

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!

Using R and the tidyverse for 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.”

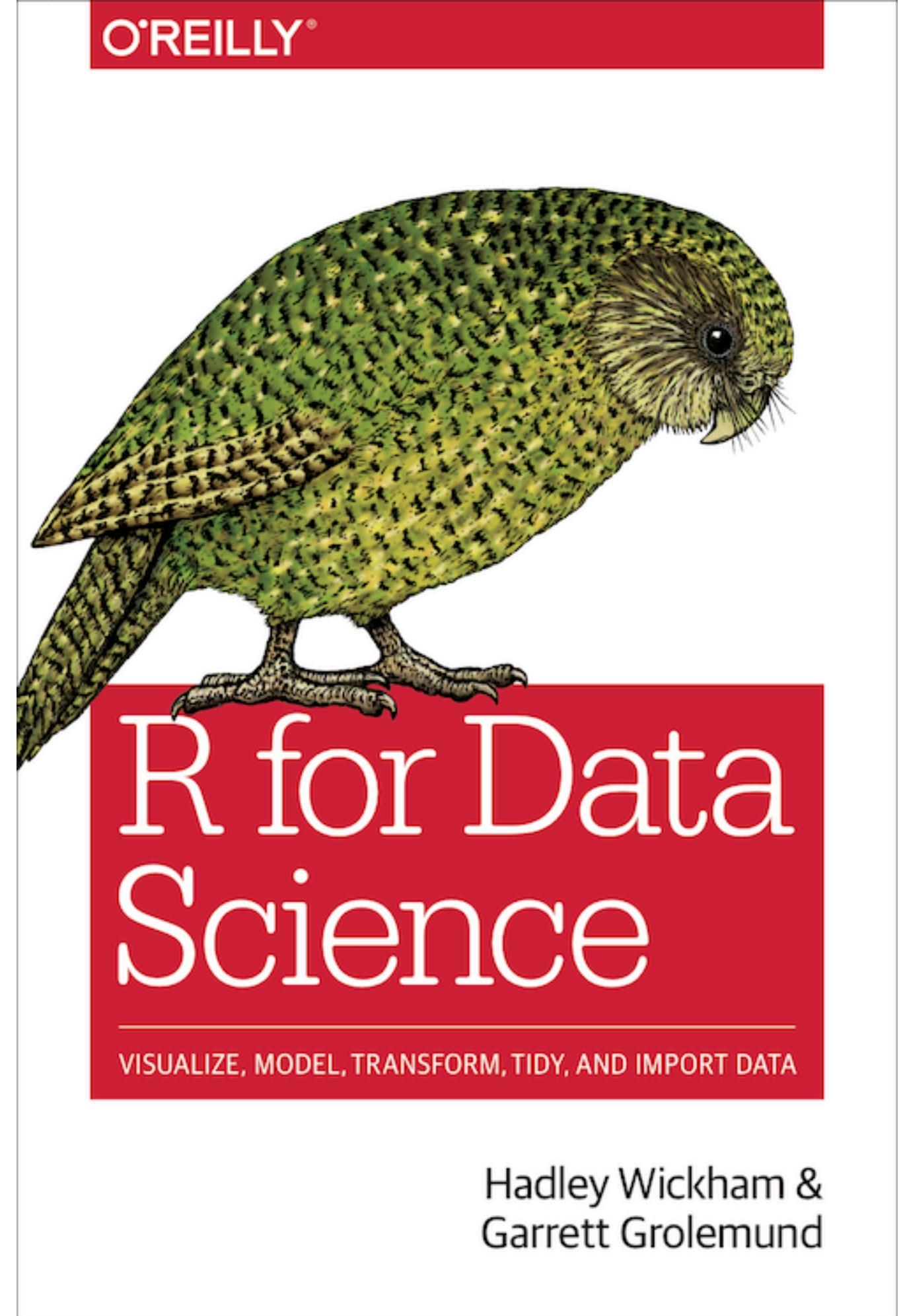


Orange sticky note: “I’m
stuck, can someone
please help me?”

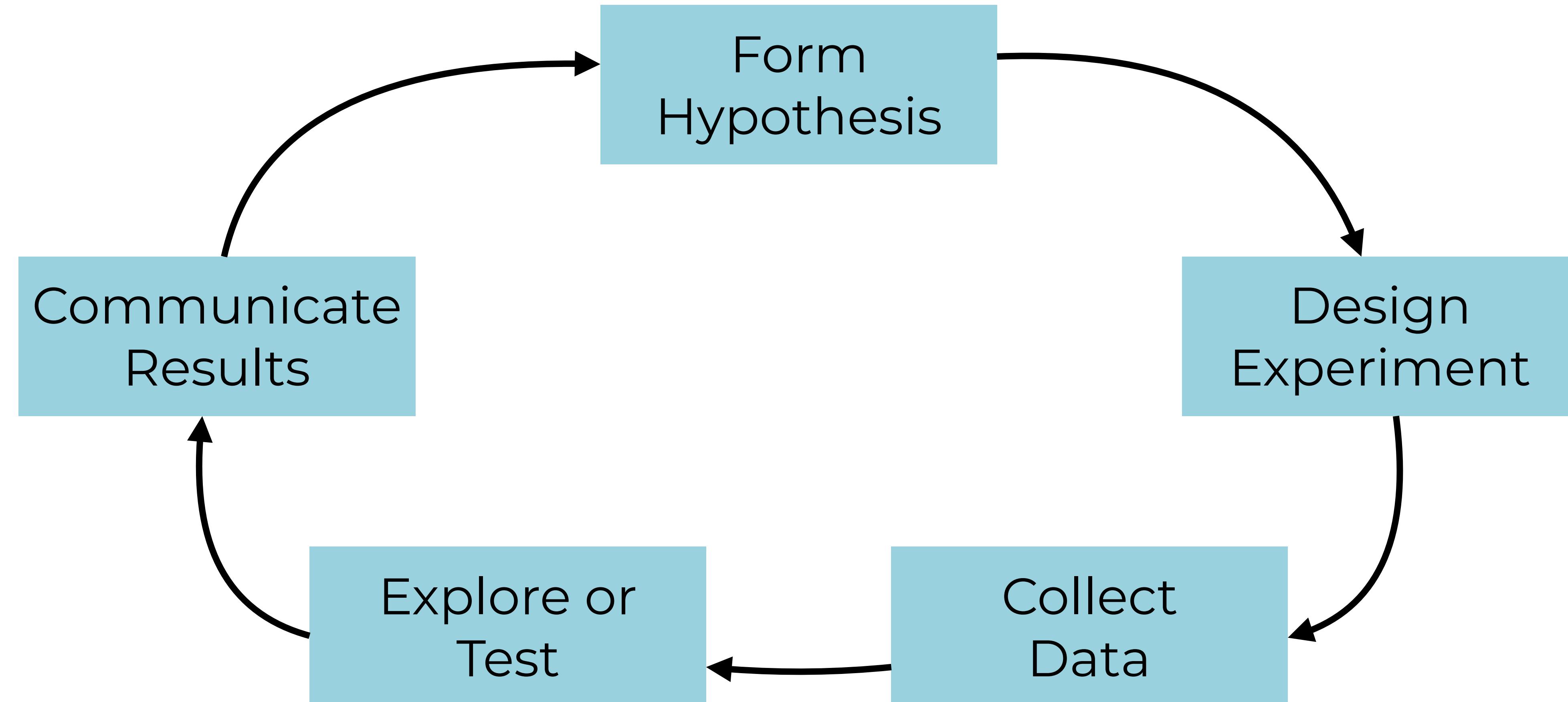


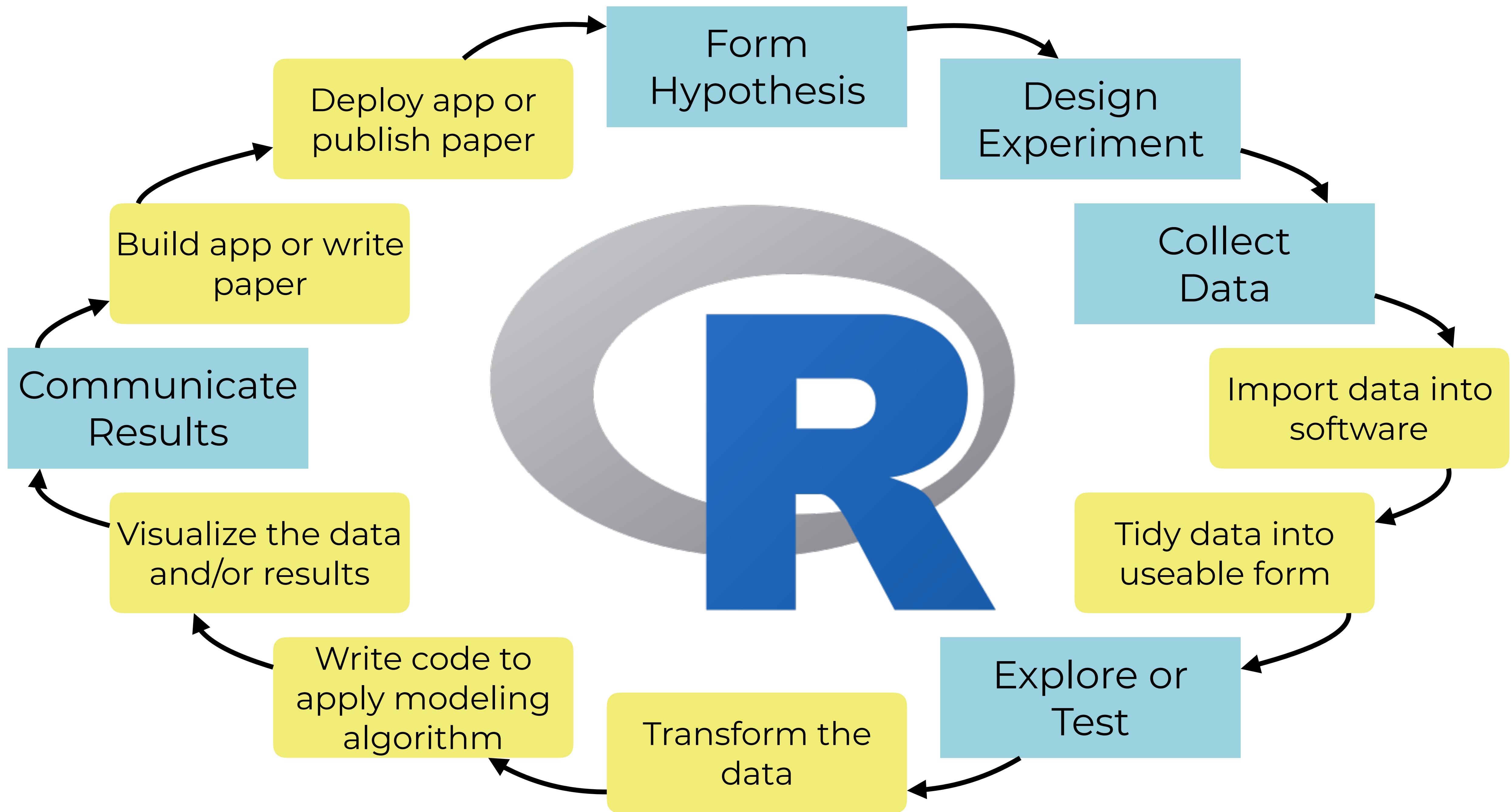
This workshop is based heavily on
R for Data Science
<https://r4ds.had.co.nz>

Links to the relevant
sections of the book

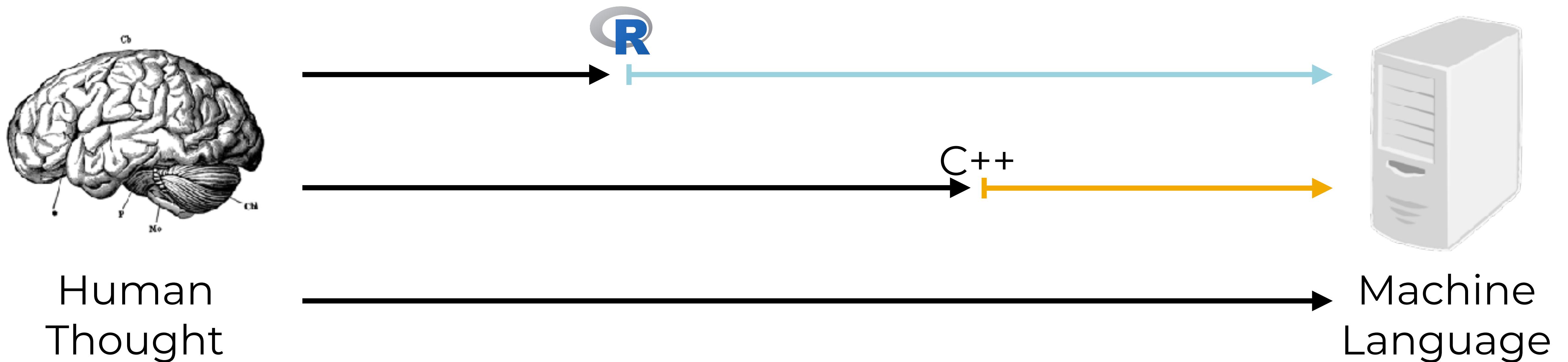


Data Science





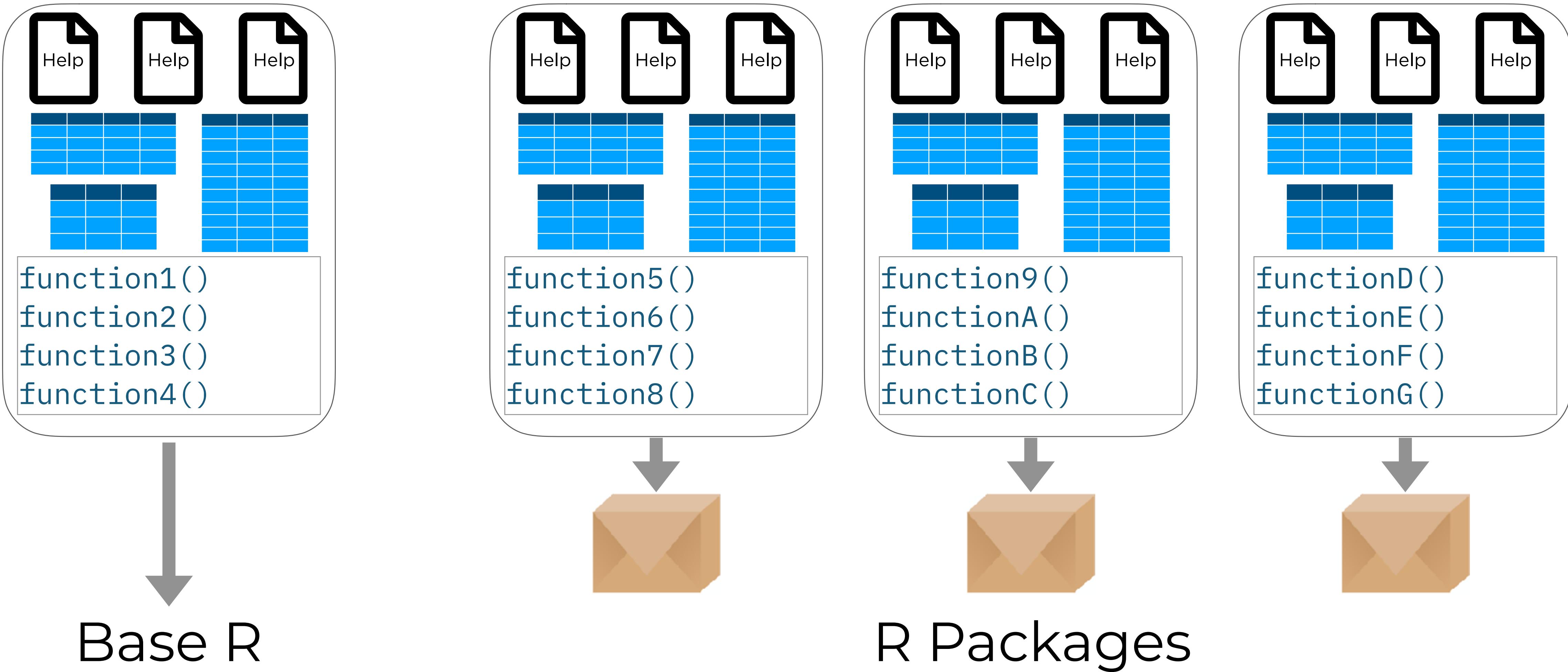
R: A computer language for scientists



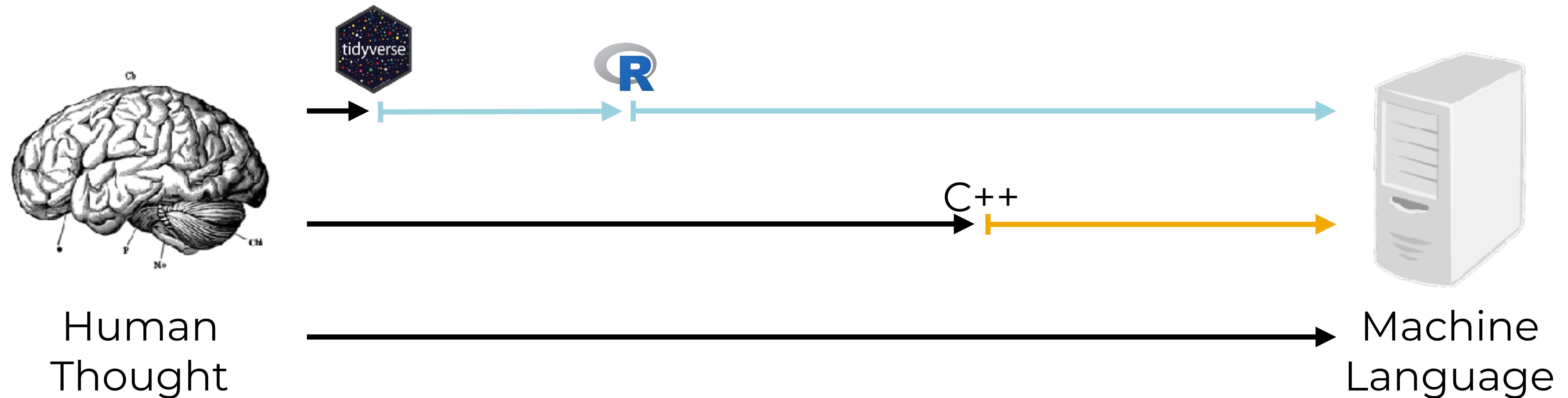
A dramatic photograph of a waterfall cascading over dark, jagged rocks into turbulent white water. The waterfall is the central focus, with its spray and mist catching some light. The background is a dark, overcast sky.

bit of success

R Packages



tidyverse: Ecosystem to unify data science tasks





gentle hill of striving

Using Packages

1

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

Downloads files to computer

1 x per Computer

2

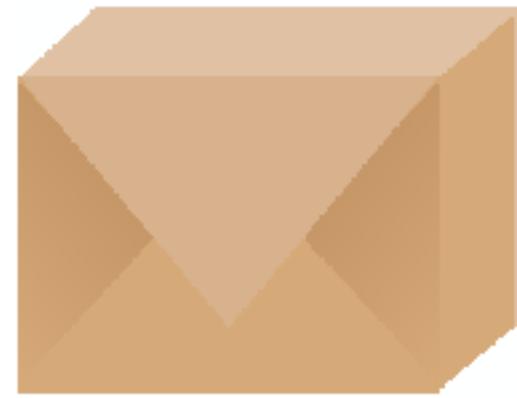
```
library("foo")
```

Loads package

1 x per R Session



tidyverse



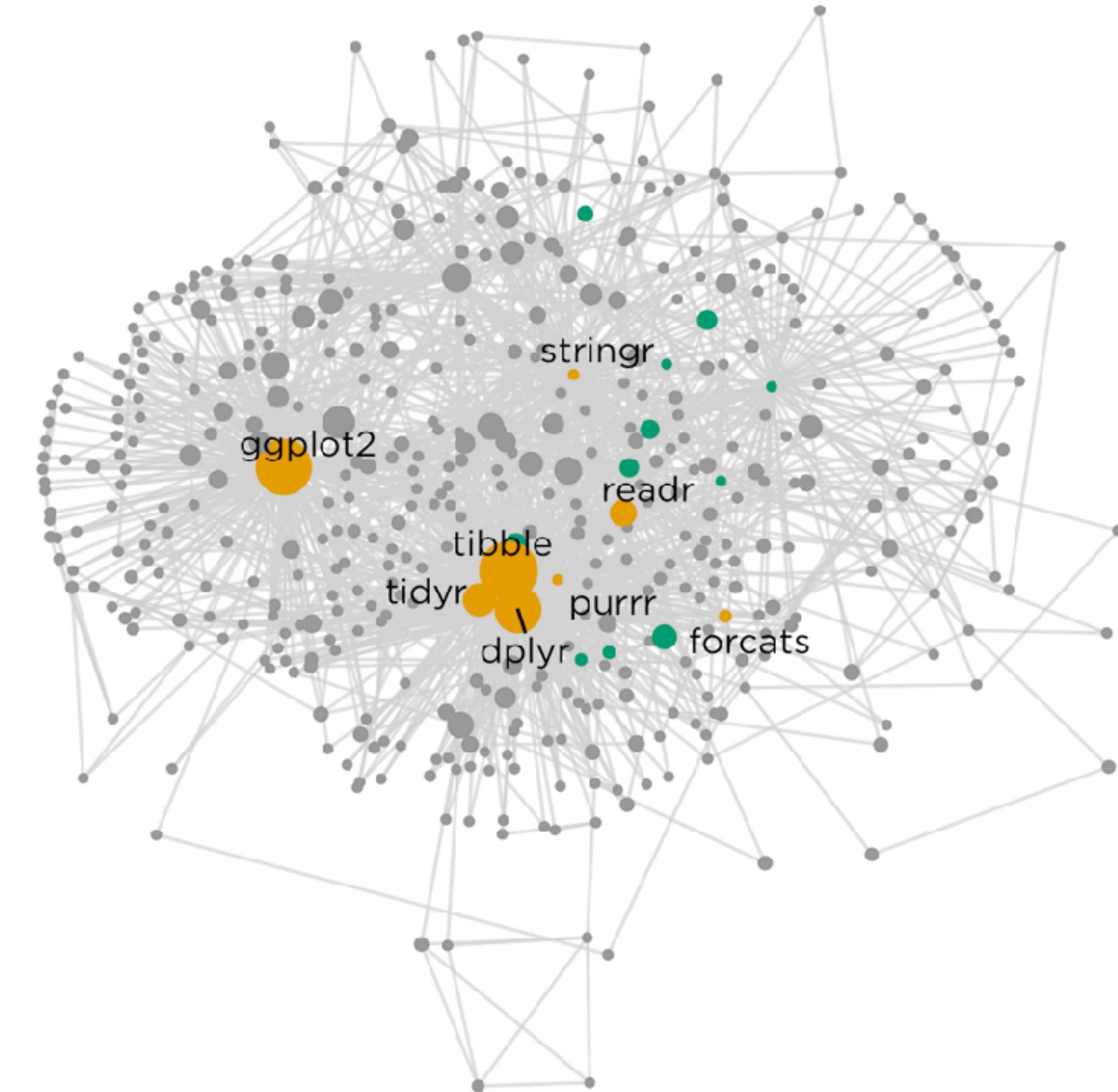
An R package that serves as a short cut for installing and loading components of the tidyverse.

```
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

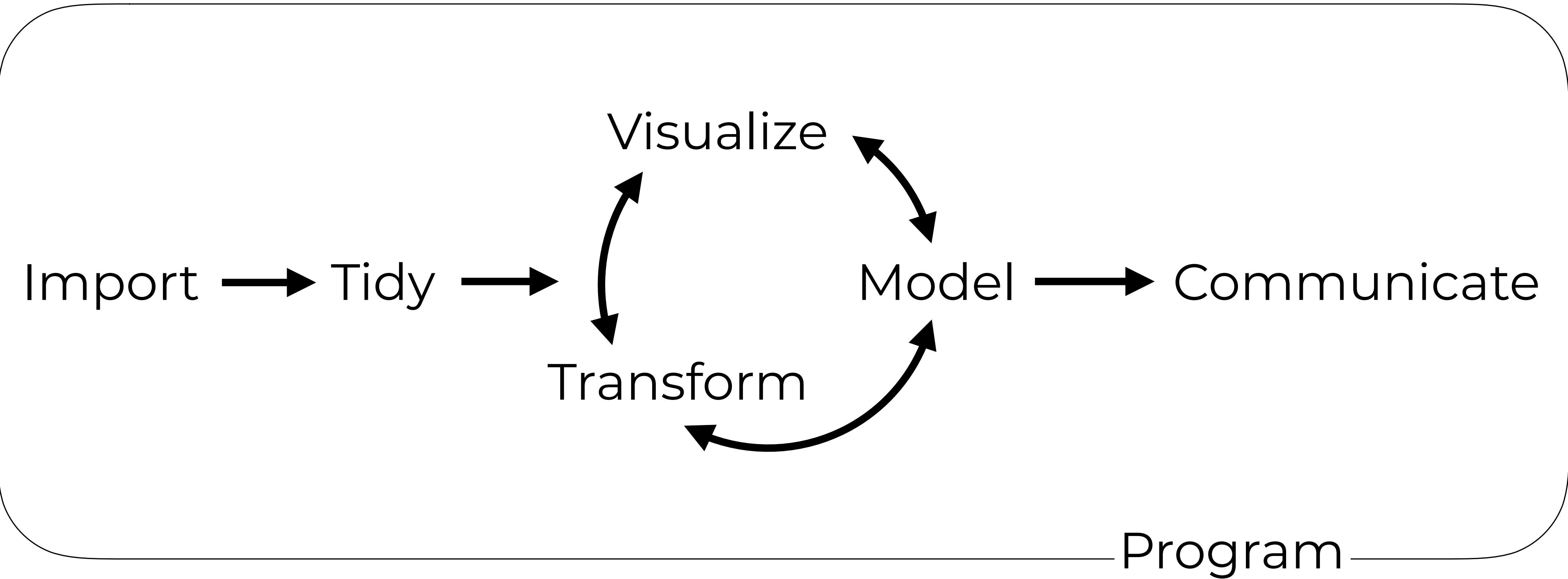
```
install.packages("ggplot2")
install.packages("tibble")
install.packages("tidyr")
install.packages("readr")
install.packages("purrr")
install.packages("dplyr")
install.packages("stringr")
install.packages("forcats")
install.packages("broom")
install.packages("dbplyr")
install.packages("haven")
install.packages("hms")
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

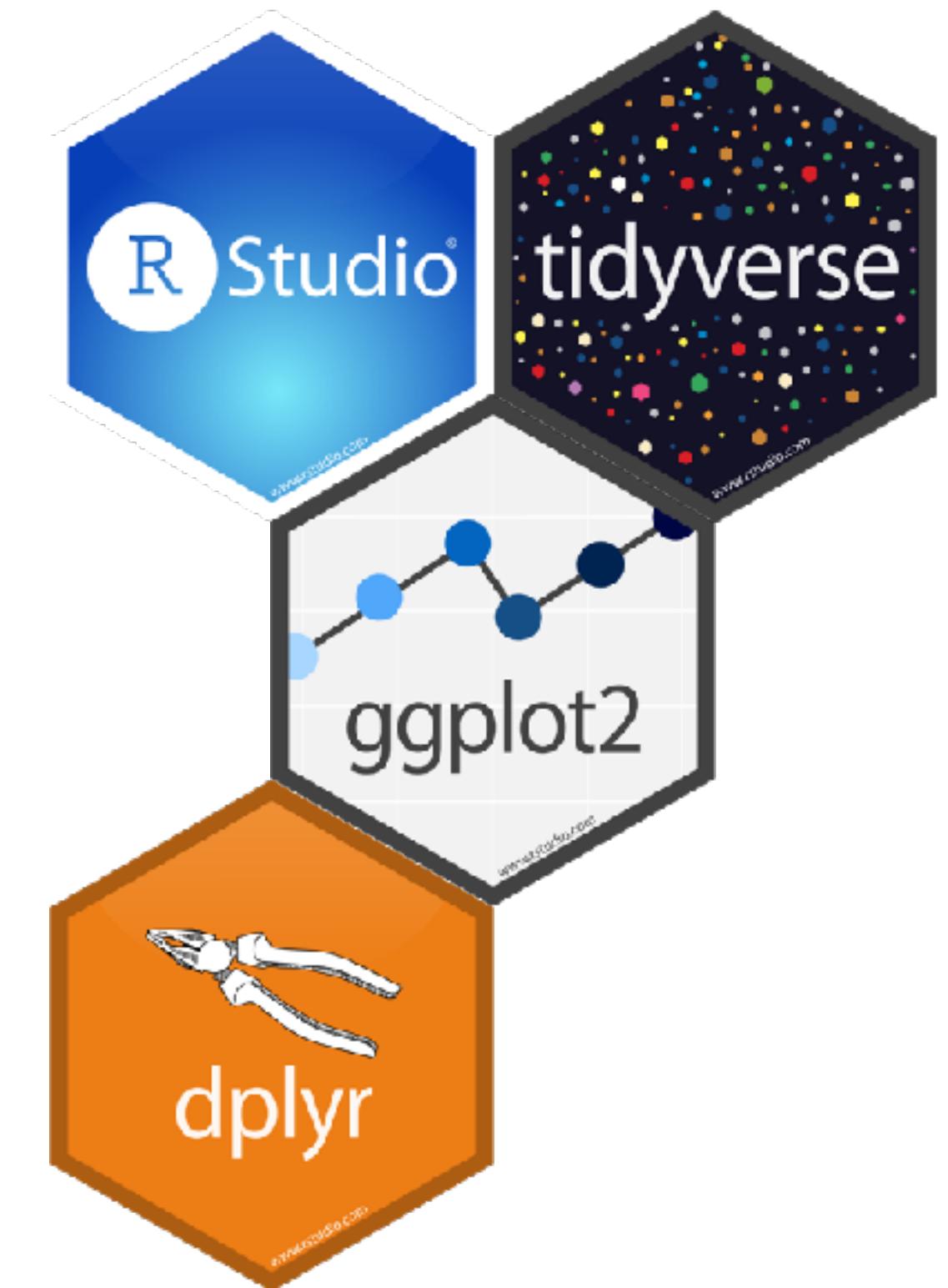
```
library("ggplot2")
library("tibble")
library("tidyr")
library("readr")
library("purrr")
library("dplyr")
library("stringr")
library("forcats")
```





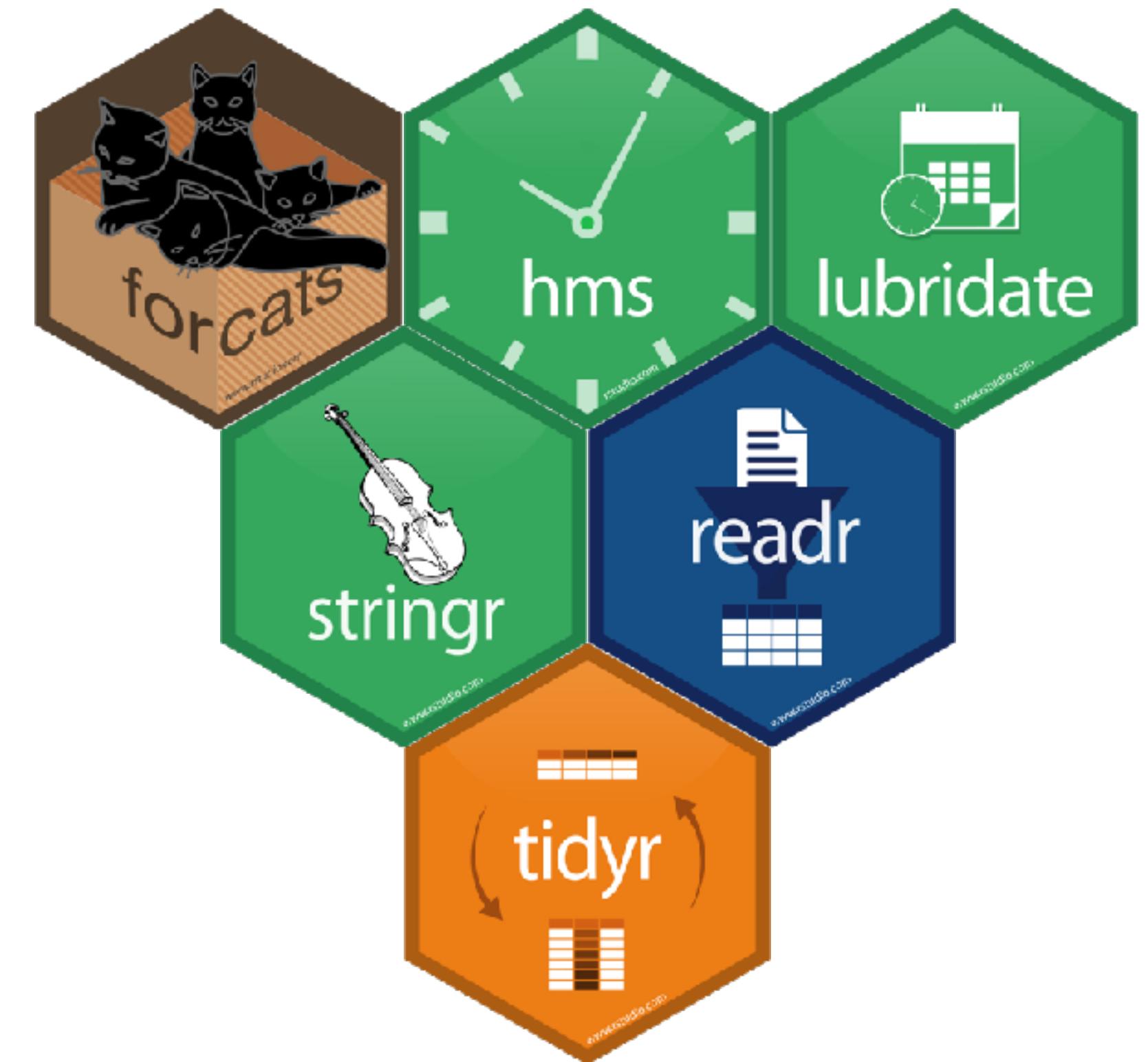
Day 1

9:30-10:30	Introduction to R and RStudio
10:30-11:00	Morning Break
11:00-12:30	Data Visualization
12:30-2:00	Lunch
2:00-3:30	Data Manipulation
3:30-4:00	Afternoon Break
4:00-5:00	Data Manipulation



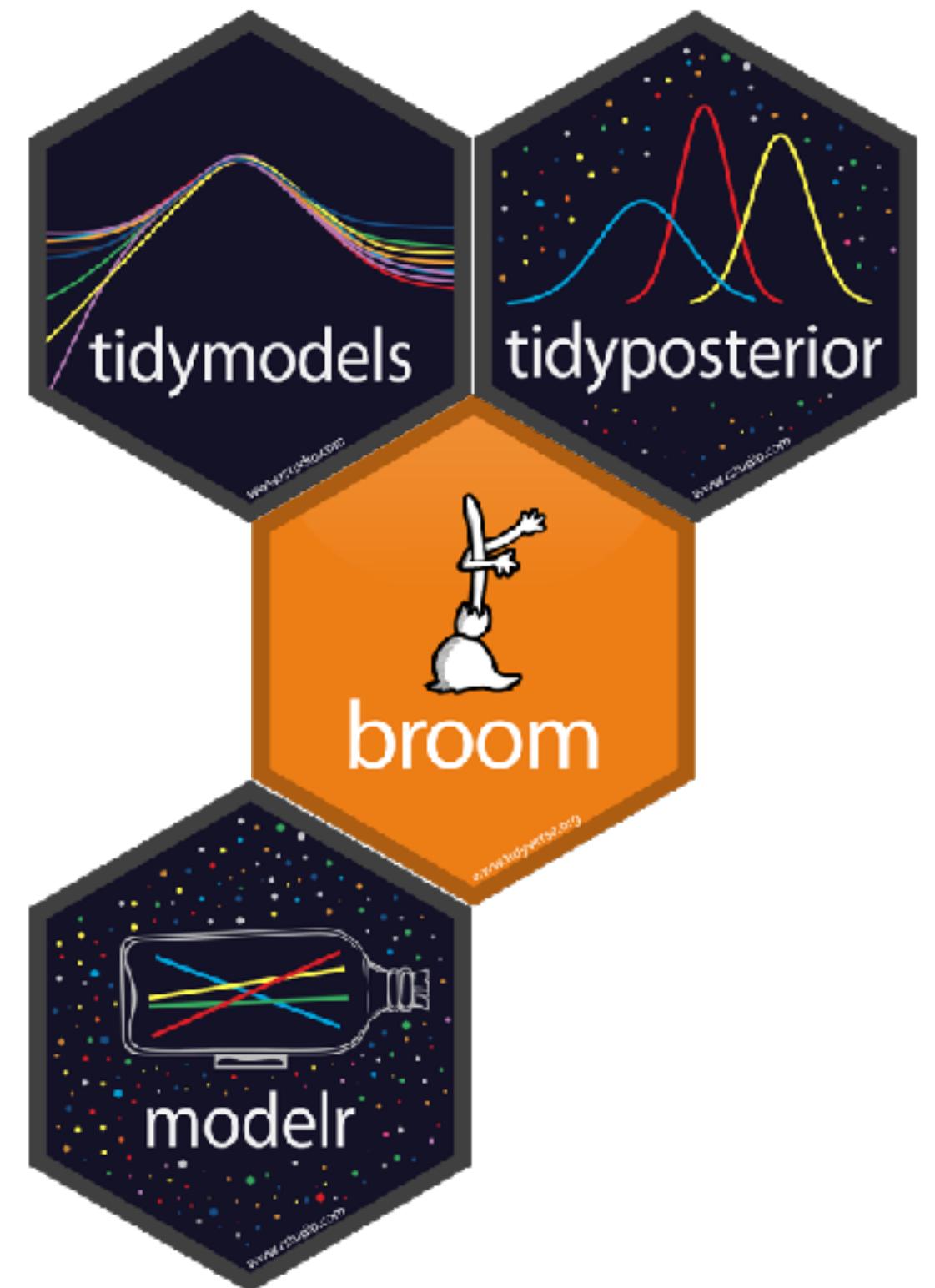
Day 2

9:00-10:30	Data Types
10:30-11:00	Morning Break
11:00-12:30	Import Data
12:30-2:00	Lunch
2:00-3:30	Tidy Data
3:30-4:00	Afternoon Break
4:00-5:00	Case Study



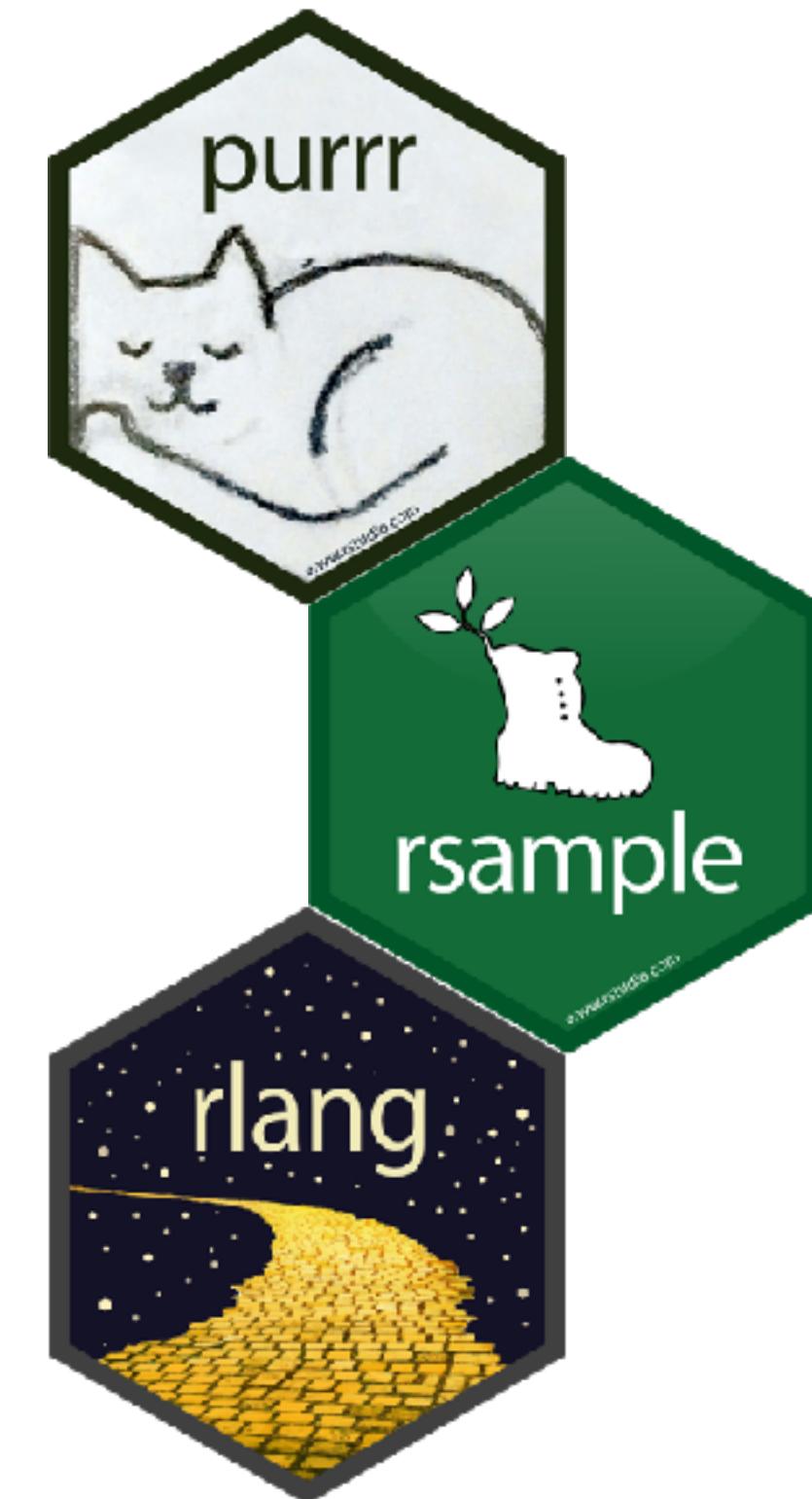
Day 3

9:00-10:30	Modeling Data
10:30-11:00	Morning Break
11:00-12:30	Tidying Models
12:30-2:00	Lunch
2:00-3:30	Visualizing Models
3:30-4:00	Afternoon Break
4:00-5:00	Case Study



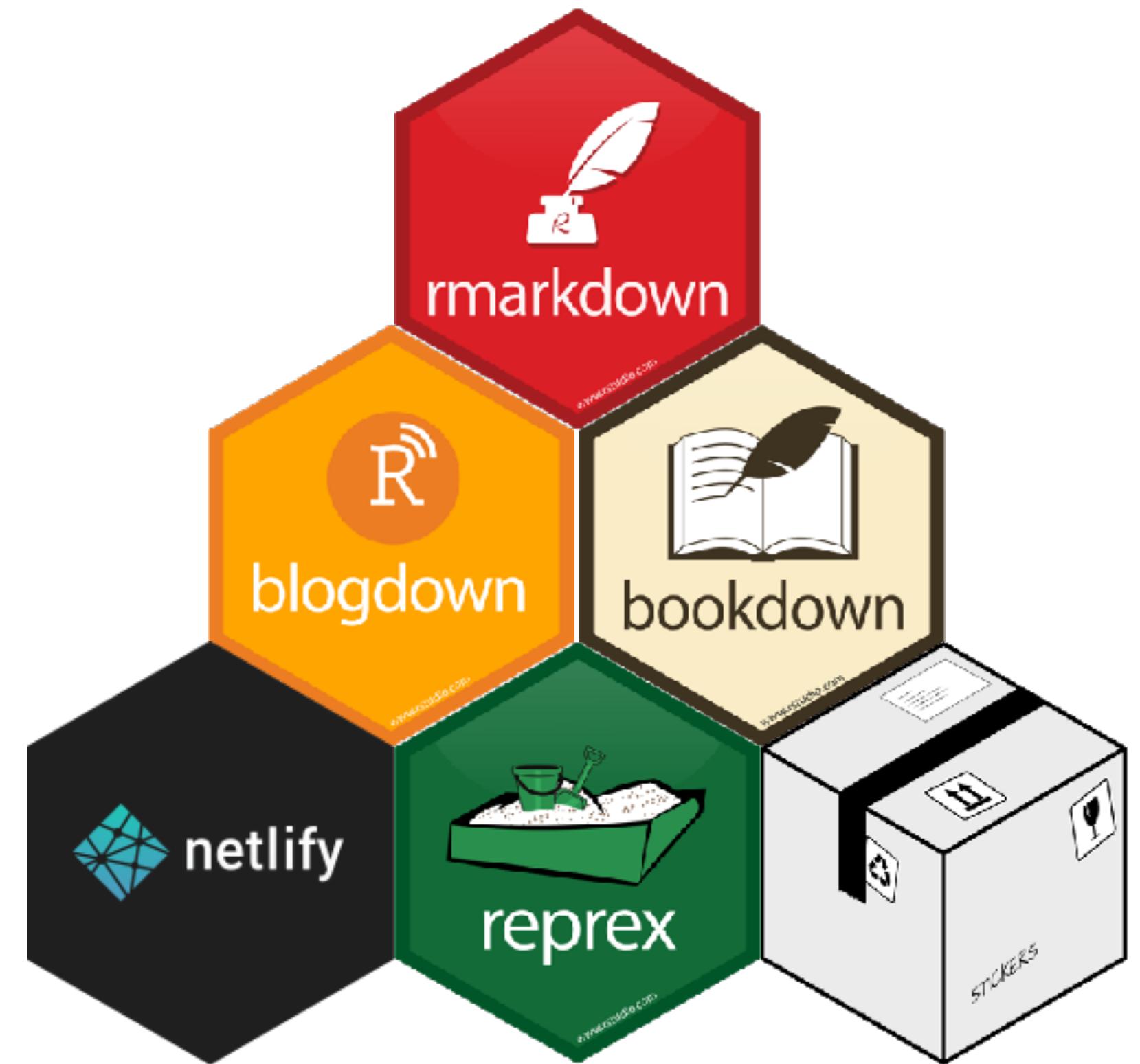
Day 4

9:00-10:30	Iteration
10:30-11:00	Morning Break
11:00-12:30	Multiple Models
12:30-2:00	Lunch
2:00-3:30	Writing Functions
3:30-4:00	Afternoon Break
4:00-5:00	Case Study



Day 5

9:00-10:30	Introduction to R Markdown
10:30-11:00	Morning Break
11:00-12:30	Multiple R Markdown Documents
12:30-2:00	Lunch
2:00-3:30	TBD
3:30-4:00	Afternoon Break
4:00-5:00	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

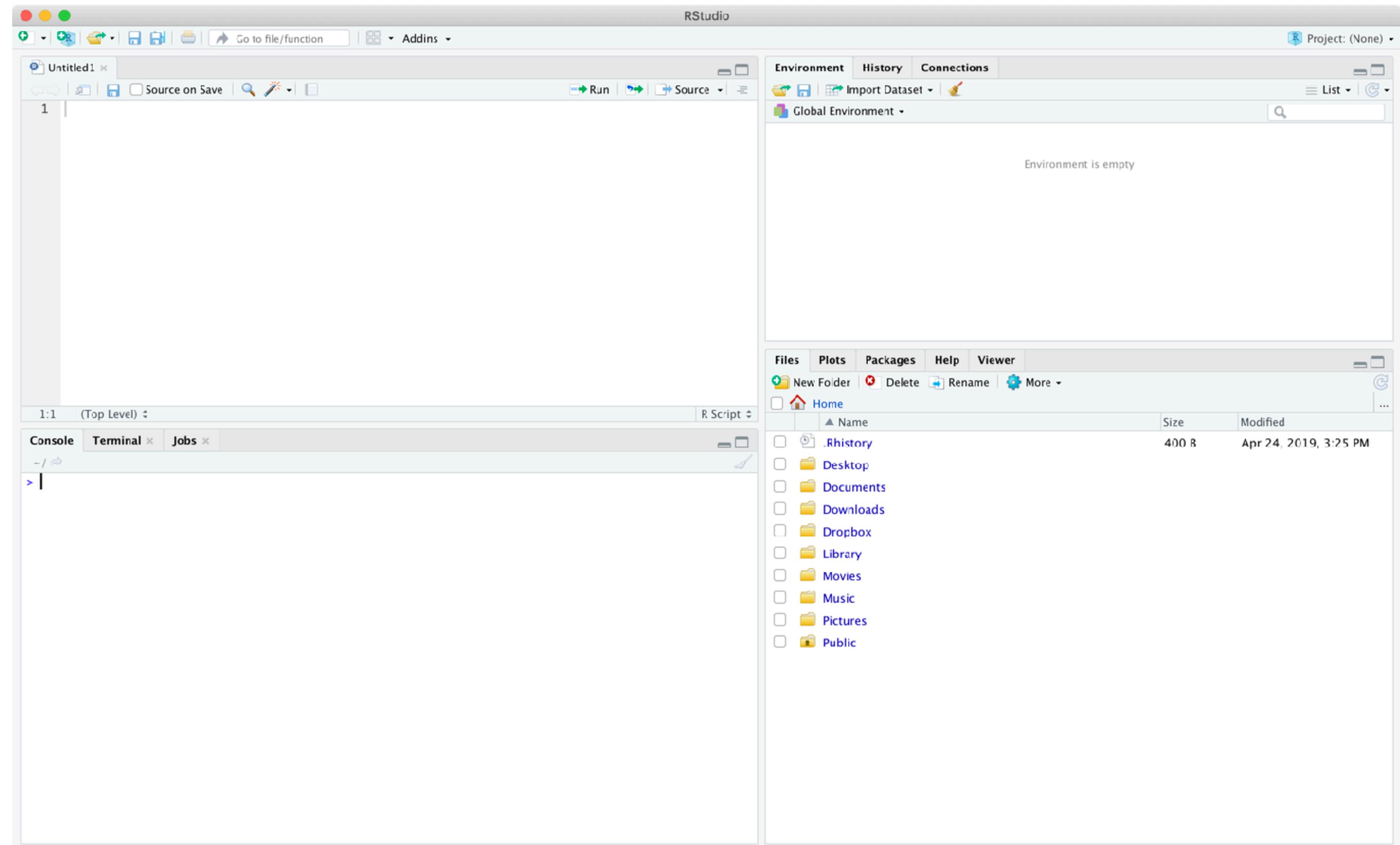
The screenshot shows the RStudio interface with an R Notebook open. The notebook contains the following Rmd code:

```
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
```

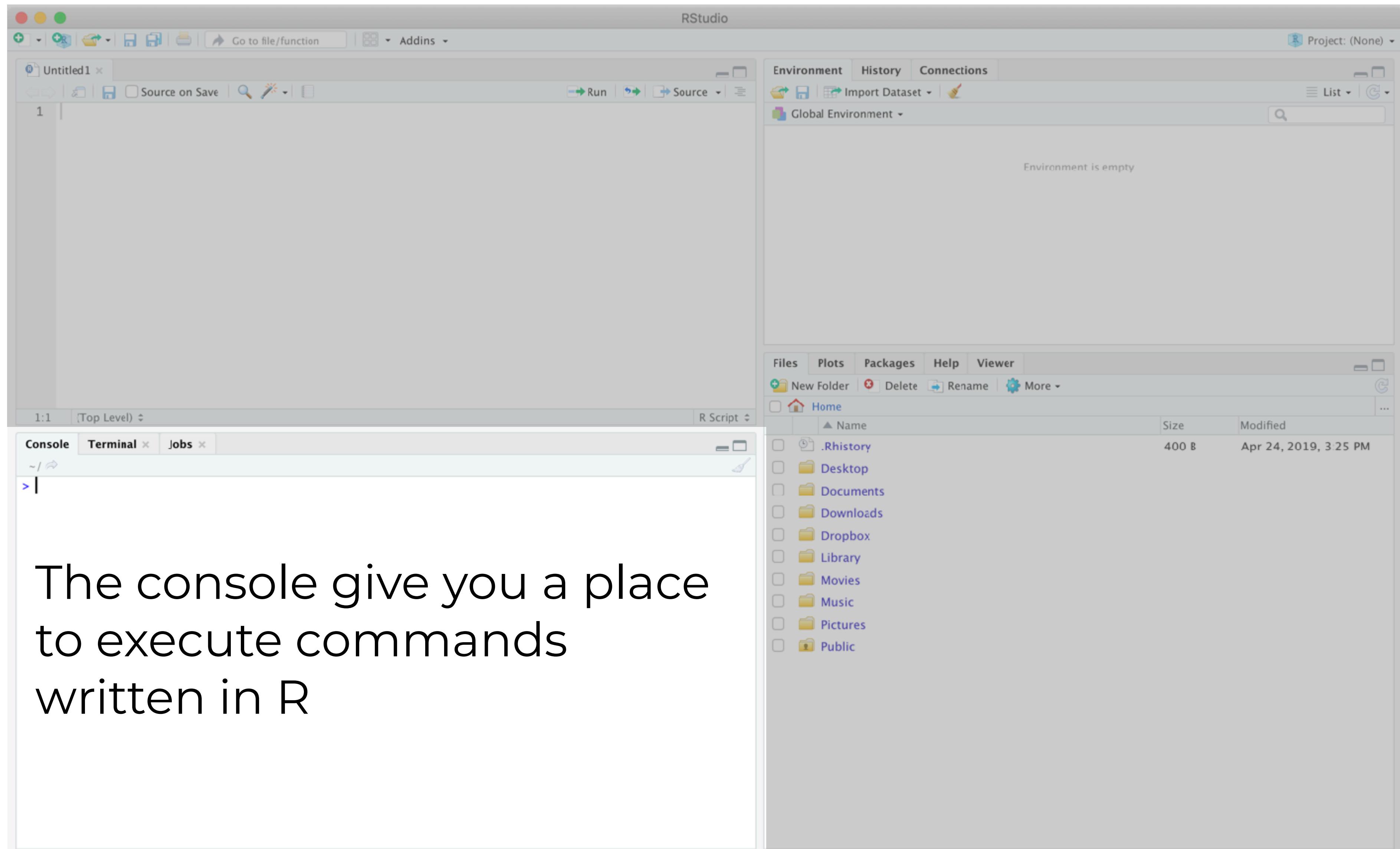
Annotations with arrows point to specific elements:

- 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".

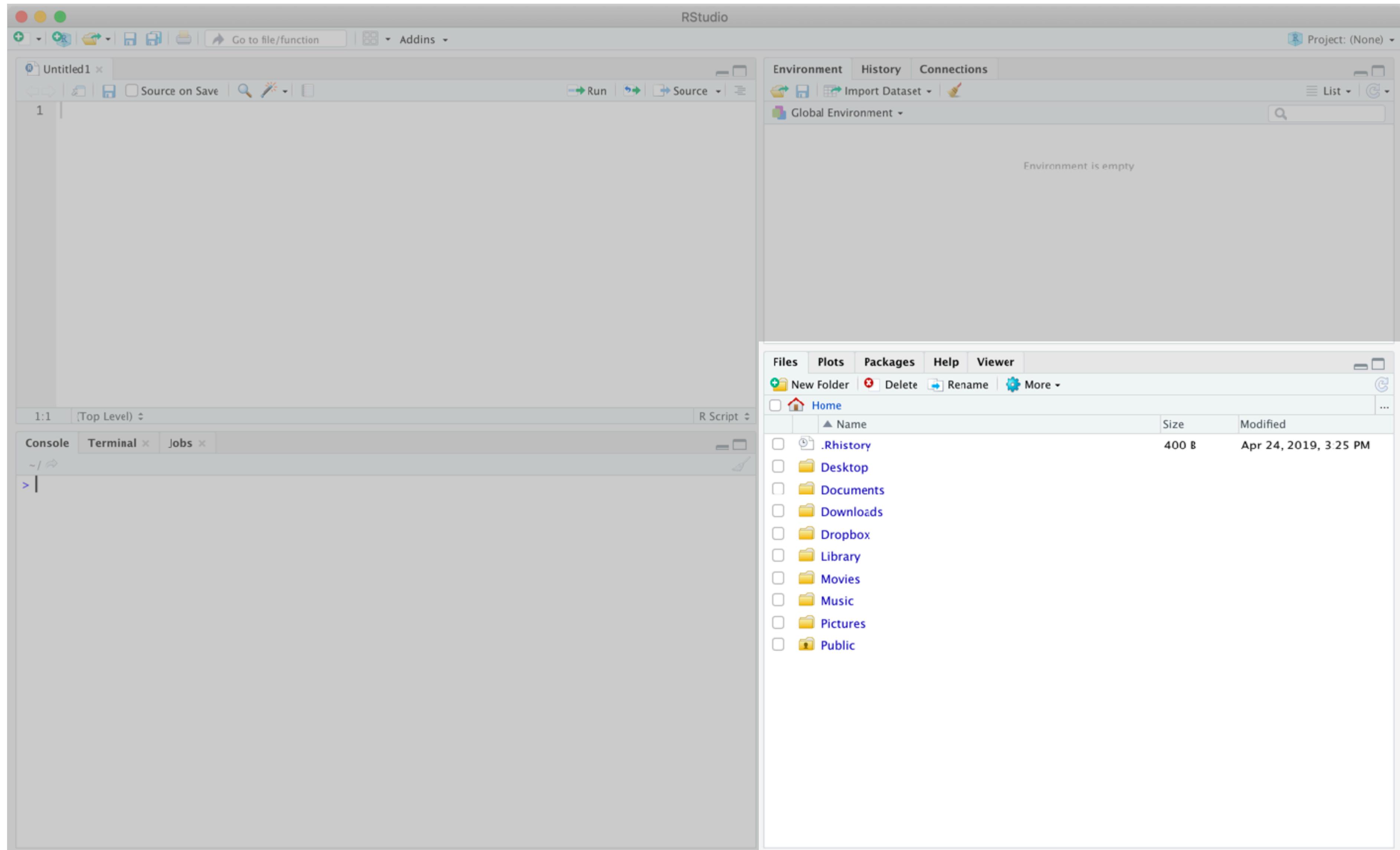
RStudio



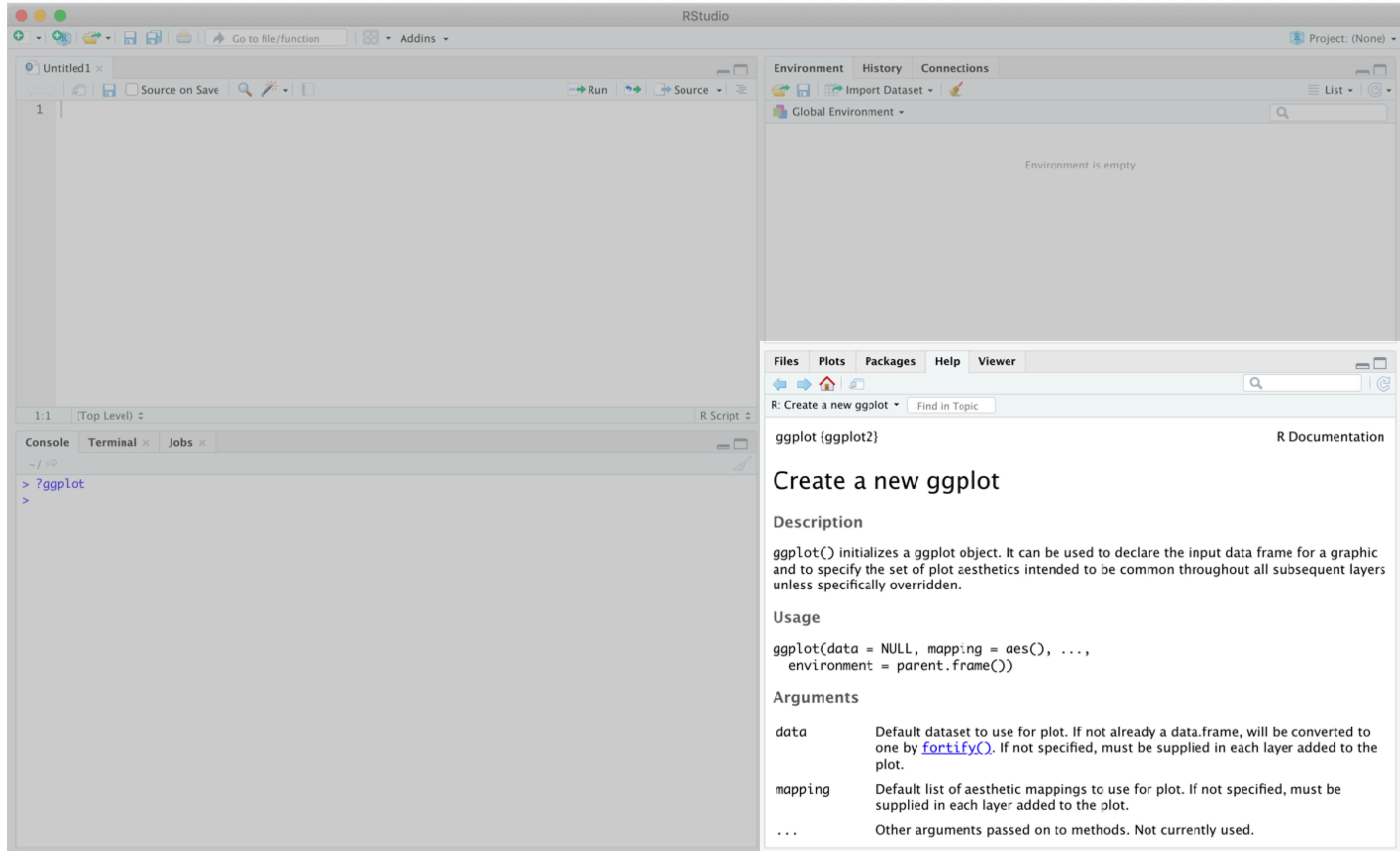
RStudio



RStudio



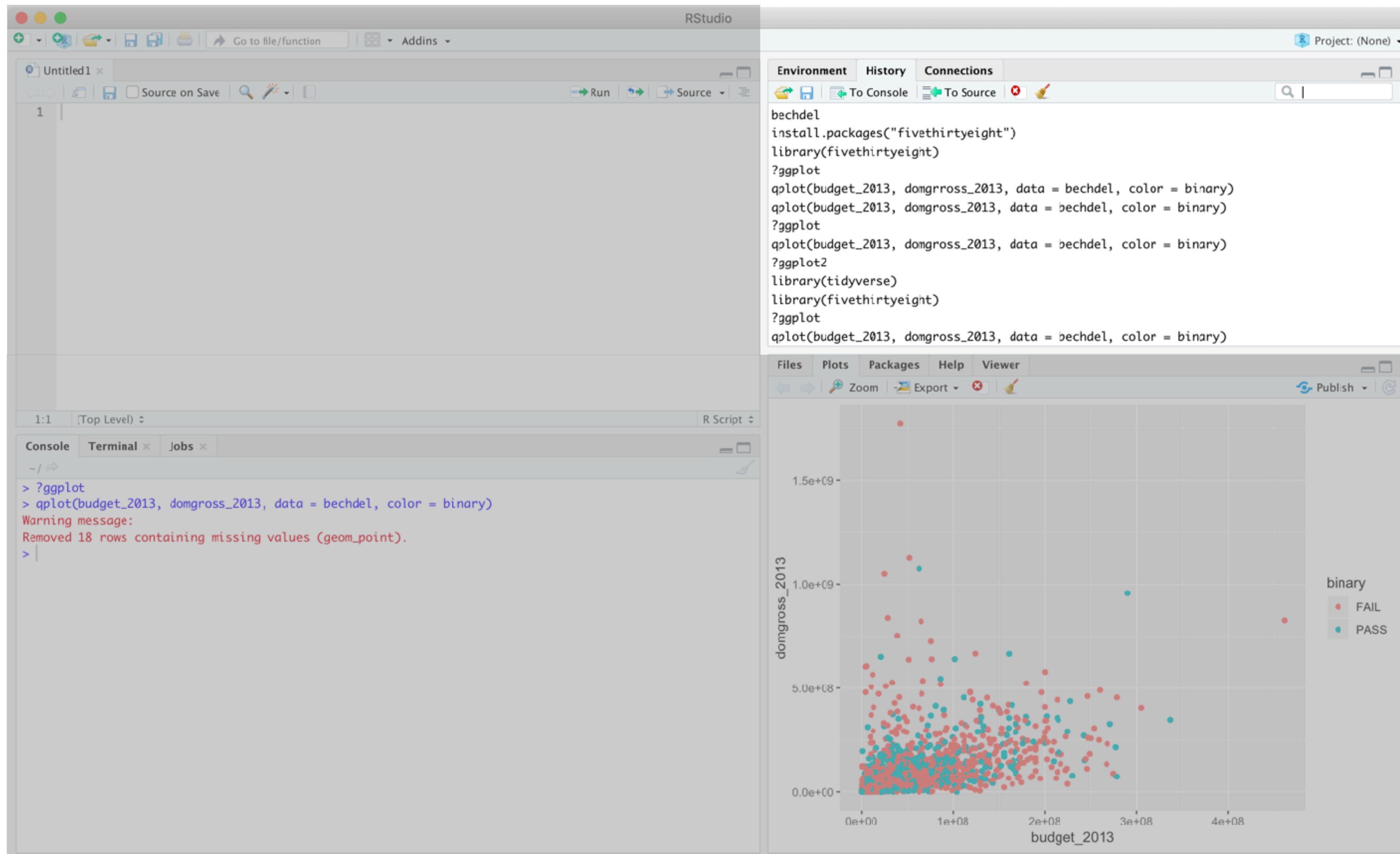
RStudio



RStudio



RStudio



RStudio

The screenshot displays the RStudio interface with the following components:

- Code Editor (Left Panel):** An R Markdown document titled "Untitled". The code includes R Markdown syntax (e.g., `---`, `# R Markdown`), R code chunks (e.g., `qplot`, `summary(cars)`), and a warning message from the console about missing values.
- Console (Bottom Left):** Shows the R command `?ggplot` followed by `qplot(budget_2013, domgross_2013, data = bechdel, color = binary)`. A warning message is displayed: "Warning message: Removed 18 rows containing missing values (geom_point)."
- Environment (Top Right):** Shows the session environment with packages like `fivethirtyeight` loaded and functions like `bechdel` and `qplot` used.
- Plots (Bottom Right):** A scatter plot titled "domgross_2013" versus "budget_2013". The x-axis ranges from 0e+00 to 4e+08, and the y-axis ranges from 0.0e+00 to 1.5e+09. Data points are colored by the "binary" variable, with red dots representing "FAIL" and teal dots representing "PASS".

Using R and the tidyverse for Data Science

