

# Welcome to the Tidyverse



# Your Turn

Form groups of 2-4 people. Introduce yourself to your group members. Tell them:

1. Who you are
2. What you do with data
3. How long you have been using R



**HELLO**  
my name is

**Garrett**



@StatGarrett

O'REILLY®

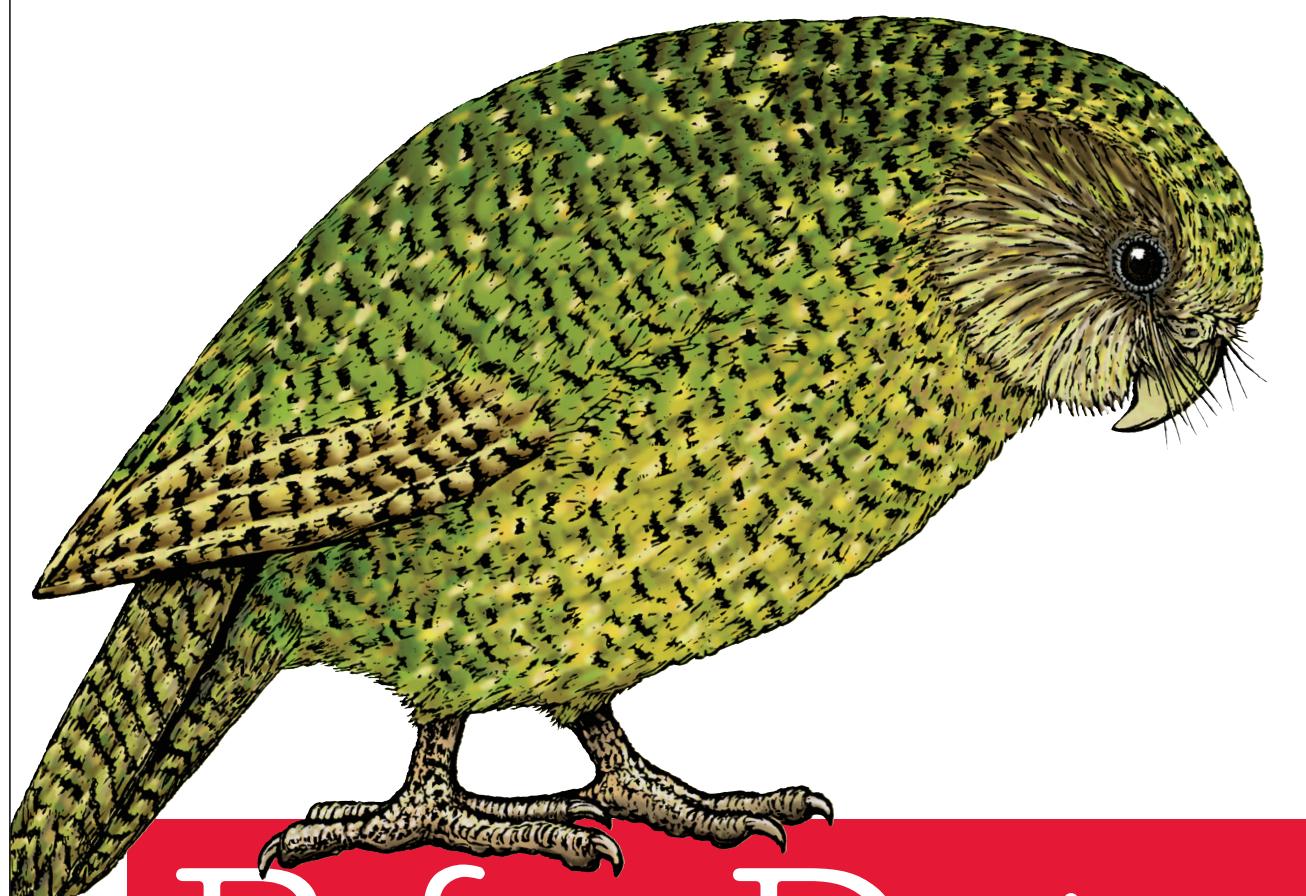


# Hands-On Programming with R

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

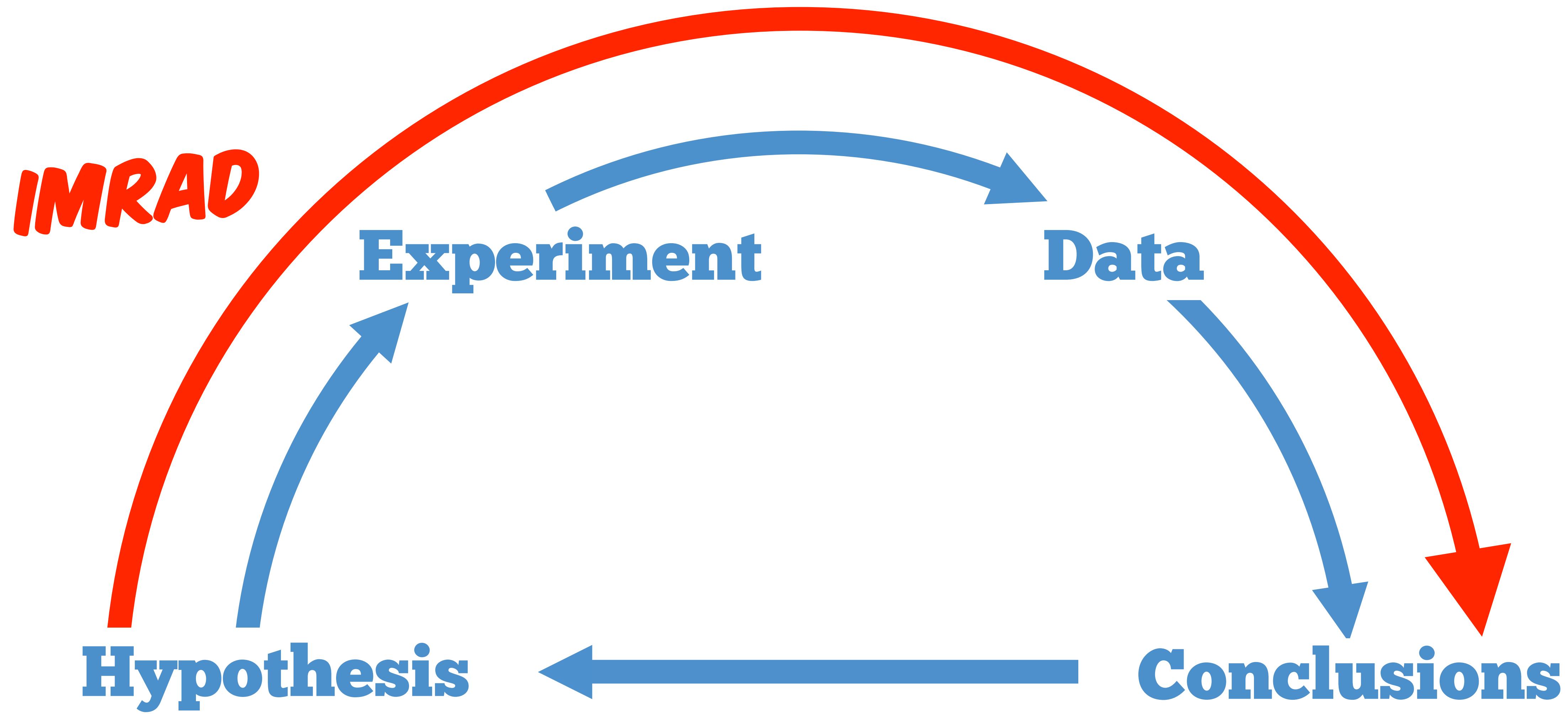
What answer was discovered?

## And Discussion

What does the answer imply?



# (Experimental) Science



# Pop Quiz

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

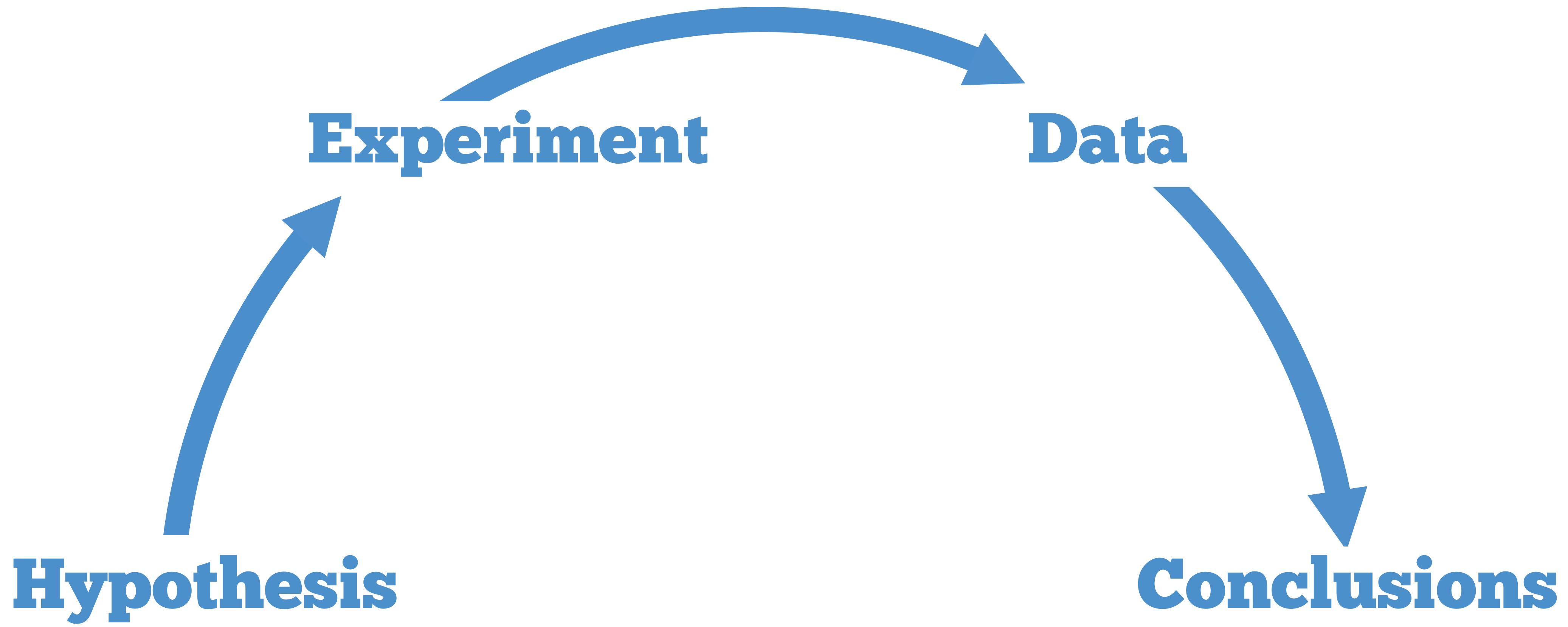


# Pop Quiz

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

**Science with non-trivial data?**

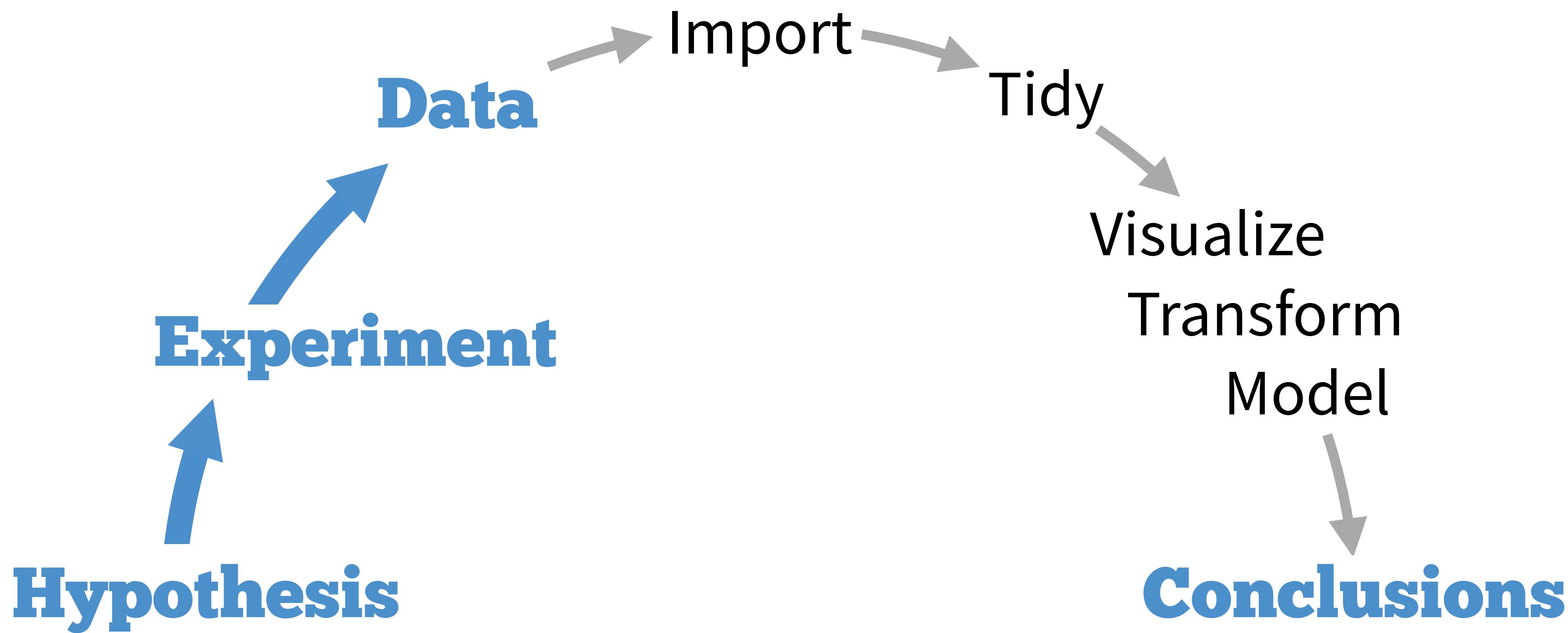
# Data Science



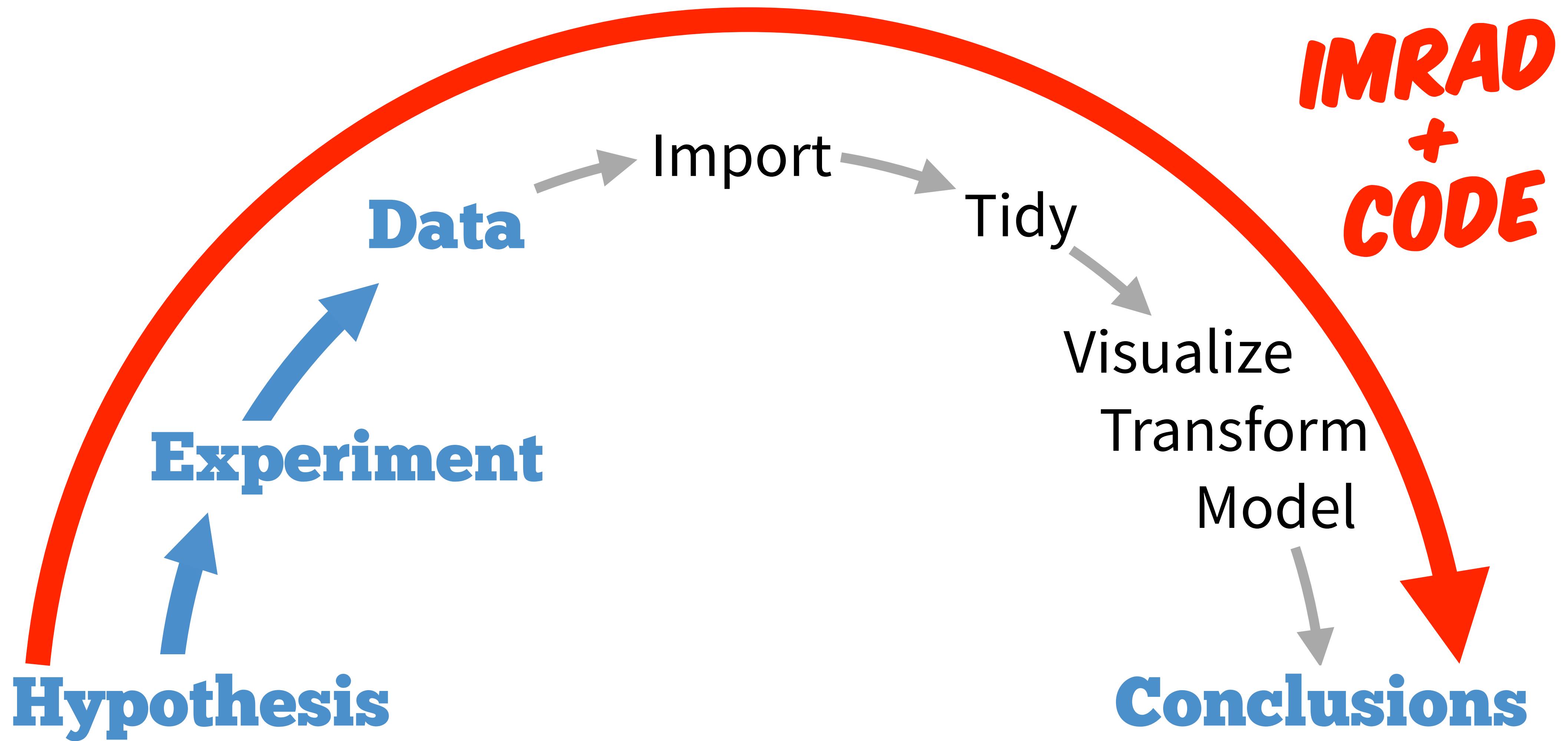
# Data Science



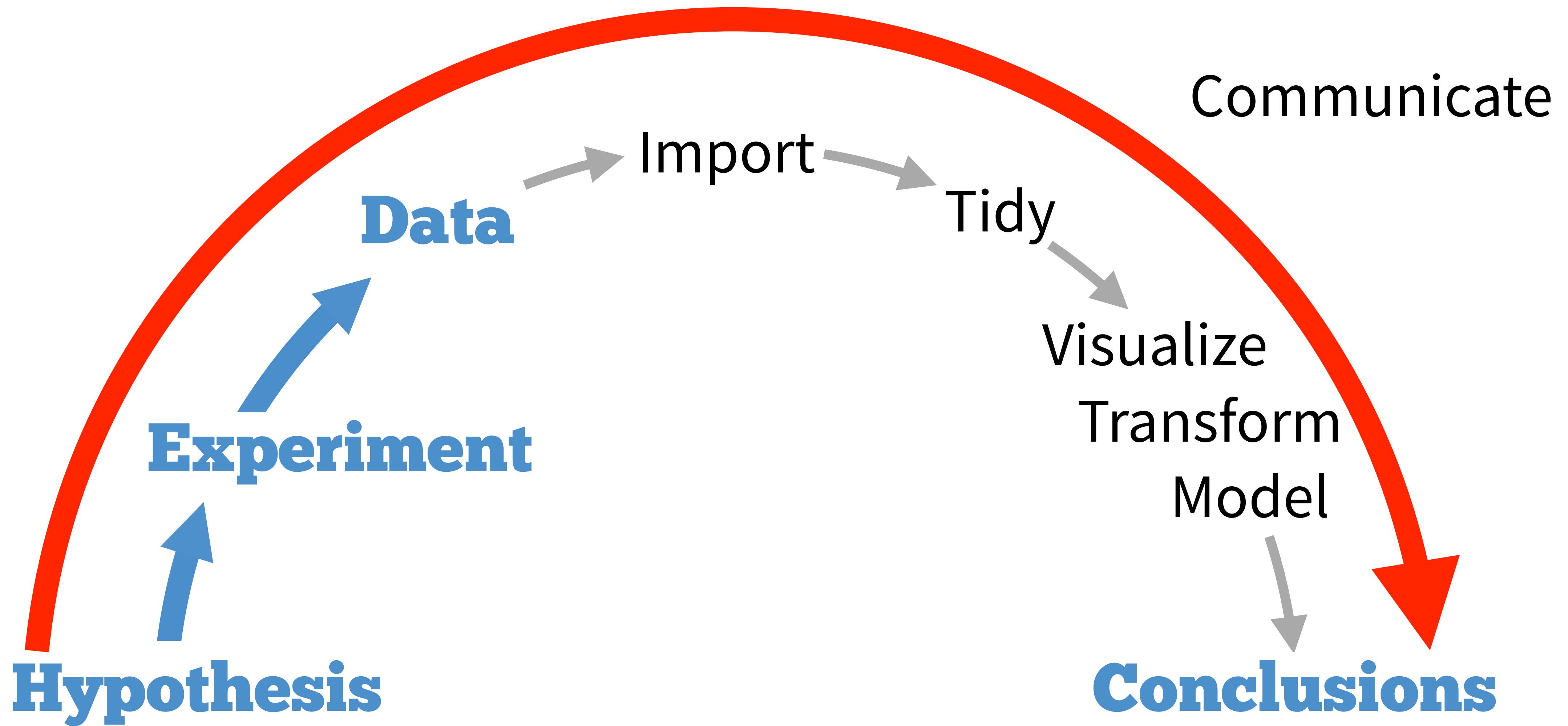
# Data Science



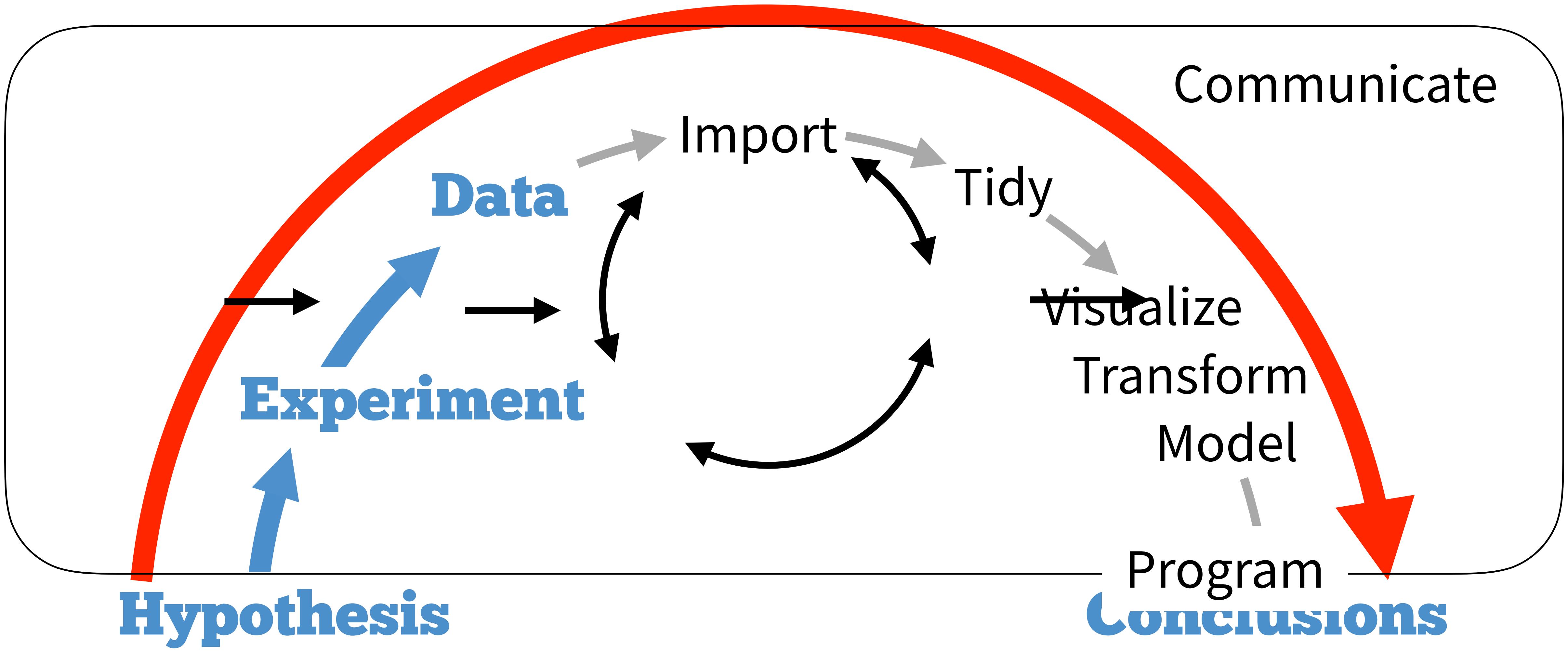
# Data Science



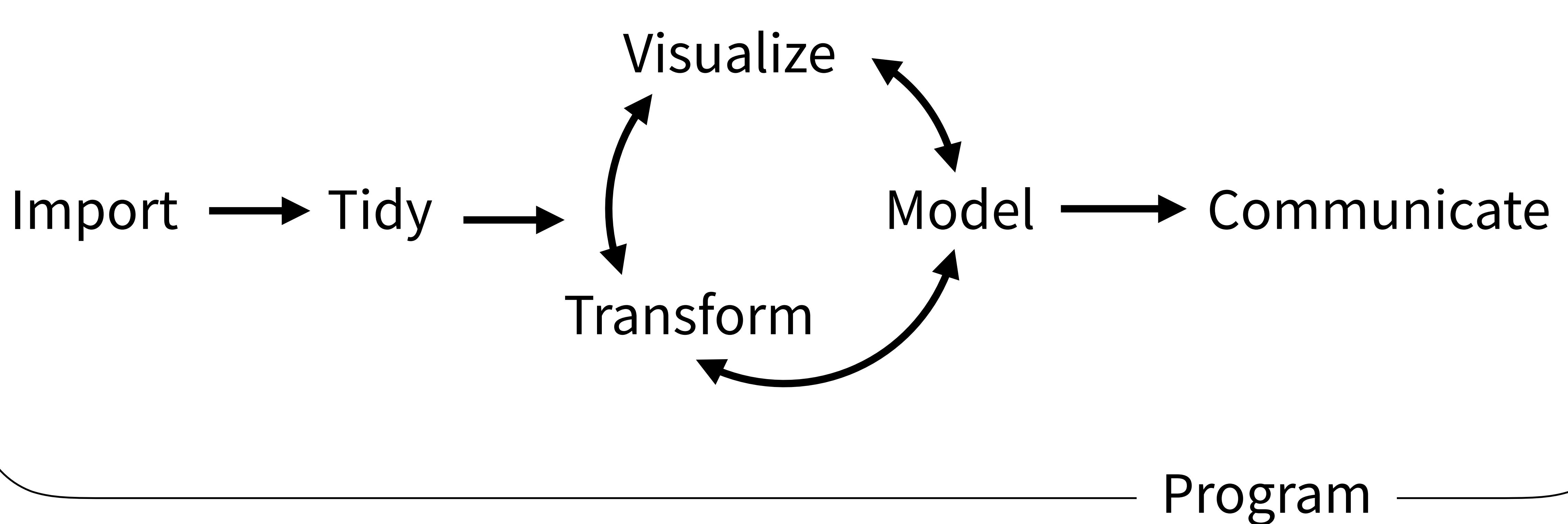
# Data Science



# Data Science



# (Applied) Data Science





# Outline

Introduction and  
Visualize Data

9:00 - 10:30

Morning Break

10:30 - 10:50

Transform Data

10:50 - 12:20

Lunch

12:20 - 1:50

Model Data

1:50 - 3:20

Afternoon Break

3:20 - 3:40

Reproducible  
Reporting

3:40 - 5:10

# R and RStudio



# Your Turn

Go here and log in for the class materials

<https://rstudio.cloud/project/491221>

↑  
**s**

↑  
**NO S**



# Demo

<https://rstudio.cloud/project/491221>



A language



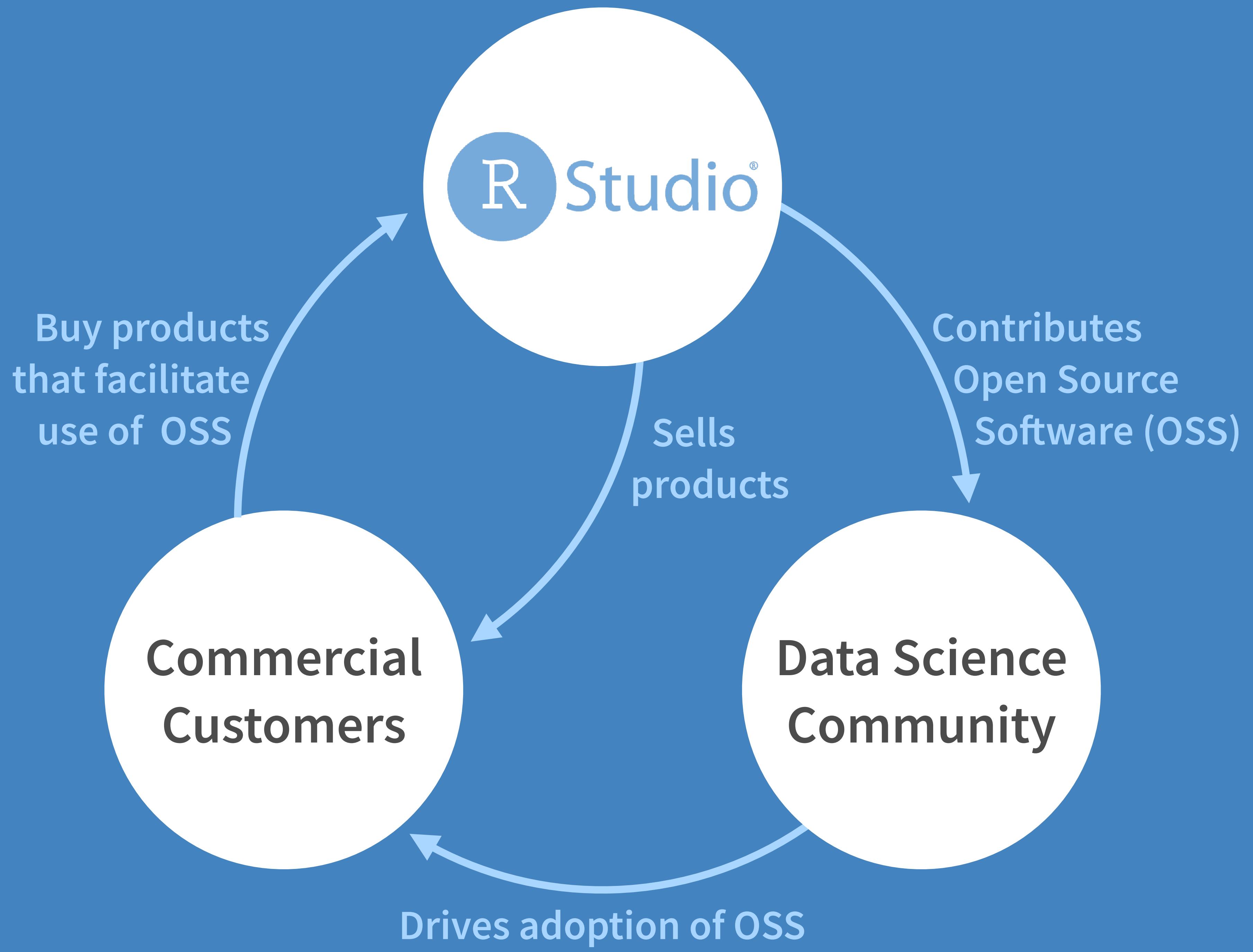
<https://rstudio.cloud/project/491221>

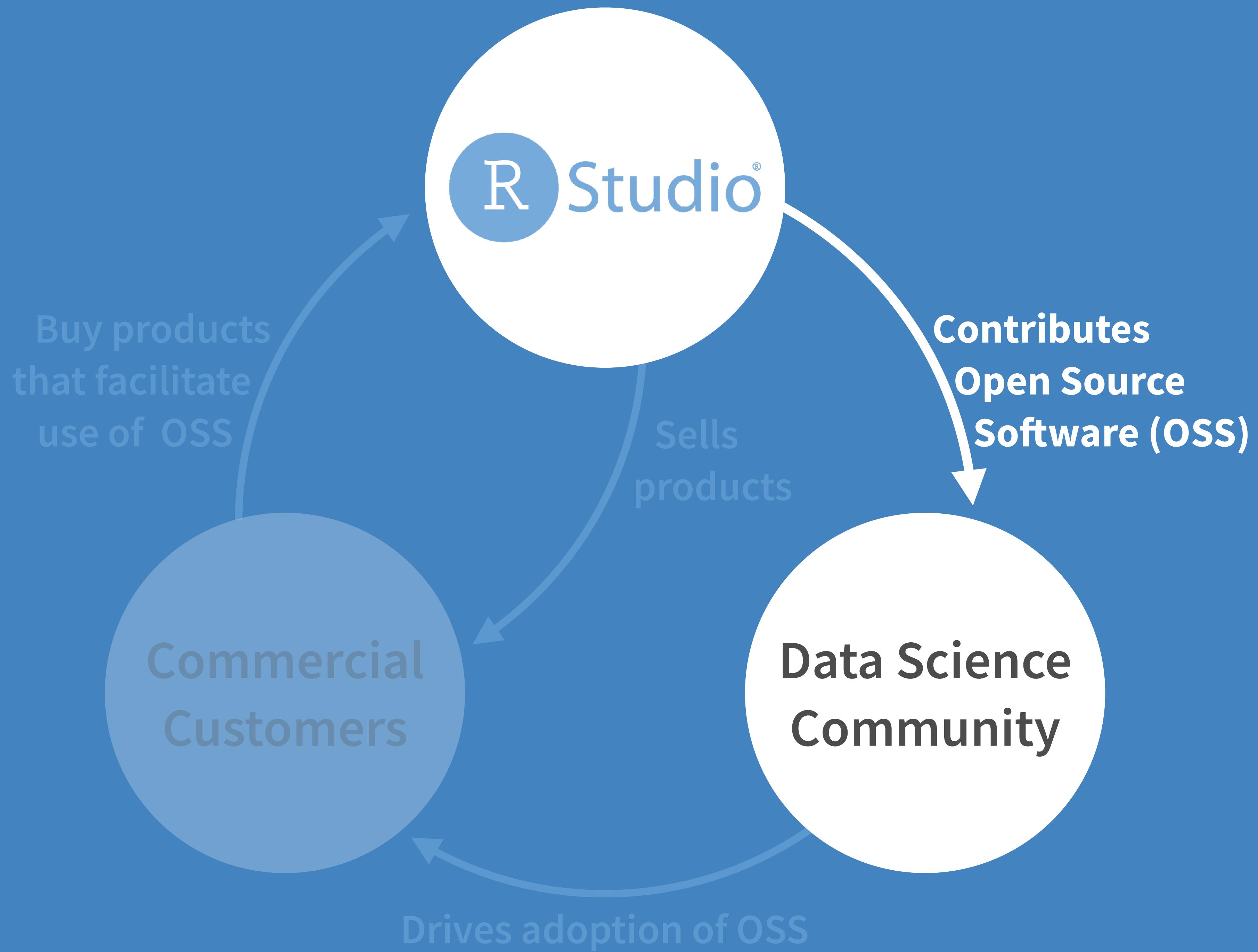


A language



Writing software





# 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 Markdown

A large, semi-transparent watermark of the R logo is positioned in the bottom right corner. The logo consists of a dark green circle containing a white letter 'R' with a diagonal line through it.

R

# Your Turn

Go to **01-Introduction.**

Open **01-RMarkdown-Exercises.Rmd**. Read through the file and do everything it tells you to do.



# R Markdown

An authoring format for Data Science.

The screenshot shows the RStudio interface with an R Markdown file open. The code editor pane contains the following R Markdown 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
```

The R console pane below shows the output of the R code:

```
[1] -1.2  1.0 -0.5  0.9 -0.6 -1.1 -1.5
```

Three callout boxes point to specific parts of the interface:

- A grey box points to the code editor area: **Code goes in a chunk**
- A grey box points to the run button in the toolbar: **Click to run code in chunk**
- A dark grey box points to the R console output: **Code result**

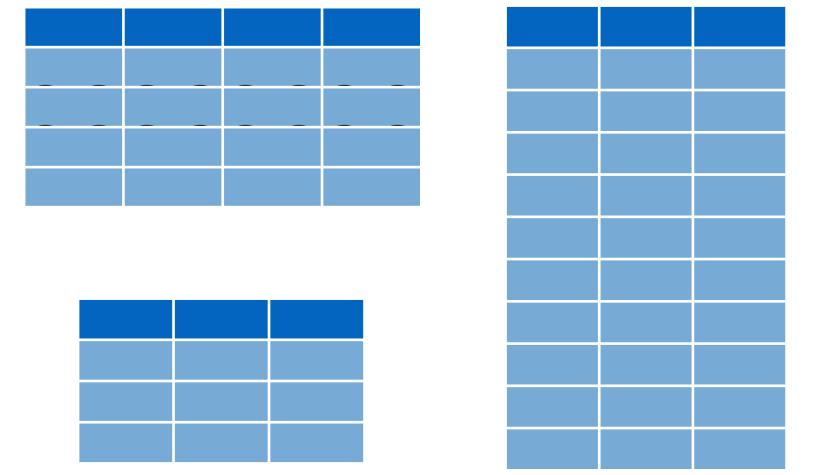
# R Packages



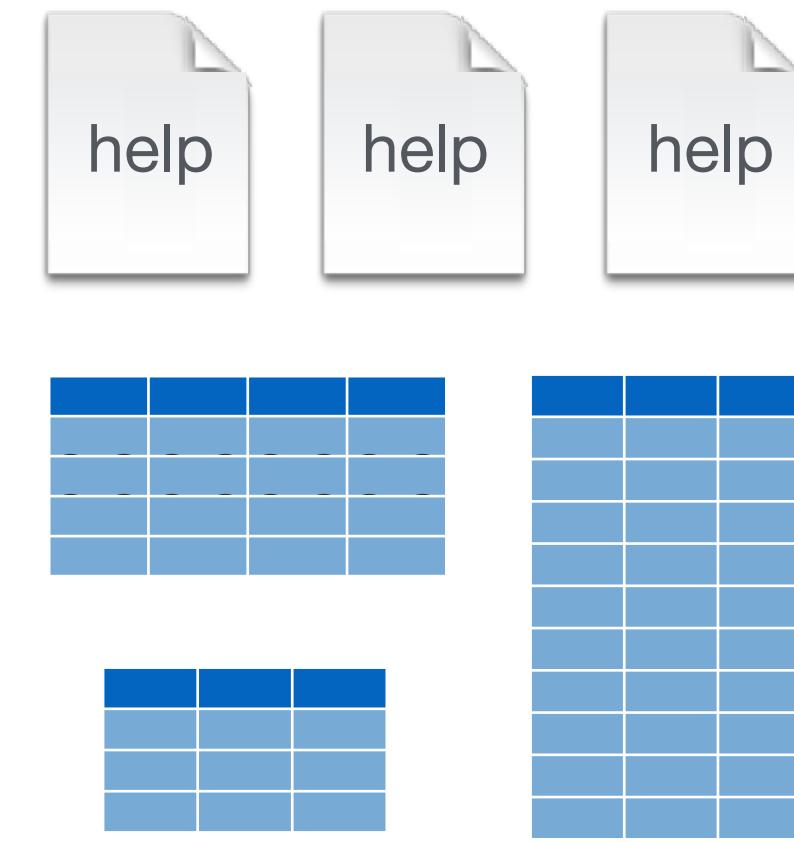
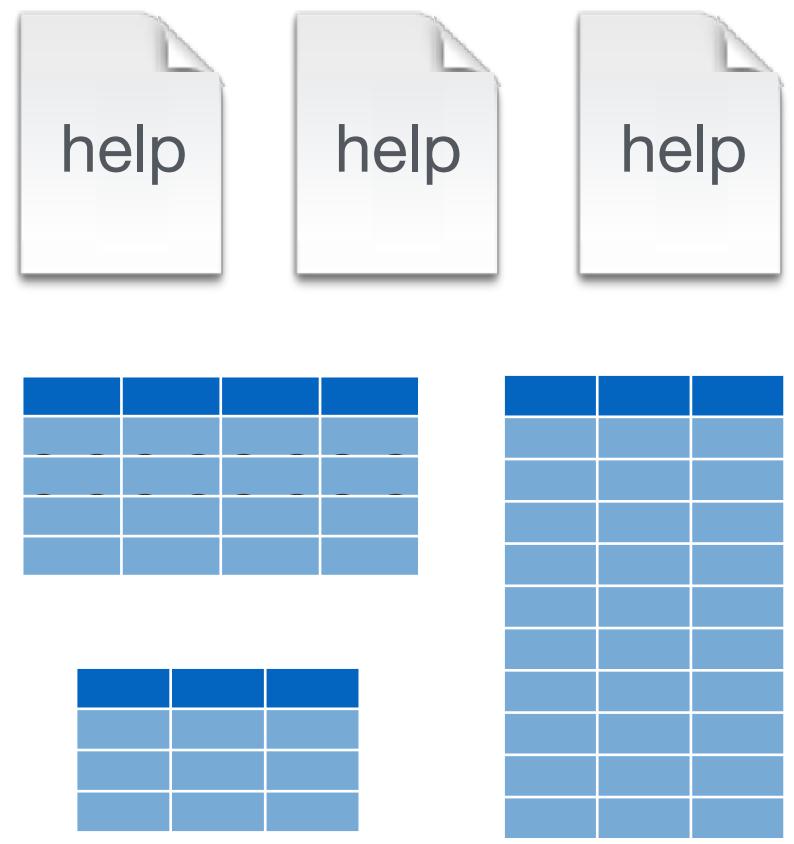
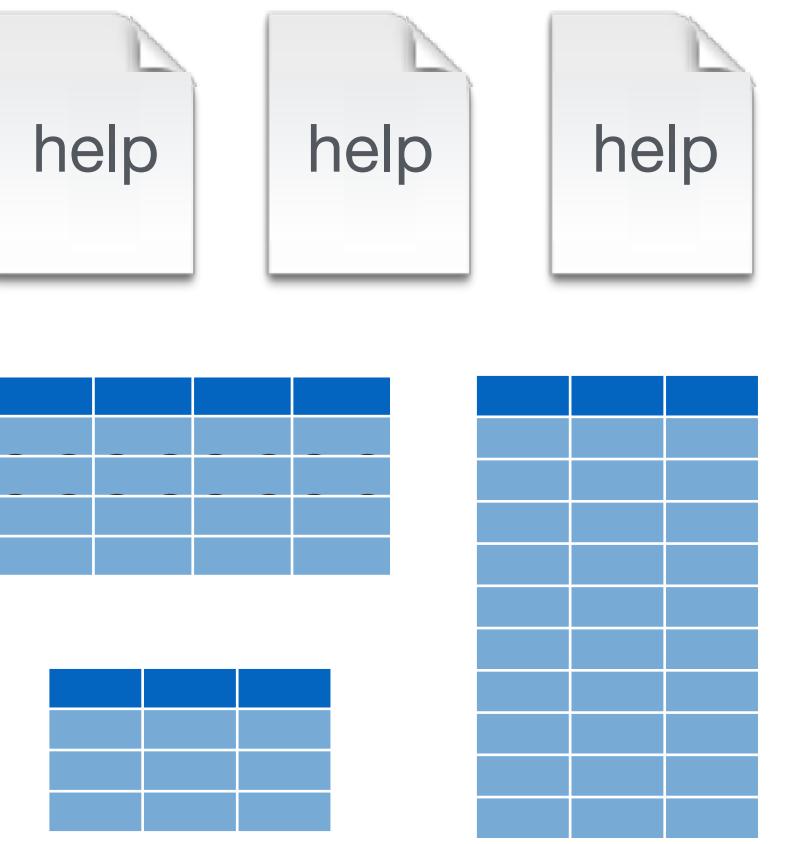
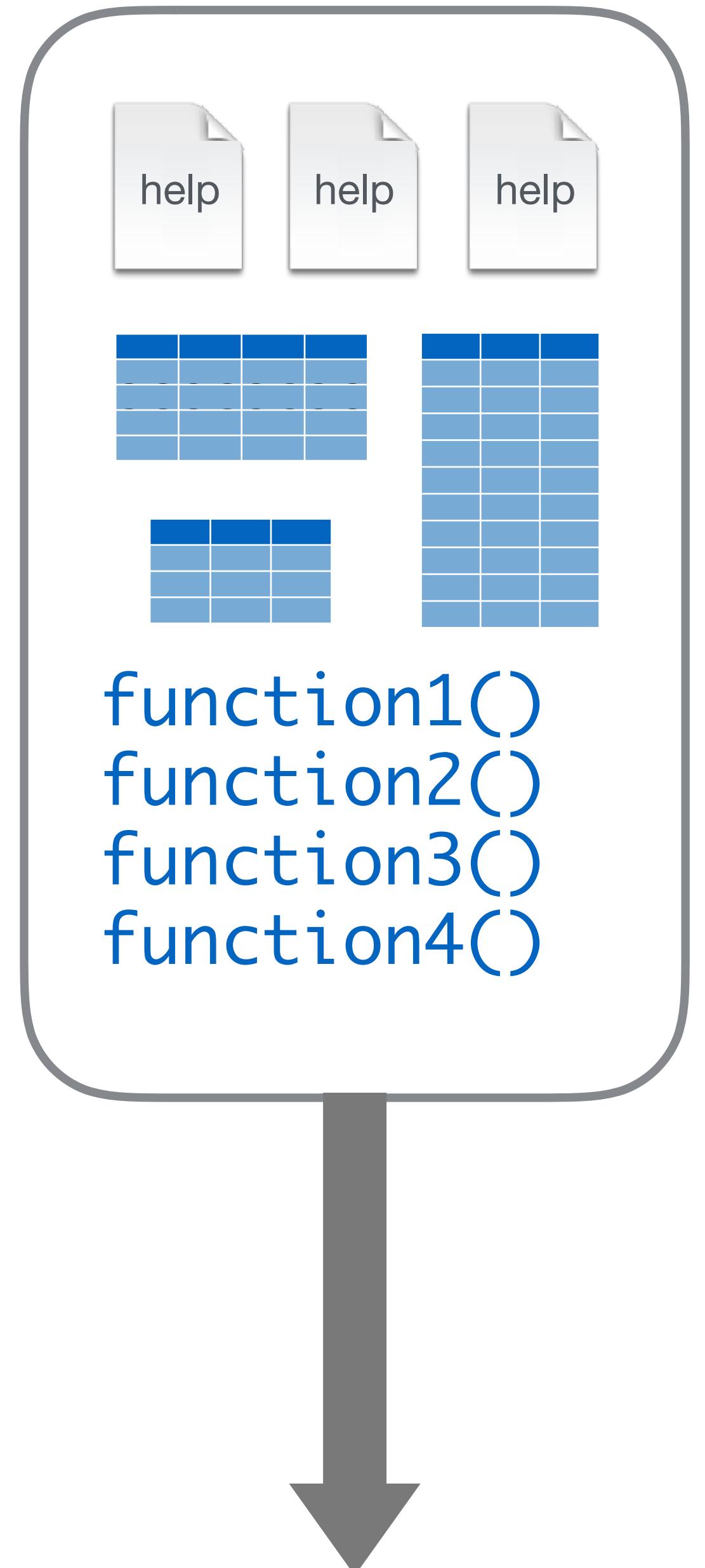
# Your Turn

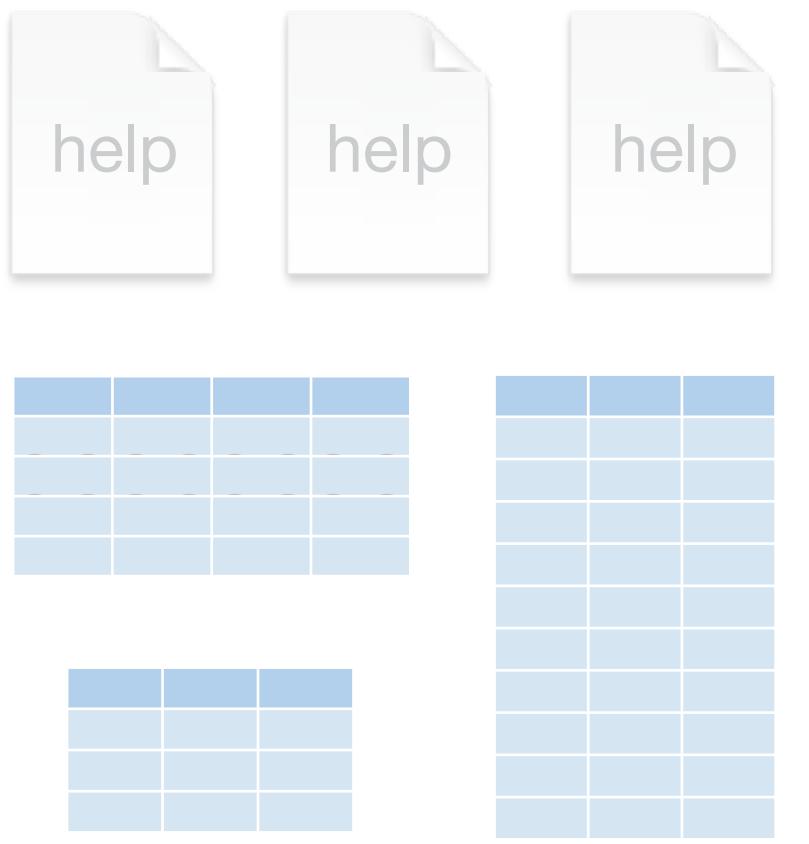
(If you know) explain to your group what an R package is.



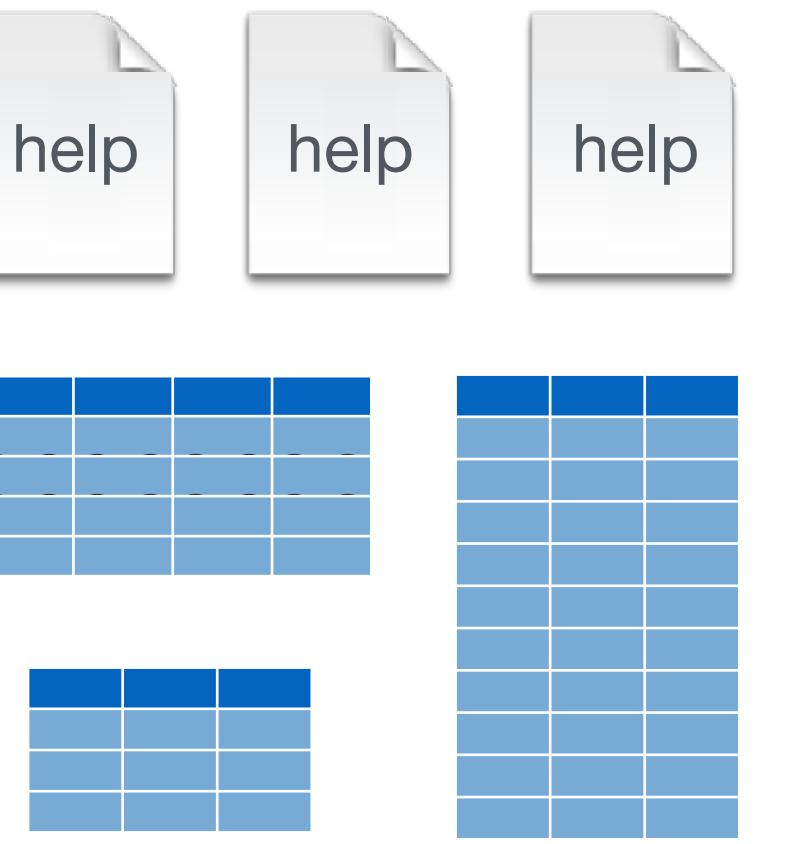


function1()  
function2()  
function3()  
function4()

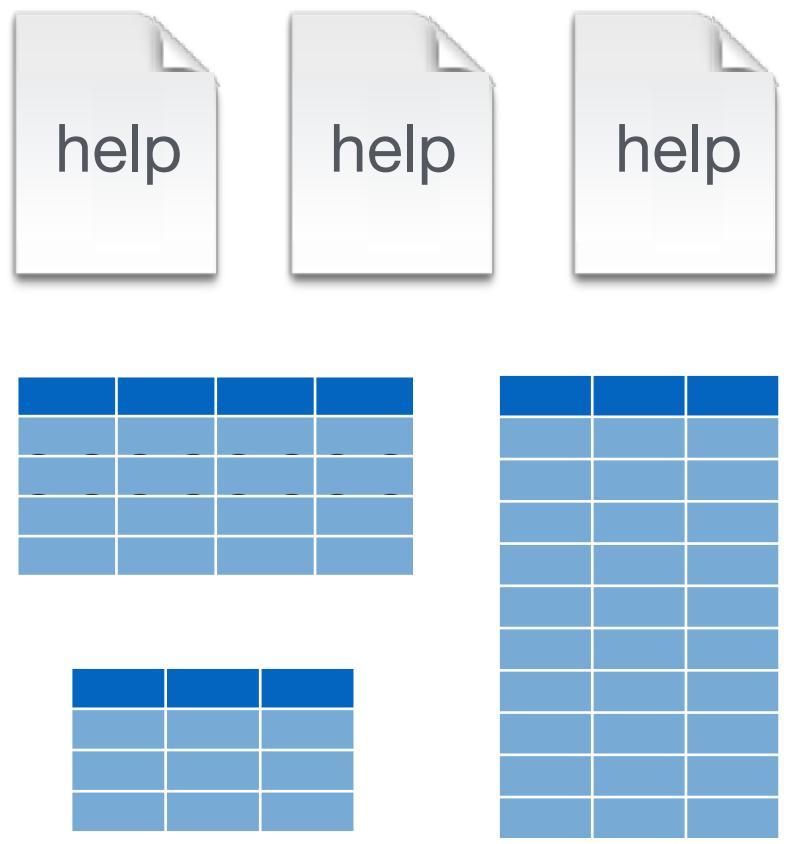




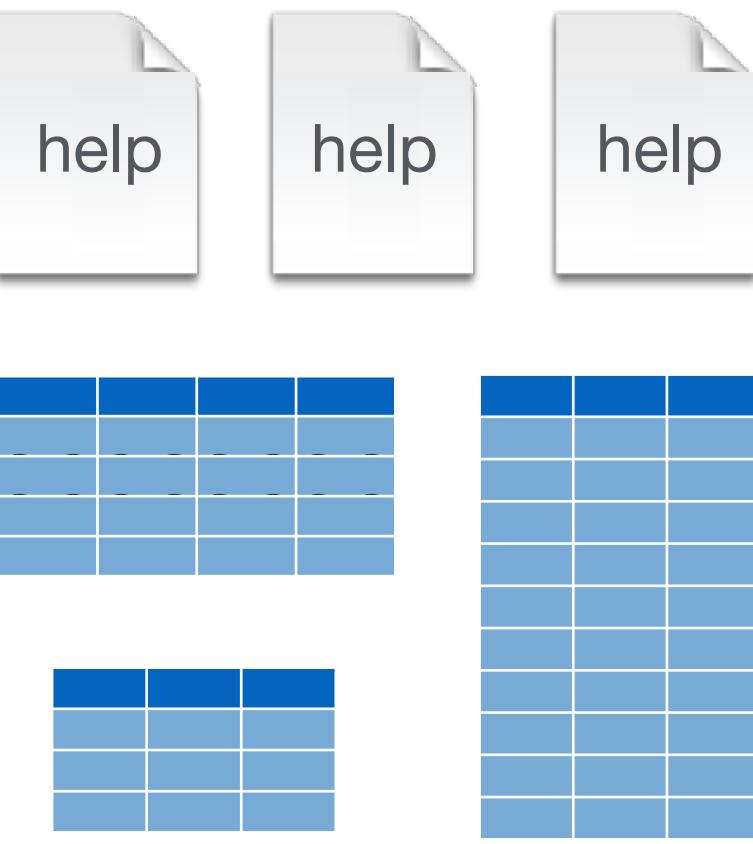
function1()  
function2()  
function3()  
function4()



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

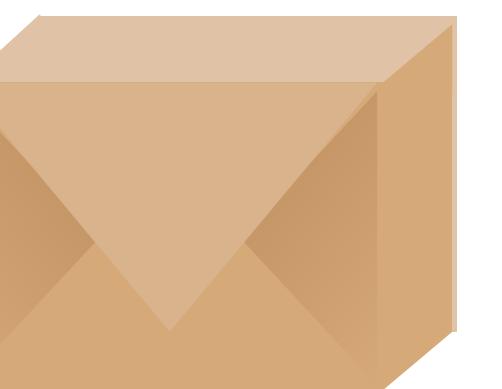


function9()  
functionA()  
functionB()  
functionC()



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

Base R



42

R Packages



The Comprehensive R Archive x Garrett

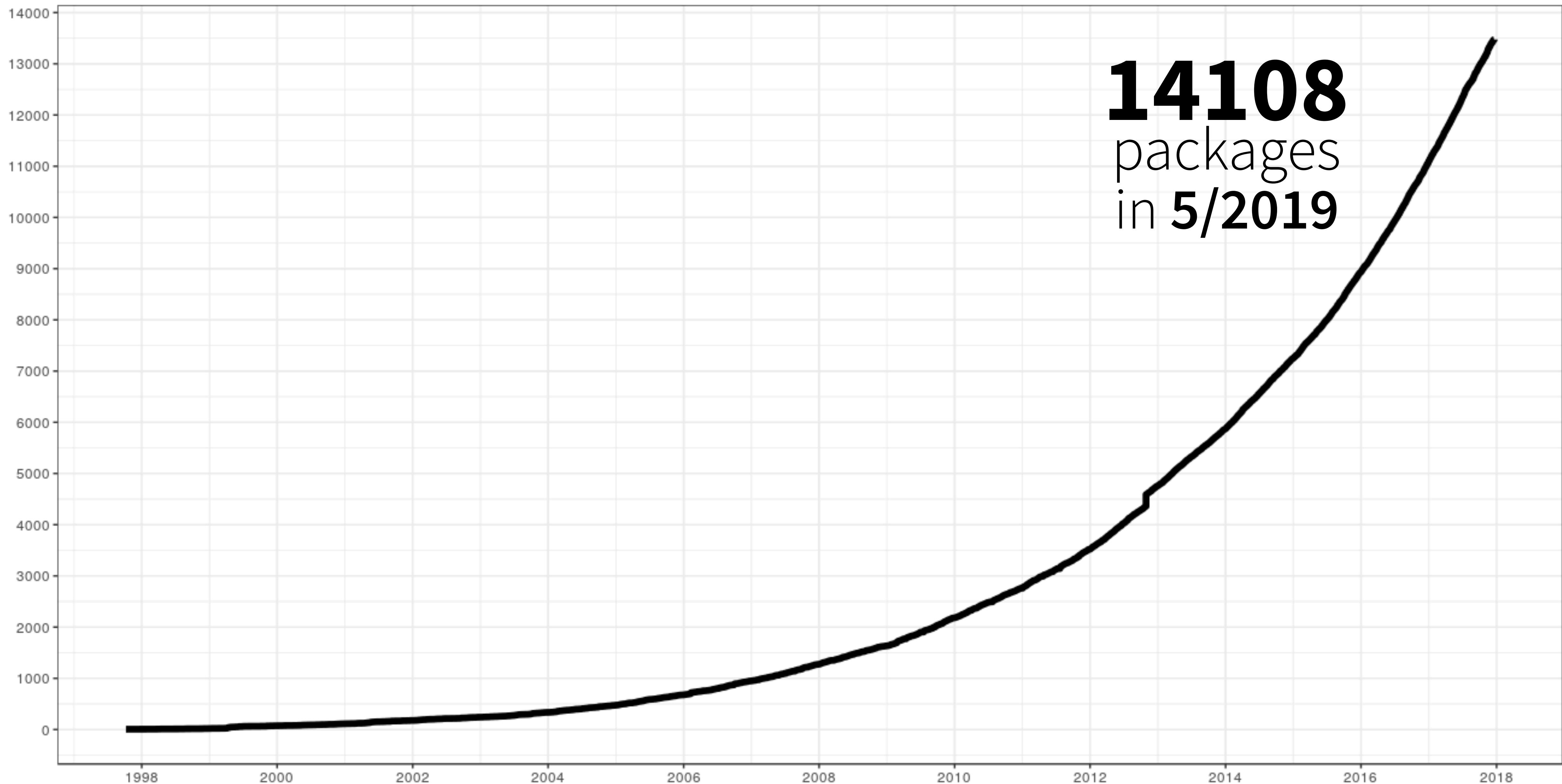
Secure | https://cran.r-project.org

# 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

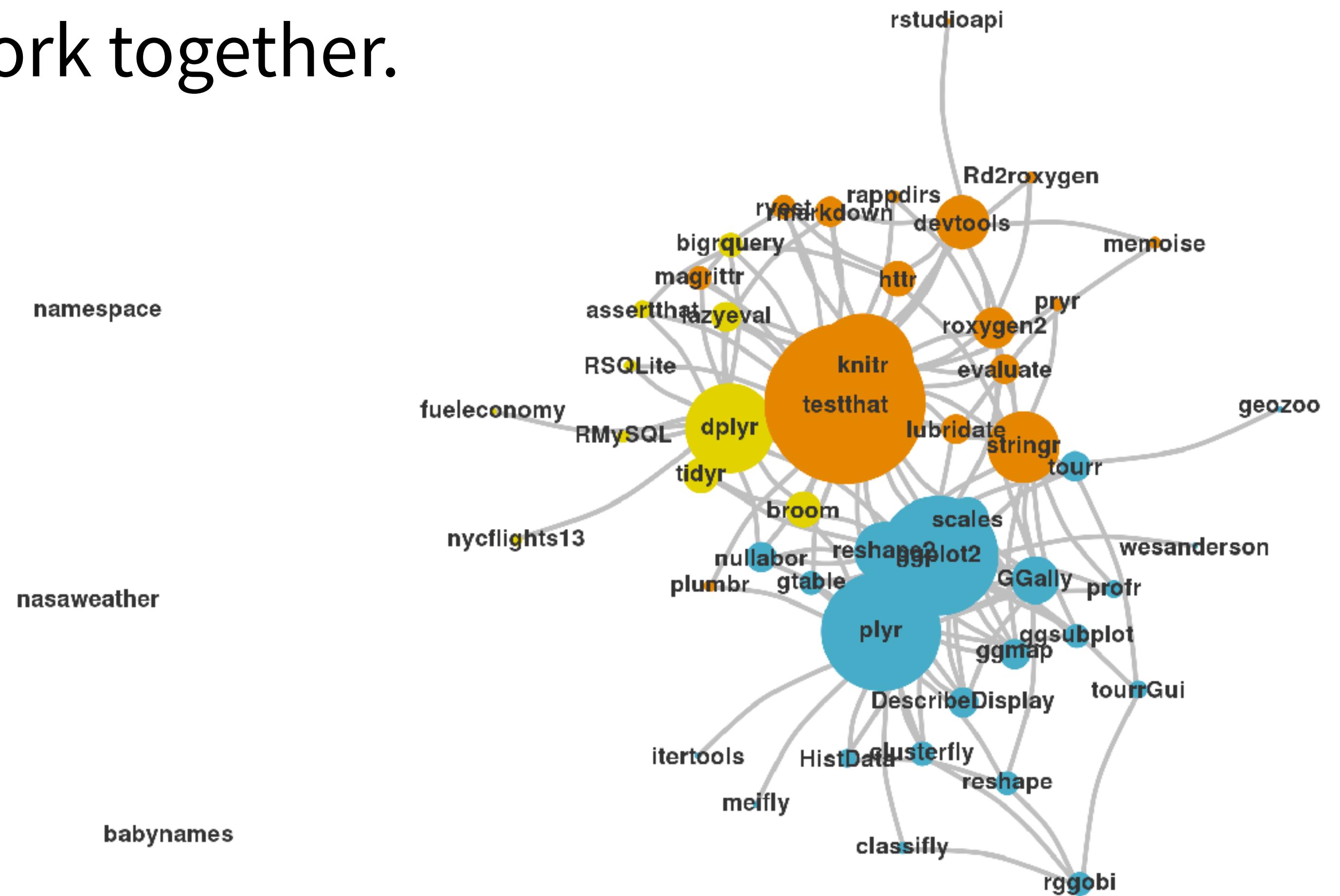
<a href="#">A3</a>	Accurate, Adaptable, and Accessible Error Metrics for Predictive Models
<a href="#">abyyR</a>	Access to Abbyy Optical Character Recognition (OCR) API
<a href="#">abc</a>	Tools for Approximate Bayesian Computation (ABC)
<a href="#">ABCAnalysis</a>	Computed ABC Analysis
<a href="#">abc.data</a>	Data Only: Tools for Approximate Bayesian Computation (ABC)
<a href="#">abcdeFBA</a>	ABCDE_FBA: A-Biologist-Can-Do-Everything of Flux Balance Analysis with this package
<a href="#">ABCOptim</a>	Implementation of Artificial Bee Colony (ABC) Optimization
<a href="#">ABCp2</a>	Approximate Bayesian Computational Model for Estimating P2
<a href="#">ABC.RAP</a>	Array Based CpG Region Analysis Pipeline
<a href="#">abcrf</a>	Approximate Bayesian Computation via Random Forests
<a href="#">abctools</a>	Tools for ABC Analyses
<a href="#">abd</a>	The Analysis of Biological Data
<a href="#">abf2</a>	Load Gap-Free Axon ABF2 Files
<a href="#">ABHgenotypeR</a>	Easy Visualization of ABH Genotypes
<a href="#">abind</a>	Combine Multidimensional Arrays
<a href="#">abjutils</a>	Useful Tools for Jurimetric Analysis Used by the Brazilian Jurimetrics Association
<a href="#">abn</a>	Modelling Multivariate Data with Additive Bayesian Networks
<a href="#">abodOutlier</a>	Angle-Based Outlier Detection

Number of R packages ever published on CRAN



# The Tidyverse

A collection of modern R packages that share common philosophies, embed best practices, and are designed to work together.



# Using packages

**1**

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

Downloads files to computer

**1 x per computer**

# Pop Quiz!

The tidyverse contains the following packages.  
How would you install them?

ggplot2

dplyr

tidyr

readr

purrr

tibble

hms

stringr

lubridate

forcats

DBI

haven

httr

jsonlite

readxl

rvest

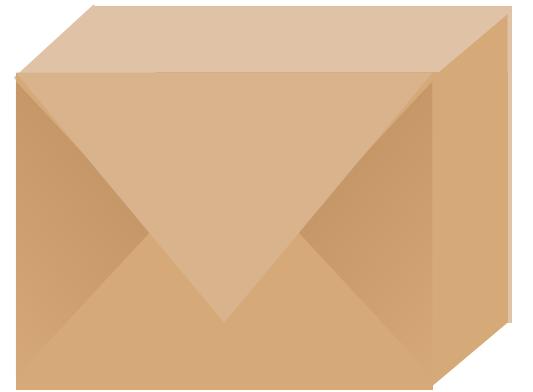
xml2

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")
```

```
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")
```

I've already installed all of  
the packages you need

# 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")
```

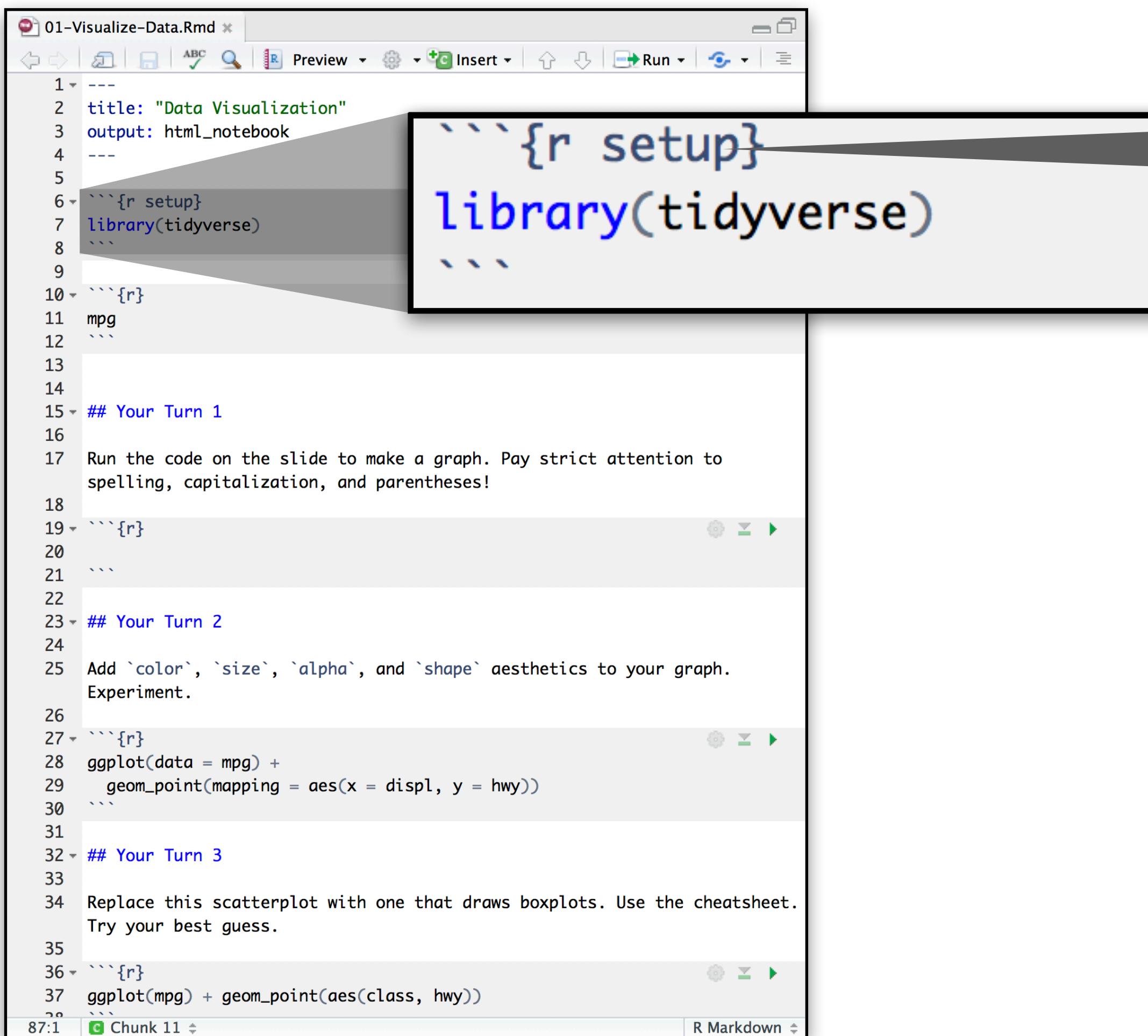
```
library("tidyverse")
```

does the equivalent of

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

# Setup

The setup chunk is always run once before anything else

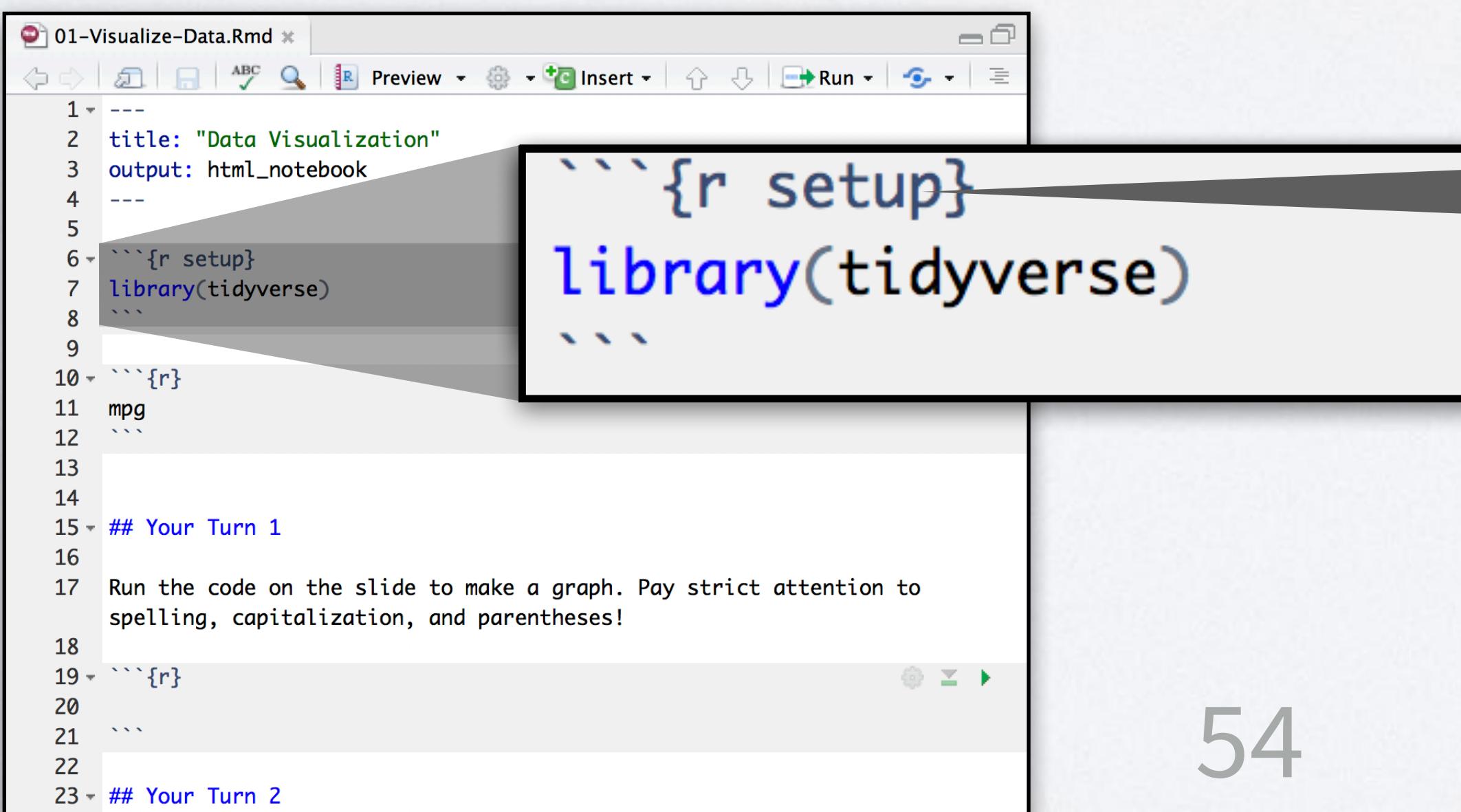


```
1 ---  
2 title: "Data Visualization"  
3 output: html_notebook  
4 ---  
5  
6 ```{r setup}  
7 library(tidyverse)  
8 ```  
9  
10 ```{r}  
11 mpg  
12  
13  
14  
15 ## Your Turn 1  
16  
17 Run the code on the slide to make a graph. Pay strict attention to  
spelling, capitalization, and parentheses!  
18  
19 ```{r}  
20  
21  
22  
23 ## Your Turn 2  
24  
25 Add `color`, `size`, `alpha`, and `shape` aesthetics to your graph.  
Experiment.  
26  
27 ```{r}  
28 ggplot(data = mpg) +  
29   geom_point(mapping = aes(x = displ, y = hwy))  
30  
31  
32 ## Your Turn 3  
33  
34 Replace this scatterplot with one that draws boxplots. Use the cheatsheet.  
Try your best guess.  
35  
36 ```{r}  
37 ggplot(mpg) + geom_point(aes(class, hwy))  
38  
87:1 | Chunk 11 | R Markdown
```

chunk labels are optional,  
the setup label is special

# Your Turn

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



```
1 ---  
2 title: "Data Visualization"  
3 output: html_notebook  
4 ---  
5  
6 ```{r setup}  
7 library(tidyverse)  
8 ```  
9  
10 ```{r}  
11 mpg  
12 ```  
13  
14  
15 ## Your Turn 1  
16  
17 Run the code on the slide to make a graph. Pay strict attention to  
spelling, capitalization, and parentheses!  
18  
19 ```{r}  
20 ```  
21  
22  
23 ## Your Turn 2
```

chunk labels are optional,  
the setup label is special





# tidyverse.org

The screenshot shows the homepage of tidyverse.org. At the top, there's a navigation bar with links for Packages, Articles, Learn, Help, and Contribute. Below the navigation, there's a large graphic featuring hexagonal icons for various R packages: dplyr (orange, with a pliers icon), ggplot2 (grey, with a line plot icon), readr (blue, with a file icon), purrr (white with a cat icon), tibble (black, with a grid icon), and tidyr (orange, with a circular arrow icon). To the right of the graphic, the text reads: "R packages for data science. The tidyverse is an opinionated collection of R packages designed for data science. All packages share an underlying philosophy and common APIs." Below this, there's a section titled "Install the complete tidyverse with:" containing the code: `install.packages("tidyverse")`.

Tidyverse

Packages Articles Learn Help Contribute

dplyr

ggplot2

readr

tibble

purrr

tidyr

R packages for data science

The tidyverse is an opinionated collection of R packages designed for data science. All packages share an underlying philosophy and common APIs.

Install the complete tidyverse with:

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

# Your Turn 0

Navigate up to the **02-Visualize** folder.

**Open 02-Visualize-Exercises.Rmd**

Add a setup chunk that loads the  
tidyverse packages.

