

Welcome!

1. Get connected to the wifi
KUGUEST
2. Get the course materials
`usethis::use_course("http://bit.ly/tidyds19")`
3. Follow instructions in **setup.R**
4. Stuck? Please ask for help!

Tidy Data Science

Jake Thompson

 wjakethompson.com
 [@wjakethompson](https://twitter.com/wjakethompson)



HELLO
my name is

Jake



@wjakethompson

HELLO
my name is

Paul

HELLO
my name is

Jeff

HELLO
my name is

Andrew

HELLO
my name is

charlie



Hadley Wickham

Chief Scientist
RStudio

 @hadleywickham

[Data Science in the tidyverse](#)
[rstudio::conf\(2019\)](#)

Amelia McNamara

Assistant Professor
University of St. Thomas
 @AmeliaMN



Charlotte Wickham

Data Scientist, Instructor
Oregon State University
 @CVWickham

[Data Science in the tidyverse](#)
[rstudio::conf\(2018\)](#)



Garrett Grolemund

Master Instructor
RStudio
 @StatGarrett

[Master the Tidyverse](#)

Your Turn

Introduce yourself to your neighbors

- Who are you?
- What do you do with data?
- How would you describe your experience with R?



No sticky note:
“I’m happily
working on it.”



Blue sticky note:
“I’m finished and
ready to move on.”

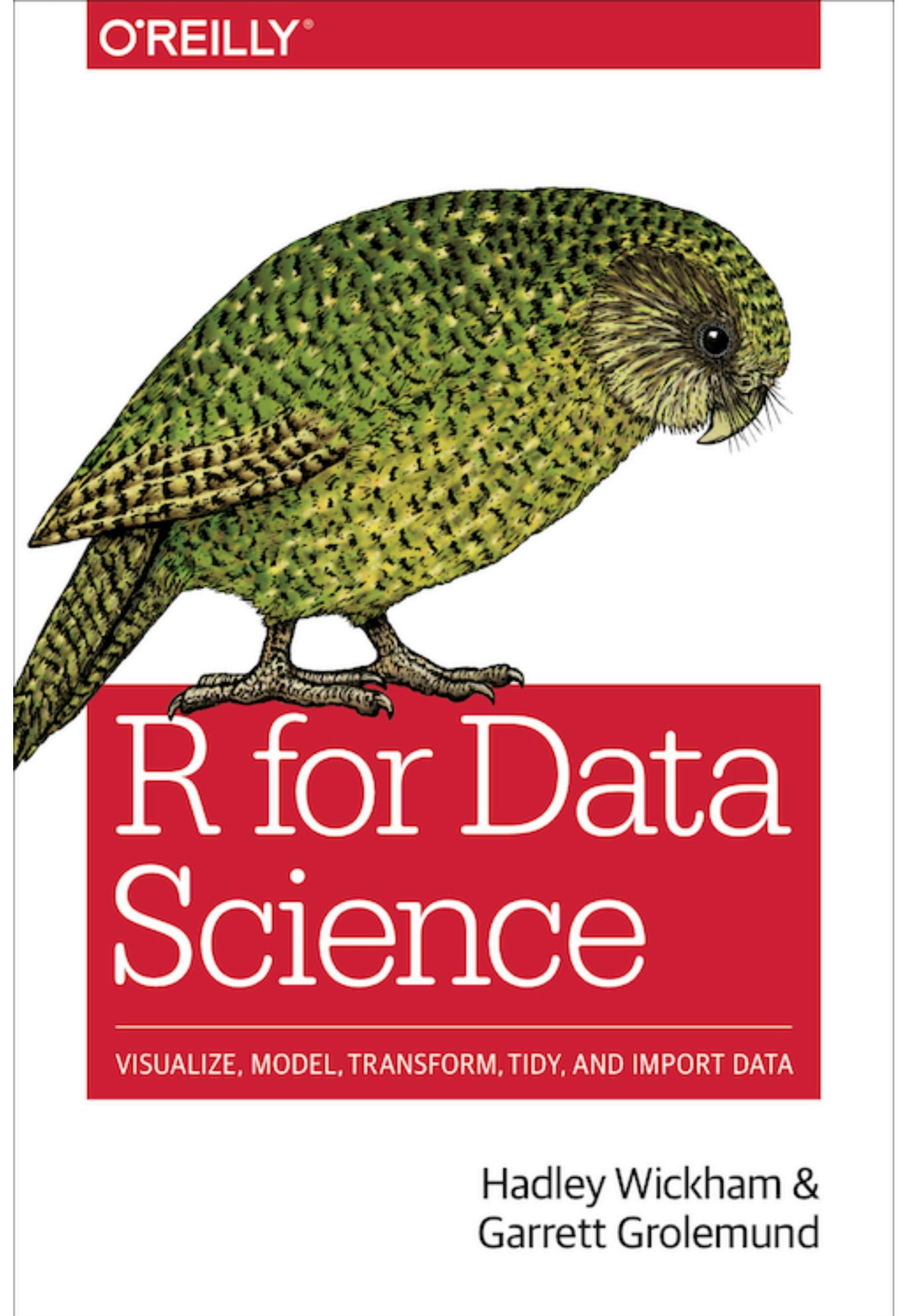


Orange sticky note: “I’m
stuck, can someone
please help me?”

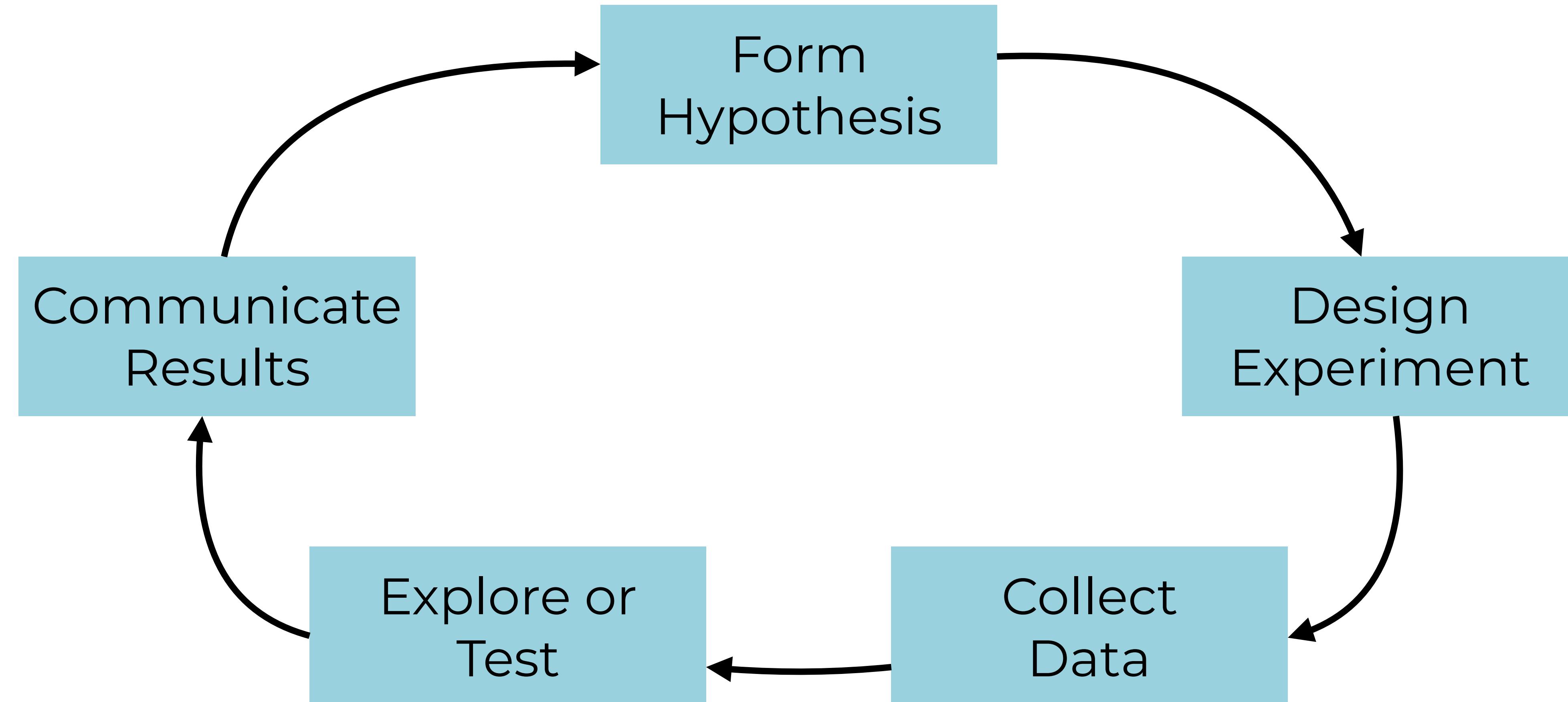


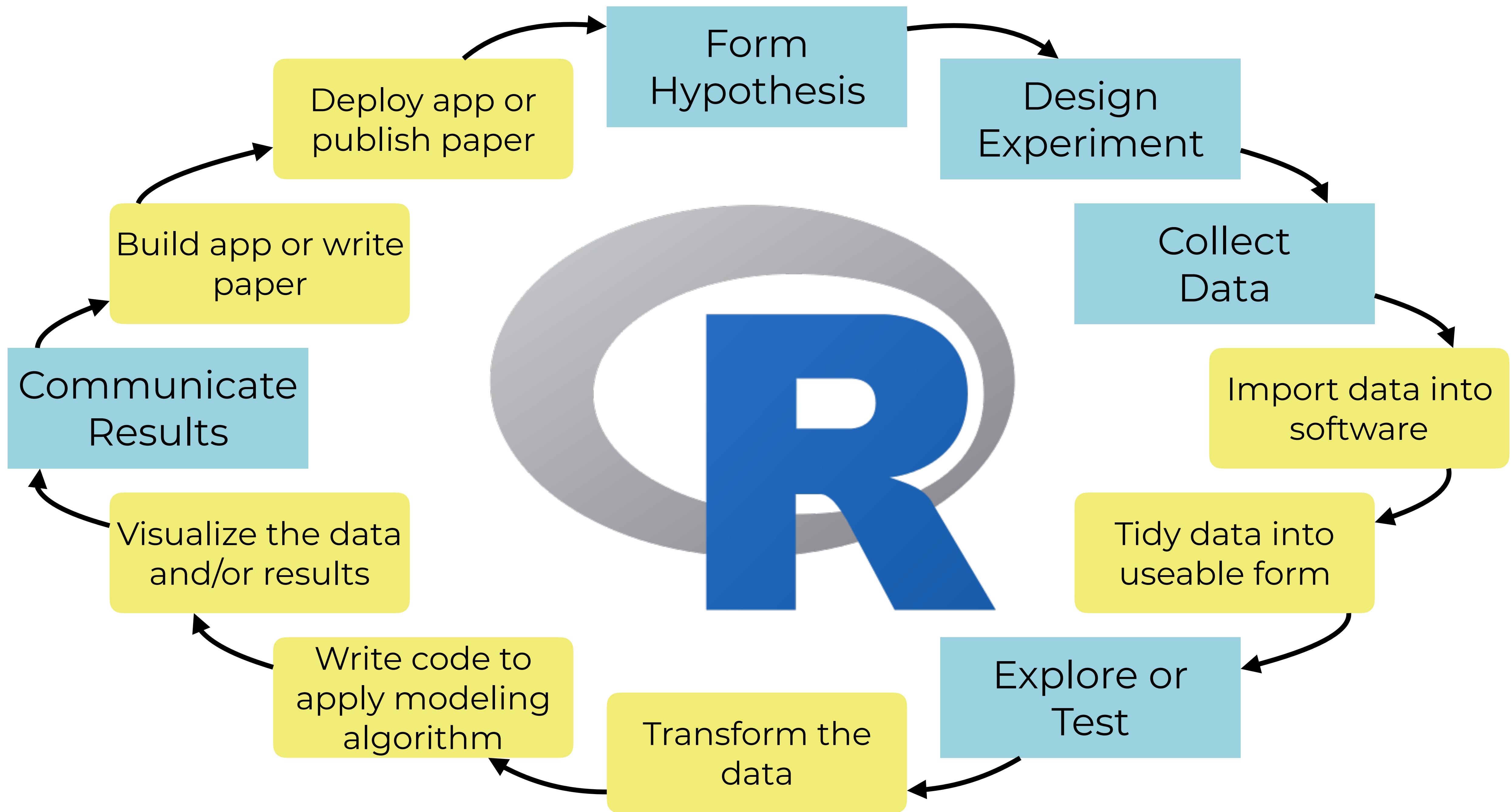
This workshop is based heavily on
R for Data Science
<https://r4ds.had.co.nz>

Links to the relevant
sections of the book

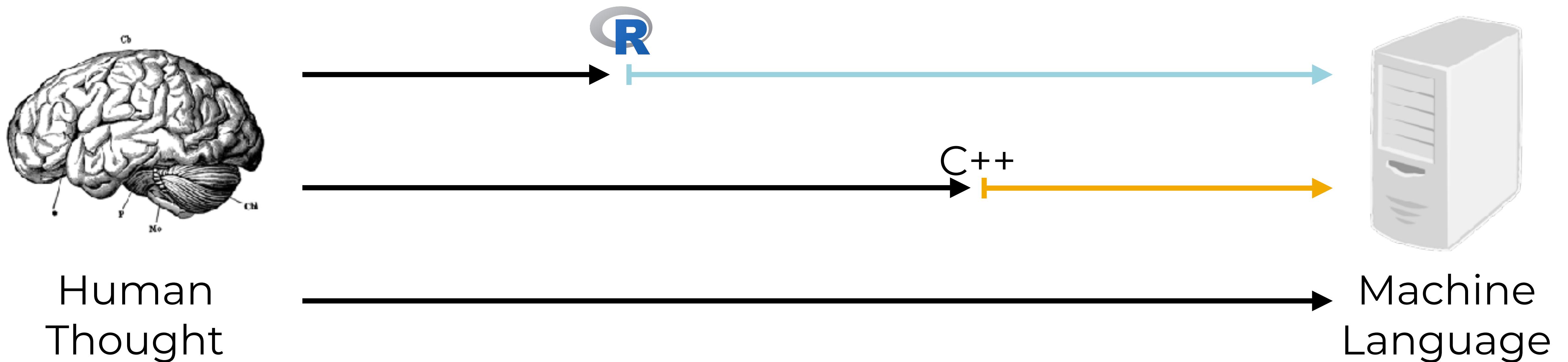


Data Science





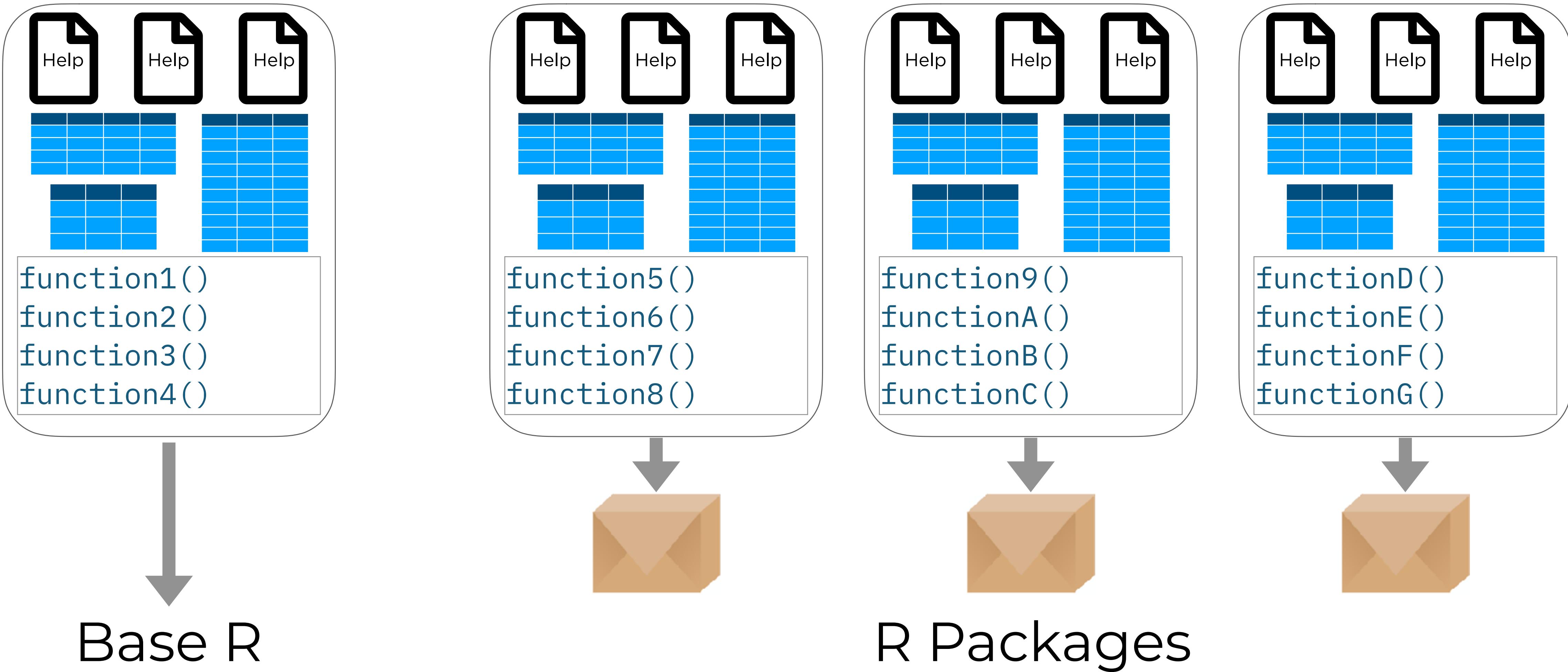
R: A computer language for scientists



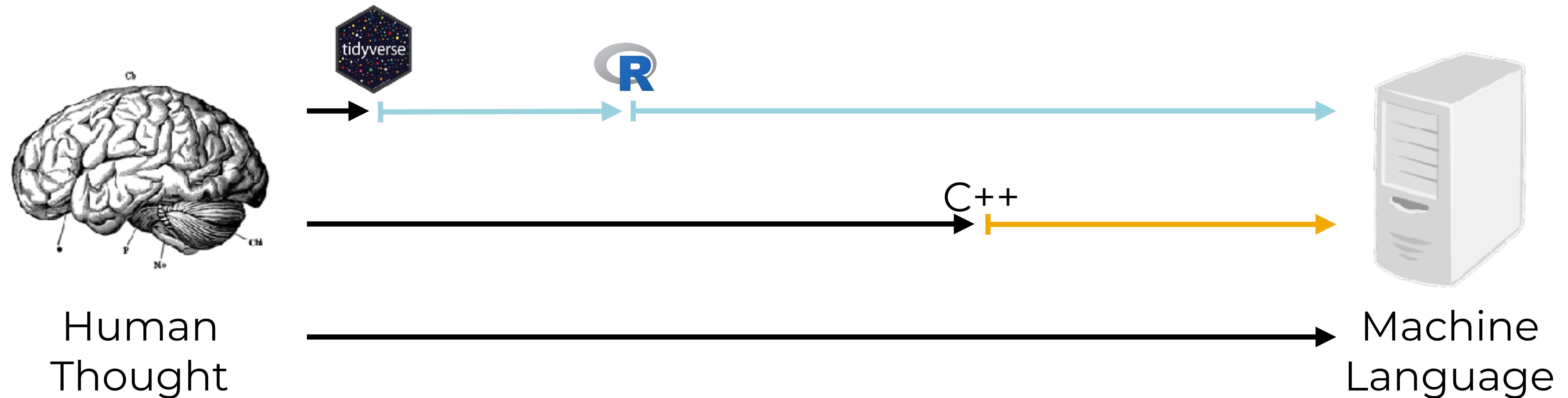
A dramatic photograph of a waterfall cascading over dark, jagged rocks into turbulent white water. The waterfall is the central focus, with its spray and mist catching some light. The background is a dark, overcast sky.

bit of success

R Packages



tidyverse: Ecosystem to unify data science tasks





Using Packages

1

```
install.packages("foo")
```

Downloads files to computer

1 x per Computer

2

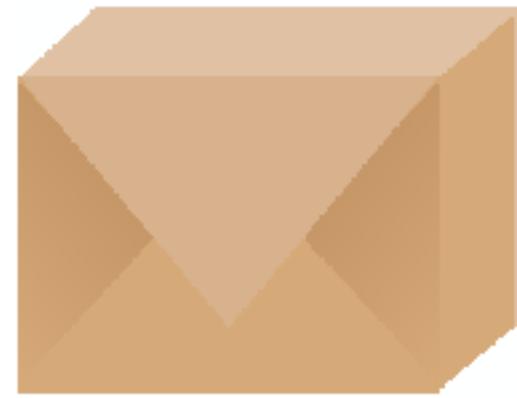
```
library("foo")
```

Loads package

1 x per R Session



tidyverse



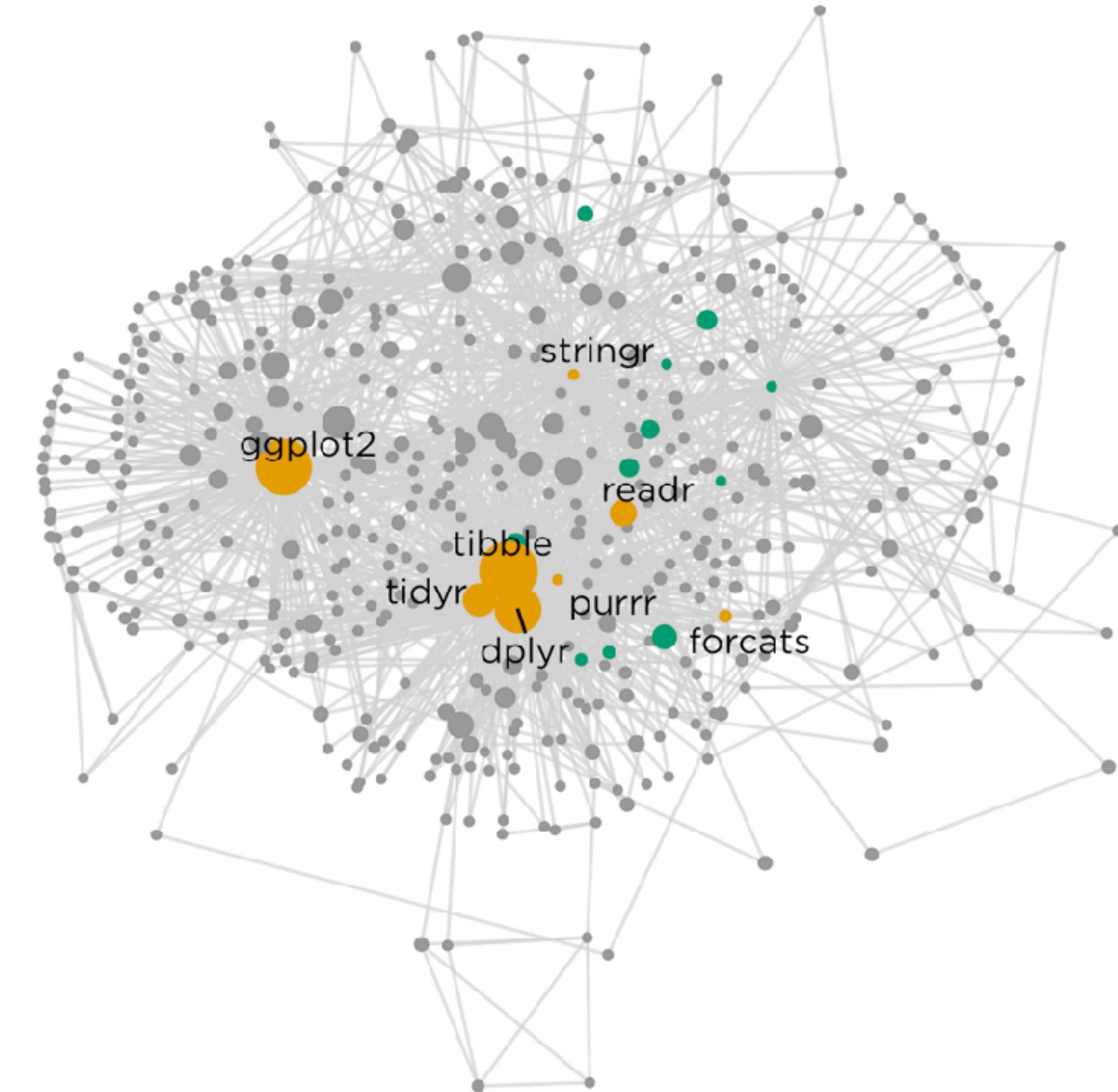
An R package that serves as a short cut for installing and loading components of the tidyverse.

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

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

does the equivalent of

```
install.packages("ggplot2")
install.packages("tibble")
install.packages("tidyr")
install.packages("readr")
install.packages("purrr")
install.packages("dplyr")
install.packages("stringr")
install.packages("forcats")
install.packages("broom")
install.packages("dbplyr")
install.packages("haven")
install.packages("hms")
install.packages("httr")
install.packages("jsonlite")
install.packages("lubridate")
install.packages("magrittr")
install.packages("modelr")
install.packages("readxl")
install.packages("reprex")
install.packages("rlang")
...
...
```



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

does the equivalent of

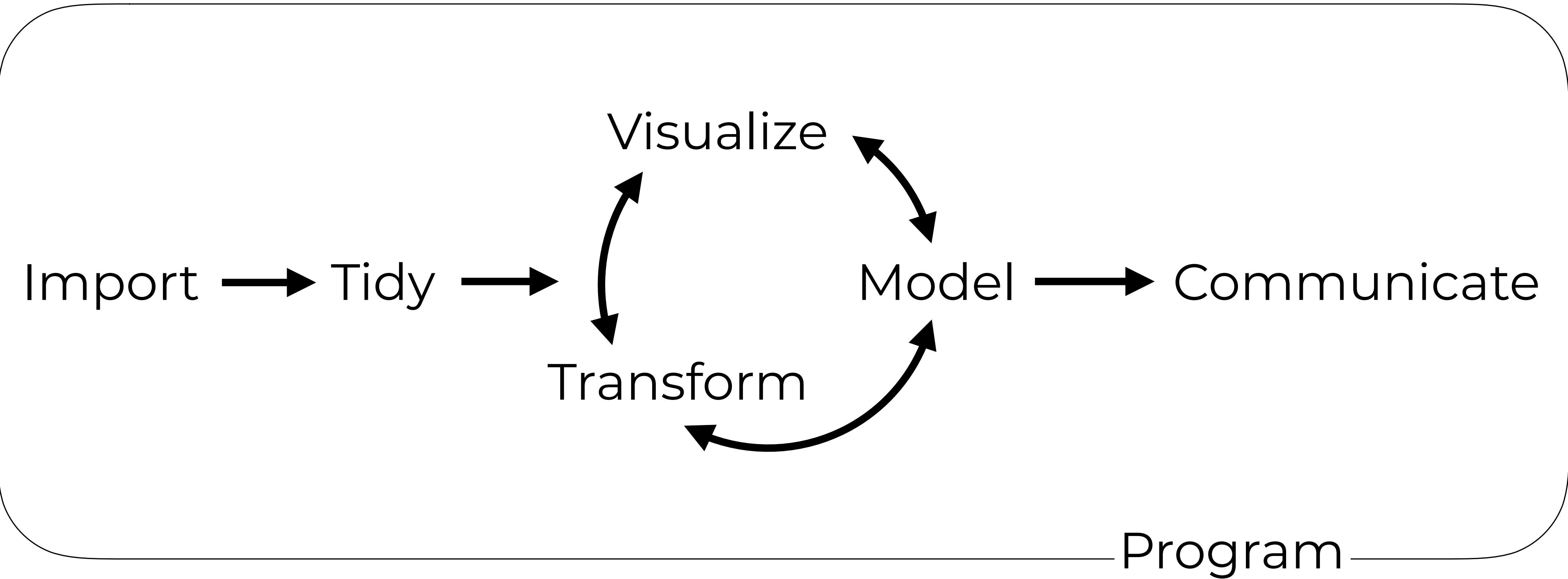
```
install.packages("ggplot2")
install.packages("tibble")
install.packages("tidyr")
install.packages("readr")
install.packages("purrr")
install.packages("dplyr")
install.packages("stringr")
install.packages("forcats")
install.packages("broom")
install.packages("dbplyr")
install.packages("haven")
install.packages("hms")
install.packages("httr")
install.packages("jsonlite")
install.packages("lubridate")
install.packages("magrittr")
install.packages("modelr")
install.packages("readxl")
install.packages("reprex")
install.packages("rlang")
...
```

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

does the equivalent of

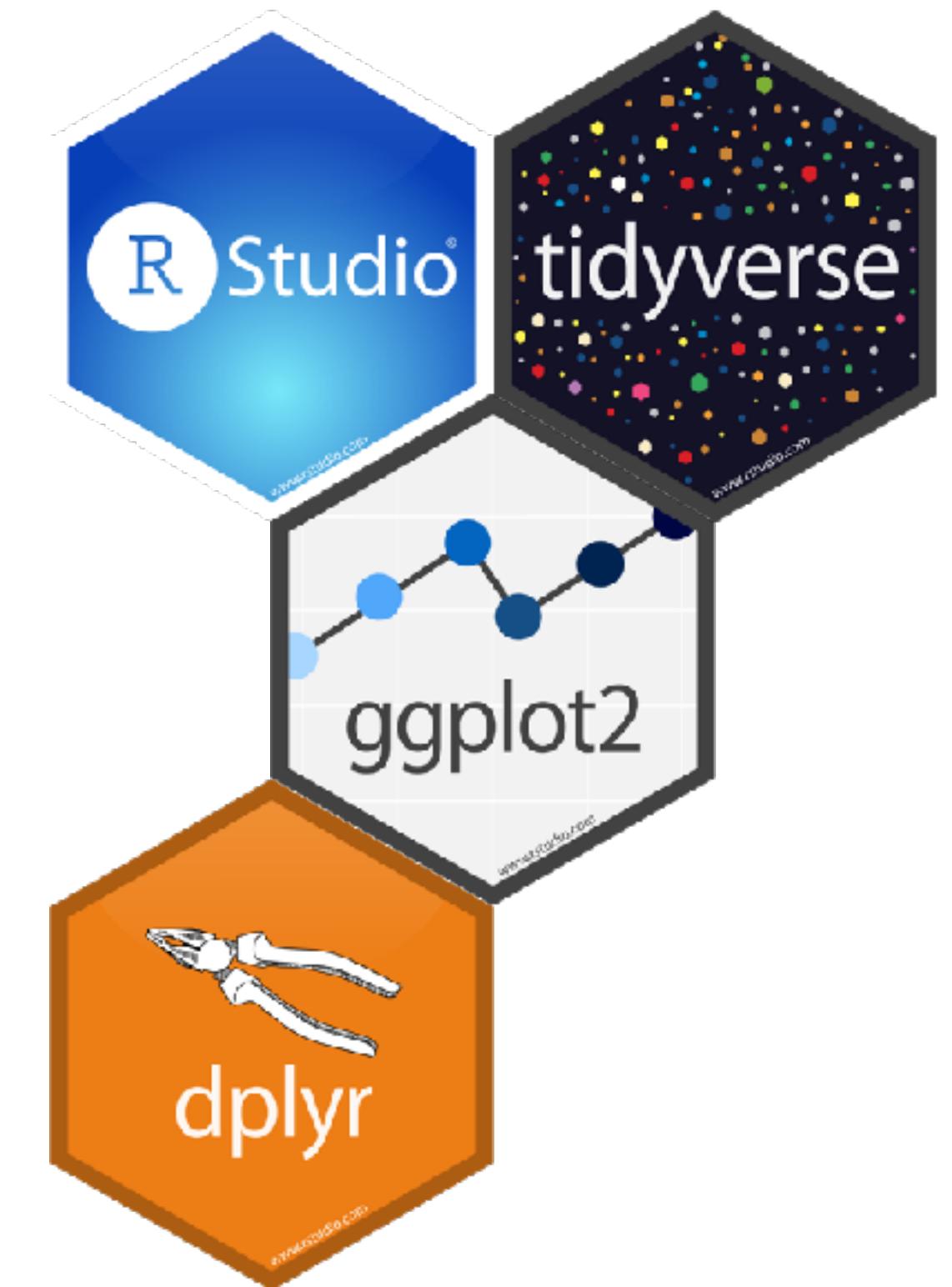
```
library("ggplot2")
library("tibble")
library("tidyr")
library("readr")
library("purrr")
library("dplyr")
library("stringr")
library("forcats")
```





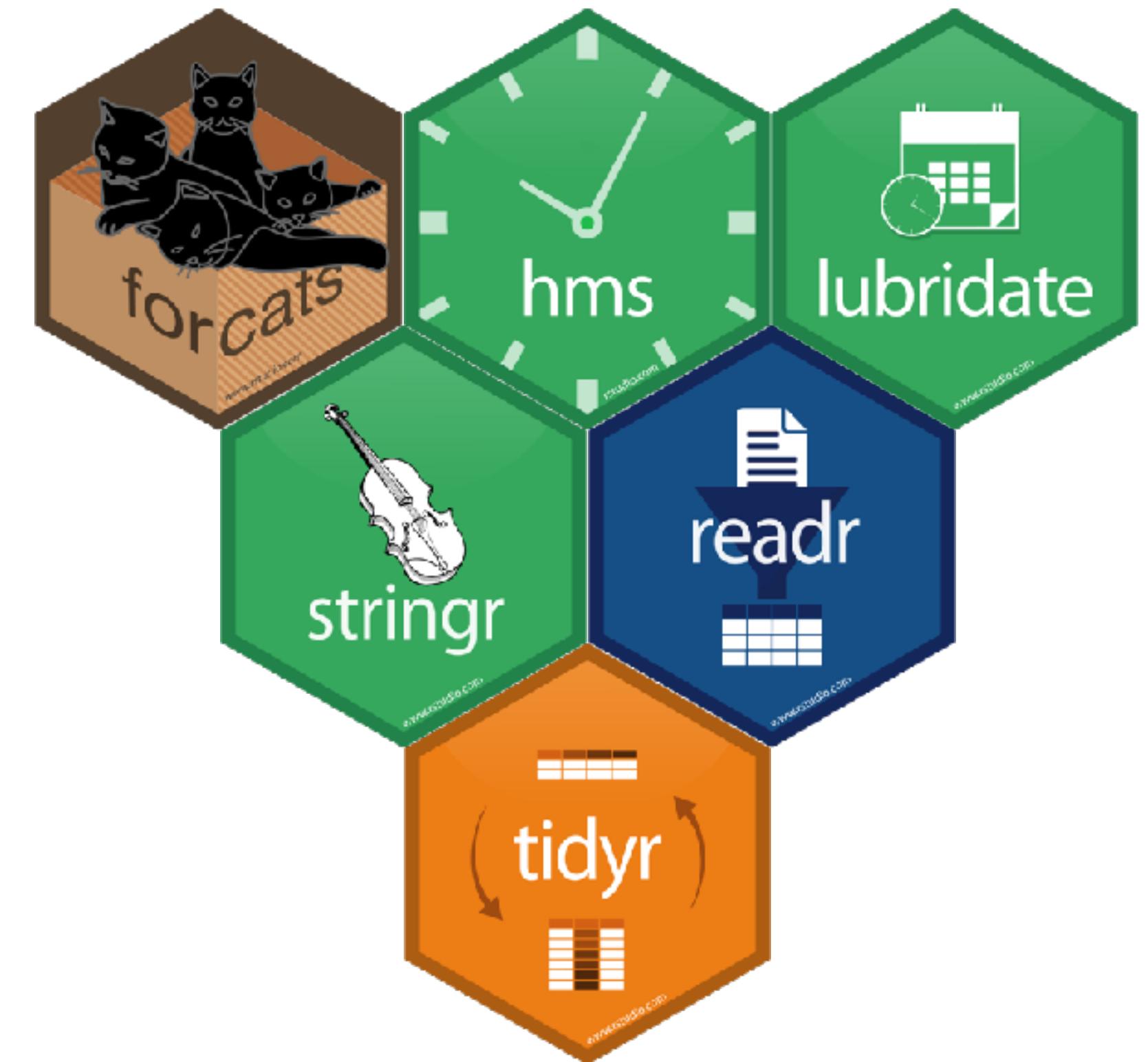
Day 1

9:30-10:15	Introduction to R and RStudio
10:15-10:30	Morning Break
10:30-12:00	Data Visualization
12:00-1:30	Lunch
1:30-2:30	Data Manipulation
2:30-2:45	Afternoon Break
2:45-4:00	Data Manipulation



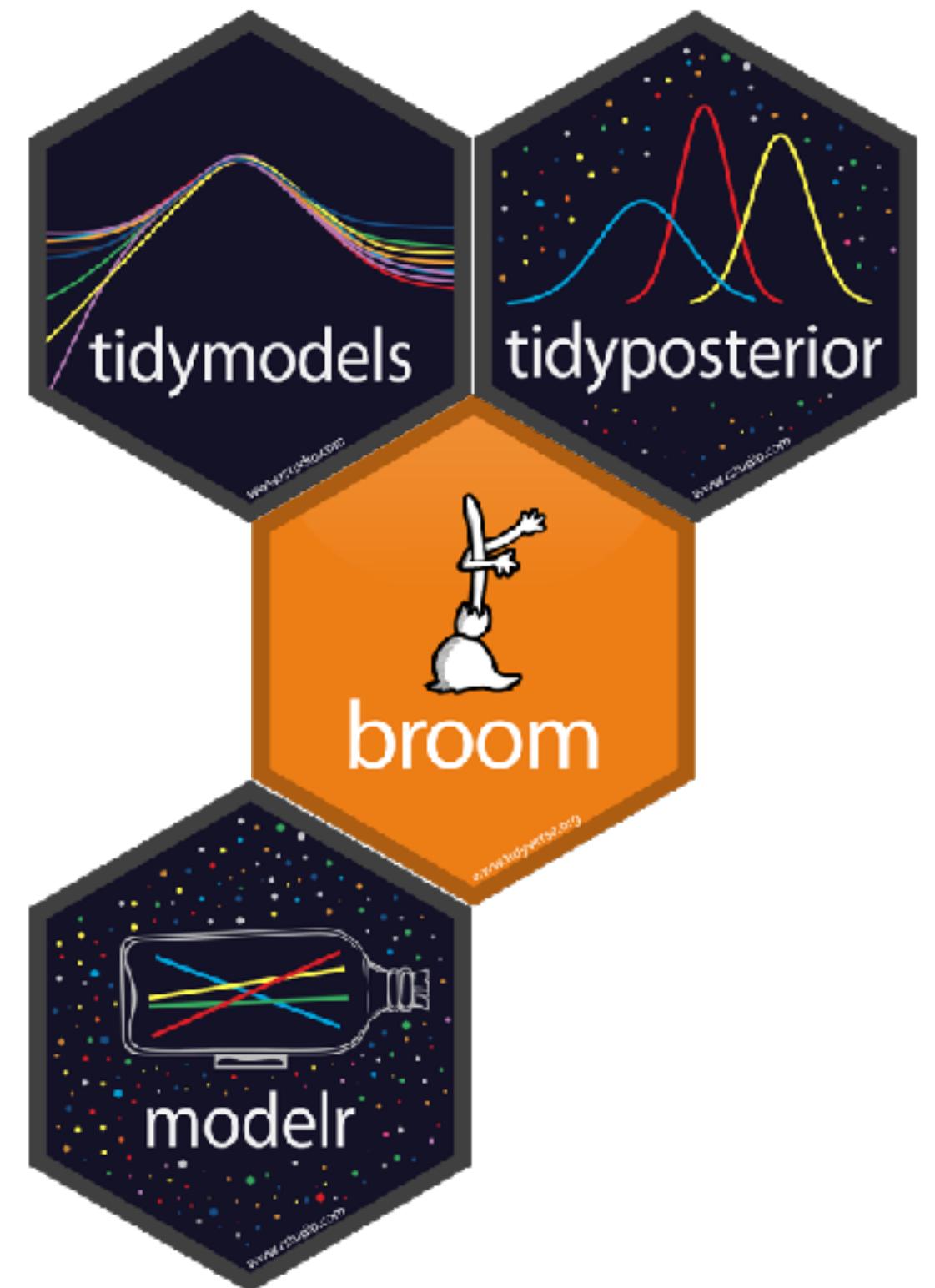
Day 2

9:00-10:15	Data Types
10:15-10:30	Morning Break
10:30-12:00	Import Data
12:00-1:30	Lunch
1:30-2:45	Tidy Data
2:45-3:00	Afternoon Break
3:00-4:00	Tidy Data



Day 3

9:00-10:00	Case Study
10:00-10:15	Morning Break
10:15-12:00	Tidying Models
12:00-1:30	Lunch
1:30-2:30	Visualizing Models
2:30-2:45	Afternoon Break
2:45-4:00	Visualizing Models



Day 4

9:00-10:15

Iteration

10:15-10:30

Morning Break

10:30-12:00

Writing Functions

12:00-1:30

Lunch

1:30-2:30

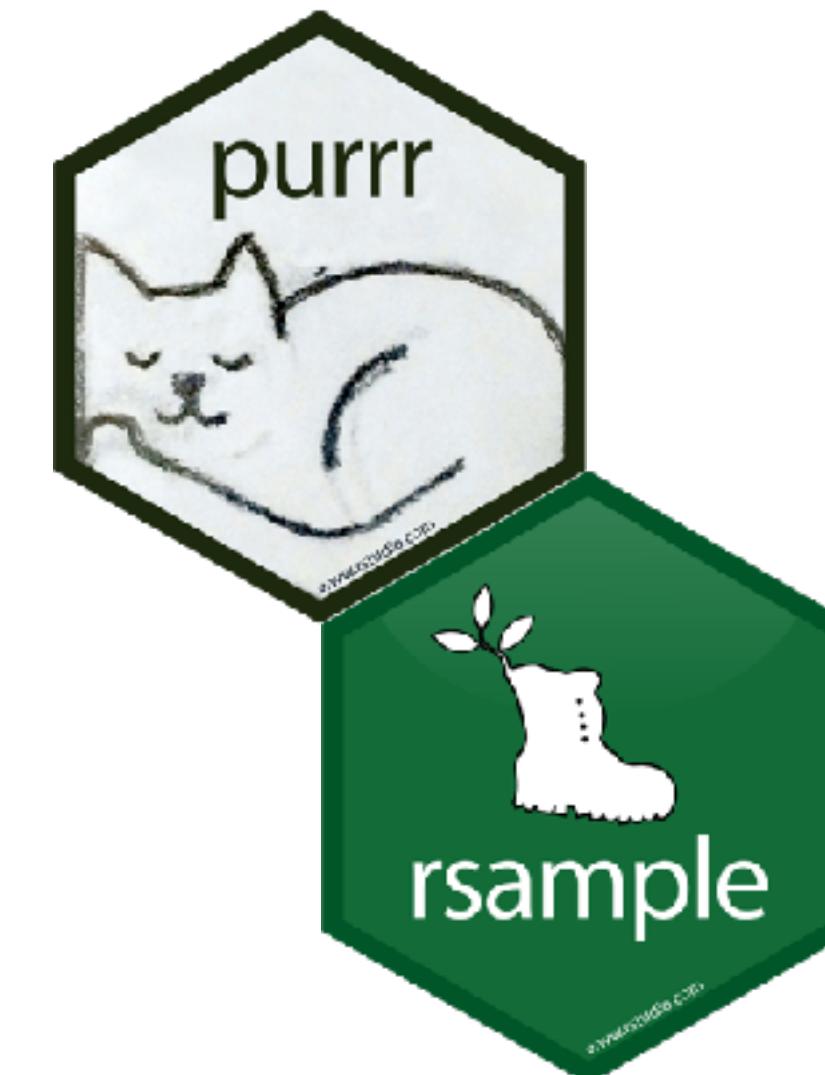
Multiple Models

2:30-2:45

Afternoon Break

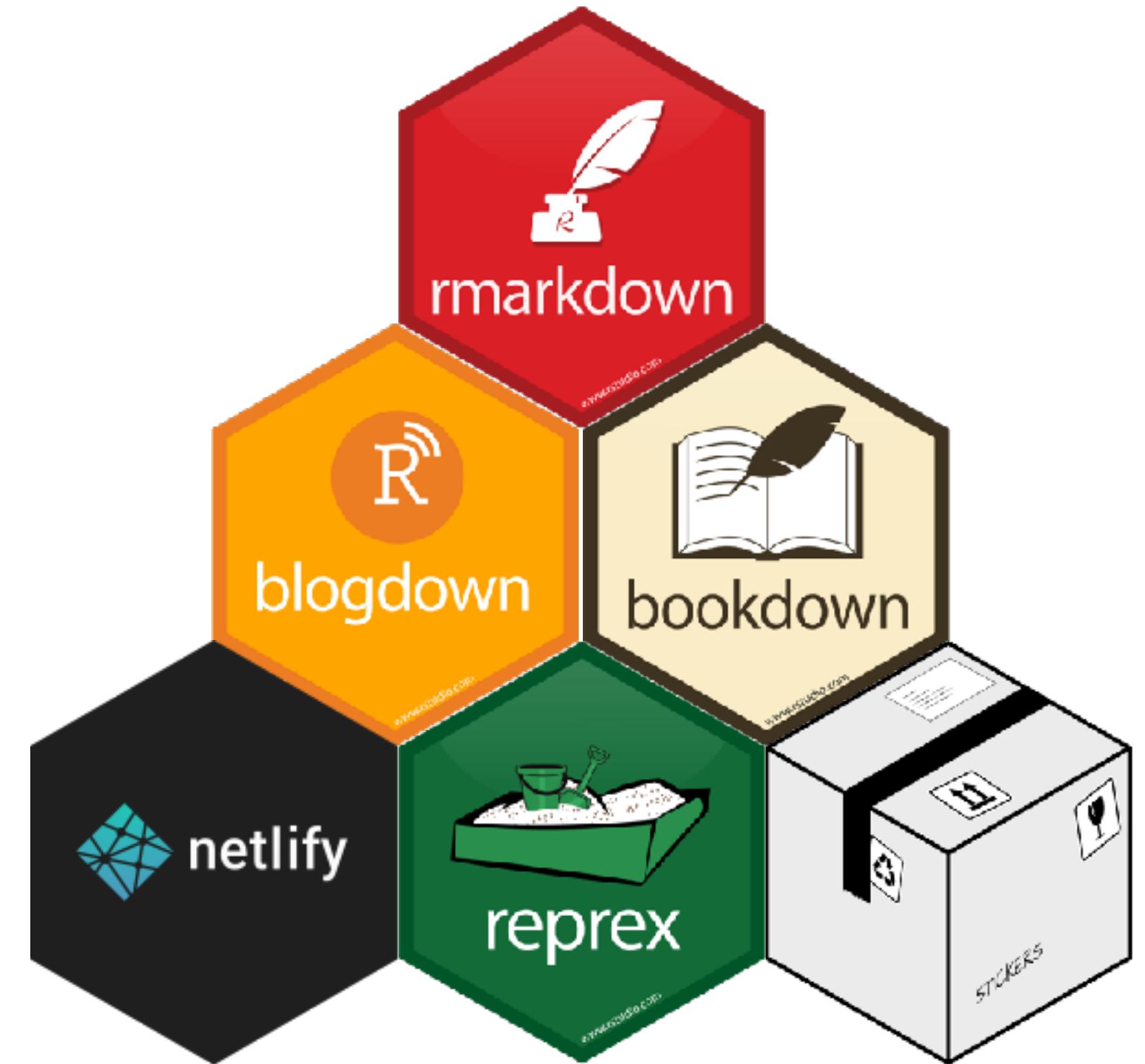
2:45-4:00

Case Study



Day 5

9:00-10:15	Introduction to R Markdown
10:15-10:30	Morning Break
10:30-12:00	Extending R Markdown
12:00-1:30	Lunch
1:30-2:30	TBD
2:30-2:45	Afternoon Break
2:45-4:00	TBD





RStudio

- Software program
- IDE: Integrated Development Environment
 - Write R code
 - Run R code
 - Analyze data with R
- Text editor, version control, debugging tools, and more...



Your Turn

Instructions with screenshots at

From R

`usethis::use_course(bit.ly/)`

Download zip file

From the Web

github.com/wjakethompson/tidyds-2019

Download zip file

Then, open **00-Getting-started.Rmd** and look around.



R Notebooks

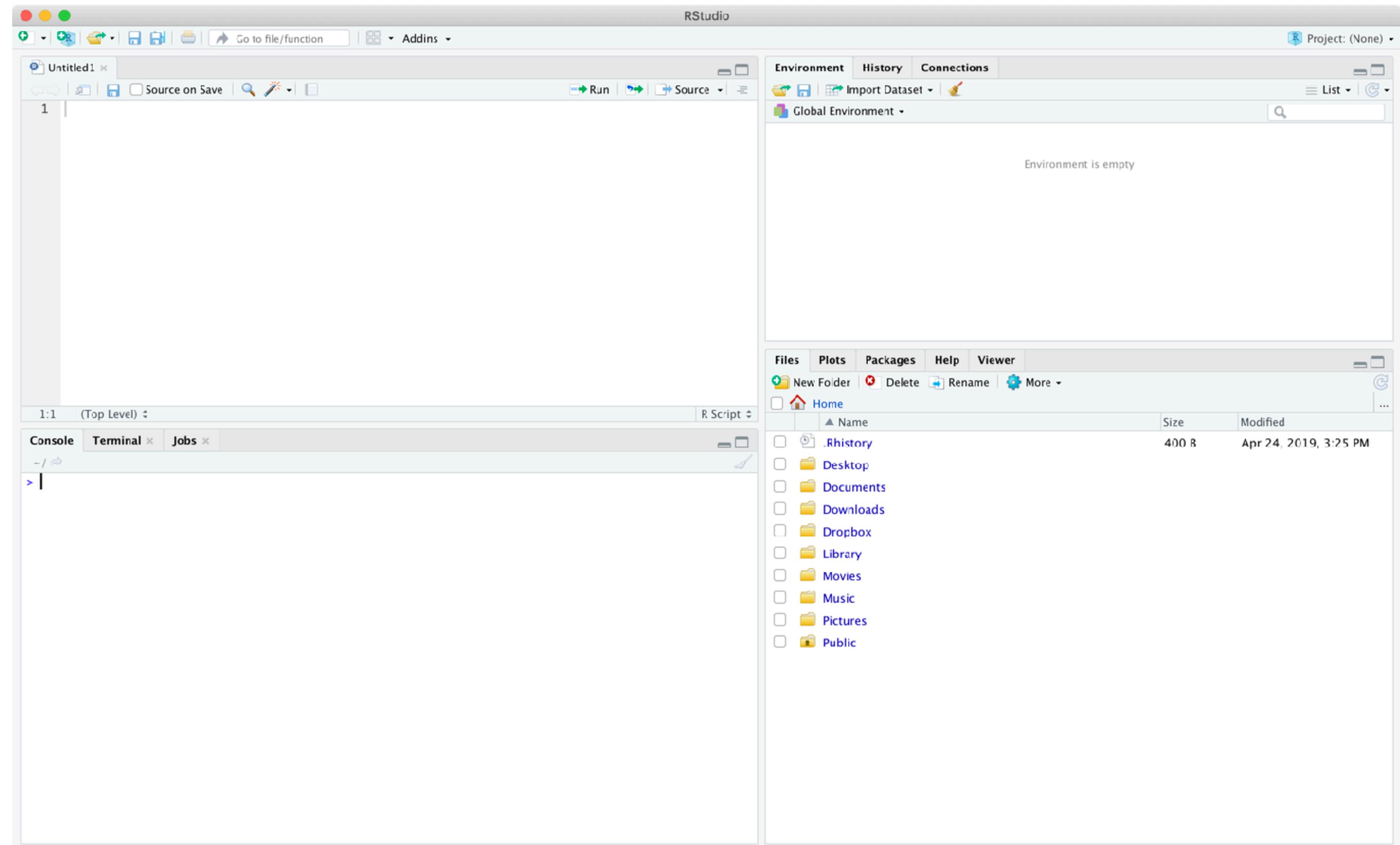
The screenshot shows the RStudio interface with an R Notebook open. The notebook contains the following Rmd code:

```
1 ---  
2 title: "R Notebook"  
3 output: html_notebook  
4 ---  
5  
6 Text written in **markdown**  
7  
8 ```{r}  
9 # code written in R  
10 (x <- rnorm(7))  
11 ```  
12  
13 Text written in _markdown_  
14  
15 ```{r}  
16 # code written in R  
17 hist(x)  
18```  
18:4 (Top Level) ▾  
Console Terminal Jobs  
16:20 C Chunk 2 ▾ R Markdown
```

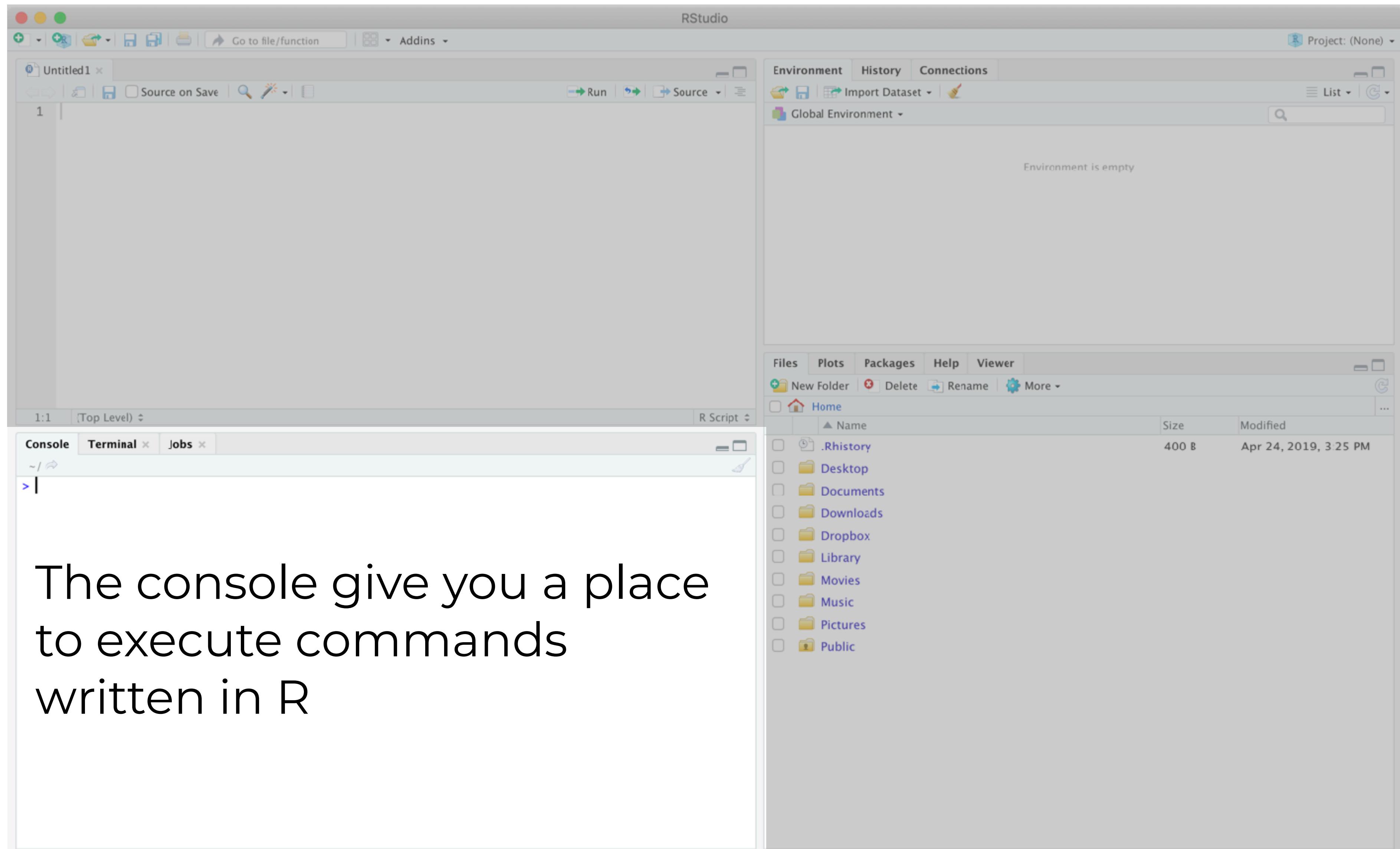
Annotations with arrows point to specific elements:

- A callout bubble points to the green triangle icon in the toolbar above the code editor, labeled "Click to run all code chunks above".
- A callout bubble points to the green triangle icon in the chunk header bar, labeled "Click to run code in a chunk".
- A callout bubble points to the code result area below the chunks, labeled "Code result".

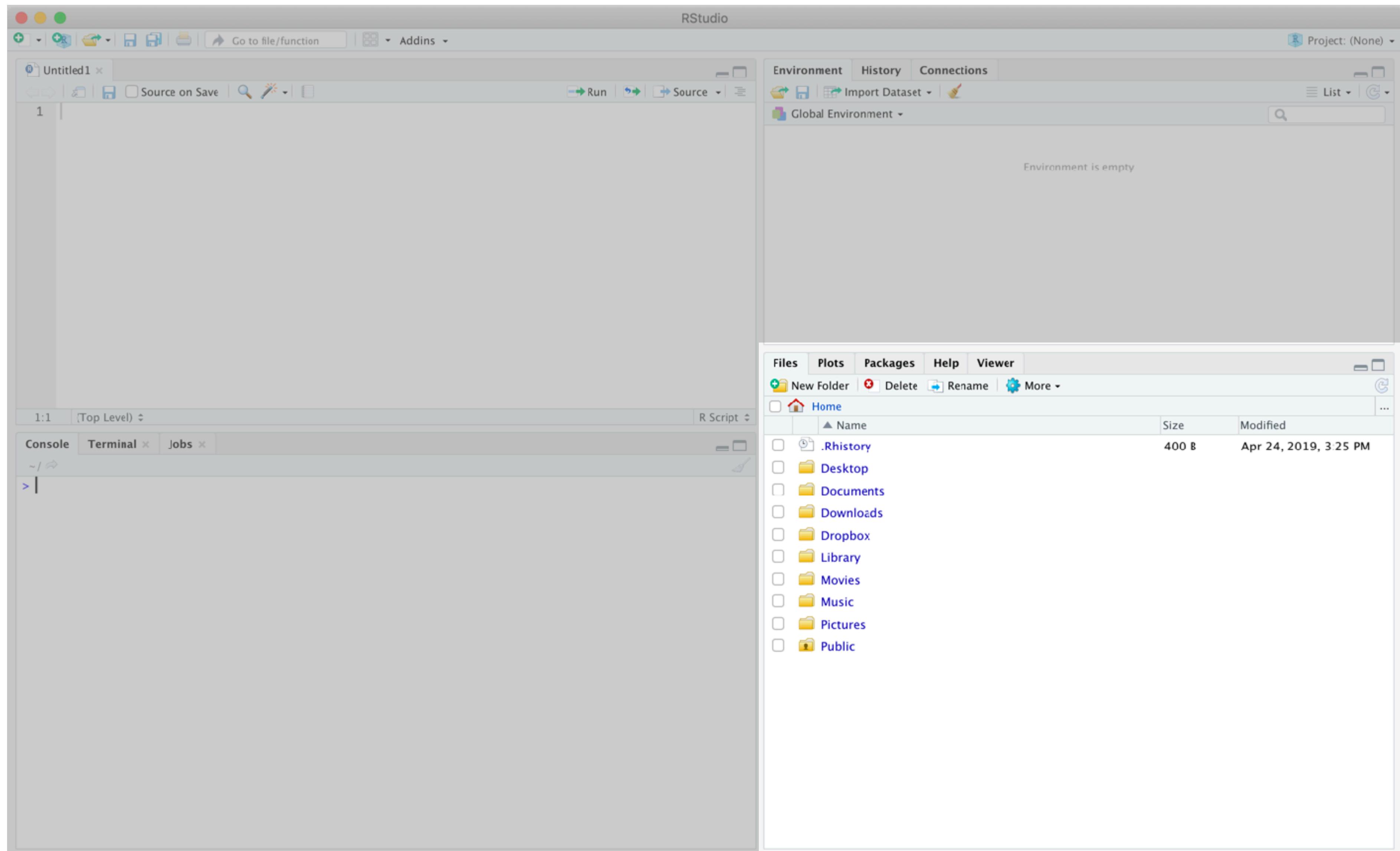
RStudio



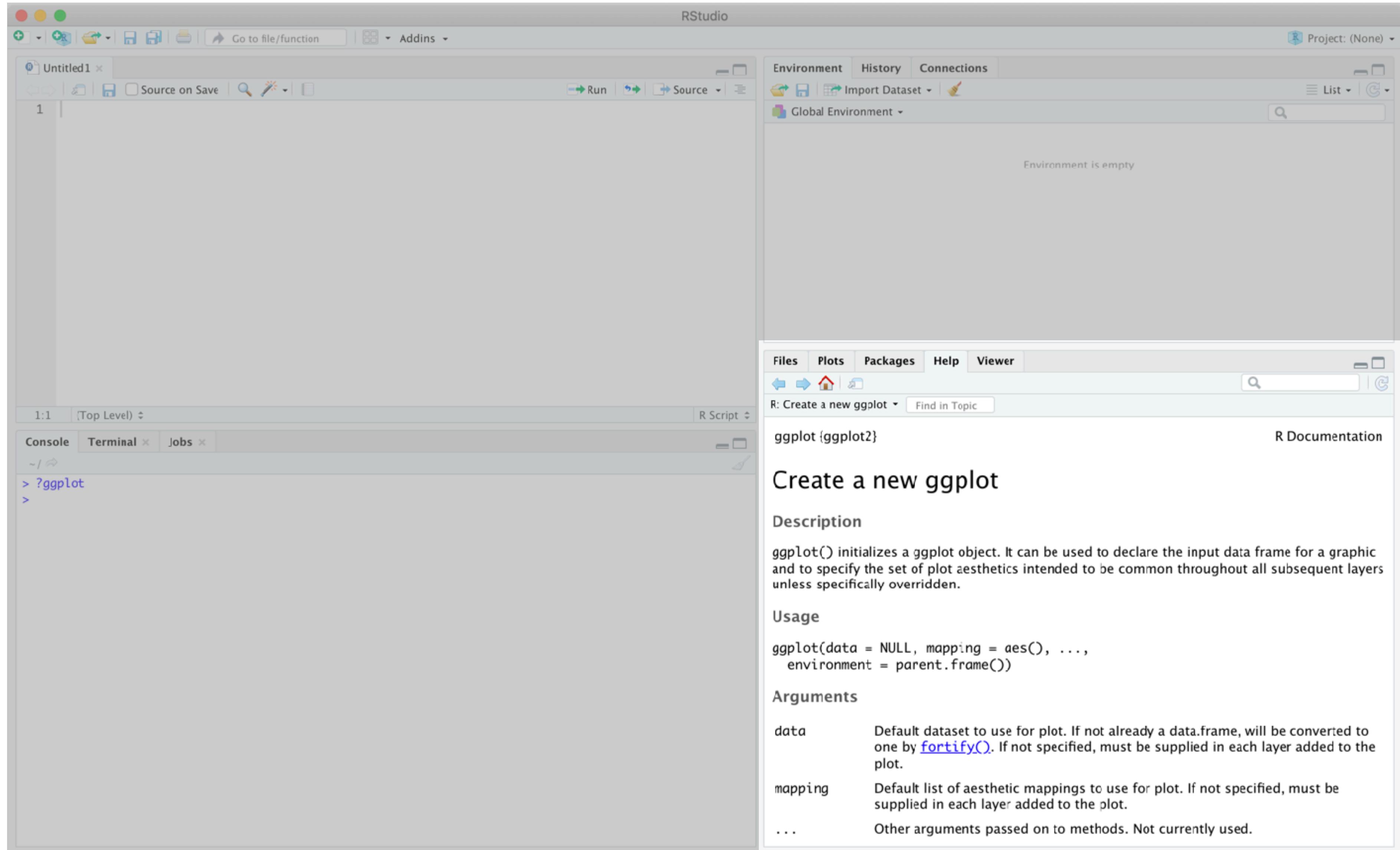
RStudio



RStudio



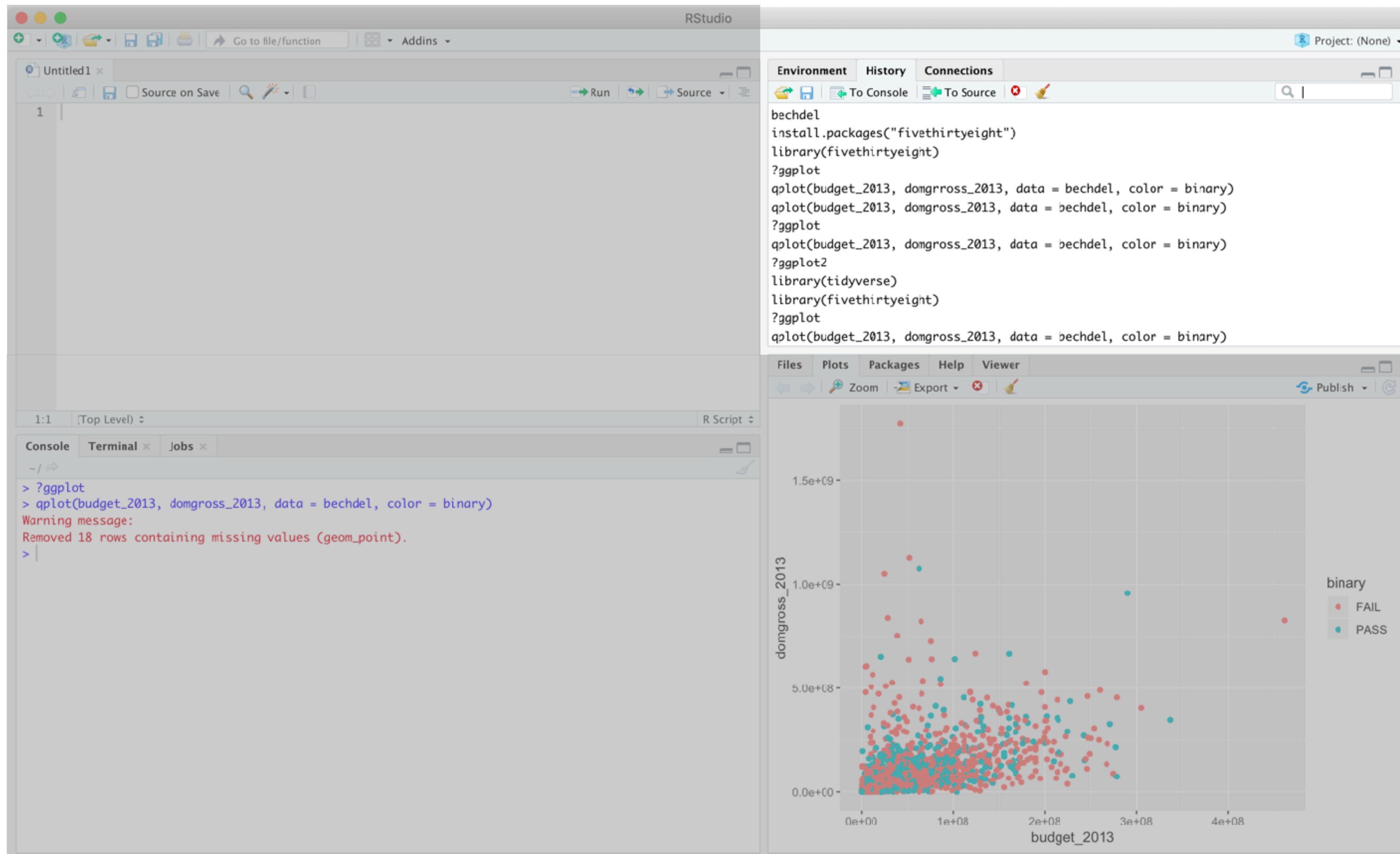
RStudio



RStudio



RStudio



RStudio

The screenshot displays the RStudio IDE interface. On the left, the R Markdown editor shows a code block for generating an HTML document, followed by R Markdown text and a code chunk for the 'cars' dataset. The R Markdown dropdown indicates the file type is 'R Markdown'. Below it, the Console window shows R code for ggplot and qplot, resulting in a warning message about missing values. On the right, the Environment tab of the workspace shows several R commands related to 'bechdel' and 'fivethirtyeight' packages. A scatter plot is displayed, comparing 'domgross_2013' on the y-axis against 'budget_2013' on the x-axis. The plot uses color to indicate the 'binary' status: red dots represent 'FAIL' and teal dots represent 'PASS'.

```
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 content as well as the output of any embedded R code chunks within the document. You can embed an R code chunk like this:  
15  
16 ```{r cars}  
17 summary(cars)  
18  
19  
20 ## To learn more about R Markdown, see http://rmarkdown.rstudio.com.  
2:1 # Untitled #
```

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

Environment History Connections

bechdel
install.packages("fivethirtyeight")
library(fivethirtyeight)
?ggplot
qplot(budget_2013, domgross_2013, data = bechdel, color = binary)
qplot(budget_2013, domgross_2013, data = bechdel, color = binary)
?ggplot
qplot(budget_2013, domgross_2013, data = bechdel, color = binary)
?ggplot2
library(tidyverse)
library(fivethirtyeight)
?ggplot
qplot(budget_2013, domgross_2013, data = bechdel, color = binary)

Files Plots Packages Help Viewer

Domgross_2013

binary

- FAIL
- PASS

budget_2013

Tidy Data Science

