

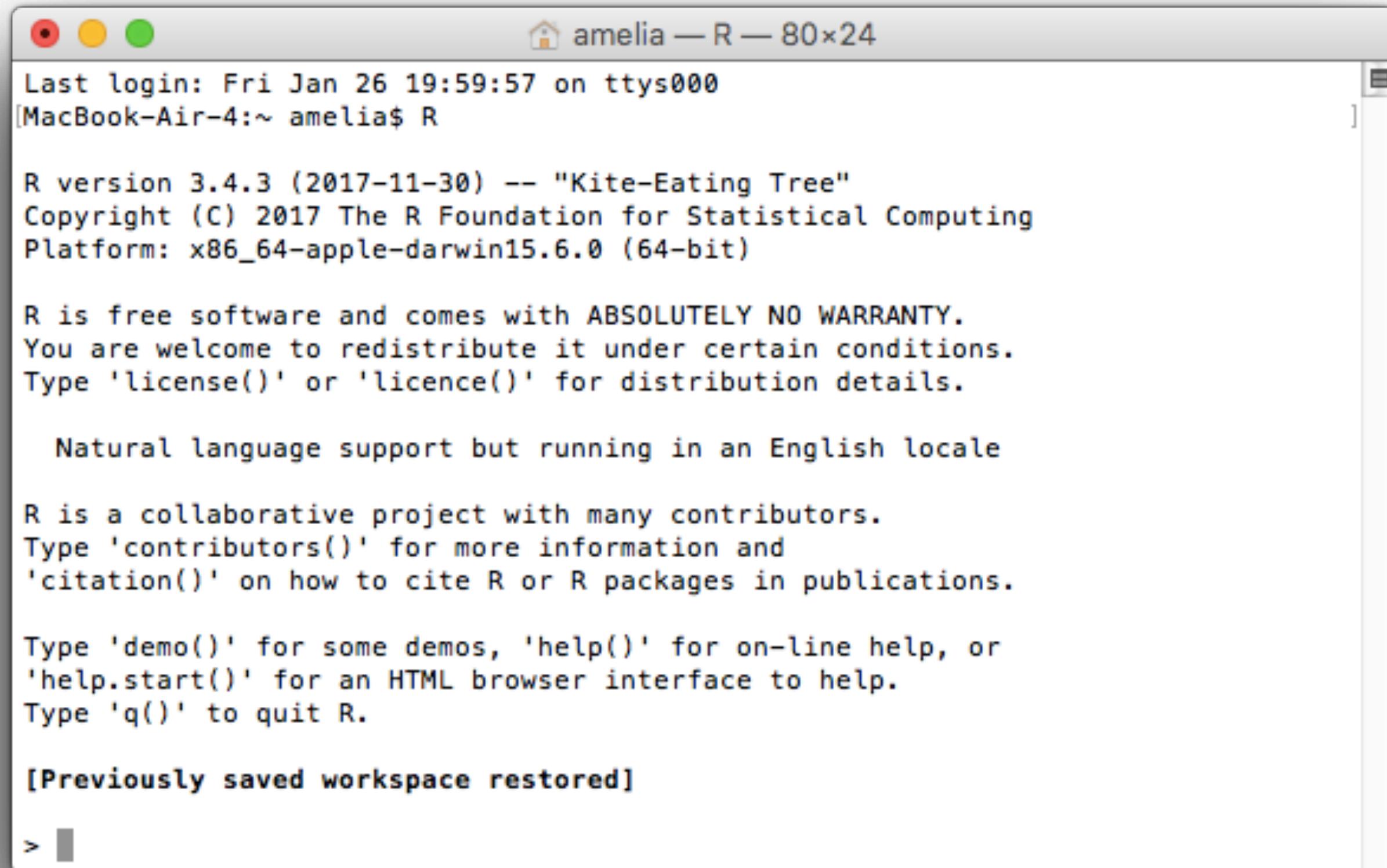
These materials adapted by Amelia McNamara from
the RStudio [CC BY-SA](#) materials Introduction to R
(2014) and [Master the Tidyverse](#) (2017).

Introduction to R & RStudio: deck 01: Getting started

Amelia McNamara



R: a computer programming language



The image shows a screenshot of an R terminal window titled "amelia — R — 80x24". The window displays the standard R startup message, which includes the last login information ("Last login: Fri Jan 26 19:59:57 on ttys000 [MacBook-Air-4:~ amelia\$ R]"), the R version ("R version 3.4.3 (2017-11-30) -- \"Kite-Eating Tree\""), copyright information ("Copyright (C) 2017 The R Foundation for Statistical Computing"), the platform ("Platform: x86_64-apple-darwin15.6.0 (64-bit)"), the license notice ("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 ("Natural language support but running in an English locale"), collaborative project information ("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."), help information ("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."), and a workspace restoration message ("[Previously saved workspace restored]"). A cursor is visible at the bottom of the terminal window.

```
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

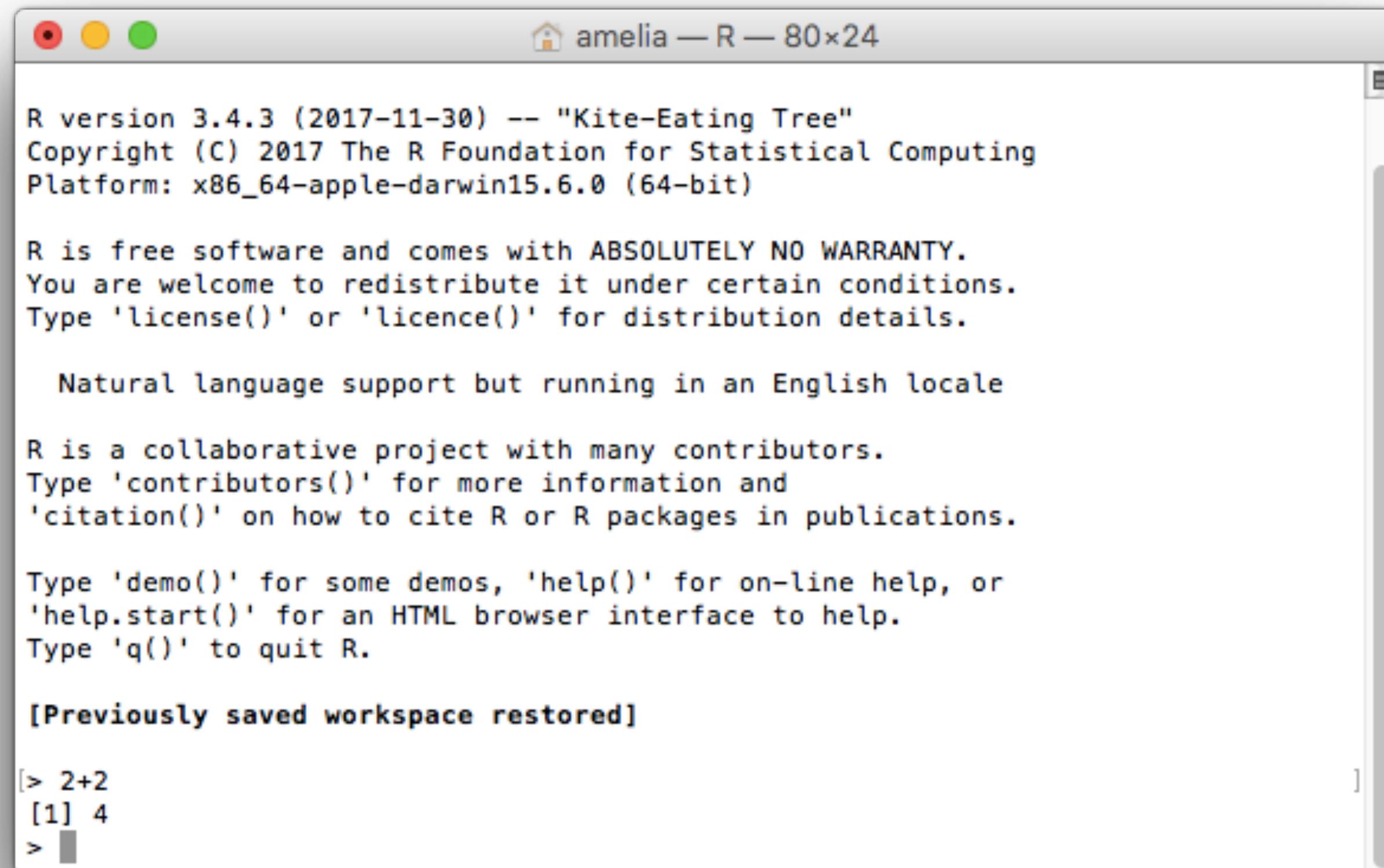
R is a collaborative project with many contributors.
Type 'contributors()' for more information and
'citation()' on how to cite R or R packages in publications.

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.

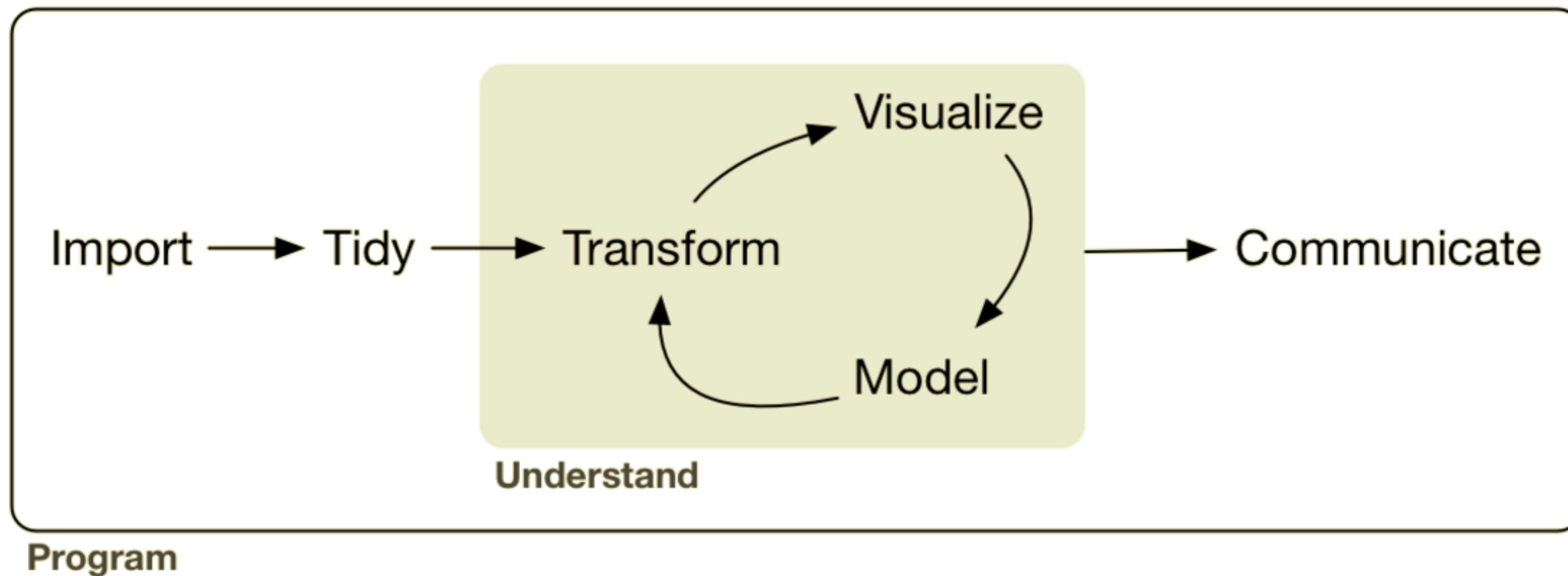
[Previously saved workspace restored]

[> 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

R: designed for data



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

```
> 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

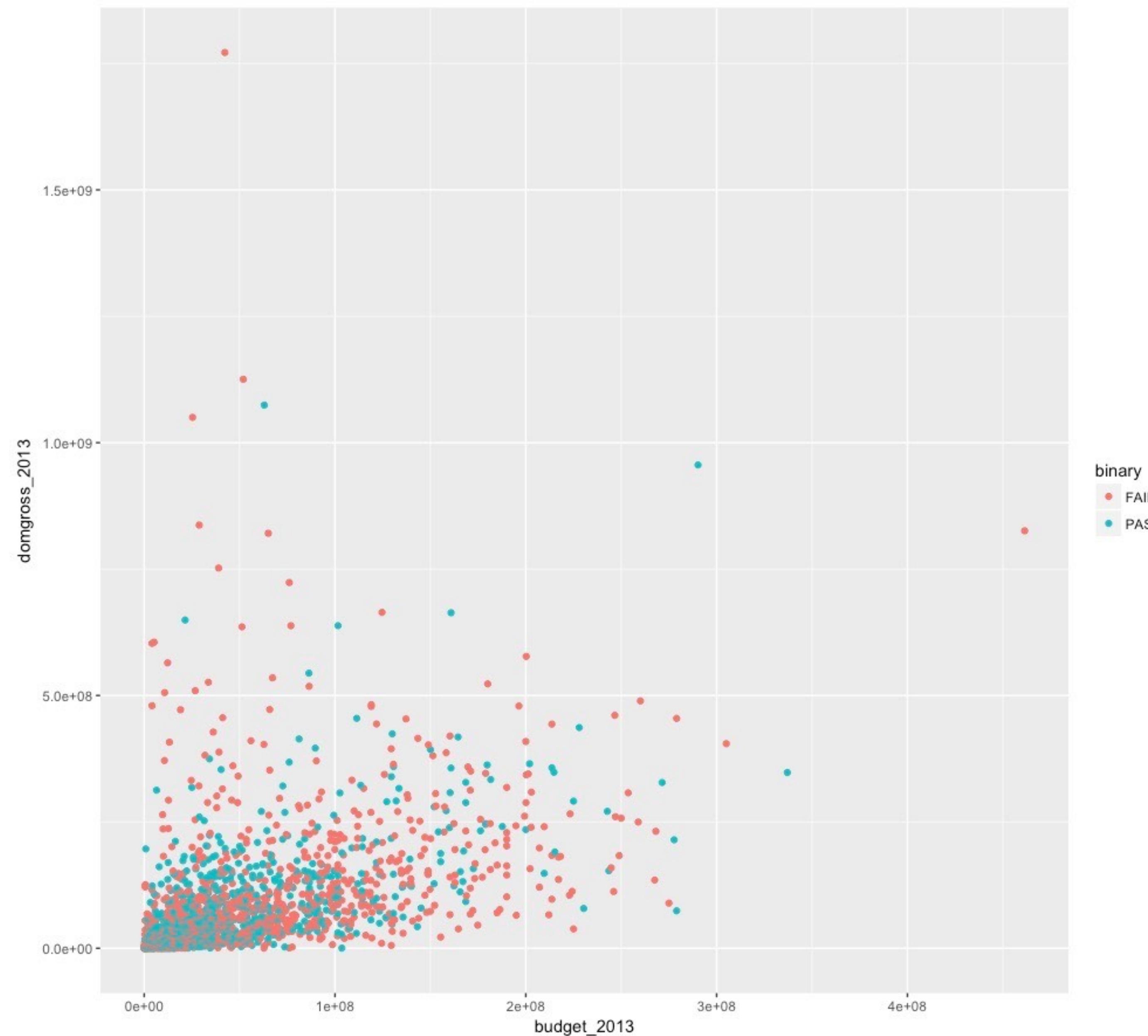
n obs: 1794

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 70 300

Year released

1,940 1,970 2,014

Minimum number of Oscar wins (all categories)

0 4

Dollars at Box Office (millions)

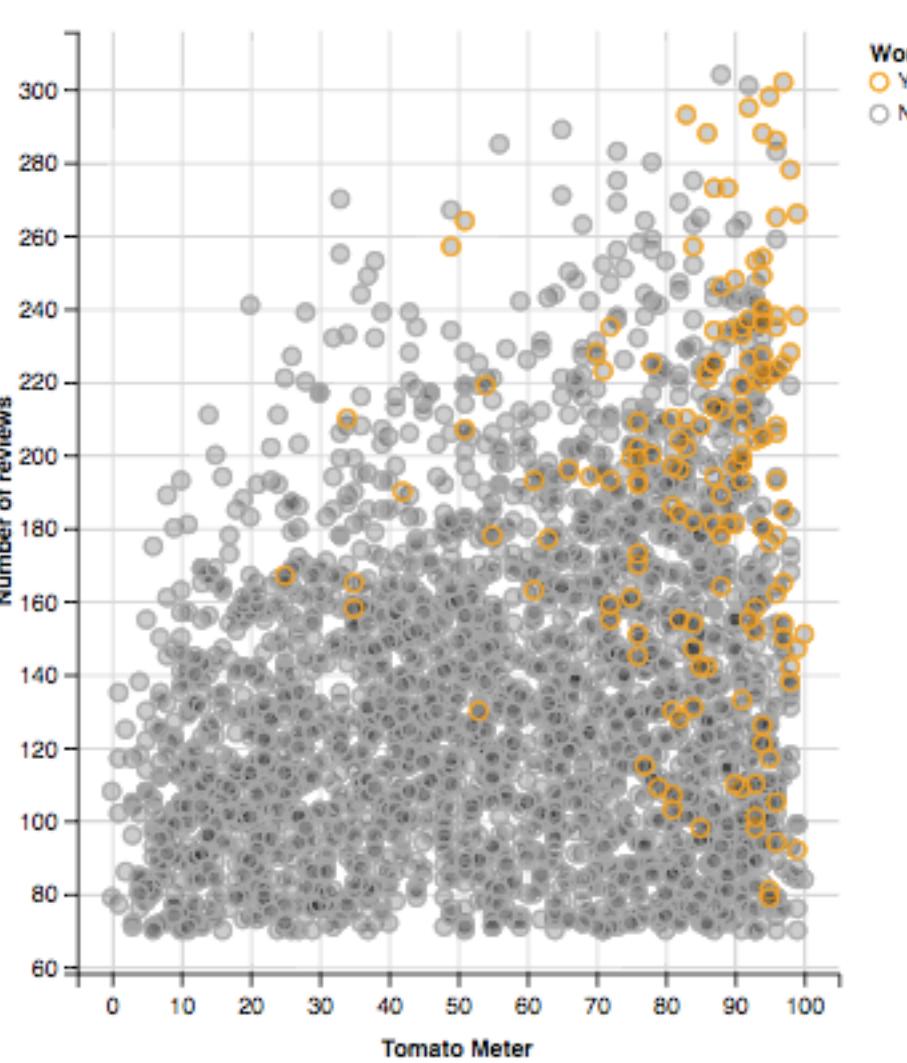
0 800

Genre (a movie can have multiple genres)

All

Director name contains (e.g., Miyazaki)

Cast names contains (e.g. Tom Hanks)



Number of movies selected:
2758

X-axis variable

Tomato Meter

Y-axis variable

Number of reviews

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).

The screenshot shows a web application titled "intRo" running on a Mac OS X system. The browser window has tabs for "intRo" and "www.intro-stats.com". The main content area displays a data analysis interface for the "MPG" dataset.

Sidebar:

- Data
- Sources** (highlighted)
- Transform
- Summaries
- Graphical
- Numerical
- Inference
- Contingency
- Regression
- T test

Main Content Area:

Choose Dataset: MPG

Show 10 entries

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

manufacterer model displ year cyl trans drv cty hwy fl

Showing 1 to 10 of 234 entries

Previous 1 2 3 4 5 ...

24 Next

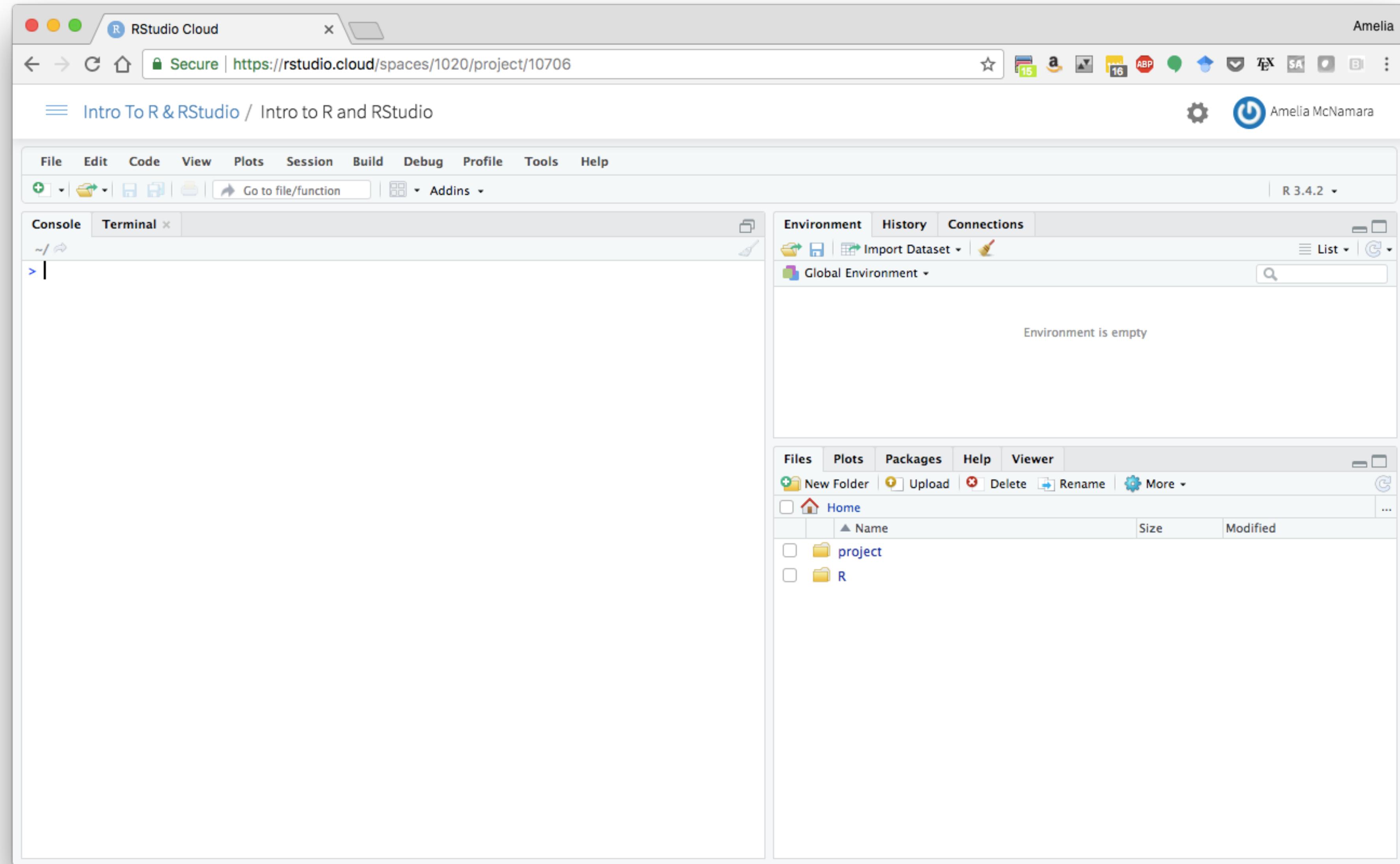
<http://www.intro-stats.com/>



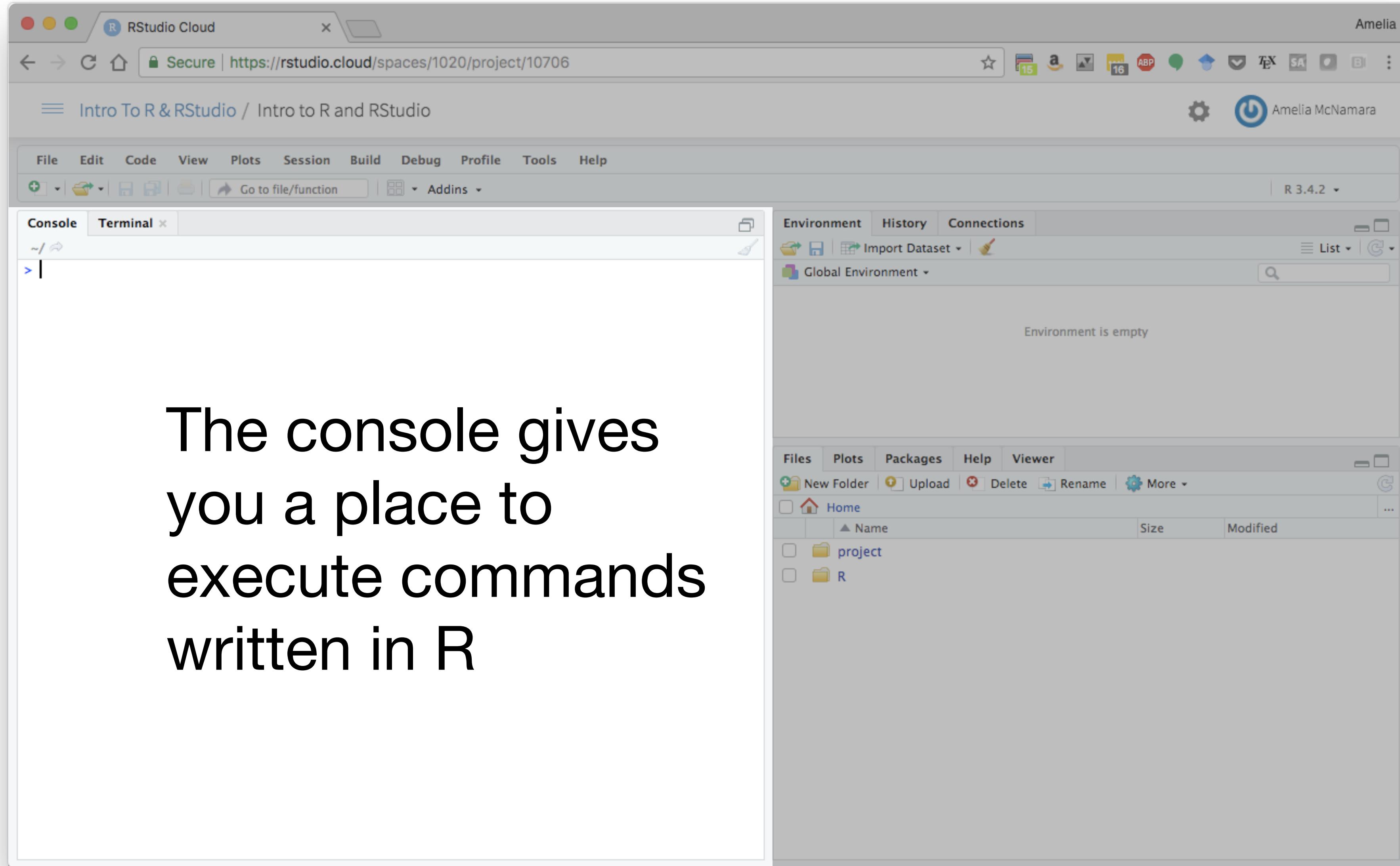
RStudio: a software program

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

RStudio



RStudio

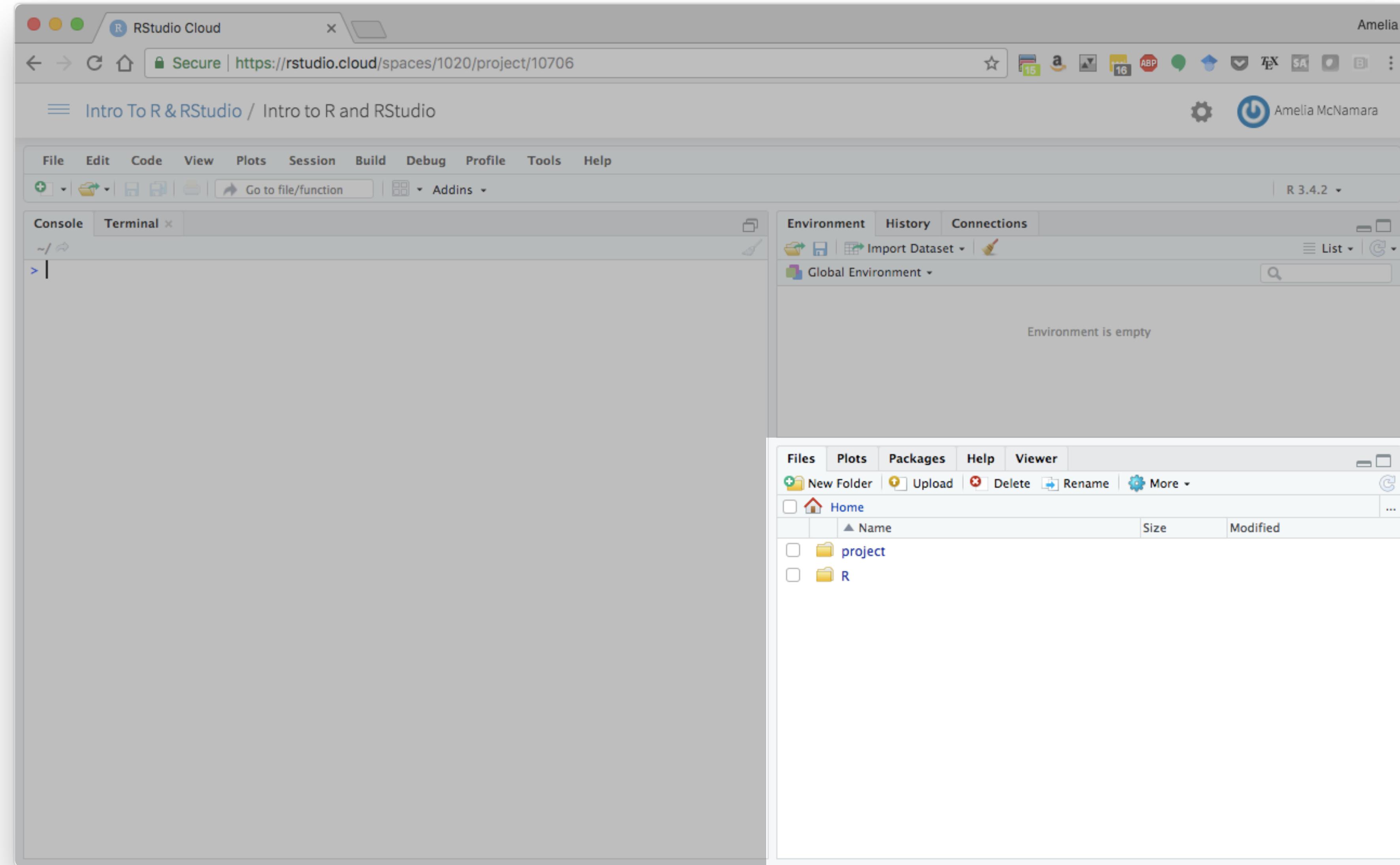


The screenshot shows the RStudio Cloud interface. The top navigation bar includes a back button, forward button, refresh button, and a secure connection indicator. The title bar says "RStudio Cloud" and "Secure | https://rstudio.cloud/spaces/1020/project/10706". The user "Amelia McNamara" is logged in. The main window has several panes:

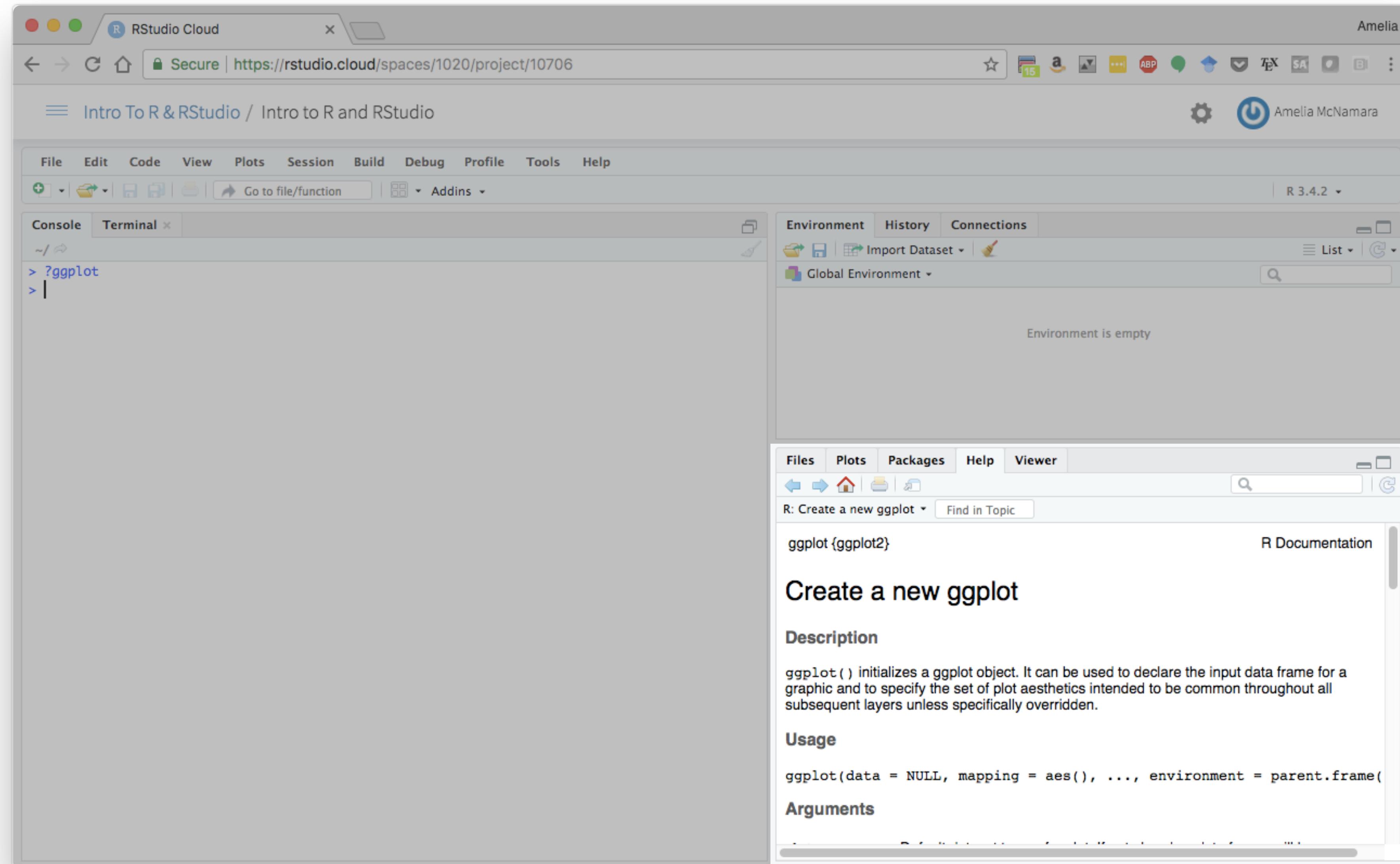
- Console**: Shows a command prompt starting with "> |".
- Environment**: Shows the "Global Environment" tab with the message "Environment is empty".
- Files**: Shows a file tree with "Home" selected, containing "project" and "R" folders.
- Plots**: Not visible in the screenshot.
- Packages**: Not visible 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

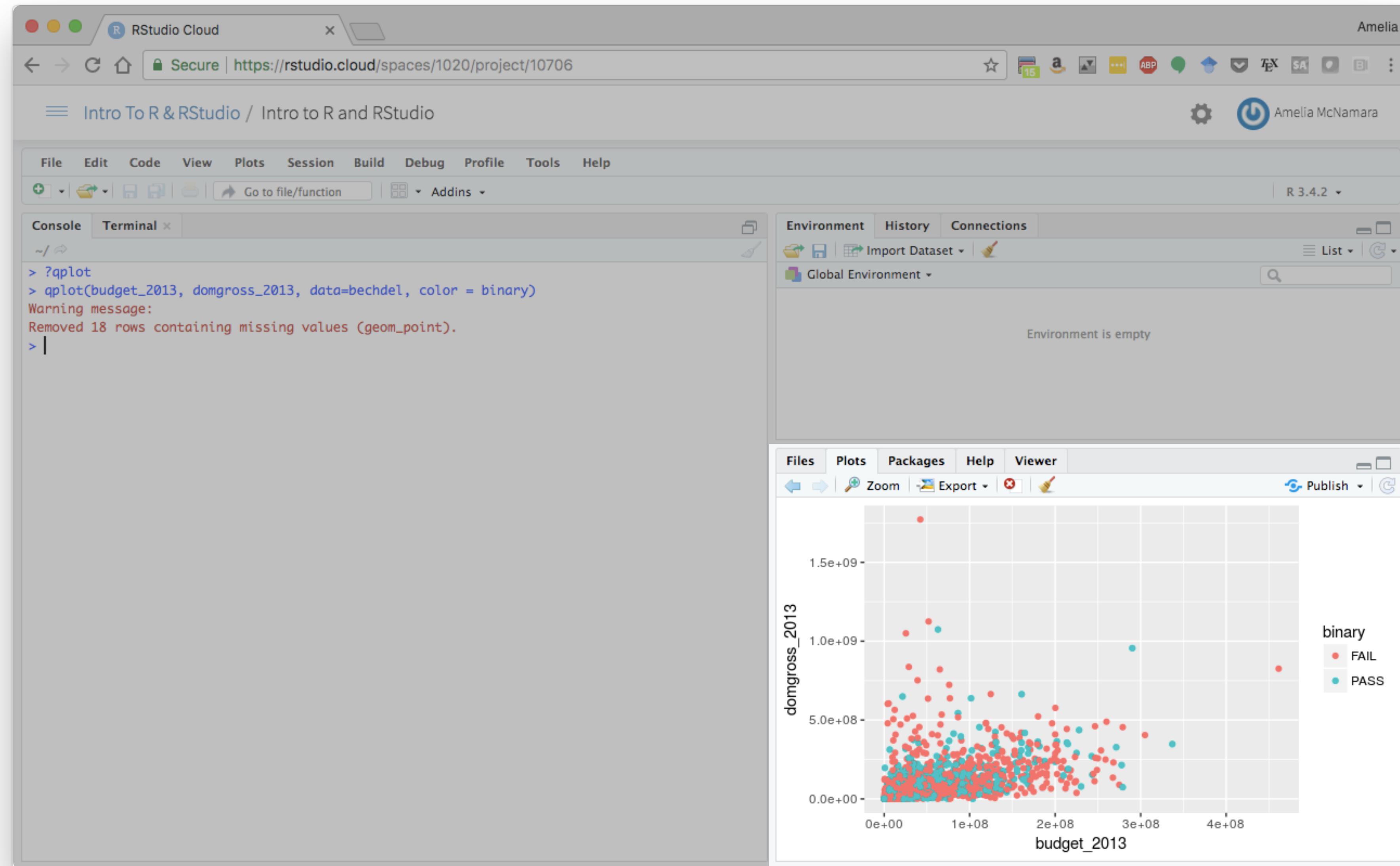
RStudio



RStudio



RStudio



RStudio

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

Intro To R & RStudio / Intro to R and RStudio

Console Terminal

```
> ?qplot
> qplot(budget_2013, domgross_2013, data=bechdel, color = binary)
Warning message:
Removed 18 rows containing missing values (geom_point).
> |
```

Environment History Connections

```
dechdel %>% skim(domgross_2013)
library(skimr)
data(bechdel)
bechdel %>% skim(domgross_2013)
bechdel %>% skim(clean_test)
qplot(budget_2013, domgross_2013, data=bechdel, color = binary)
lm(domgross_2013~budget_2013, data=bechdel)
?qplot
qplot(budget_2013, domgross_2013, data=bechdel, color = binary)
```

Files Plots Packages Help Viewer

domgross_2013

budget_2013

binary

- FAIL
- PASS

RStudio

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

Intro To R & RStudio / Intro to R and RStudio

File Edit Code View Plots Session Build Debug Profile Tools Help

Untitled1 x Go to file/function Addins R 3.4.2

```
1 ---  
2 title: "Untitled"  
3 output: html_document  
4 ---  
5  
6 ```{r setup, include=FALSE}  
7 knitr::opts_chunk$set(echo = TRUE)  
8  
9  
10 ## R Markdown  
11  
12 This is an R Markdown document. Markdown is a simple formatting syntax for authoring  
HTML, PDF, and MS Word documents. For more details on using R Markdown see  
http://rmarkdown.rstudio.com.  
13  
14 When you click the **Knit** button a document will be generated that includes both  
2:1 # Untitled
```

Console Terminal

```
> ?qplot  
> qplot(budget_2013, domgross_2013, data=bechdel, color = binary)  
Warning message:  
Removed 18 rows containing missing values (geom_point).  
>
```

Environment History Connections

Import Dataset Global Environment

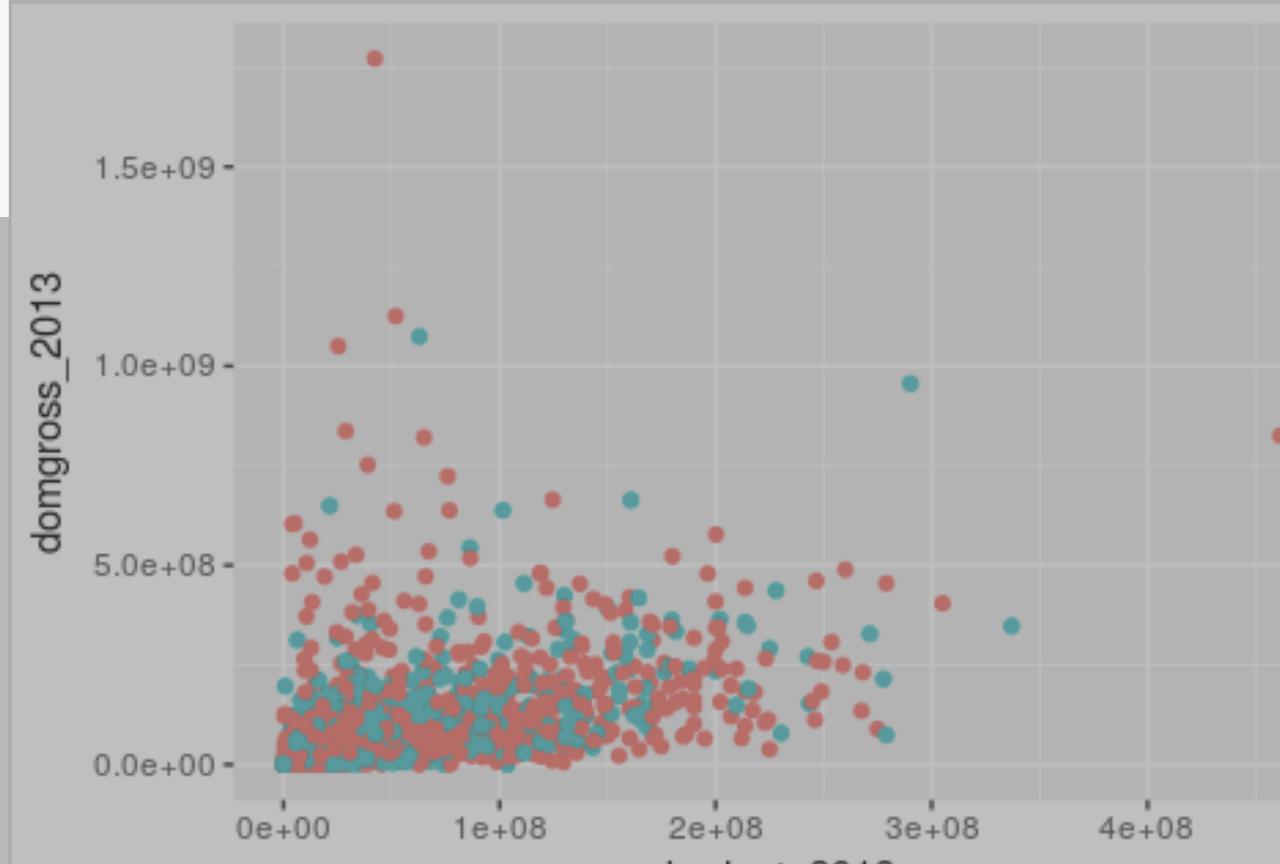
Environment is empty

Files Plots Packages Help Viewer

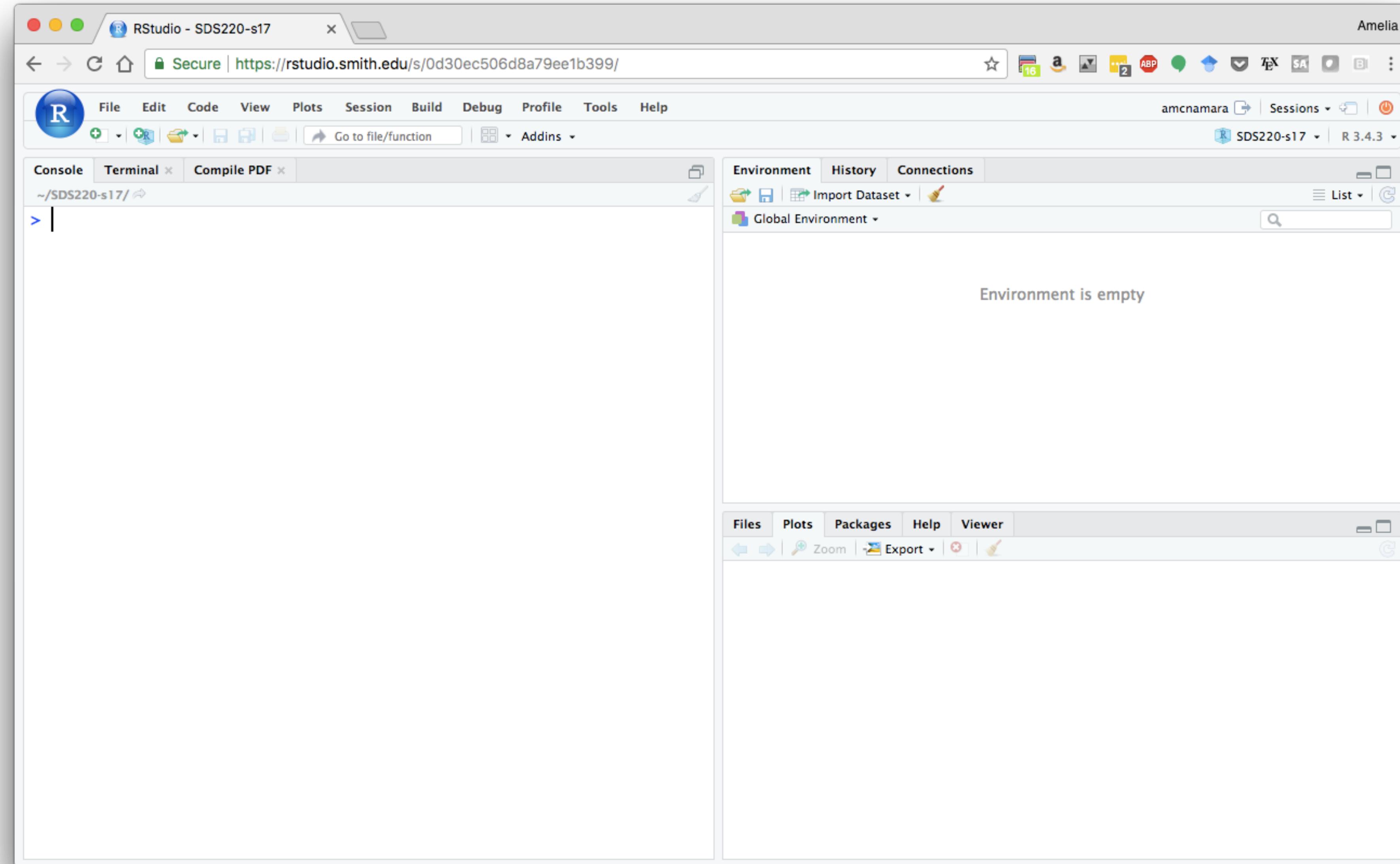
domgross_2013 budget_2013

binary

- FAIL
- PASS



RStudio

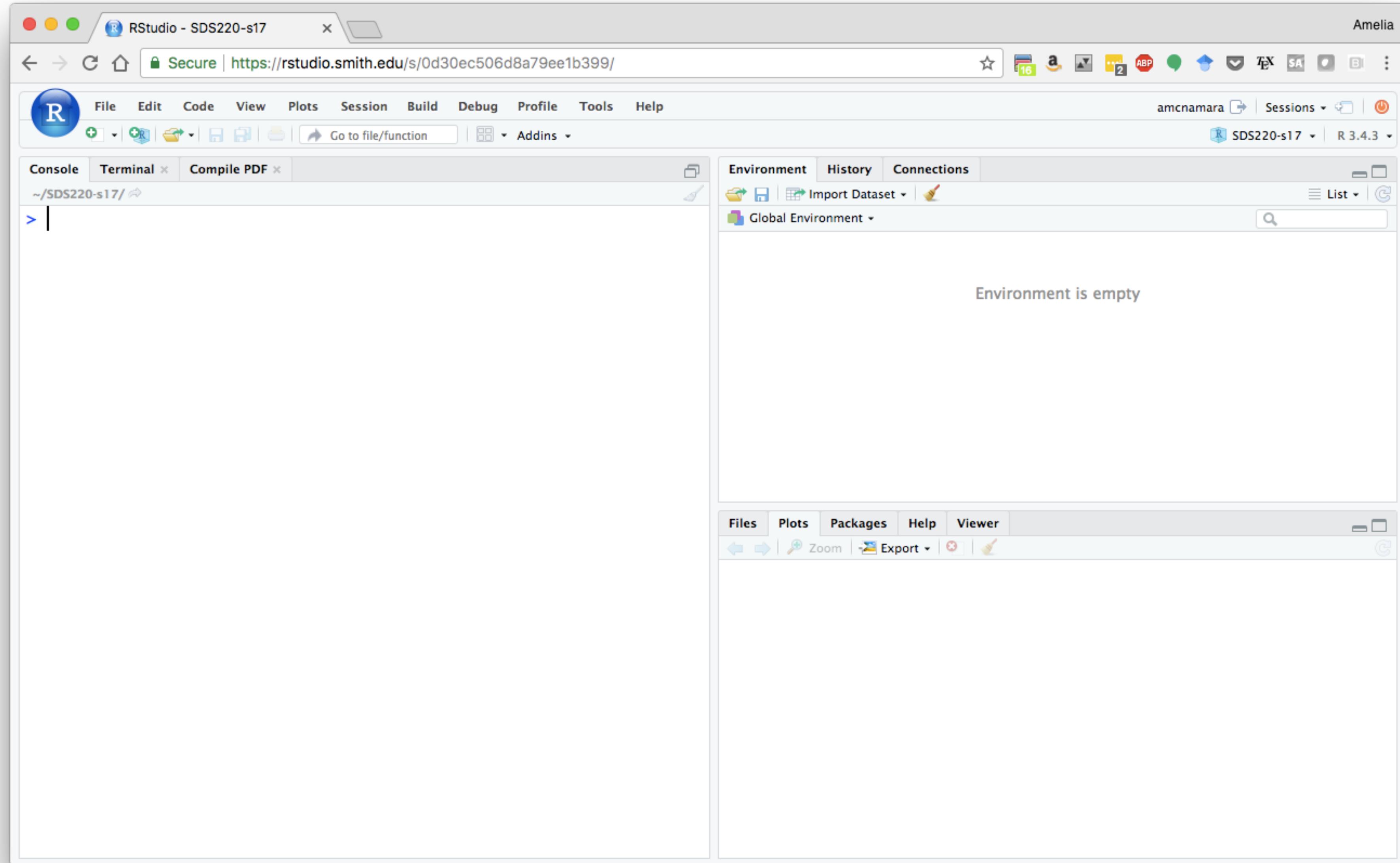


RStudio: ways to use

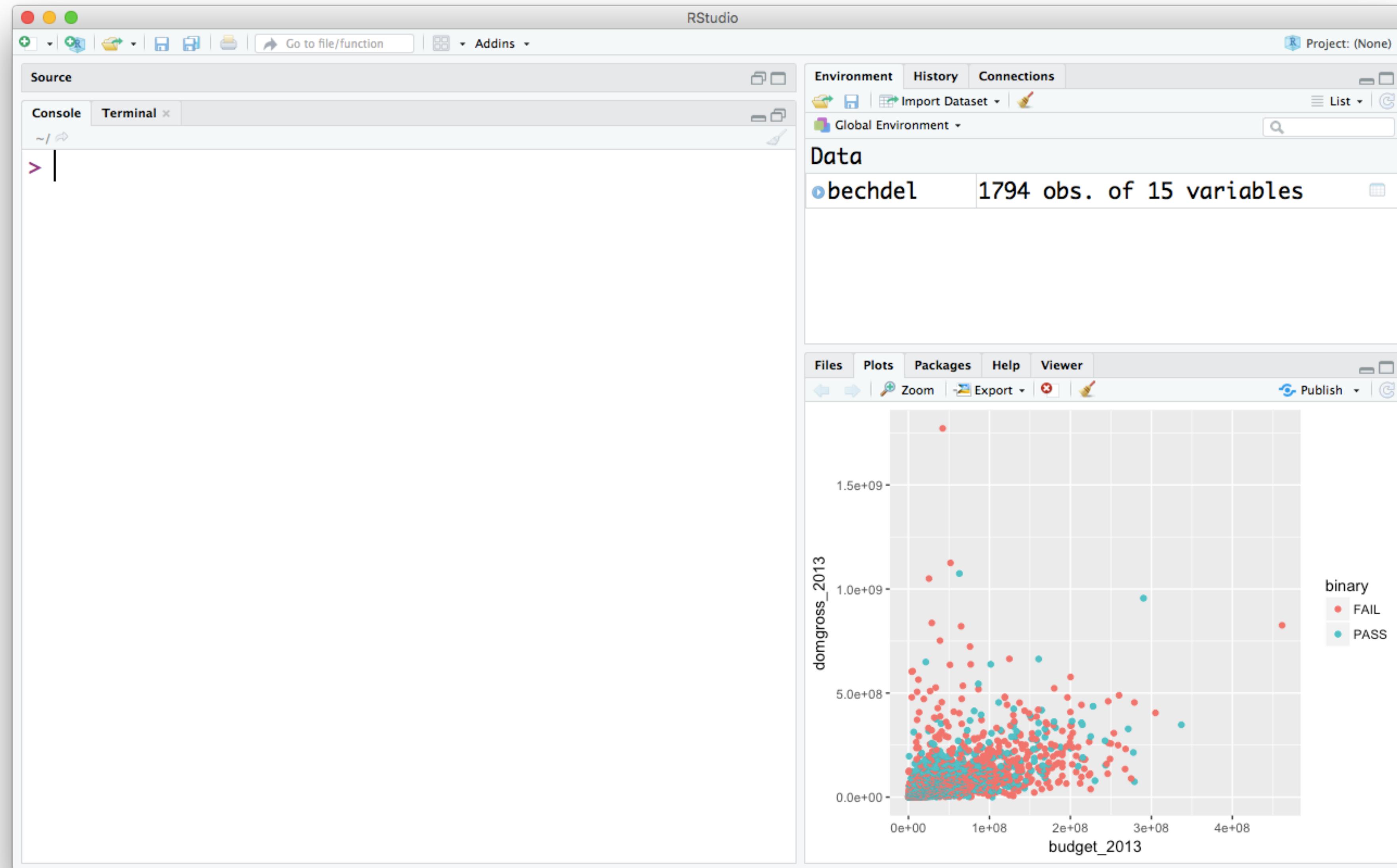
We're using RStudio Cloud, which allows you to log in through a web browser and do your work there.

But, there are other versions of RStudio.

RStudio: server edition



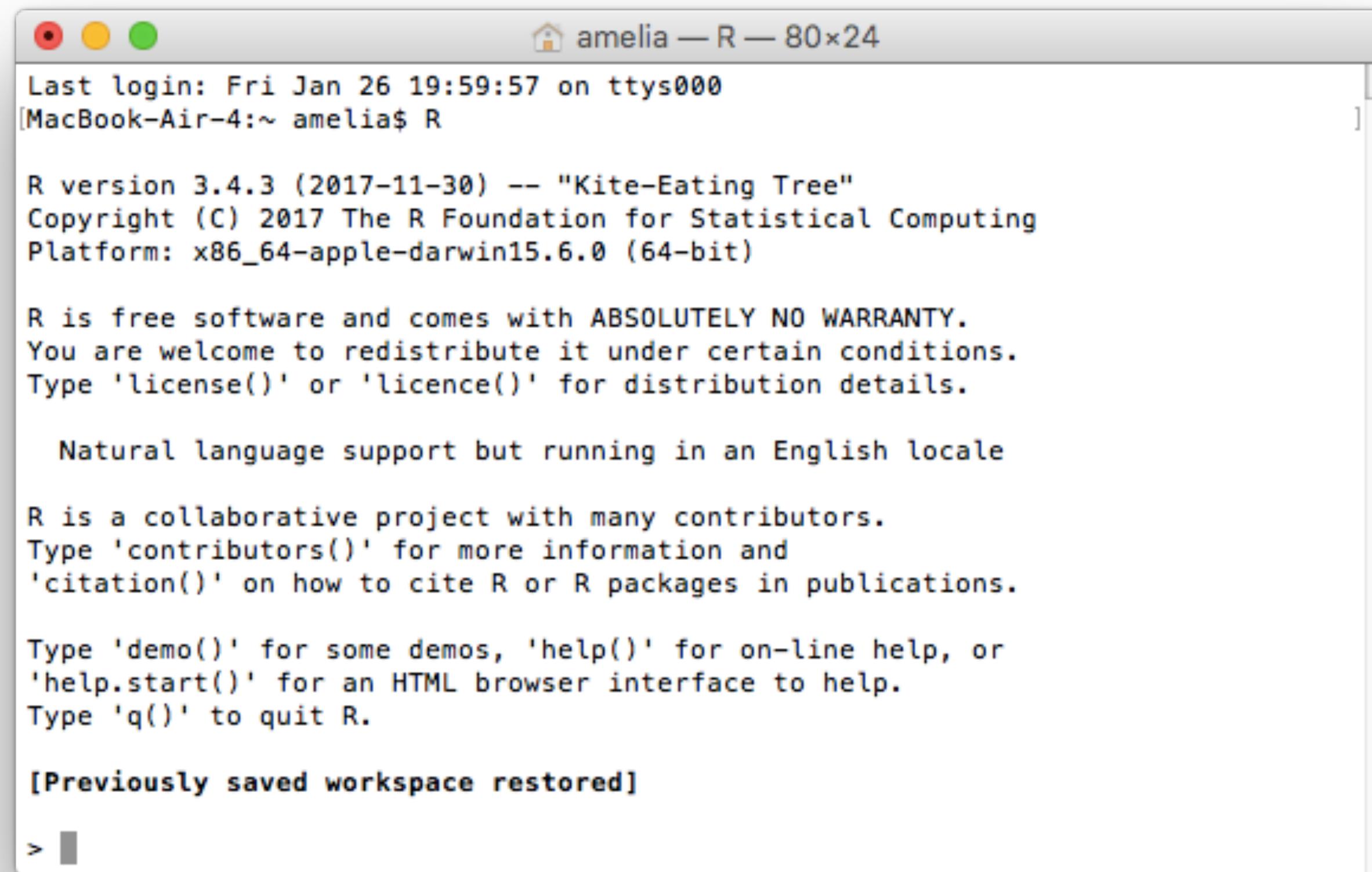
RStudio: desktop edition



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

Learning things can be frustrating!

Ask questions!

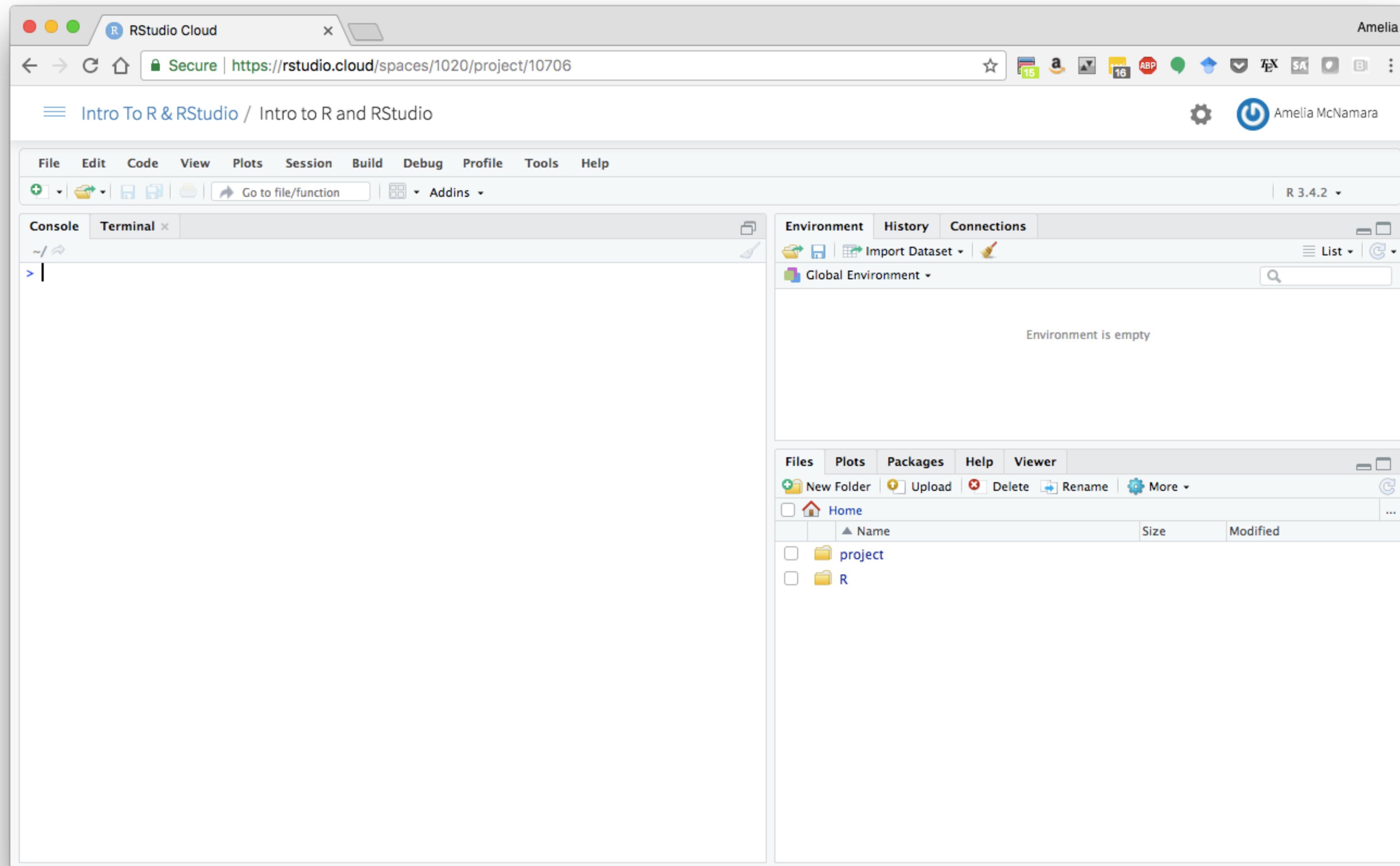
Practice!

Seriously: practice!

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

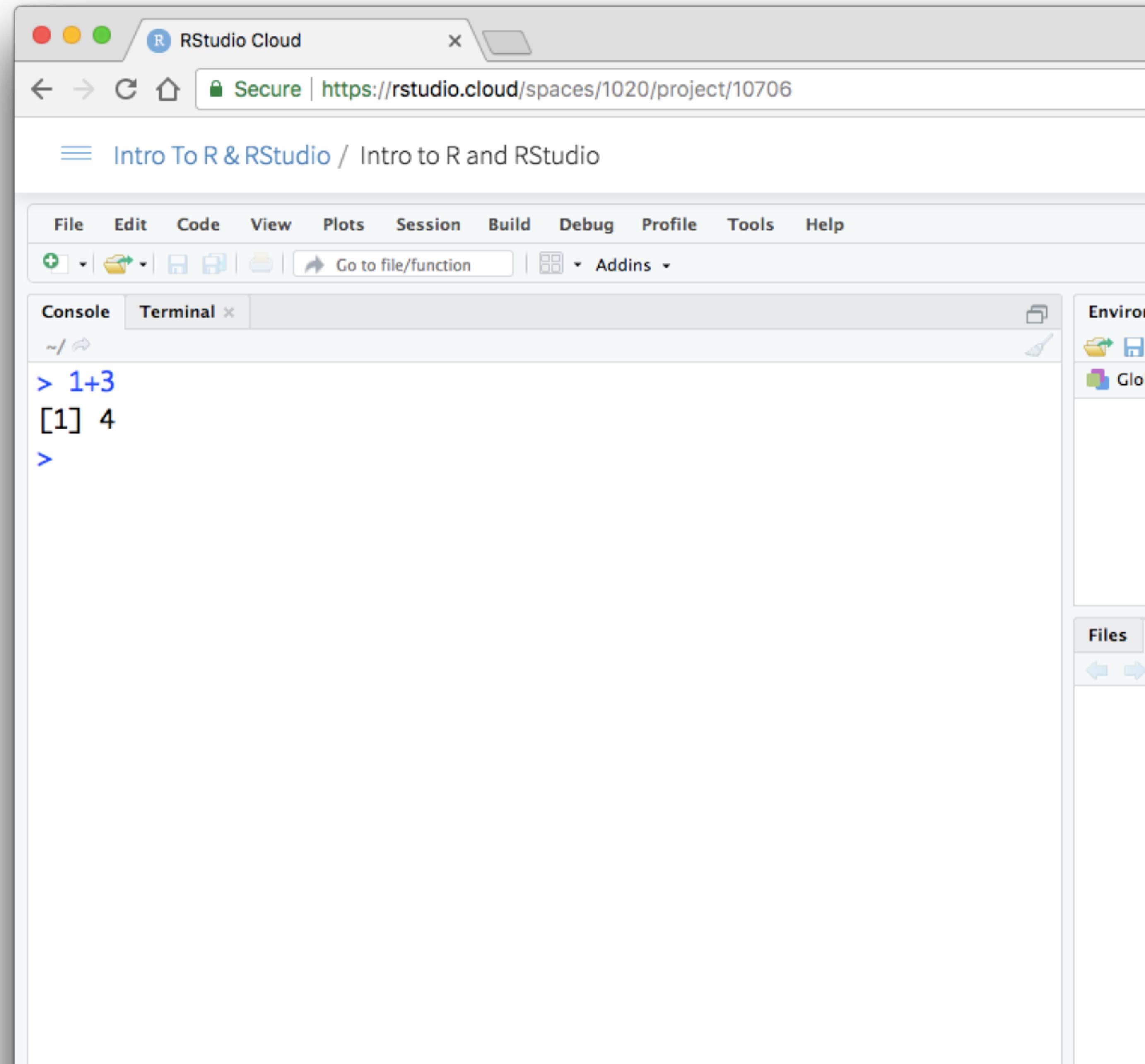
RStudio

Getting started



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)



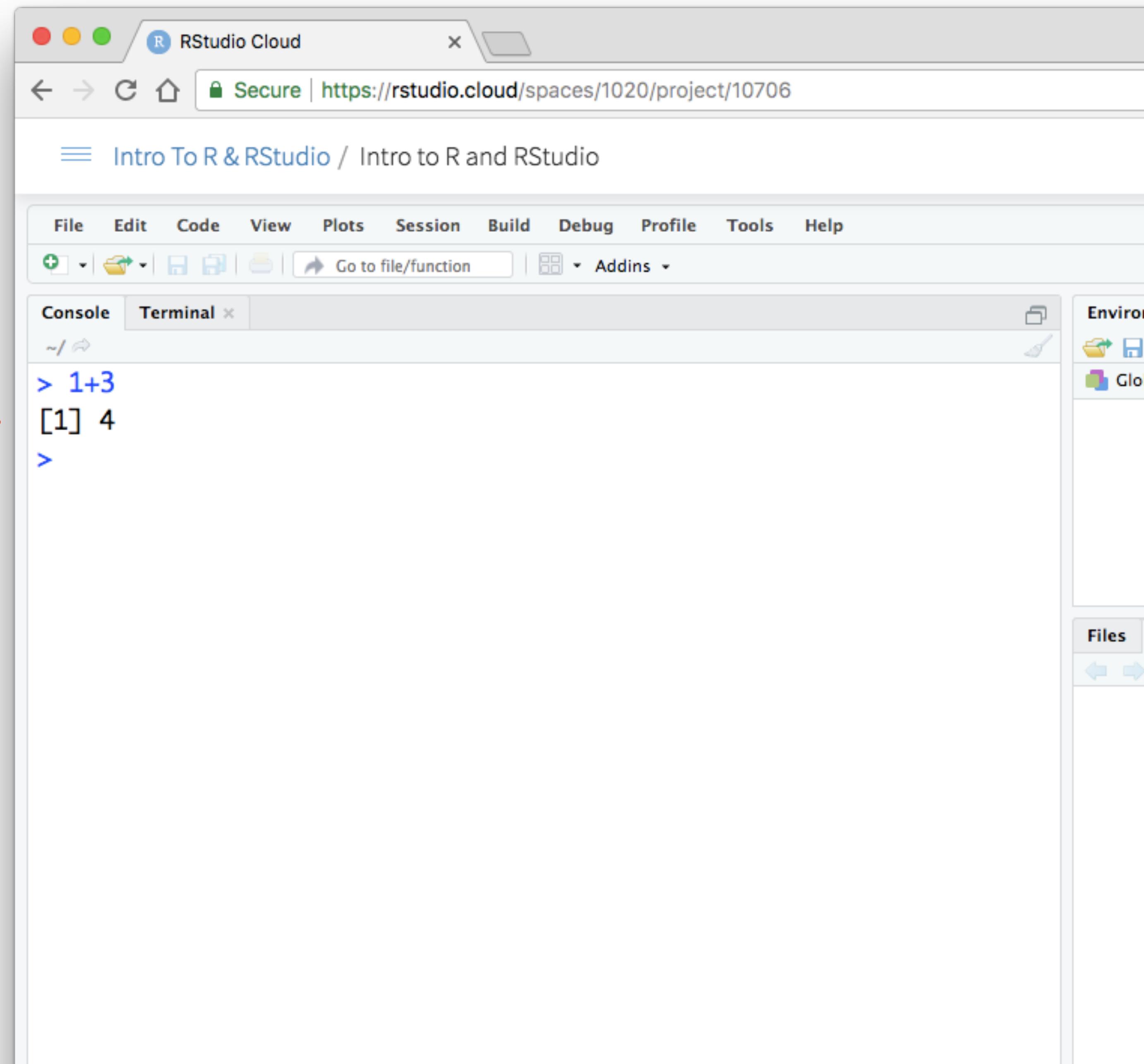
A screenshot of the RStudio Cloud interface. The title bar says "RStudio Cloud". The address bar shows a secure connection to "https://rstudio.cloud/spaces/1020/project/10706". The main area displays the "Intro To R & RStudio / Intro to R and RStudio" page. Below the header is a standard RStudio menu bar with File, Edit, Code, View, Plots, Session, Build, Debug, Profile, Tools, and Help. A toolbar follows with icons for file operations like New, Open, Save, and Addins. The central workspace shows a "Console" tab active, displaying the command "> 1+3" followed by the output "[1] 4". A second tab labeled "Terminal" is visible. On the right side, there are panels for "Environment", "Files", and "Global Options".

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

Output

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

Output →
New prompt →



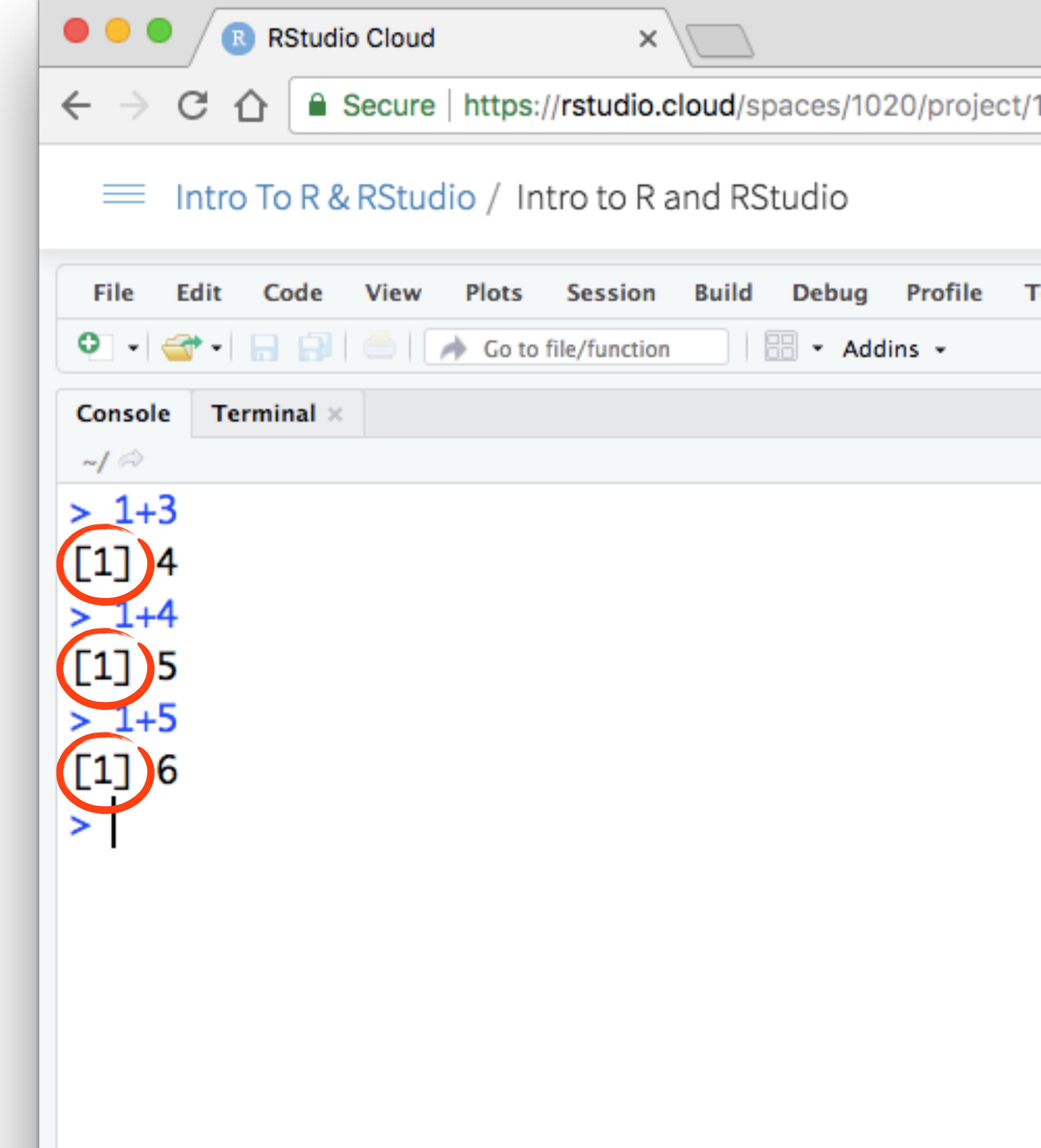
The screenshot shows the RStudio Cloud interface. The title bar says "RStudio Cloud". The address bar is "Secure | https://rstudio.cloud/spaces/1020/project/10706". The main area shows a "Console" tab with the following text:
> 1+3
[1] 4
>
The "Console" tab is highlighted in blue. To the right of the console, there are several tabs: "Console", "Terminal", "Plots", "Session", "Build", "Debug", "Profile", "Tools", and "Help". Below the tabs is a toolbar with icons for file operations like "New", "Open", "Save", etc. On the far right, there are vertical panels for "Environment", "Files", and "Global Options".

[1]

R displays an index next to the output.

Just ignore this.

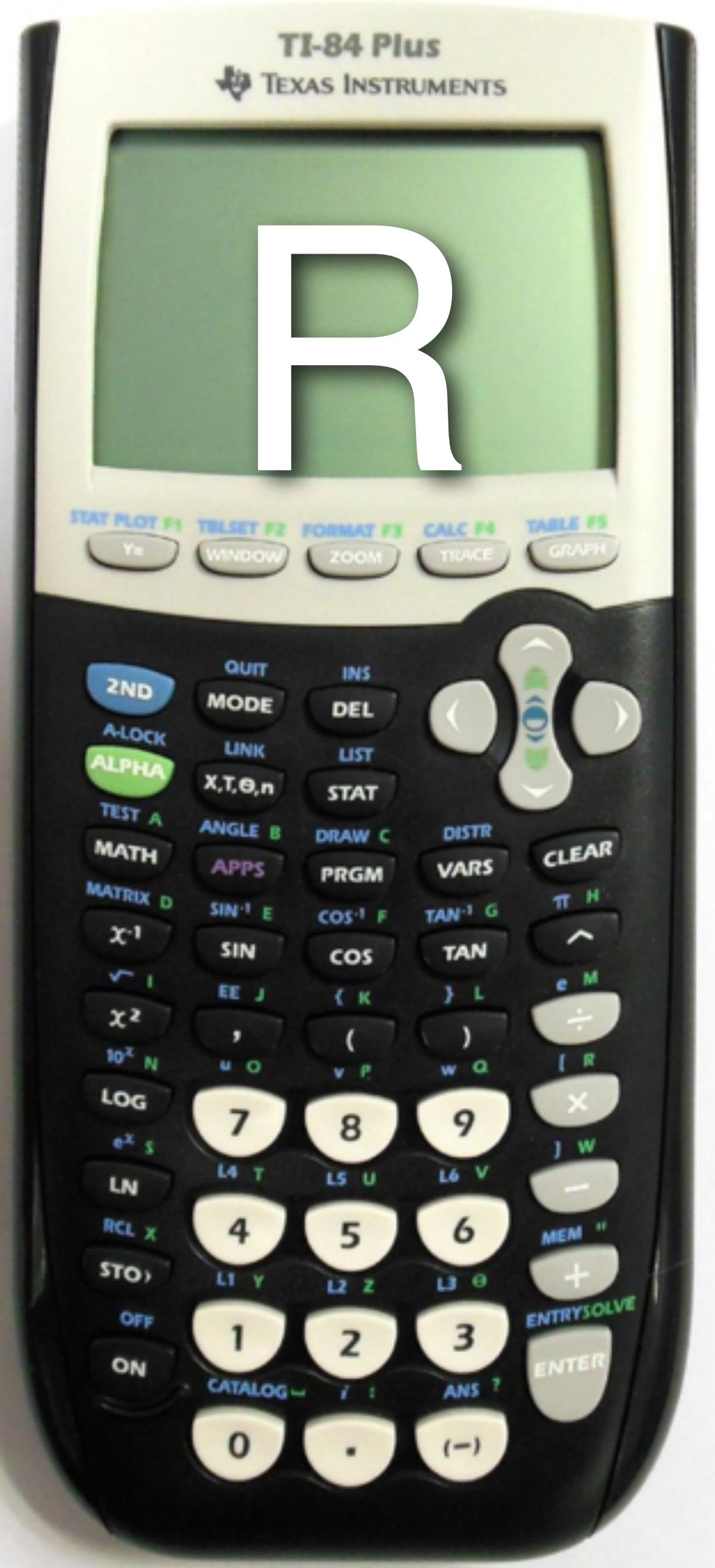
Somewhat helpful when R returns more than one value in the output.



The screenshot shows the RStudio Cloud interface with a secure connection. The title bar indicates "Intro To R & RStudio / Intro to R and RStudio". The menu bar includes File, Edit, Code, View, Plots, Session, Build, Debug, Profile, and Tools. Below the menu is a toolbar with various icons. The main area is the Console tab, which contains the following R session history:

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

The first three outputs, "[1] 4", "[1] 5", and "[1] 6", are circled in red, highlighting the index notation that is typically ignored when dealing with single values.



R is like a fancy calculator
on your computer

$5 + 5$

10

$4 - 1$

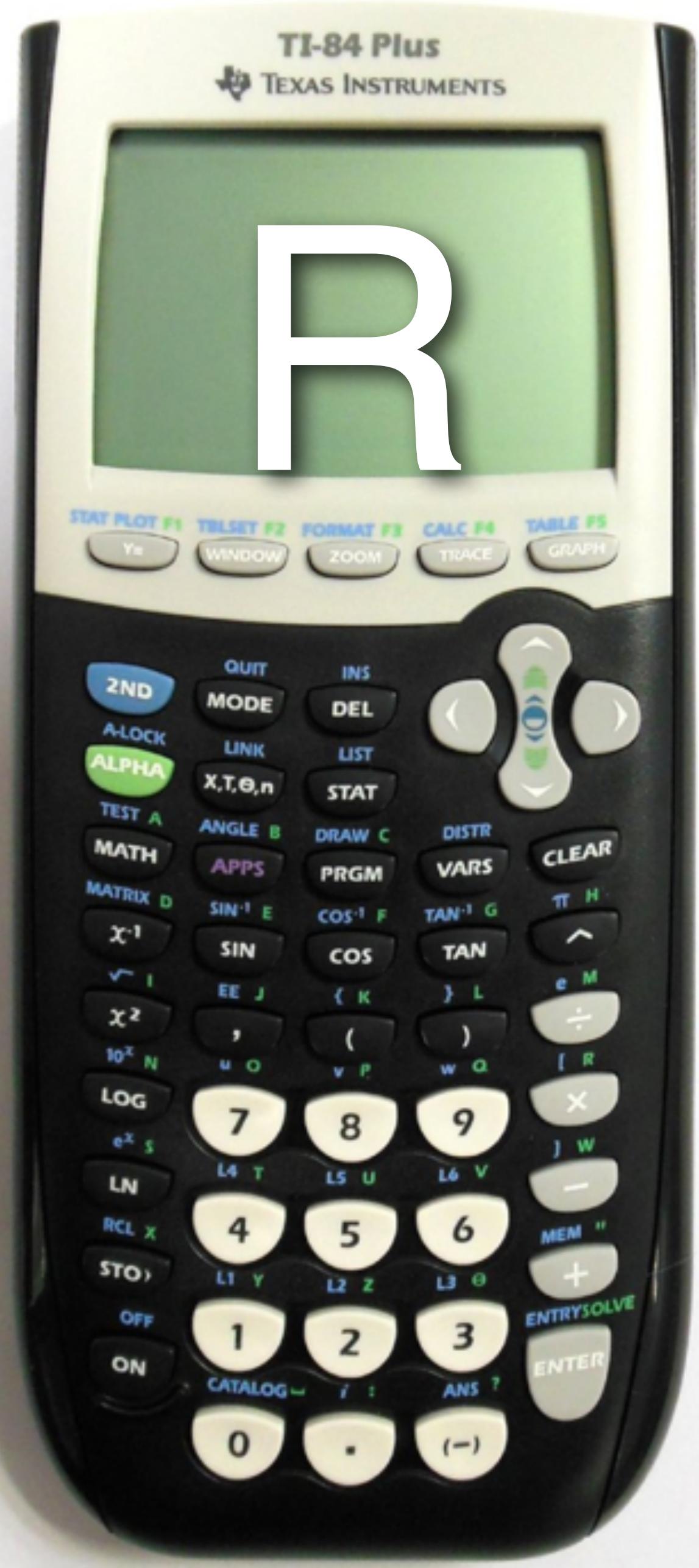
3

$1 * 2$

2

$4 ^ 2$

16



It can do algebra

$$a <- 1$$

$$b <- 2$$

$$a + b$$

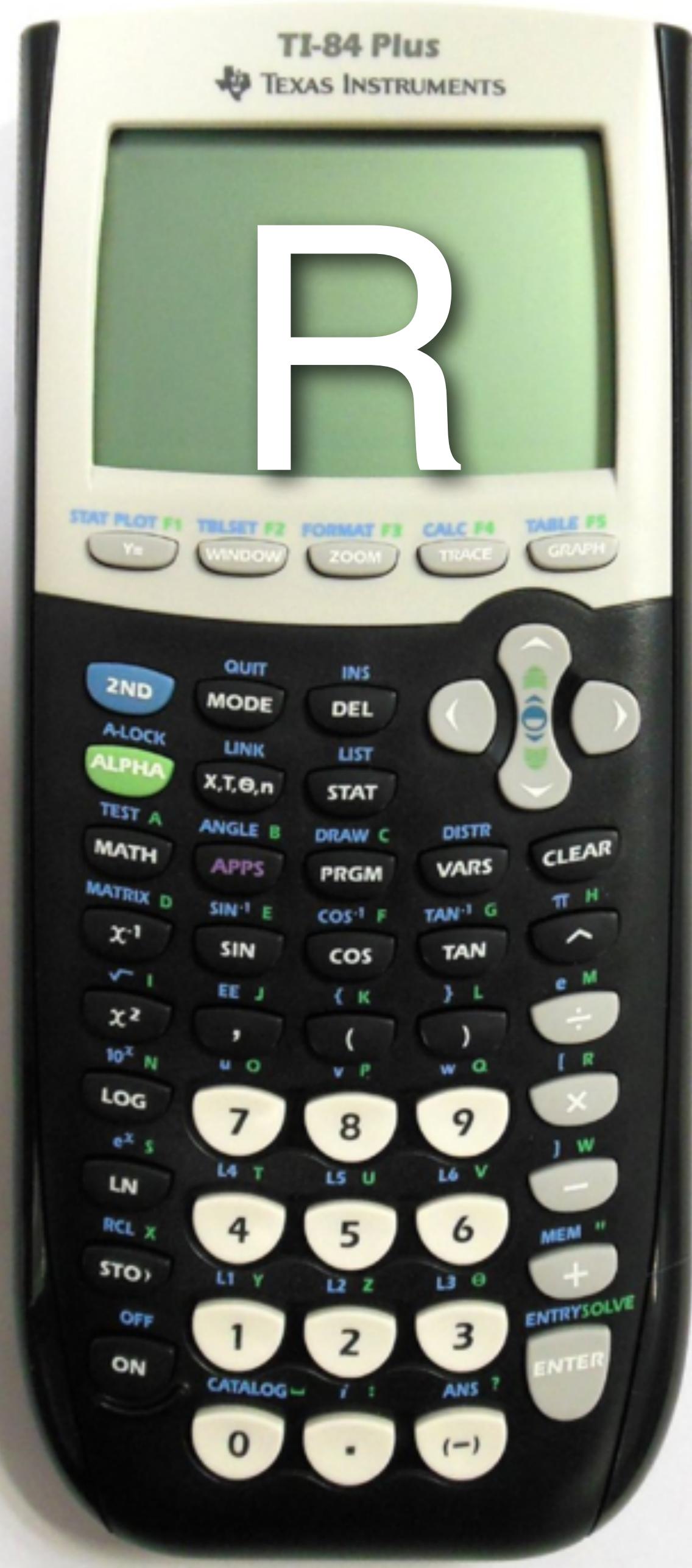
$$\# 3$$

$$A <- 3$$

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)

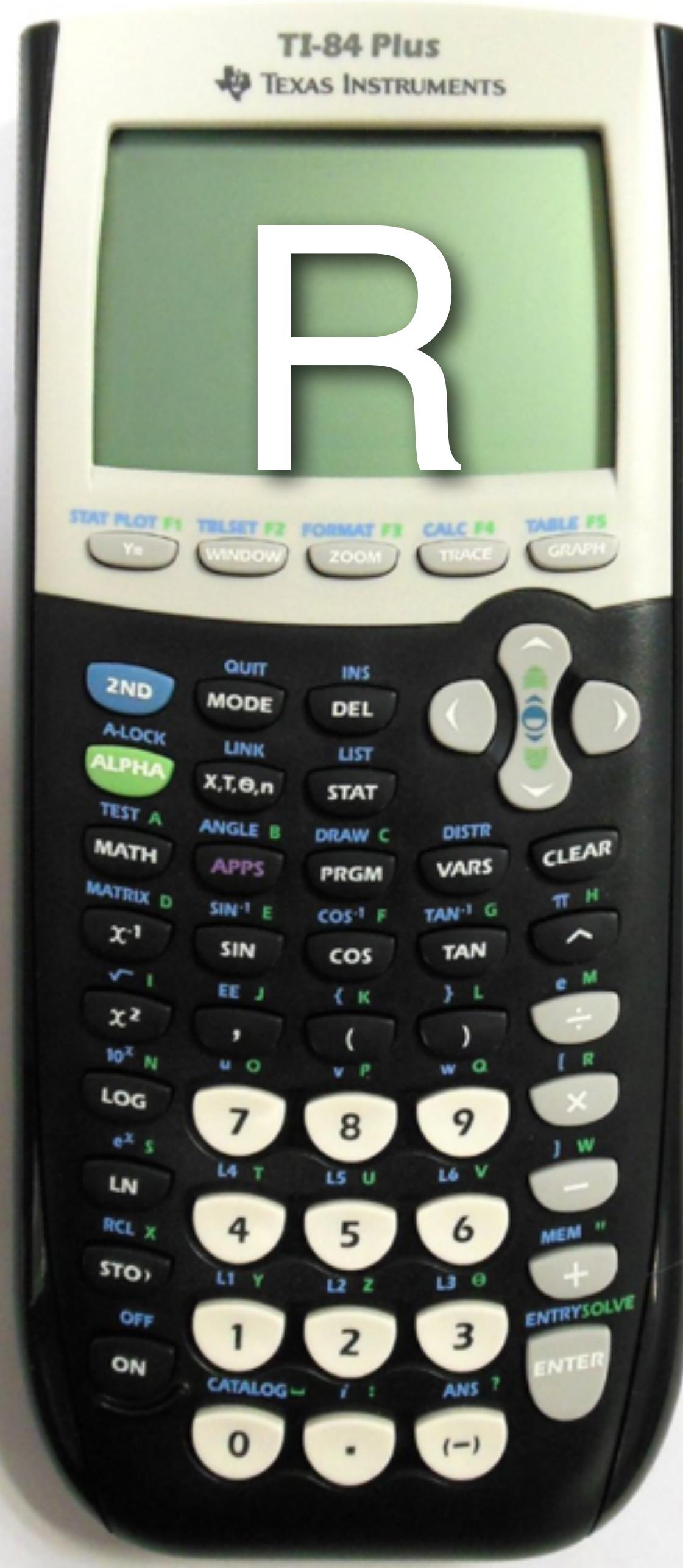
$$3! = 3 \times 2 \times 1$$

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

square root of 9
sqrt(9)
3

Your turn

What do you think this will return?

Your turn

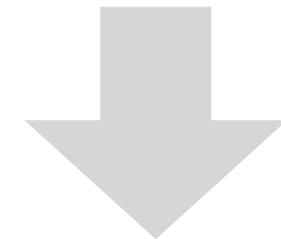
What do you think this will return?

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

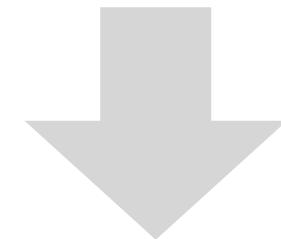
00 : 30

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

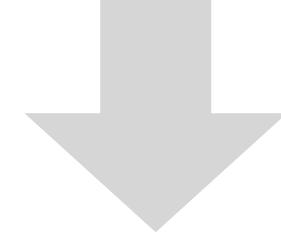
```
factorial(round(2.0015) + 1)
```



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



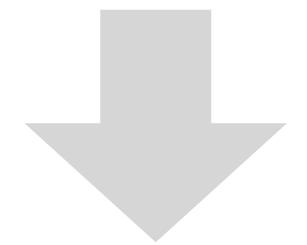
```
factorial(3)
```



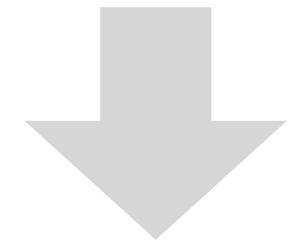
6

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

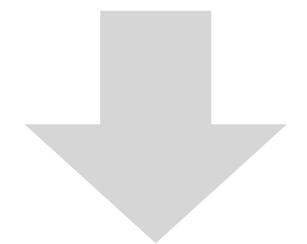
`factorial(round(2.0015) + 1)`



`factorial(2 + 1)`



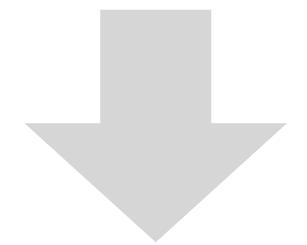
`factorial(3)`



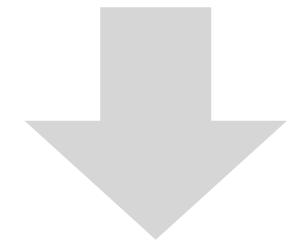
6

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

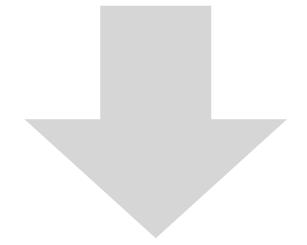
factorial(round(2.0015) + 1)



factorial(2 + 1)



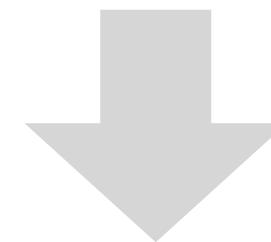
factorial(3)



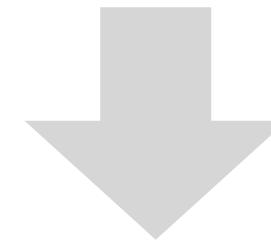
6

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

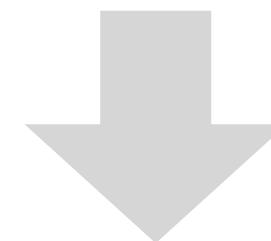
factorial(round(2.0015) + 1)



factorial(2 + 1)



factorial(3)



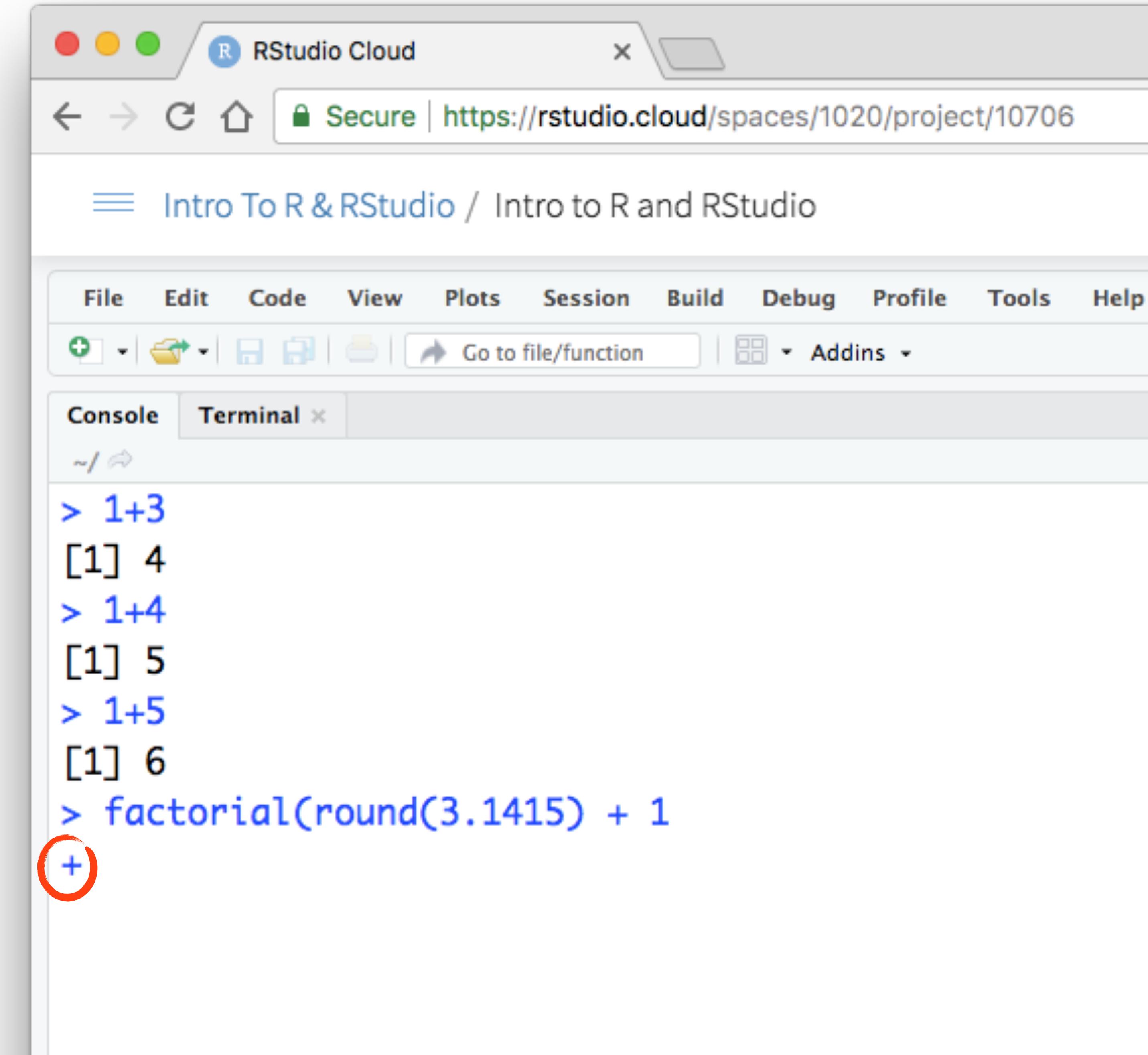
6

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

+ 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 web-based RStudio Cloud interface. The top bar includes the RStudio Cloud logo and a secure connection indicator. Below the header is a navigation bar with links to 'Intro To R & RStudio' and 'Intro to R and RStudio'. The main area is a code editor with tabs for 'Console' and 'Terminal'. The 'Console' tab is active, displaying the following R session:

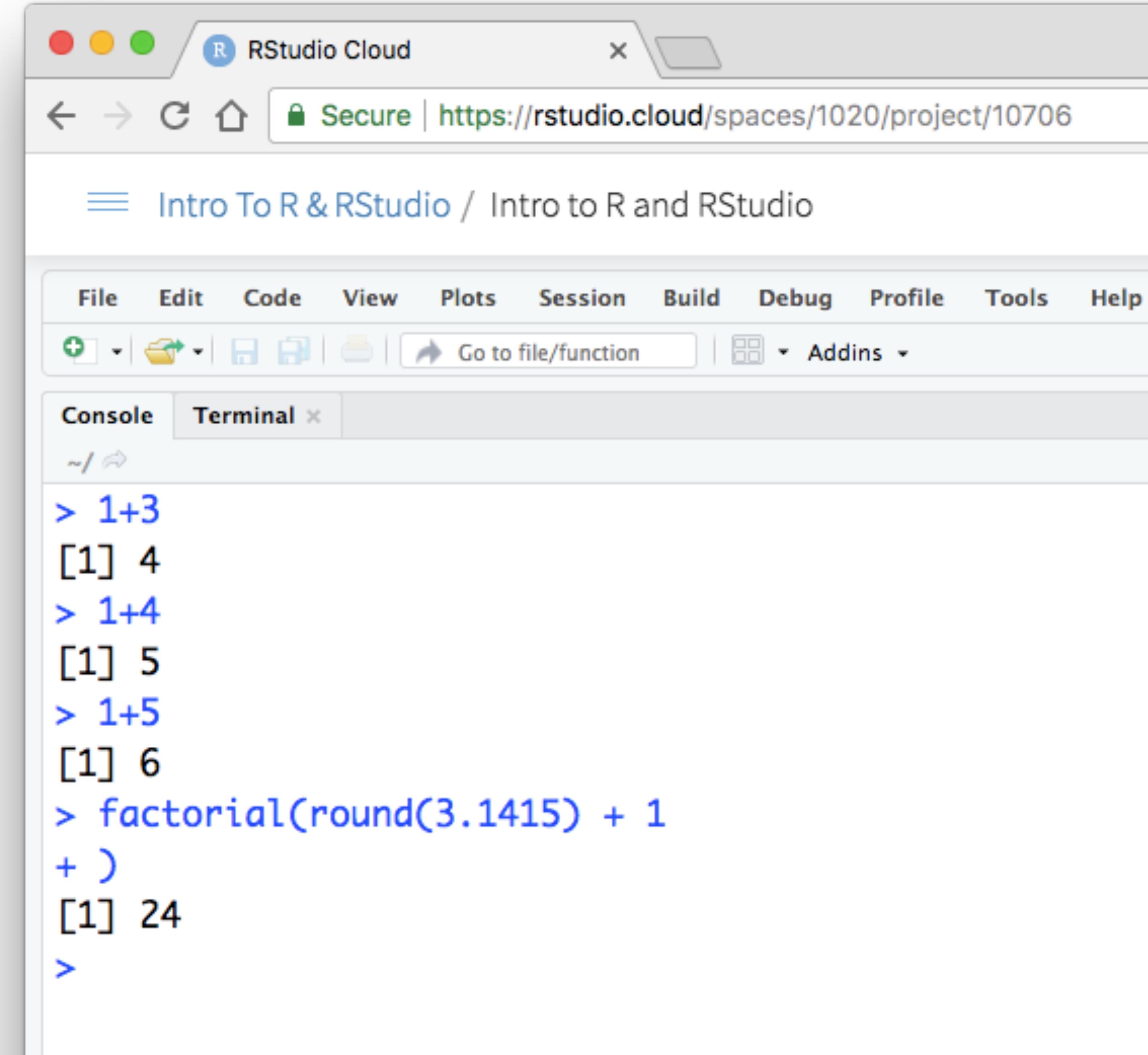
```
> 1+3  
[1] 4  
> 1+4  
[1] 5  
> 1+5  
[1] 6  
> factorial(round(3.1415) + 1  
+
```

The final line, '+', is circled in red at the bottom left of the console output.

+ 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 web-based RStudio Cloud interface. The title bar says "RStudio Cloud". The address bar is secure with "https://rstudio.cloud/spaces/1020/project/10706". The page 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 icons for new file, open file, save, and go to file/function. The "Console" tab is selected. The console output is as follows:

```
> 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

rmarkdown

It is easier to compose your code in an rmarkdown document than in the command line, and rmarkdown allows you to keep text with your code.

RStudio Cloud

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

Intro To R & RStudio / Intro to R and RStudio

File Edit Code View Plots Session Build Debug Profile Tools Help

01-Structures.Rmd x

```
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
```

Environment History Connections

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

1:1 R Notebook R Markdown

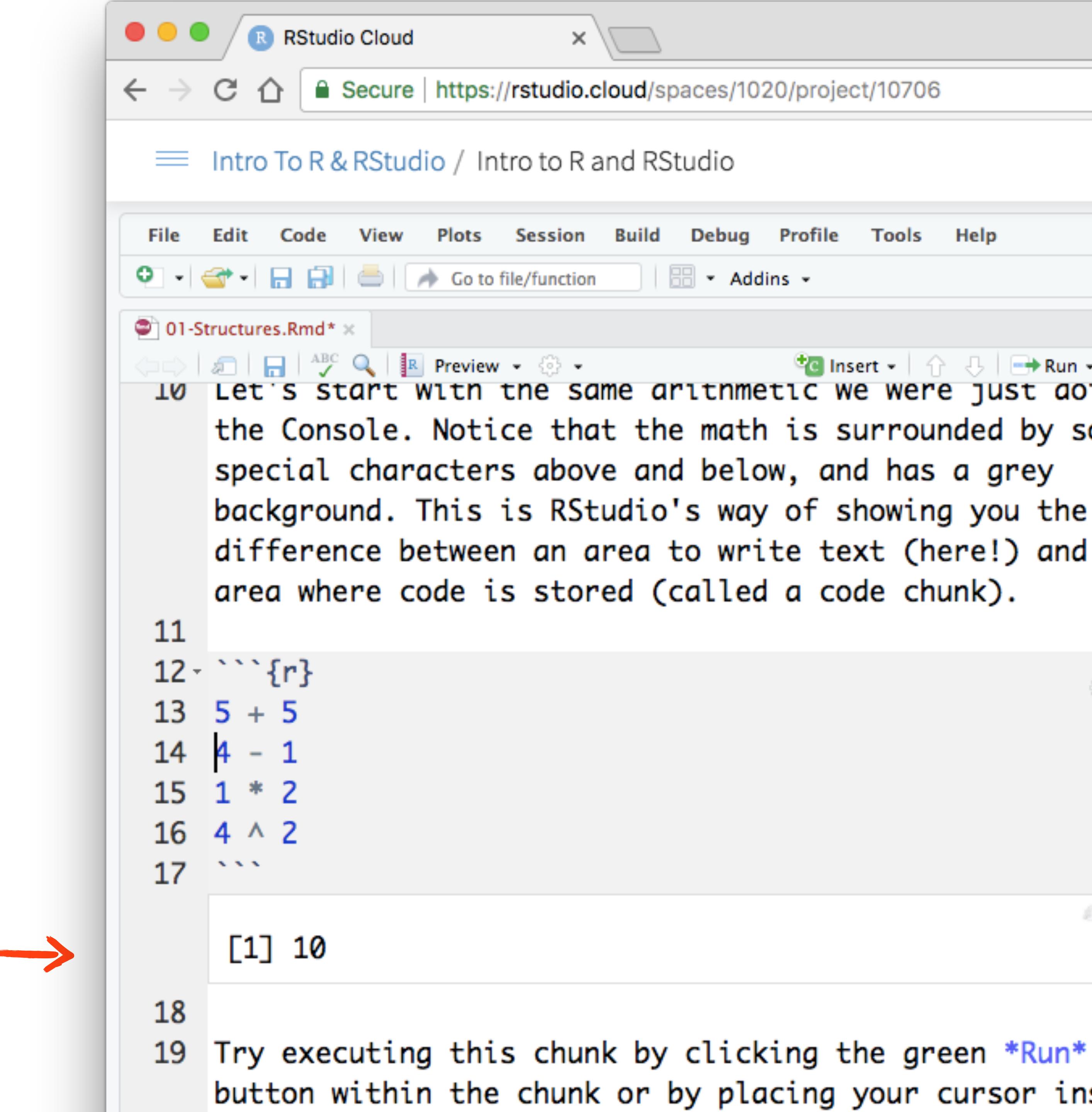
Console

The screenshot displays the RStudio Cloud interface. The main window shows an R Notebook titled "R Notebook". The code editor contains R code for setting up the notebook and performing arithmetic operations. A red box highlights the code editor area. The environment pane indicates that the global environment is empty. The file browser shows three R Markdown files in the "code" directory: "01-Structures.Rmd", "02-Syntax.Rmd", and "03-Programming.Rmd", all modified on Jan 27, 2018.

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.

The screenshot shows the RStudio Cloud interface. At the top, there's a toolbar with various icons. Below it is a header bar with the title "Intro To R & RStudio / Intro to R and RStudio" and a user profile for "Amelia McNamara". The main area contains a code editor with R Markdown syntax. A red box highlights the "Console" tab at the bottom left of the interface. In the background, there's a file browser showing files like "01-Structures.Rmd", "02-Syntax.Rmd", and "03-Programming.Rmd".

**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. The top bar includes the RStudio Cloud logo, a secure connection indicator, and the URL <https://rstudio.cloud/spaces/1020/project/10706>. The main area displays a file named "01-Structures.Rmd". A red arrow points from the text in the left block to the code chunk in the RStudio interface. The code chunk at line 10 contains the following R code:

```
10 Let's start with the same arithmetic we were just doing in the Console. Notice that the math is surrounded by 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 5 + 5  
13 4 - 1  
14 1 * 2  
15 4 ^ 2  
16  
17
```

The output of the code chunk is shown in the console at the bottom:

```
[1] 10
```

Line 19 contains instructions: "Try executing this chunk by clicking the green *Run* button within the chunk or by placing your cursor in...".

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
```

```
16 4 ^ 2
```

```
17 ````
```

```
[1] 10
```

18

19 Try executing this chunk by clicking the green ***Run*** button within the chunk or by placing your cursor inside it and pressing ***Cmd+Shift+Enter***

Global Environment

Files	Plots	Packages	Help
New Folder	Delete	R	
Home > Dropbox > Intro_to			
<hr/>			
Name			
			..
			.Rhistory
			04-Syntax.nb.html
			04-Syntax.Rmd
			03-DataTypes.nb.html
			03-DataTypes.Rmd
			02-Visualization.Rmd
			01-Intro.Rmd
			solutions

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  
16 4 ^ 2  
17 ...
```

You can also run just one line of code, by placing your cursor on the line and hitting Command Enter

```
[1] 10
```

Output for the one line you've run

18

19 Try executing this chunk by clicking the green *Run* button within the chunk or by placing your cursor inside it and pressing *Cmd+Shift+Enter*

Files Plots Packages Help

New Folder Delete R

Home > Dropbox > Intro_to

Name
..
.Rhistory
04-Syntax.nb.html
04-Syntax.Rmd
03-DataTypes.nb.html
03-DataTypes.Rmd
02-Visualization.Rmd
01-Intro.Rmd
solutions

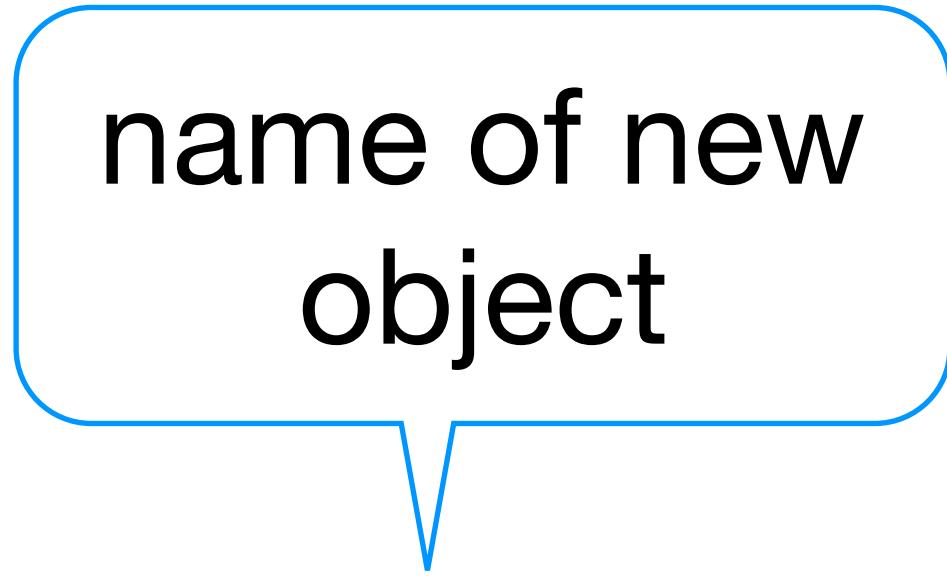
R objects

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 with the following details:

- Cloud** tab is active.
- Address Bar:** secure | <https://rstudio.cloud/spaces/1020/project/10706>
- Toolbar:** Includes icons for file operations, preview, and various tools.
- Session Header:** RStudio / Intro to R and RStudio, Amelia McNamara.
- Menu Bar:** View, Plots, Session, Build, Debug, Profile, Tools, Help.
- Environment Pane:** Shows the "Global Environment" tab. A red circle highlights the "Values" section, which contains the following data:

	Values
a	1
A	3
b	2
- Files Pane:** Shows a folder structure: Home > Day1 > code. It lists three files: 01-Structures.Rmd, 02-Syntax.Rmd, and 03-Programming.Rmd, all modified on Jan 27, 2018, at 5:19 PM.

What objects are in your environment right now?

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
```

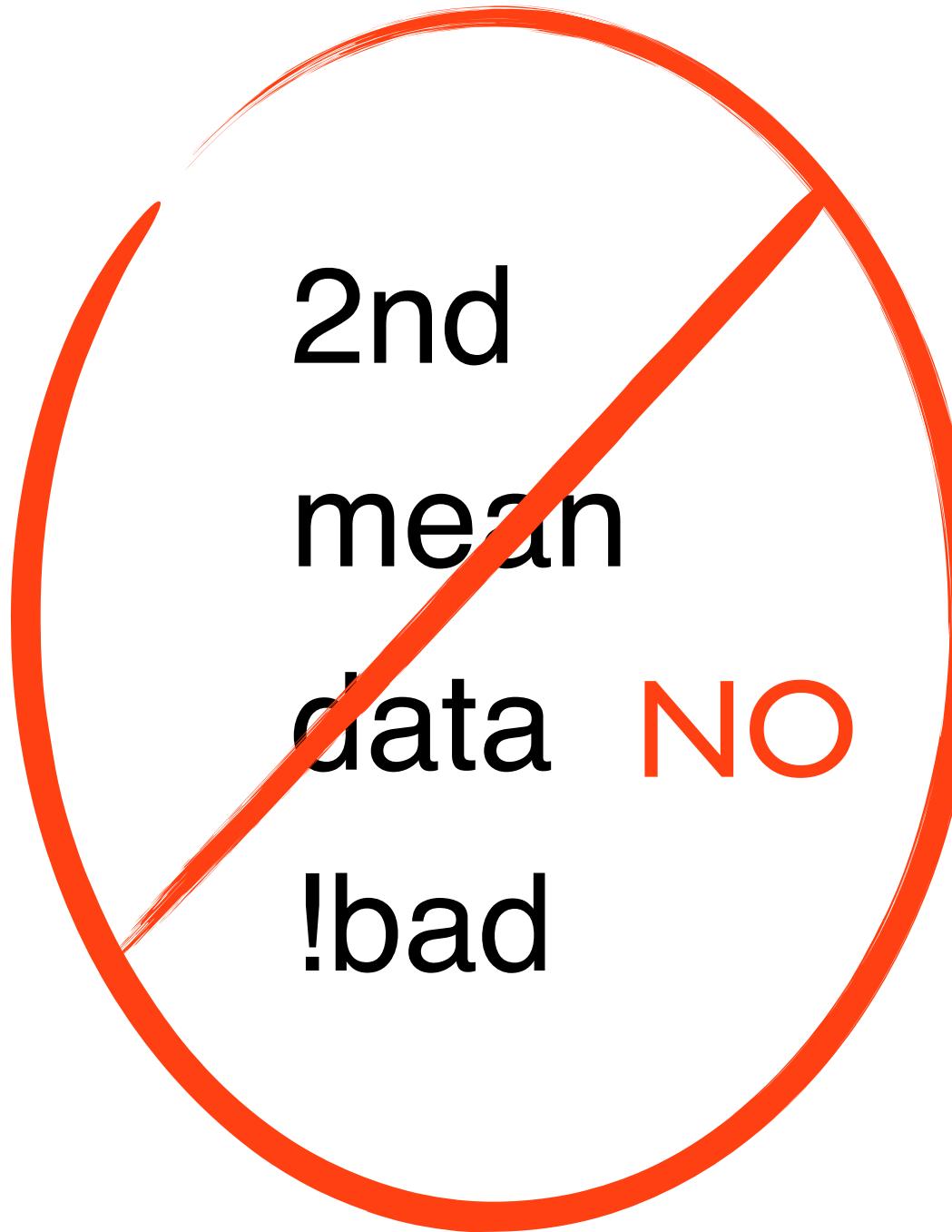
Object names

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	CDC_data
b	finalModel
FOO	more_pi yeah!
my_var	withoutOver64
.day	

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
```

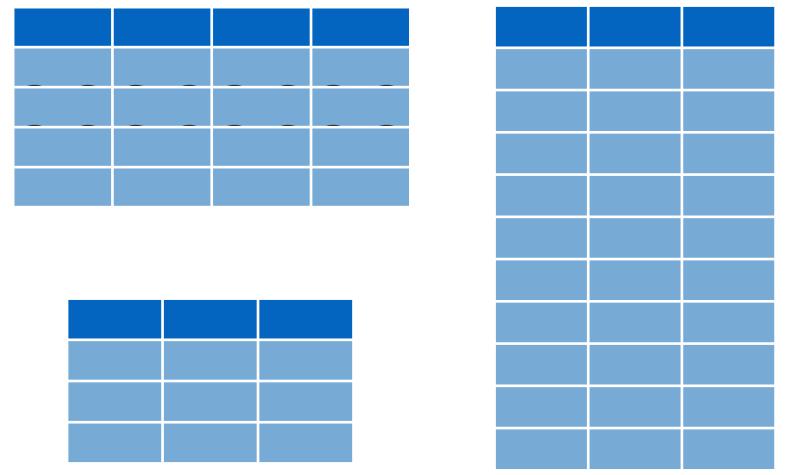
```
# 1
```

```
rm(pi)
```

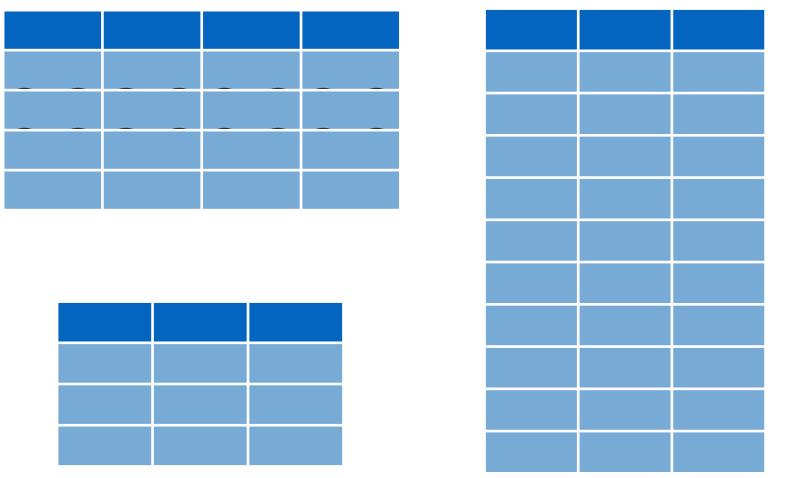
```
pi
```

```
# 3.141593
```

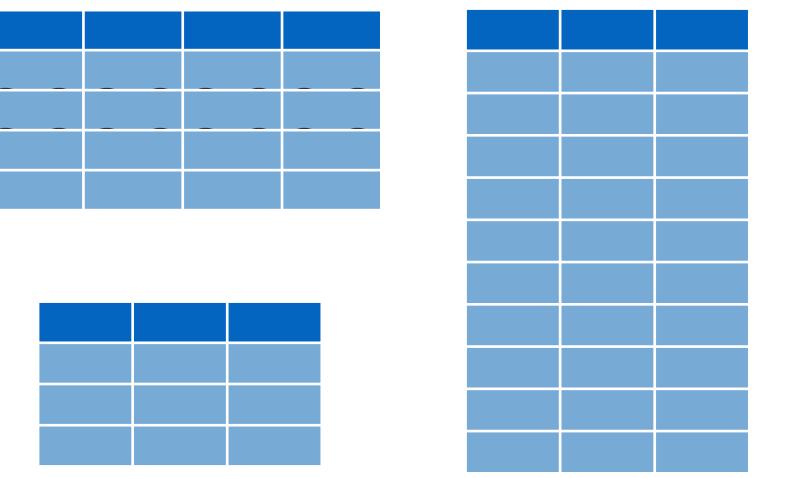
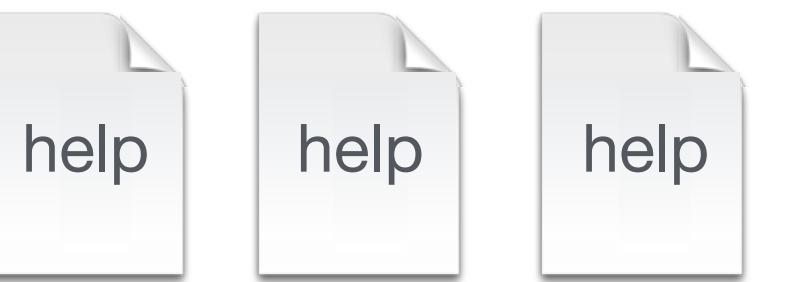
R packages



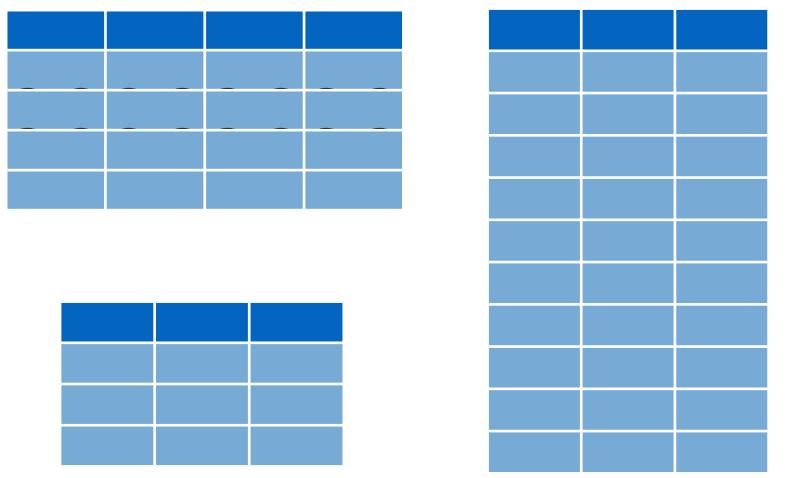
function1()
function2()
function3()
function4()



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



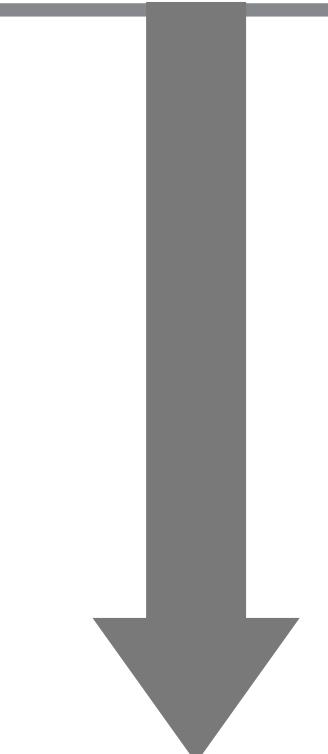
function5()
function6()
function7()
function8()



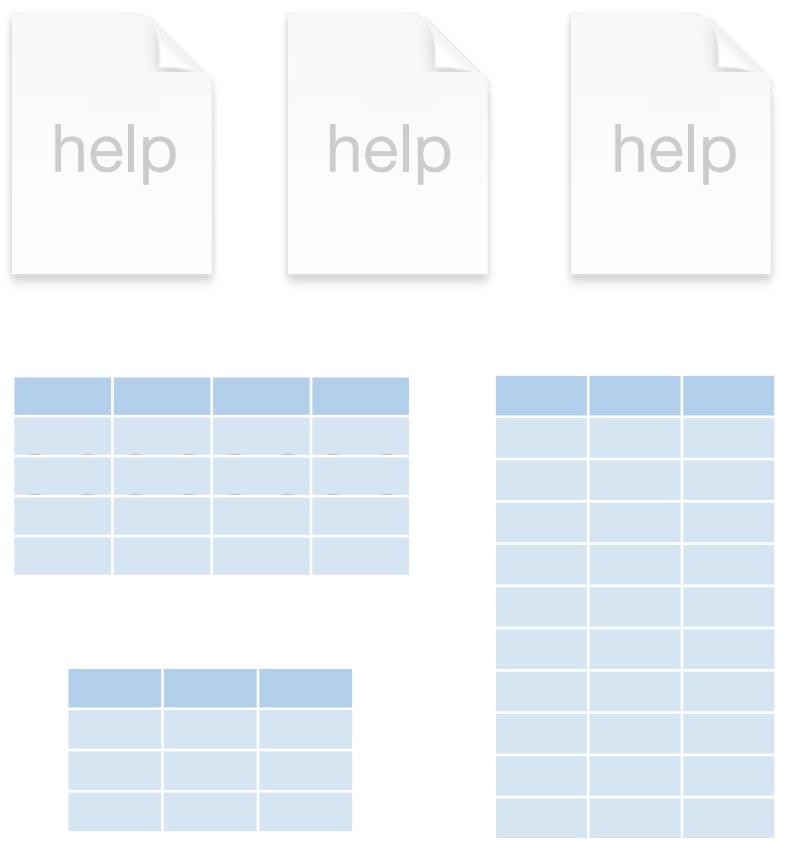
function9()
functionA()
functionB()
functionC()



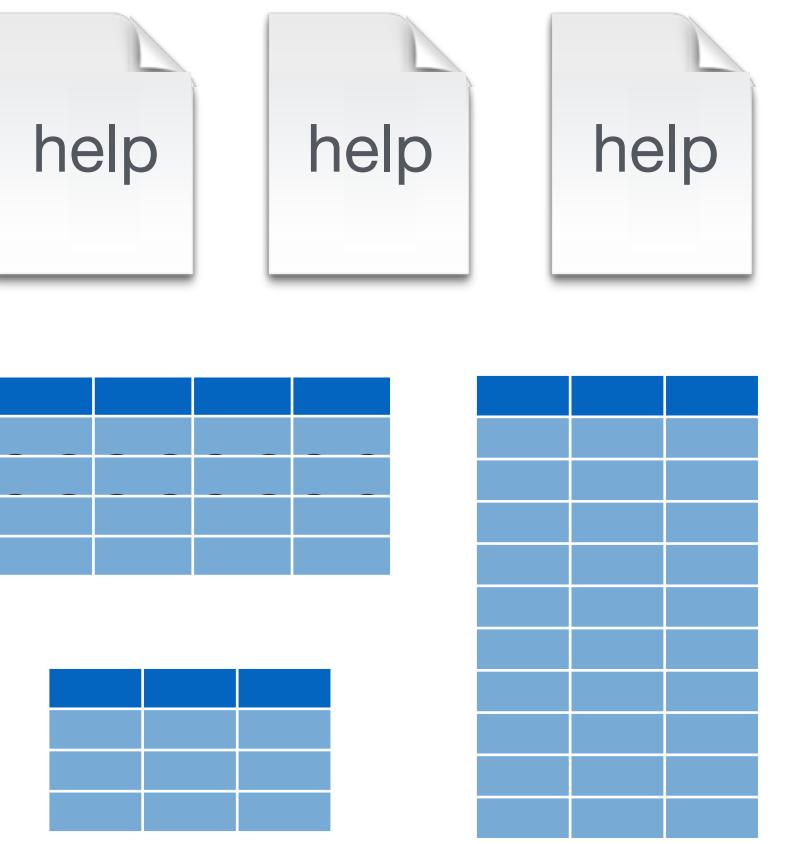
functionD()
functionE()
functionF()
functionG()



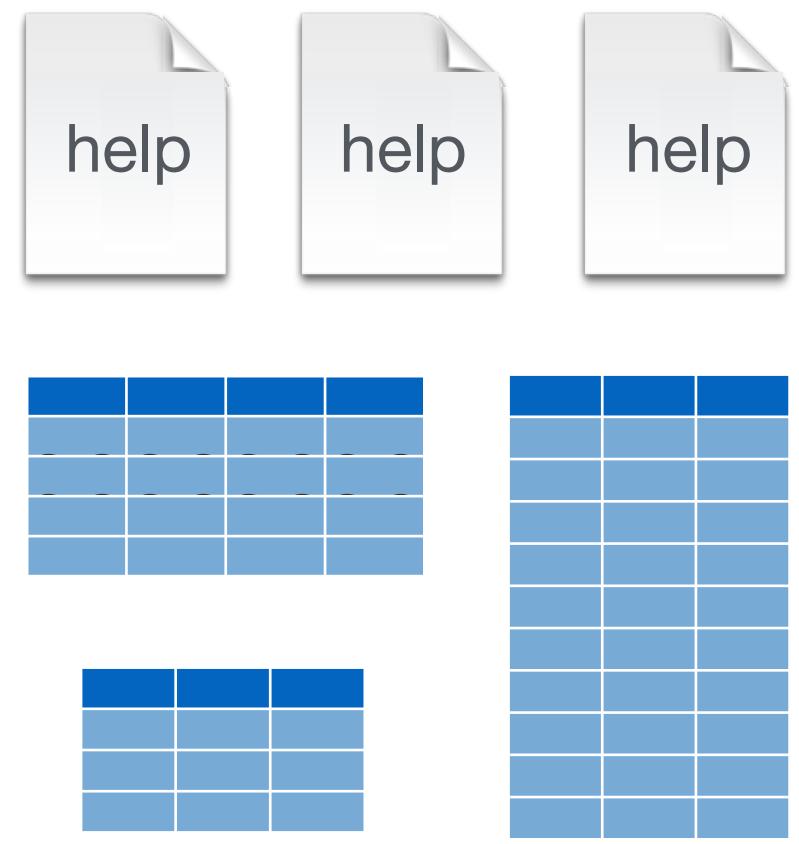
Base R



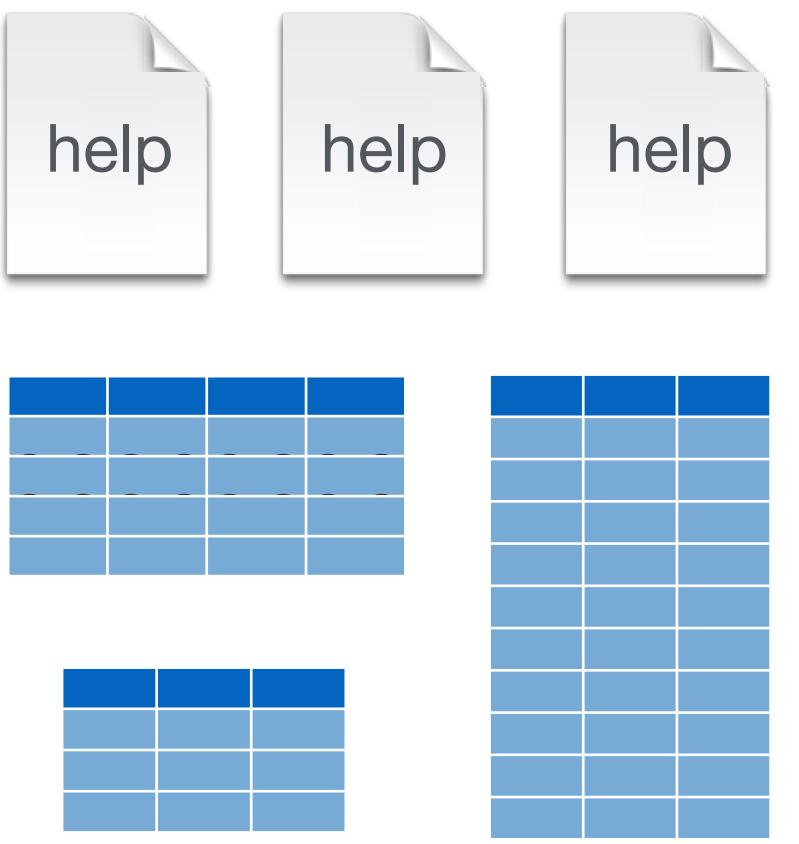
function1()
function2()
function3()
function4()



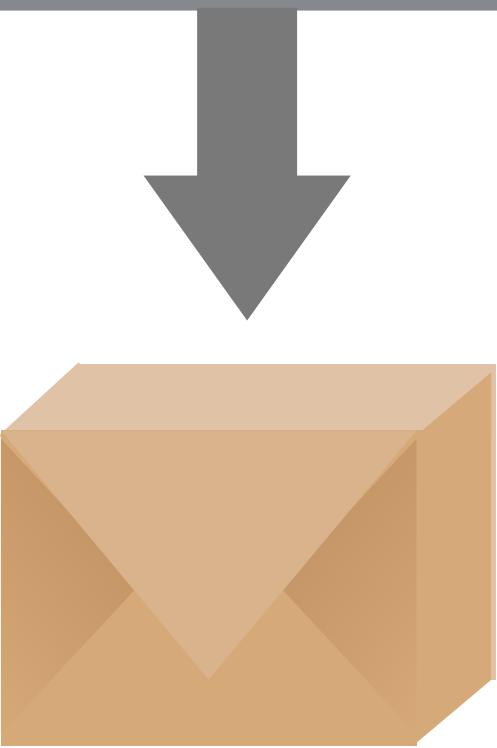
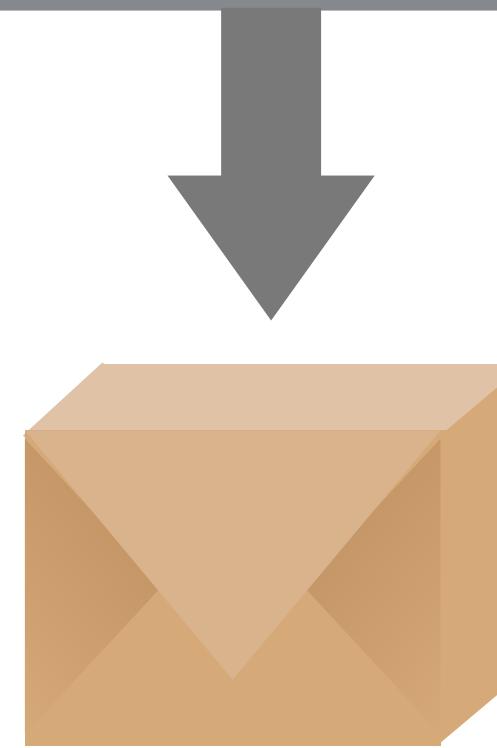
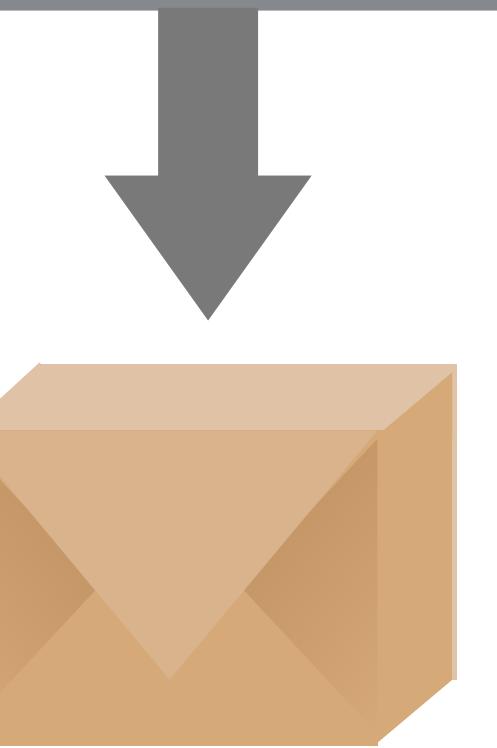
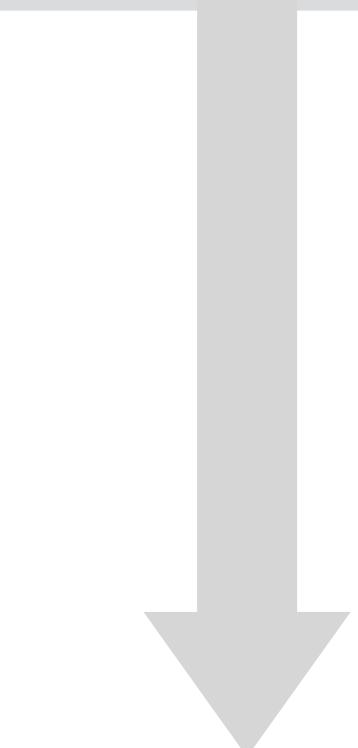
function5()
function6()
function7()
function8()



function9()
functionA()
functionB()
functionC()



functionD()
functionE()
functionF()
functionG()



Base R

R Packages

The Comprehensive R Archive

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



CRAN

[Mirrors](#)

[What's new?](#)

[Task Views](#)

[Search](#)

About R

[R Homepage](#)

[The R Journal](#)

Software

[R Sources](#)

[R Binaries](#)

[Packages](#)

[Other](#)

Documentation

[Manuals](#)

[FAQs](#)

[Contributed](#)

[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

[accelerometry](#) Functions for Processing Minute-to-Minute Accelerometer Data

[acelmissing](#) 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.

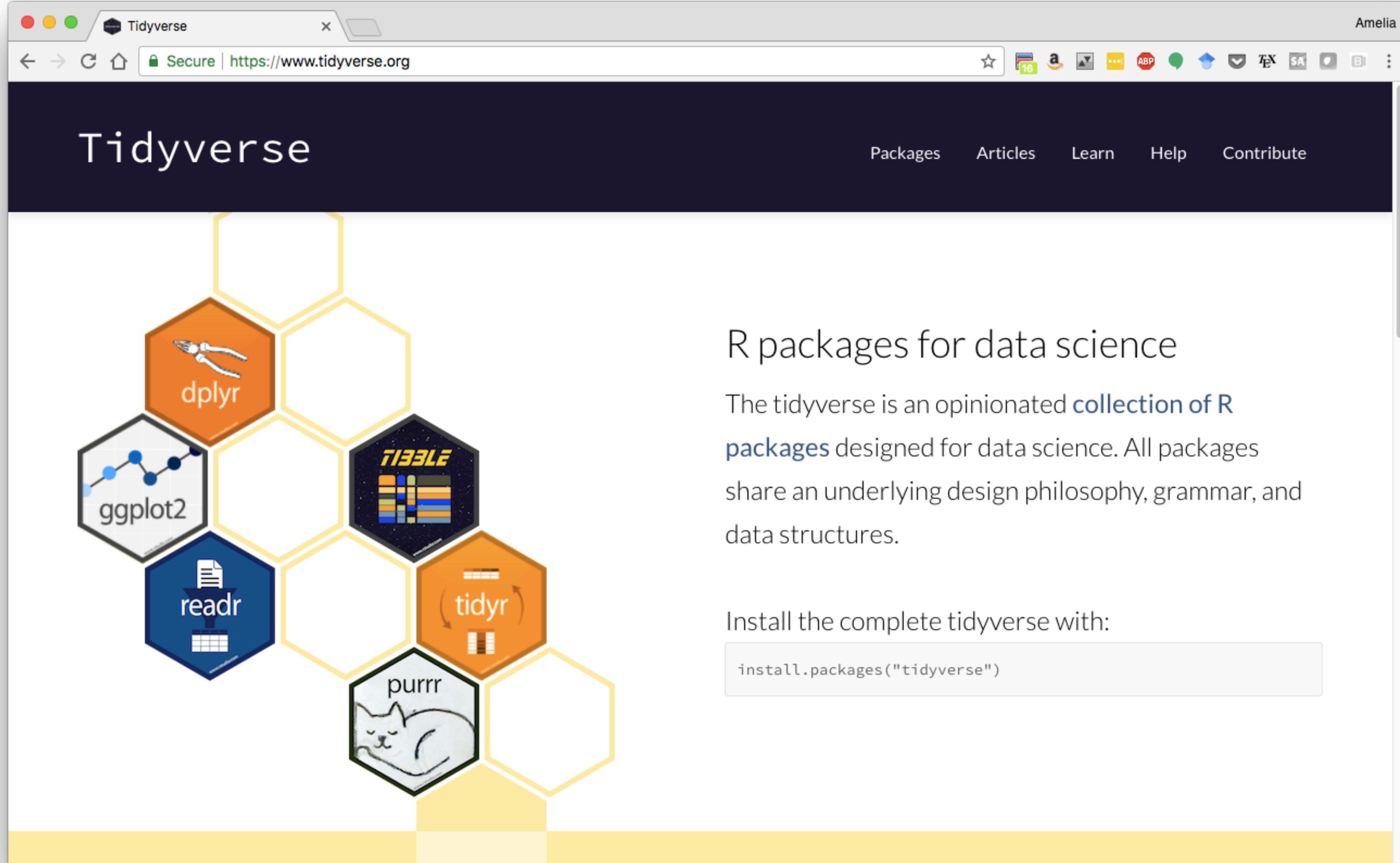
```
library(ggplot2)
```

Loads package

1 x per R Session

I've done this
for you

The tidyverse

A screenshot of a web browser displaying the Tidyverse website at <https://www.tidyverse.org>. The page features a dark header with the word "Tidyverse" and navigation links for "Packages", "Articles", "Learn", "Help", and "Contribute". Below the header is a large graphic composed of yellow hexagons containing icons for various R packages: dplyr (orange hexagon with pliers), ggplot2 (grey hexagon with a line plot), readr (blue hexagon with a file icon), purrr (grey hexagon with a cat icon), tibble (dark blue hexagon with a grid icon), and tidyr (orange hexagon with a circular arrow icon). To the right of the graphic, the text "R packages for data science" is displayed, followed by a detailed description of the tidyverse as an opinionated collection of R packages designed for data science. A code snippet for installing the tidyverse is shown in a box: `install.packages("tidyverse")`.

Tidyverse

Packages Articles Learn Help Contribute

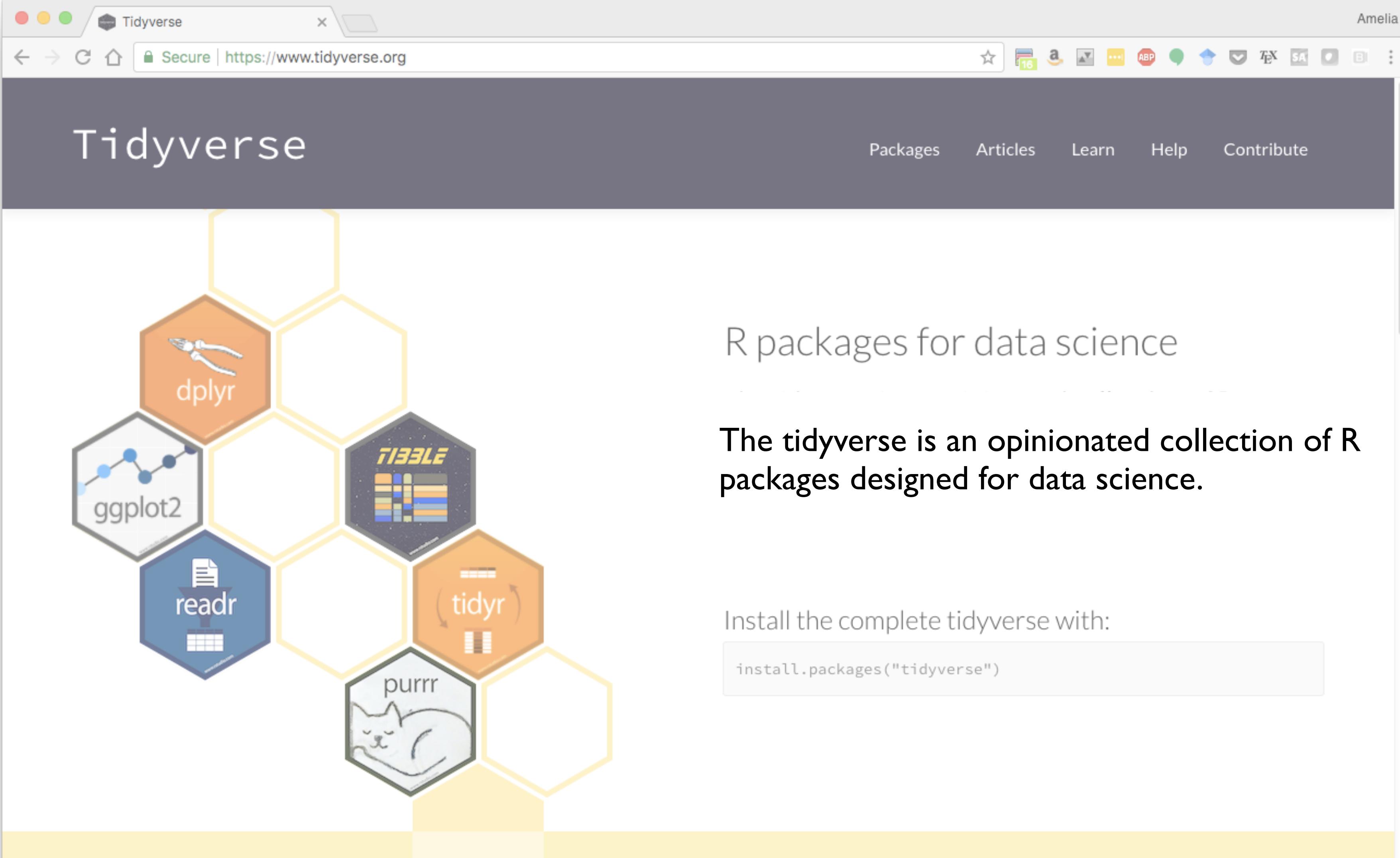


R packages for data science

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")
```

The image shows a screenshot of a web browser displaying the official Tidyverse website at <https://www.tidyverse.org>. The page has a dark grey header with the word "Tidyverse" in white. Below the header is a navigation bar with links for "Packages", "Articles", "Learn", "Help", and "Contribute". The main content area features a large graphic of six hexagons arranged in a hexagonal pattern, each containing a different R package icon: dplyr (orange, with a pair of pliers), ggplot2 (grey, with a line plot), readr (blue, with a file icon), purrr (grey, with a cat icon), tibble (dark grey, with a grid icon), and tidyr (orange, with a circular arrow icon). To the right of the graphic, the text "R packages for data science" is displayed. Below this, a bold statement reads "The tidyverse is an opinionated collection of R packages designed for data science." Further down, instructions for installing the tidyverse are provided: "Install the complete tidyverse with:" followed by the code "install.packages("tidyverse")".

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$efficient[mtcars$mpg>30] <- TRUE`
`mtcars$efficient[mtcars$mpg<30] <- FALSE`

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, data=mtcars)`
`lattice::bwplot(~disp, data=mtcars)`

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

two continuous variables:
`lattice::xyplot(mpg~disp, data=mtcars)`

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

one continuous, one categorical:
`lattice::histogram(~disp|cyl, data=mtcars)`
`lattice::bwplot(cyl~disp, data=mtcars)`

The variety of R syntaxes give
you many ways to “say” the
same thing

read across the cheatsheet to see how different
syntaxes approach the same problem

Tidyverse syntax

```
data %>% goal(x)
```

SUMMARY STATISTICS:

one continuous variable:
`mtcars %>% dplyr::summarize(mean(mpg))`

one categorical variable:
`mtcars %>% dplyr::group_by(cyl) %>%
dplyr::summarize(n())`

two categorical variables:
`mtcars %>% dplyr::group_by(cyl, am) %>%
dplyr::summarize(n())`

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

the pipe

PLOTTING:

one continuous variable:
`ggplot2::qplot(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))`