

An introduction to R using the tidyverse

@ BC Stats

Charlotte Wickham

December 2017

cwickham@gmail.com

Introductions

HELLO

my name is

Charlotte

HELLO

my name is

Stephanie

HELLO

my name is

Andy

HELLO

my name is

Sam

Day 2

**Introduction and
Visualize Data**

9:00-10:15

Morning Break

10:15 - 10:30

Visualize Data

10:30 - 12:00

Lunch

12:00 - 1:00

Transform Data

1:00 - 2:30

Afternoon Break

2:30 - 2:45

Tidy Data

2:45 - 4:00

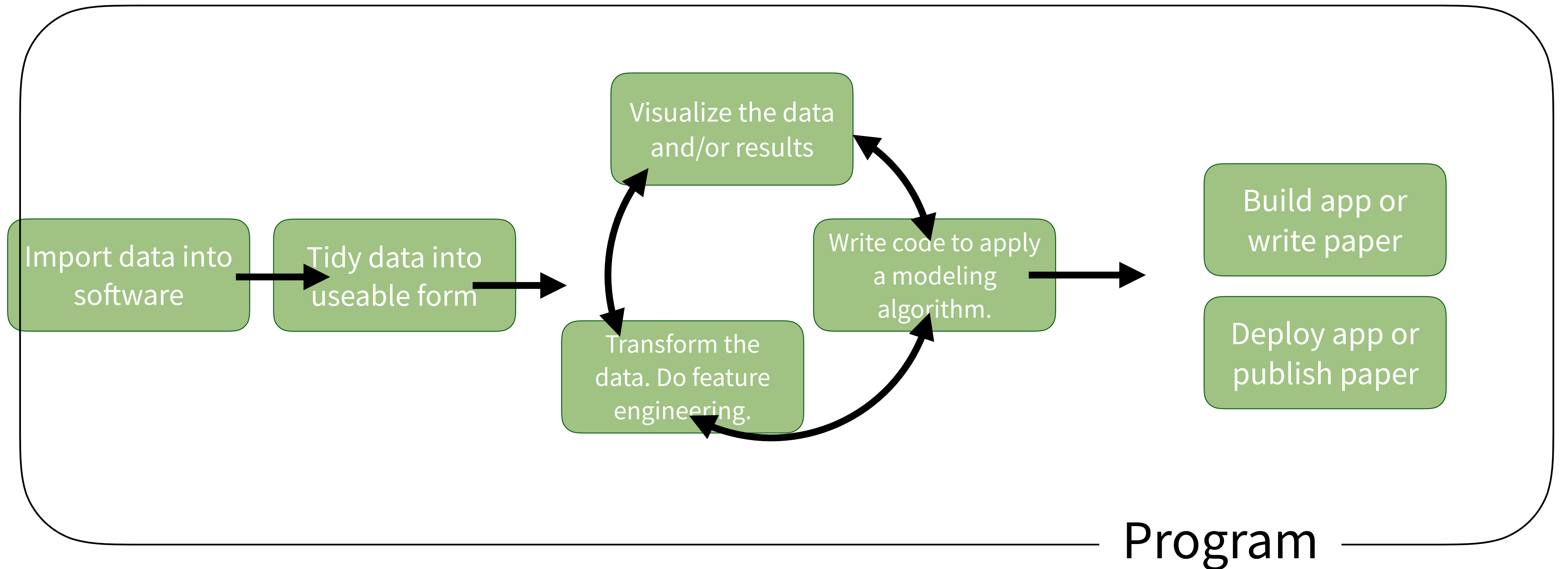
Your Turn

Re-Introduce yourself to your neighbours. Tell them:

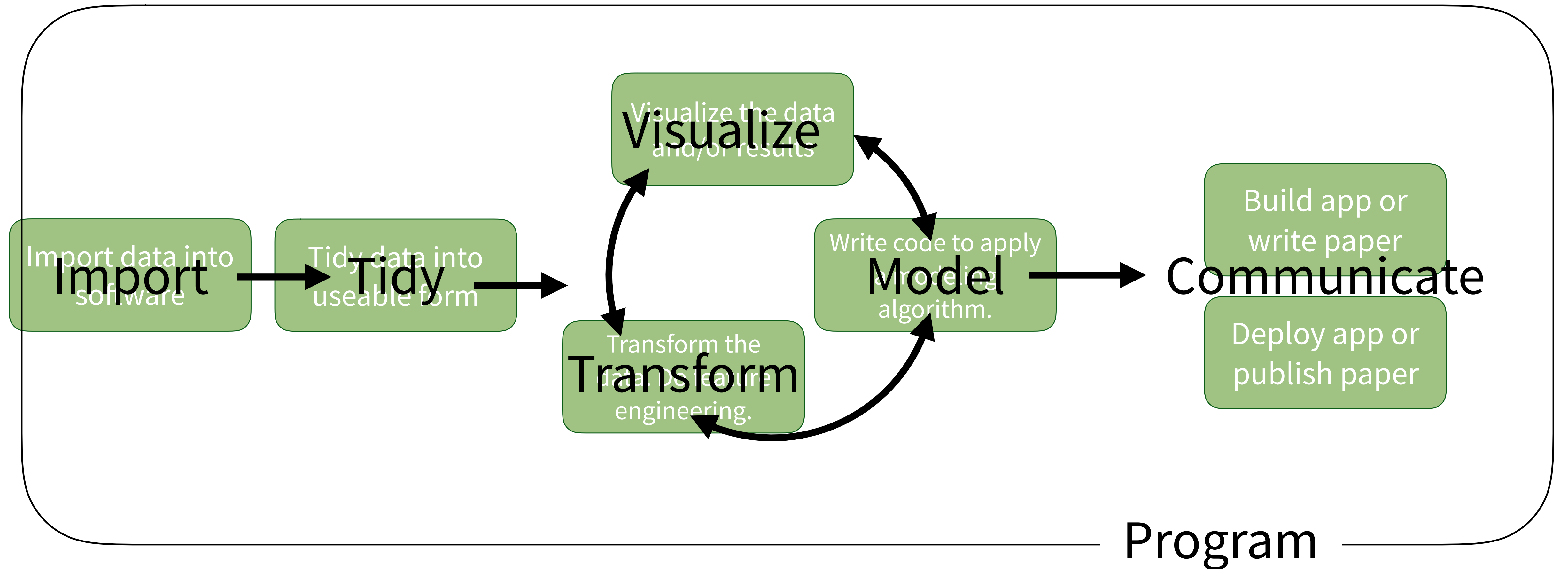
1. Who you are
2. What you do with data
3. How long you have been using R

05:00

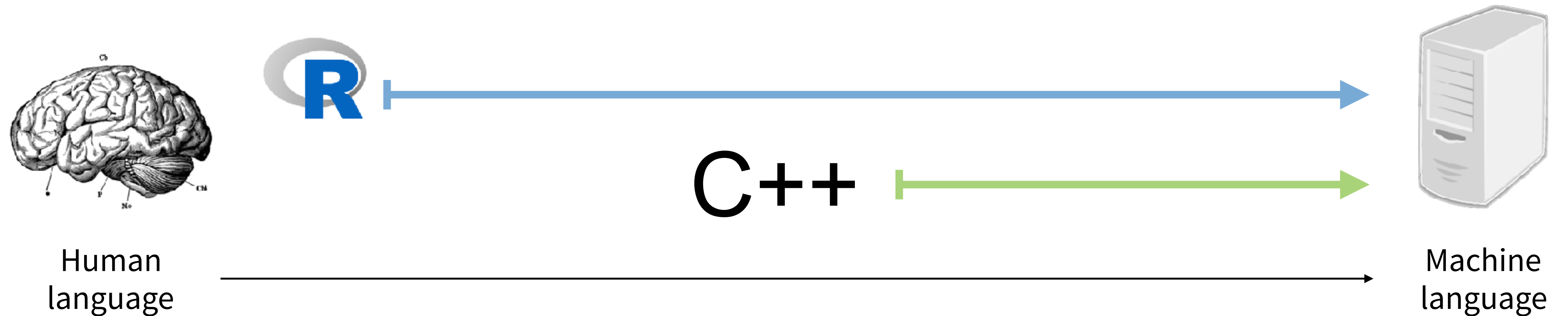
(Applied) Data Science



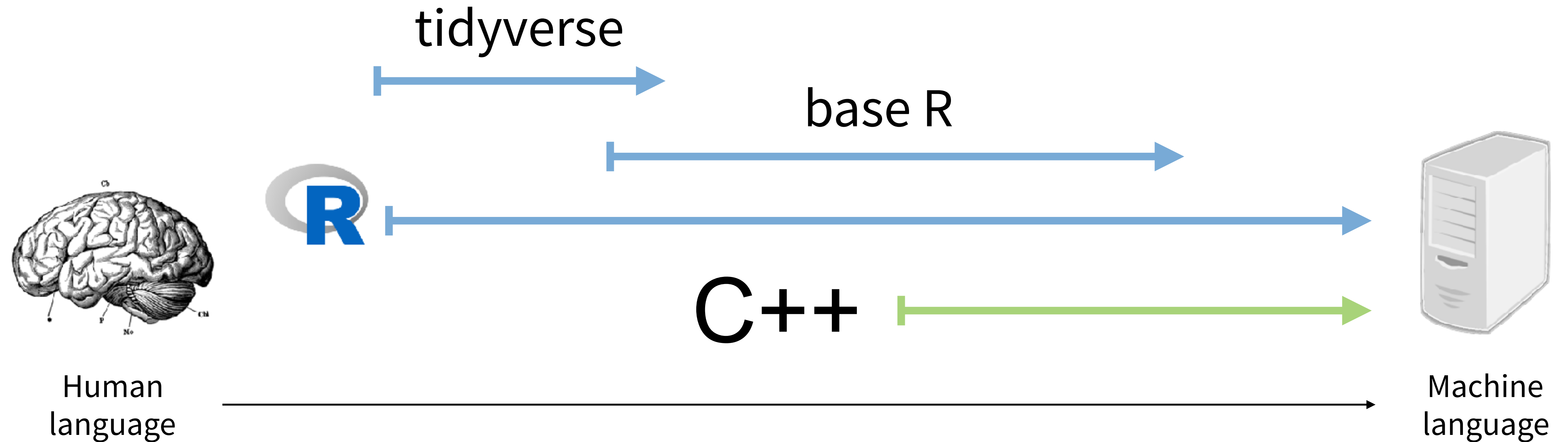
(Applied) Data Science



R - A computer language for scientists

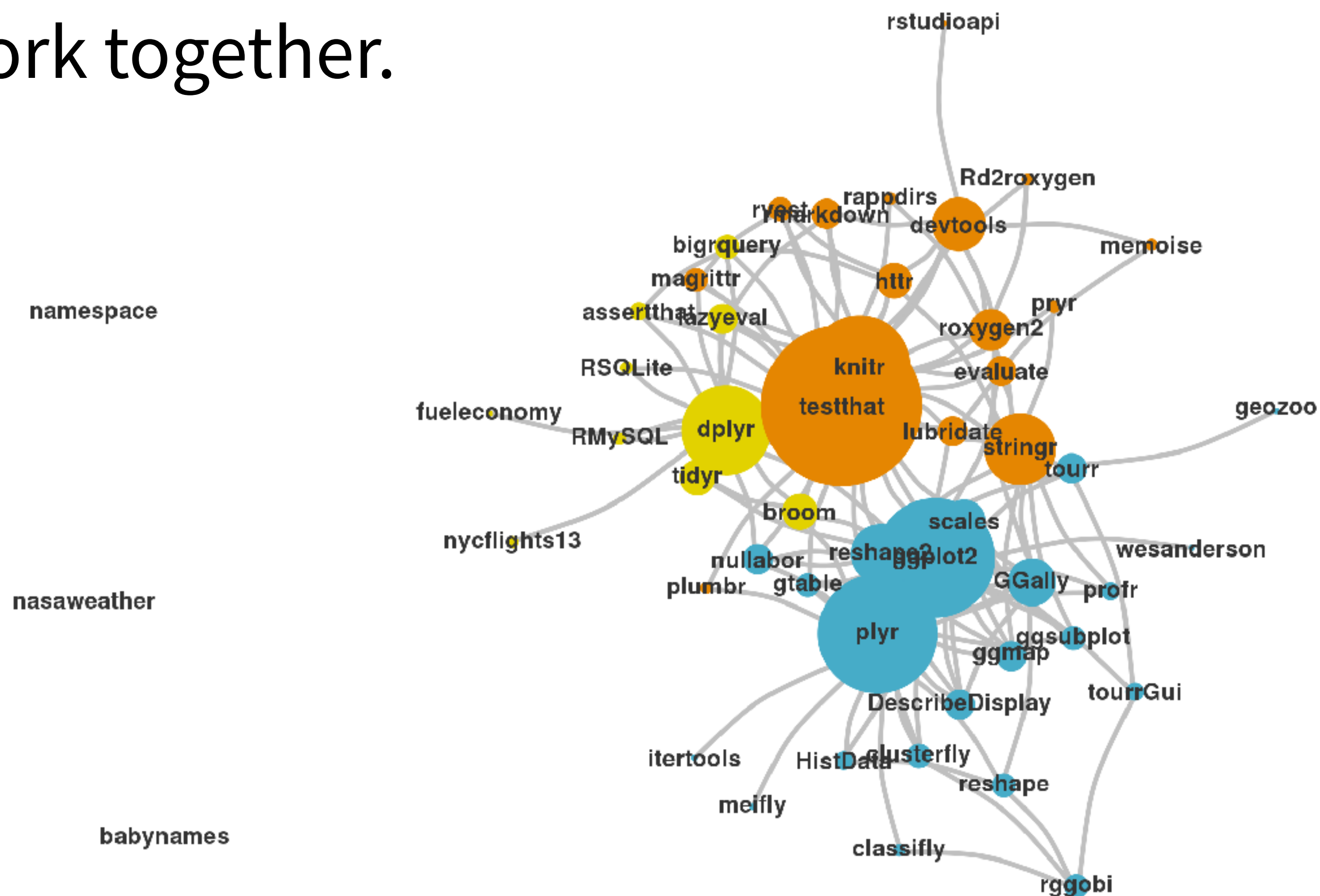


R - A computer language for scientists

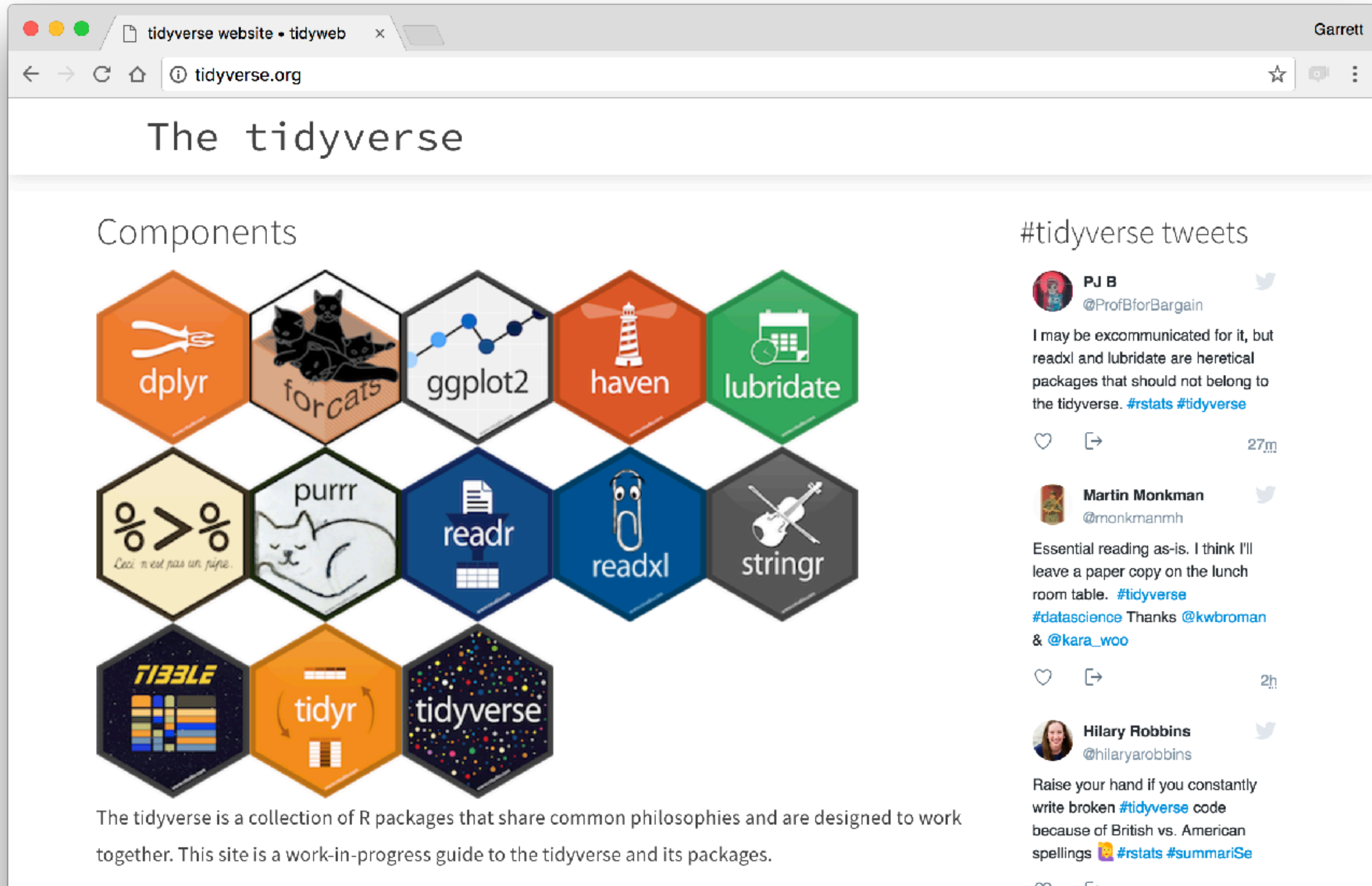


The Tidyverse

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



tidyverse.org

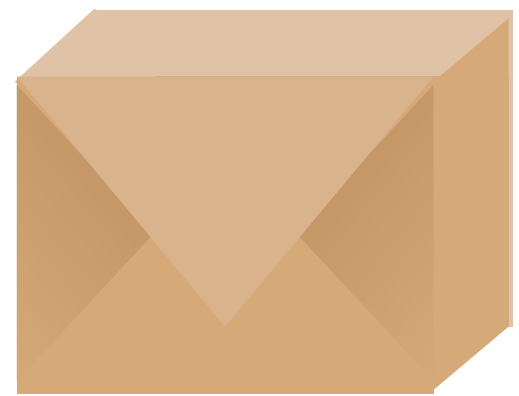


The screenshot shows a web browser window with the URL `tidyverse.org`. The page title is "The tidyverse". Under the heading "Components", there is a grid of 14 hexagonal icons representing different R packages: `dplyr` (orange, pliers), `forcats` (orange, cats), `ggplot2` (grey, network graph), `haven` (orange, lighthouse), `lubridate` (green, calendar), `%>%` (yellow, code symbols), `purrr` (light blue, cat), `readr` (dark blue, document), `readxl` (dark blue, paperclip), `stringr` (dark grey, violin), `TIBBLE` (dark blue, grid), `tidyr` (orange, arrows), and `tidyverse` (dark blue, stars). Below the grid, a paragraph states: "The tidyverse is a collection of R packages that share common philosophies and are designed to work together. This site is a work-in-progress guide to the tidyverse and its packages."

The right sidebar, titled "#tidyverse tweets", displays three tweets:

- PJ B** (@ProfBforBargain) - 27m: "I may be excommunicated for it, but `readxl` and `lubridate` are heretical packages that should not belong to the tidyverse. [#rstats](#) [#tidyverse](#)"
- Martin Monkman** (@monkmanmh) - 2h: "Essential reading as-is. I think I'll leave a paper copy on the lunch room table. [#tidyverse](#) [#datascience](#) Thanks [@kwbroman](#) & [@kara_woo](#)"
- Hilary Robbins** (@hilaryarobbins) - 5m: "Raise your hand if you constantly write broken [#tidyverse](#) code because of British vs. American spellings 🇬🇧 [#rstats](#) [#summarise](#)"

tidyverse



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

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

R Packages

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("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")
```

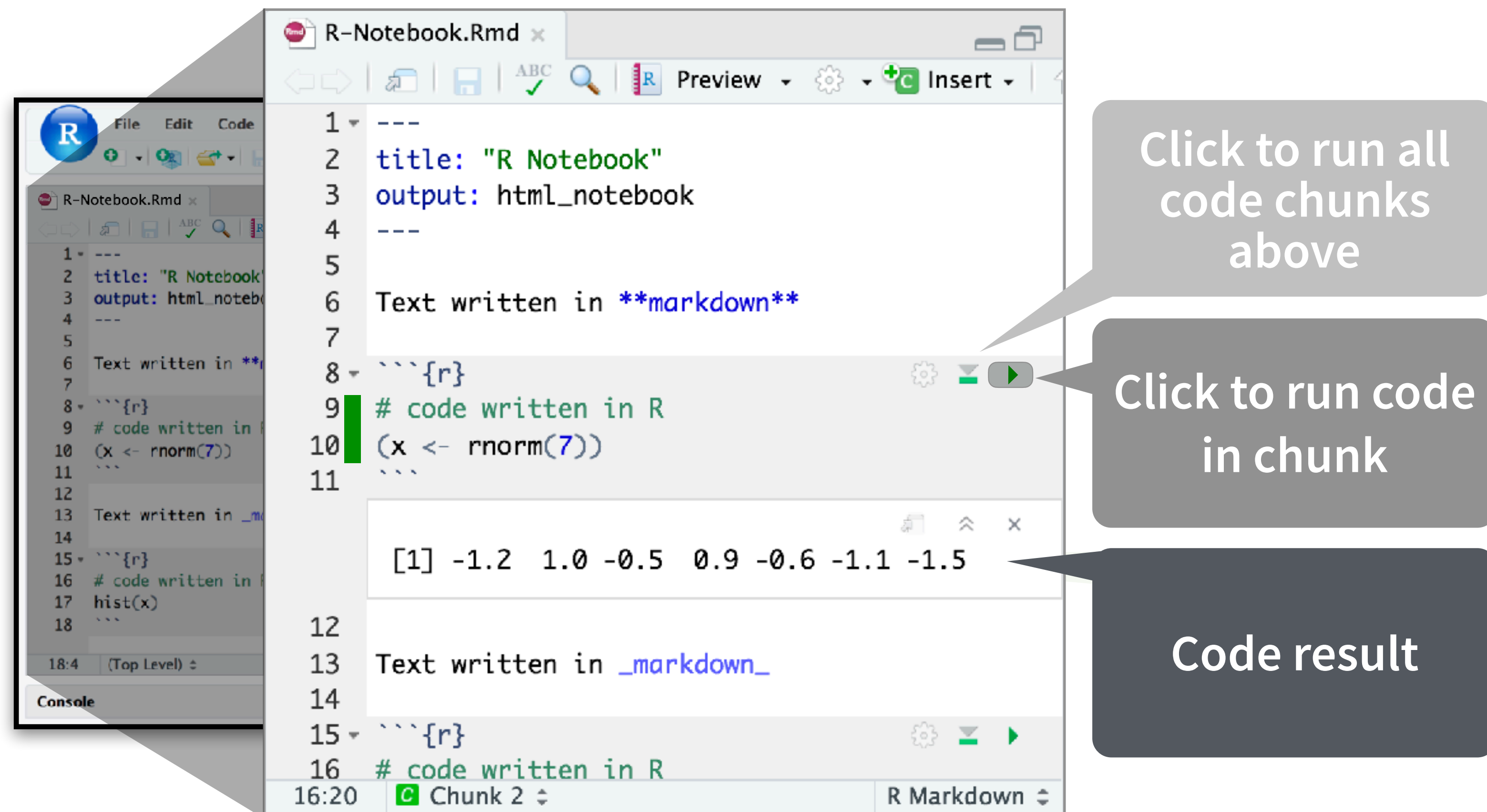
does the equivalent of

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

R Notebooks

R Notebooks

An authoring format for Data Science.



The screenshot displays the RStudio interface with an R Notebook open. The notebook content is as follows:

```
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 ```
```

Annotations with callouts:

- Click to run all code chunks above**: Points to the 'Run All' button (a green play icon with a downward arrow) in the toolbar above the first code chunk.
- Click to run code in chunk**: Points to the 'Run Chunk' button (a green play icon) in the toolbar to the right of the first code chunk.
- Code result**: Points to the output of the first code chunk, which is a vector of seven random numbers: `[1] -1.2 1.0 -0.5 0.9 -0.6 -1.1 -1.5`.

The status bar at the bottom indicates '16:20', 'Chunk 2', and 'R Markdown'.

Your Turn

Open **03-Visualize-Data.Rmd**.

01:00