

Master the Tidyverse



Garrett Grolemond

Data Scientist, Educator

January 2017

RStudio

HELLO

my name is

Garrett

 @StatGarrett

O'REILLY®



R for Data Science

VISUALIZE, MODEL, TRANSFORM, TIDY, AND IMPORT DATA

Hadley Wickham &
Garrett Grolemund

HELLO

my name is

Sean

 @lopp_sean

Day 1

**Introduction and
Data Visualization**

9:00 - 10:30

Morning Break

10:30 - 10:45

Data Visualization

10:45 - 12:00

Lunch

12:00 - 1:00

Transforming Data

1:00 - 3:15

Afternoon Break

3:15 - 3:30

Tidy Data

3:30 - 5:00

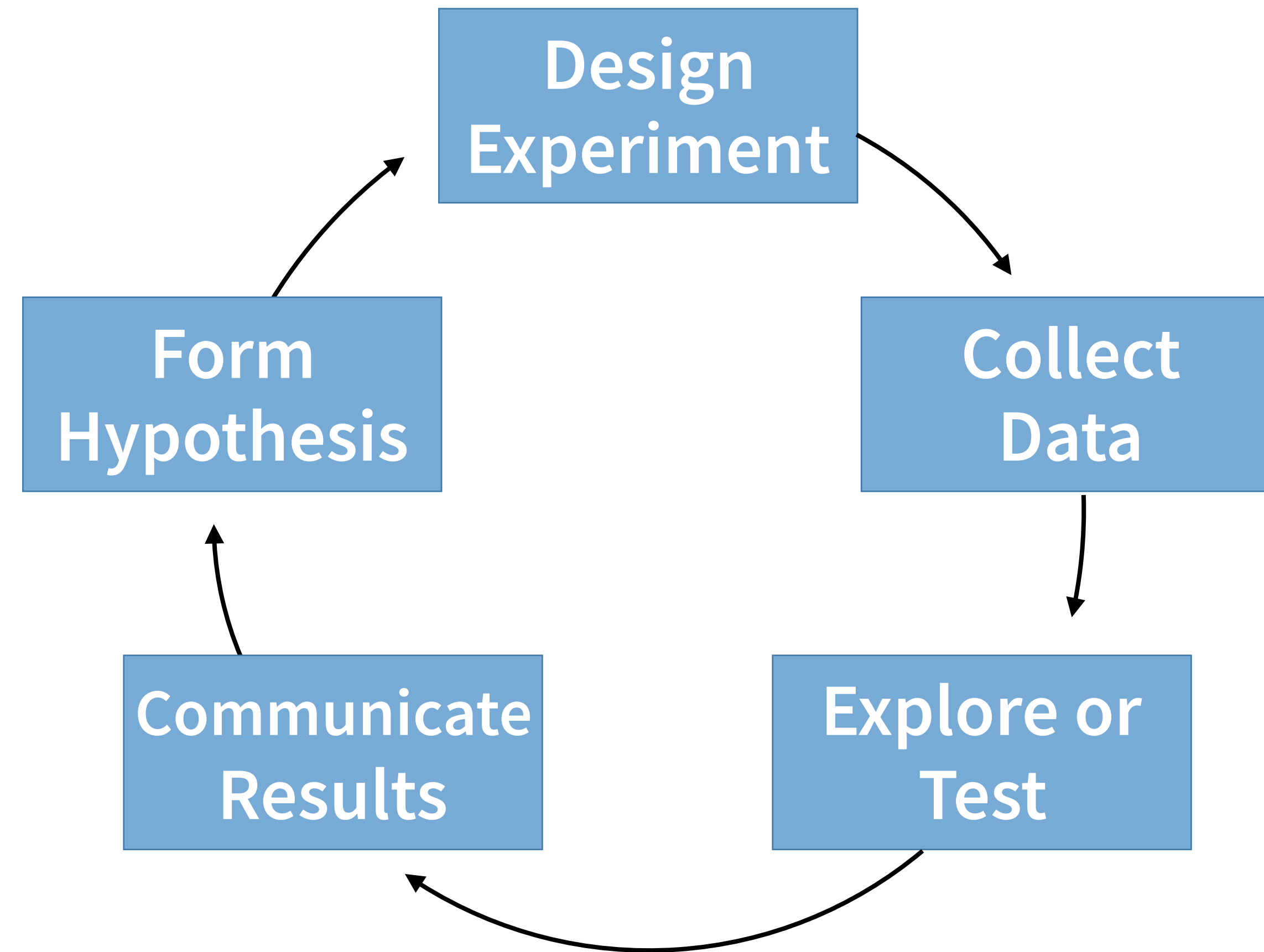
Your Turn

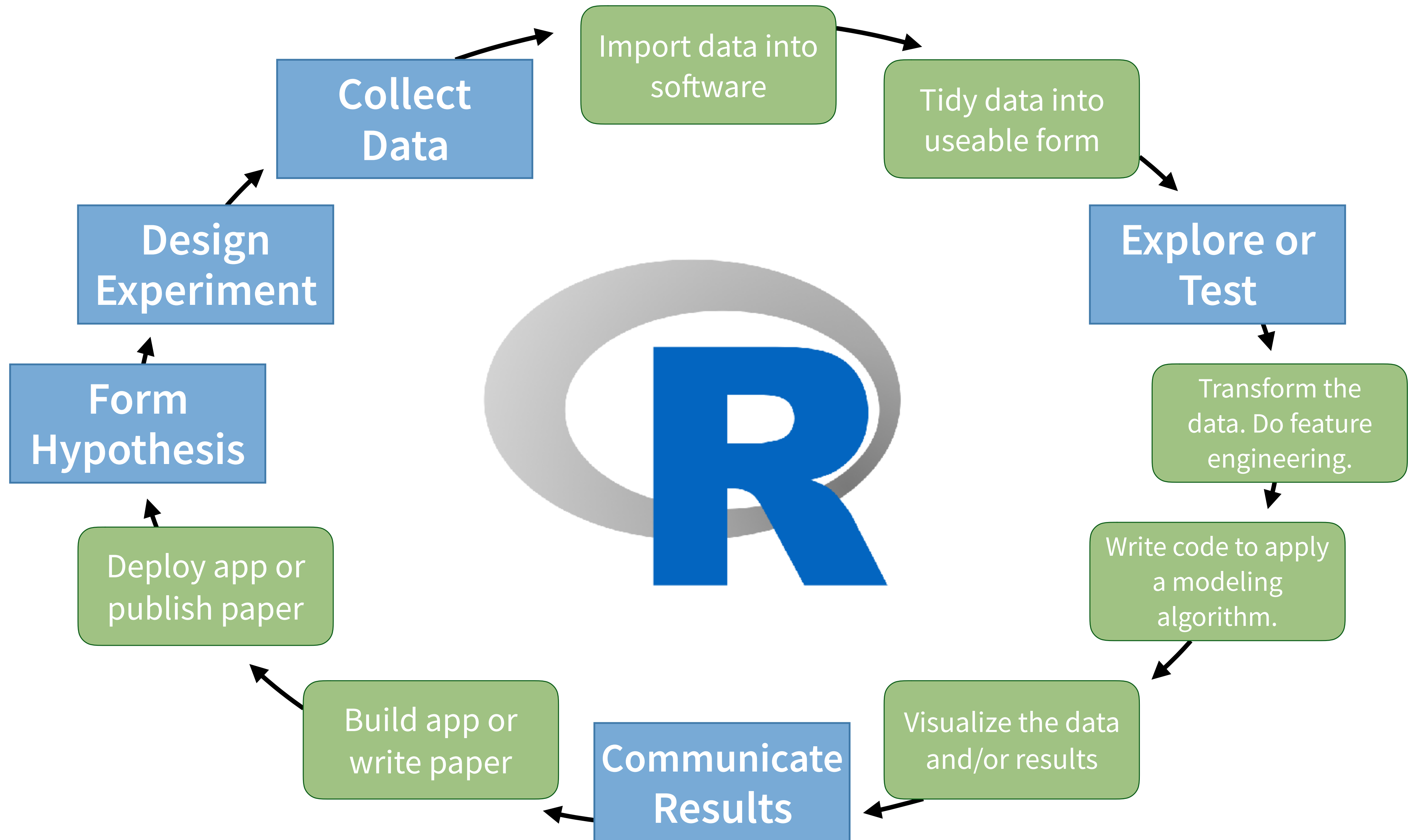
Form groups of 2-4 people. Introduce yourself to your group members. Tell them:

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

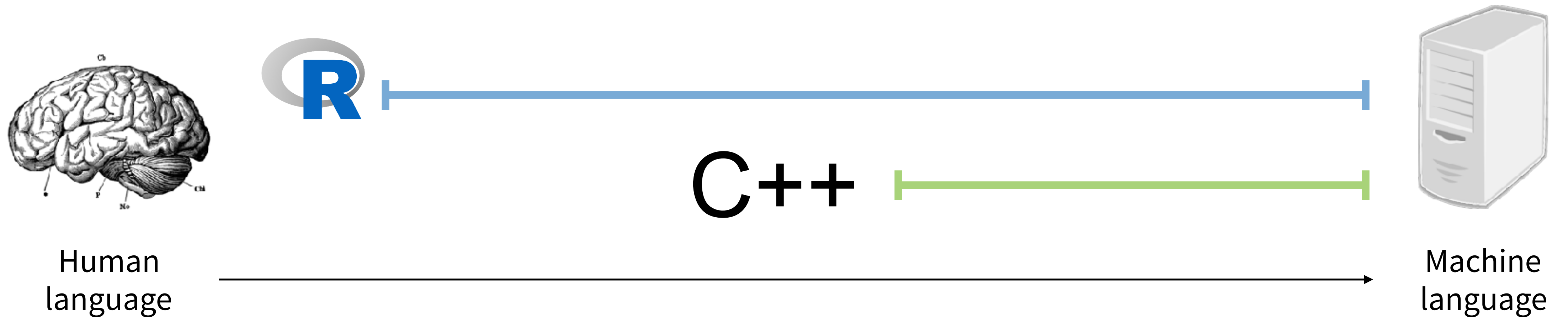
05:00

"Data Science"

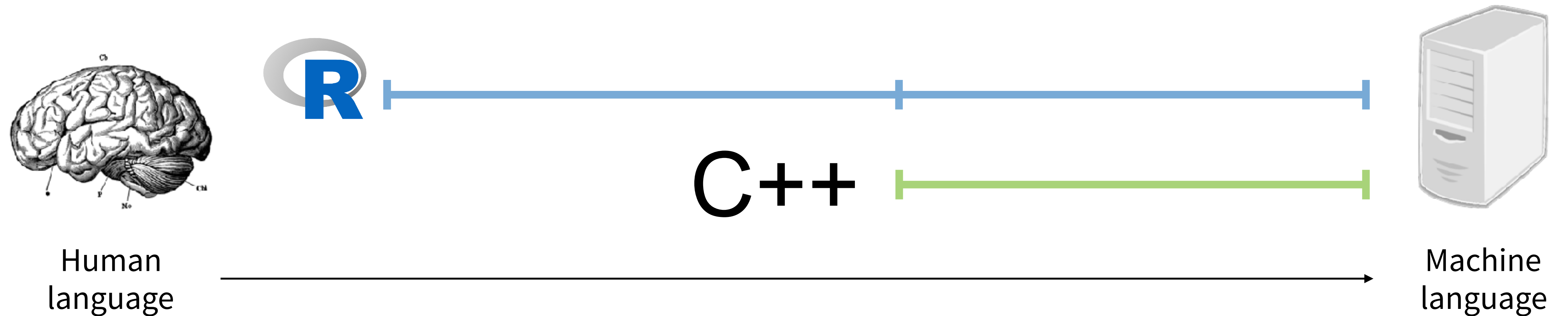


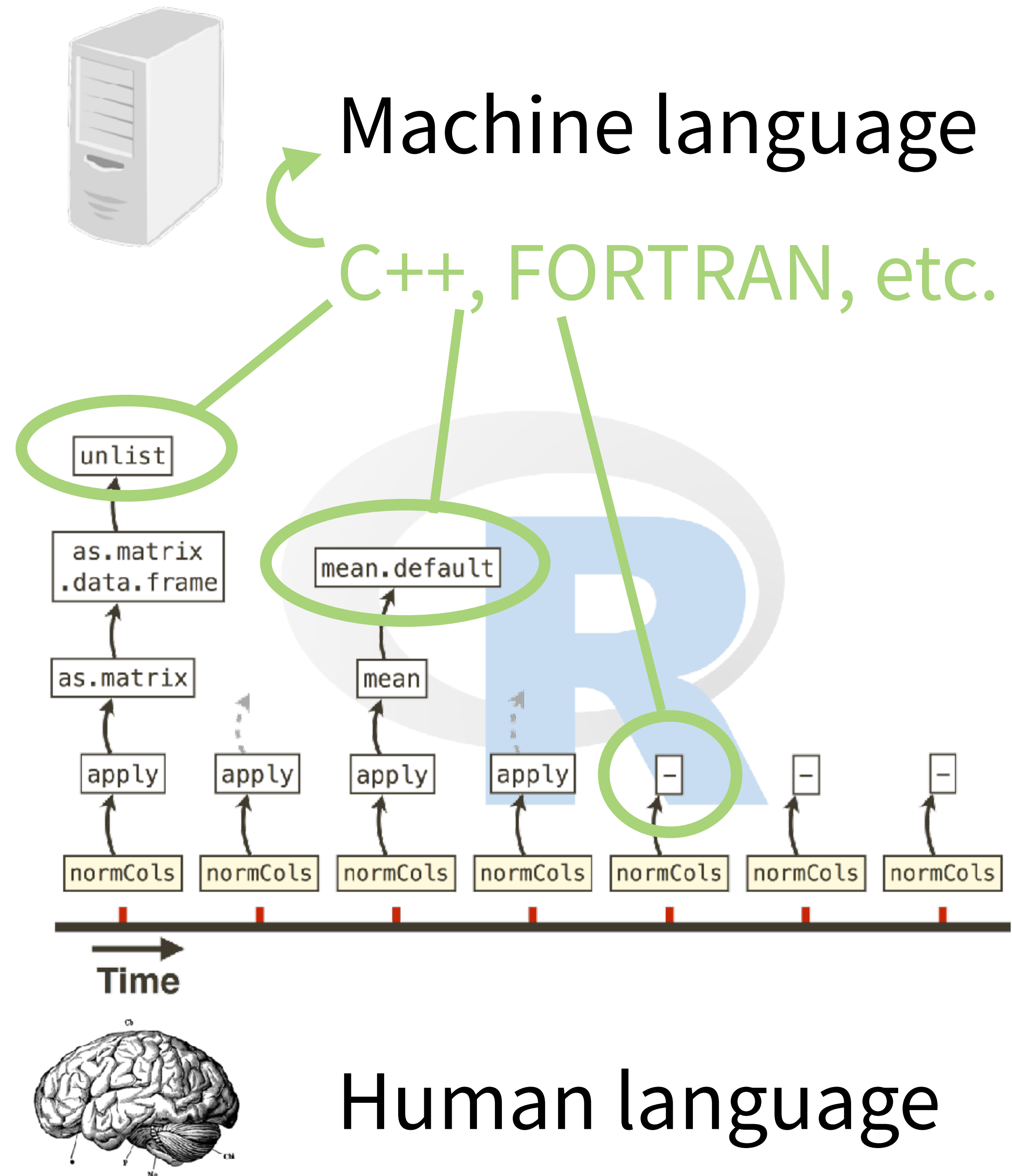


R - A computer language for scientists

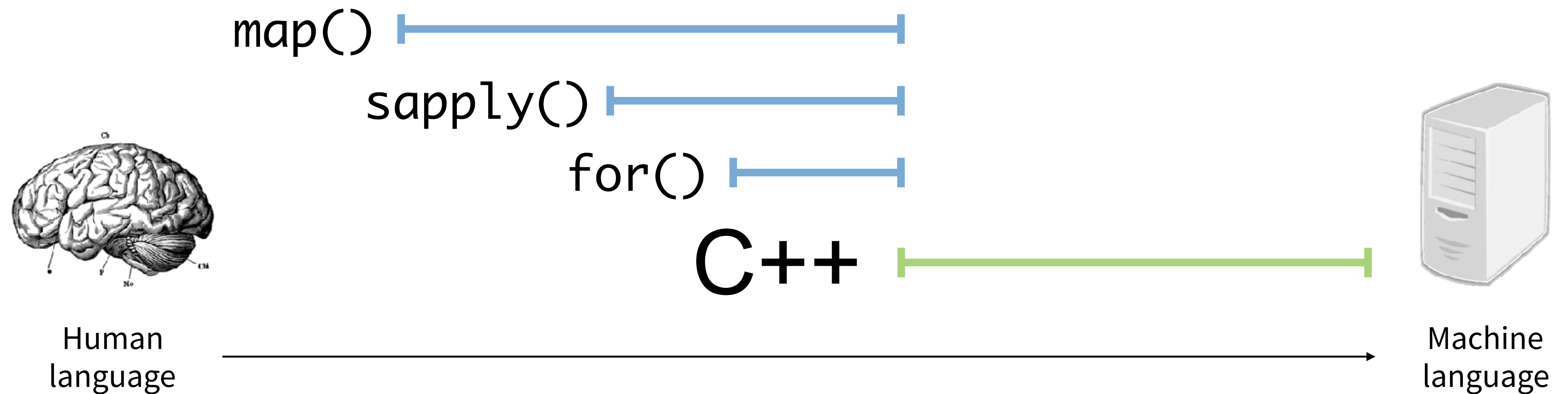


R - A computer language for scientists





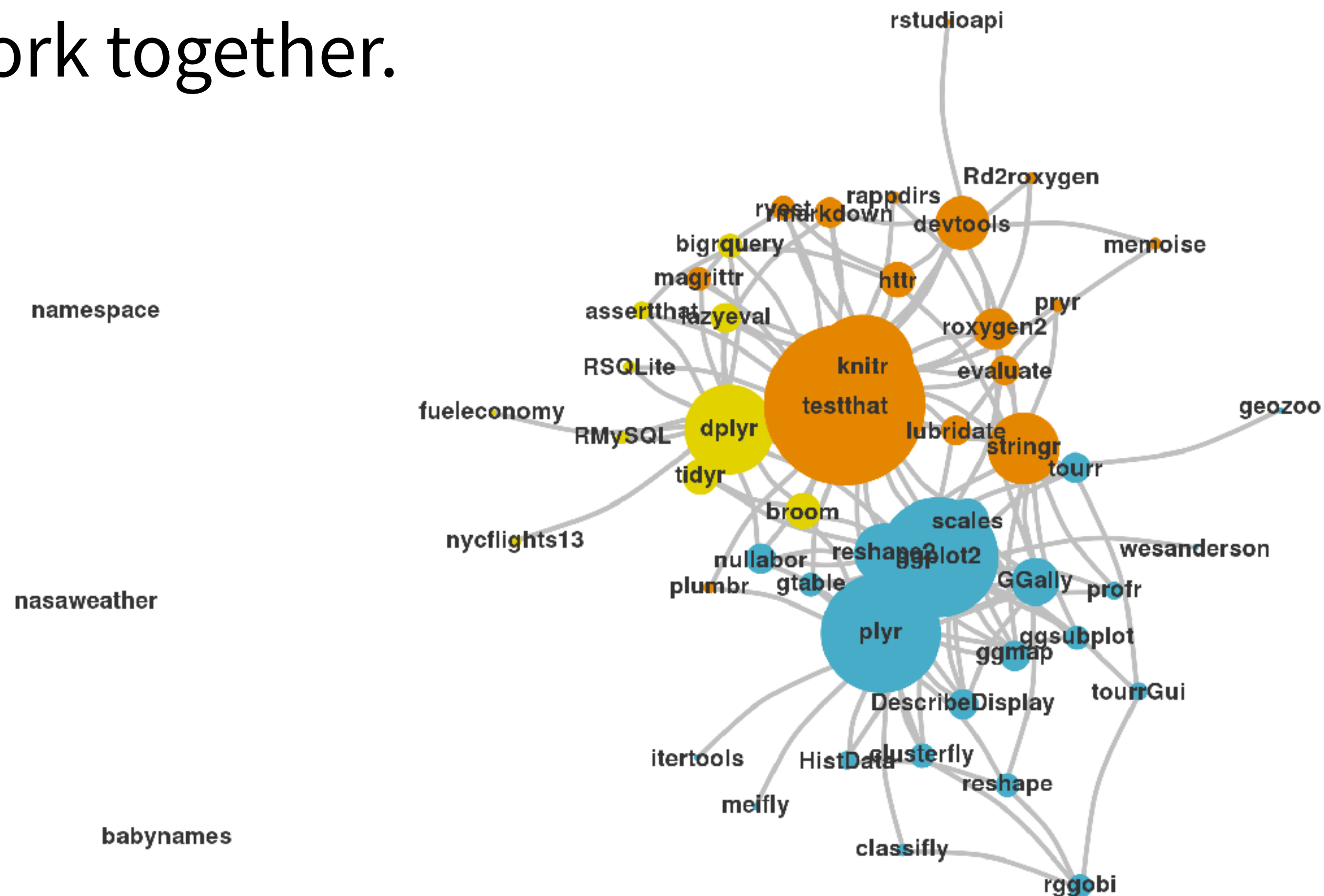
R - A computer language for scientists

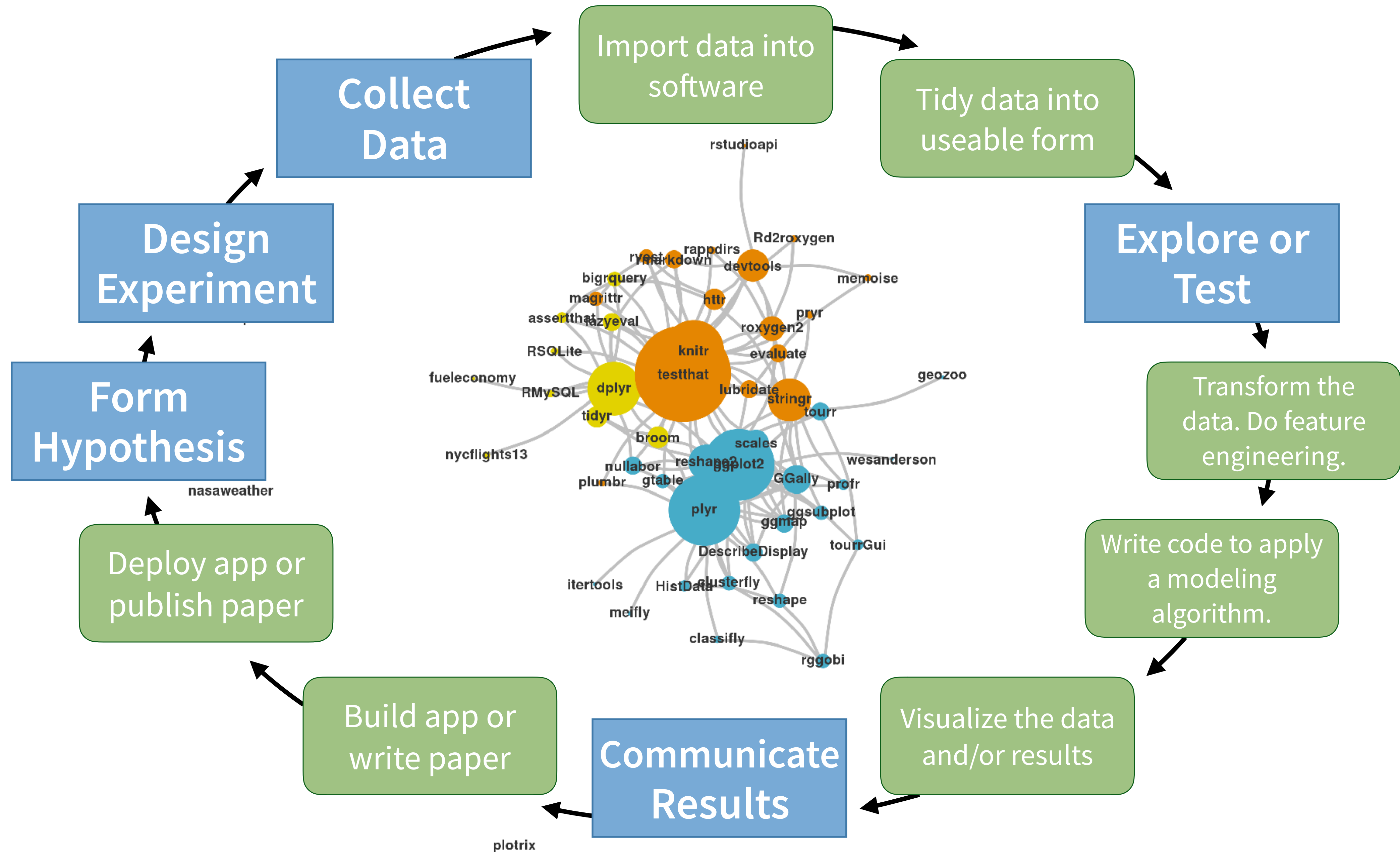




The Tidyverse

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





Tidy tools

Functions are easiest to use when they are:

1. **Simple** - They do one thing, and they do it well
2. **Composable** - They can be combined with other functions for multi-step operations
3. **Smart** - They can use R objects as input.

Tidy functions do these things in a specific way.

Tidy data

Tidy functions all expect and return the same data structure, known as **tidy data**:

Your Turn

Discuss in your group:

"What is the difference between a number and a piece of data?"

03:00

"Data are not just numbers,
they are numbers with a context."

- George Cobb and David Moore (1997)

Subject (car)	Speed (mph)	Distance (ft)
1	25	85
2	24	120
3	24	93
4	24	92

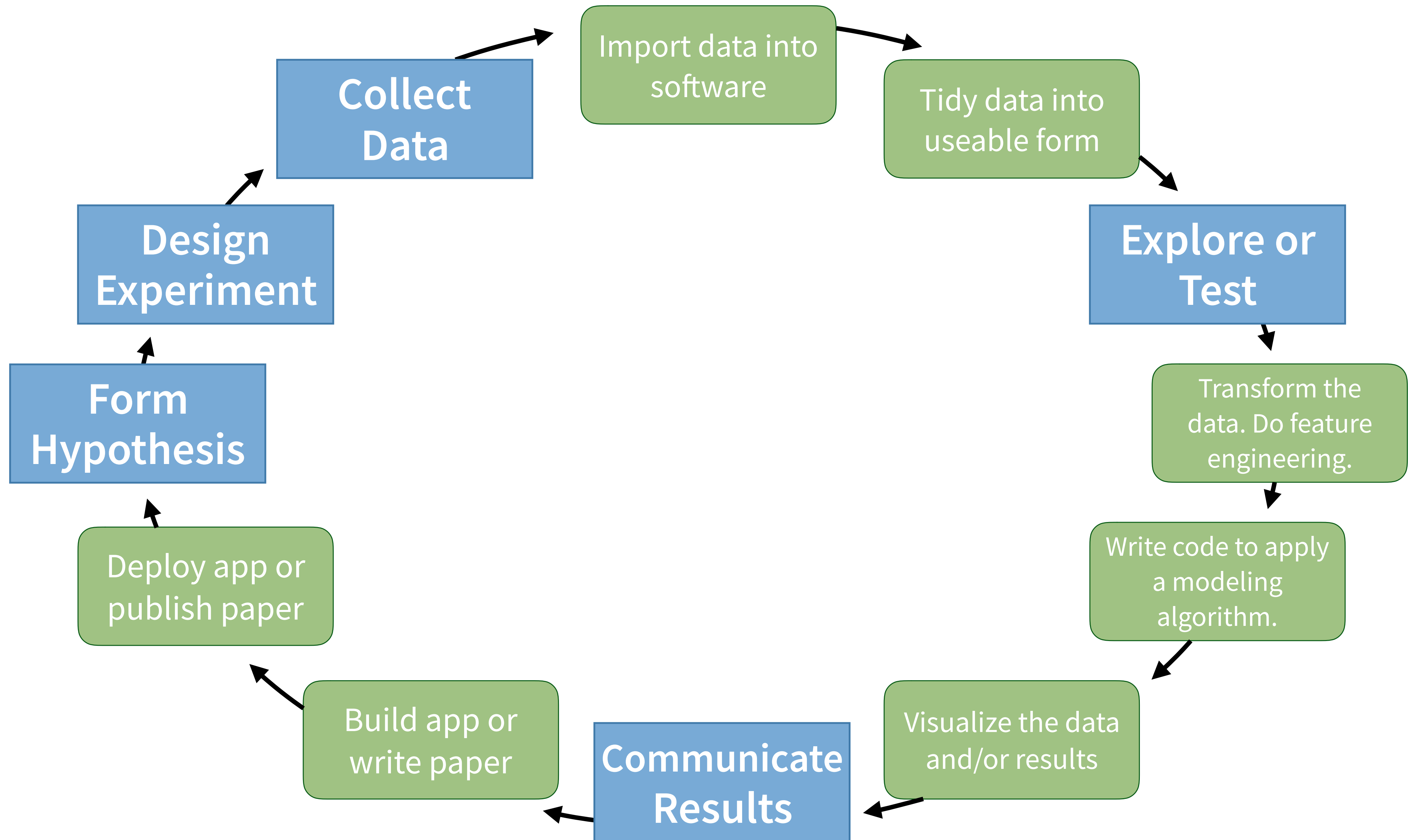
Every value is associated with:

- A variable
- An observation (case)

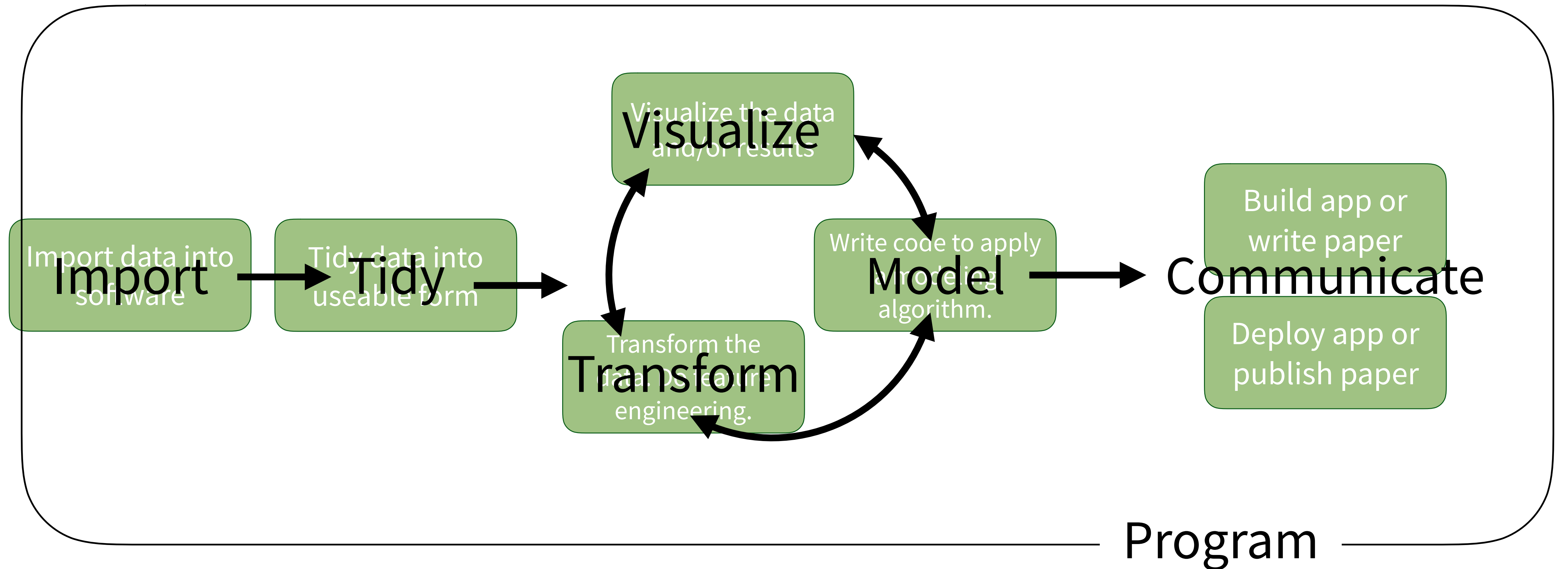
Tidy data

Tidy functions all expect and return the same data structure, known as **tidy data**:

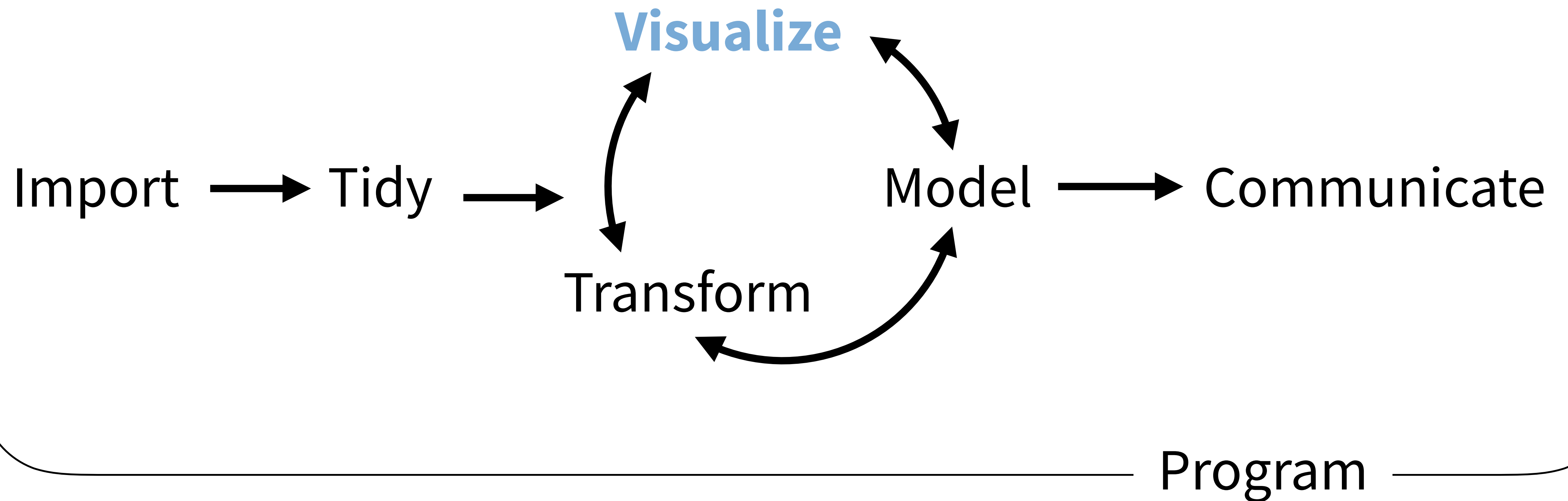
1. A **data frame** that contains
2. **variables** in the **columns** and
3. **cases** in the **rows**.



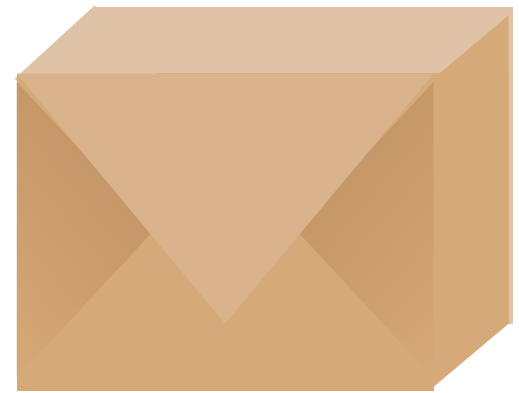
(Applied) Data Science



(Applied) Data Science



tidyverse



An R package that serves as a short cut for installing and loading the components of the 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")
```

```
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")
```

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

does the equivalent of

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

Your Turn

Individually:

Run the code below to load the core tidyverse packages, including ggplot2, dplyr, and tidyr.

```
library(tidyverse)
```

01:00

Master the Tidyverse



Garrett Grolemond

Data Scientist, Educator

January 2017

RStudio