

# Welcome to the Tidyverse



**HELLO**  
my name is

**Aymeric**

 [@aymeric-stamm](https://www.linkedin.com/in/aymeric-stamm)

[aymeric.stamm@cnrs.fr](mailto:aymeric.stamm@cnrs.fr)

**HELLO**  
my name is

**Manon**

 [@manon-simonot](https://www.linkedin.com/in/manon-simonot)

[manon.simonot@univ-nantes.fr](mailto:manon.simonot@univ-nantes.fr)

O'REILLY®



# Hands-On Programming with R

WRITE YOUR OWN FUNCTIONS AND SIMULATIONS

Garrett Grolemund  
Foreword by Hadley Wickham

O'REILLY®



# R for Data Science

VISUALIZE, MODEL, TRANSFORM, TIDY, AND IMPORT DATA

Hadley Wickham &  
Garrett Grolemund

The R Series

## R Markdown

The Definitive Guide



Yihui Xie  
J. J. Allaire  
Garrett Grolemund

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

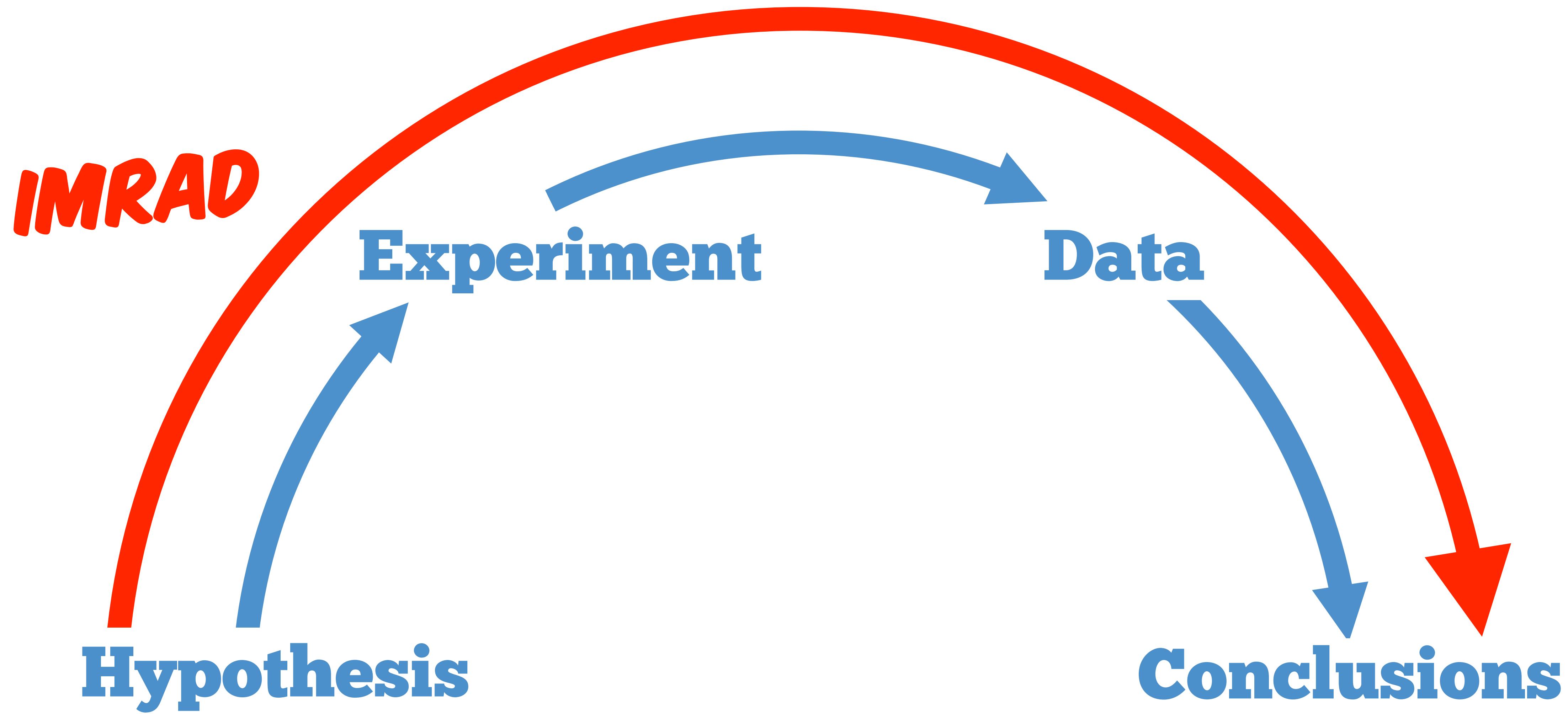
What answer was discovered?

## And Discussion

What does the answer imply?



# (Experimental) Science

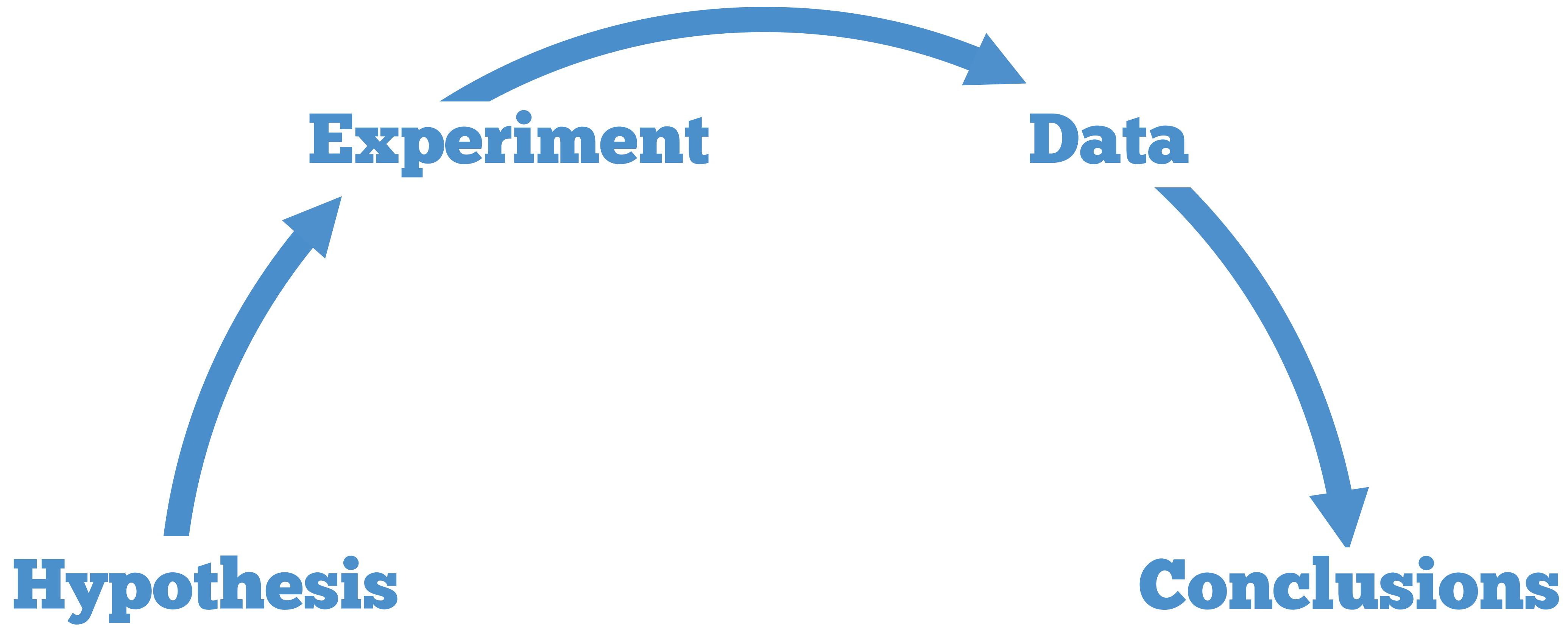


# Pop Quiz

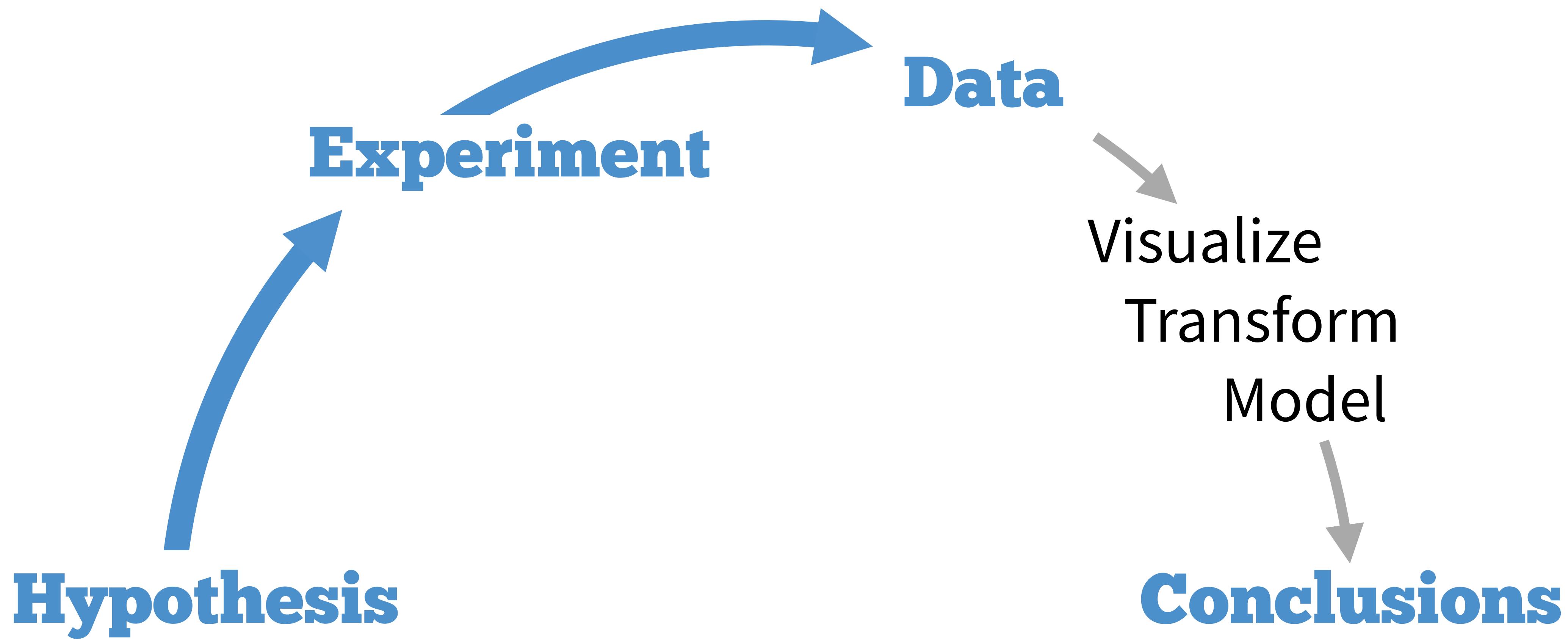
What is **data science**? Poll your neighbors.



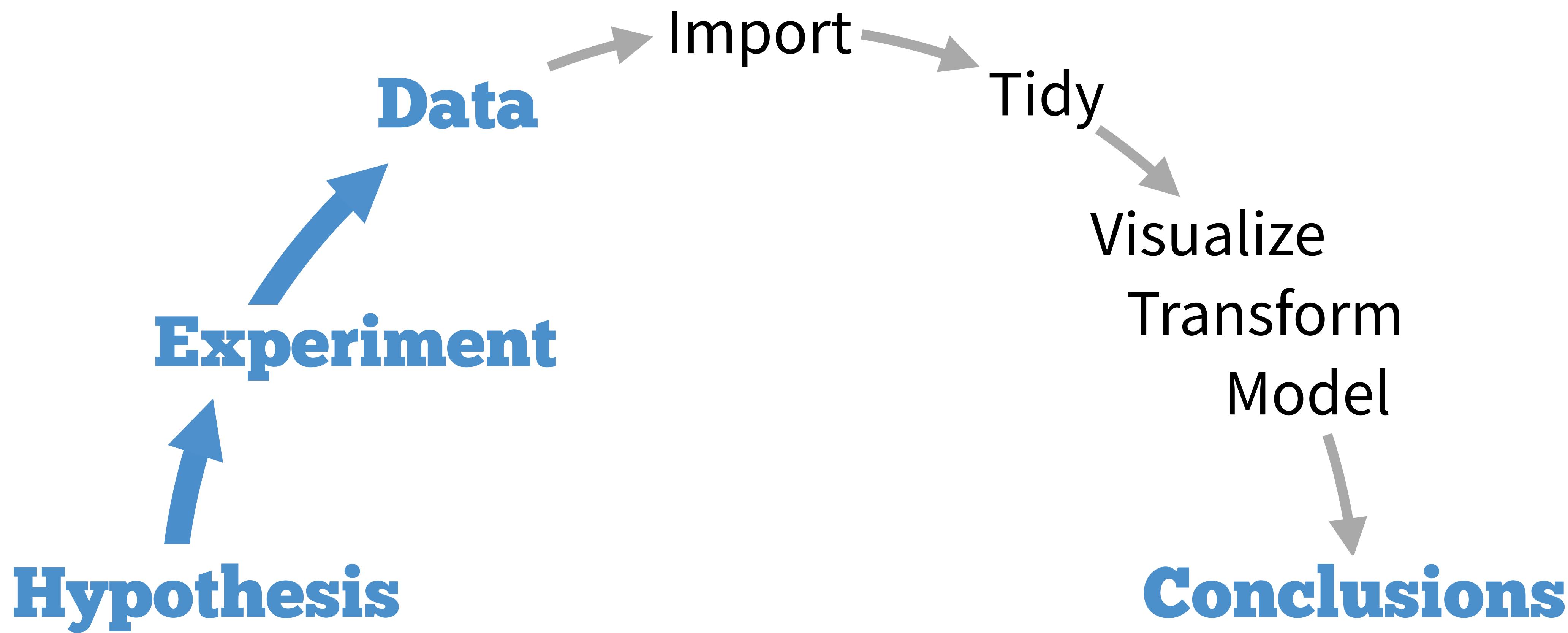
# Data Science



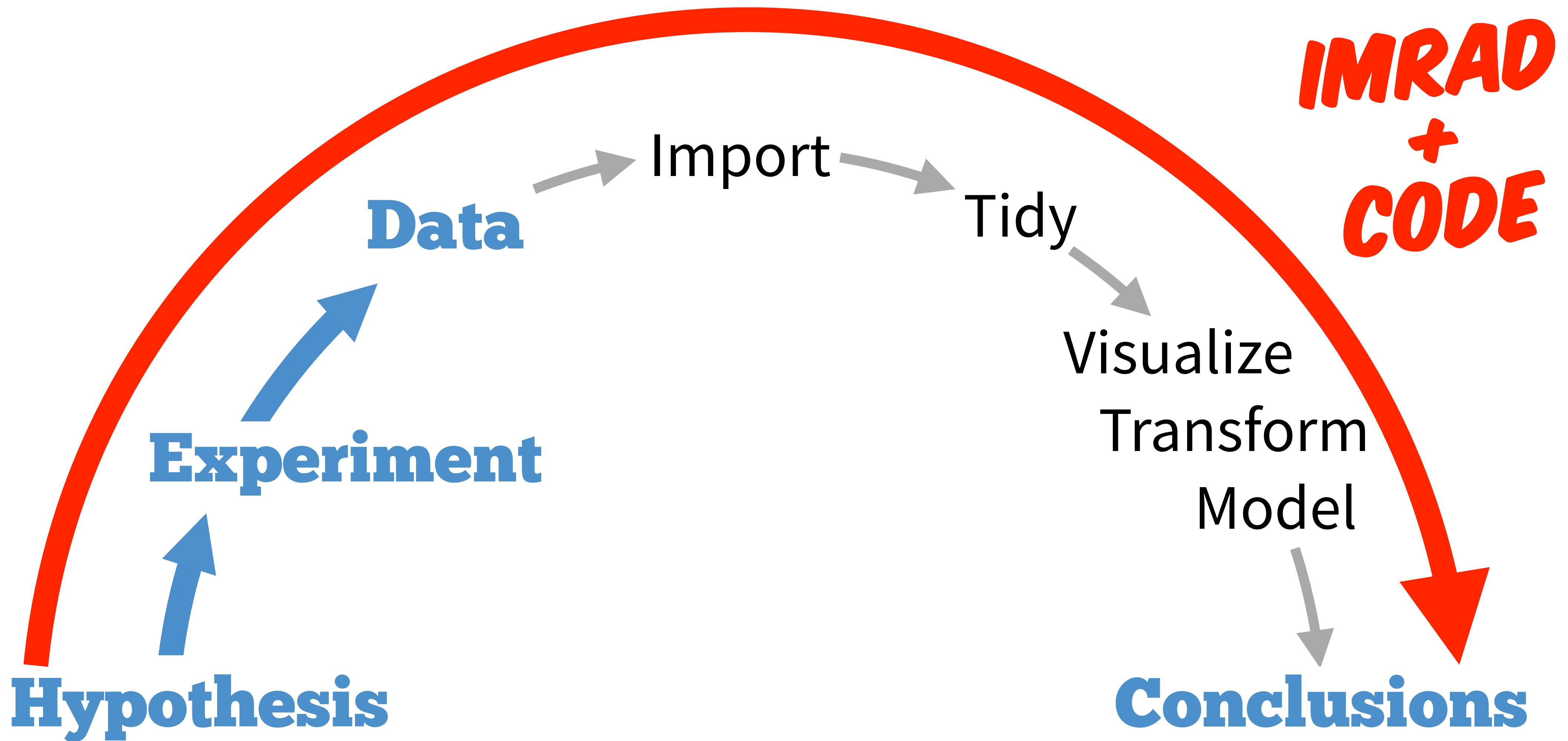
# Data Science



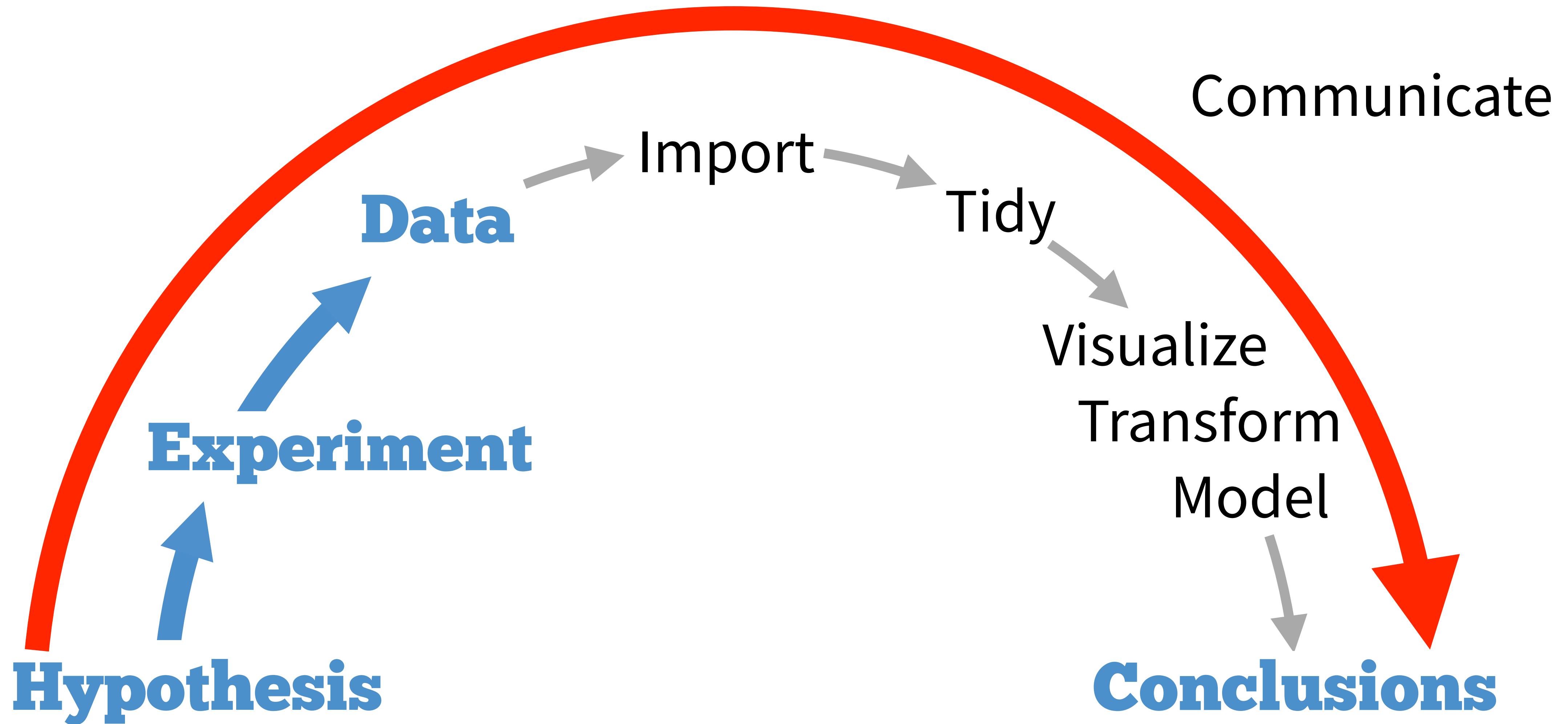
# Data Science



# Data Science



# Data Science



# Data Science

Import

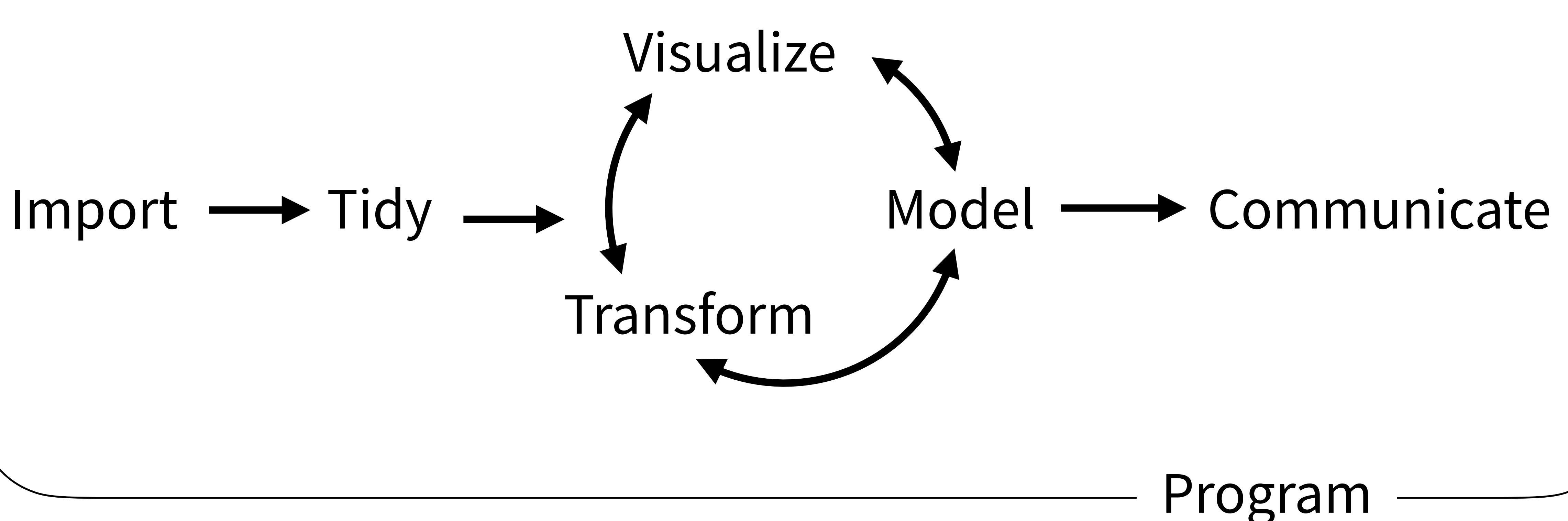
Tidy

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/>



# R the language

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

# R the language

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

**Objects:** x <- 22/7

A name  
without quotes

< 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

"1"

"one"

one

---

number

# Warm Up

Which of these are numbers?

1

number

"1"

words (strings)

"one"

one

# Warm Up

Which of these are numbers?

1

number

"1"

words (strings)

"one"

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)`

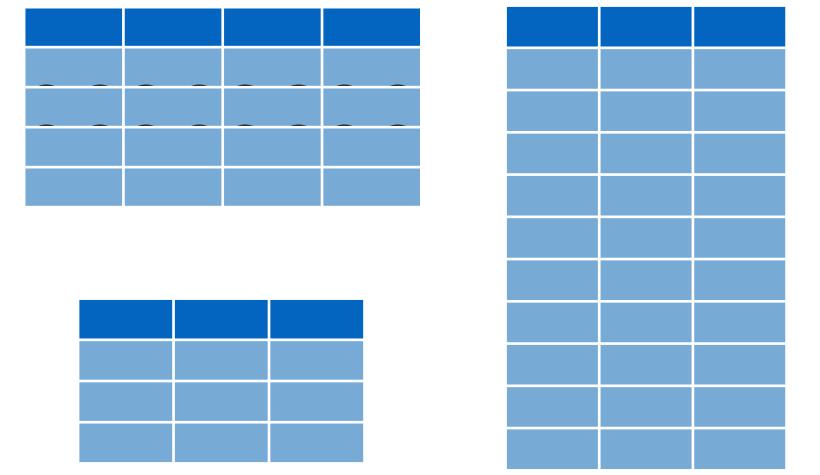
`log("1")`

`log("one")`

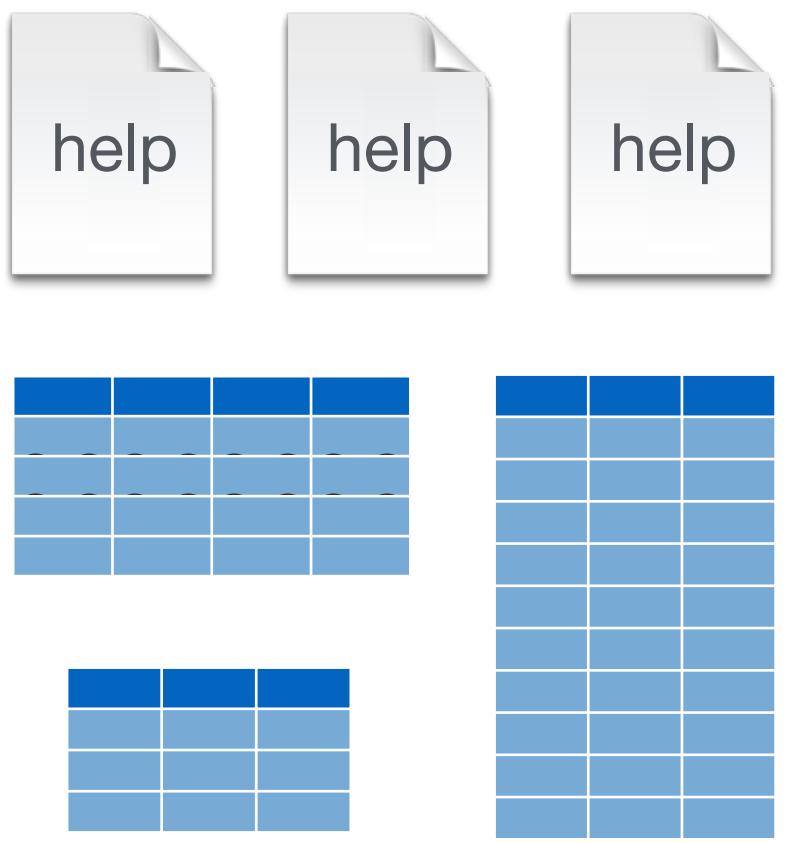
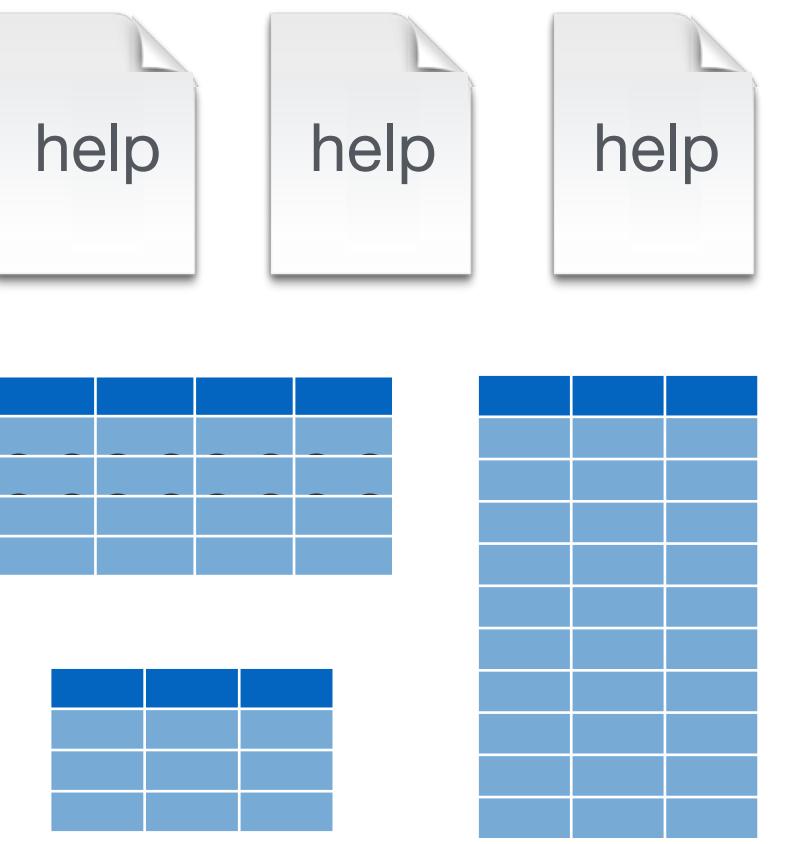
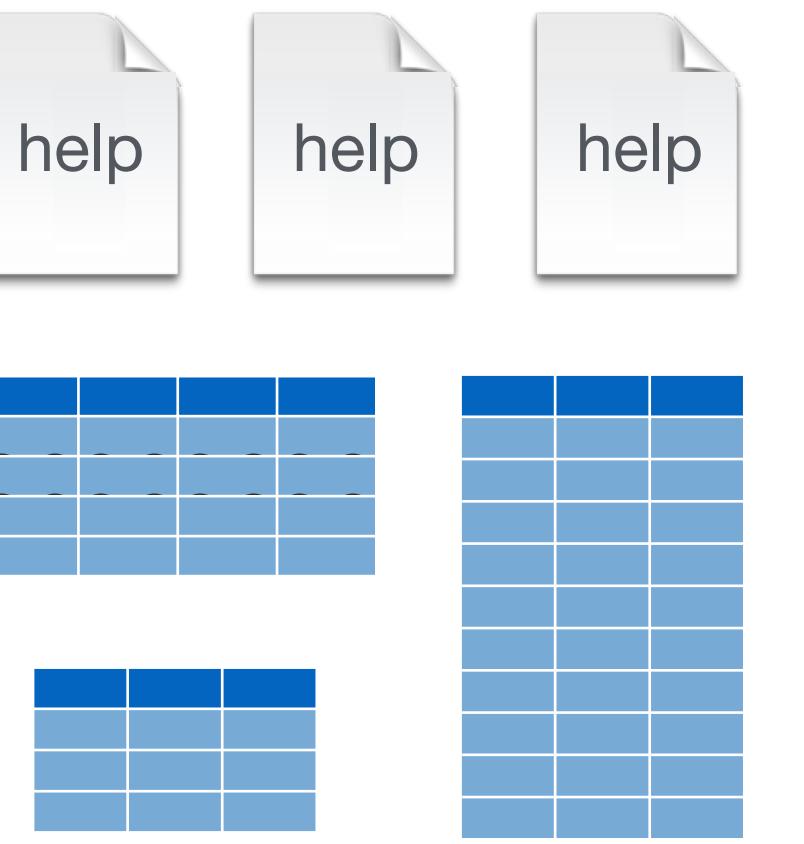
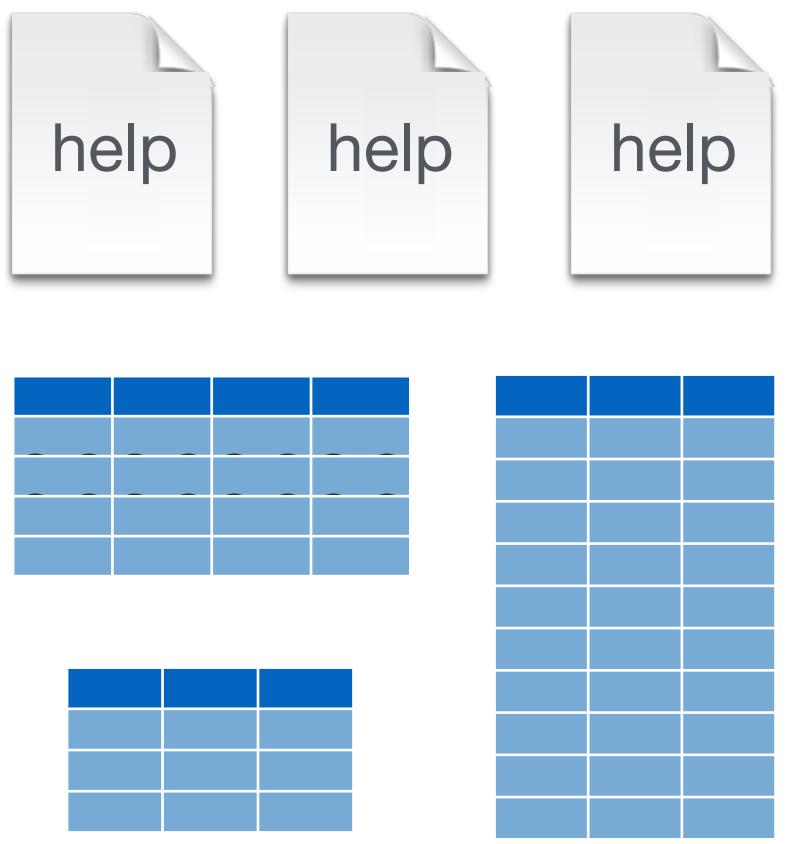
`log(one)`

# R Packages

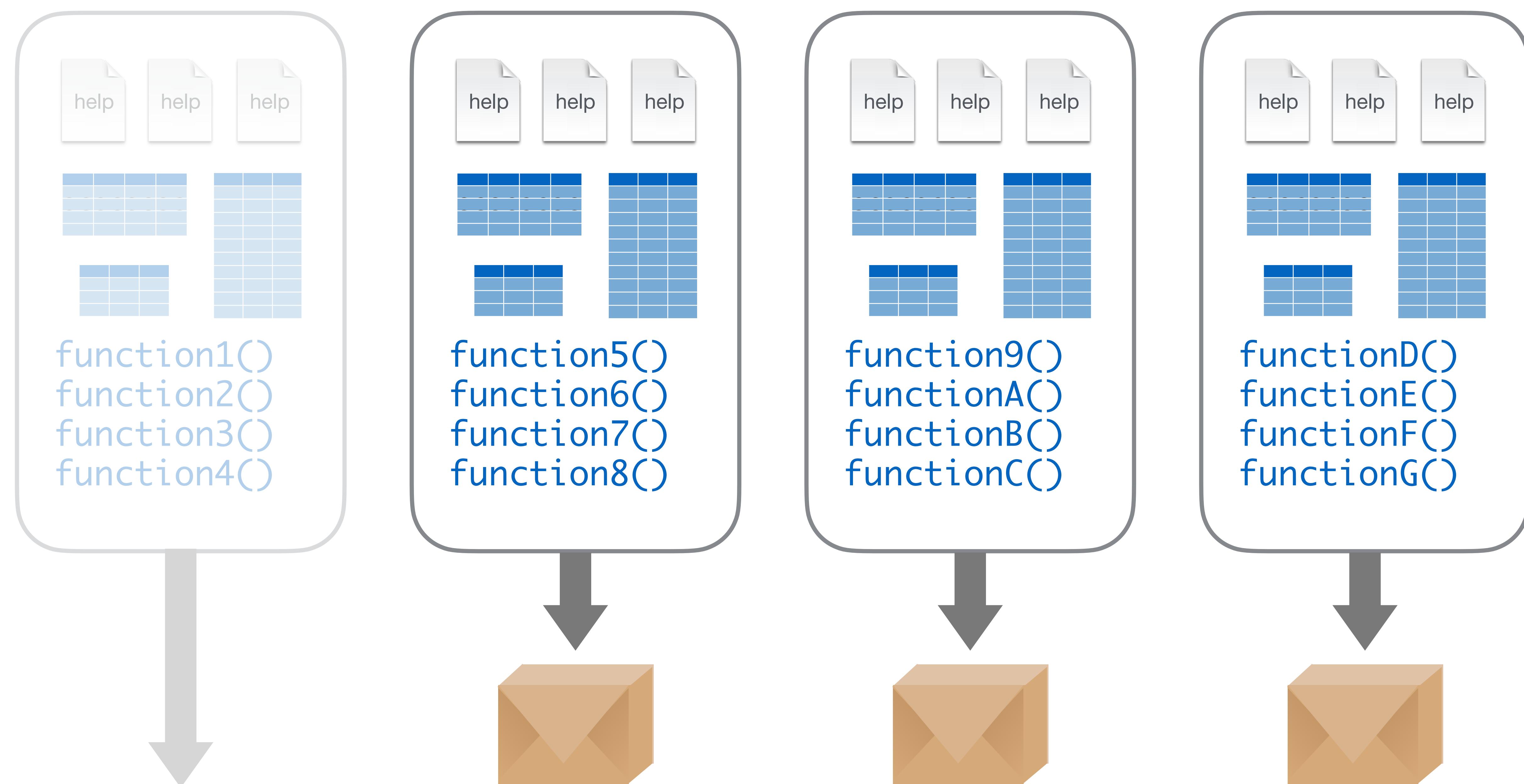




function1()  
function2()  
function3()  
function4()



Base R

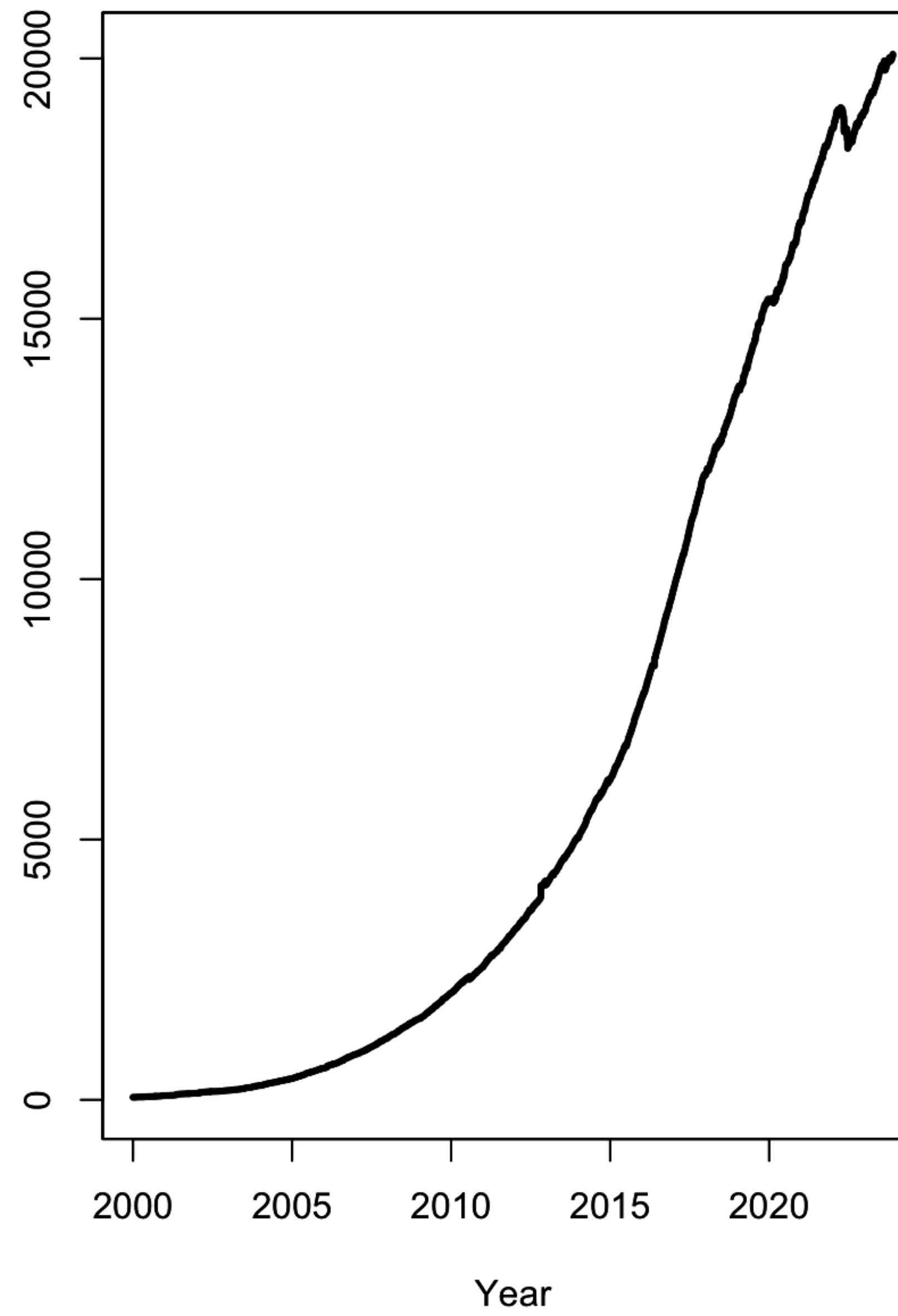


Base R

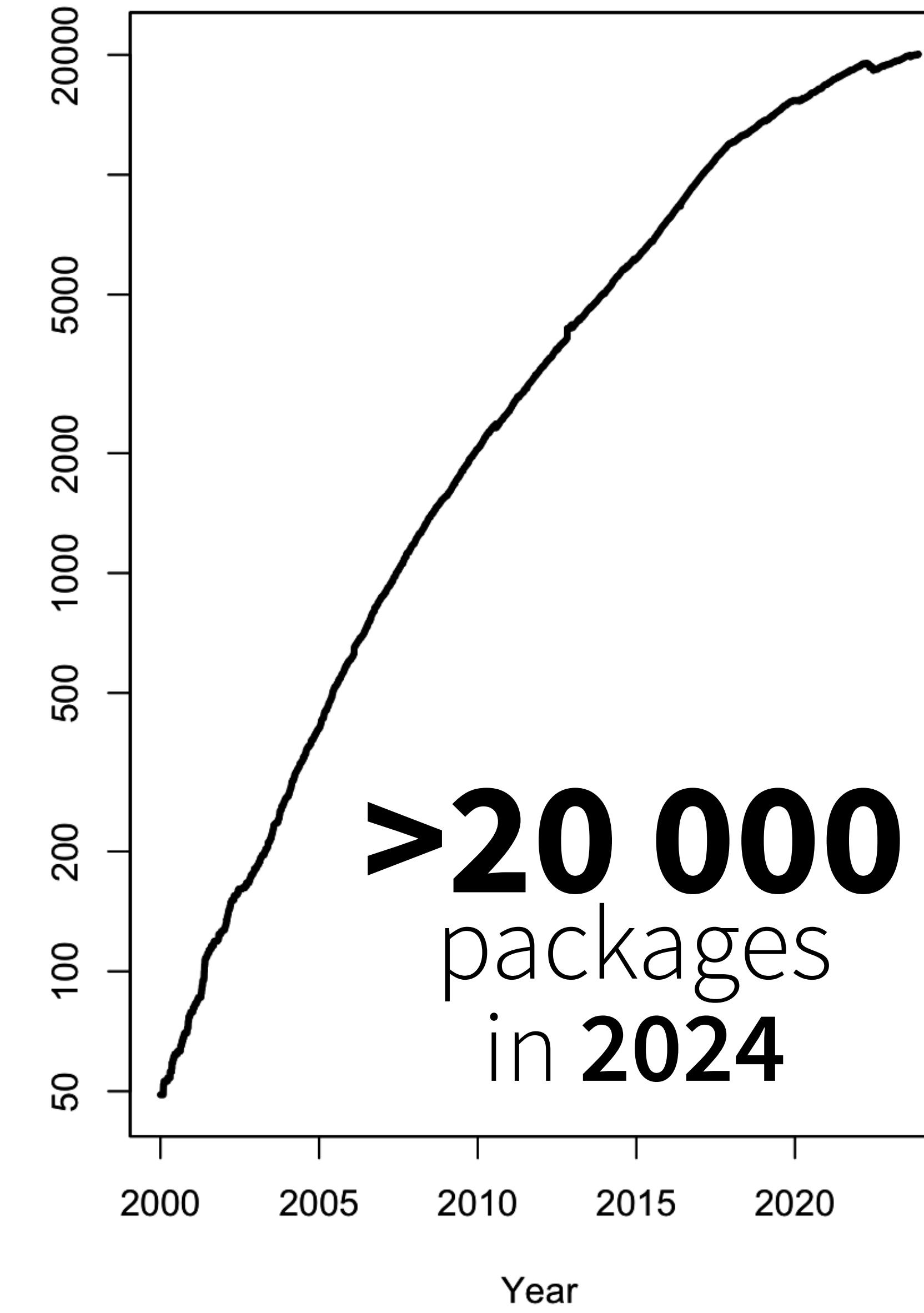
32

R Packages

**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](#)

[AalenJohansen](#)

[AATtools](#)

[ABACUS](#)

[abasequence](#)

[abbreviate](#)

[abc](#)

[abc.data](#)

[ABC.RAP](#)

[ABCanalysis](#)

[abclass](#)

[ABCOptim](#)

[ABCp2](#)

[abcrf](#)

[abcrlda](#)

[abctools](#)

Accurate, Adaptable, and Accessible Error Metrics for Predictive Models

Conditional Aalen-Johansen Estimation

Reliability and Scoring Routines for the Approach-Avoidance Task

Apps Based Activities for Communicating and Understanding Statistics

Coding 'ABA' Patterns for Sequence Data

Readable String Abbreviation

Tools for Approximate Bayesian Computation (ABC)

Data Only: Tools for Approximate Bayesian Computation (ABC)

Array Based CpG Region Analysis Pipeline

Computed ABC Analysis

Angle-Based Large-Margin Classifiers

Implementation of Artificial Bee Colony (ABC) Optimization

Approximate Bayesian Computational Model for Estimating P2

Approximate Bayesian Computation via Random Forests

Asymptotically Bias-Corrected Regularized Linear Discriminant Analysis

Tools for ABC Analyses

# tidyverse.org

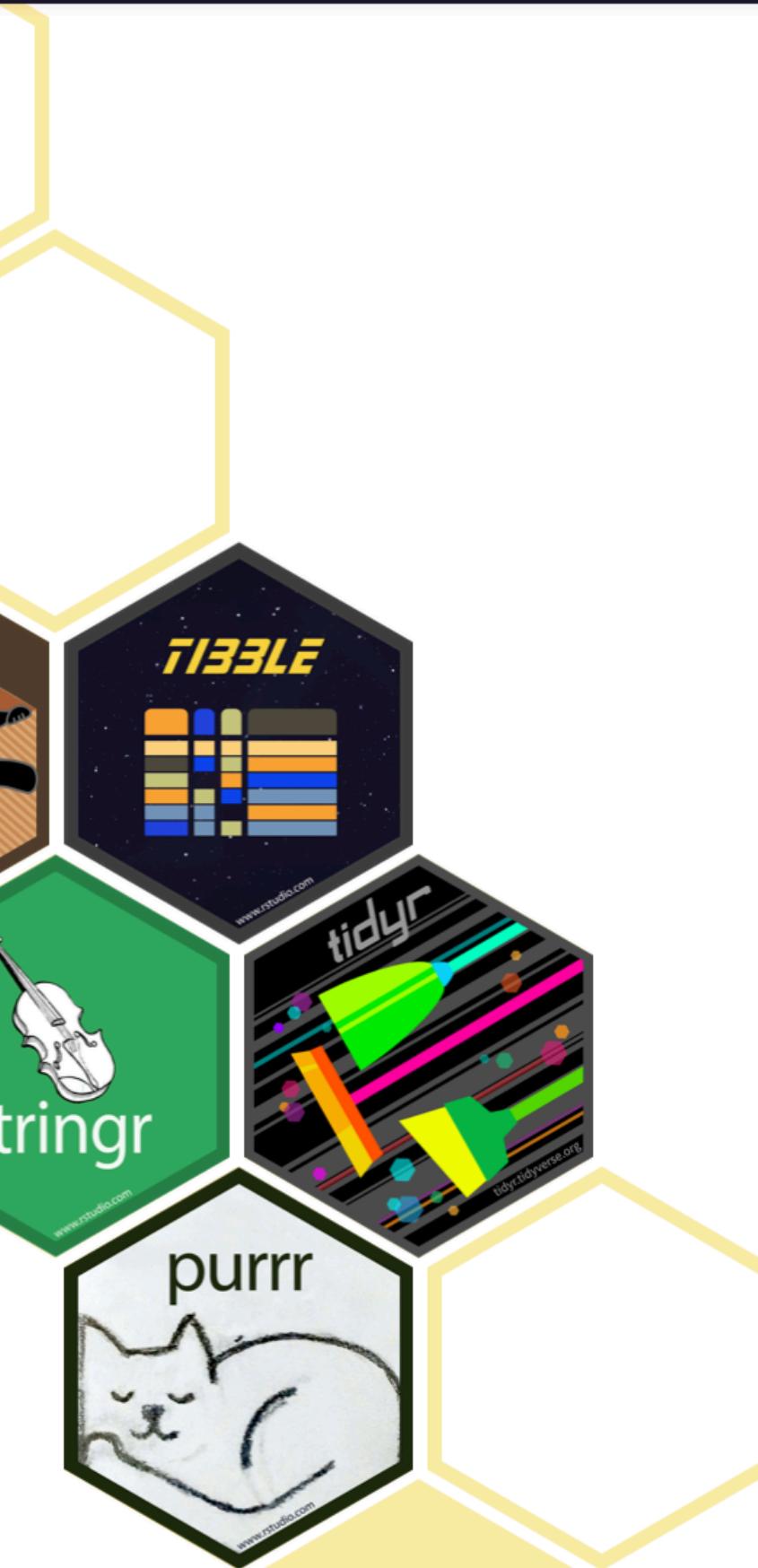
The screenshot shows the tidyverse.org website. The top navigation bar includes links for Packages, Blog, Learn, Help, and Contribute. Below the navigation bar, the word "Tidyverse" is prominently displayed. The main content area features a hexagonal grid of nine R package logos: dplyr, ggplot2, readr, purrr, forcats, stringr, and tidyverse. To the right of the grid, there is descriptive text about the tidyverse and a code block for installing it.

## 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

dplyr

tidyr

readr

purrr

tibble

hms

stringr

lubridate

forcats

DBI

haven

httr

jsonlite

readxl

rvest

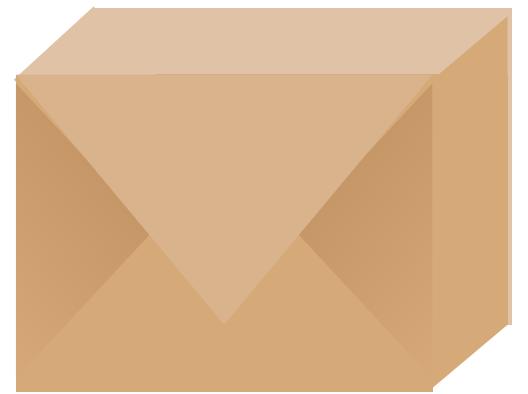
xml2

modelr

tidyverse

```
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.



# Quarto

An authoring format for Data Science.

The screenshot shows an RStudio interface with a Quarto document titled "authoring-with-quarto.qmd". The "Source" tab is selected, displaying the following code:

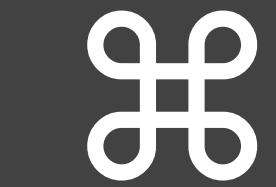
```
1 ---  
2 title: "Authoring with Quarto"  
3 format: html  
4 ---  
5  
6 Text written in **markdown**.  
7  
8 ```{r}  
9 # Code written in R  
10 (x <- 1:10)  
11 ...  
12  
13 Text written in *markdown*.  
14  
15
```

A gray callout points to the code in line 10 with the text "Code goes in a chunk". Another gray callout points to the "Run" button in the toolbar with the text "Click to run code in chunk". A dark gray callout points to the output window at the bottom with the text "Code result". The output window displays the R console output: [1] 1 2 3 4 5 6 7 8 9 10.

# Code chunks

Insert a chunk of R code with

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



+

Opt

+

A white lowercase letter 'i' inside a dark rounded square button.

(Mac)

Ctrl

+

Alt

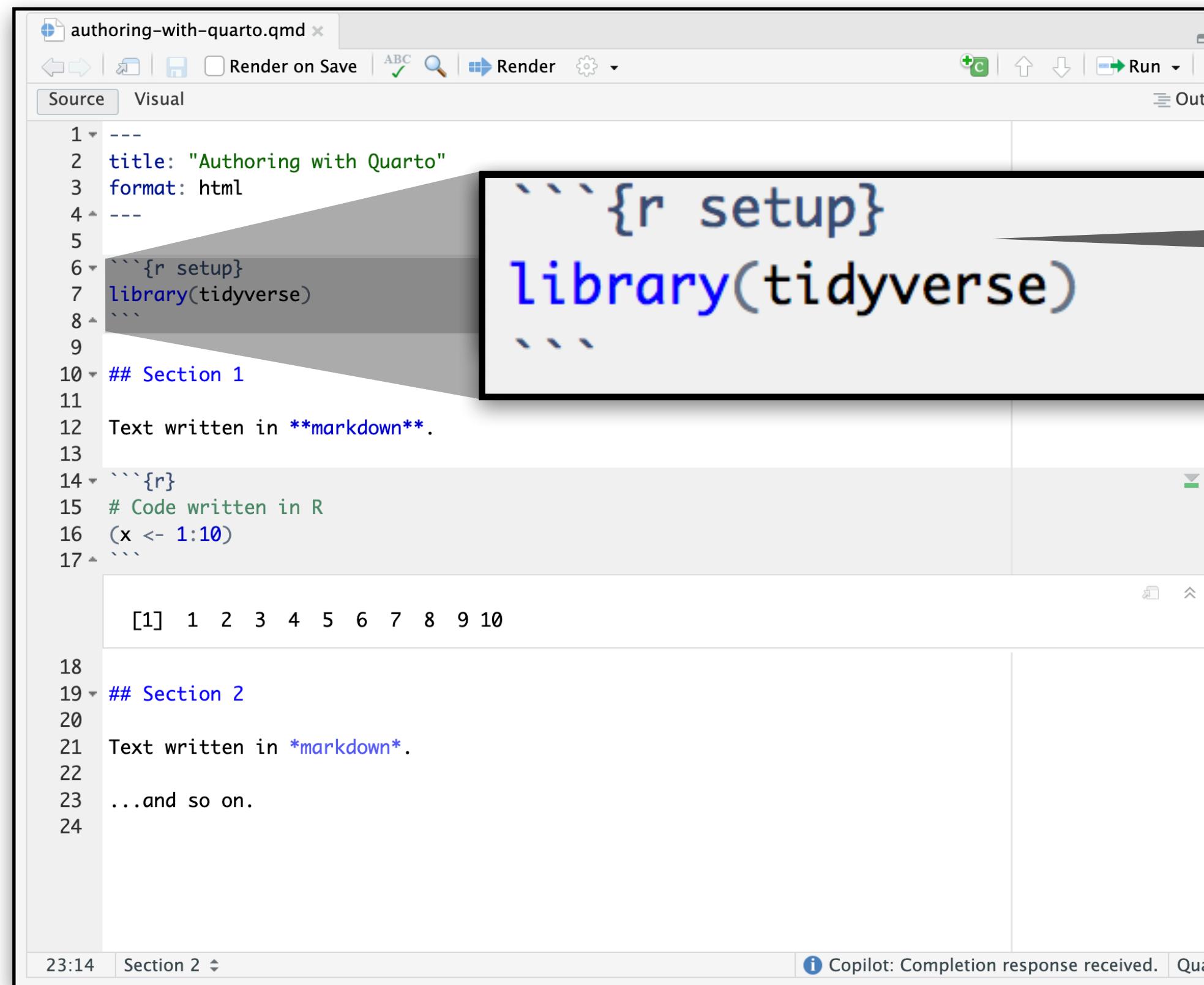
+

A white lowercase letter 'i' inside a dark rounded square button.

(PC)

# Setup

The setup chunk is always run once before anything else



A screenshot of the Quarto authoring interface. The top bar shows the file name "authoring-with-quarto.qmd". Below the bar, there are tabs for "Source" and "Visual". The "Source" tab is active, displaying the following code:

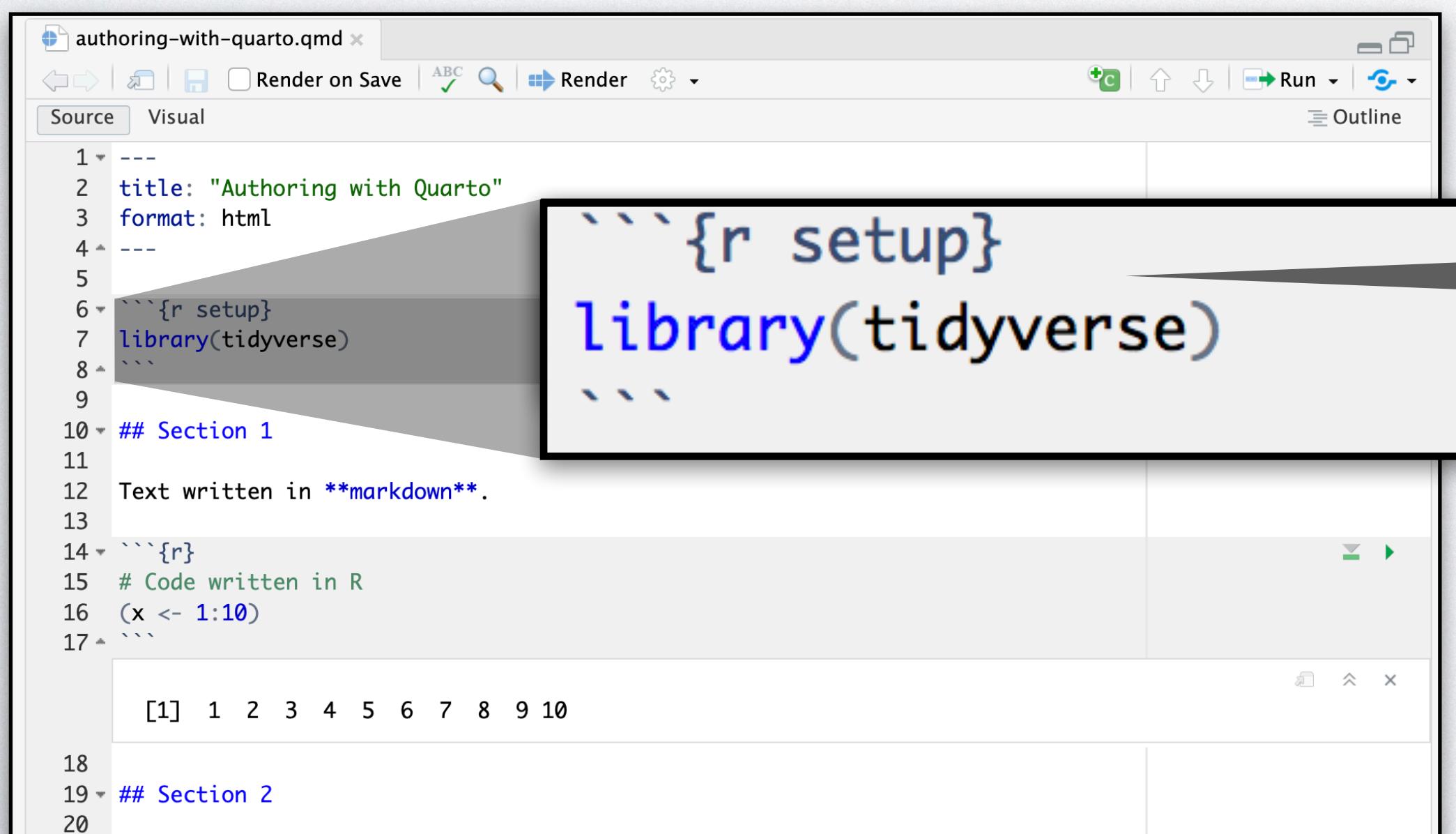
```
1 ---  
2 title: "Authoring with Quarto"  
3 format: html  
4 ---  
5  
6 ```{r setup}  
7 library(tidyverse)  
8 ...  
9  
10 ## Section 1  
11  
12 Text written in **markdown**.  
13  
14 ...  
15 # Code written in R  
16 (x <- 1:10)  
17 ...  
18  
19 ## Section 2  
20  
21 Text written in *markdown*.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

The code block containing the `library` command is highlighted with a gray box. A callout bubble points from this box to a text box containing the code snippet: ````{r setup}  
library(tidyverse)  
````. Another callout bubble points from the same code snippet to the text: "chunk labels are optional, the setup label is special".

At the bottom of the interface, the status bar shows "23:14 Section 2" and "Copilot: Completion response received. Quarto".

# 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.



The screenshot shows a Quarto editor window titled "authoring-with-quarto.qmd". The "Source" tab is selected, displaying the following R code:

```
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
```

A callout box highlights the `library(tidyverse)` line, which is enclosed in a ````{r setup}` block. The line is preceded by a comment `## Section 1`.

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.

