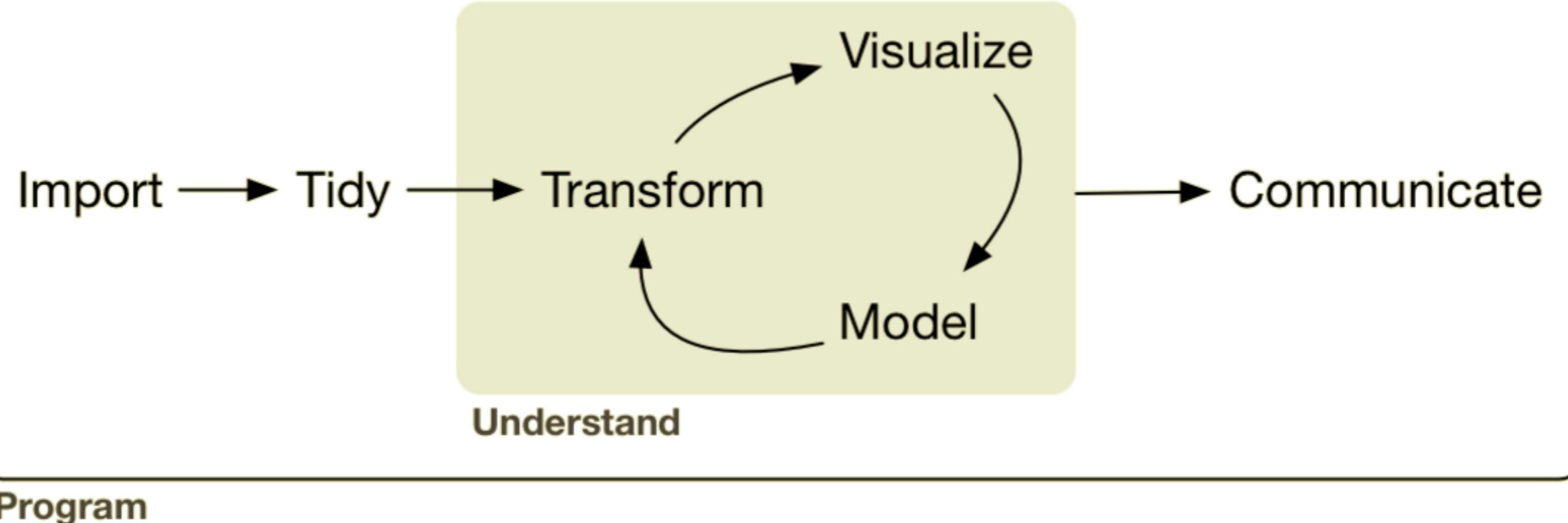
manufacturer model displ year cyl trans drv cty hwy fl

Showing 1 to 10 of 234 entries

Previous 1 2 3 4 5 ...

24 Next

<http://www.intro-stats.com/>



From *R for Data Science* by Hadley Wickham and Garrett Grolemund.



RStudio: a software program

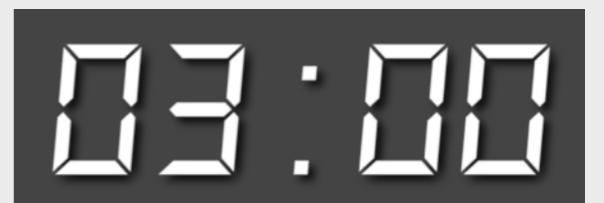
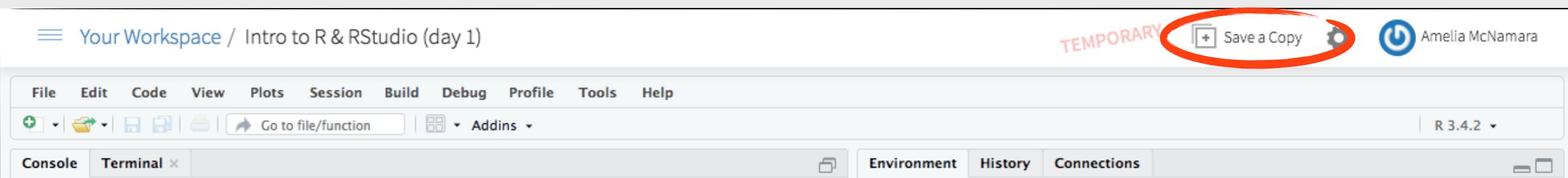
1. like Microsoft Word, Excel, etc.
2. built to help you write R code, run R code, and analyze data with R
3. text editor, version control, keyboard shortcuts, debugging tools, and much more

Your turn

It's time to log in to RStudio, if you haven't already. Go to http://bit.ly/R_and_RStudio_day1

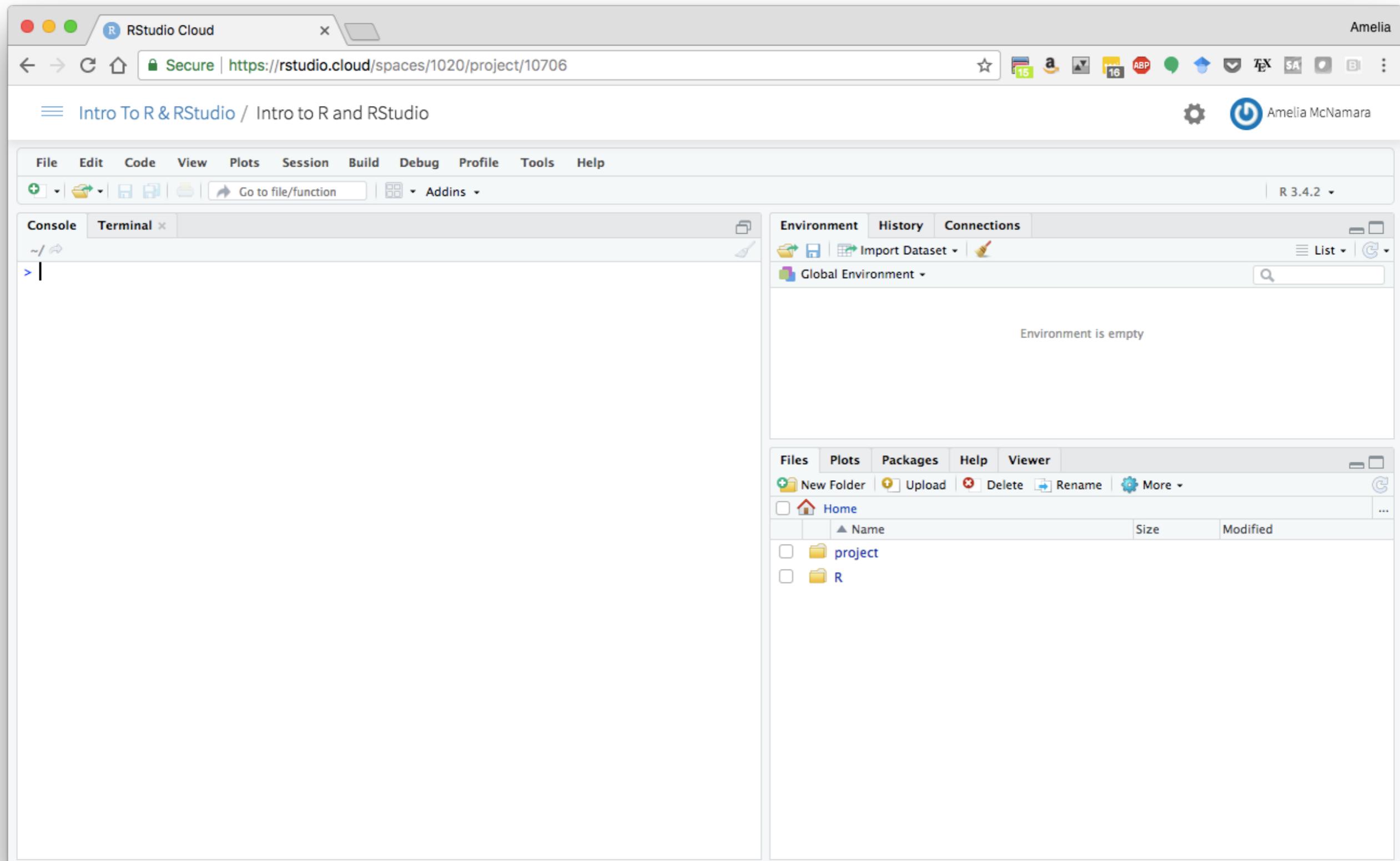
Make an account

Click Save a Copy



RStudio

http://bit.ly/R_and_RStudio_day1



RStudio

http://bit.ly/R_and_RStudio_day1

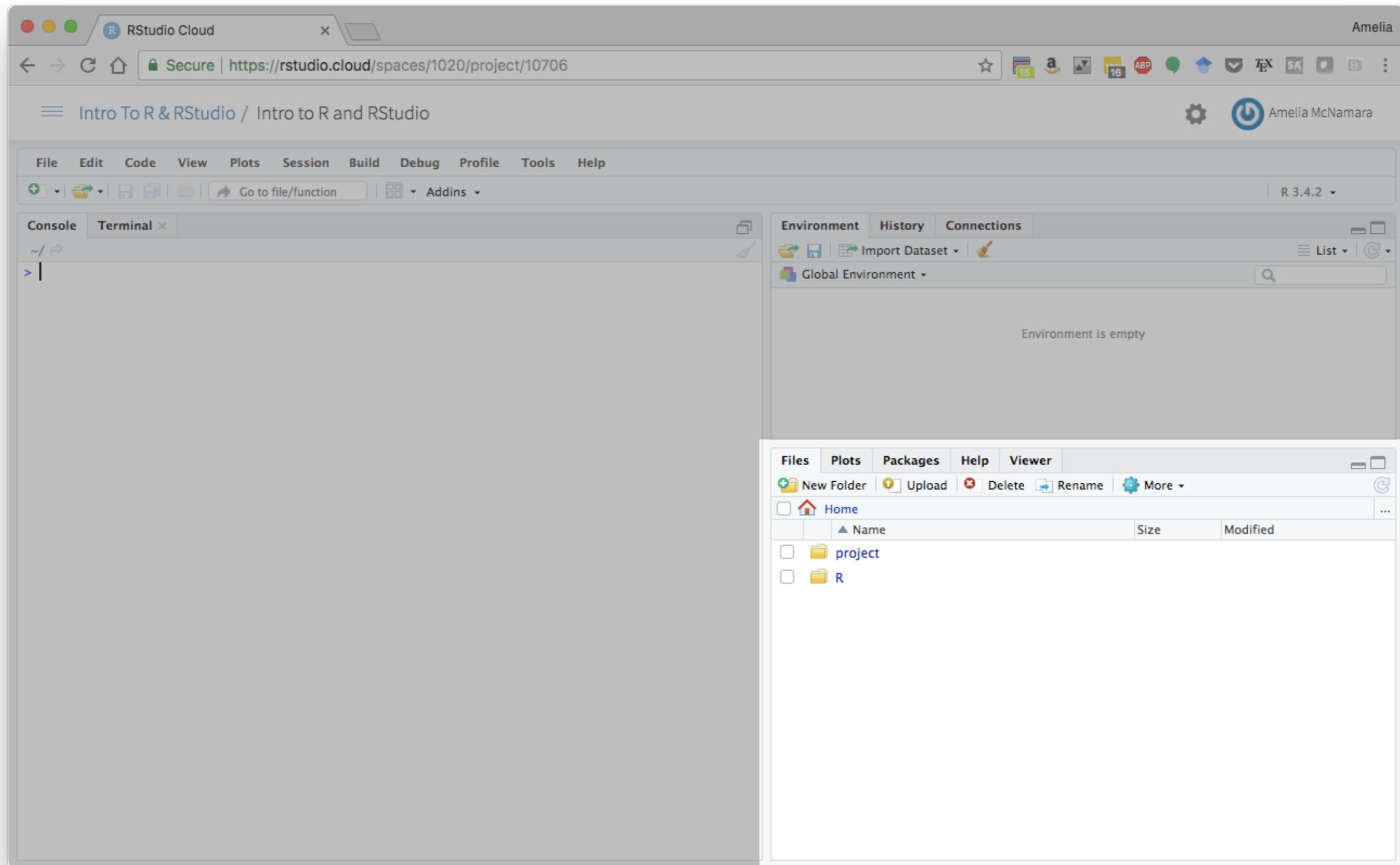
The screenshot shows the RStudio Cloud interface. The top navigation bar includes 'RStudio Cloud' and the user 'Amelia'. The address bar shows a secure connection to <https://rstudio.cloud/spaces/1020/project/10706>. The main menu bar has options like File, Edit, Code, View, Plots, Session, Build, Debug, Profile, Tools, and Help. Below the menu is a toolbar with various icons. The interface is divided into several panes:

- Console**: Shows a command line input field starting with '> |'.
- Environment**: Shows the Global Environment, which is currently empty.
- Files**: Shows a file tree with 'Home' selected, containing 'project' and 'R' folders.
- Plots**: Not visible in the screenshot.
- Packages**: Not visible in the screenshot.
- Help**: Not visible in the screenshot.
- Viewer**: Not visible in the screenshot.

The console gives you a place to execute commands written in R

RStudio

http://bit.ly/R_and_RStudio_day1



RStudio

http://bit.ly/R_and_RStudio_day1

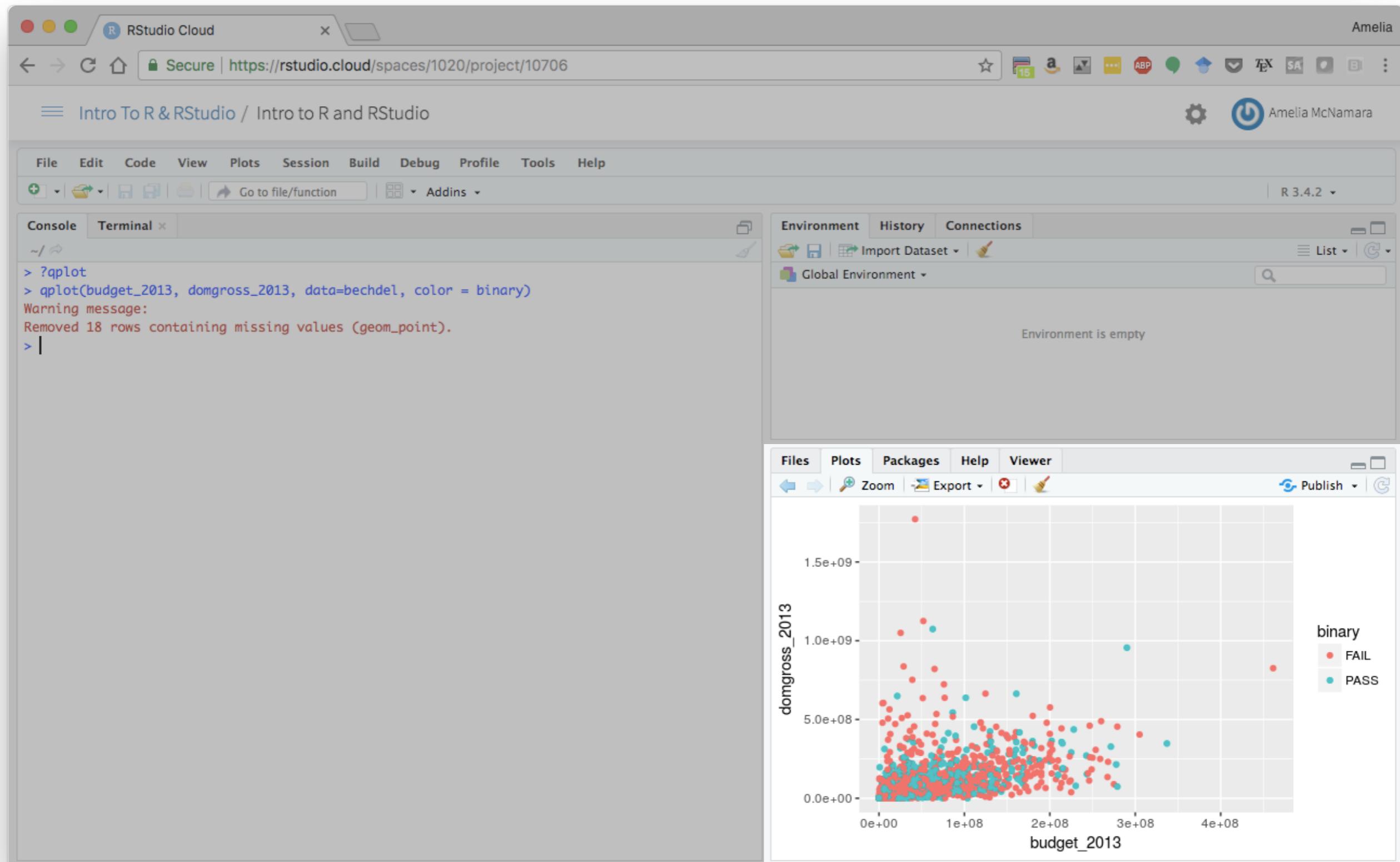
The screenshot shows the RStudio Cloud interface. The top navigation bar includes the RStudio Cloud logo, a user profile for "Amelia", and a secure connection indicator. The main menu bar has options like File, Edit, Code, View, Plots, Session, Build, Debug, Profile, Tools, and Help. Below the menu is a toolbar with various icons for file operations. The left pane contains a "Console" tab with the command `> ?ggplot` and a "Terminal" tab. The right pane has tabs for "Environment", "History", and "Connections". The "Environment" tab shows the message "Environment is empty". The bottom right pane is a "Documentation" viewer for the `ggplot` function from the `ggplot2` package. It displays the function's usage, description, and arguments. The usage is shown as:

```
ggplot(data = NULL, mapping = aes(), ..., environment = parent.frame()
```

The description and arguments sections are also visible.

RStudio

http://bit.ly/R_and_RStudio_day1



RStudio

http://bit.ly/R_and_RStudio_day1

The screenshot shows the RStudio Cloud interface with the following components:

- Header:** RStudio Cloud, Secure connection to https://rstudio.cloud/spaces/1020/project/10706, User: Amelia.
- Toolbar:** File, Edit, Code, View, Plots, Session, Build, Debug, Profile, Tools, Help.
- Console:** Shows R code and its output:

```
> ?qplot
> qplot(budget_2013, domgross_2013, data=bechdel, color = binary)
Warning message:
Removed 18 rows containing missing values (geom_point).
> |
```
- Environment:** Displays the R session environment with objects like `bechdel`, `domgross_2013`, and `skimr`.
- Plots:** A scatter plot titled "domgross_2013" vs "budget_2013". The x-axis ranges from 0e+00 to 4e+08, and the y-axis ranges from 0.0e+00 to 1.5e+09. The plot shows data points colored by the `binary` factor, with red dots representing "FAIL" and teal dots representing "PASS".

RStudio

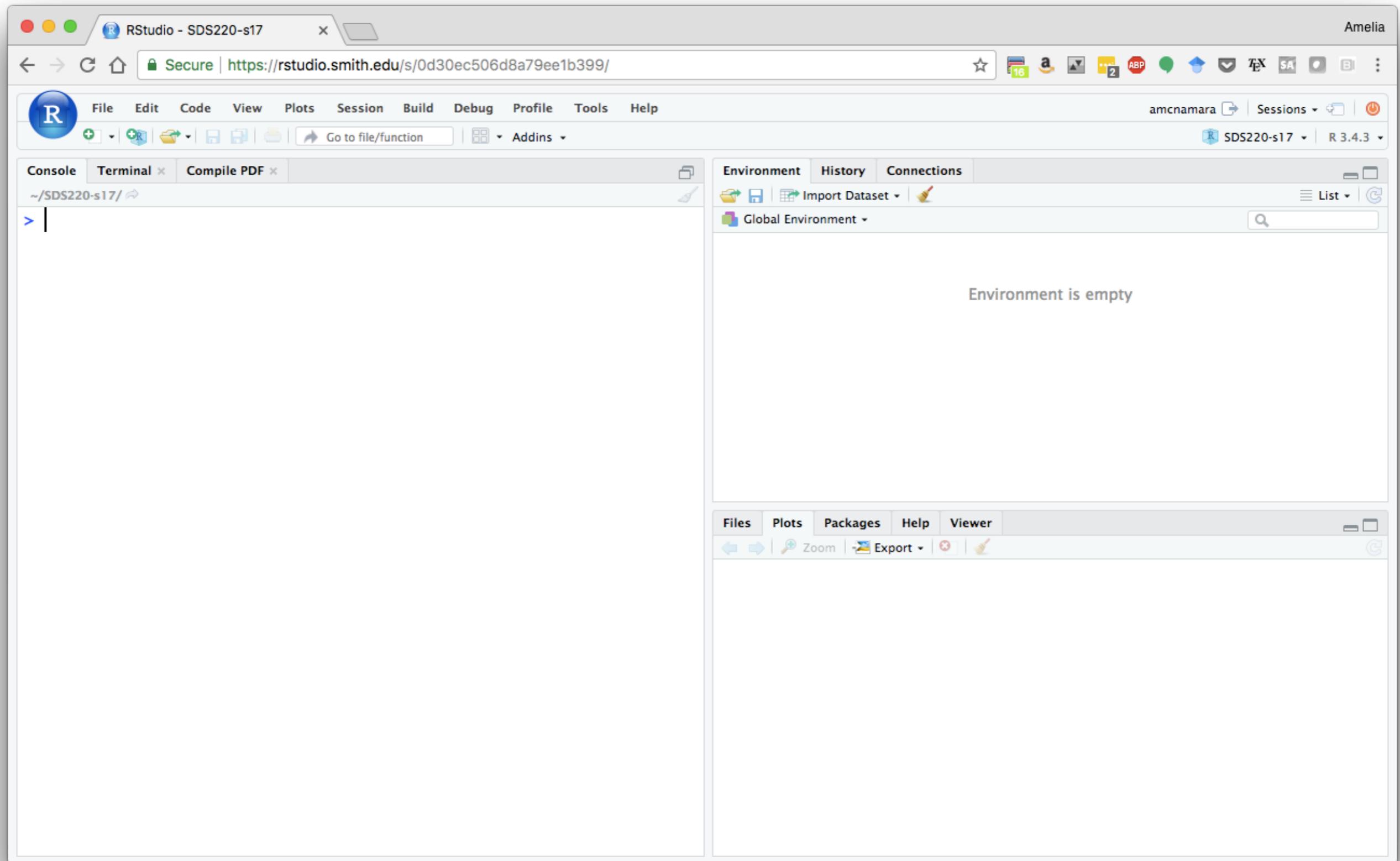
http://bit.ly/R_and_RStudio_day1

The screenshot shows the RStudio Cloud interface with the following components:

- Header:** RStudio Cloud, Secure connection to https://rstudio.cloud/spaces/1020/project/10706, User: Amelia.
- Toolbar:** File, Edit, Code, View, Plots, Session, Build, Debug, Profile, Tools, Help.
- Code Editor:** Untitled1 (R Markdown). The code includes setup for knitr, a brief explanation of R Markdown, and a note about the Knit button.
- Environment:** Global Environment (empty).
- Plots:** A scatter plot titled "domgross_2013" vs "budget_2013". The x-axis ranges from 0e+00 to 4e+08, and the y-axis ranges from 0.0e+00 to 1.5e+09. Data points are colored by "binary" status: FAIL (red) and PASS (teal).
- Console:** Shows R code execution for qplot and qplot(budget_2013, domgross_2013, data=bechdel, color = binary), resulting in a warning message about removed rows.

RStudio

http://bit.ly/R_and_RStudio_day1

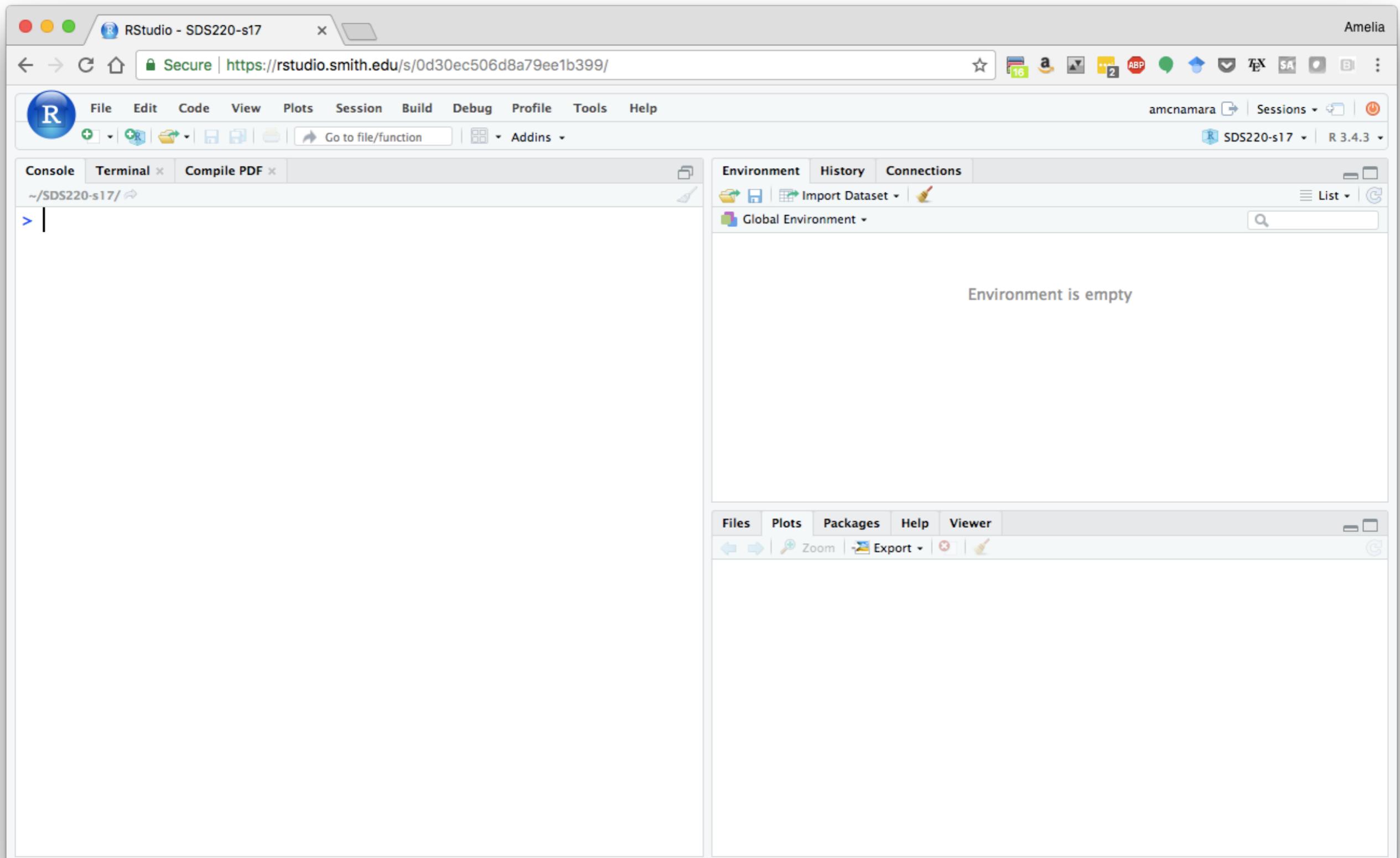


RStudio: ways to use

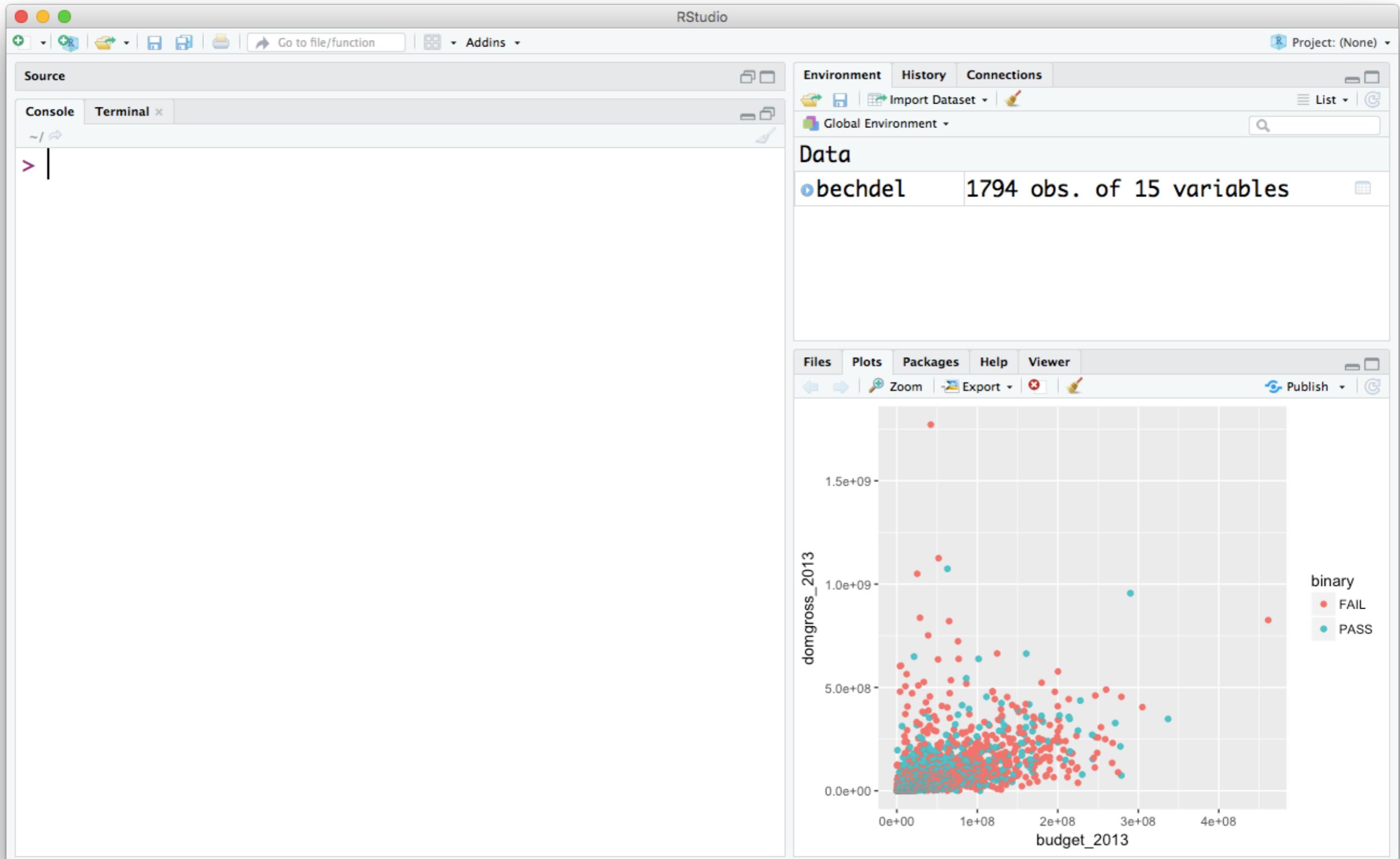
We're using RStudio Cloud, which allows you to log in through a web browser and do your work there.

But, there are other versions of RStudio.

RStudio: server edition



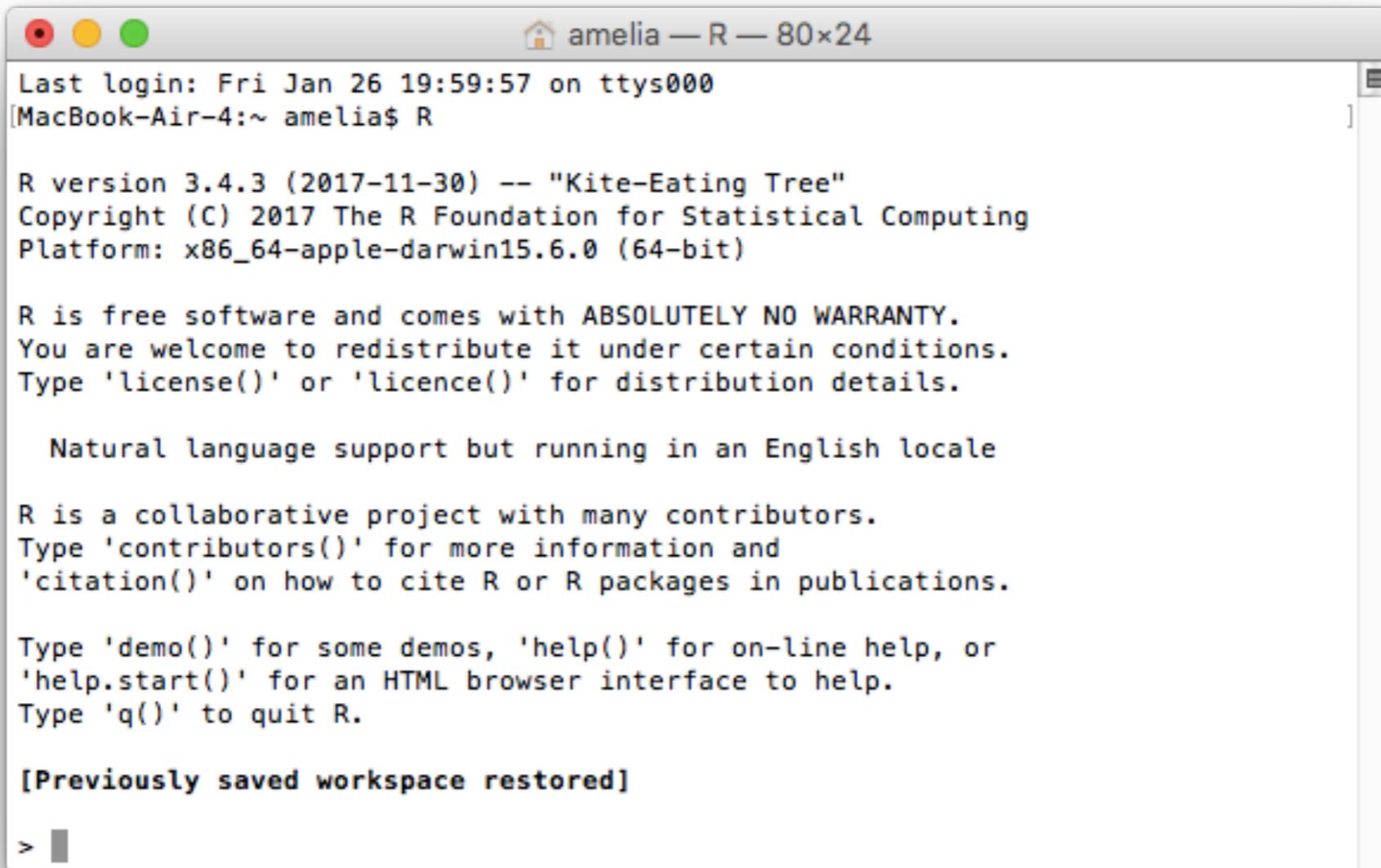
RStudio: desktop edition



Transferability

Everything you learn here will work in any version of RStudio

It will even work in the basic console version of R



```
Last login: Fri Jan 26 19:59:57 on ttys000
[MacBook-Air-4:~ amelia$ R

R version 3.4.3 (2017-11-30) -- "Kite-Eating Tree"
Copyright (C) 2017 The R Foundation for Statistical Computing
Platform: x86_64-apple-darwin15.6.0 (64-bit)

R is free software and comes with ABSOLUTELY NO WARRANTY.
You are welcome to redistribute it under certain conditions.
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.

[Previously saved workspace restored]
> 
```

Tips

Ask questions!

Practice!

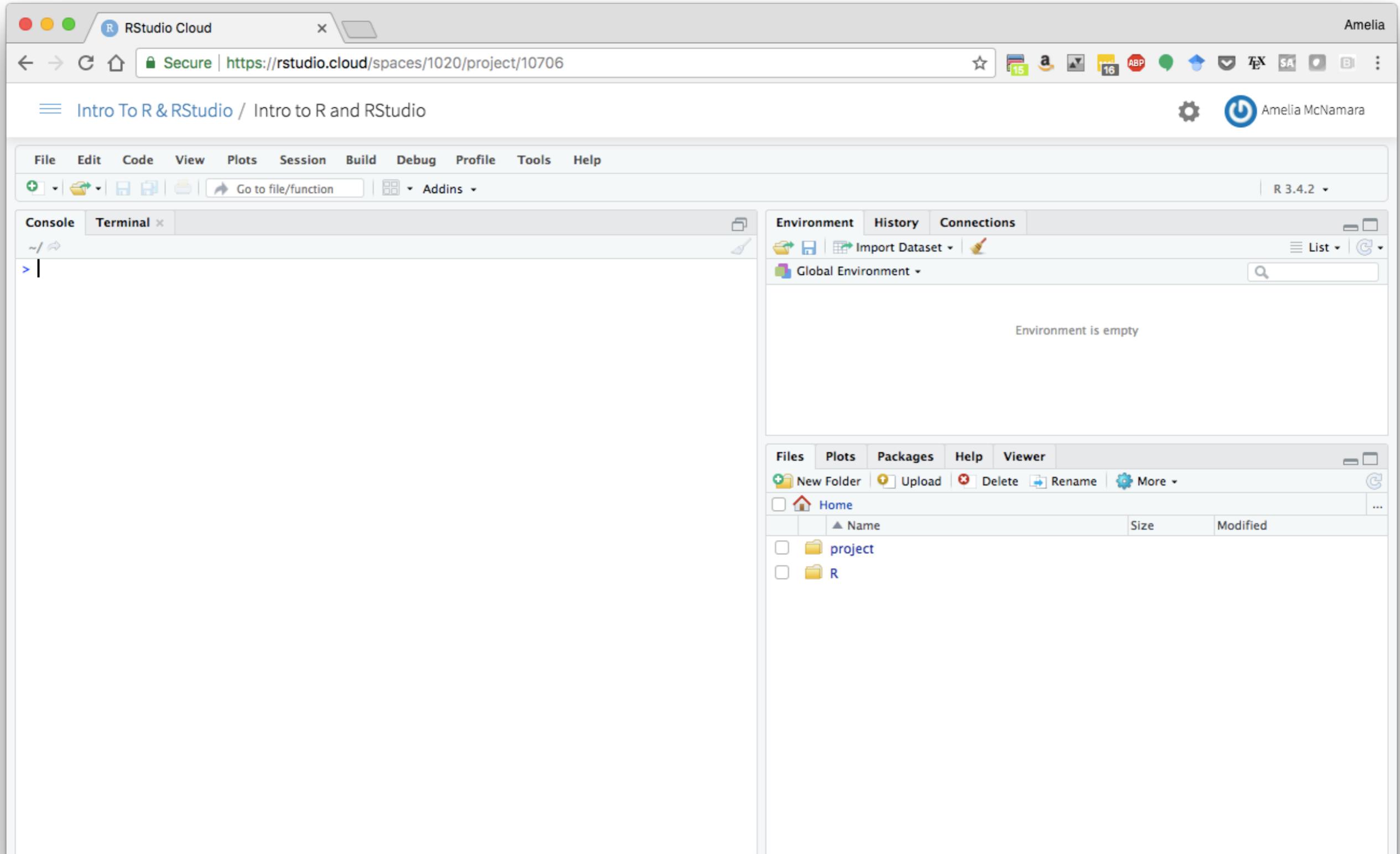
Seriously: practice!

And practice consciously: make a prediction,
then test it, then reflect.

RStudio

Getting started

http://bit.ly/R_and_RStudio_day1



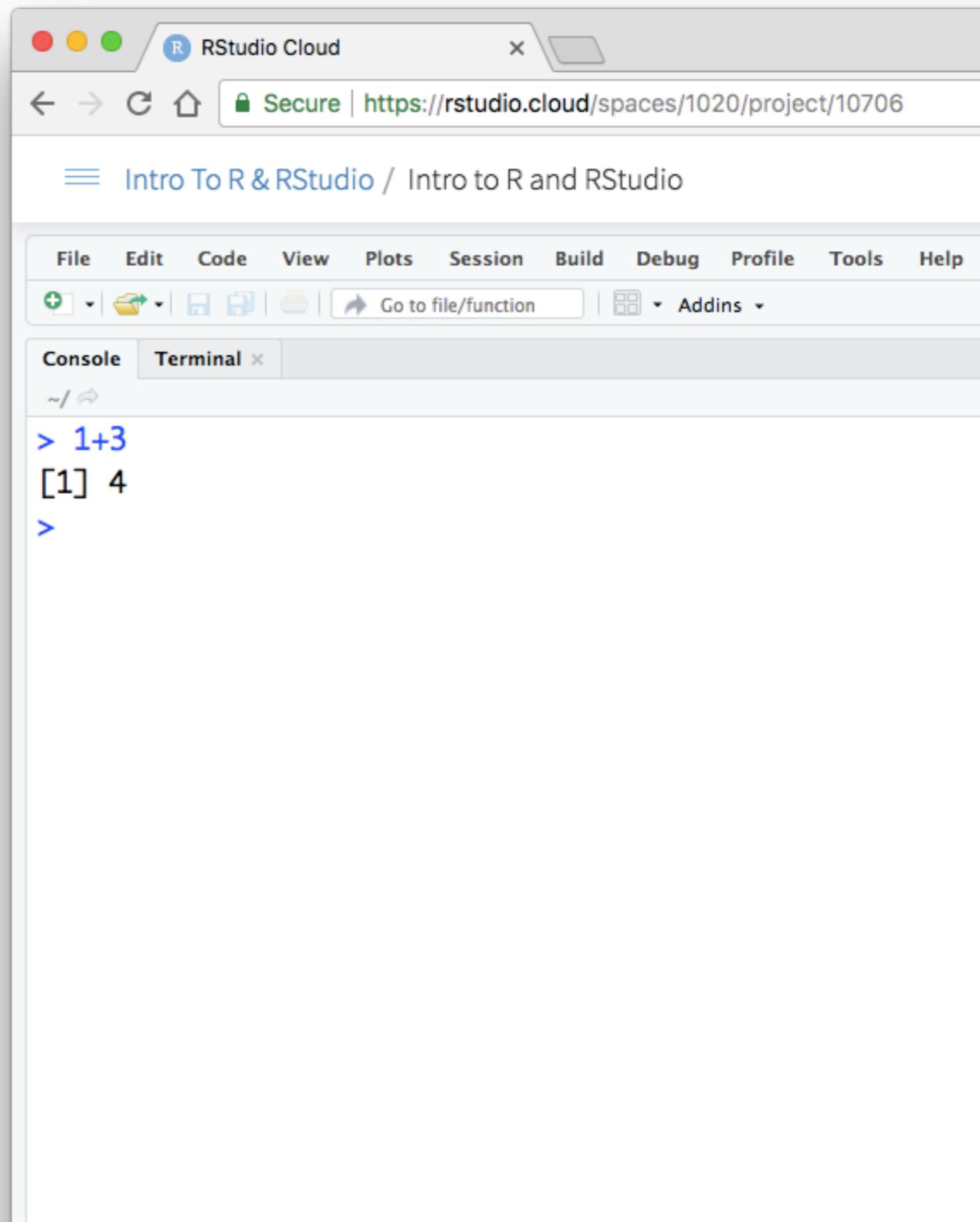
The console gives you a place to execute commands written in R

Type commands on the line that begins with a > sign (known as the prompt)

Output

Output →
New prompt →

When you hit enter,
RStudio will run
your command and
display any output
below it



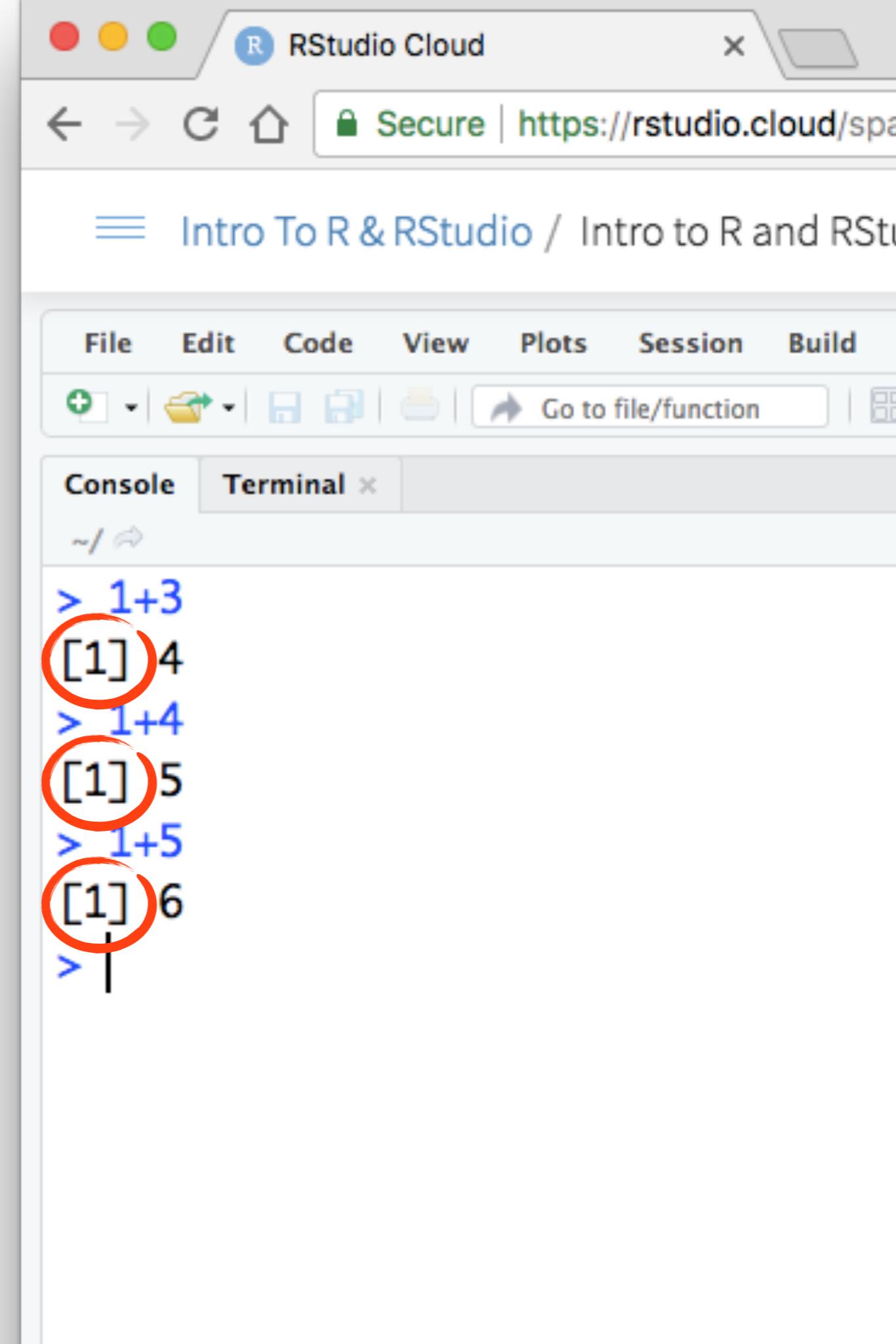
A screenshot of the RStudio Cloud interface. The title bar says "RStudio Cloud". The address bar shows a secure connection at "https://rstudio.cloud/spaces/1020/project/10706". The main window title is "Intro To R & RStudio / Intro to R and RStudio". The menu bar includes File, Edit, Code, View, Plots, Session, Build, Debug, Profile, Tools, and Help. Below the menu is a toolbar with various icons. The main area has tabs for "Console" and "Terminal", with "Console" selected. The console output shows the following text:
> 1+3
[1] 4
>

[1]

R displays an index next to the output.

Just ignore this.

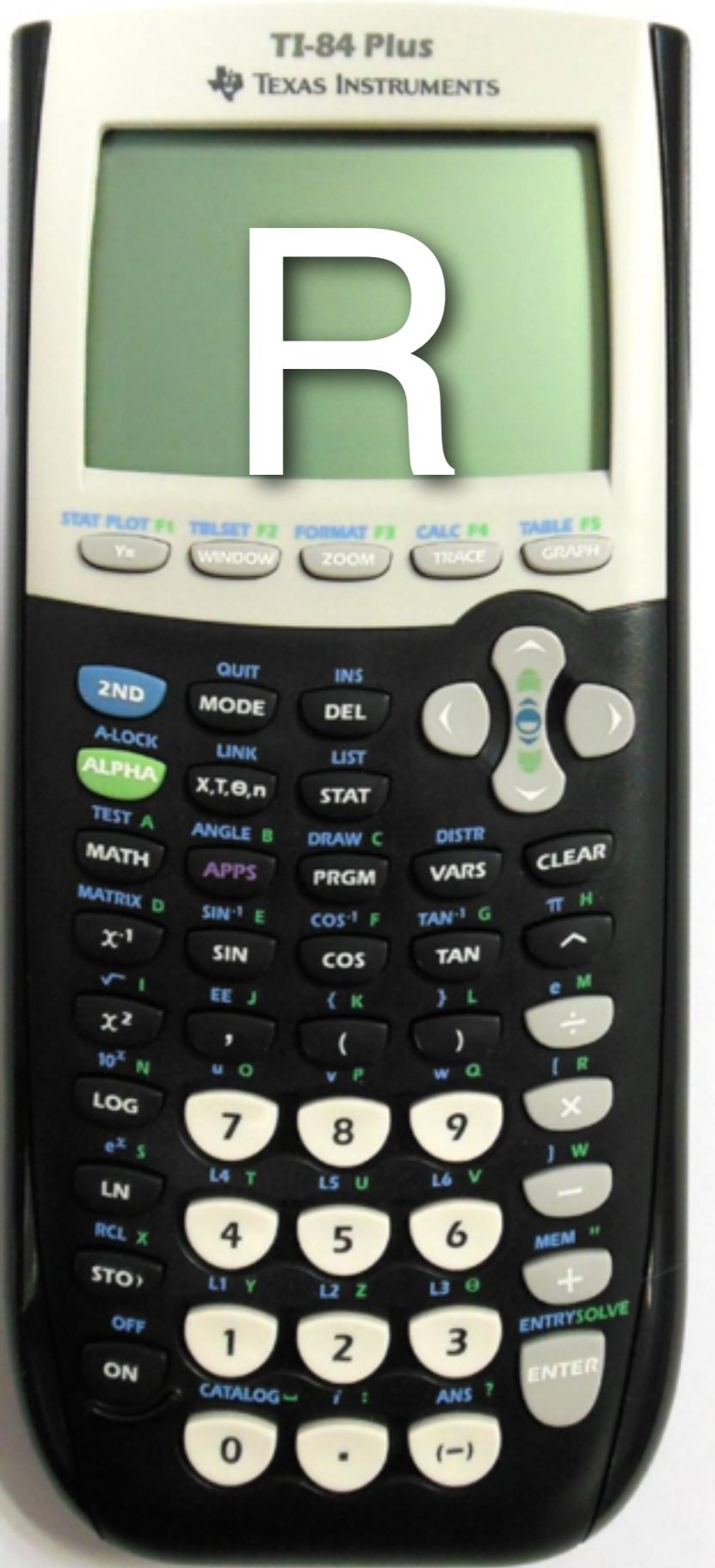
Somewhat helpful when R returns more than one value in the output.



The screenshot shows the RStudio Cloud interface with a browser header and a menu bar. The main area is a console window titled "Console". It contains the following R session history:

```
> 1+3  
[1] 4  
> 1+4  
[1] 5  
> 1+5  
[1] 6  
>
```

The output lines for each addition operation (i.e., the lines containing "[1]" followed by a value) are circled in red.



R is like a fancy calculator
on your computer

$$5 + 5$$

10

$$4 - 1$$

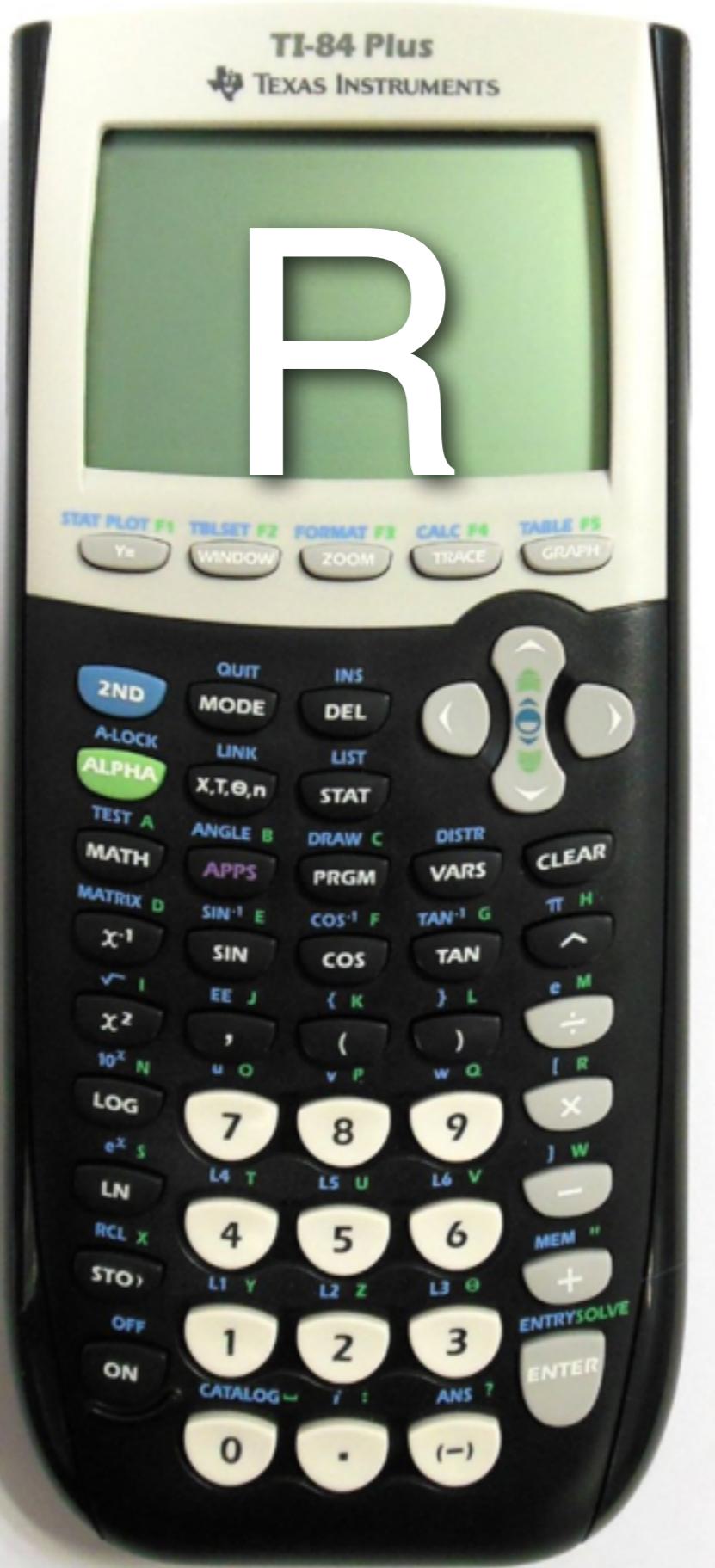
3

$$1 * 2$$

2

$$4 ^ 2$$

16



It can do algebra

$$a <- 1$$

$$b <- 2$$

$$a + b$$

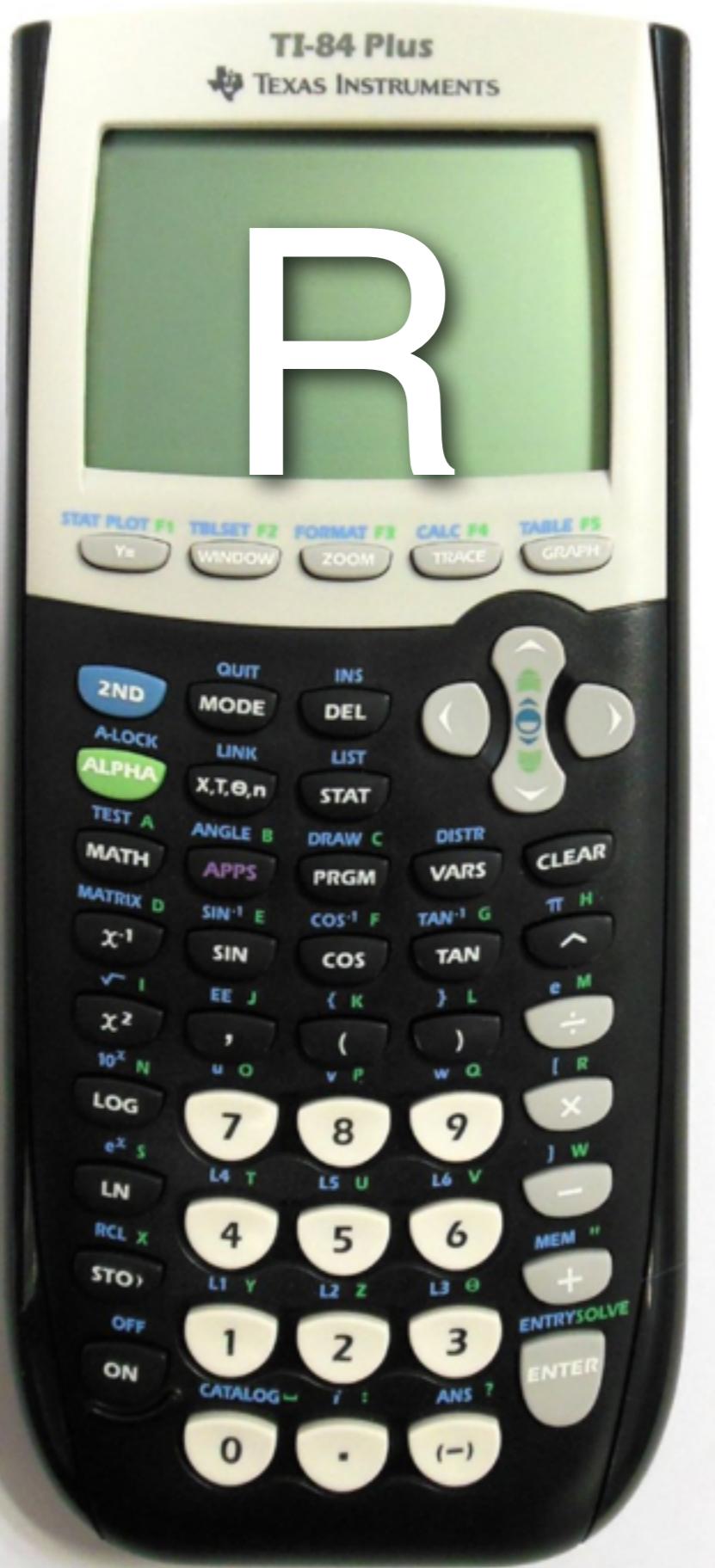
3

$$A <- 3$$

It cares about capitalization

$$a + b - A$$

0



And it has functions that let you do more sophisticated manipulations

round(3.1415)

3

factorial(3)

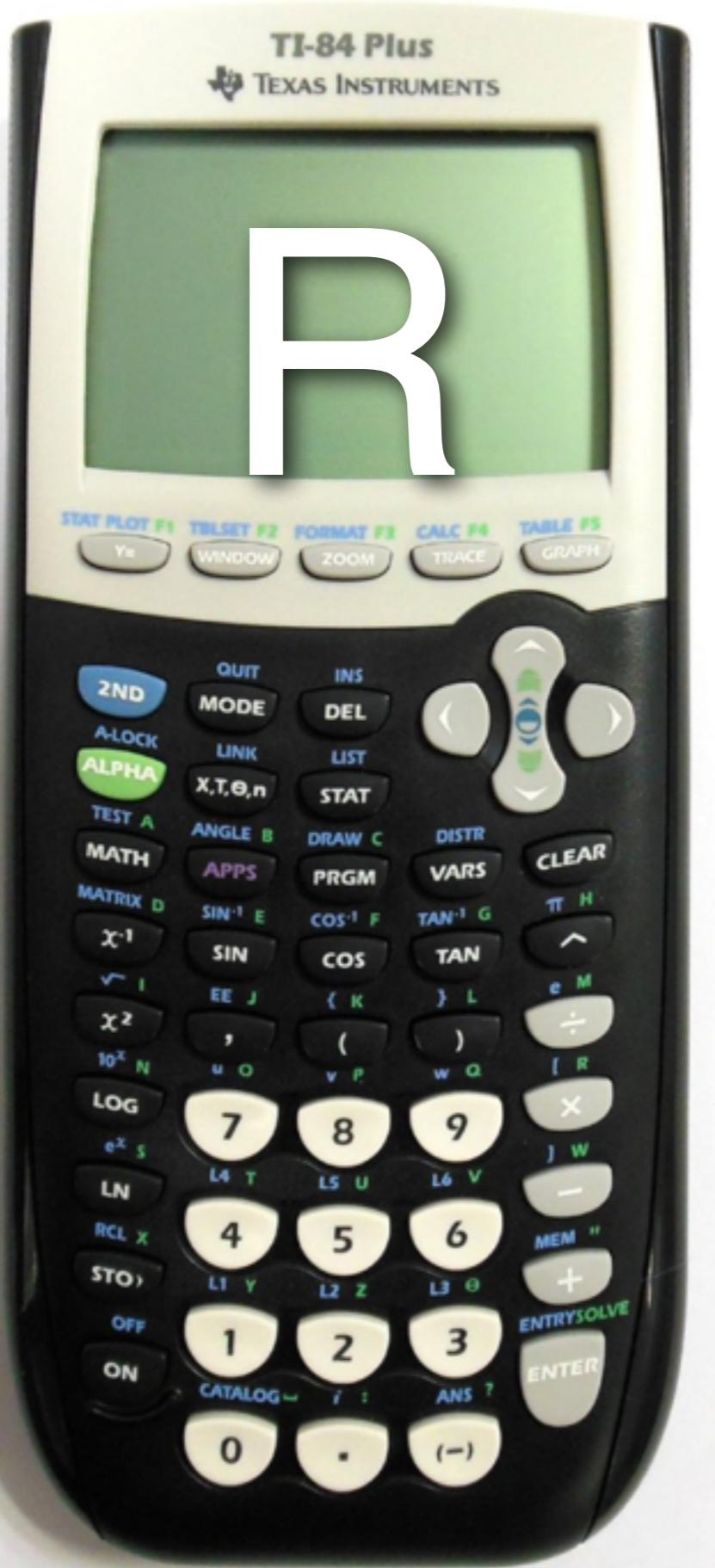
$$3! = 3 \times 2 \times 1$$

6

sqrt(9)

square root

3



Most of the cool stuff in R comes from functions. Like $f(x)$ ("f of x") functions in R have names, parentheses, and arguments

factorial(3)
6

factorial of 3

sqrt(9)
3

square root of 9

Your turn

What do you think this will return?

Your turn

What do you think this will return?

`factorial(round(2.0015) + 1)`

00 : 30

R always works from the innermost parenthesis to the outermost (just like a calculator).

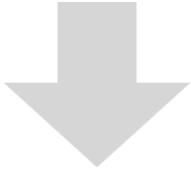
```
factorial(round(2.0015) + 1)
```



```
factorial(2 + 1)
```



```
factorial(3)
```



6

R always works from the innermost parenthesis to the outermost (just like a calculator).

```
factorial(round(2.0015) + 1)
```



```
factorial(2 + 1)
```



```
factorial(3)
```



6

R always works from the innermost parenthesis to the outermost (just like a calculator).

`factorial(round(2.0015) + 1)`



`factorial(2 + 1)`



`factorial(3)`



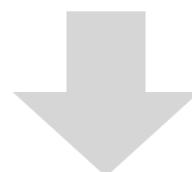
6

R always works from the innermost parenthesis to the outermost (just like a calculator).

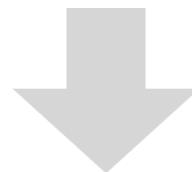
`factorial(round(2.0015) + 1)`



`factorial(2 + 1)`



`factorial(3)`



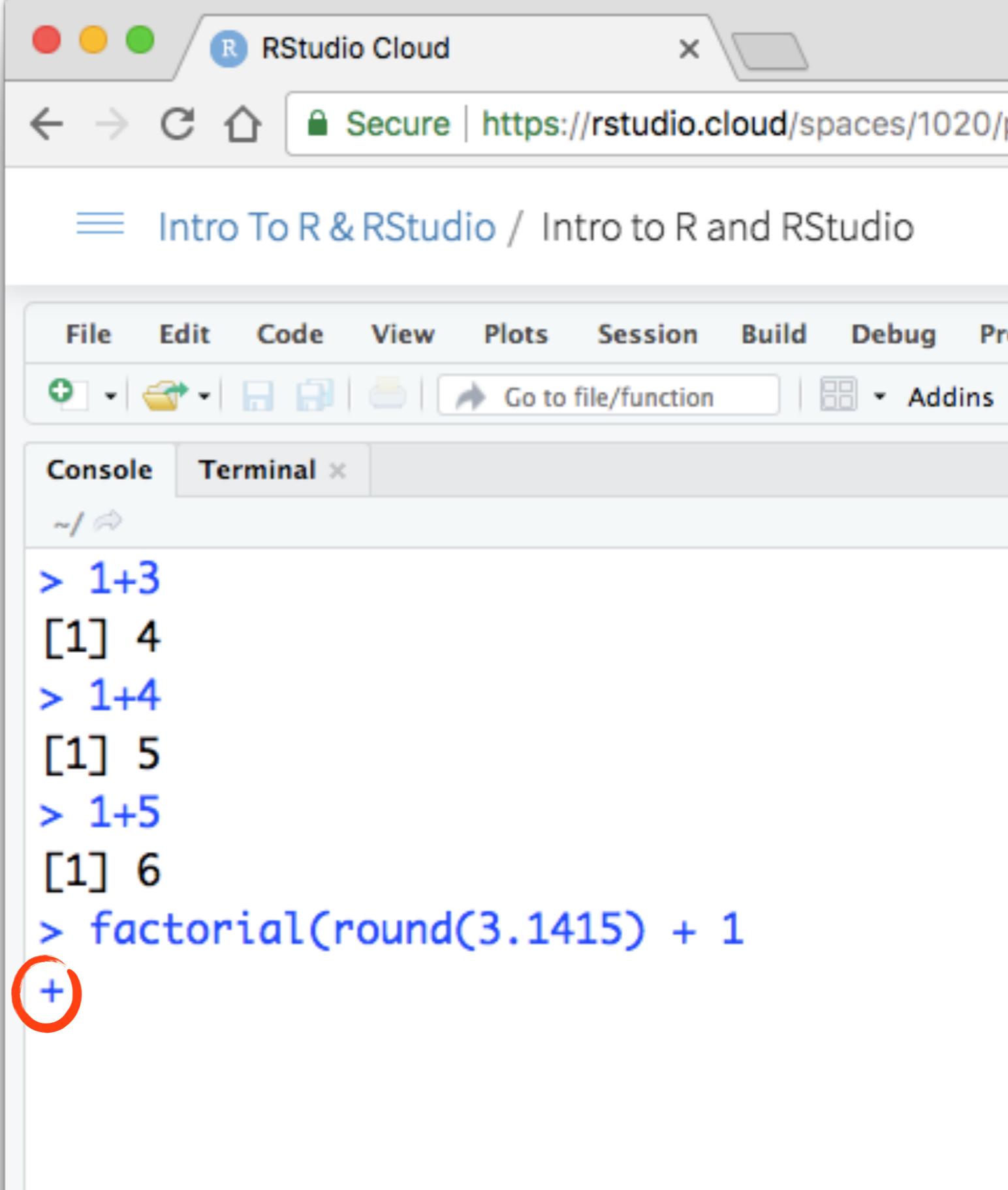
6

$$3! = 3 \times 2 \times 1$$

+ prompt

If your prompt turns into a "+", R thinks you haven't finished your previous command.

Either finish the command, or press escape.



The screenshot shows a browser window for RStudio Cloud at the URL <https://rstudio.cloud/spaces/1020/>. The title bar indicates it's a secure connection. Below the title bar is a navigation bar with icons for back, forward, search, and home. The main content area shows a terminal session in RStudio. The terminal tab is selected, showing the following interaction:

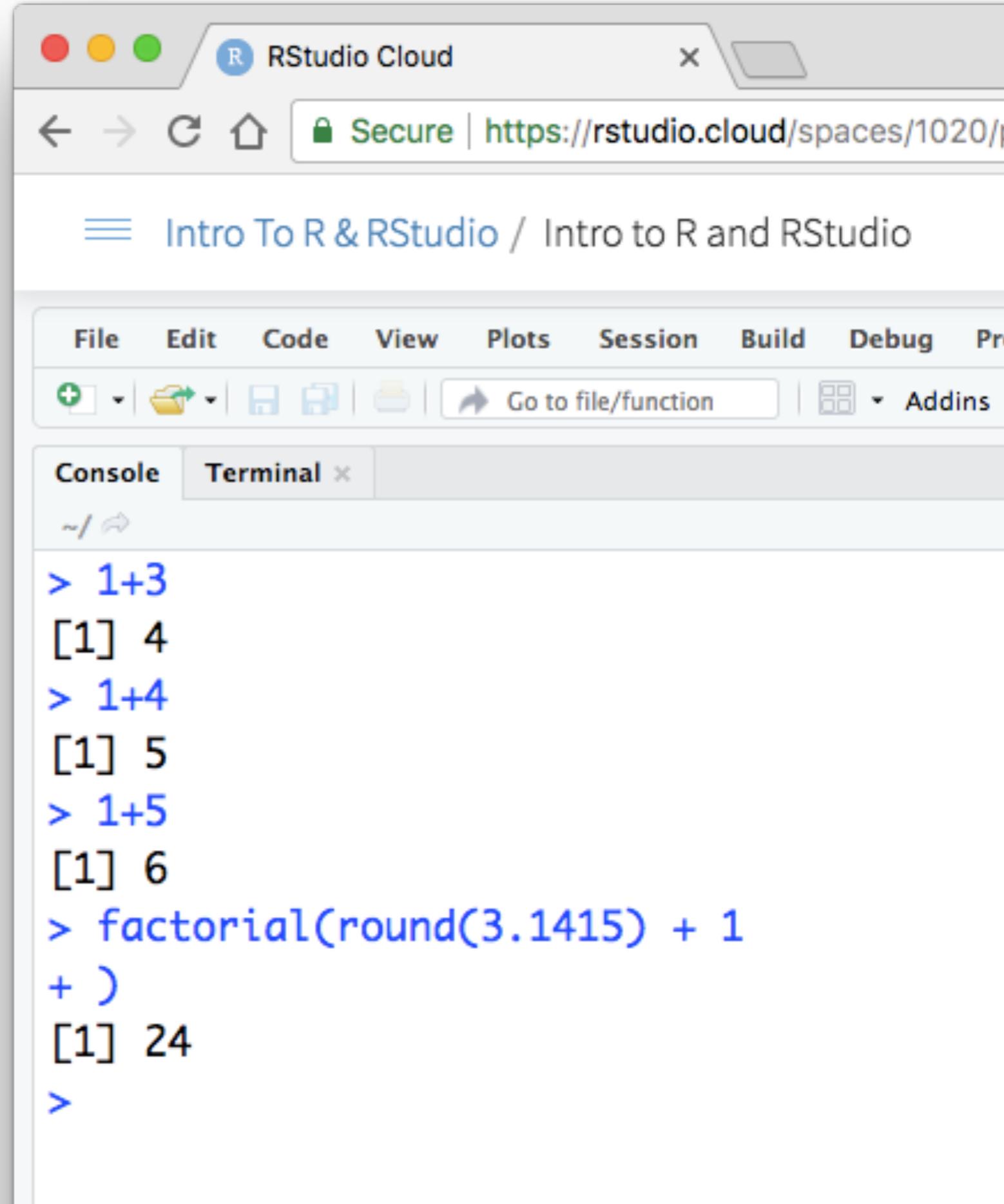
```
> 1+3  
[1] 4  
> 1+4  
[1] 5  
> 1+5  
[1] 6  
> factorial(round(3.1415) + 1
```

The last command, `factorial(round(3.1415) + 1`, has a red circle drawn around the final character, a '+', which is the focus of the slide's message about an incomplete command.

+ prompt

If your prompt turns into a "+", R thinks you haven't finished your previous command.

Either finish the command, or press escape.



The screenshot shows the RStudio Cloud interface. The title bar says "RStudio Cloud". The address bar is secure with "https://rstudio.cloud/spaces/1020/". The navigation bar includes "Intro To R & RStudio / Intro to R and RStudio". The menu bar has "File", "Edit", "Code", "View", "Plots", "Session", "Build", "Debug", and "Pr...". Below the menu is a toolbar with icons for file operations like new, open, save, and addins. The main area shows a "Terminal" tab active, displaying R session history:

```
~/
> 1+3
[1] 4
> 1+4
[1] 5
> 1+5
[1] 6
> factorial(round(3.1415) + 1
+ )
[1] 24
>
```

Your turn

Open RStudio and try the following tasks:

1. Pick a number and add 2 to it
2. Multiply the result by 3
3. Subtract 6 from the result of step 2
4. Divide the result of step 3 by 3

$10 + 2$

12

$12 * 3$

36

$36 - 6$

30

$30 / 3$

10

Workflow

R notebooks

It is easier to compose your code in an R Notebook than in the command line, and Notebooks allow you to keep text with your code.

We'll begin with a notebook I have started for you, called **01-Structures.Rmd**

Look in the folder **project**, then **code**, then click on

01-Intro.Rmd

RStudio Cloud Amelia

Secure | https://rstudio.cloud/spaces/1020/project/10706

Intro To R & RStudio / Intro to R and RStudio Amelia McNamara

File Edit Code View Plots Session Build Debug Profile Tools Help

Go to file/function Addins

R 3.4.2

01-Structures.Rmd

```
1 ---  
2 title: "R Notebook"  
3 output: html_notebook  
4 editor_options:  
5   chunk_output_type: inline  
6 ---  
7  
8 # Arithmetic and algebra  
9  
10 Let's start with the same arithmetic we were just doing in  
the Console. Notice that the math is surrounded by some  
special characters above and below, and has a grey  
background. This is RStudio's way of showing you the  
difference between an area to write text (here!) and an  
area where code is stored (called a code chunk).  
11  
12 `~`{r}  
13 5 + 5  
14 4 - 1  
15 1 * 2
```

R Markdown

Environment History Connections

Import Dataset Global Environment

Environment is empty

Files Plots Packages Help Viewer

New Folder Upload Delete Rename More

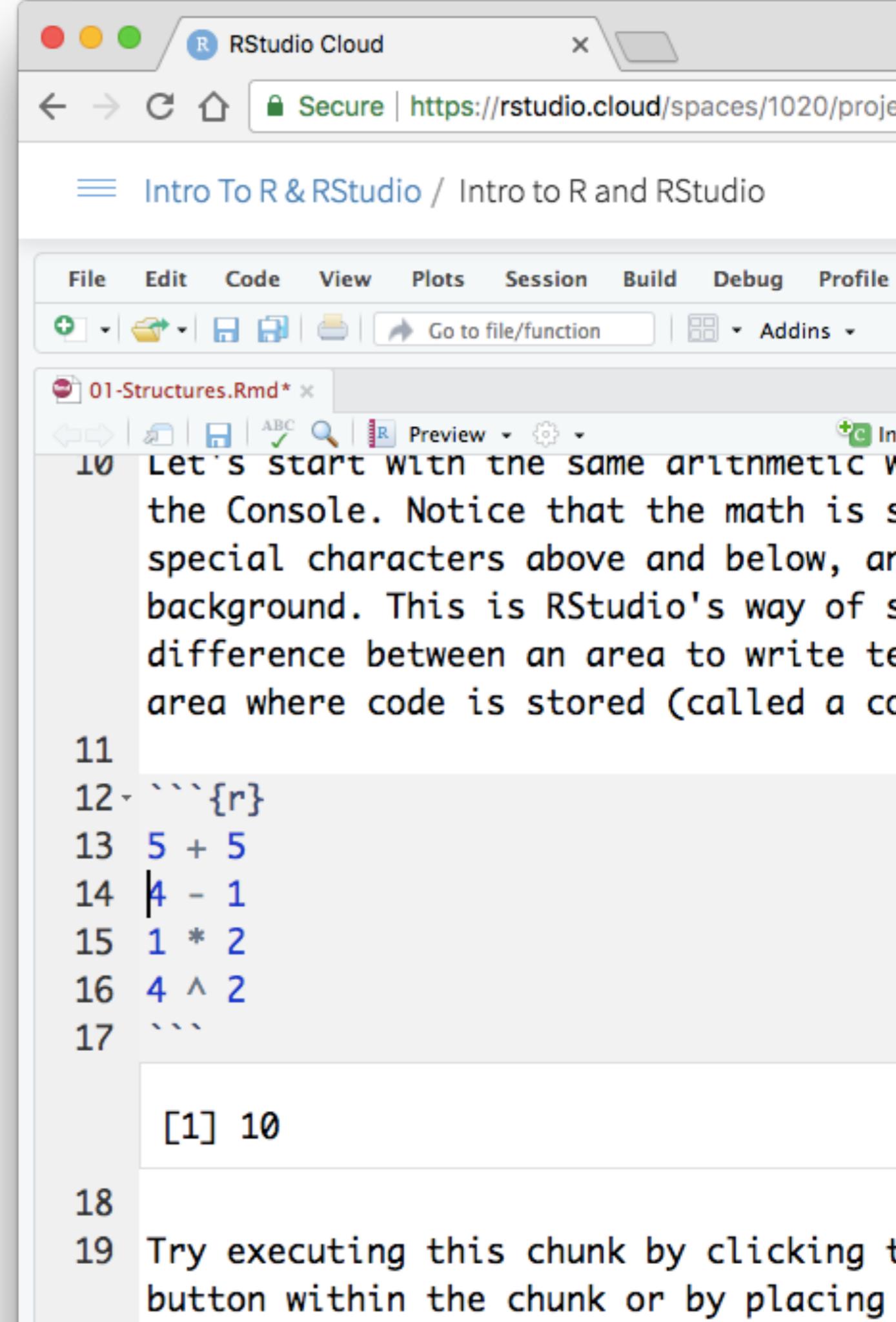
Home > Day1 > code

Name	Size	Modified
..		
01-Structures.Rmd	1.3 KB	Jan 27, 2018, 5:19 PM
02-Syntax.Rmd	869 B	Jan 27, 2018, 5:20 PM
03-Programming.Rmd	869 B	Jan 27, 2018, 5:20 PM

Console

The screenshot shows the RStudio Cloud interface. The main window displays an R Notebook file named '01-Structures.Rmd'. The code chunk from line 12 to 15 is highlighted with a red box. The text in the code chunk discusses the difference between text and code in the R Notebook. The right side of the interface shows the Environment pane (empty), the Files pane (listing '01-Structures.Rmd', '02-Syntax.Rmd', and '03-Programming.Rmd' files), and the Plots pane.

**Do what the text
instructs, and run a line
of the code. Notice how
results display
immediately below the
chunk, just like they did
in the Console.**



The screenshot shows the RStudio Cloud interface with the following details:

- Header:** RStudio Cloud, Secure | https://rstudio.cloud/spaces/1020/project
- Breadcrumbs:** Intro To R & RStudio / Intro to R and RStudio
- Toolbar:** File, Edit, Code, View, Plots, Session, Build, Debug, Profile
- File:** 01-Structures.Rmd*
- Code Preview:** Preview tab is active.
- Code Content:** Lines 10 through 17 show arithmetic operations:

```
10 Let's start with the same arithmetic we used in the Console. Notice that the math is stored in the code chunk, and the results are displayed immediately below the chunk, just like they did in the Console.  
11  
12 5 + 5  
13 4 - 1  
14 1 * 2  
15 4 ^ 2  
16  
17
```
- Output:** Line 18 shows the result of the addition: [1] 10
- Text Below Output:** Try executing this chunk by clicking the Run button within the chunk or by placing the cursor in the code and pressing Ctrl+Shift+Enter.

RStudio Cloud Amelia

Secure | https://rstudio.cloud/spaces/1020/project/10706

Intro To R & RStudio / Intro to R and RStudio Amelia McNamara

File Edit Code View Plots Session Build Debug Profile Tools Help

Addins Go to file/function R 3.4.2

01-Structures.Rmd

```
1 ---  
2 title: R Notebook  
3 output: html_notebook  
4 editor_options:  
5   chunk_output_type: inline  
6 ---  
7  
8 # Arithmetic and algebra  
9  
10 Let's start with the same code we were just going to write in the Console. Notice that the math is surrounded by some special characters above and below, and has a grey background. This is RStudio's way of showing you the difference between an area to write text (here!) and an area where code is stored (called a code chunk).  
11  
12 ````{r}  
13 5 + 5  
14 4 - 1  
15 1 * 2
```

Environment History Connections

Import Dataset Global Environment

File New Folder Upload Delete Rename More

01-Structures.Rmd 113 KB Jan 27, 2018, 5:19 PM
02-Syntax.Rmd 869 B Jan 27, 2018, 5:20 PM
03-Programming.Rmd 869 B Jan 27, 2018, 5:20 PM

Console

The screenshot shows the RStudio Cloud interface. A large, bold black text overlay is centered in the middle of the screen, reading: "Notice that the Console has automatically minimized itself, to give you room to work in your notebook. From here out, we'll be working almost exclusively in notebooks, but all the code we write would work in the Console as well." Below this text, the RStudio interface is visible, including the code editor with R Markdown syntax, the global environment, and the file browser. A red box highlights the "Console" tab at the bottom left of the interface.

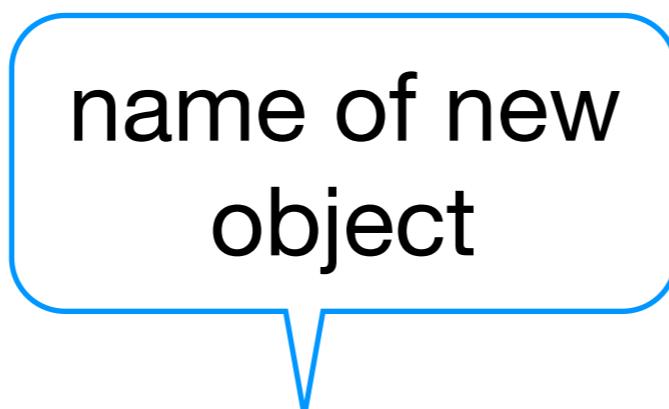
R objects

You can save information as an R object with the greater than sign followed by a minus, e.g, an arrow: <-

```
the_answer <- 42
```

You can save information as an R object with the greater than sign followed by a minus, e.g, an arrow: <-

name of new
object



```
the_answer <- 42
```

You can save information as an R object with the greater than sign followed by a minus, e.g, an arrow: <-

assignment
operator,
"gets"

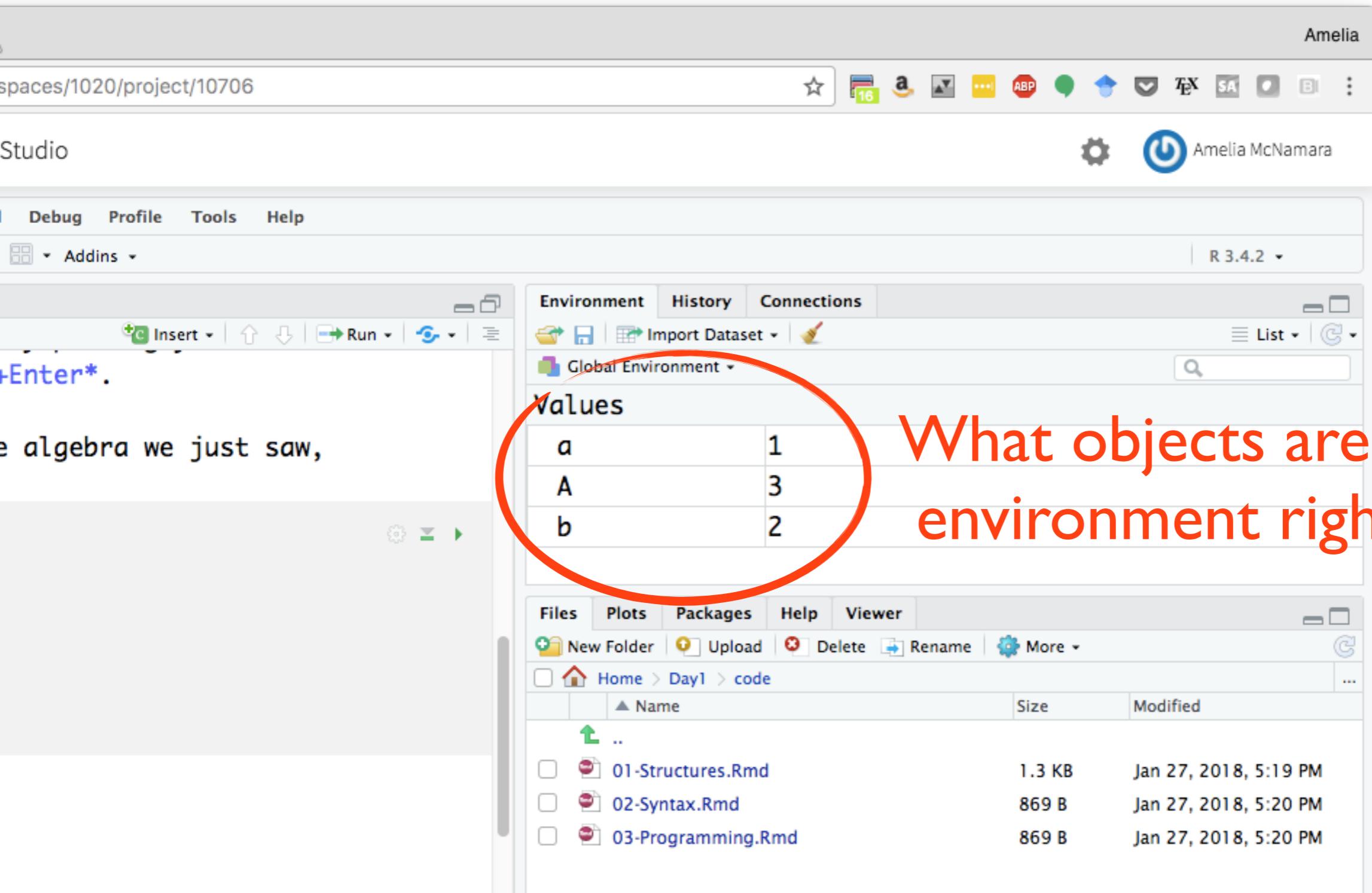
the_answer <- 42

You can save information as an R object with the greater than sign followed by a minus, e.g, an arrow: <-

information
to store in the
object

```
the_answer <- 42
```

When you create an R object, you'll see it appear in your environment pane



The screenshot shows the RStudio interface. The top bar displays the title "Amelia" and the path "spaces/1020/project/10706". The toolbar includes icons for file operations like Insert, Run, and Save. The main window has tabs for "Environment", "History", and "Connections", with "Environment" selected. A red circle highlights the "Values" table in the Environment pane, which contains three rows: "a" with value 1, "A" with value 3, and "b" with value 2. To the right of the table, a large red text overlay asks, "What objects are in your environment right now?". Below the Environment pane, the "Files" tab is active, showing a directory structure under "Home > Day1 > code" with files "01-Structures.Rmd", "02-Syntax.Rmd", and "03-Programming.Rmd".

What objects are in your environment right now?

Name	Value
a	1
A	3
b	2

Name	Size	Modified
01-Structures.Rmd	1.3 KB	Jan 27, 2018, 5:19 PM
02-Syntax.Rmd	869 B	Jan 27, 2018, 5:20 PM
03-Programming.Rmd	869 B	Jan 27, 2018, 5:20 PM

Common R workflow

Save output of one function as an R object
to use in a second function.

```
more_pi <- round(3.1415) + 1
```

```
more_pi
```

```
# 4
```

```
factorial(more_pi)
```

```
# 24
```

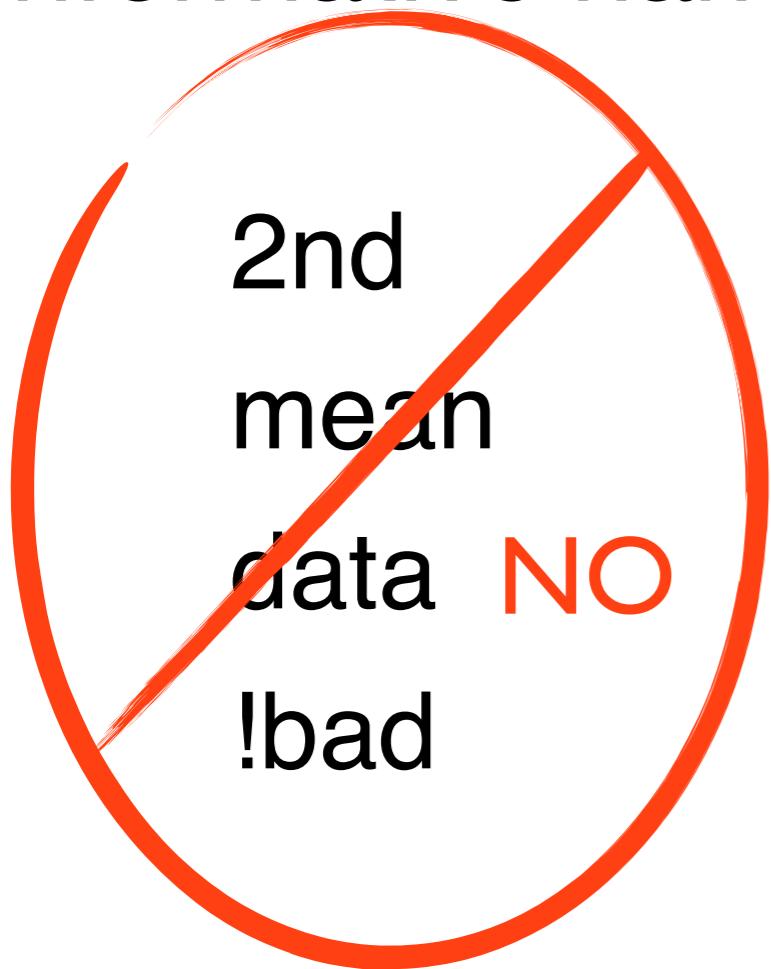
Object names

Object names cannot begin with numbers

They cannot contain spaces

It is wise to avoid names already in use

Informative names are better than generic ones



a
b
FOO
my_var
.day

CDC_data
finalModel
more_pi **yeah!**
withoutOver64

eh

Capitalization matters

R will treat each of these as a different object

cdc_data

finalmodel

sum

CDC_data

finalModel

SUM

rm

You can remove an object from your environment with rm

```
more_pi
```

```
# 4
```

```
rm(more_pi)
```

```
more_pi
```

```
# Error: object 'more_pi' not found
```

This can be useful if you overwrite an object that comes with R

```
pi
```

```
# 3.141593
```

```
pi <- 1
```

```
pi
```

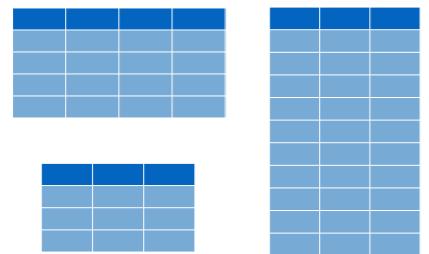
```
# 1
```

```
rm(pi)
```

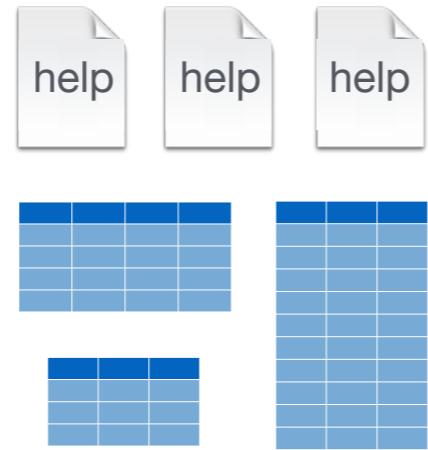
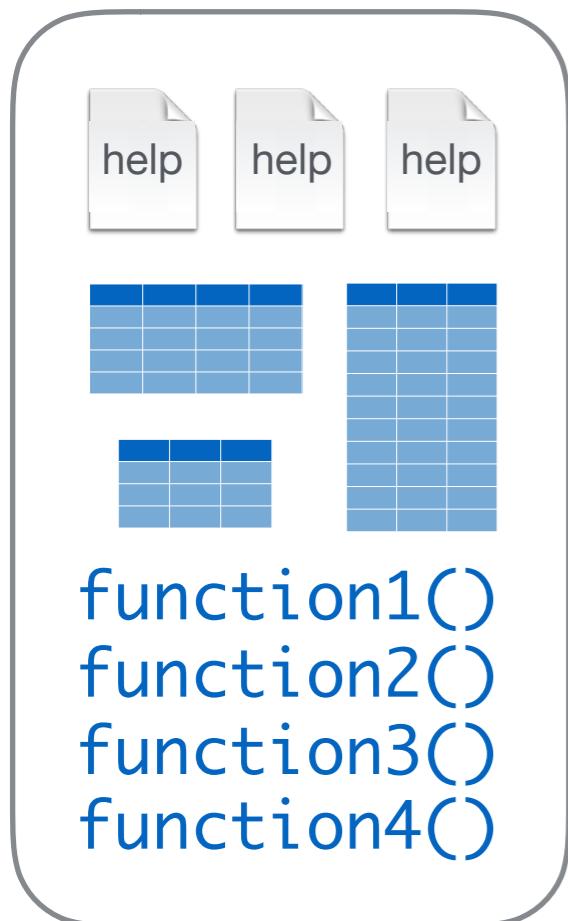
```
pi
```

```
# 3.141593
```

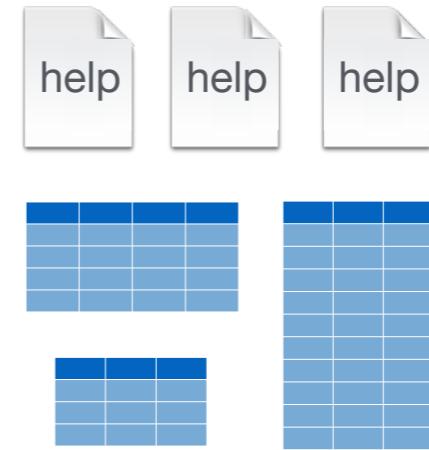
R packages



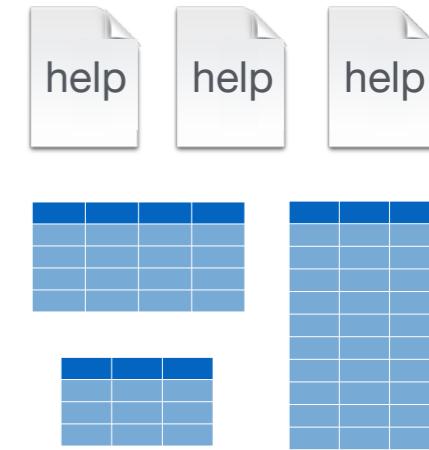
function1()
function2()
function3()
function4()



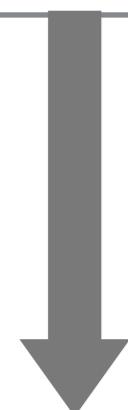
function5O
function6O
function7O
function8O



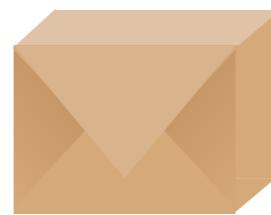
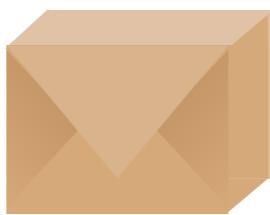
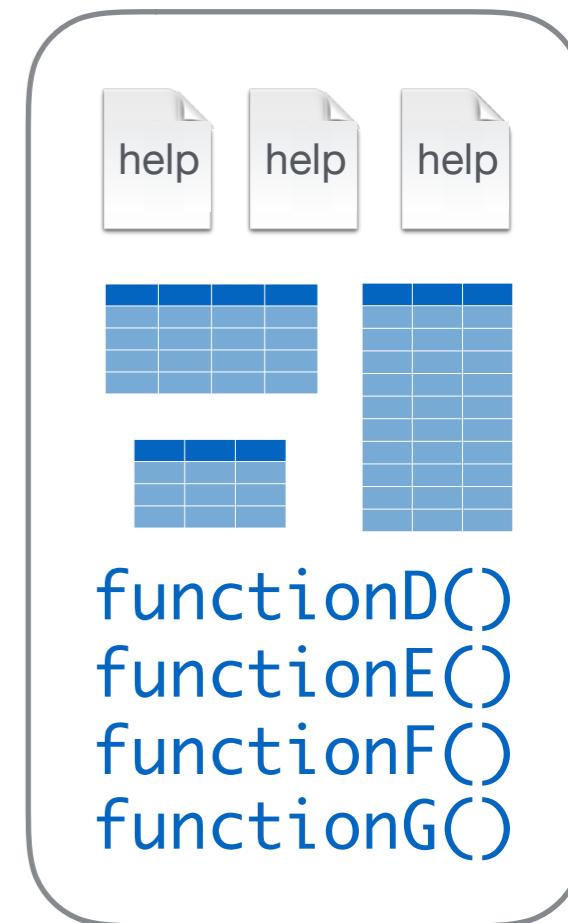
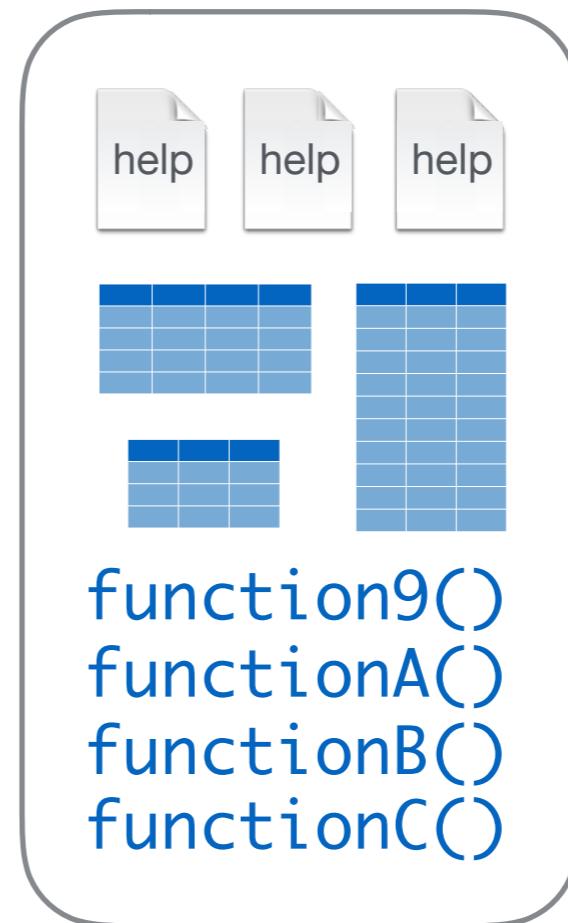
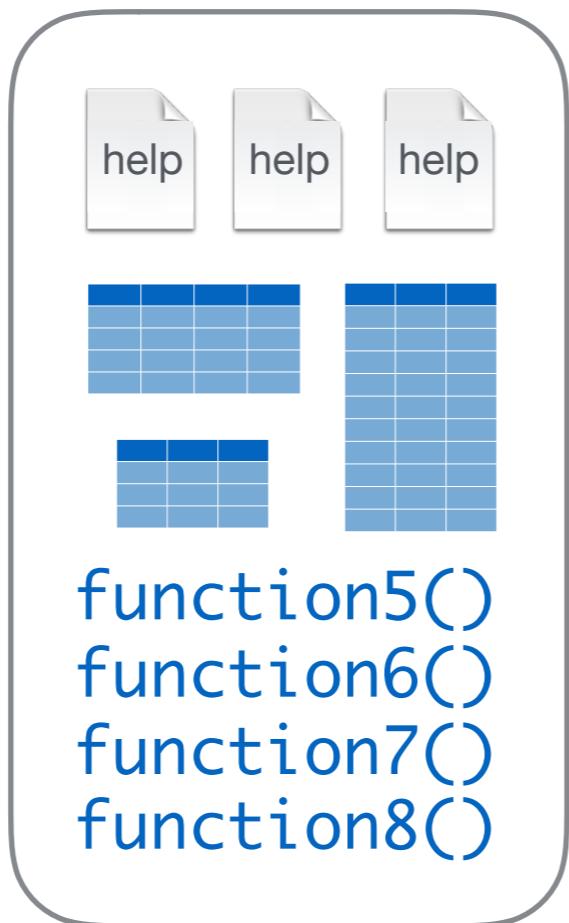
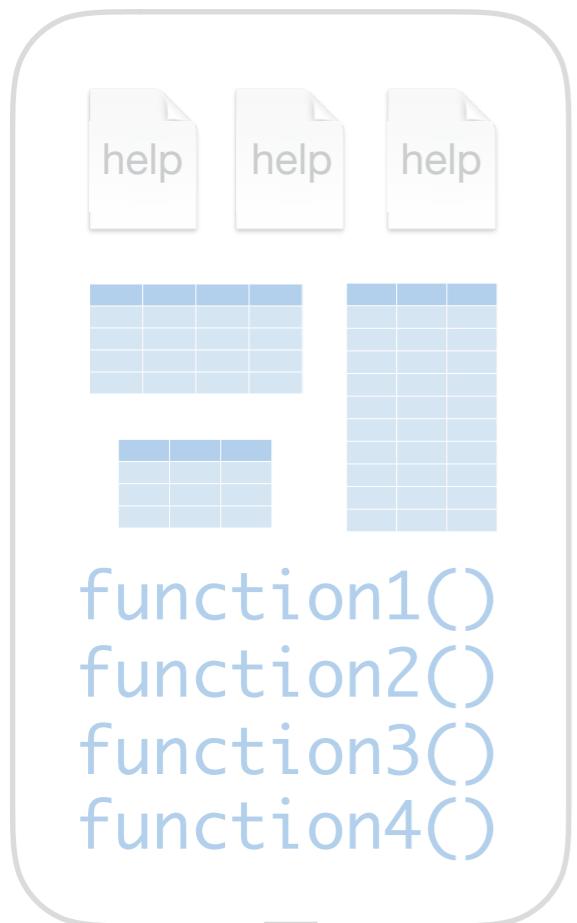
function9O
functionAO
functionBO
functionCO



functionDO
functionEO
functionFO
functionGO



Base R



Base R

R Packages

The Comprehensive R Archive X Amelia

Secure | https://cran.r-project.org



Available CRAN Packages By Name

[A](#) [B](#) [C](#) [D](#) [E](#) [F](#) [G](#) [H](#) [I](#) [J](#) [K](#) [L](#) [M](#) [N](#) [O](#) [P](#) [Q](#) [R](#) [S](#) [T](#) [U](#) [V](#) [W](#) [X](#) [Y](#) [Z](#)

A3	Accurate, Adaptable, and Accessible Error Metrics for Predictive Models
abbyyR	Access to Abbyy Optical Character Recognition (OCR) API
abc	Tools for Approximate Bayesian Computation (ABC)
abc.data	Data Only: Tools for Approximate Bayesian Computation (ABC)
ABC.RAP	Array Based CpG Region Analysis Pipeline
ABCanalysis	Computed ABC Analysis
abcdeFBA	ABCDE_FBA: A-Biologist-Can-Do-Everything of Flux Balance Analysis with this package
ABCOptim	Implementation of Artificial Bee Colony (ABC) Optimization
ABCp2	Approximate Bayesian Computational Model for Estimating P2
abcrf	Approximate Bayesian Computation via Random Forests
abctools	Tools for ABC Analyses
abd	The Analysis of Biological Data
abe	Augmented Backward Elimination
abf2	Load Gap-Free Axon ABF2 Files
ABHgenotypeR	Easy Visualization of ABH Genotypes
abind	Combine Multidimensional Arrays
abjutils	Useful Tools for Jurimetical Analysis Used by the Brazilian Jurimetrics Association
abn	Modelling Multivariate Data with Additive Bayesian Networks
abodOutlier	Angle-Based Outlier Detection
ABPS	The Abnormal Blood Profile Score to Detect Blood Doping
AbsFilterGSEA	Improved False Positive Control of Gene-Permuting GSEA with Absolute Filtering
AbSim	Time Resolved Simulations of Antibody Repertoires
abundant	High-Dimensional Principal Fitted Components and Abundant Regression
ACA	Abrupt Change-Point or Aberration Detection in Point Series
acc	Exploring Accelerometer Data
accelrometry	Functions for Processing Minute-to-Minute Accelerometer Data
accelmissing	Missing Value Imputation for Accelerometer Data
AcceptanceSampling	Creation and Evaluation of Acceptance Sampling Plans
ACCLMA	ACC & LMA Graph Plotting

Using packages

1.

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

Downloads files to computer

1 x per computer

2.

```
library(ggplot2)
```

Loads package

1 x per R Session

I've done this
for you for this
workshop

The tidyverse

Tidyverse

Secure | https://www.tidyverse.org

Packages Articles Learn Help Contribute

R packages for data science

The tidyverse is an opinionated [collection of R packages](#) designed for data science. All packages share an underlying design philosophy, grammar, and data structures.

Install the complete tidyverse with:

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

Tidyverse

Secure | https://www.tidyverse.org

Packages Articles Learn Help Contribute

R packages for data science

The tidyverse is an opinionated collection of R packages designed for data science.

Install the complete tidyverse with:

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

tidyverse



An R package that serves as a short cut for installing and loading the components of the tidyverse.

```
library("tidyverse")
```

does the equivalent of

```
library("ggplot2")
library("dplyr")
library("tidyr")
library("readr")
library("purrr")
library("tibble")
```

R Syntax Comparison :: CHEAT SHEET

Dollar sign syntax

```
goal(data$x, data$y)
```

SUMMARY STATISTICS:

one continuous variable:
`mean(mtcars$mpg)`

one categorical variable:
`table(mtcars$cyl)`

two categorical variables:
`table(mtcars$cyl, mtcars$am)`

one continuous, one categorical:
`mean(mtcars$mpg[mtcars$cyl==4])`
`mean(mtcars$mpg[mtcars$cyl==6])`
`mean(mtcars$mpg[mtcars$cyl==8])`

PLOTTING:

one continuous variable:
`hist(mtcars$disp)`

```
boxplot(mtcars$disp)
```

one categorical variable:
`barplot(table(mtcars$cyl))`

two continuous variables:
`plot(mtcars$disp, mtcars$mpg)`

two categorical variables:
`mosaicplot(table(mtcars$am, mtcars$cyl))`

one continuous, one categorical:
`histogram(mtcars$disp[mtcars$cyl==4])`
`histogram(mtcars$disp[mtcars$cyl==6])`
`histogram(mtcars$disp[mtcars$cyl==8])`

```
boxplot(mtcars$disp[mtcars$cyl==4])
boxplot(mtcars$disp[mtcars$cyl==6])
boxplot(mtcars$disp[mtcars$cyl==8])
```

WRANGLING:

subsetting:
`mtcars[mtcars$mpg>30]`

making a new variable:
`mtcars$efficient`
`mtcars$efficien`

Formula syntax

```
goal(y~x|z, data=data, group=w)
```

SUMMARY STATISTICS:
one continuous variable:
`mosaic::mean(~mpg, data=mtcars)`

one categorical variable:
`mosaic::tally(~cyl, data=mtcars)`

two categorical variables:
`mosaic::tally(cyl~am, data=mtcars)`

one continuous, one categorical:
`mosaic::mean(mpg~cyl, data=mtcars)`

tilde

PLOTTING:

one continuous variable:
`lattice::histogram(~disp,`

```
lattice::bwplot(~disp,
```

one categorical variable:
`mosaic::bargraph(~cyl,`

two continuous variables:
`lattice::xyplot(mpg~disp,`

two categorical variables:
`mosaic::facet_grid(.~am,`

one continuous, one categorical:
`ggplot(data=mtcars, aes(x=disp, y=mpg))`

`geom="point")`

`data=mtcars)`

`data=mtcars)`

I've given you a copy of this cheatsheet in the "cheatsheets" folder of your workspace, and I have paper copies for anyone who wants one!

Please read across the cheatsheet to see how different R syntaxes give different ways to "say" the same thing.



SMITH

Tidyverse syntax

SUMMARY STATISTICS:

one continuous variable:
`mtcars %>%`

one categorical variable:
`mtcars %>%`

two categorical variables:
`mtcars %>%`

one continuous, one categorical:
`mtcars %>% group_by(cyl) %>%`
`summarize(mean(mpg))`

one continuous variable:
`ggplot(x=mpg, data=mtcars, geom = "histogram")`

```
ggplot2::qplot(y=disp, x=1, data=mtcars, geom="boxplot")
```

one categorical variable:
`ggplot2::qplot(x=cyl, data=mtcars, geom="bar")`

two continuous variables:
`ggplot2::qplot(x=disp, y=mpg, data=mtcars, geom="point")`

two categorical variables:
`ggplot2::qplot(x=factor(cyl), data=mtcars, geom="bar") +`
`facet_grid(.~am)`

one continuous, one categorical:
`ggplot2::qplot(x=disp, data=mtcars, geom = "histogram") +`
`facet_grid(.~cyl)`

```
ggplot2::qplot(y=disp, x=factor(cyl), data=mtcars,
geom="boxplot")
```

WRANGLING:

subsetting:
`mtcars %>% dplyr::filter(mpg>30)`

making a new variable:
`mtcars <- mtcars %>%`
`dplyr::mutate(efficient = if_else(mpg>30, TRUE, FALSE))`

the pipe