

An introduction to R using the tidyverse

@ BC Stats

Charlotte Wickham

July 2018

cwickham@gmail.com

To avoid package problems today

In a fresh R session, in the **Console** type:

```
.libPaths()[1]
```

If the output looks like:

```
[1] "C:/Users/[your_user_name]/R/win-library/3.4/"
```

You are good to go. Really, anything that starts with "C:/" should be fine as long as it's not "C:/Program Files/..."

If not, especially if it is either H:/... or \\SFP.IDIR.BCGOV
talk to a TA!

Introductions

HELLO

my name is

Charlotte

Day 2

**Introduction and
Visualize Data**

8:30-10:15

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 - 4:30

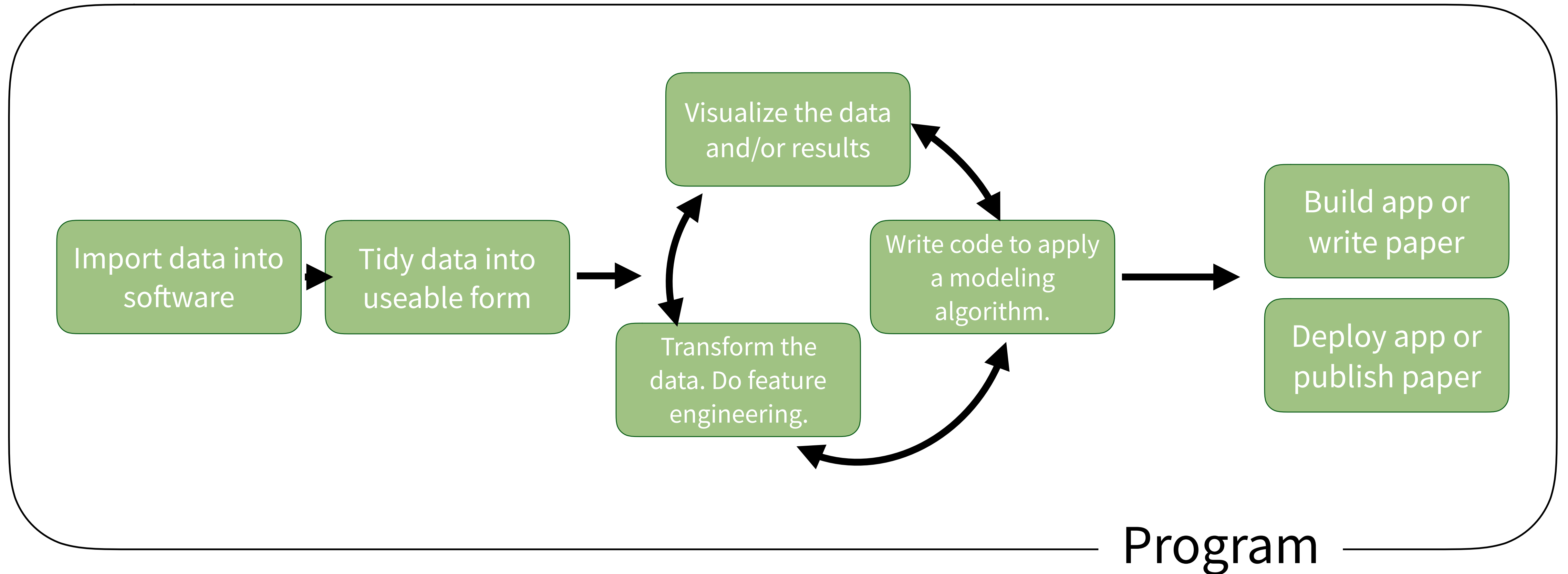
Your Turn

Re-Introduce yourself to your neighbours. Tell them:

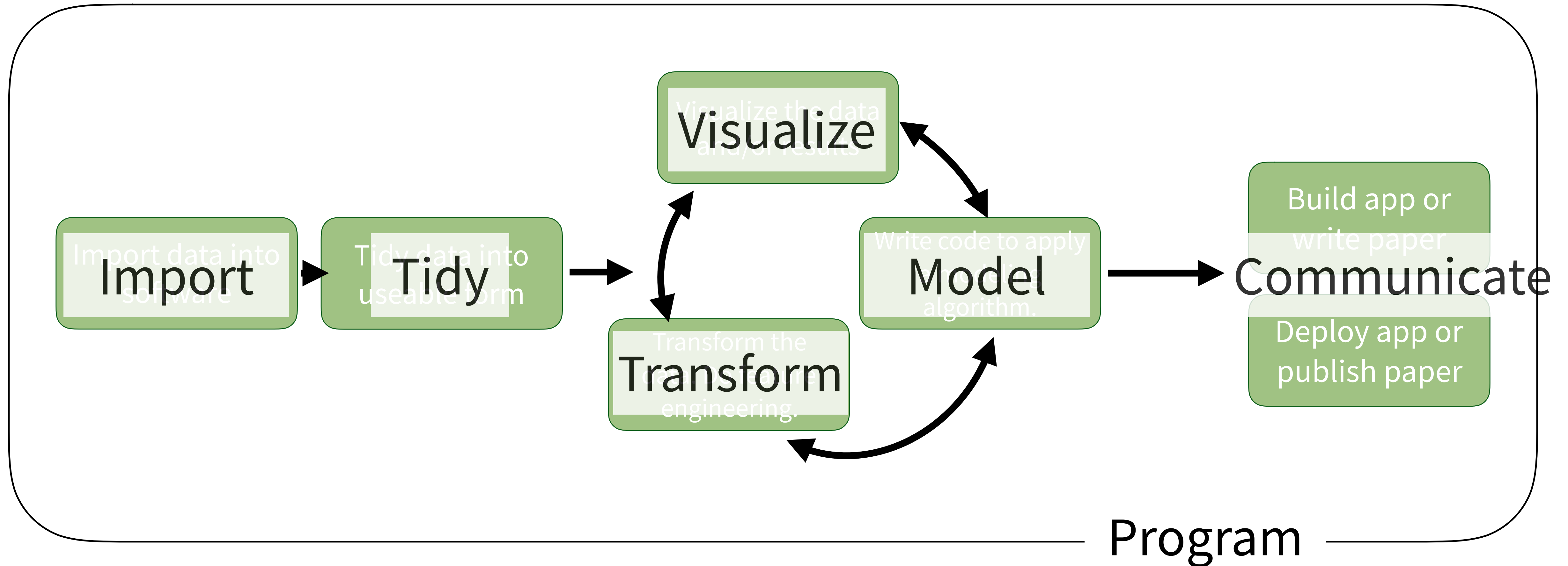
1. Who you are
2. What you do with data
3. How long you have been using R

05:00

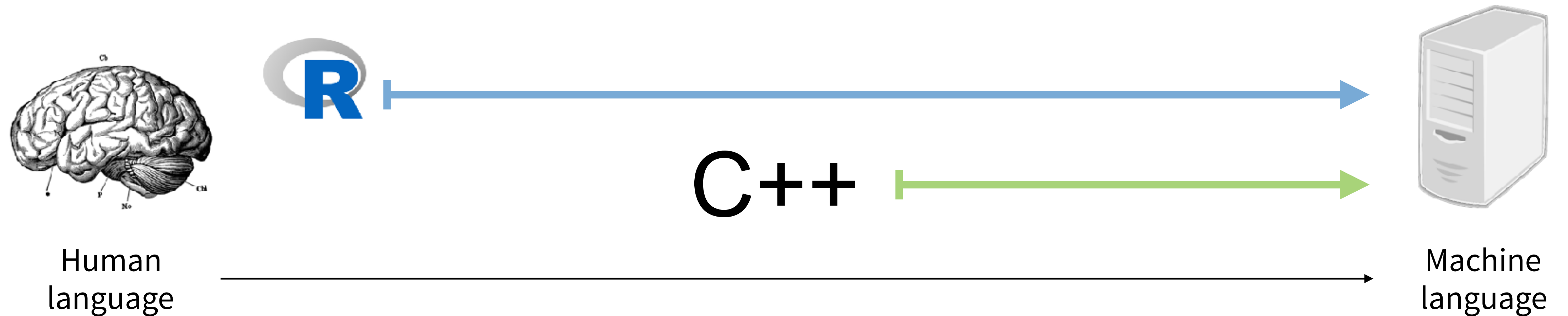
(Applied) Data Science



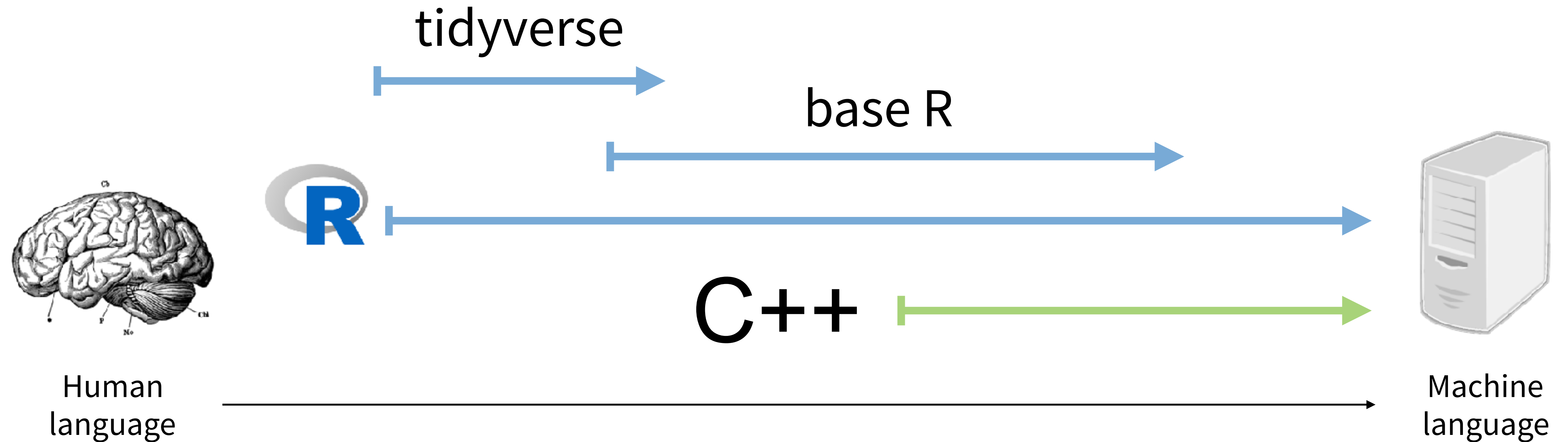
(Applied) Data Science



R - A computer language for scientists

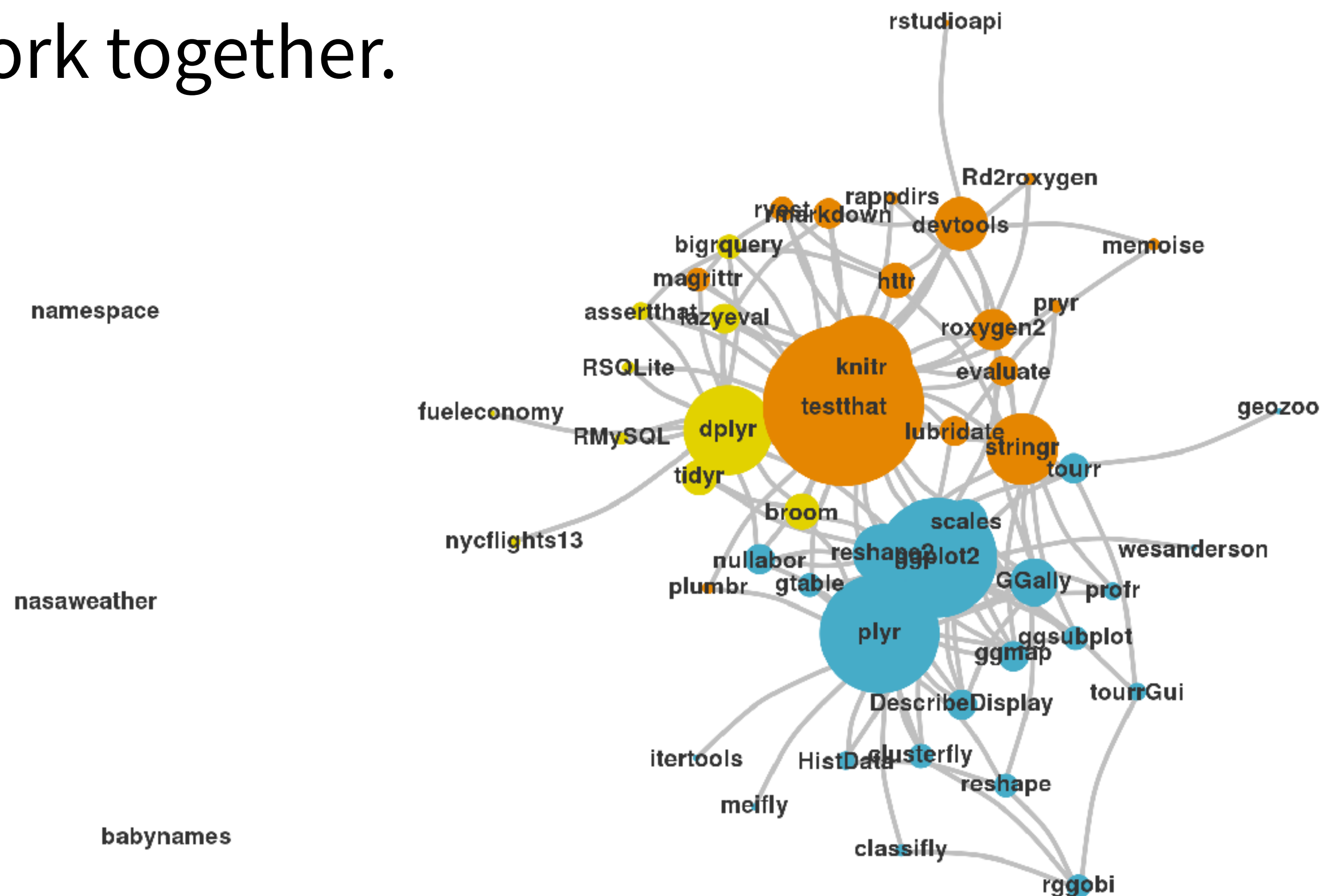


R - A computer language for scientists

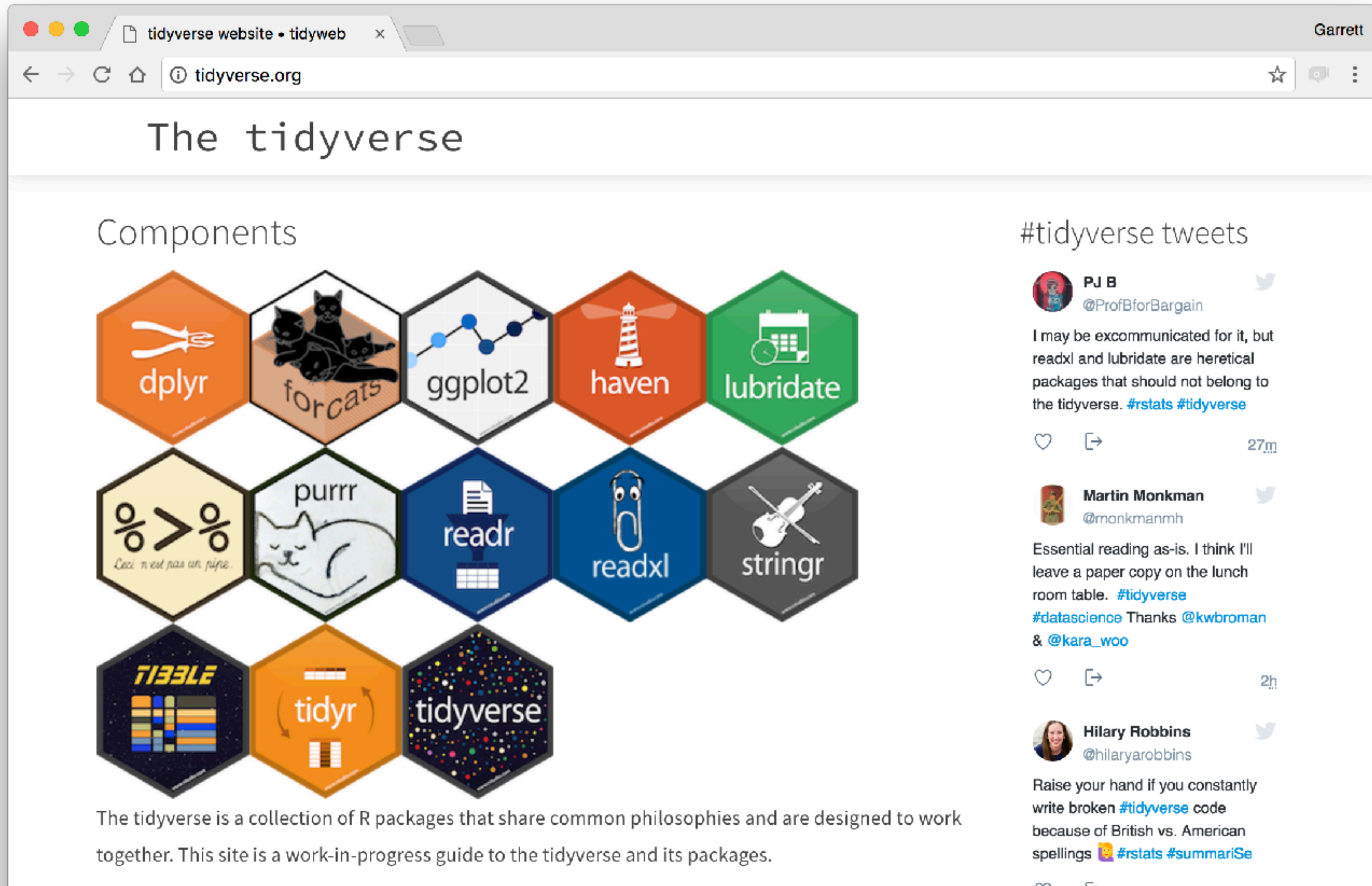


The Tidyverse

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



tidyverse.org

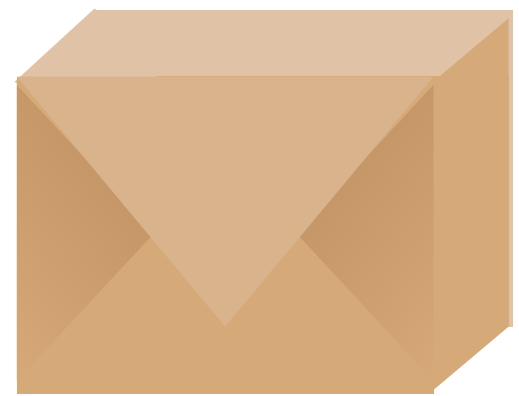


The screenshot shows a web browser window with the URL `tidyverse.org`. The page title is "The tidyverse". Under the heading "Components", there is a grid of 14 hexagonal icons representing different R packages: `dplyr` (orange, pliers), `forcats` (orange, cats), `ggplot2` (grey, network graph), `haven` (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."

On the right side, under the heading "#tidyverse tweets", there are 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")
```

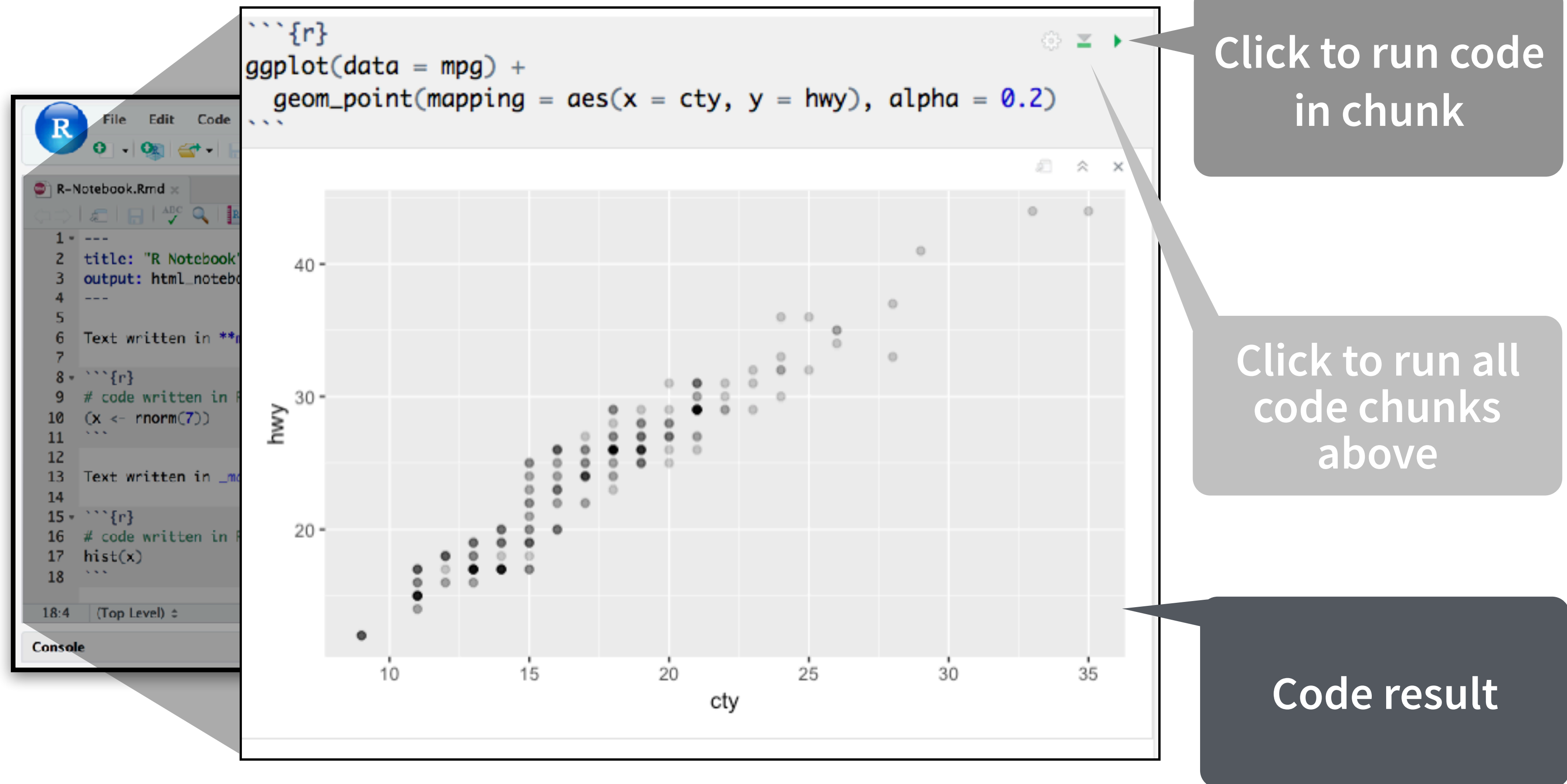
does the equivalent of

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

R Notebooks

RMarkdown

An authoring format for Data Science.



The image shows a screenshot of the RStudio interface. On the left, the R Markdown editor displays a document with several code chunks. The first chunk is highlighted, showing the following R code:

```
```{r}
ggplot(data = mpg) +
 geom_point(mapping = aes(x = cty, y = hwy), alpha = 0.2)
```
```

On the right, the output of this code chunk is displayed as a scatter plot. The x-axis is labeled 'cty' and ranges from 10 to 35. The y-axis is labeled 'hwy' and ranges from 20 to 40. The plot shows a positive correlation between city mileage and highway mileage, with points colored by engine displacement.

Three callout boxes provide instructions:

- Click to run code in chunk**: Points to the green play button icon in the top right corner of the code chunk.
- Click to run all code chunks above**: Points to the green play button icon in the top right corner of the output area.
- Code result**: Points to the scatter plot output.

Your Turn

1. Open RStudio by opening **r_intro_bc_stats.Rproj**.
(Or open RStudio then File -> Recent Projects -> r_intro_bc_stats)
2. Open **03-Visualize-Data.Rmd**.

01:00