

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

June 2018

HELLO

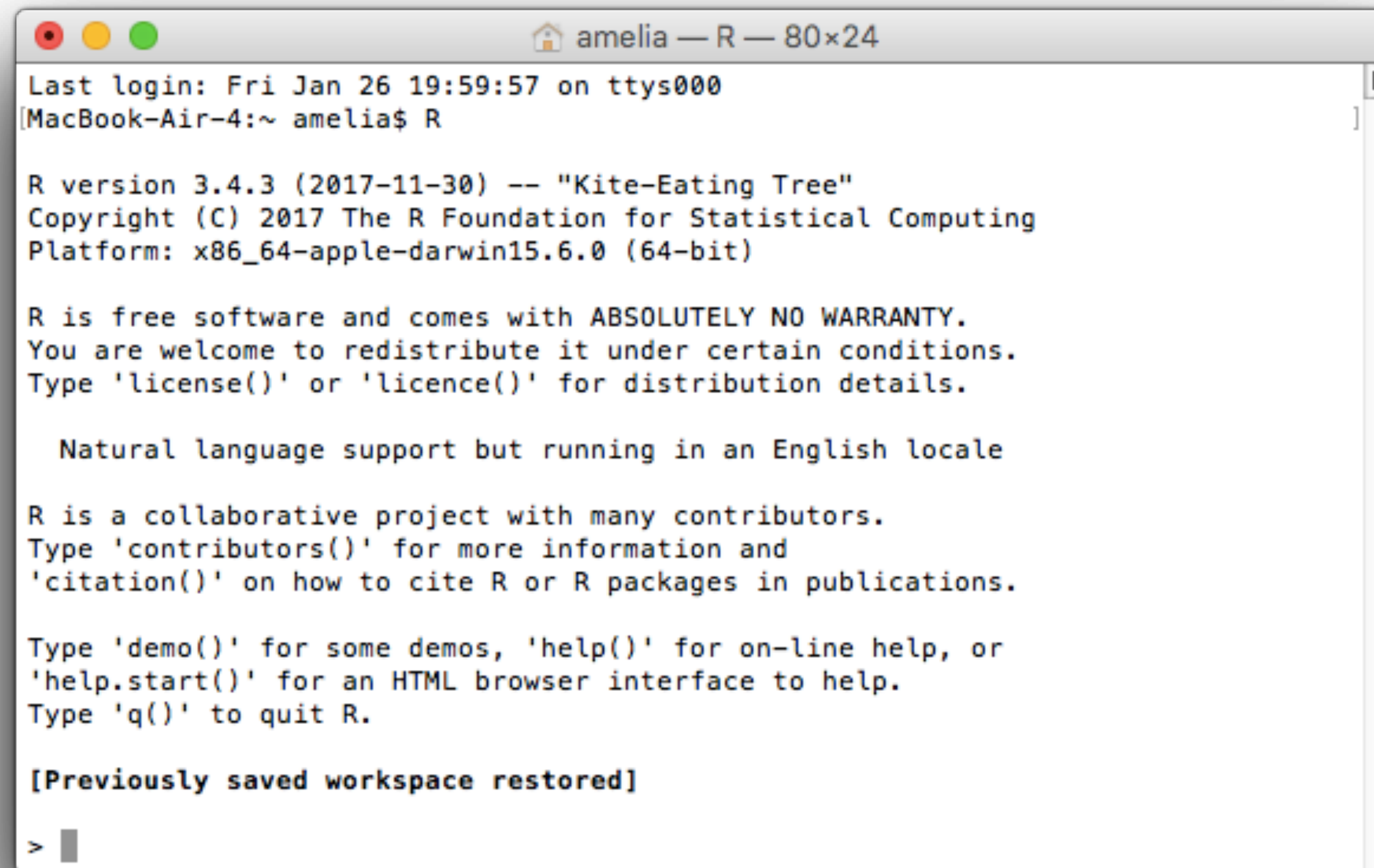
my name is

Amelia

@AmeliaMN



R: a computer programming language



```
amelia — R — 80x24
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

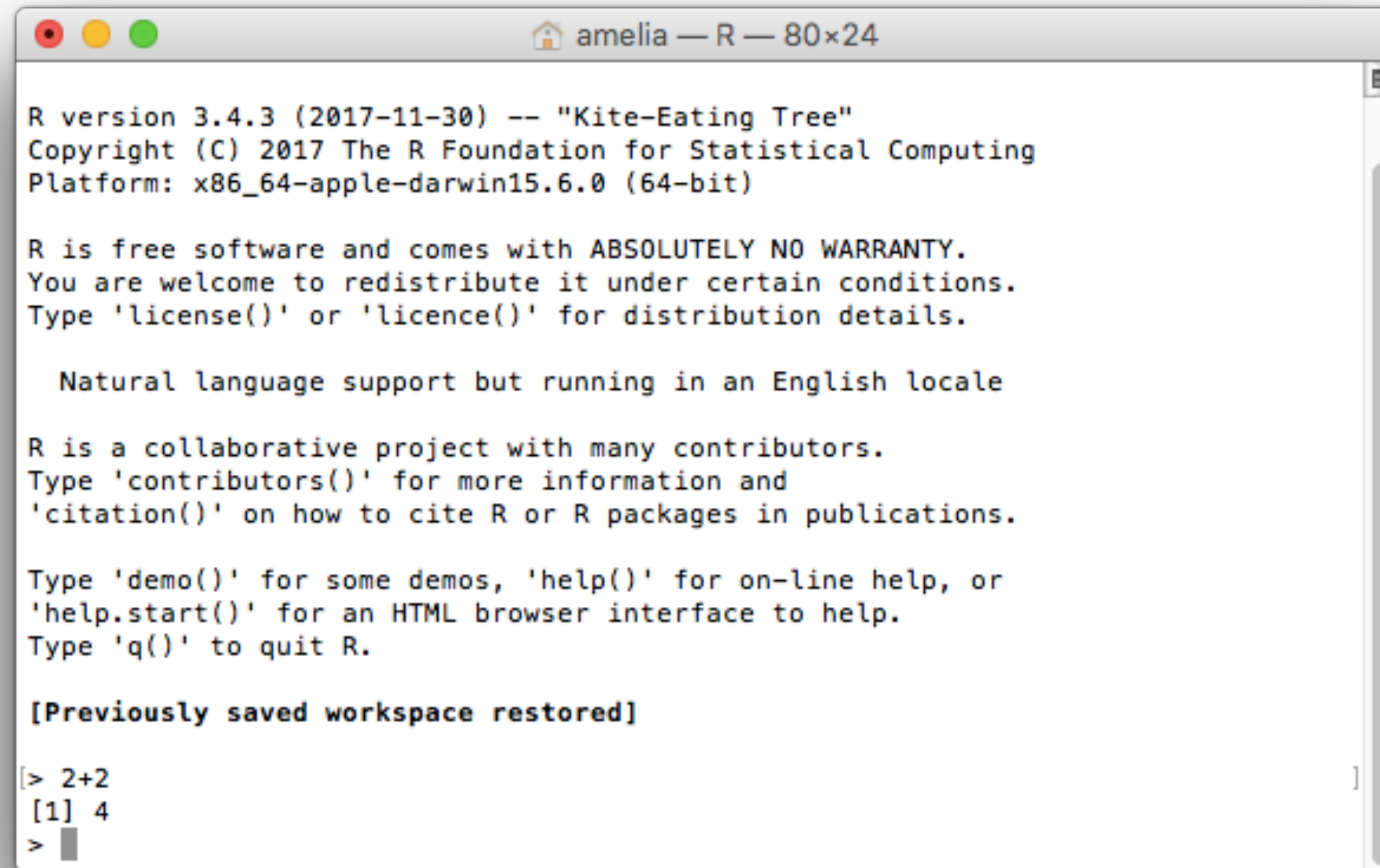
R is a collaborative project with many contributors.
Type 'contributors()' for more information and
'citation()' on how to cite R or R packages in publications.

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



```
amelia — R — 80x24
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.

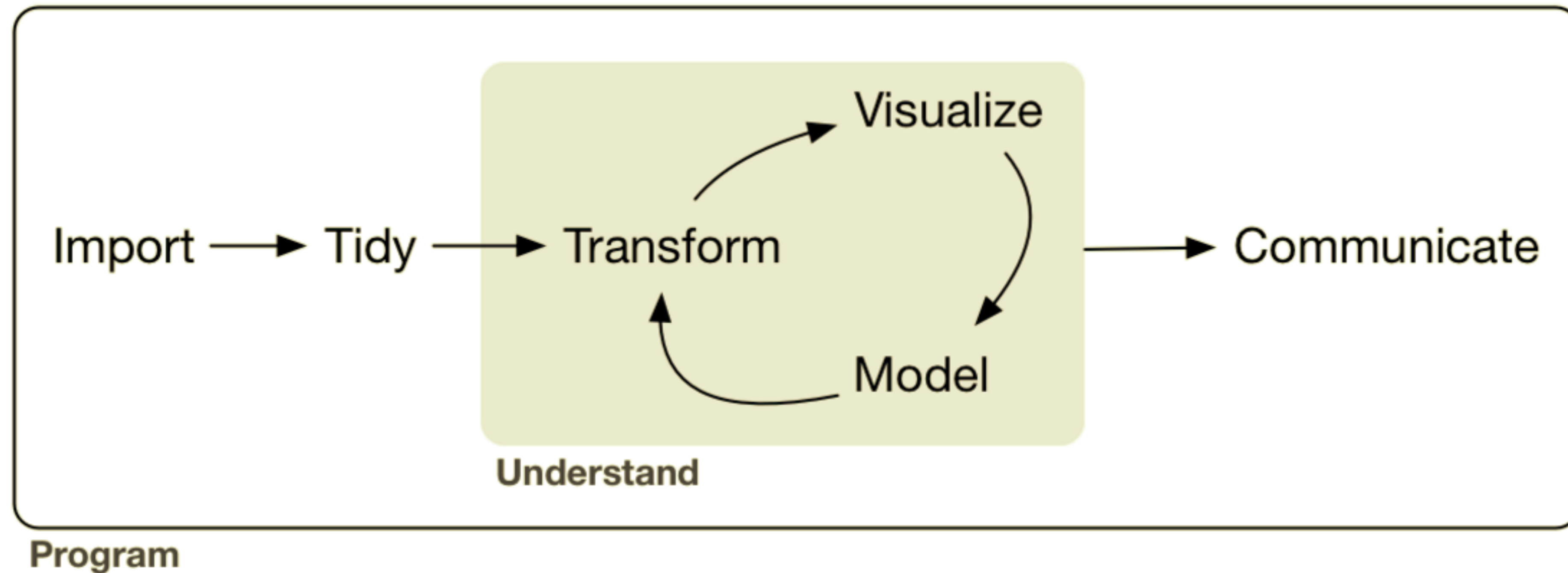
[Previously saved workspace restored]

[> 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

```
> gf_point(domgross_2013~budget_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: 111800000 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

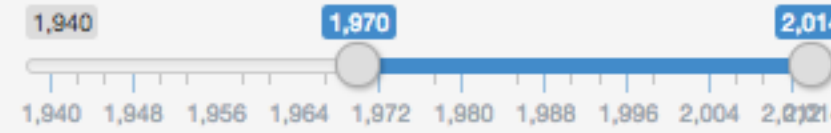
Movie explorer

Filter

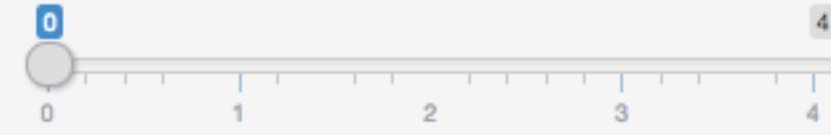
Minimum number of reviews on Rotten Tomatoes



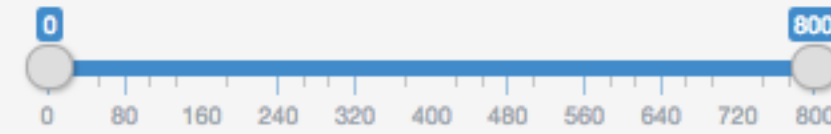
Year released



Minimum number of Oscar wins (all categories)



Dollars at Box Office (millions)



Genre (a movie can have multiple genres)

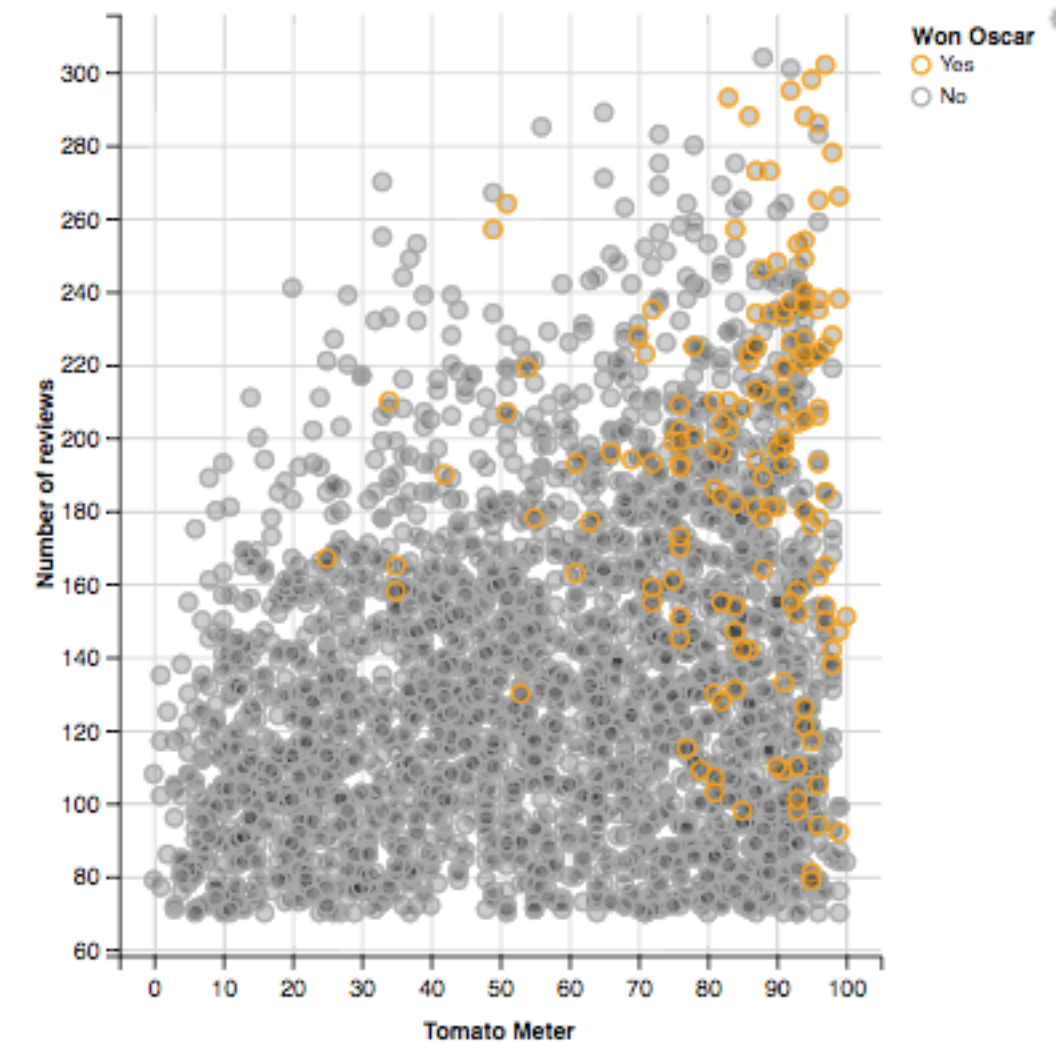
Director name contains (e.g., Miyazaki)

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

www.intro-stats.com

Data

Sources

Transform

Summaries

Graphical

Numerical

Inference

Contingency

Regression

T test

Choose Dataset

MPG

Upload Dataset

Random Subset

Save Subset

Reset Data

Download Data

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/>



RStudio: a software program

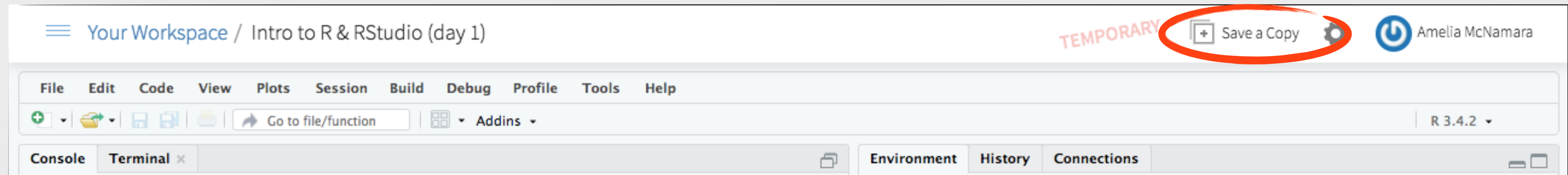
1. like Microsoft Word, Excel, etc.
2. built to help you write R code, run R code, and analyze data with R
3. text editor, version control, keyboard shortcuts, debugging tools, and much more

Your turn

It's time to log in to RStudio, if you haven't already.
Go to <http://bit.ly/statPREP-cloud>

Make an account

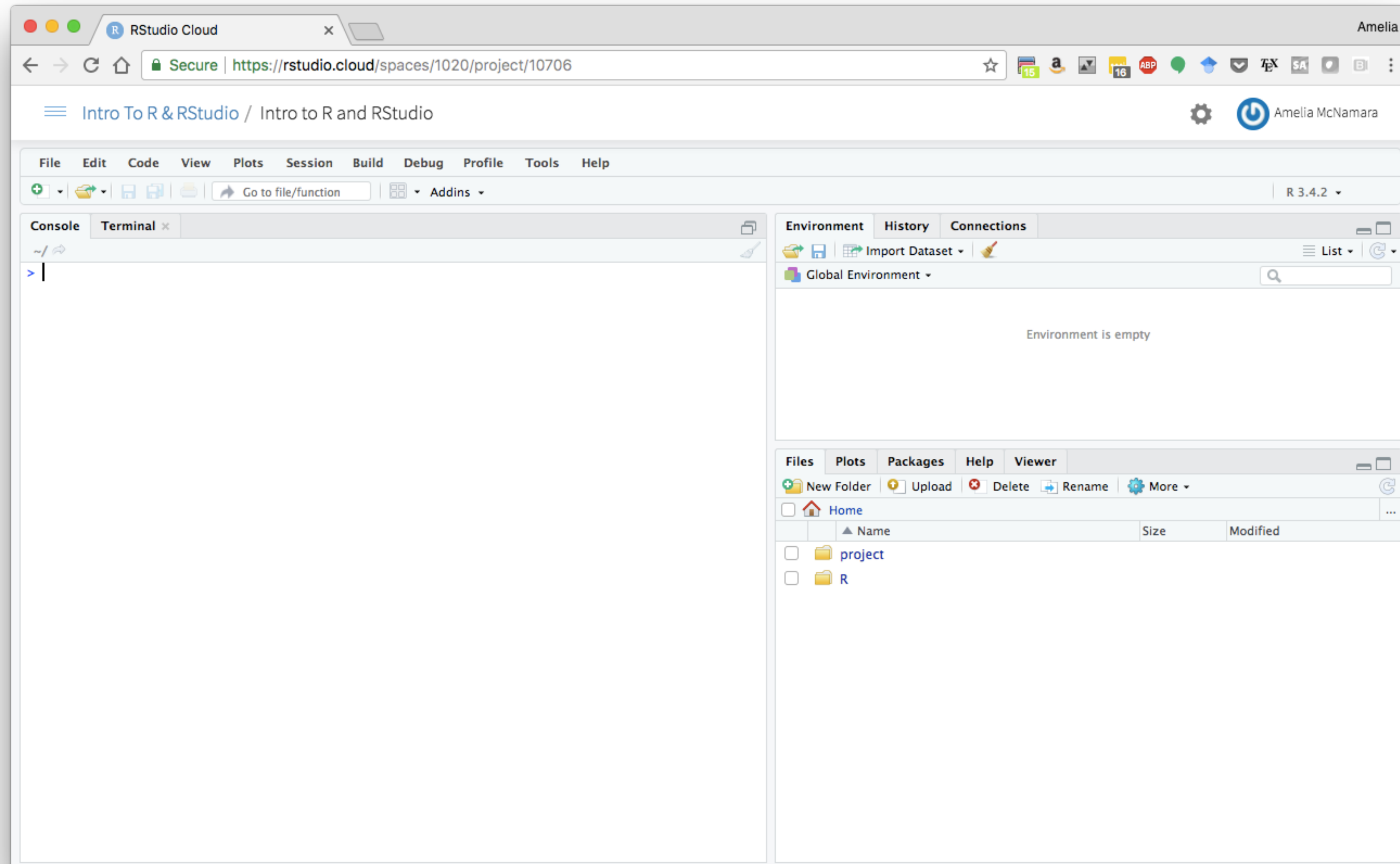
Click Save a Copy



04:00

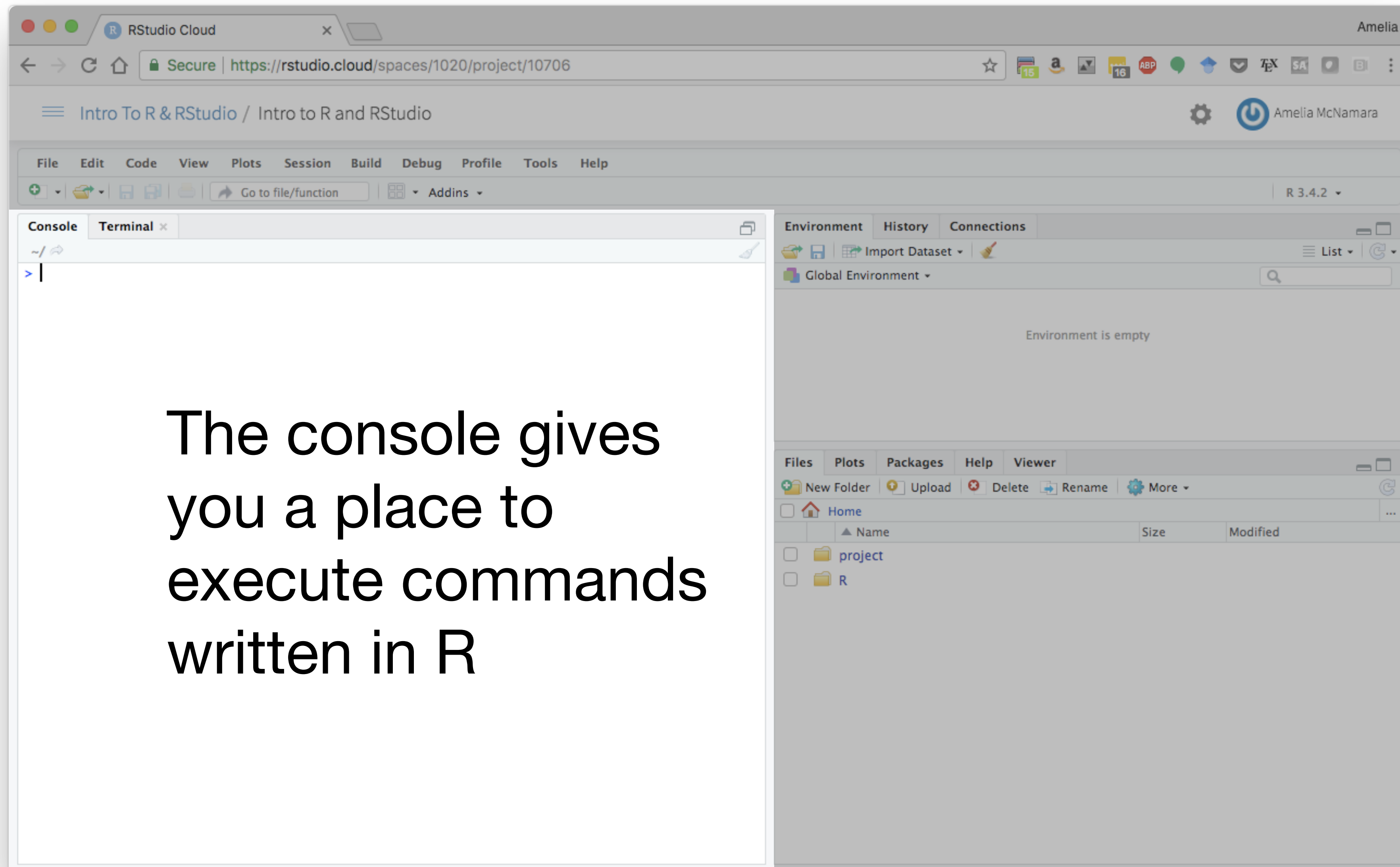
RStudio

<http://bit.ly/statPREP-cloud>



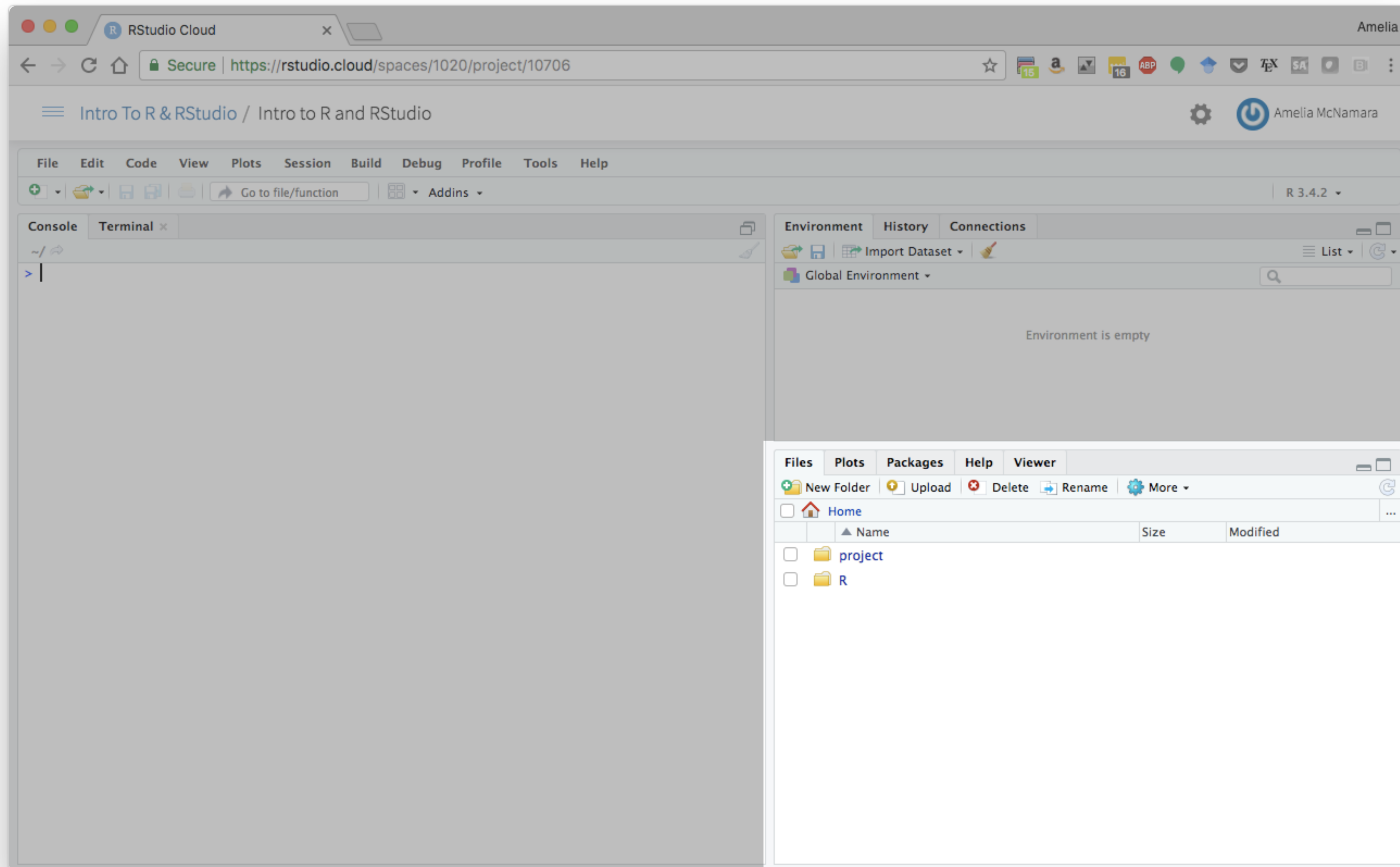
RStudio

<http://bit.ly/statPREP-cloud>



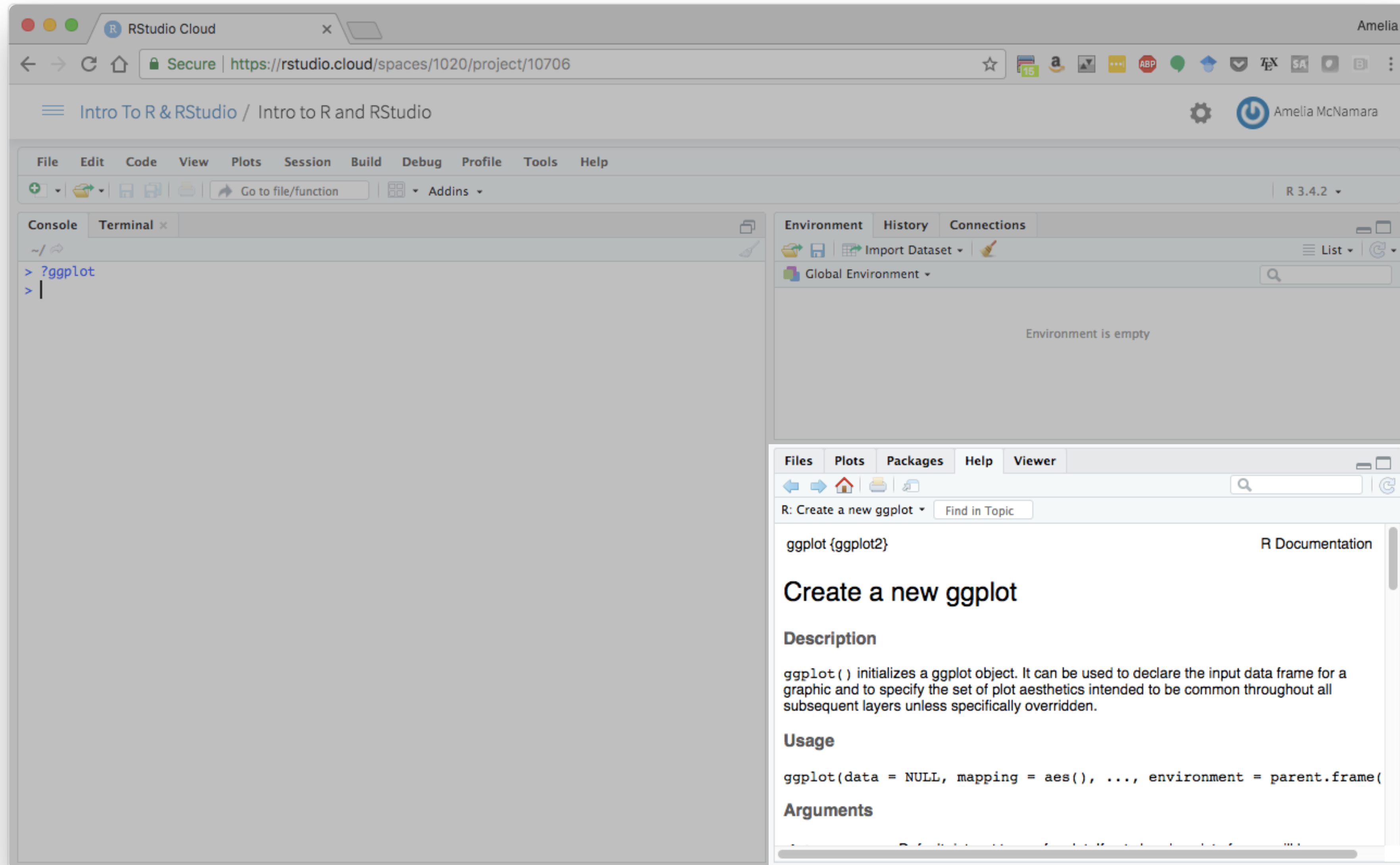
RStudio

<http://bit.ly/statPREP-cloud>



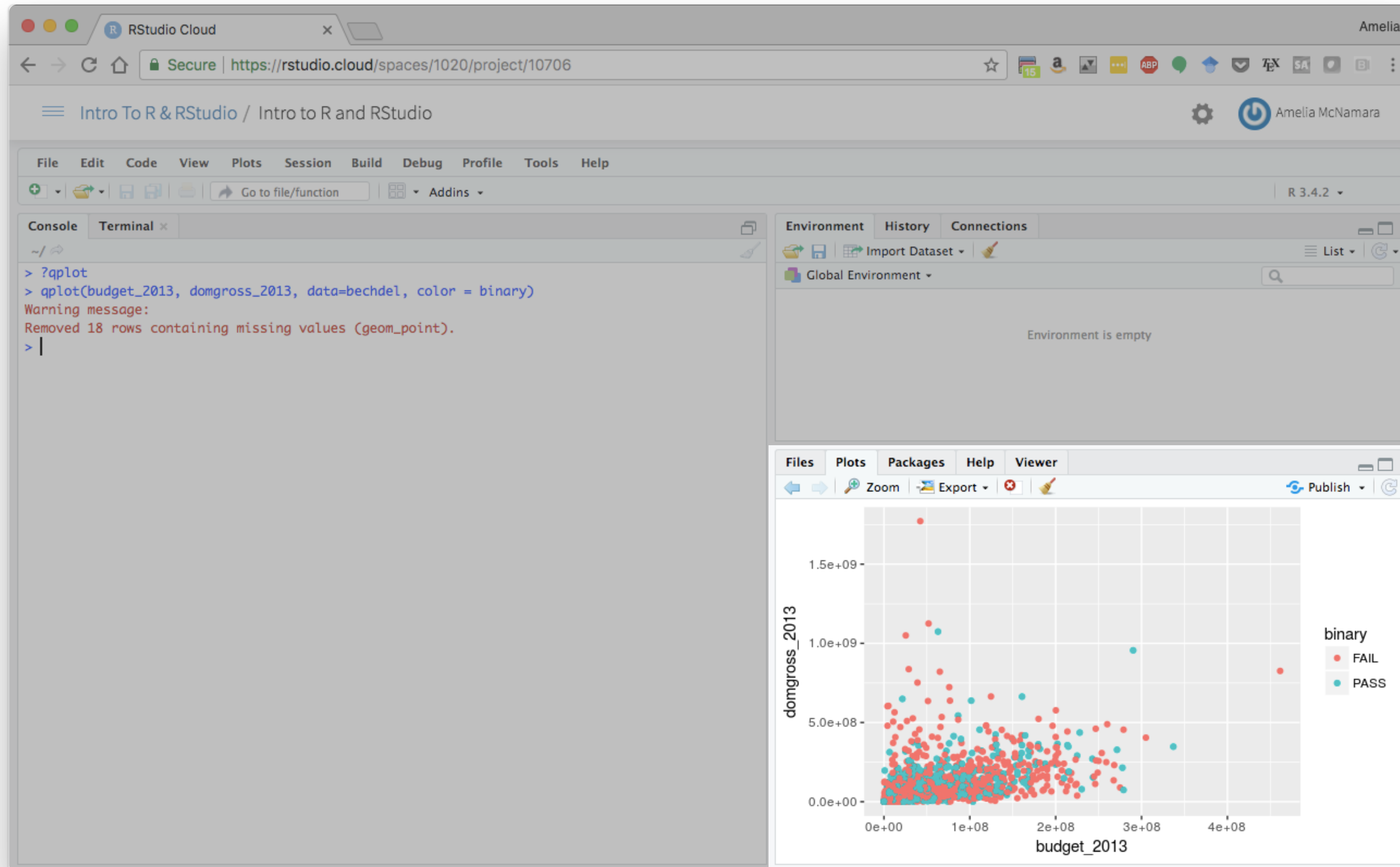
RStudio

<http://bit.ly/statPREP-cloud>



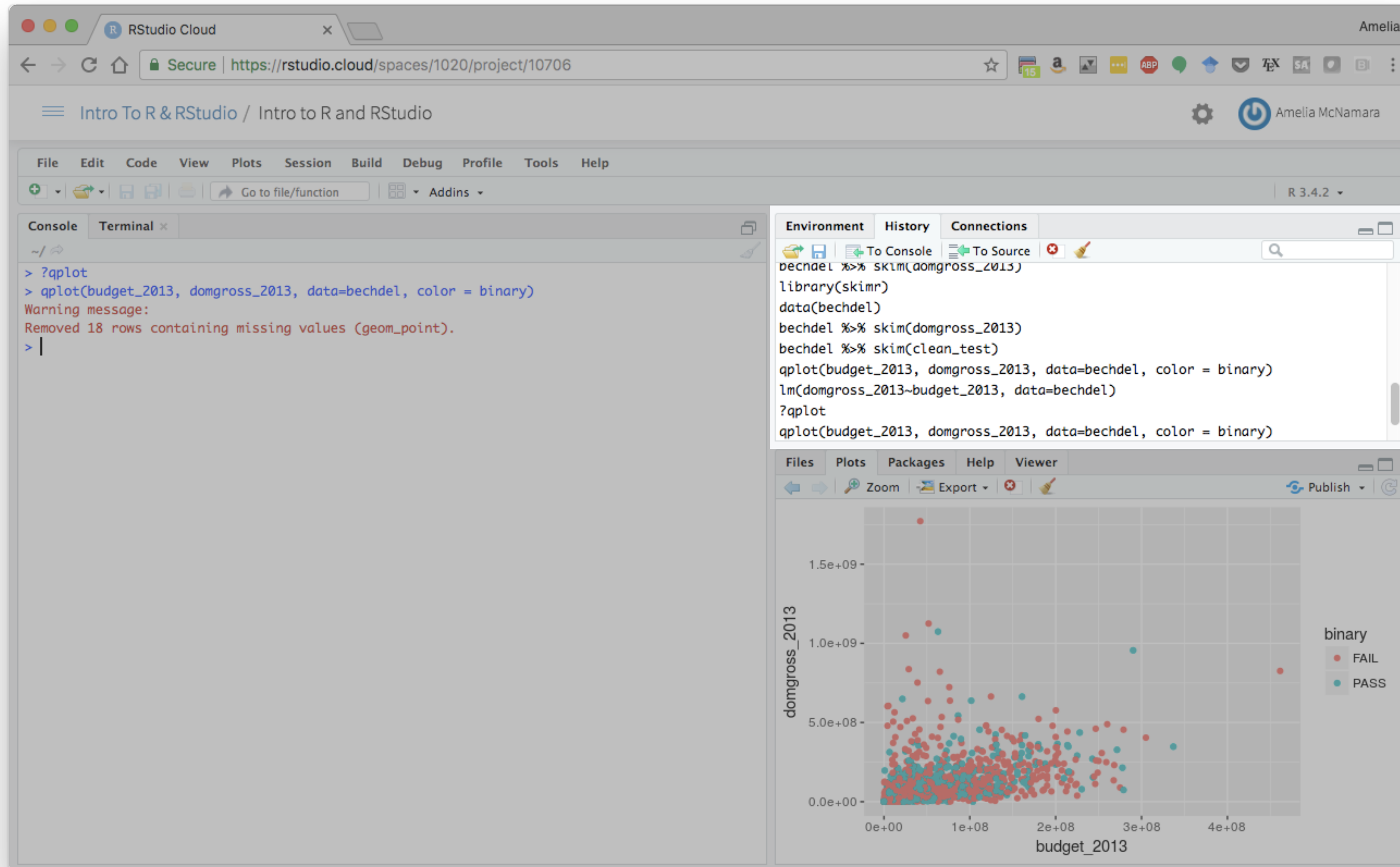
RStudio

<http://bit.ly/statPREP-cloud>



RStudio

<http://bit.ly/statPREP-cloud>



RStudio

<http://bit.ly/statPREP-cloud>

The screenshot displays the RStudio Cloud web interface. At the top, the browser address bar shows the URL `https://rstudio.cloud/spaces/1020/project/10706`. The main interface is divided into several panes:

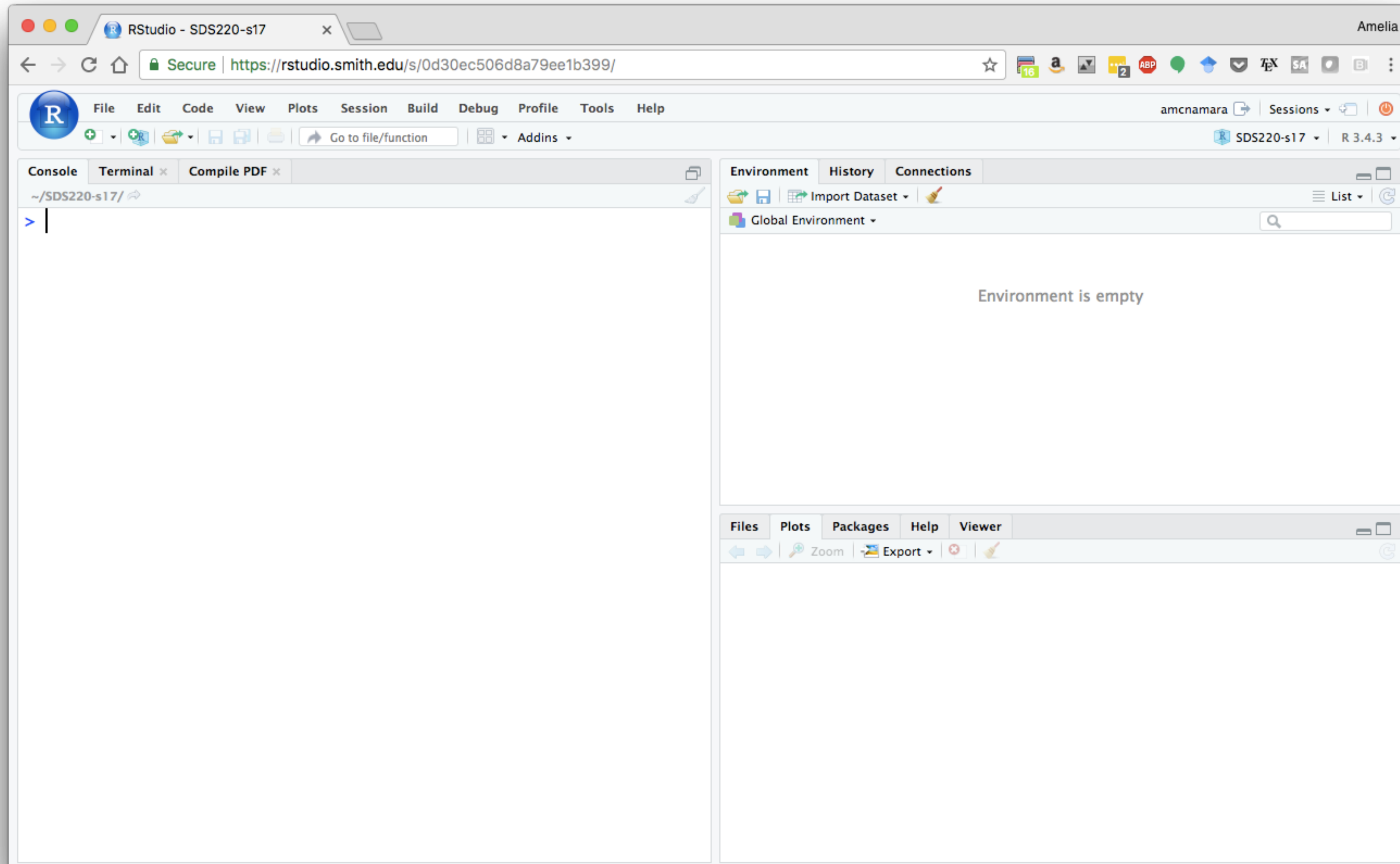
- Code Editor:** Contains an R Markdown document with the following content:

```
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
13 HTML, PDF, and MS Word documents. For more details on using R Markdown see
14 <http://rmarkdown.rstudio.com>.
15
16 When you click the Knit button a document will be generated that includes both
```
- Console:** Shows the execution of the following R code:

```
> ?qplot
> qplot(budget_2013, domgross_2013, data=bechdel, color = binary)
Warning message:
Removed 18 rows containing missing values (geom_point).
>
```
- Plots:** Displays a scatter plot of `domgross_2013` (y-axis) versus `budget_2013` (x-axis). The plot uses a binary color scheme where red points represent 'FAIL' and teal points represent 'PASS'. The x-axis ranges from 0 to 4e+08, and the y-axis ranges from 0 to 1.5e+09.
- Environment:** Shows the 'Global Environment' with the message 'Environment is empty'.

RStudio

<http://bit.ly/statPREP-cloud>

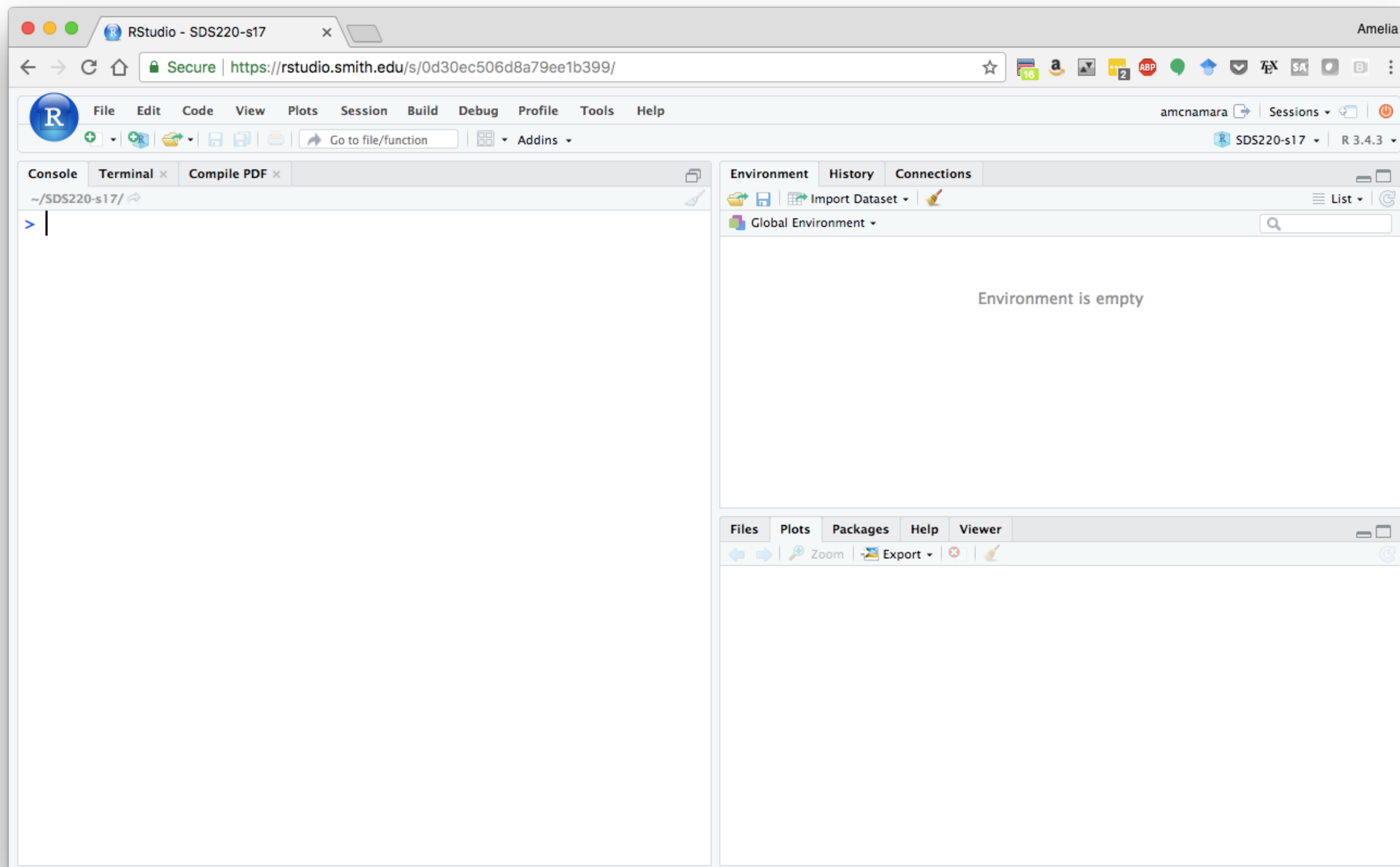


RStudio: ways to use

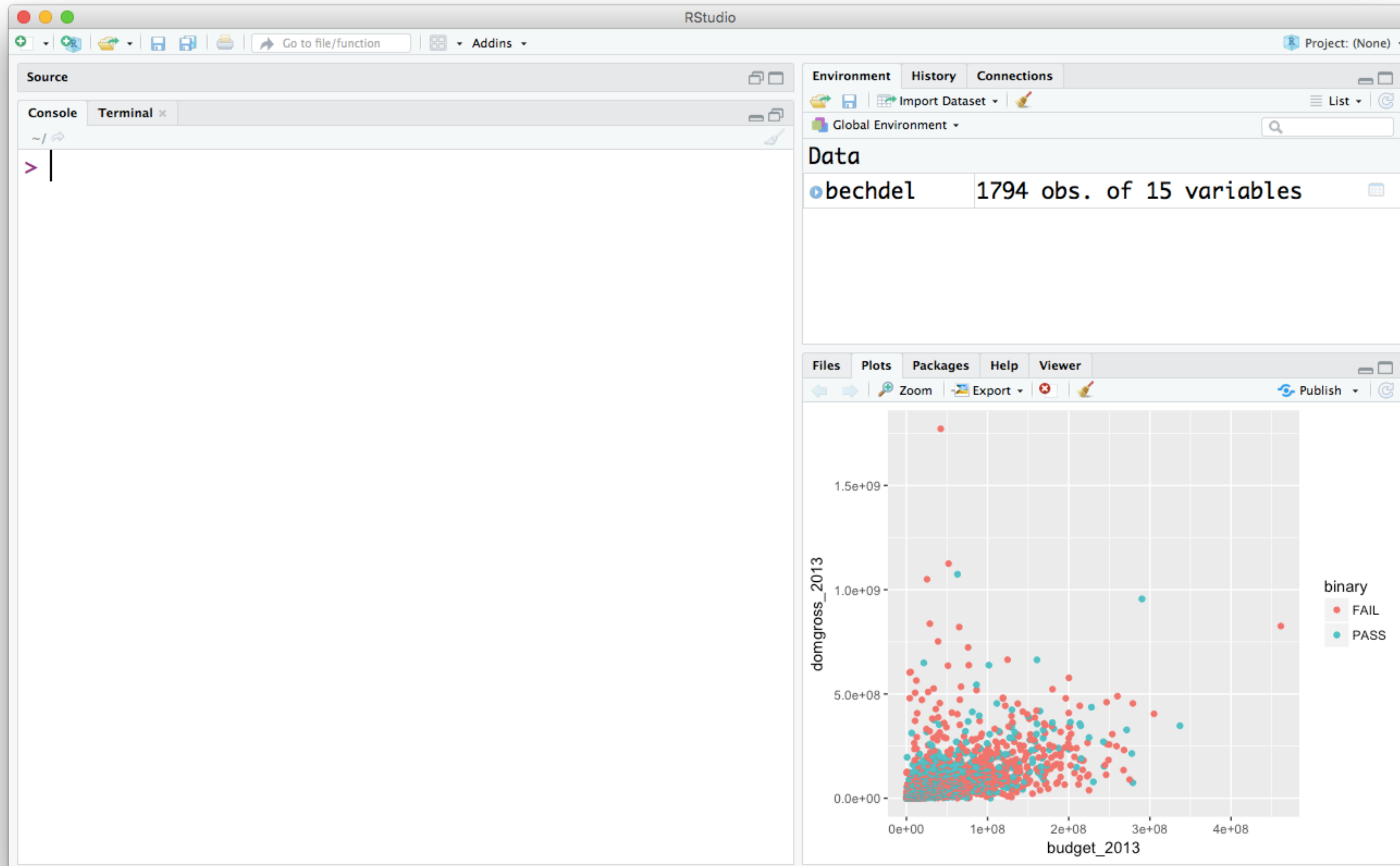
We're using RStudio Cloud, which allows you to log in through a web browser and do your work there.

But, there are other versions of RStudio.

RStudio: server edition



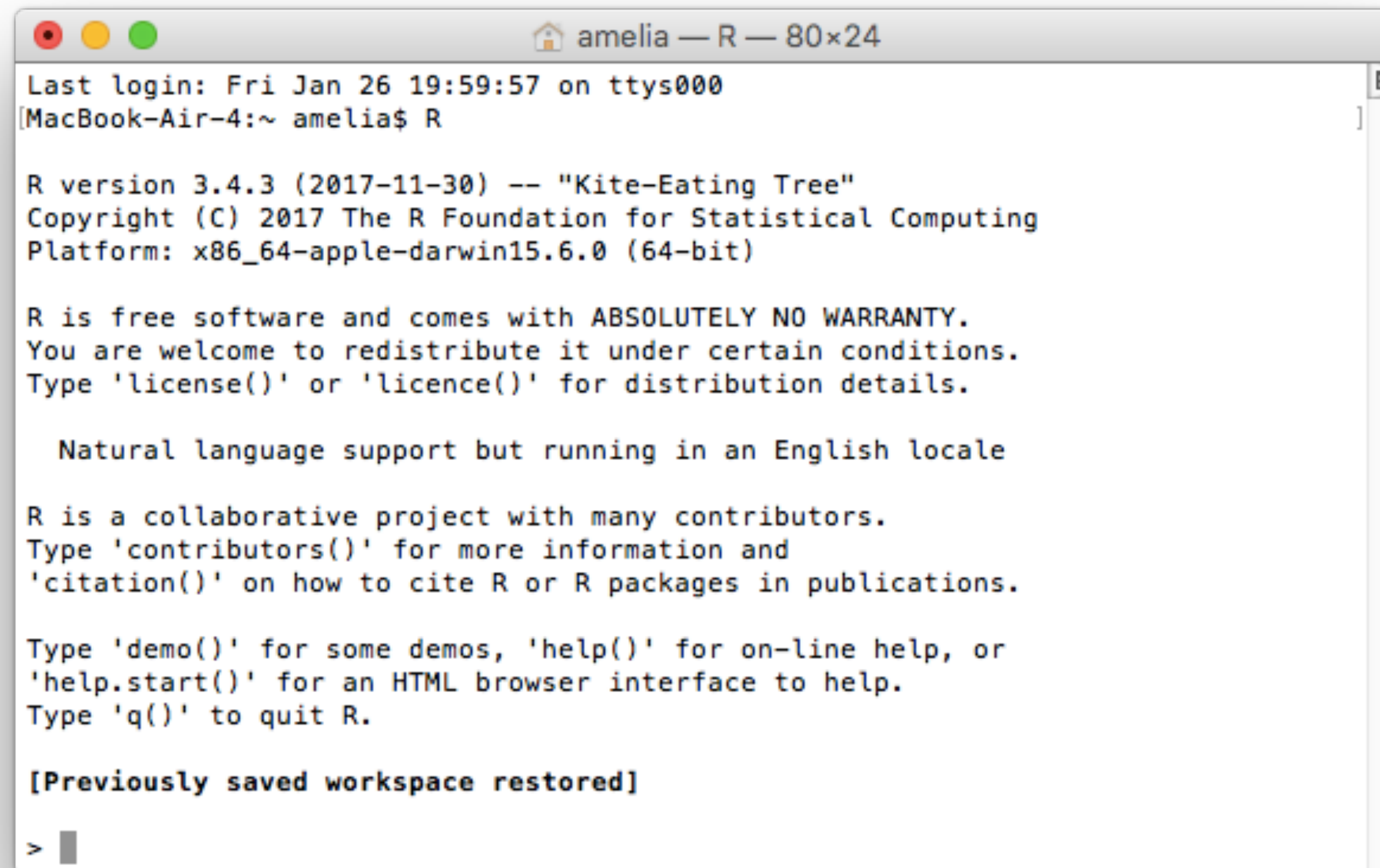
RStudio: desktop edition



Transferability

Everything you learn here will work in any version of RStudio

It will even work in the basic console version of R

A screenshot of a terminal window titled "amelia — R — 80x24". The window shows the output of running the R command in a terminal. The text displayed is: "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]", and a prompt ">".

```
amelia — R — 80x24
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

<http://bit.ly/statPREP-cloud>

The screenshot displays the RStudio Cloud web interface. At the top, the browser address bar shows the URL <https://rstudio.cloud/spaces/1020/project/10706>. The page title is "Intro To R & RStudio / Intro to R and RStudio". The user's name, "Amelia", is visible in the top right corner.

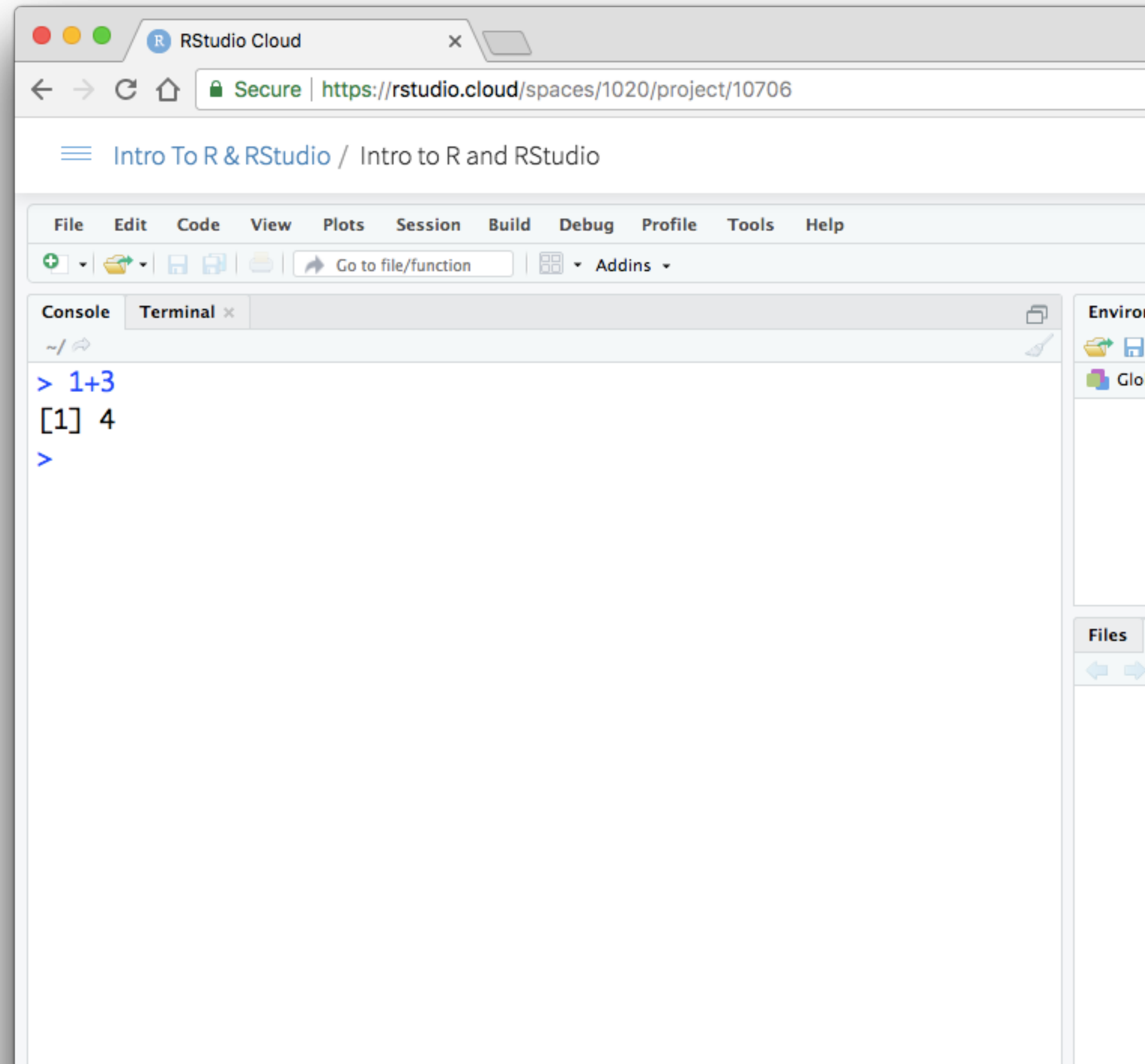
The main interface is divided into several panels:

- Menu Bar:** Includes File, Edit, Code, View, Plots, Session, Build, Debug, Profile, Tools, and Help.
- Terminal:** Located on the left, it shows a prompt `> |` in a terminal window.
- Environment Panel:** Located on the right, it shows "Global Environment" and "Environment is empty".
- Files Panel:** Located at the bottom right, it shows a file browser with a table of files and folders:

	Name	Size	Modified
<input type="checkbox"/>	Home		
<input type="checkbox"/>	project		
<input type="checkbox"/>	R		

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

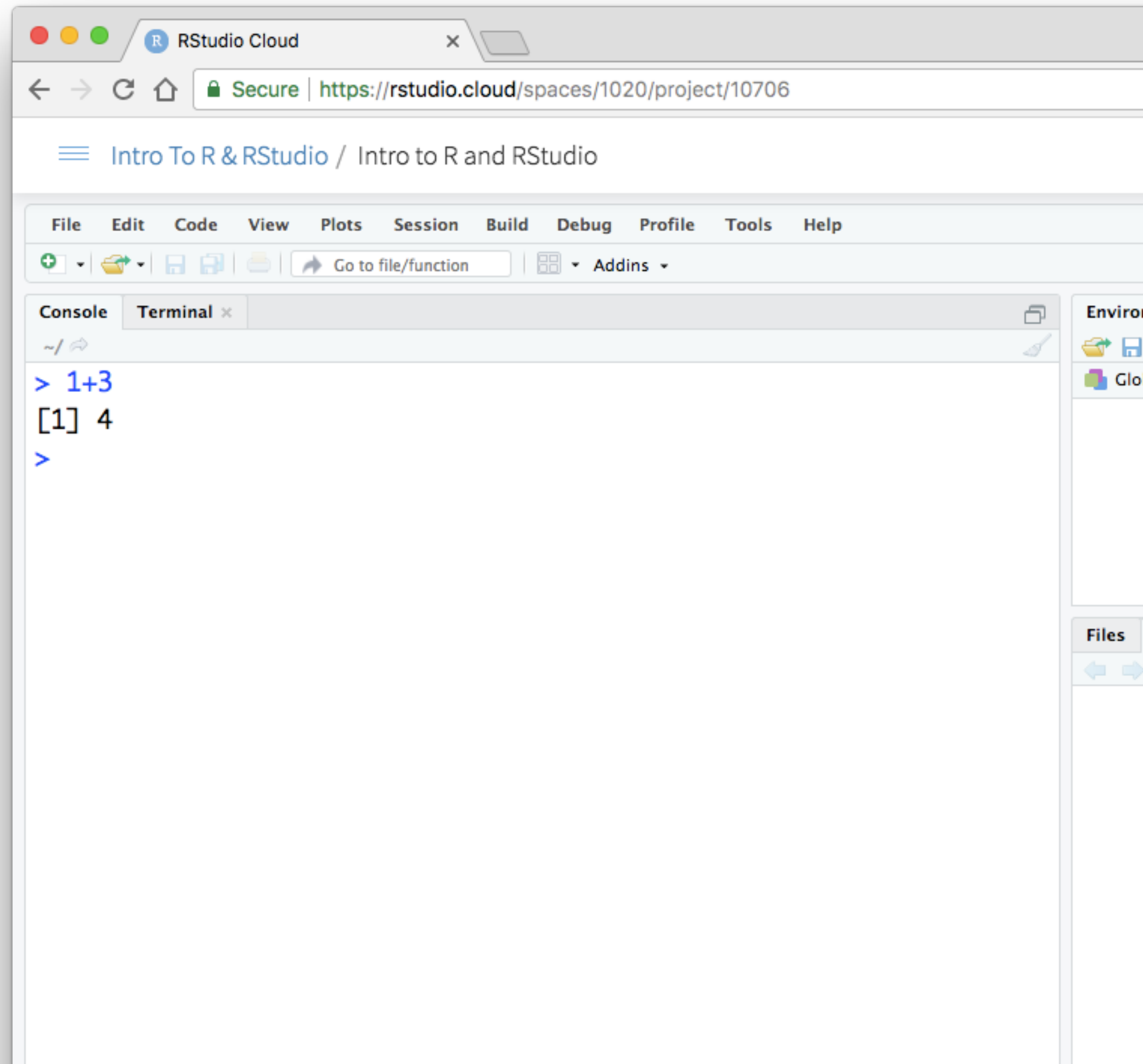
Type commands on the line that begins with a > sign (known as the prompt)



Output

Output →
New prompt →

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

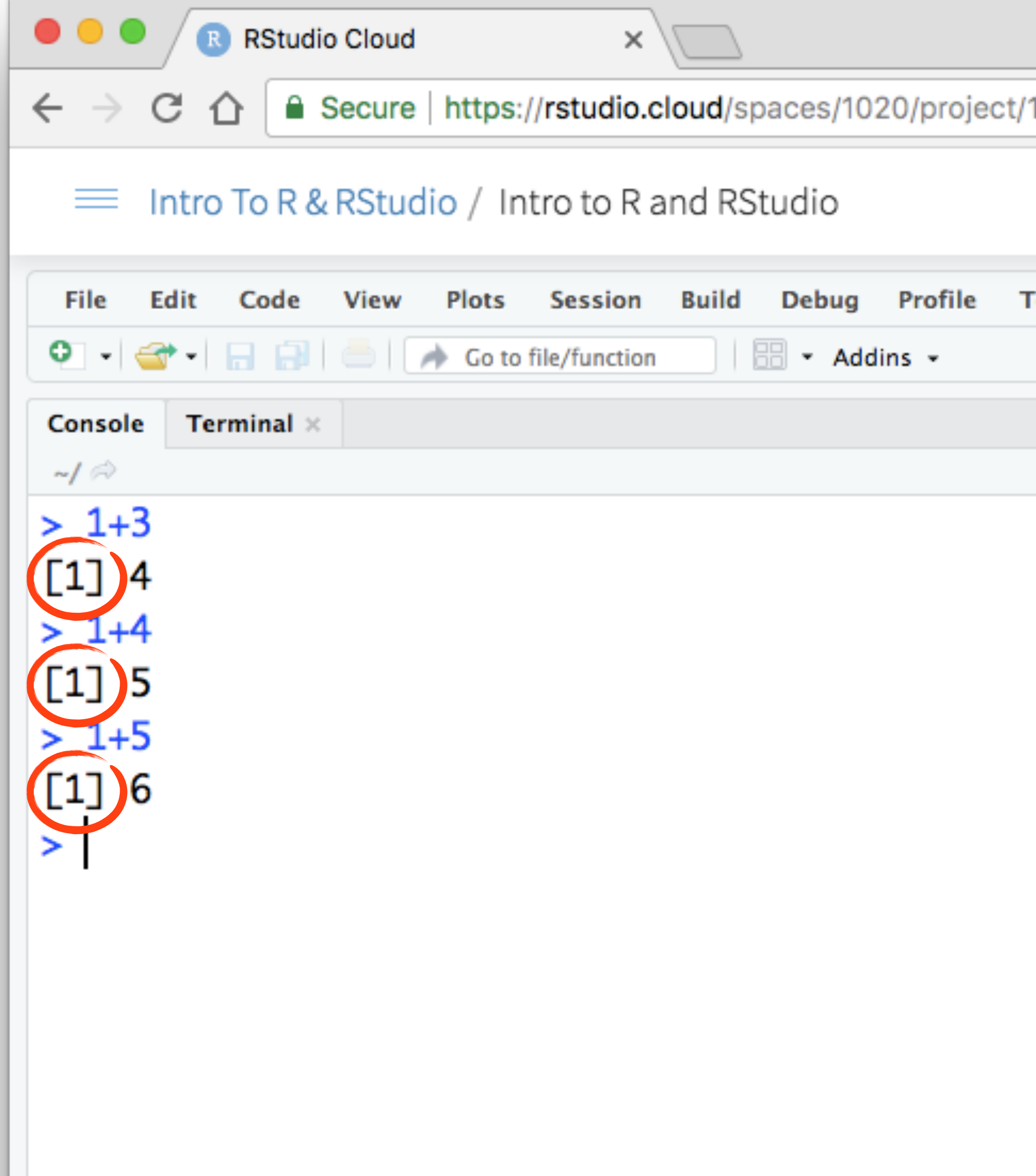


[1]

R displays an index next to the output.

Just ignore this.

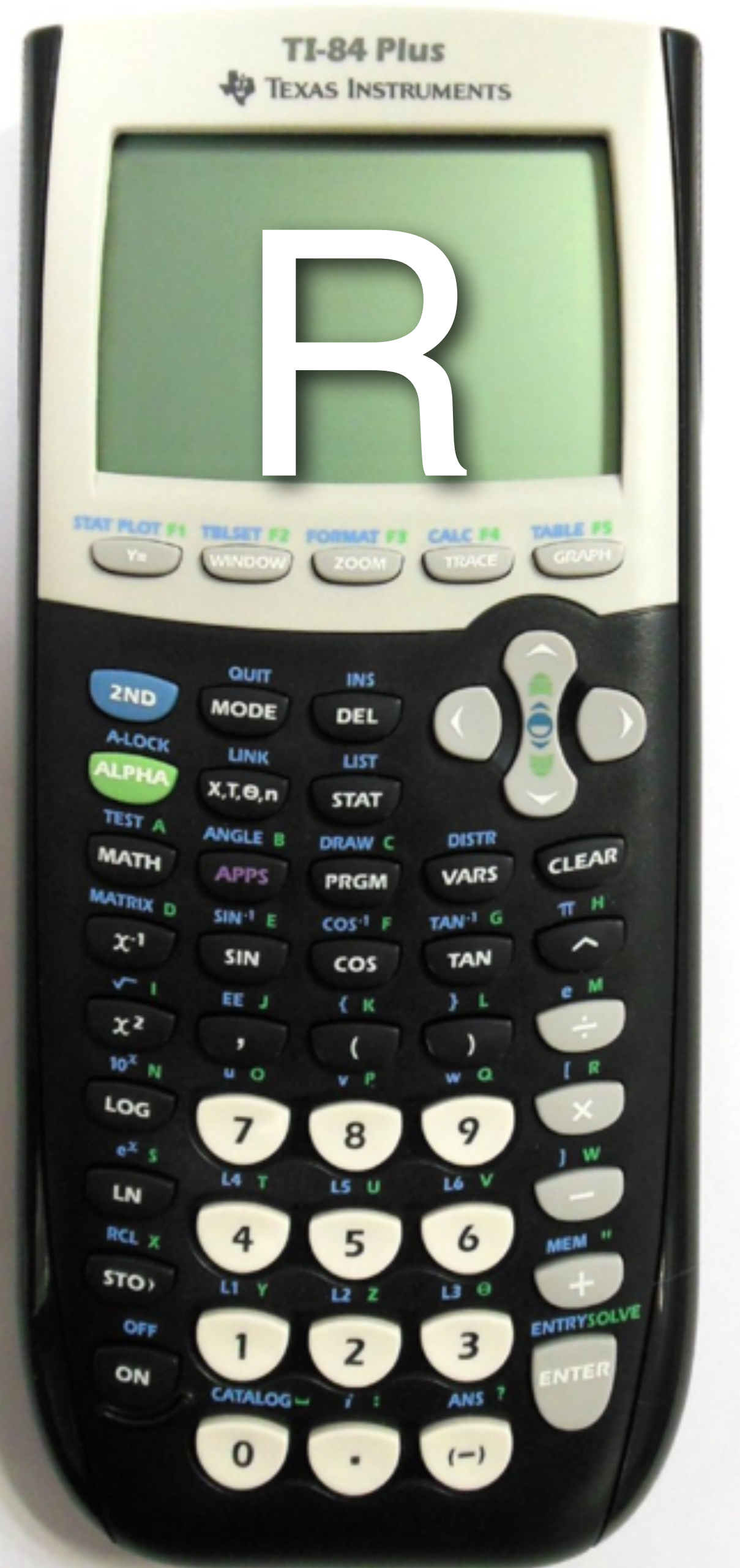
Somewhat helpful when R returns more than one value in the output.



The screenshot shows the RStudio Cloud interface. The browser address bar displays `https://rstudio.cloud/spaces/1020/project/1`. 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", and "Tools". The toolbar contains icons for file operations and a search bar labeled "Go to file/function". The console window is active, showing the following R session:

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

The output values 4, 5, and 6 are circled in red in the original image.



R is like a fancy calculator
on your computer

$$5 + 5$$

10

$$4 - 1$$

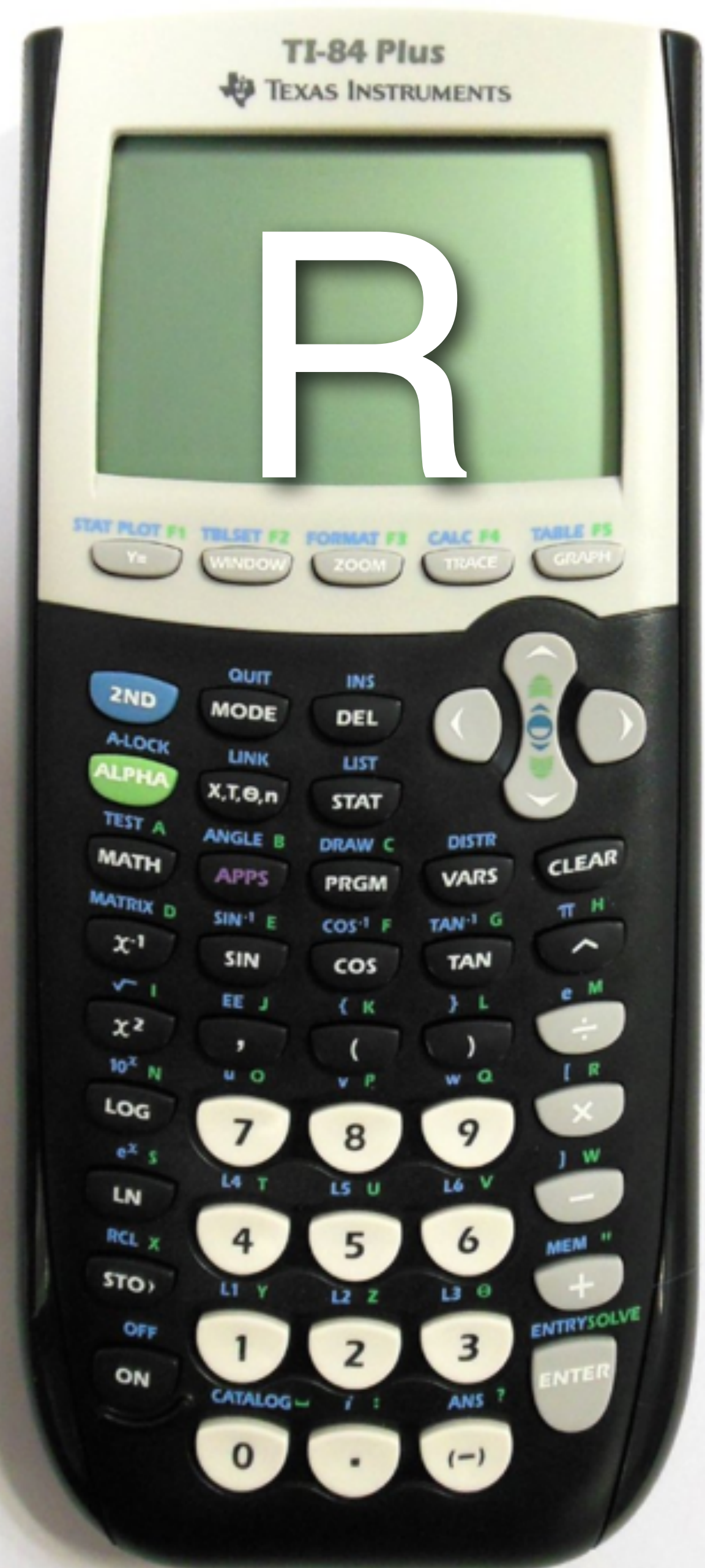
3

$$1 * 2$$

2

$$4 ^ 2$$

16



It can do algebra

$$a \leftarrow 1$$

$$b \leftarrow 2$$

$$a + b$$

3

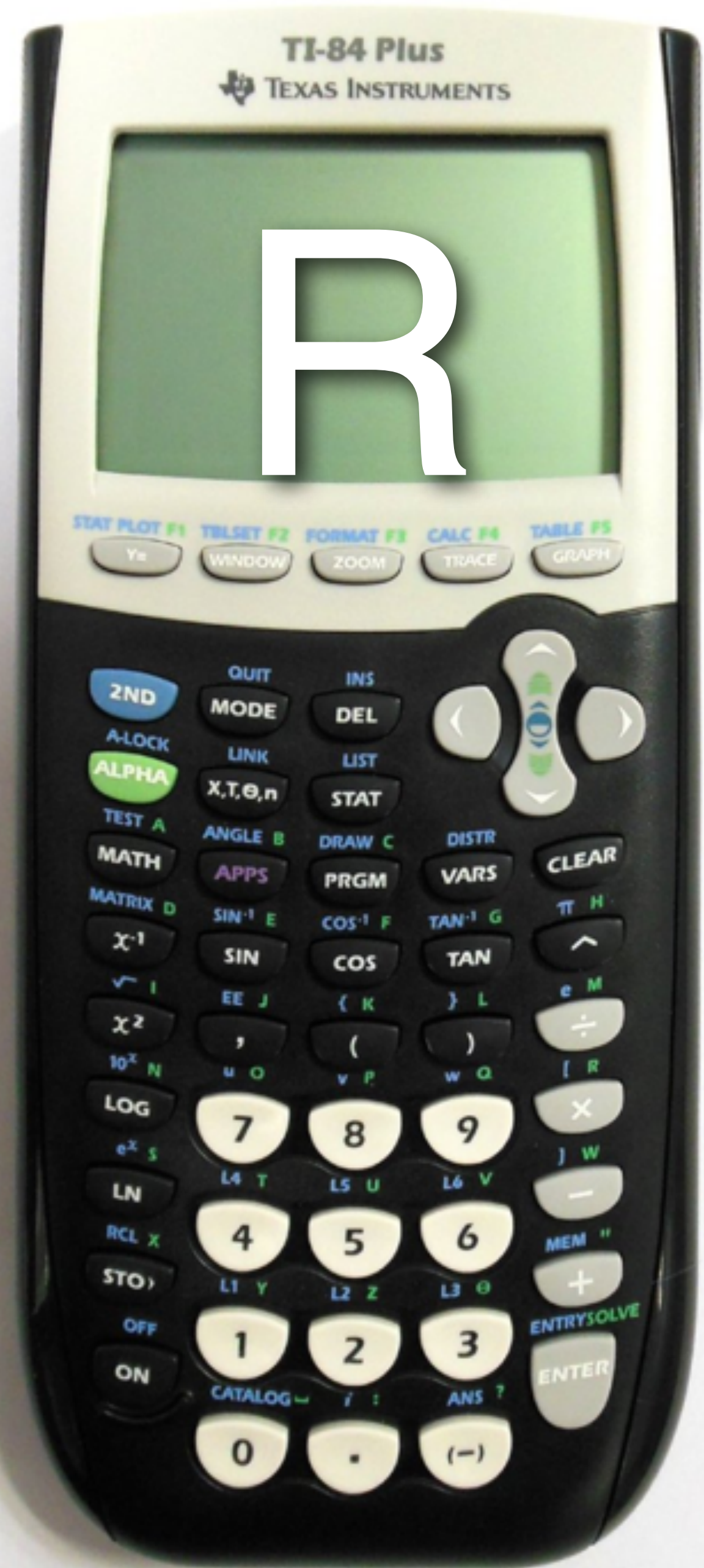
$$A \leftarrow 3$$

$$a + b - A$$

0

It cares about capitalization





And it has functions that let you do more sophisticated manipulations

`round(3.1415)`

3

`factorial(3)`

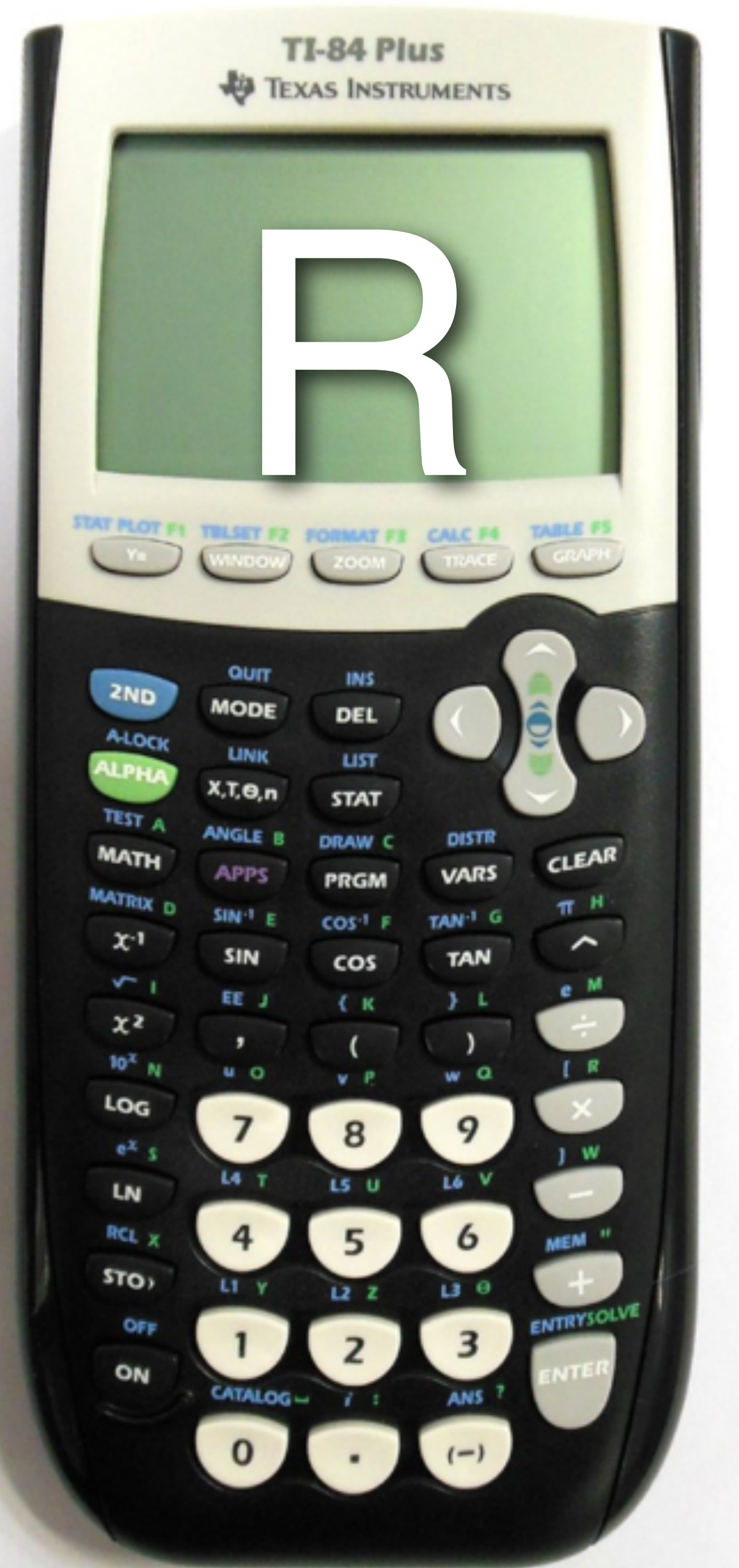
6

`sqrt(9)`

3

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

square root



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) ↙ factorial of 3

6

sqrt(9) ↙ square root of 9

3

+ prompt

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

Either finish the command, or press escape.

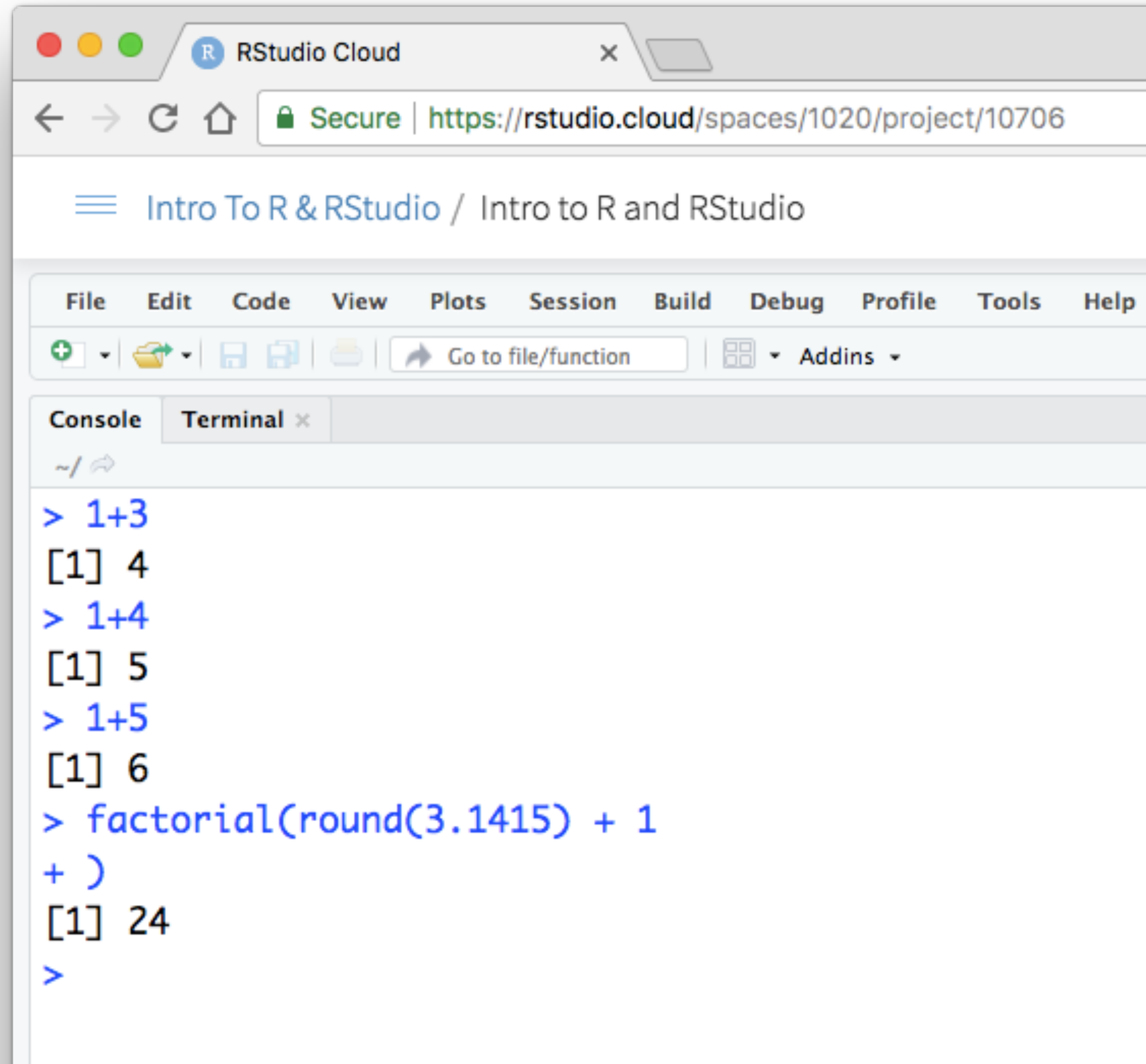


```
RStudio Cloud
Secure | https://rstudio.cloud/spaces/1020/project/10706
Intro To R & RStudio / Intro to R and RStudio
File Edit Code View Plots Session Build Debug Profile Tools Help
~/
> 1+3
[1] 4
> 1+4
[1] 5
> 1+5
[1] 6
> factorial(round(3.1415)) + 1
+
```


+ 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 the RStudio Cloud interface. The browser address bar displays `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". The toolbar contains icons for file operations and a search box. The terminal window is active, showing the following R session:

```
~/  
> 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.

We'll begin with a document I have started for you, called `01-Intro.Rmd`

RStudio Cloud

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

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
11 the Console. Notice that the math is surrounded by some
12 special characters above and below, and has a grey
13 background. This is RStudio's way of showing you the
14 difference between an area to write text (here!) and an
15 area where code is stored (called a code chunk).
```

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

R Notebook

The image shows a screenshot of the RStudio Cloud web interface. The browser address bar shows the URL `https://rstudio.cloud/spaces/1020/project/10706`. The page title is "Intro To R & RStudio / Intro to R and RStudio". The user's name, "Amelia McNamara", is visible in the top right corner. The main editor area displays a file named "01-Structures.Rmd" with the following content:

```
1 ---
2 title: Notebooks
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).
```

Below the code editor, the "Console" panel is visible and highlighted with a red circle. It shows the output of the code chunk:

```
1:1 [R Notebook]
5 + 5
4 - 1
1 * 2
```

Overlaid on the screenshot is a large text box with the following text:

Notice that the Console has automatically minimized itself, to give you room to work in your document. From here out, we'll be working almost exclusively in documents, but all the code we write would work in the Console as well.

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 web interface. The browser address bar displays `https://rstudio.cloud/spaces/1020/project/10706`. The main workspace contains a code chunk titled "01-Structures.Rmd*" with the following content:

```
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 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 ```
```

Below the code chunk, the output is displayed in a white box with a grey border:

```
[1] 10
```

The interface includes a menu bar (File, Edit, Code, View, Plots, Session, Build, Debug, Profile, Tools, Help) and a toolbar with icons for file operations and a "Run" button.

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

Home > Dropbox > Intro_to

	Name
<input type="checkbox"/>	..
<input type="checkbox"/>	.Rhistory
<input type="checkbox"/>	04-Syntax.nb.html
<input type="checkbox"/>	04-Syntax.Rmd
<input type="checkbox"/>	03-DataTypes.nb.html
<input type="checkbox"/>	03-DataTypes.Rmd
<input type="checkbox"/>	02-Visualization.Rmd
<input type="checkbox"/>	01-Intro.Rmd
<input type="checkbox"/>	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**

Global Environment

Files Plots Packages Help

New Folder Delete

Home > Dropbox > Intro_to

	Name
	..
<input type="checkbox"/>	.Rhistory
<input type="checkbox"/>	04-Syntax.nb.html
<input type="checkbox"/>	04-Syntax.Rmd
<input type="checkbox"/>	03-DataTypes.nb.html
<input type="checkbox"/>	03-DataTypes.Rmd
<input type="checkbox"/>	02-Visualization.Rmd
<input type="checkbox"/>	01-Intro.Rmd
<input type="checkbox"/>	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. The environment pane on the right displays the following table:

Values	
a	1
A	3
b	2

The table is circled in red. The RStudio interface also shows the menu bar (View, Plots, Session, Build, Debug, Profile, Tools, Help), the toolbar (Go to file/function, Addins), and the file explorer at the bottom showing a directory structure: Home > Day1 > code, with files 01-Structures.Rmd (1.3 KB), 02-Syntax.Rmd (869 B), and 03-Programming.Rmd (869 B).

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
```

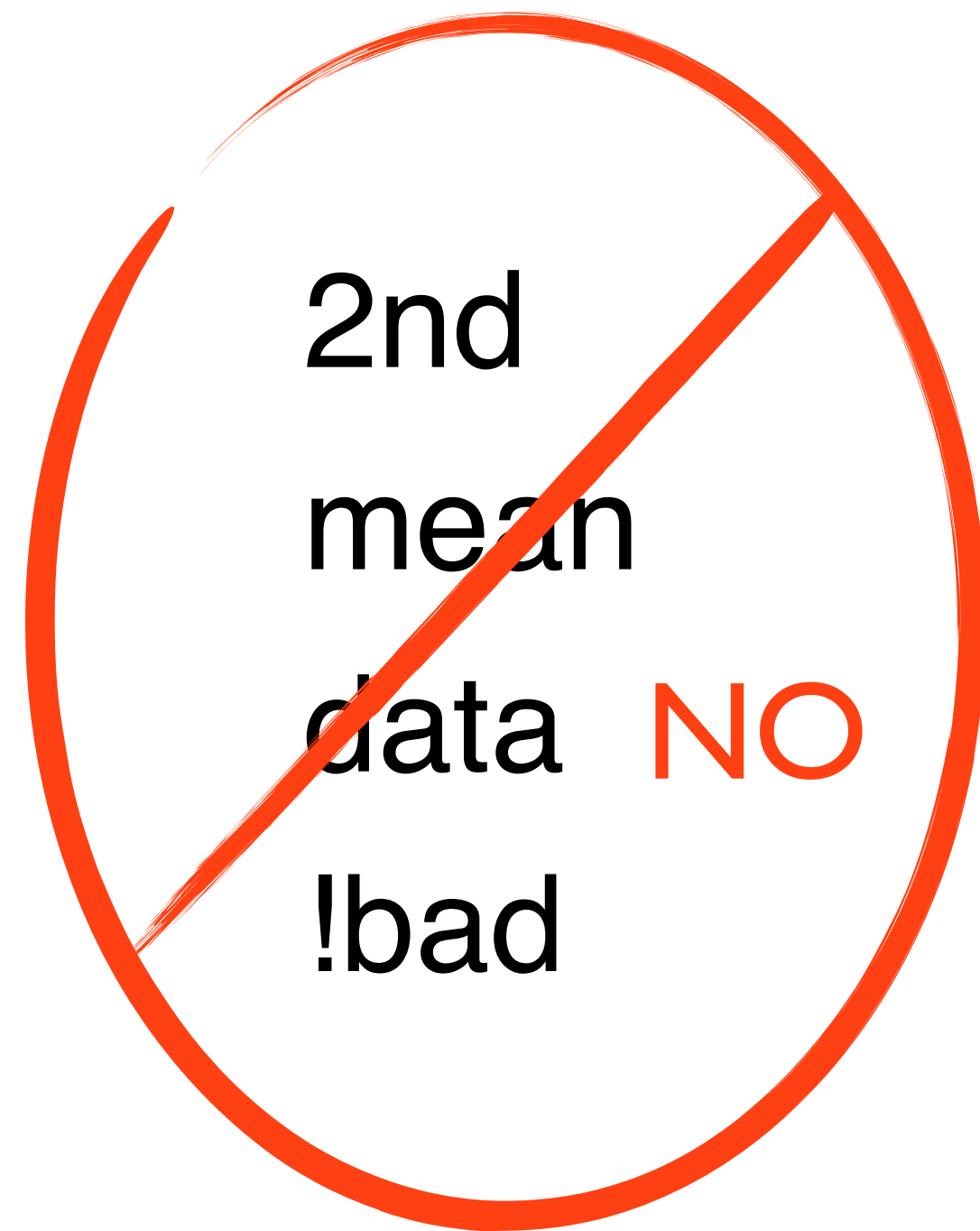
Object names

Object names cannot begin with numbers

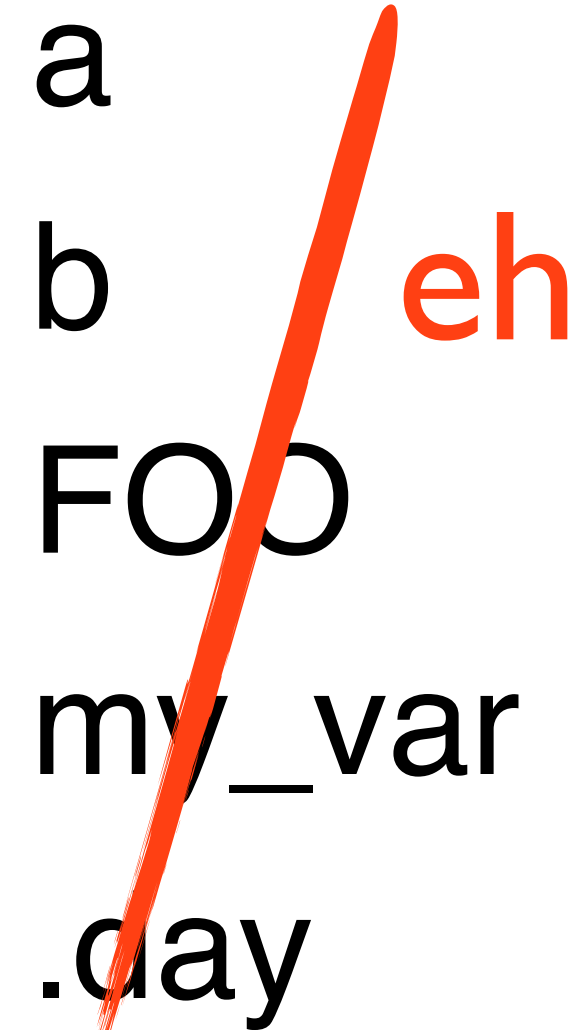
They cannot contain spaces

It is wise to avoid names already in use

Informative names are better than generic ones



2nd
mean
data **NO**
!bad



a
b **eh**
FOO
my_var
.day

CDC_data
finalModel
more_pi **yeah!**
withoutOver64

Capitalization matters

R will treat each of these as a different object

cdc_data

CDC_data

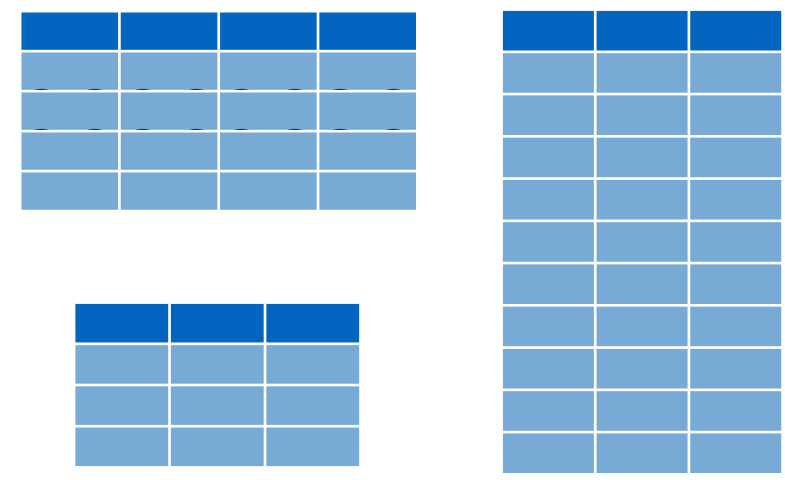
finalmodel

finalModel

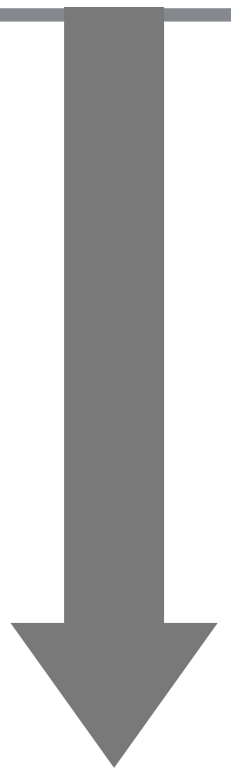
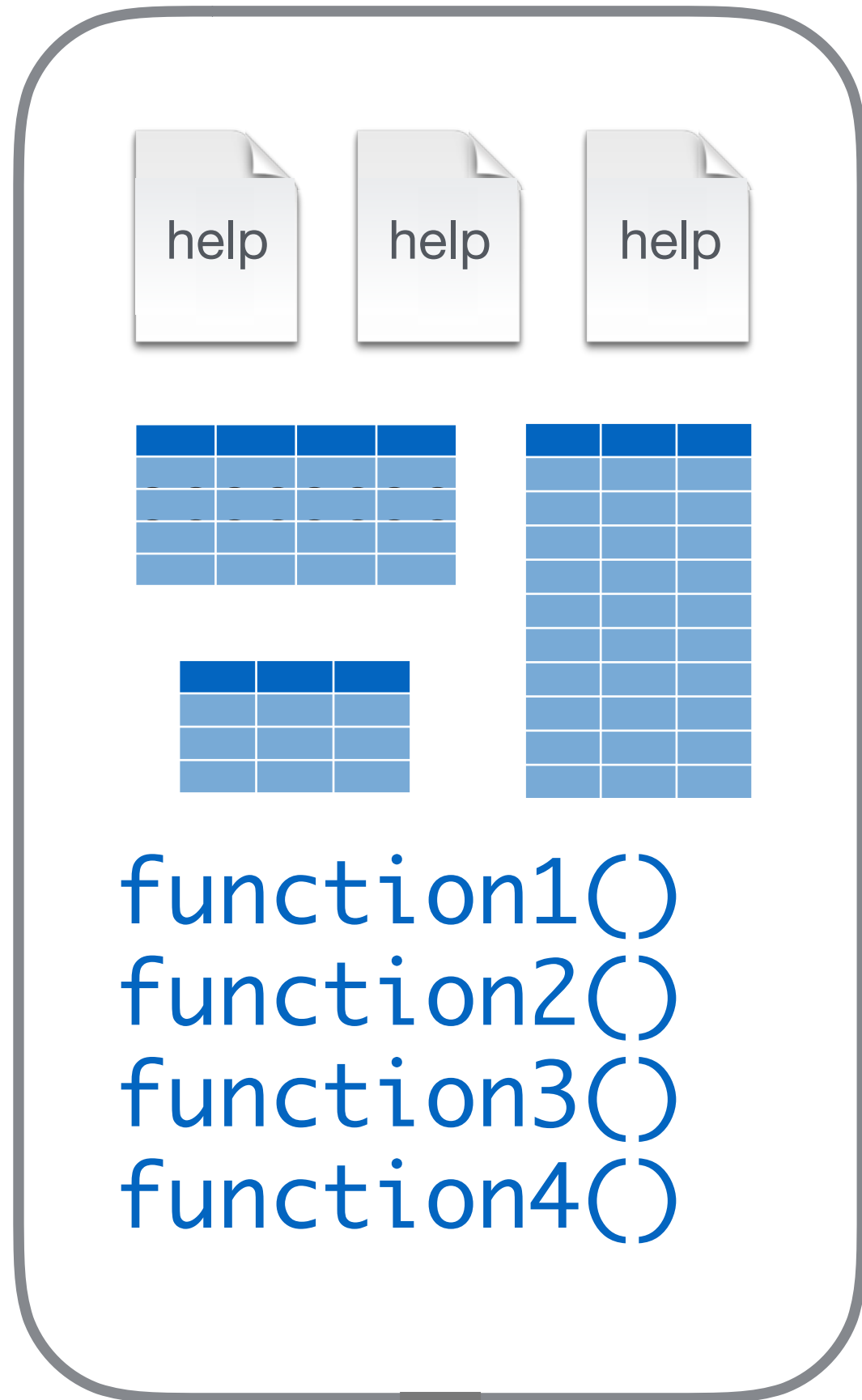
sum

SUM

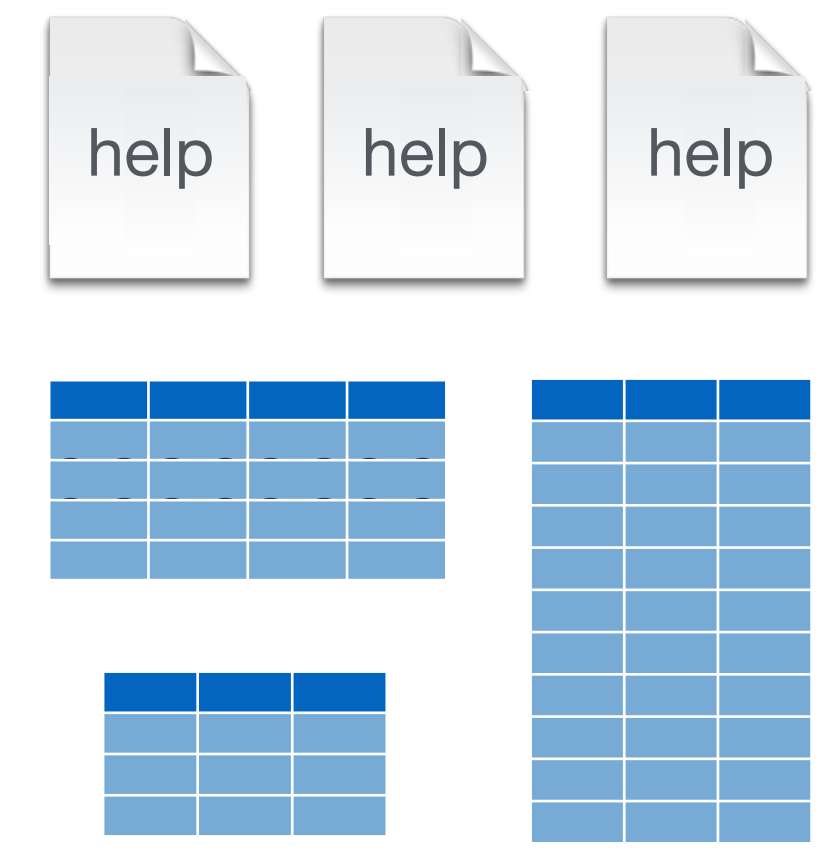
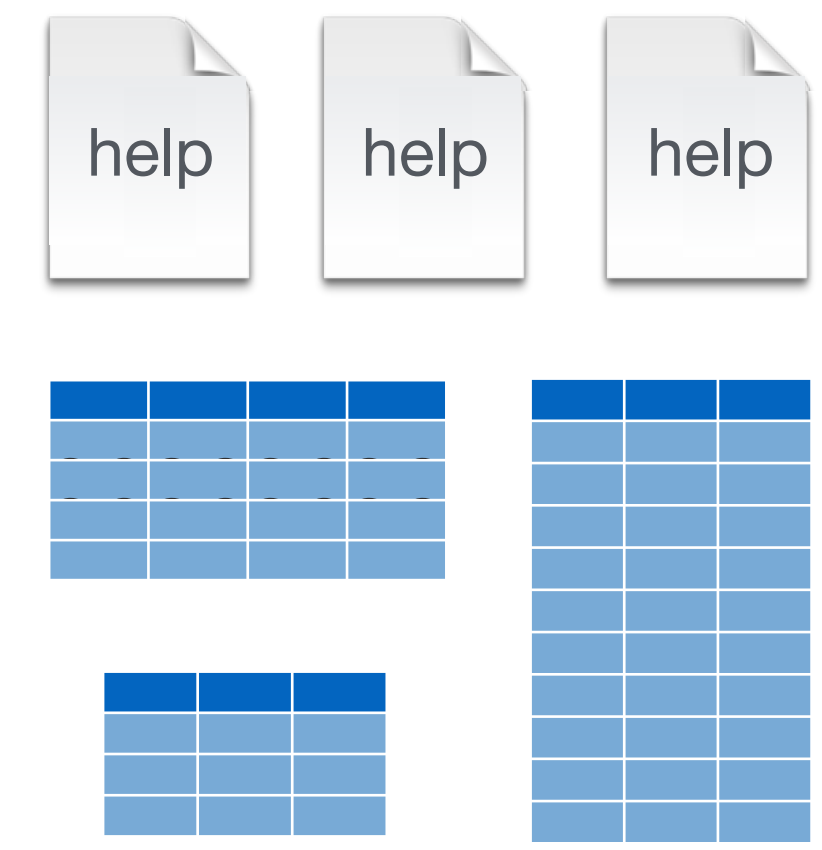
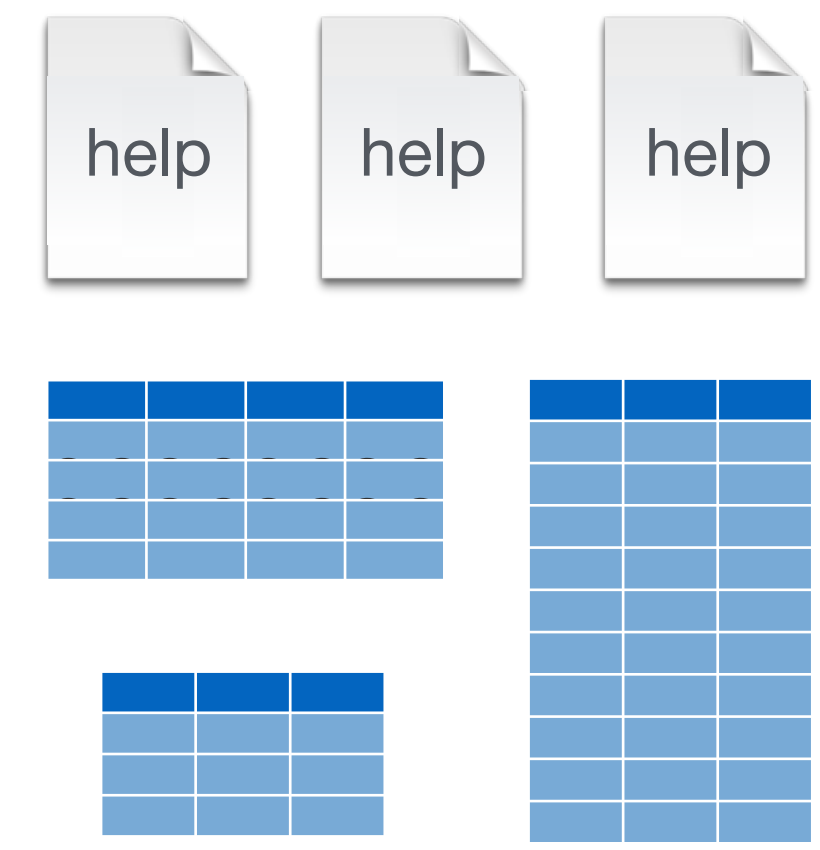
R packages

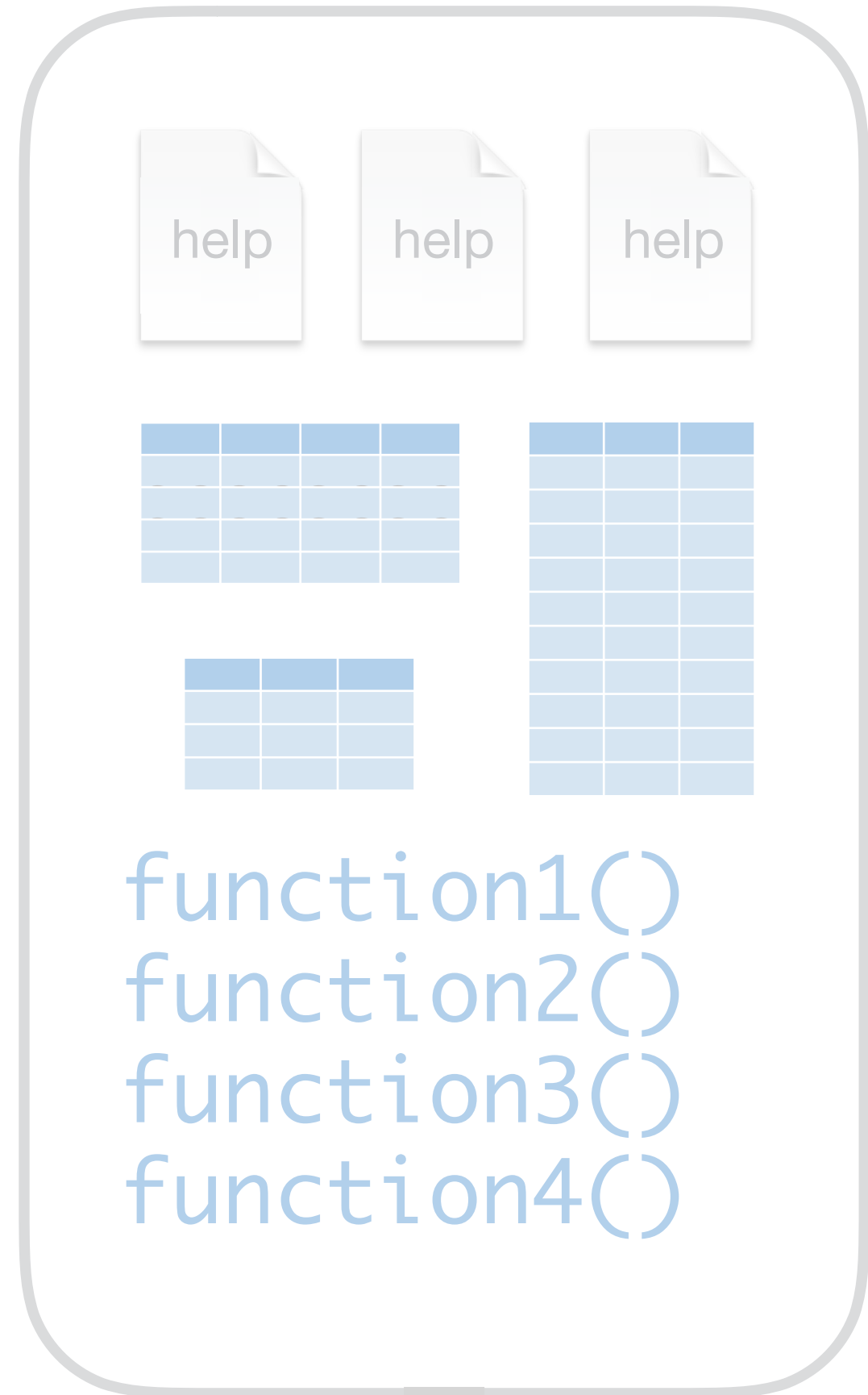


function1()
function2()
function3()
function4()

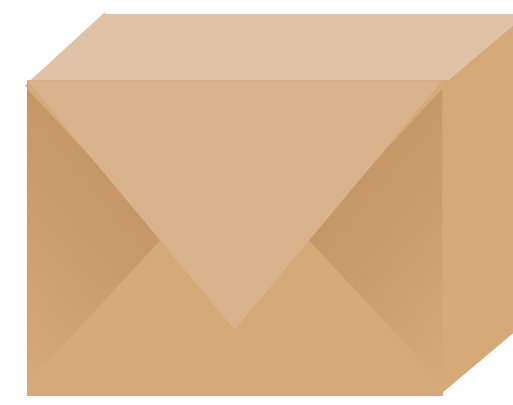
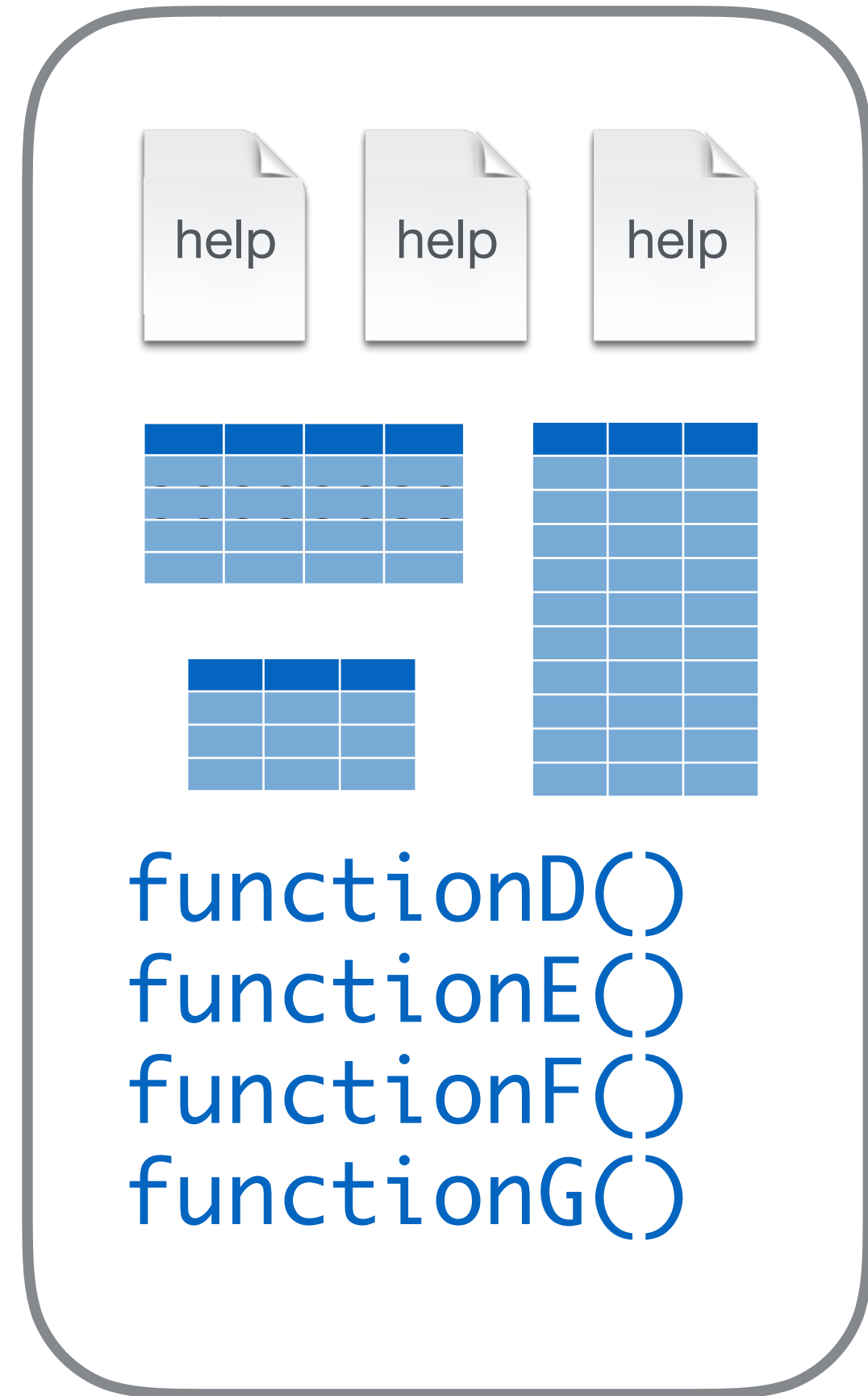
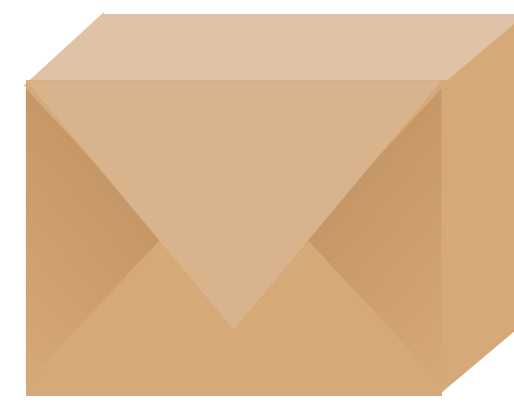
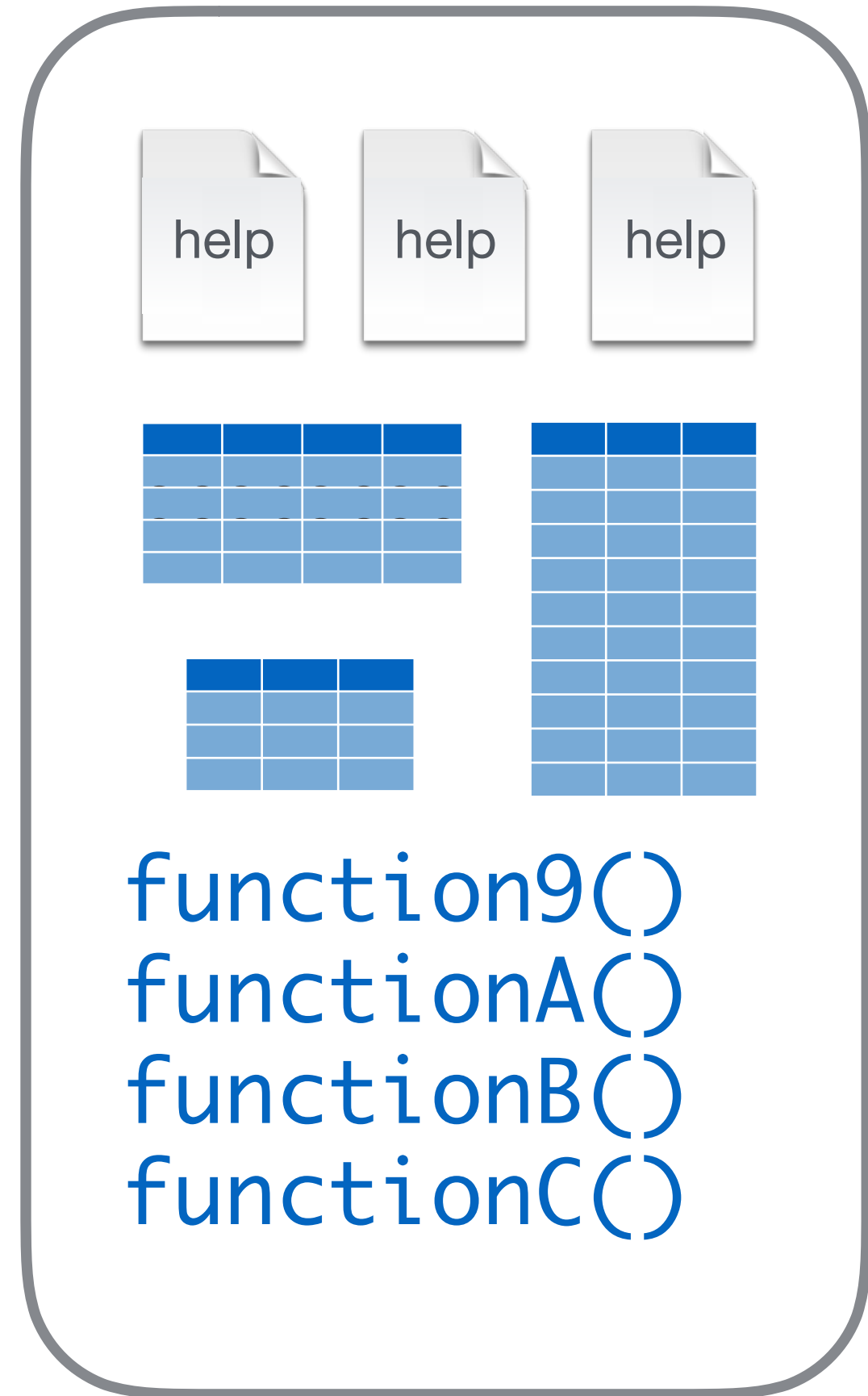
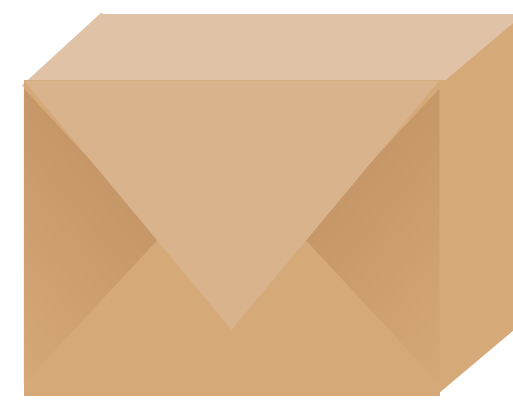
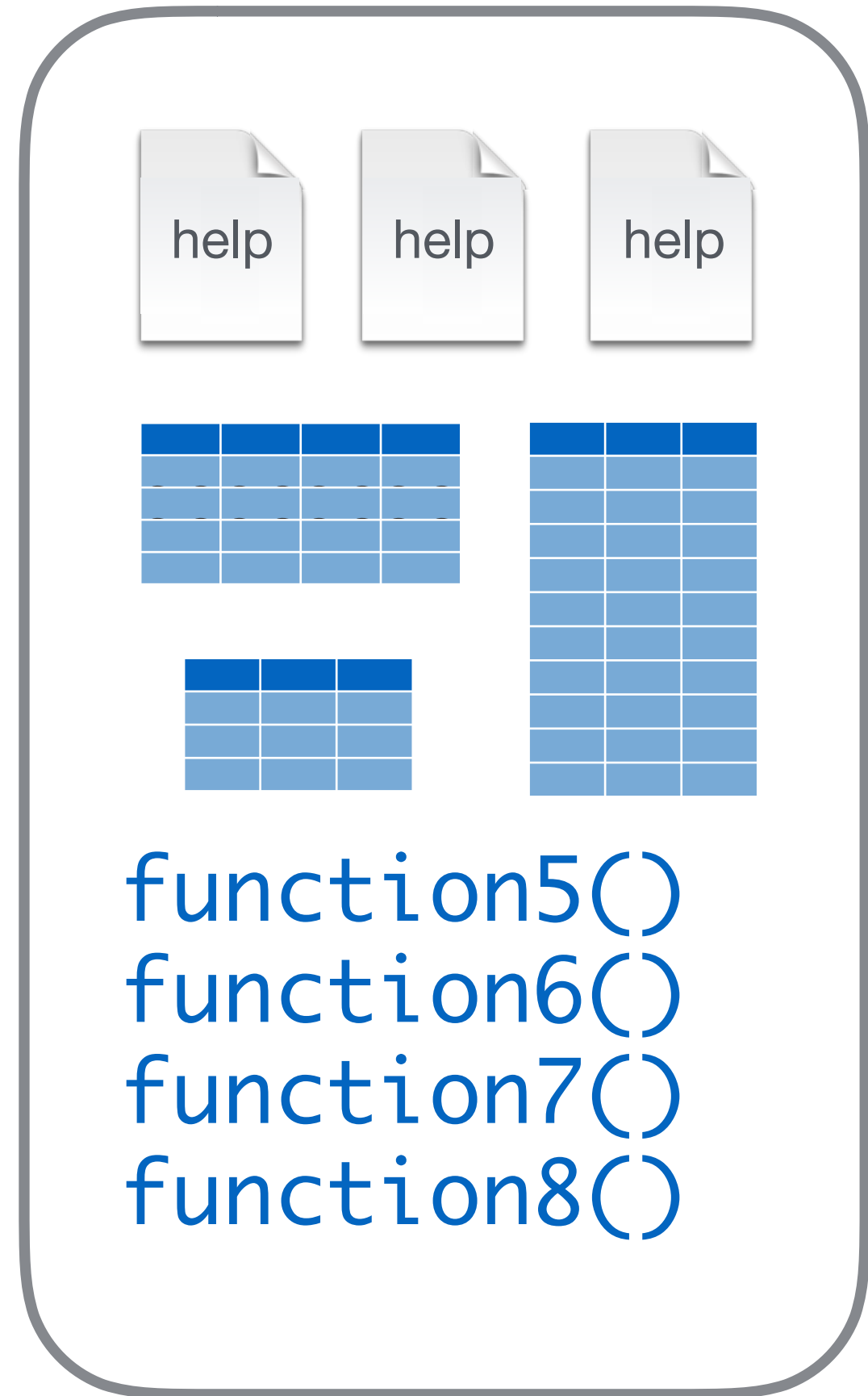


Base R






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](#)

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

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.

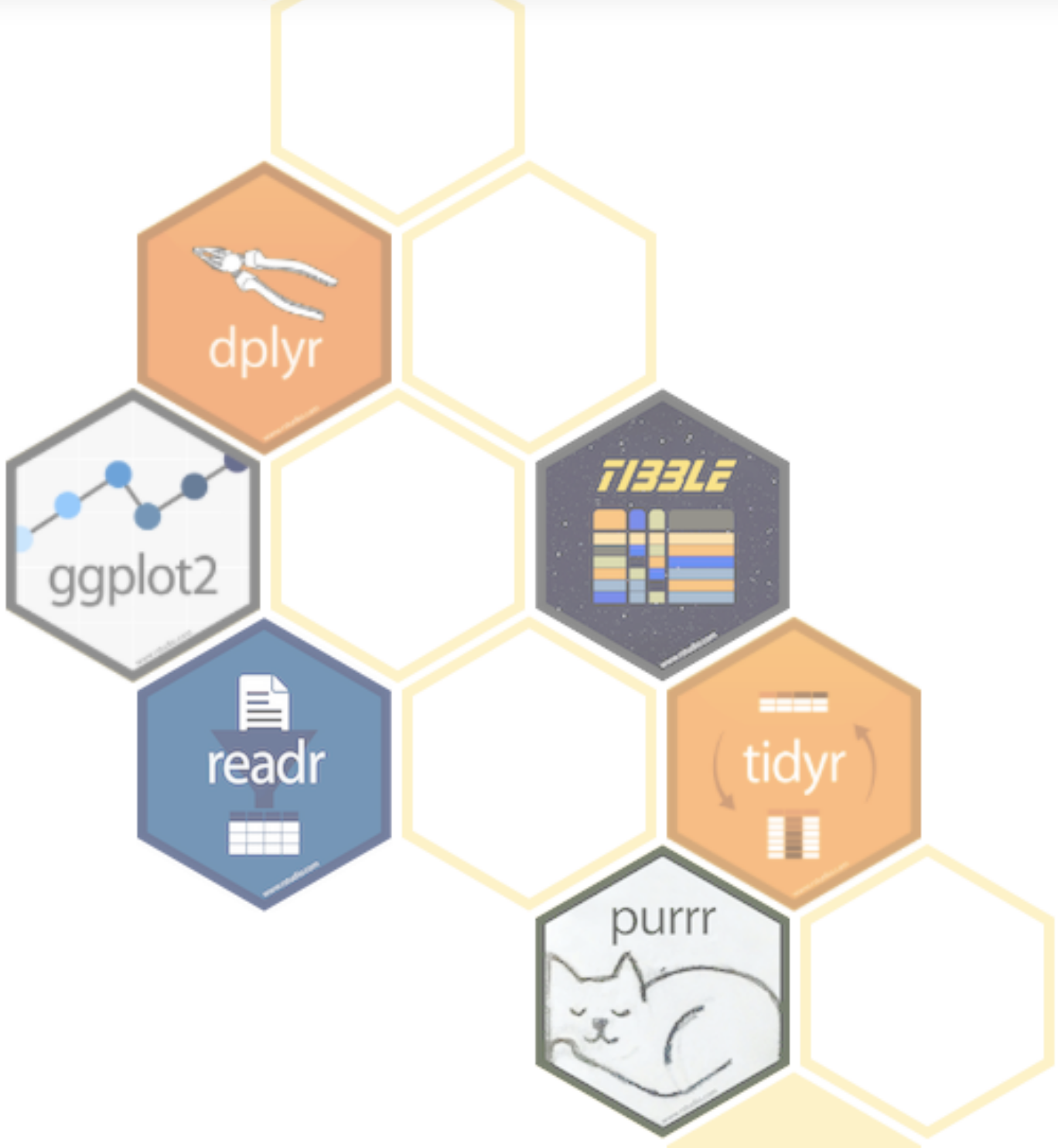
Install the complete tidyverse with:

```
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")
```

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`
`mtcars$efficient`

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::tally(cyl~am, data=mtcars)`

one continuous, one categorical:
`lattice::histogram(mpg~cyl, data=mtcars)`
`lattice::boxplot(mpg~cyl, data=mtcars)`

`lattice::boxplot(mpg~cyl, data=mtcars)`

WRANGLING:

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

making a new variable:
`mtcars$efficient = mtcars$mpg > 30`
`mtcars$efficient`

Tidyverse syntax

SUMMARY STATISTICS:

one continuous variable:
`mtcars %>% summarise(mean_mpg = mean(mpg))`

one categorical variable:
`mtcars %>% summarise(n_cyl = n_distinct(cyl))`

two categorical variables:
`mtcars %>% summarise(n_cyl_am = n_distinct(cyl, am))`

one continuous, one categorical:
`mtcars %>% summarise(mean_mpg_by_cyl = summarise(mean_mpg))`

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))`

I've given you a copy of this cheatsheet in your files

the pipe

Each R syntaxes give ways to "say" the same thing. Head across the cheatsheet to see how different syntaxes approach the same problem