

Welcome to the Tidyverse



HELLO

my name is

Aymeric

 [@aymeric-stamm](#)

aymeric.stamm@cnr.fr

HELLO

my name is

Manon

 [@manon-simonot](#)

manon.simonot@univ-nantes.fr

O'REILLY®



Hands-On Programming with R

WRITE YOUR OWN FUNCTIONS AND SIMULATIONS

Garrett Golemud
Foreword by Hadley Wickham

O'REILLY®



R for Data Science

VISUALIZE, MODEL, TRANSFORM, TIDY, AND IMPORT DATA

Hadley Wickham &
Garrett Golemud

The R Series

R Markdown

The Definitive Guide



Yihui Xie
J. J. Allaire
Garrett Golemud

CRC Press
Taylor & Francis Group
A CHAPMAN & HALL BOOK

Pop Quiz

What does **IMRAD** stand for? Poll your neighbors.

Introduction

What hypothesis was tested and why?

Methods

How was the study done?

Results

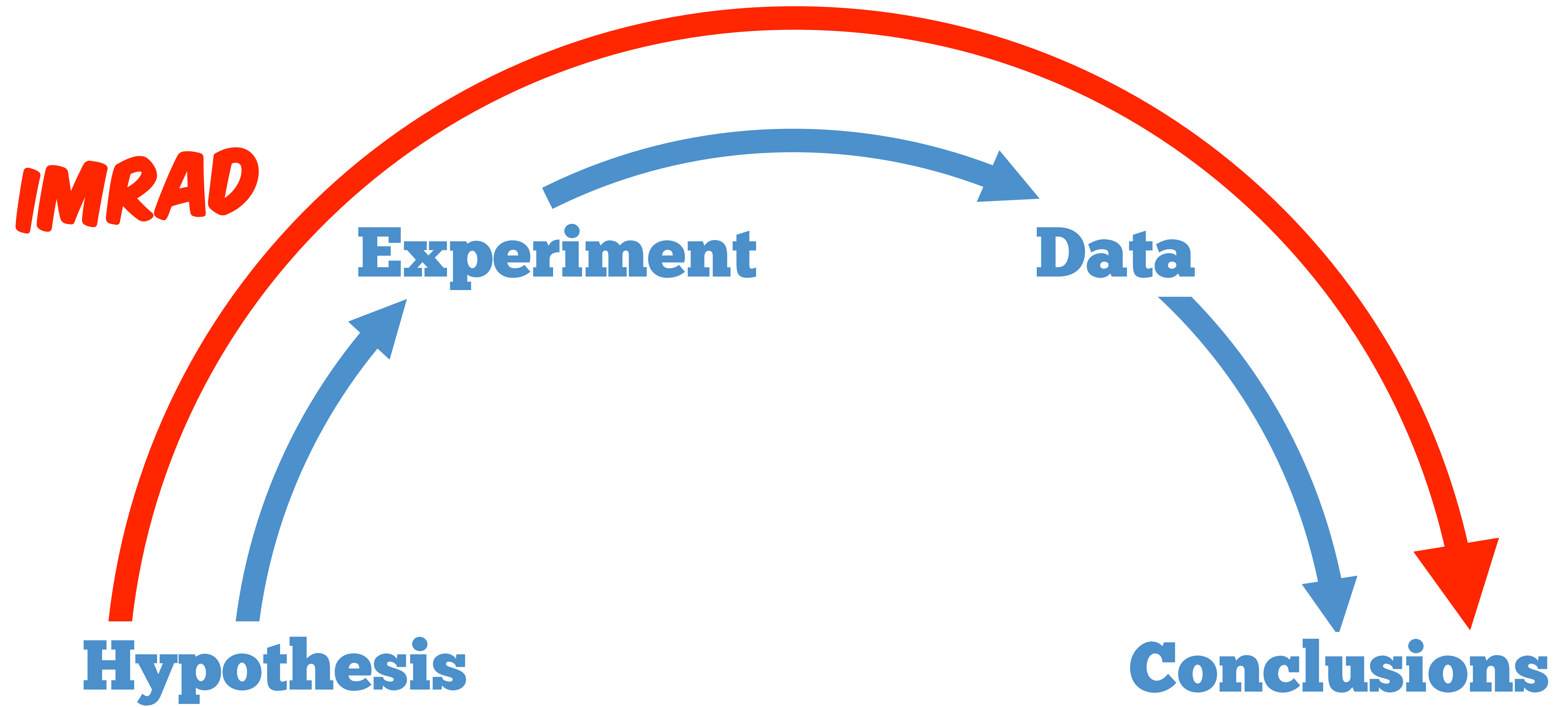
What answer was discovered?

And Discussion

What does the answer imply?

01:00

(Experimental) Science

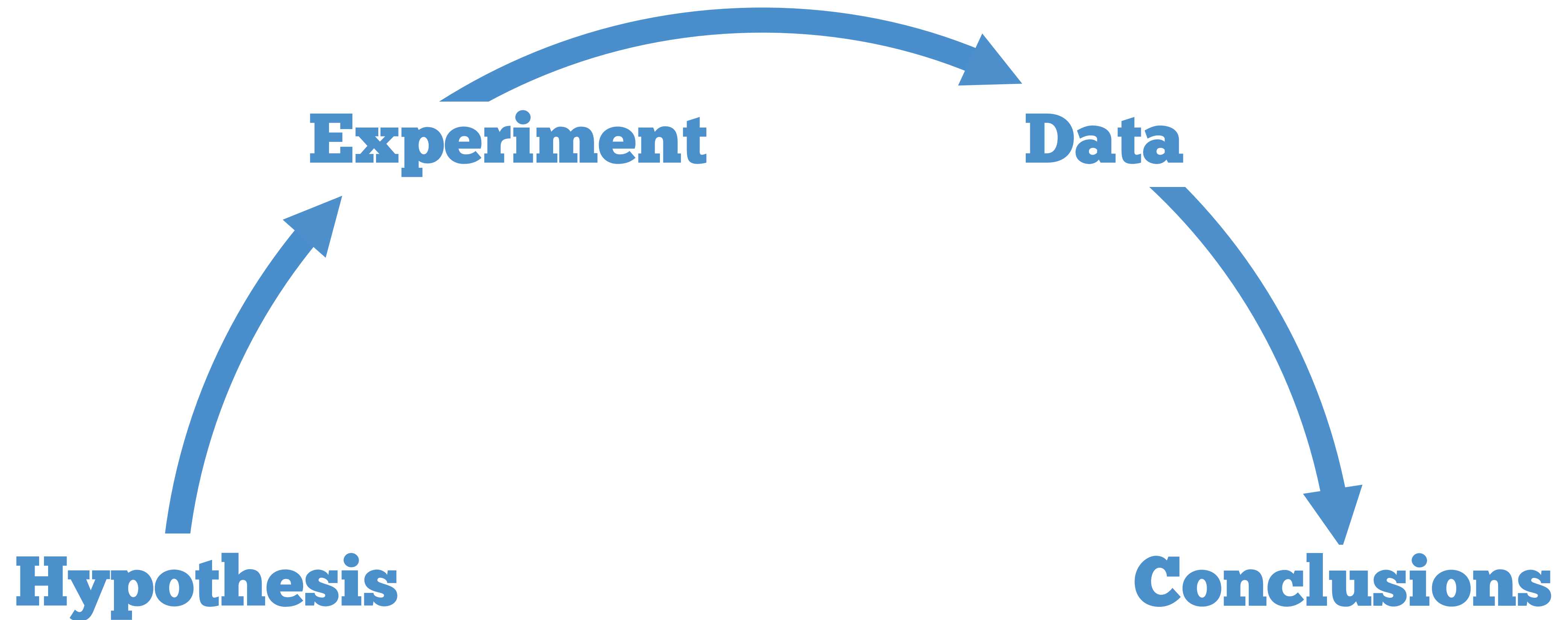


Pop Quiz

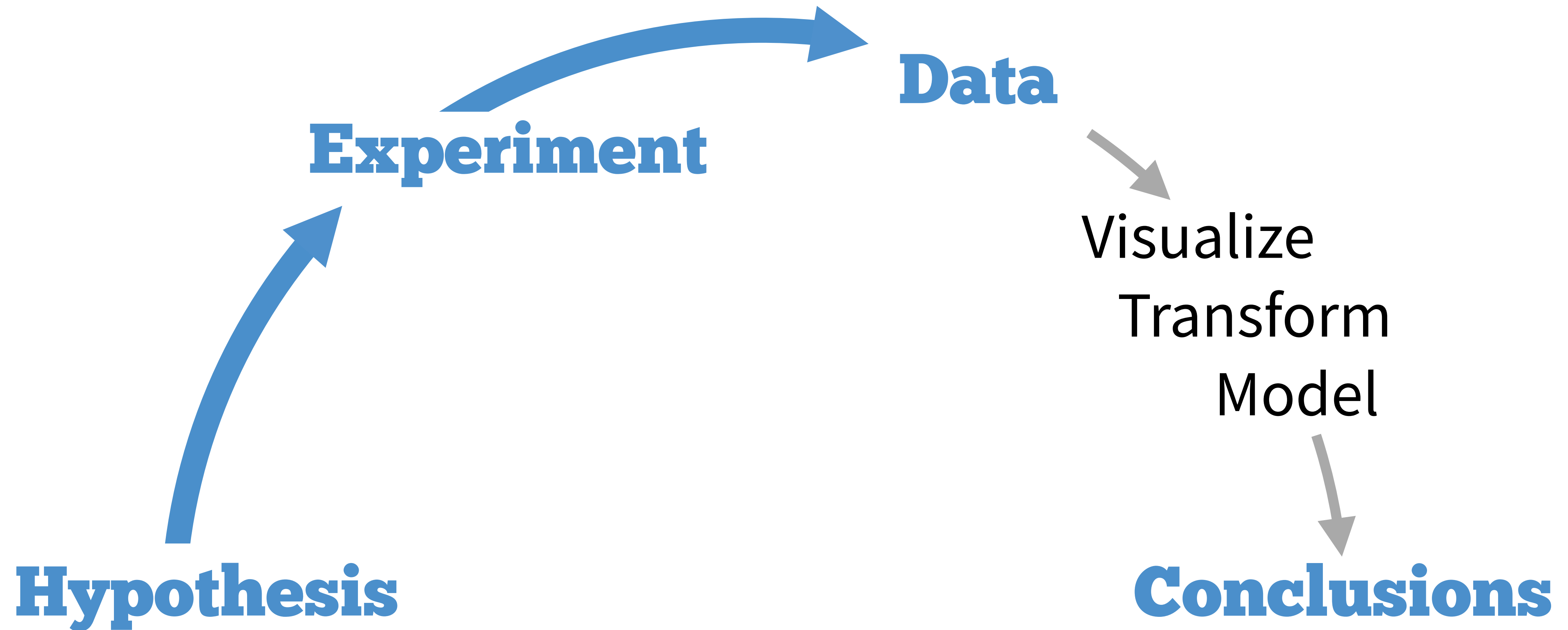
What is **data science**? Poll your neighbors.

01:00

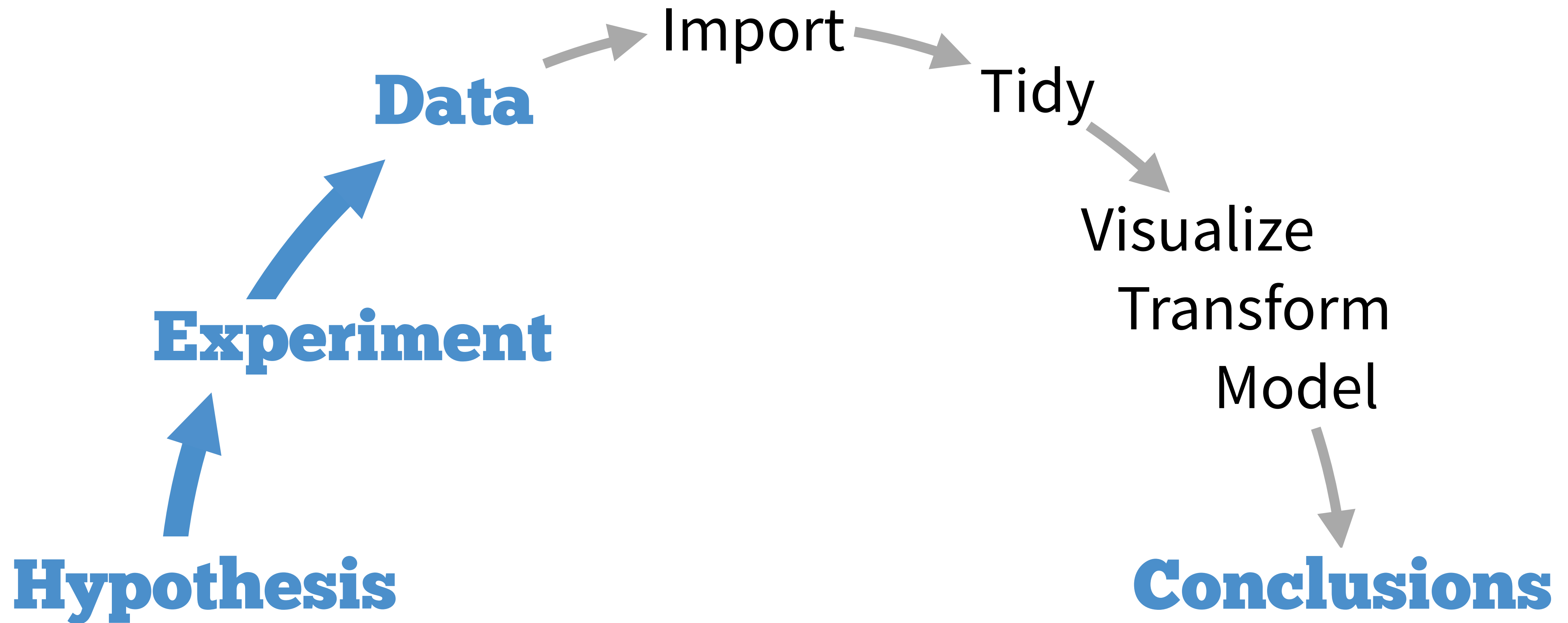
Data Science



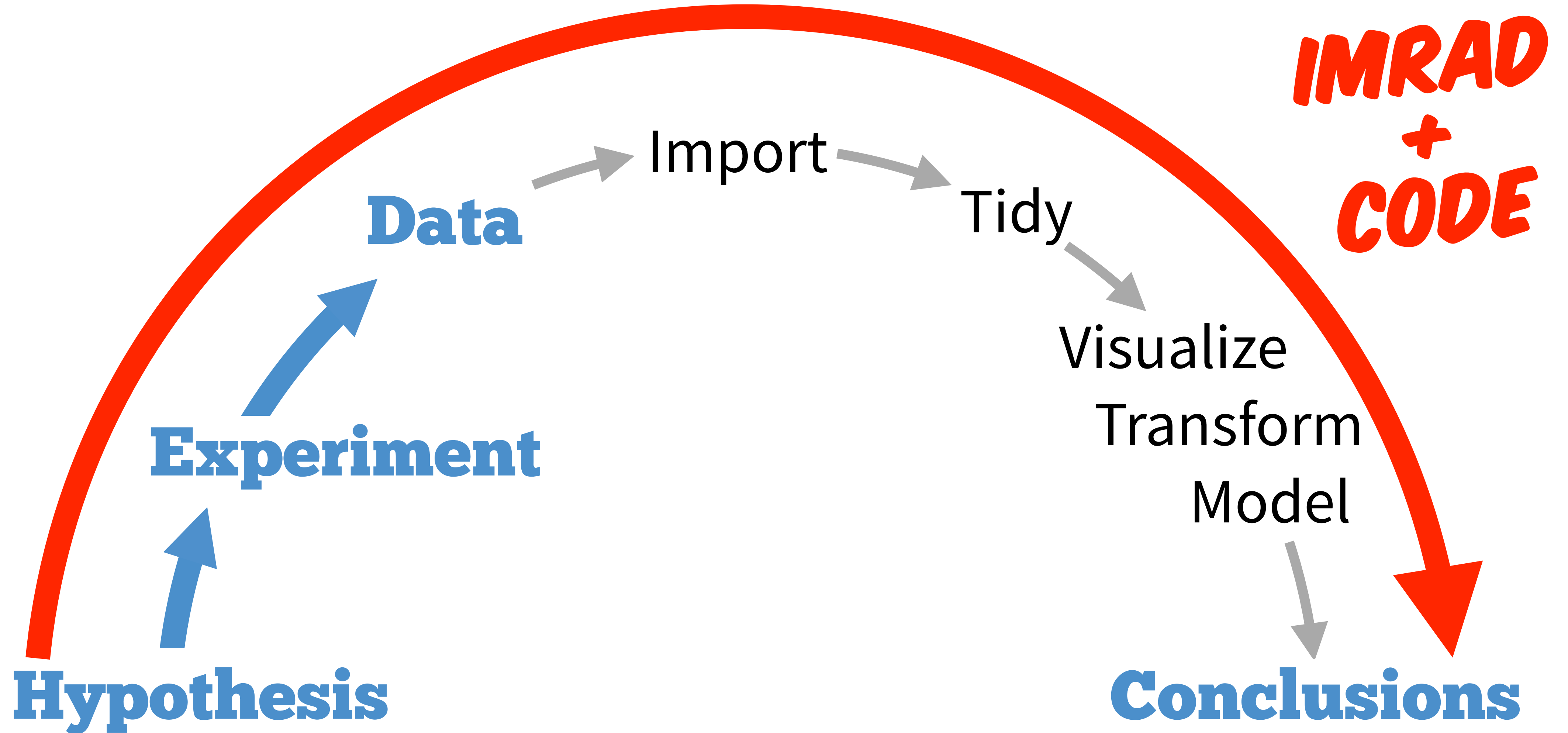
Data Science



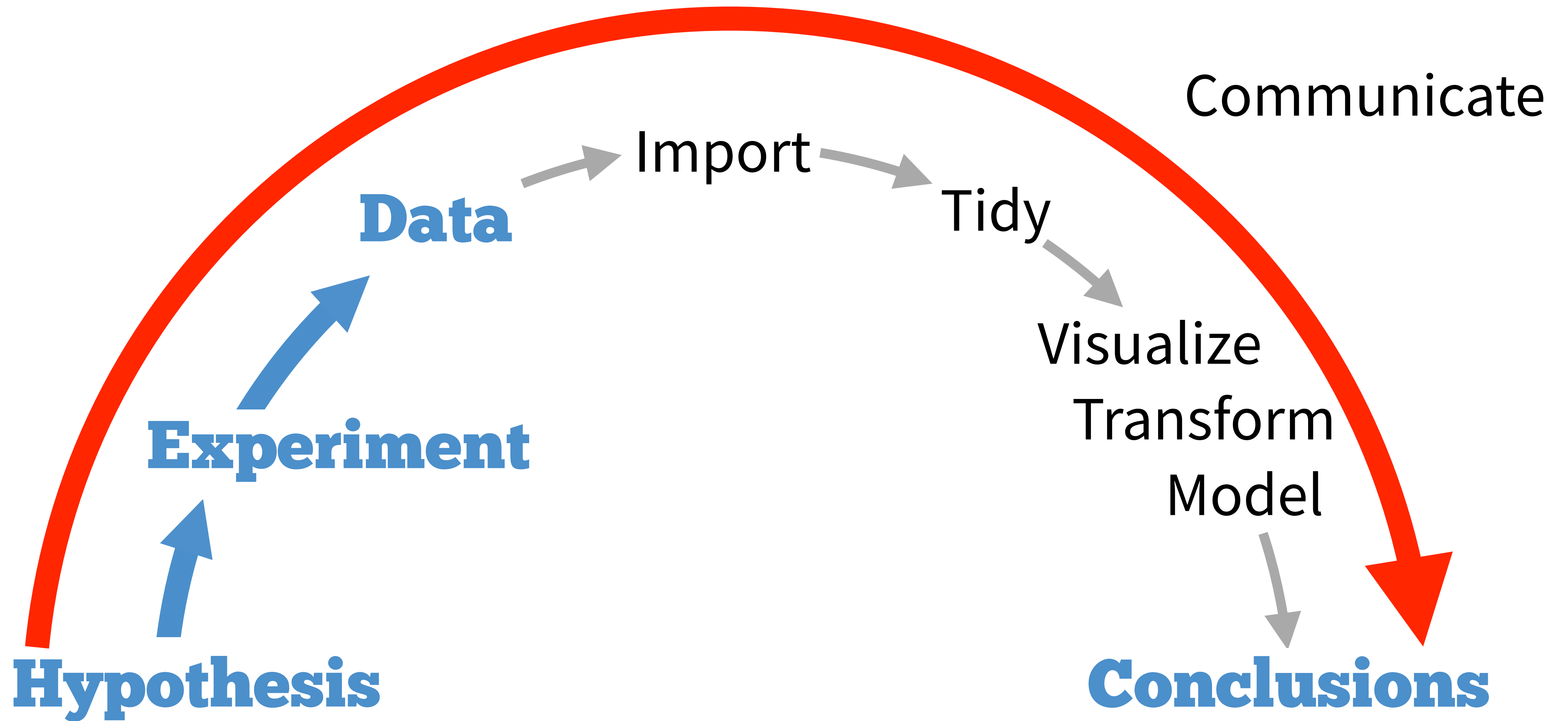
Data Science



Data Science



Data Science



Data Science

Import

Tidy

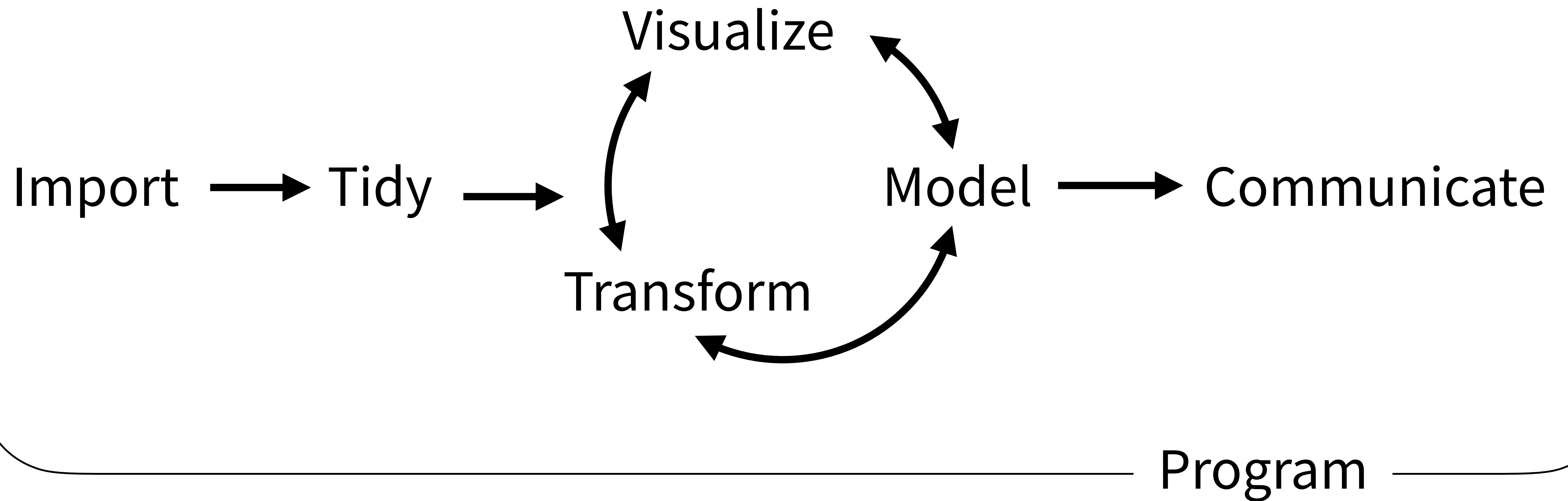
Communicate

Visualize

Transform

Model

(Applied) Data Science







A language





A language



Writing software

Outline

Introduction

Visualize Data

Transform Data

Model Data

Communicate Data

Tidy Data

Join Data

Manipulate Data Types

Manipulate Lists

Your Turn

Go here for the class material

<https://astamm.github.io/data-science-with-r/>

02:00

R the language

Values: 1, "Florida", "2010-01-25"

R the language

Values: 1, "Florida", "2010-01-25"

Objects: $x \leftarrow 22/7$

A name
without quotes

\leftarrow followed by -
(it looks like an arrow)

A value,
object, or
function result

R the language

Values: 1, "Florida", "2010-01-25"

Objects: `x <- c(22/7, 0.99, 3)`

To put multiple values in an object,
combine the values with `c()`

R the language

Values: 1, "Florida", "2010-01-25"

Objects: `x <- c(22/7, 0.99, 3)`

Functions: `round(x, digits = 3)`

A name
without
quotes

followed by
() to run the
function

Arguments:
values, objects, or
function results

Warm Up

Which of these are numbers?

1

"1"

"one"

one

Warm Up

Which of these are numbers?

1
number

"1"

"one"

one

Warm Up

Which of these are numbers?

1

number

"1" **"one"**

words (strings)

one

Warm Up

Which of these are numbers?

1

number

"1" **"one"**

words (strings)

one

object

Warm Up

Which of these will work? Suppose `one <- 1`.

`log(1)` `log("1")` `log("one")` `log(one)`

Warm Up

Which of these will work? Suppose `one <- 1`.

`log(1)`

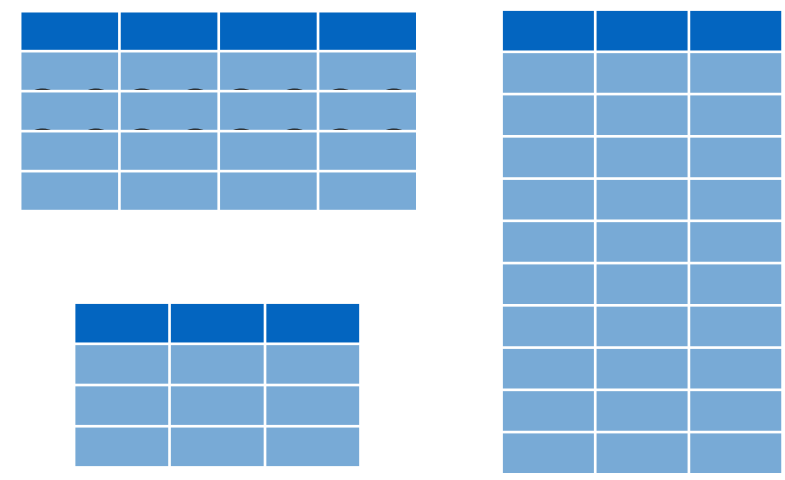
`log("1")`

`log("one")`

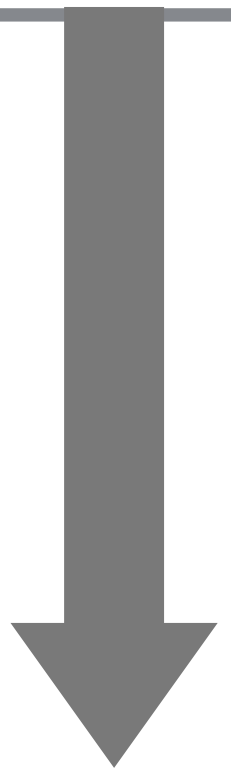
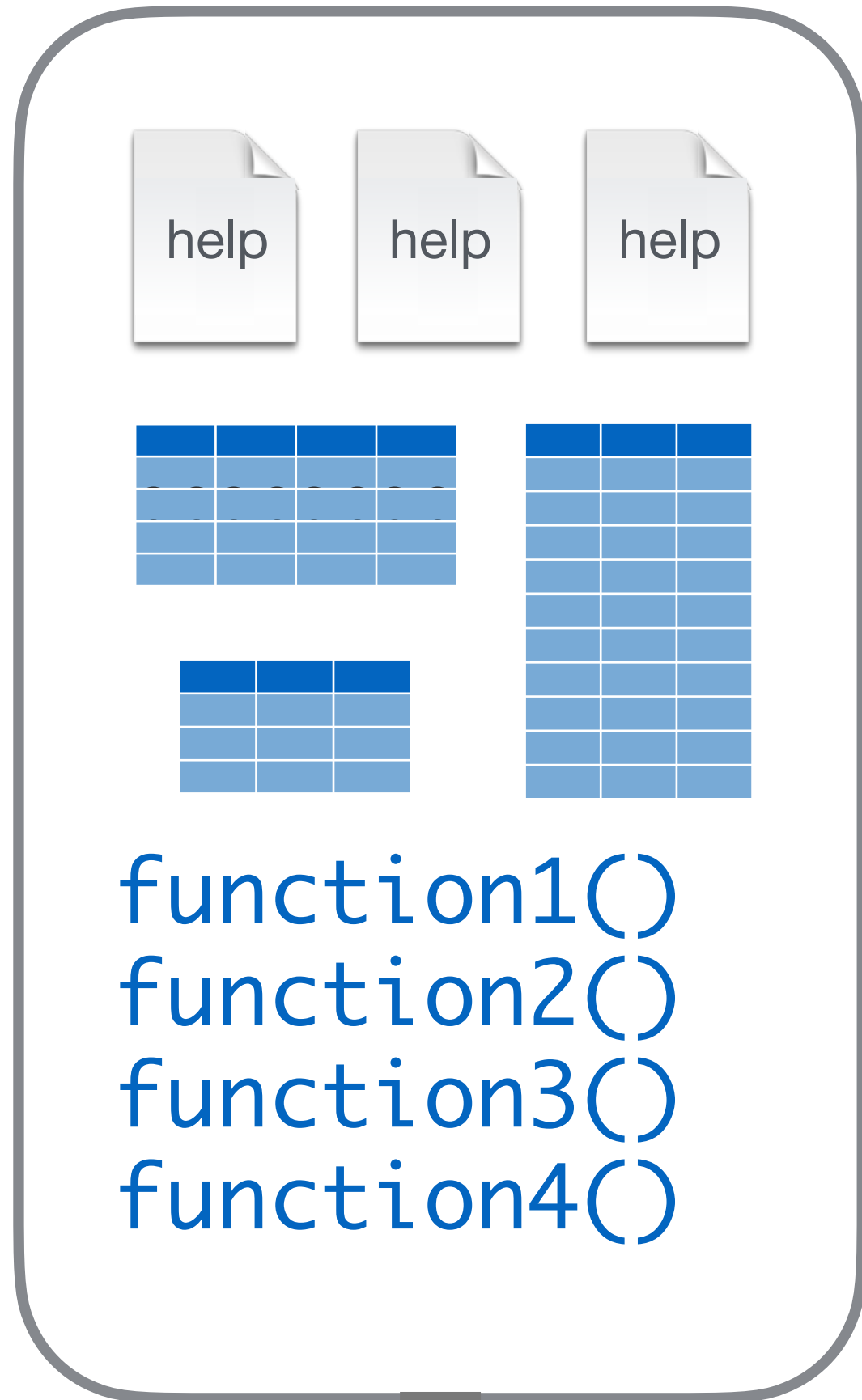
`log(one)`

R Packages

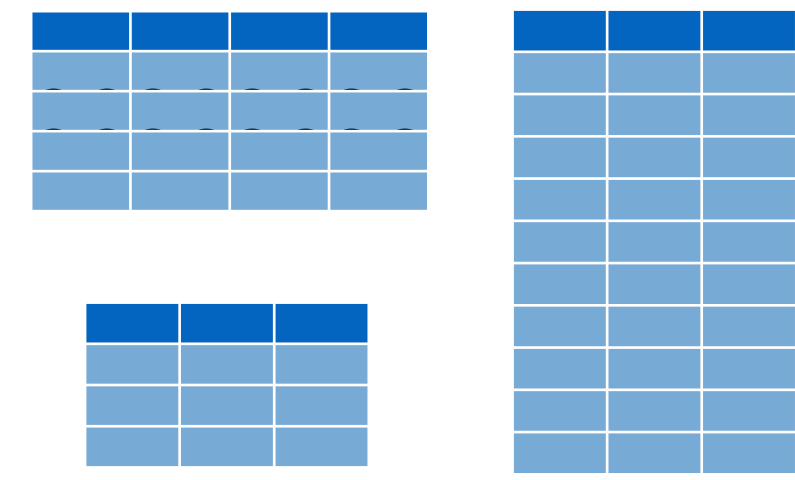




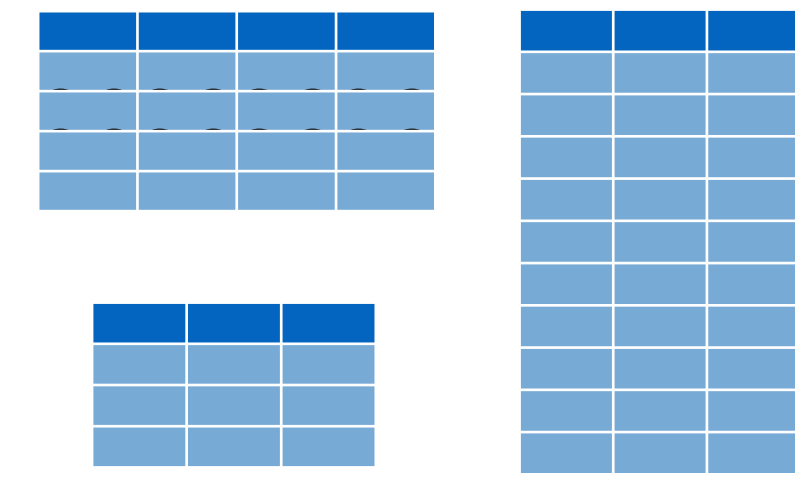
function1()
function2()
function3()
function4()



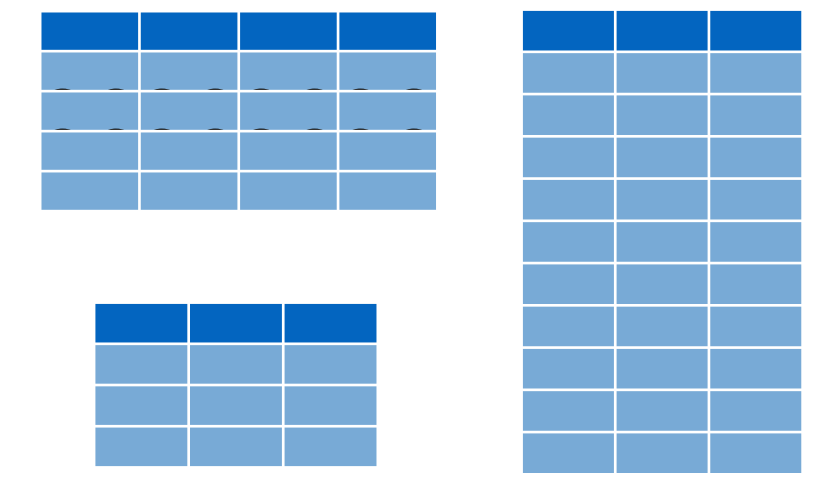
Base R



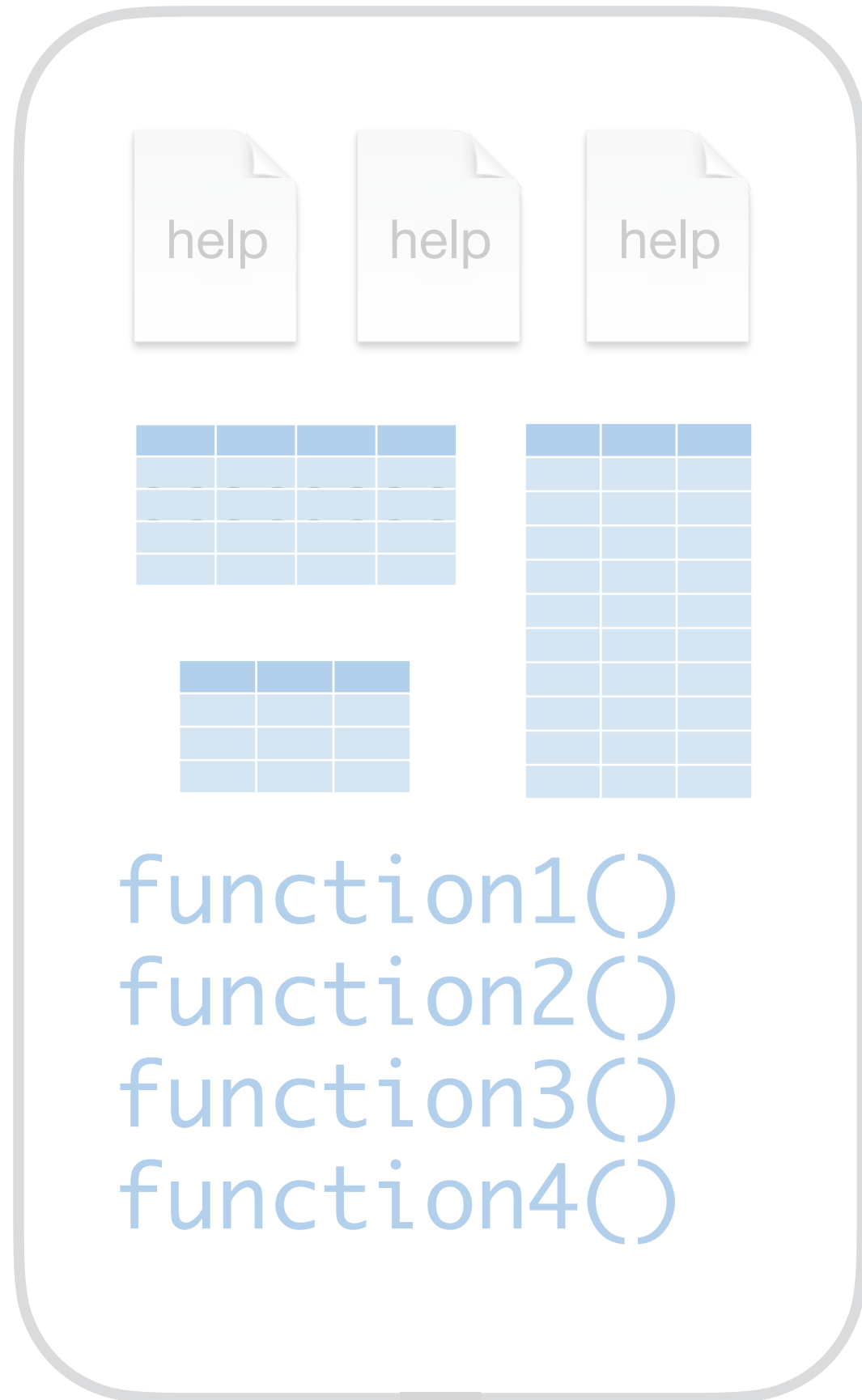
function5()
function6()
function7()
function8()



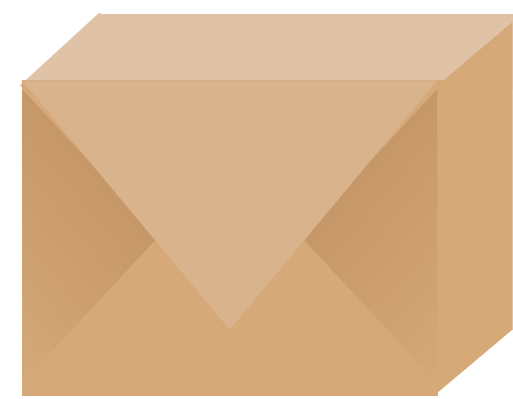
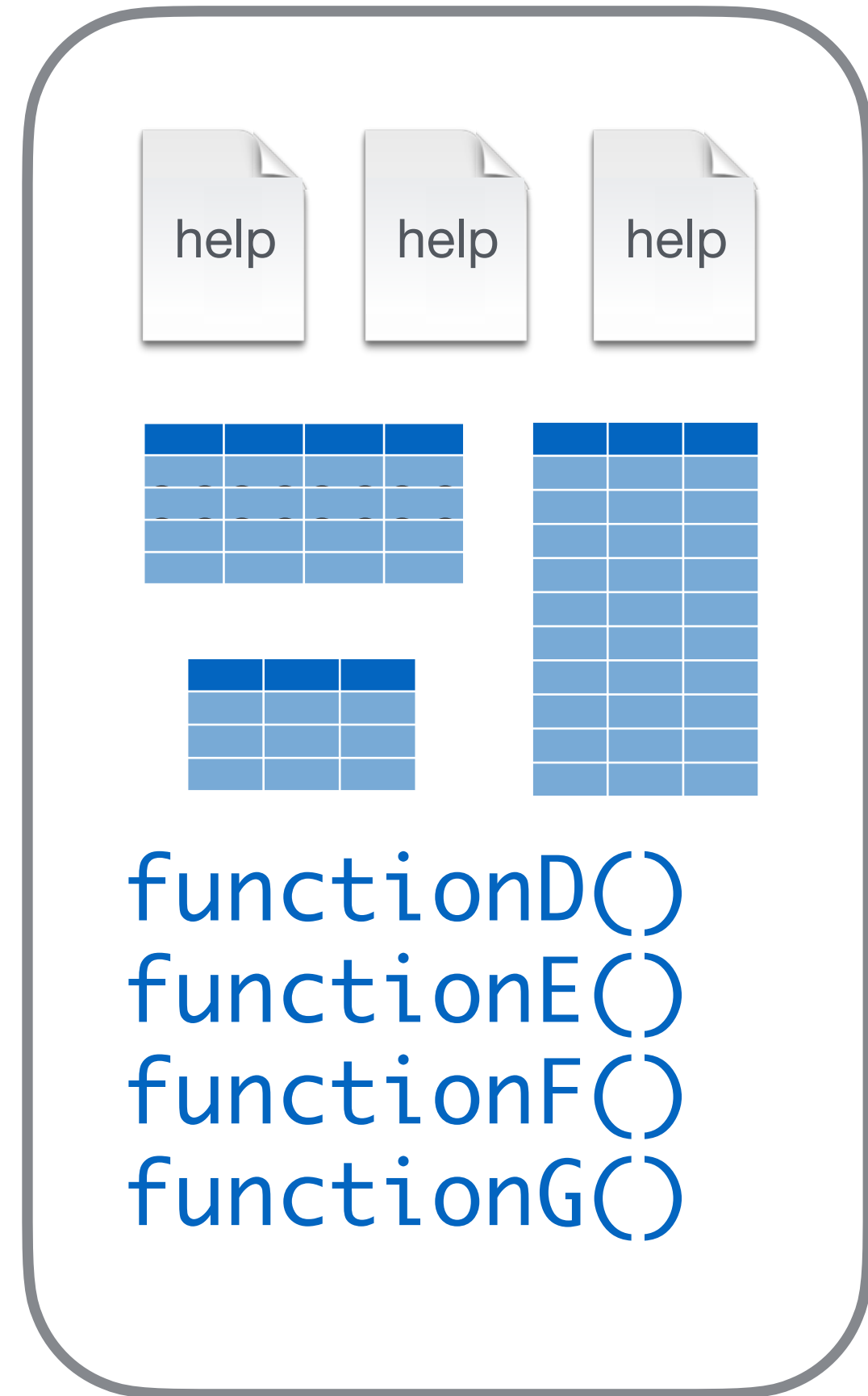
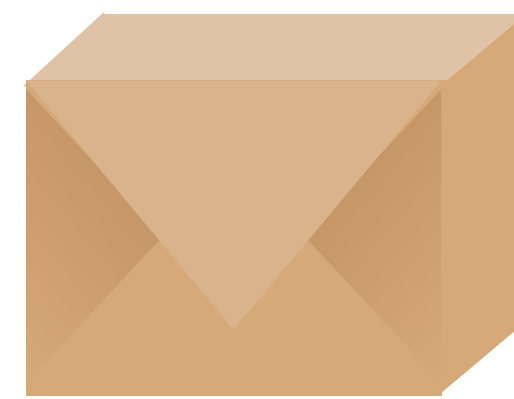
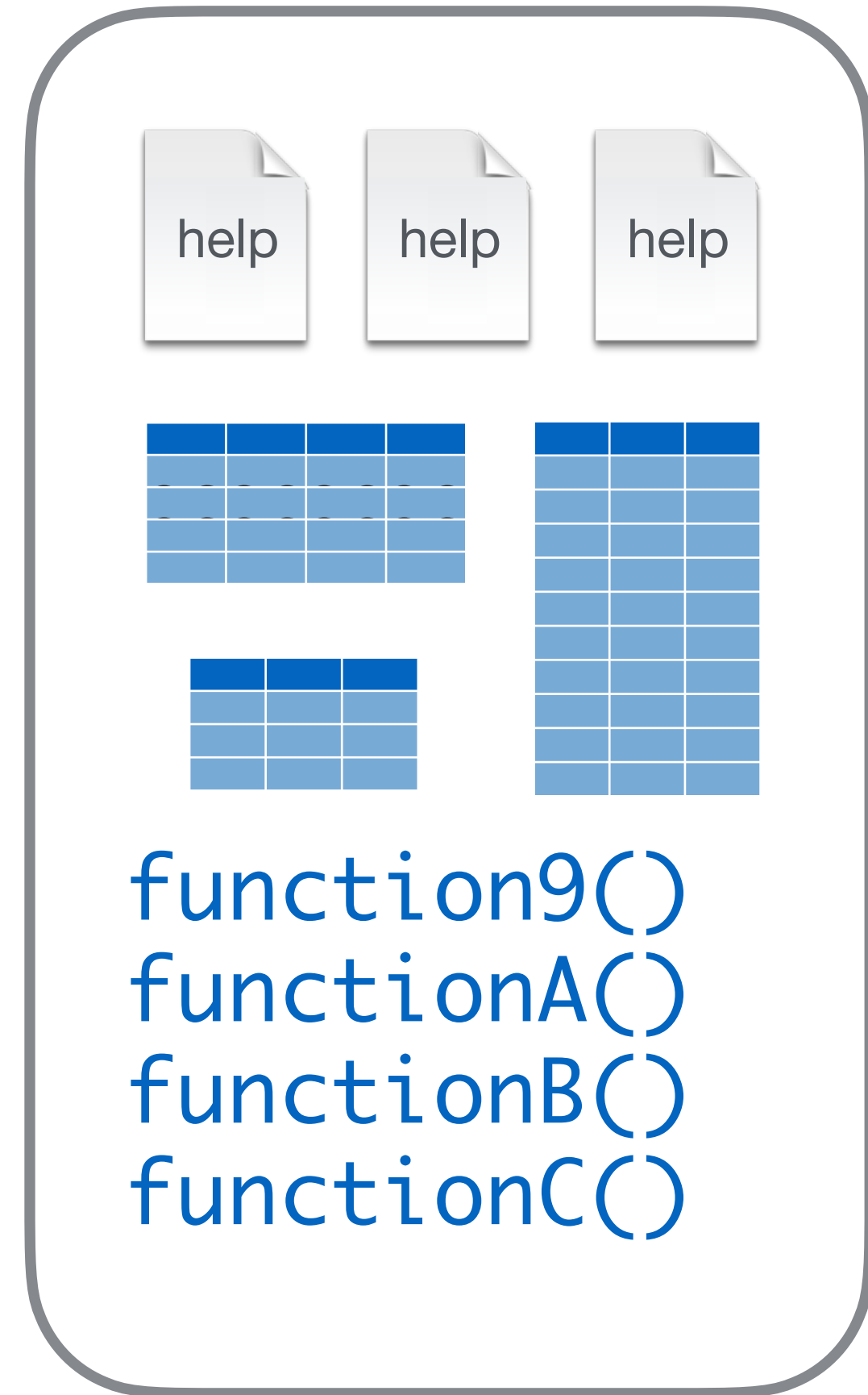
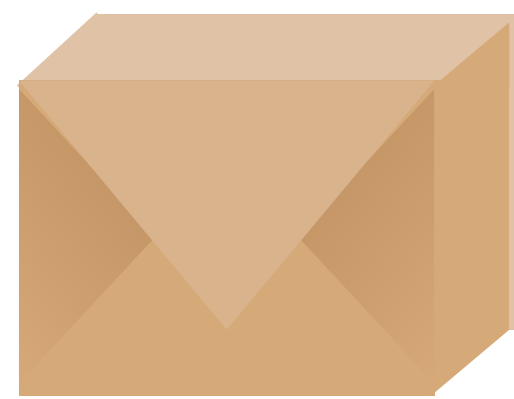
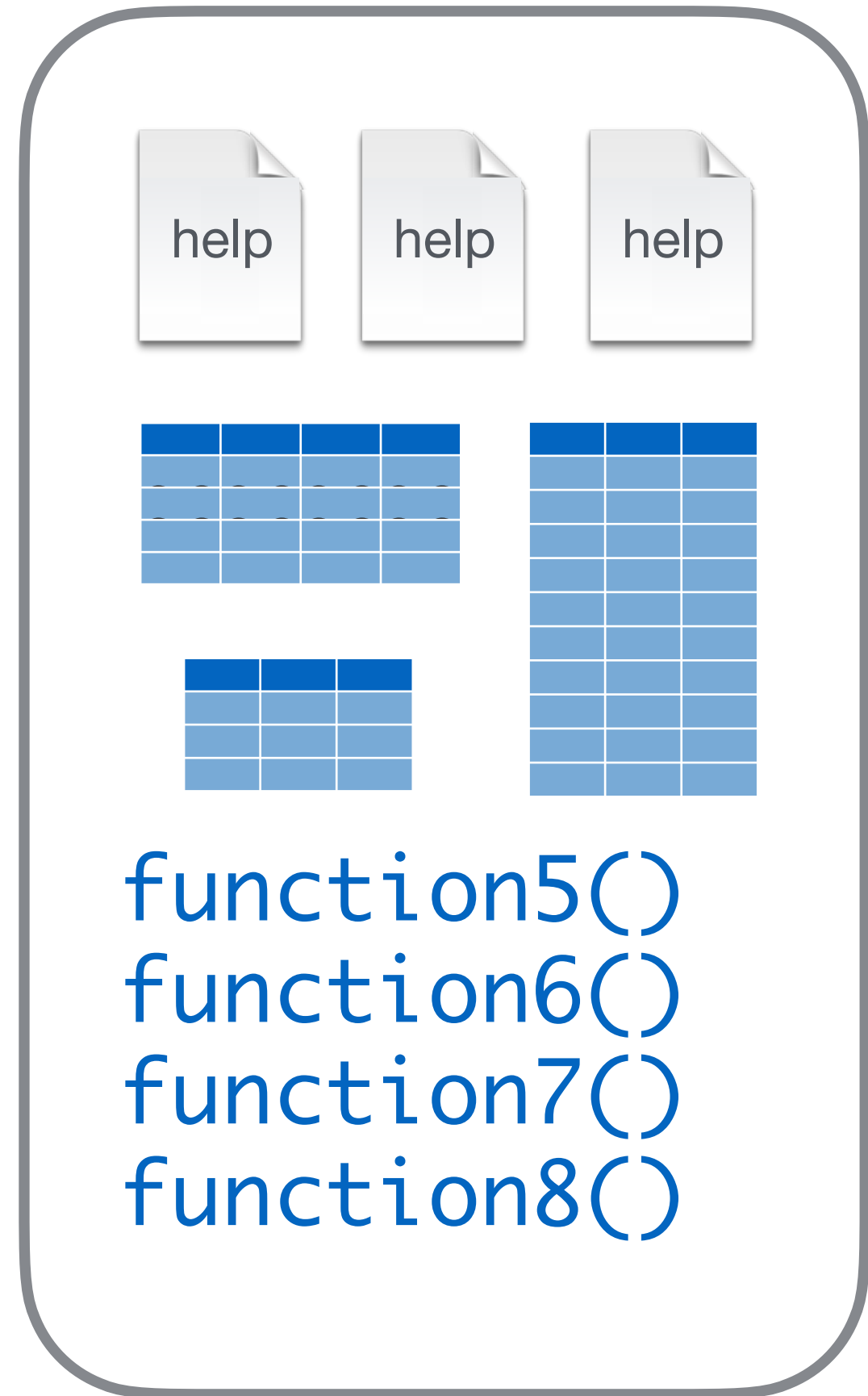
function9()
functionA()
functionB()
functionC()



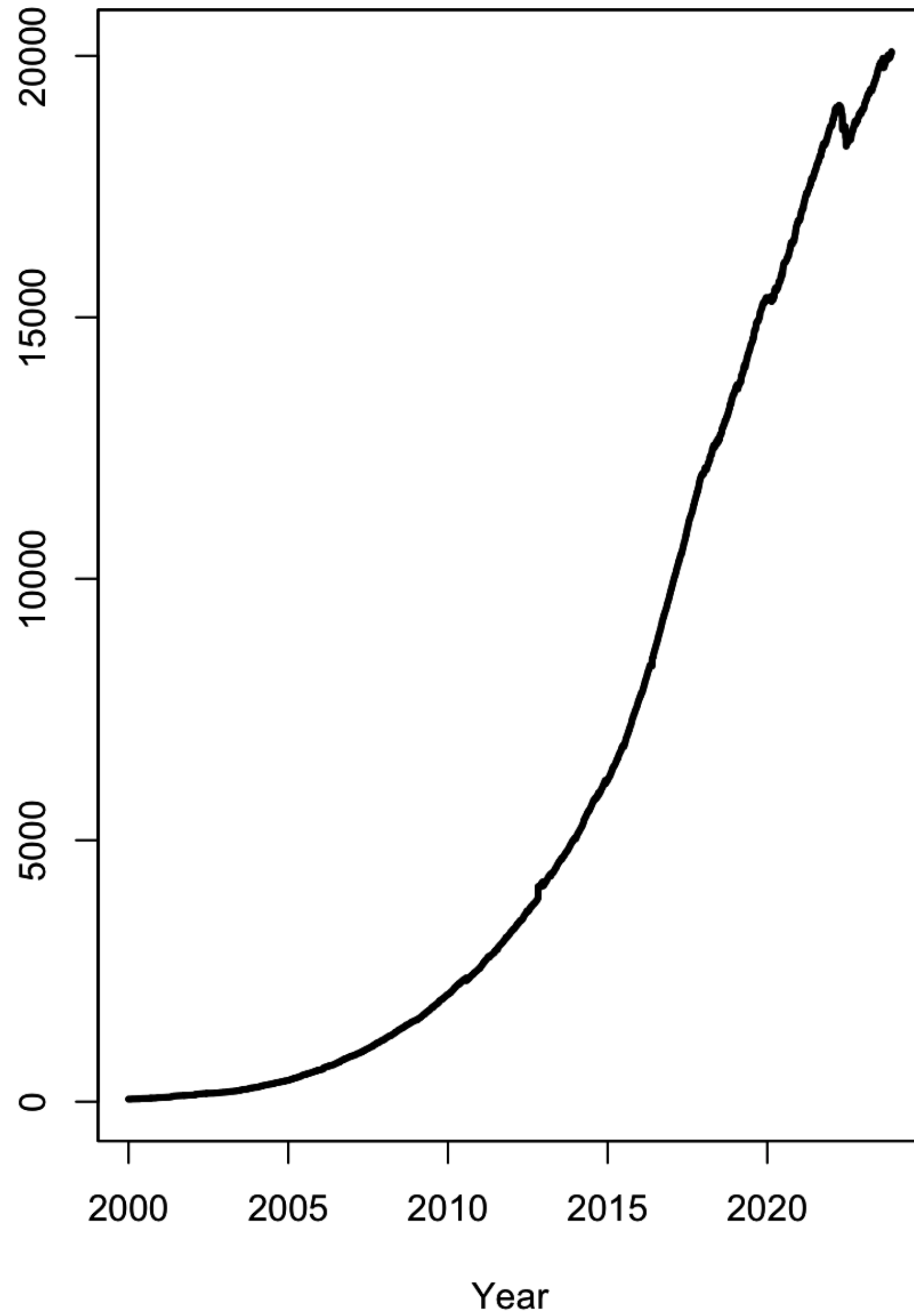
functionD()
functionE()
functionF()
functionG()



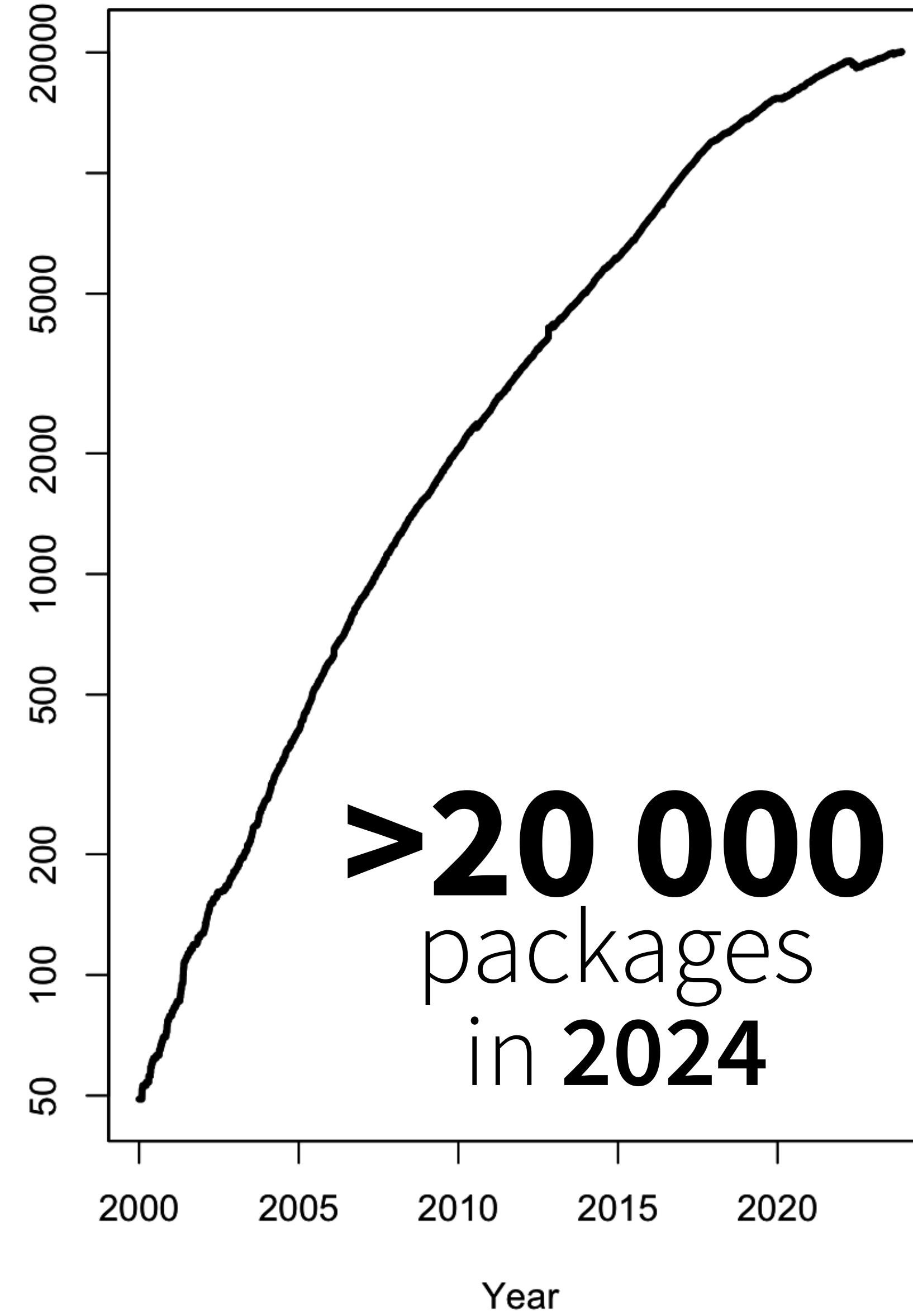
Base R



Number of CRAN Packages



Number of CRAN Packages (Log-Scale)



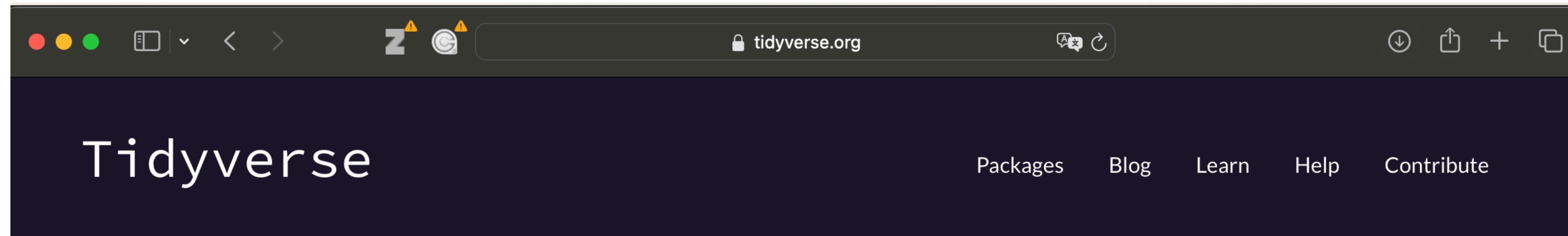


Available CRAN Packages By Name

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

A3	Accurate, Adaptable, and Accessible Error Metrics for Predictive Models
AalenJohansen	Conditional Aalen-Johansen Estimation
AATtools	Reliability and Scoring Routines for the Approach-Avoidance Task
ABACUS	Apps Based Activities for Communicating and Understanding Statistics
abasequence	Coding 'ABA' Patterns for Sequence Data
abbreviate	Readable String Abbreviation
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
abclass	Angle-Based Large-Margin Classifiers
ABCoptim	Implementation of Artificial Bee Colony (ABC) Optimization
ABCp2	Approximate Bayesian Computational Model for Estimating P2
abcrf	Approximate Bayesian Computation via Random Forests
abcrla	Asymptotically Bias-Corrected Regularized Linear Discriminant Analysis
abctools	Tools for ABC Analyses

tidyverse.org



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")
```


Using packages

1

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

Downloads files to computer

1 x per computer

Pop Quiz!

The tidyverse contains the following packages.
What function would you use to install them?

ggplot2

tibble

DBI

rvest

dplyr

hms

haven

xml2

tidyr

stringr

httr

modelr

readr

lubridate

jsonlite

tidyverse

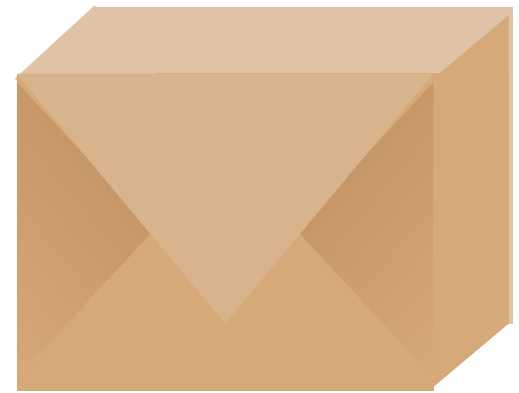
purrr

forcats

readxl


```
install.packages("ggplot2")
install.packages("dplyr")
install.packages("tidyr")
install.packages("readr")
install.packages("purrr")
install.packages("tibble")
install.packages("hms")
install.packages("stringr")
install.packages("lubridate")
install.packages("forcats")
install.packages("DBI")
install.packages("haven")
install.packages("httr")
install.packages("jsonlite")
install.packages("readxl")
install.packages("rvest")
install.packages("xml2")
install.packages("modelr")
install.packages("broom")
```


tidyverse



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

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

A shortcut!


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

does the equivalent of

```
install.packages("ggplot2")  
install.packages("dplyr")  
install.packages("tidyr")  
install.packages("readr")  
install.packages("purrr")  
install.packages("tibble")  
install.packages("hms")  
install.packages("stringr")  
install.packages("lubridate")  
install.packages("forcats")  
install.packages("DBI")  
install.packages("haven")  
install.packages("httr")  
install.packages("jsonlite")  
install.packages("readxl")  
install.packages("rvest")  
install.packages("xml2")  
install.packages("modelr")  
install.packages("broom")
```


Using packages

1

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

Downloads files to computer

1 x per computer

2

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

Loads package

1 x per R Session


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

does the equivalent of

```
install.packages("ggplot2")
install.packages("dplyr")
install.packages("tidyr")
install.packages("readr")
install.packages("purrr")
install.packages("tibble")
install.packages("stringr")
install.packages("forcats")
install.packages("lubridate")
install.packages("hms")
install.packages("DBI")
install.packages("haven")
install.packages("httr")
install.packages("jsonlite")
install.packages("readxl")
install.packages("rvest")
install.packages("xml2")
install.packages("modelr")
install.packages("broom")
```

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

does the equivalent of

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


Quarto



Your Turn

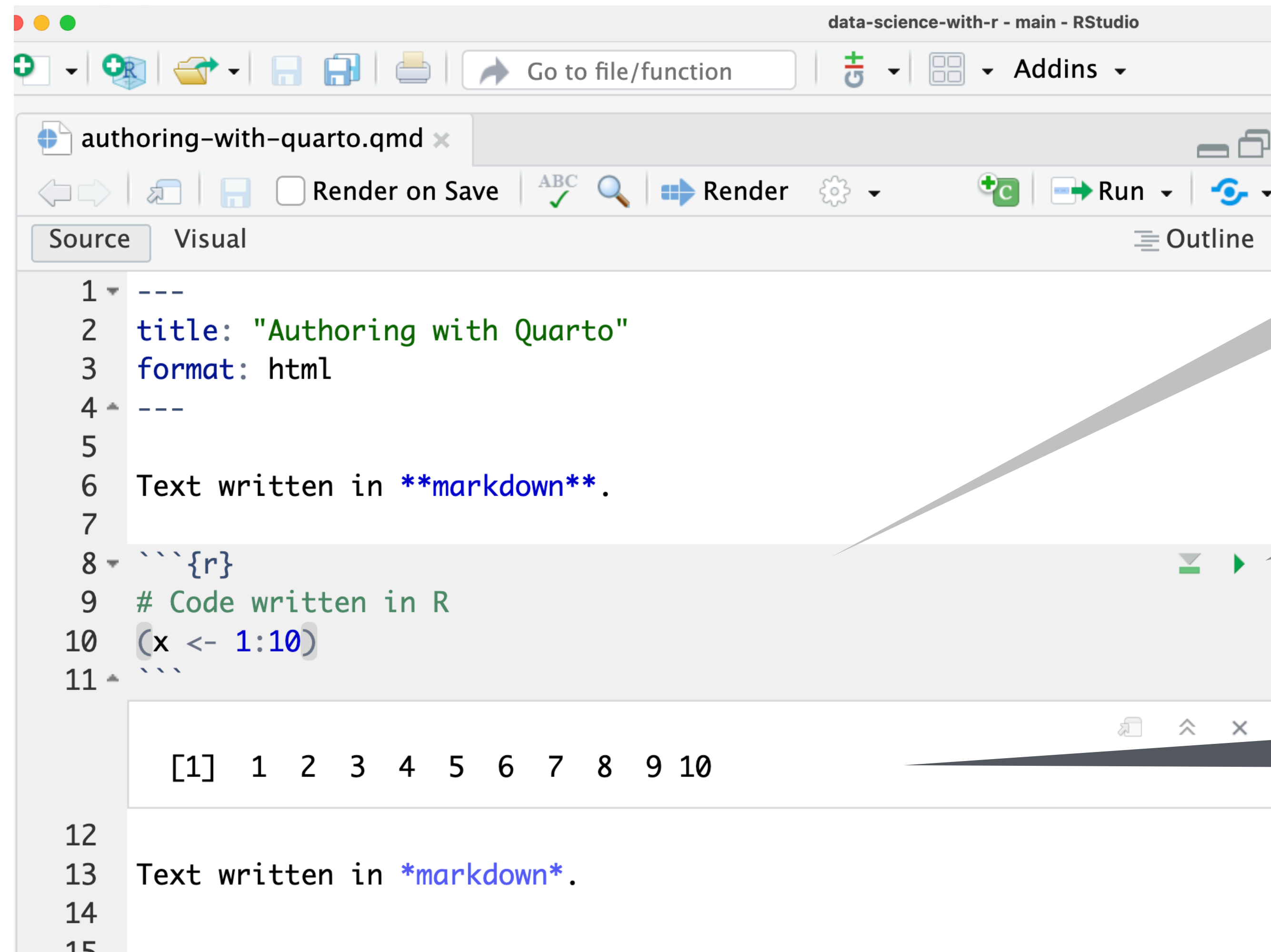
Navigate to the main page of the class: **<https://astamm.github.io/data-science-with-r/>**.

Download **01-Introduction-Exercises.qmd** from the outline table and open it. Read through the file and do everything it tells you to do.

05:00

Quarto

An authoring format for Data Science.



The screenshot shows the RStudio interface with a Quarto document open. The document contains the following content:

```
1 ---  
2 title: "Authoring with Quarto"  
3 format: html  
4 ---  
5  
6 Text written in markdown.  
7  
8 ```{r}  
9 # Code written in R  
10 (x <- 1:10)  
11 ```  
12  
13 Text written in markdown.  
14  
15
```

The code chunk on lines 8-11 is highlighted in light green. Below the code, the console output shows the result of the R code: [1] 1 2 3 4 5 6 7 8 9 10.

Code goes in a chunk

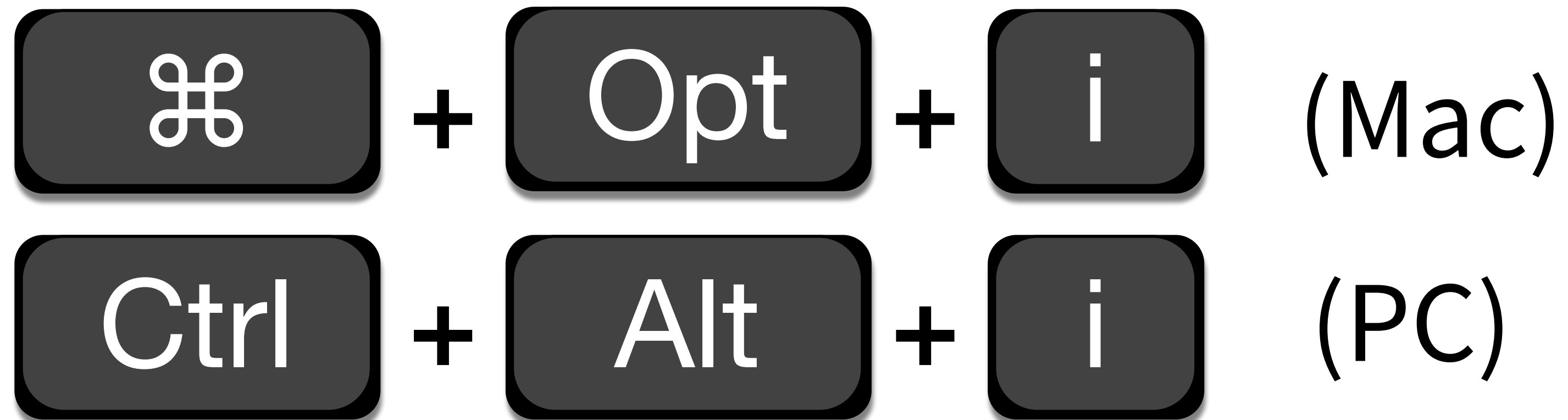
Click to run code in chunk

Code result

Code chunks

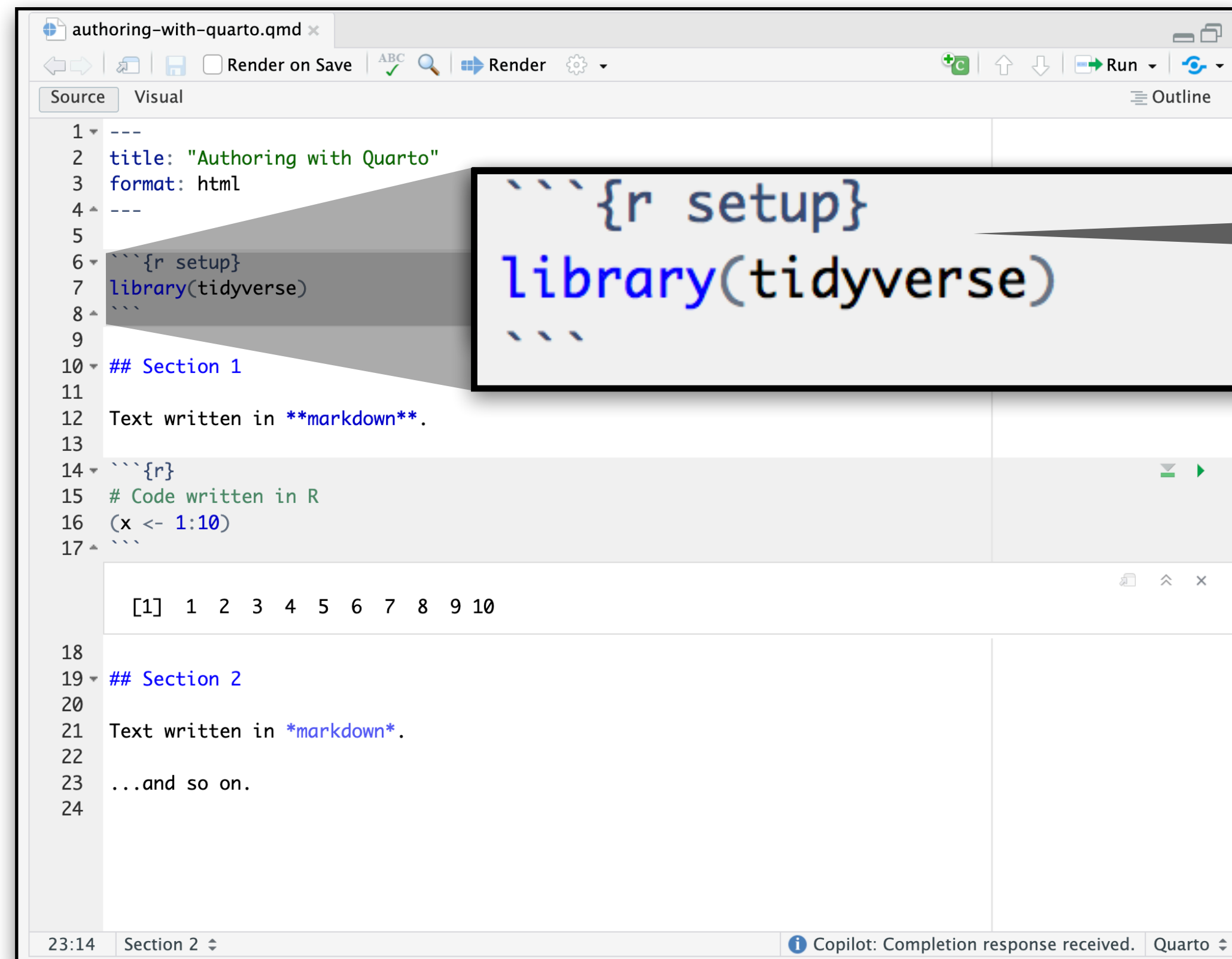
Insert a chunk of R code with

```
` `` {r}  
# some code  
` ``
```



Setup

The setup chunk is always run once before anything else



```
1 ---  
2 title: "Authoring with Quarto"  
3 format: html  
4 ---  
5  
6 ```{r setup}  
7 library(tidyverse)  
8 ```  
9  
10 ## Section 1  
11  
12 Text written in markdown.  
13  
14 ```{r}  
15 # Code written in R  
16 (x <- 1:10)  
17 ```  
18  
19 ## Section 2  
20  
21 Text written in markdown.  
22  
23 ...and so on.  
24
```

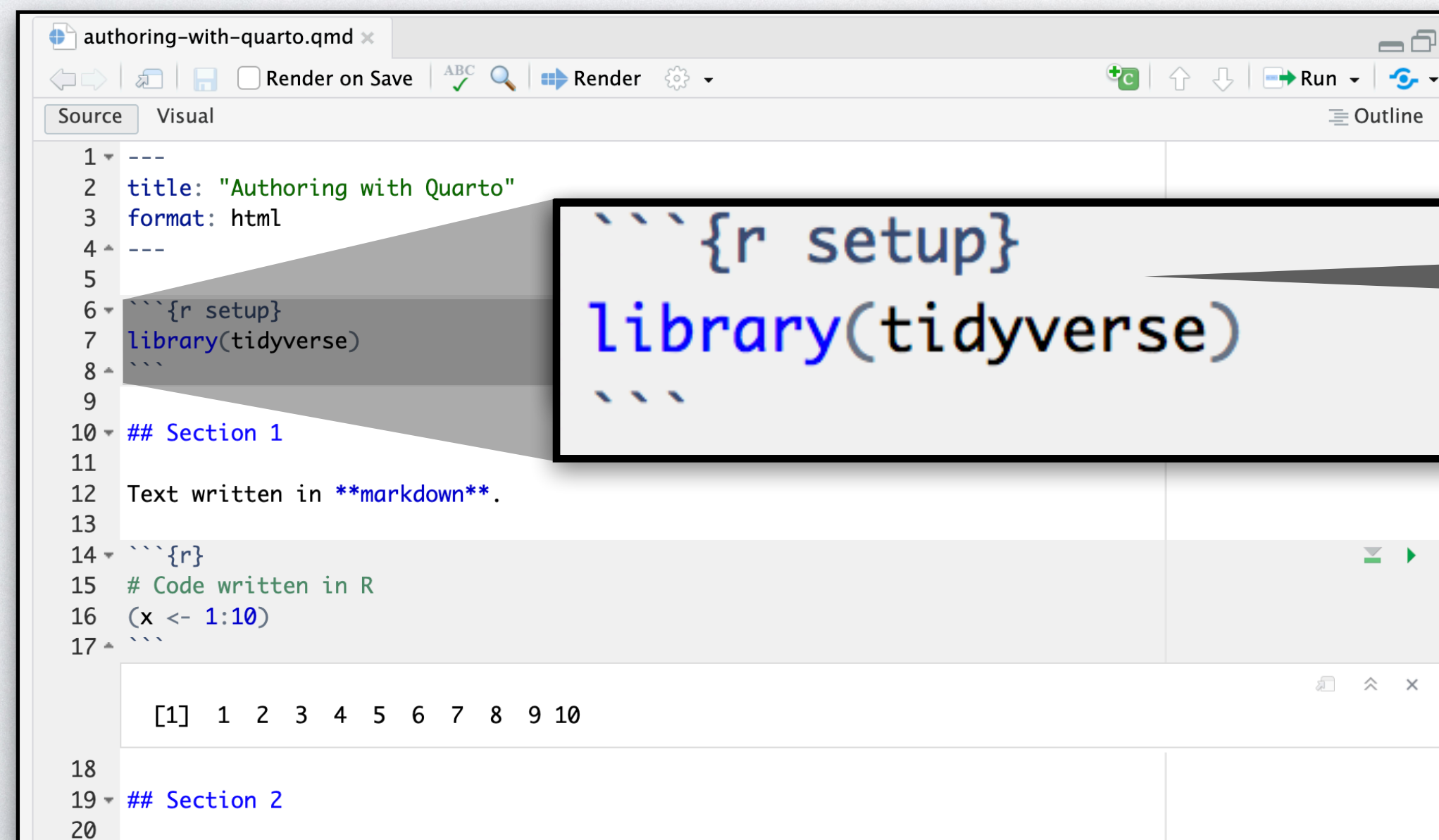
[1] 1 2 3 4 5 6 7 8 9 10

```
```{r setup}  
library(tidyverse)
```
```

chunk labels are optional,
the setup label is special

Your Turn

Add a setup chunk to the top of **01-Introduction-Exercises.qmd**. Use it to load the tidyverse package, then uncomment and run the **ggplot2** chunk at the bottom of your file.



```
1 ---  
2 title: "Authoring with Quarto"  
3 format: html  
4 ---  
5  
6 ```{r setup}  
7 library(tidyverse)  
8 ```  
9  
10 ## Section 1  
11  
12 Text written in markdown.  
13  
14 ```{r}  
15 # Code written in R  
16 (x <- 1:10)  
17 ```  
18  
19 ## Section 2  
20
```

[1] 1 2 3 4 5 6 7 8 9 10

```
```{r setup}  
library(tidyverse)
```
```

chunk labels are optional,
the setup label is special

02:00



Your Turn

Navigate to the main page of the class: **<https://astamm.github.io/data-science-with-r/>**.

Download **02-Visualize-Exercises.qmd** from the outline table and open it.

Add a setup chunk that loads the tidyverse packages.

02:00