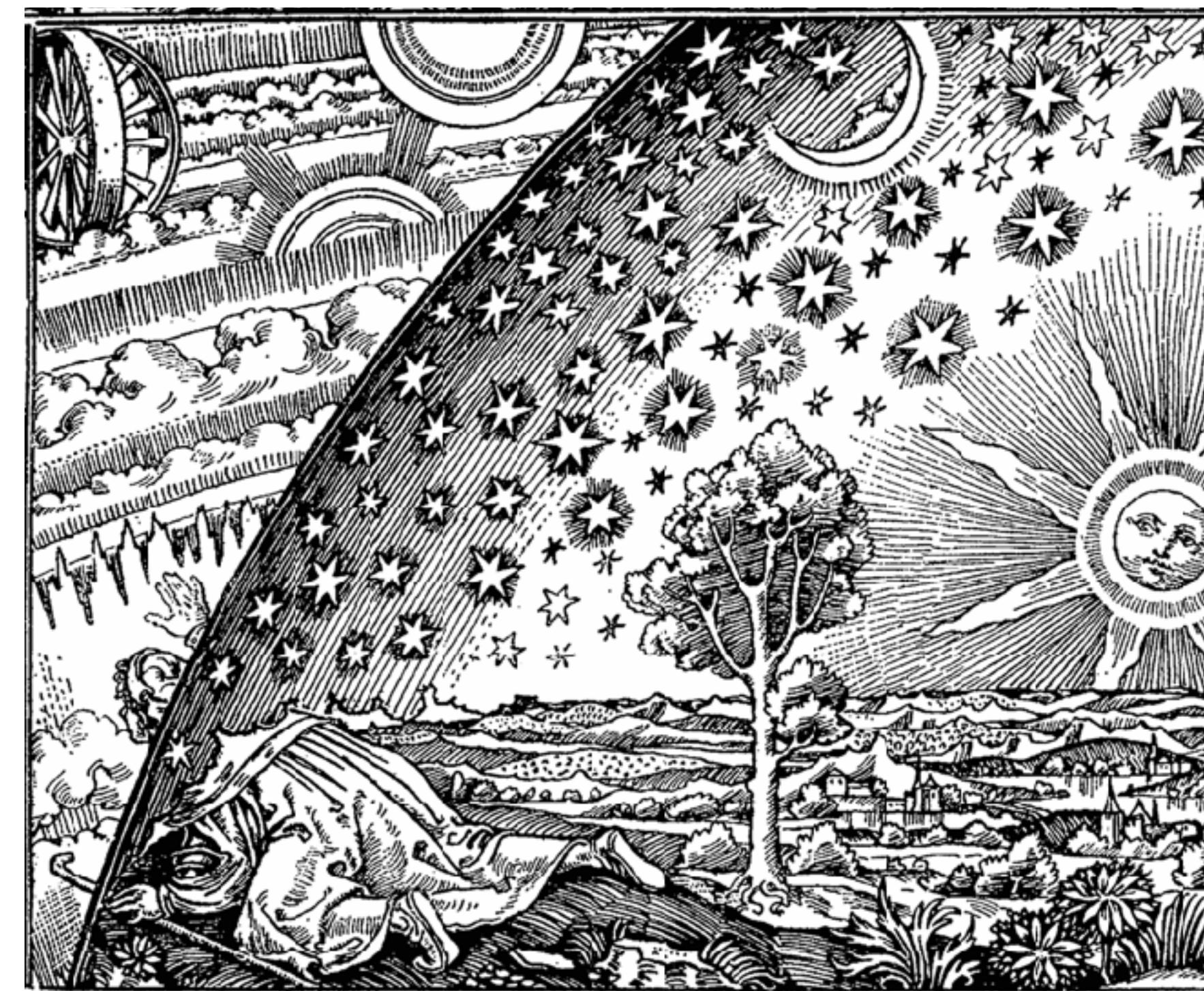


Master the Tidyverse



Garrett Grolemund

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

Nathan



@nwstephens

HELLO

my name is

Dallam

Day 1

Introduction and
Visualize Data

8:30 - 10:15

Morning Break

10:15 - 10:30

Transform Data

10:30 - 12:30

Lunch

12:30 - 2:00

Transform Data

2:00 - 3:15

Afternoon Break

3:15 - 3:30

Tidy Data

3:30 - 5:00

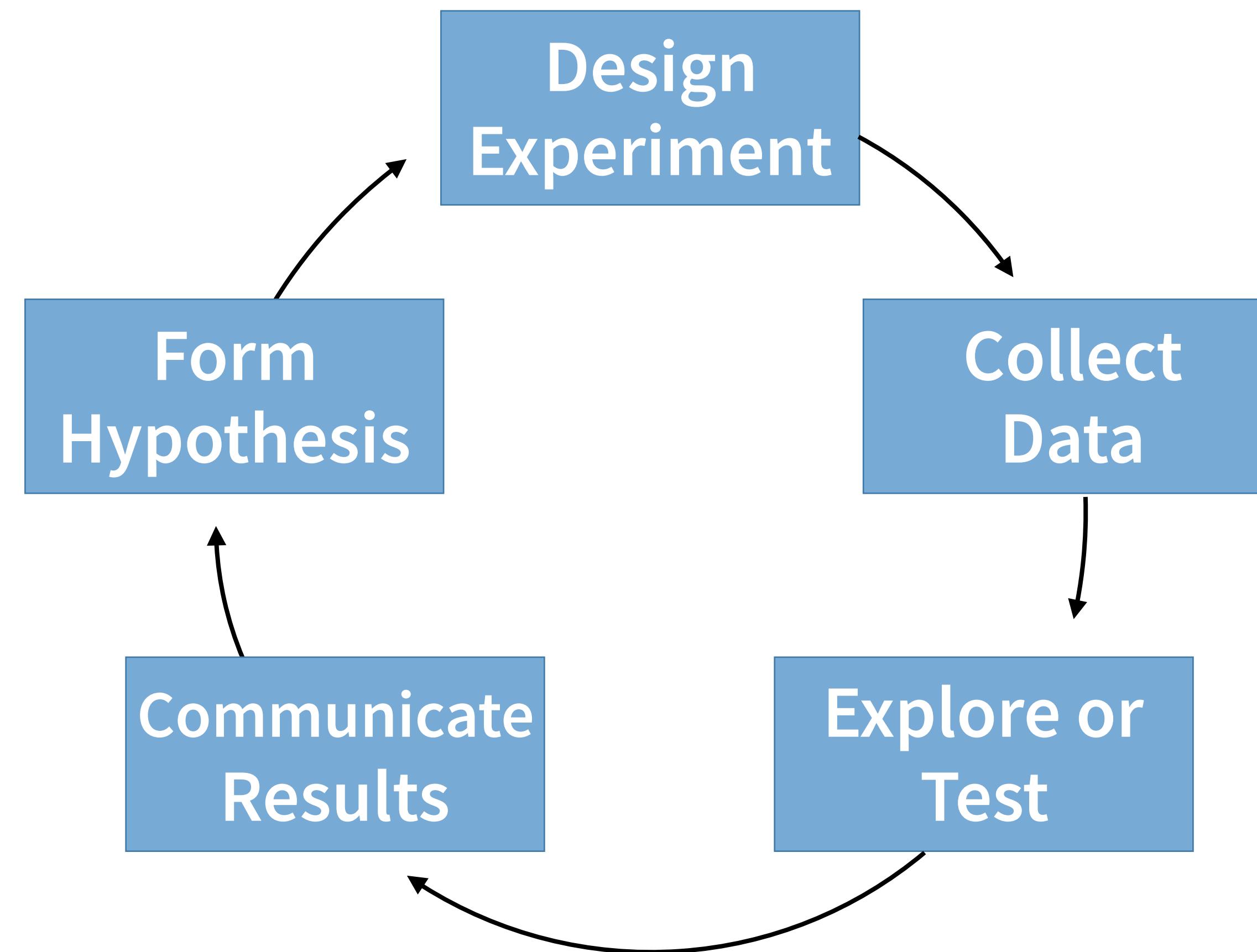
Your Turn

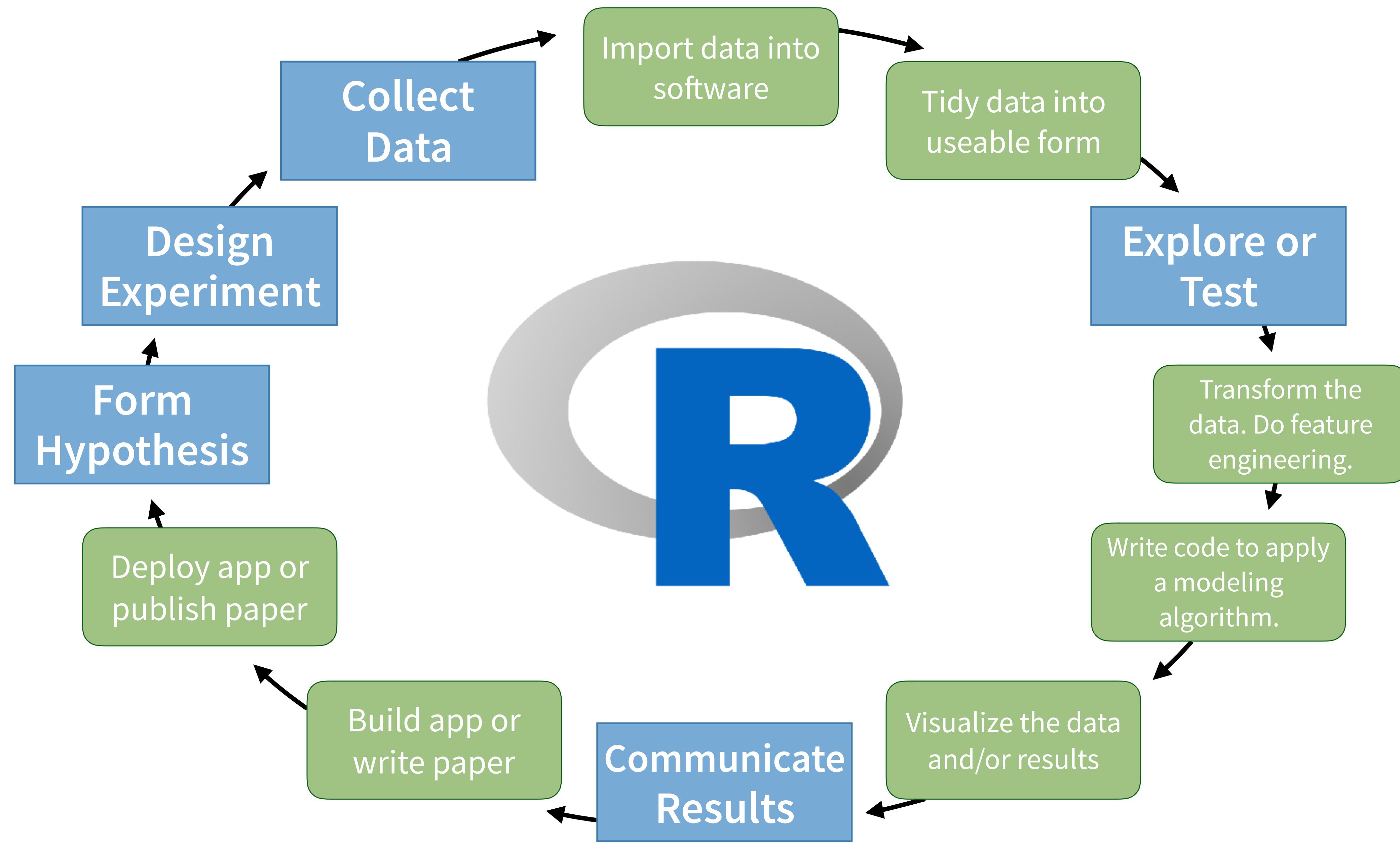
Form groups of 2-4 people. Introduce yourself to your group members. Tell them:

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

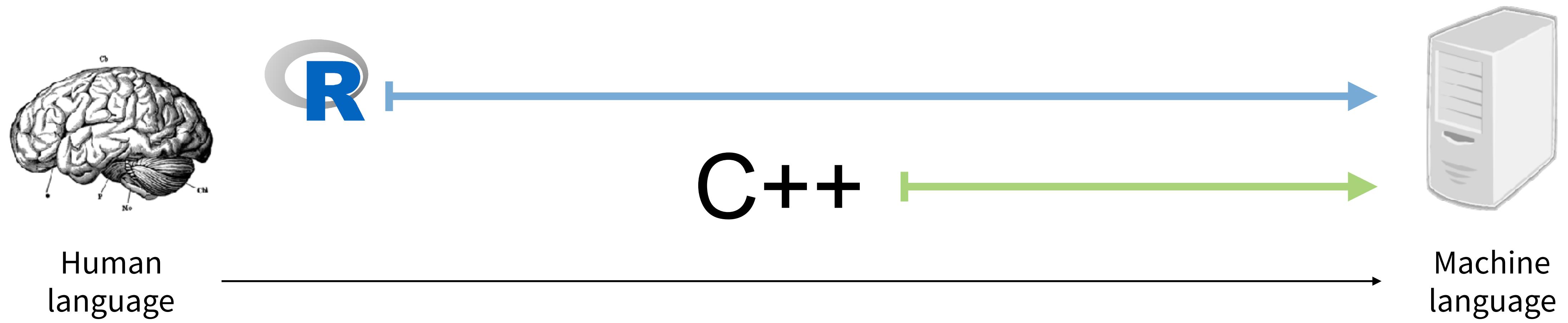


"Data Science"

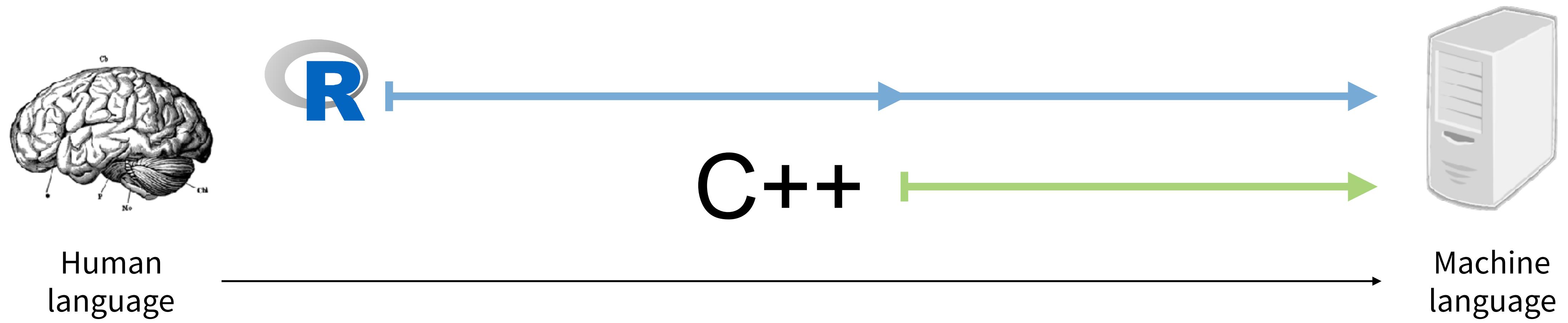




R - A computer language for scientists



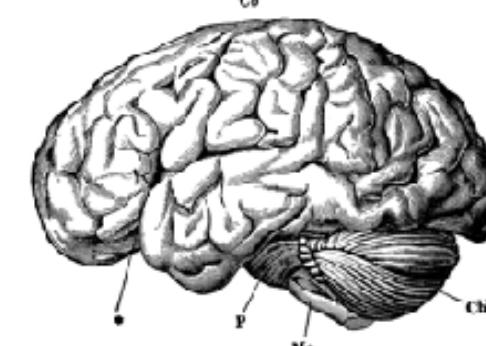
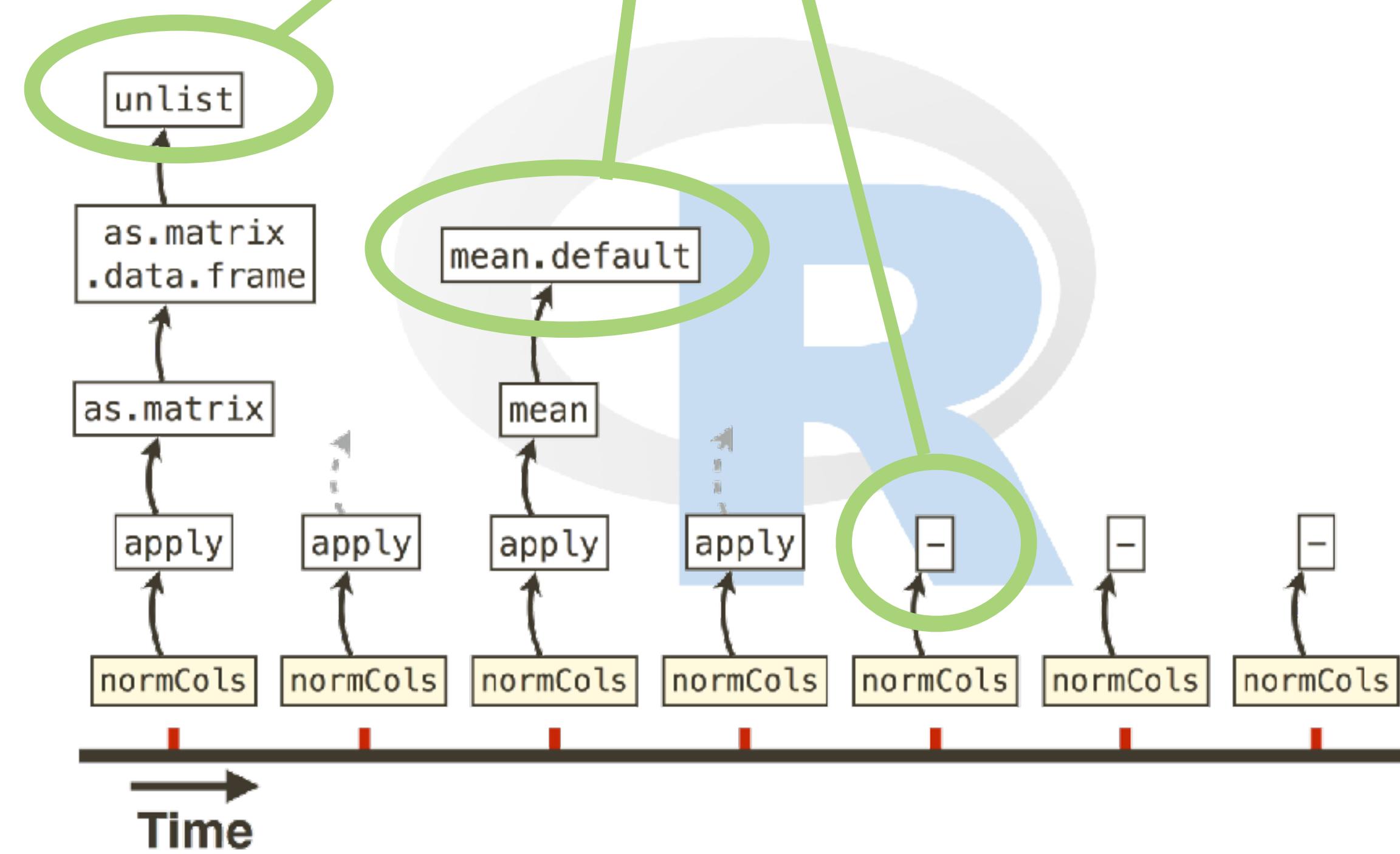
R - A computer language for scientists





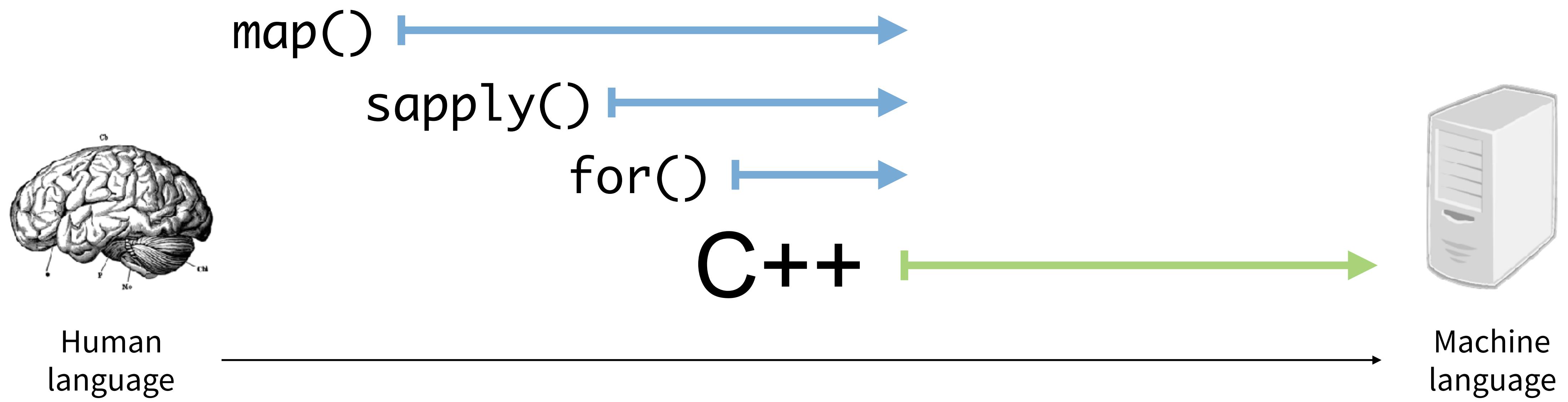
Machine language

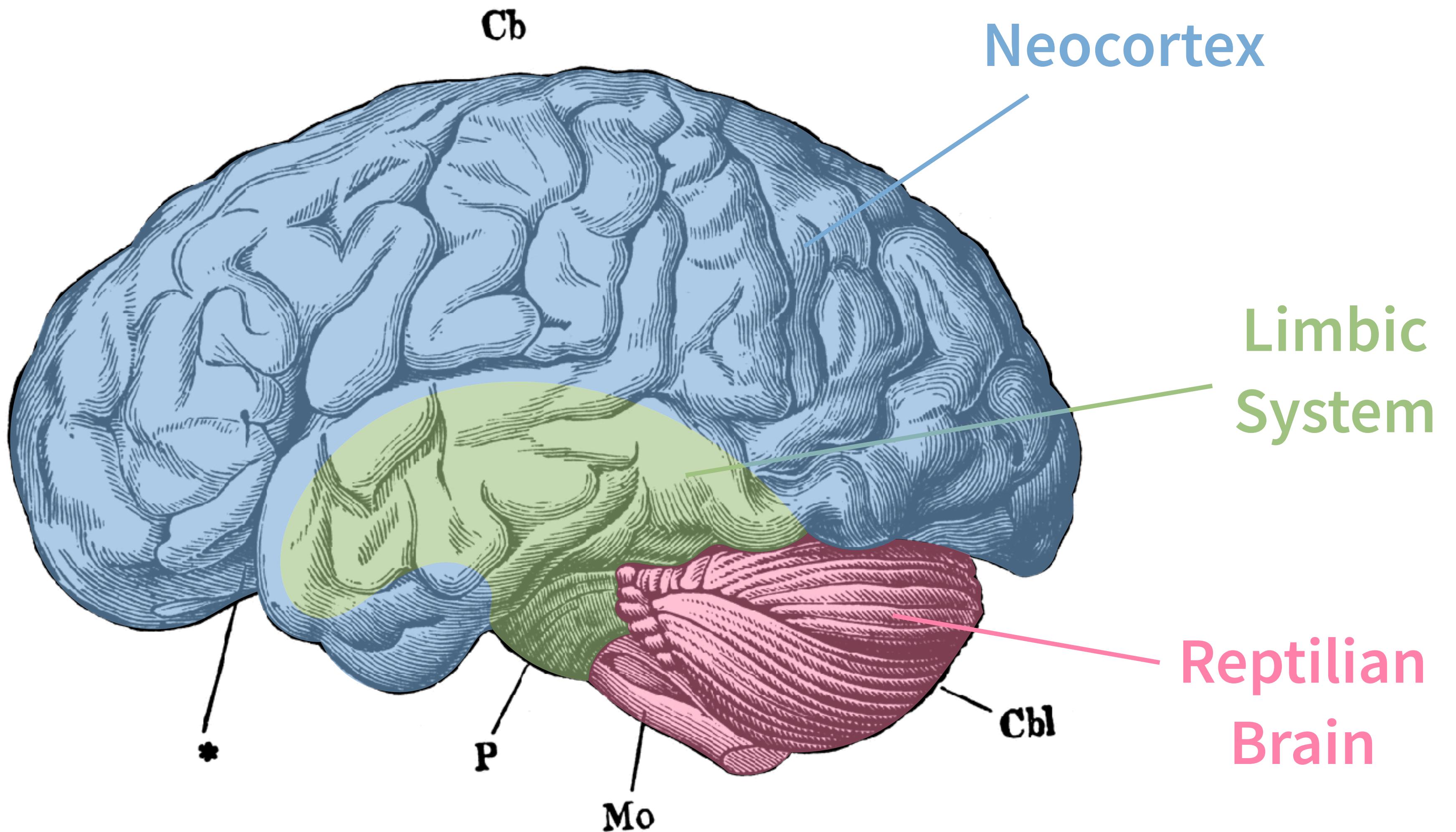
C++, FORTRAN, etc.



Human language

R - A computer language for scientists

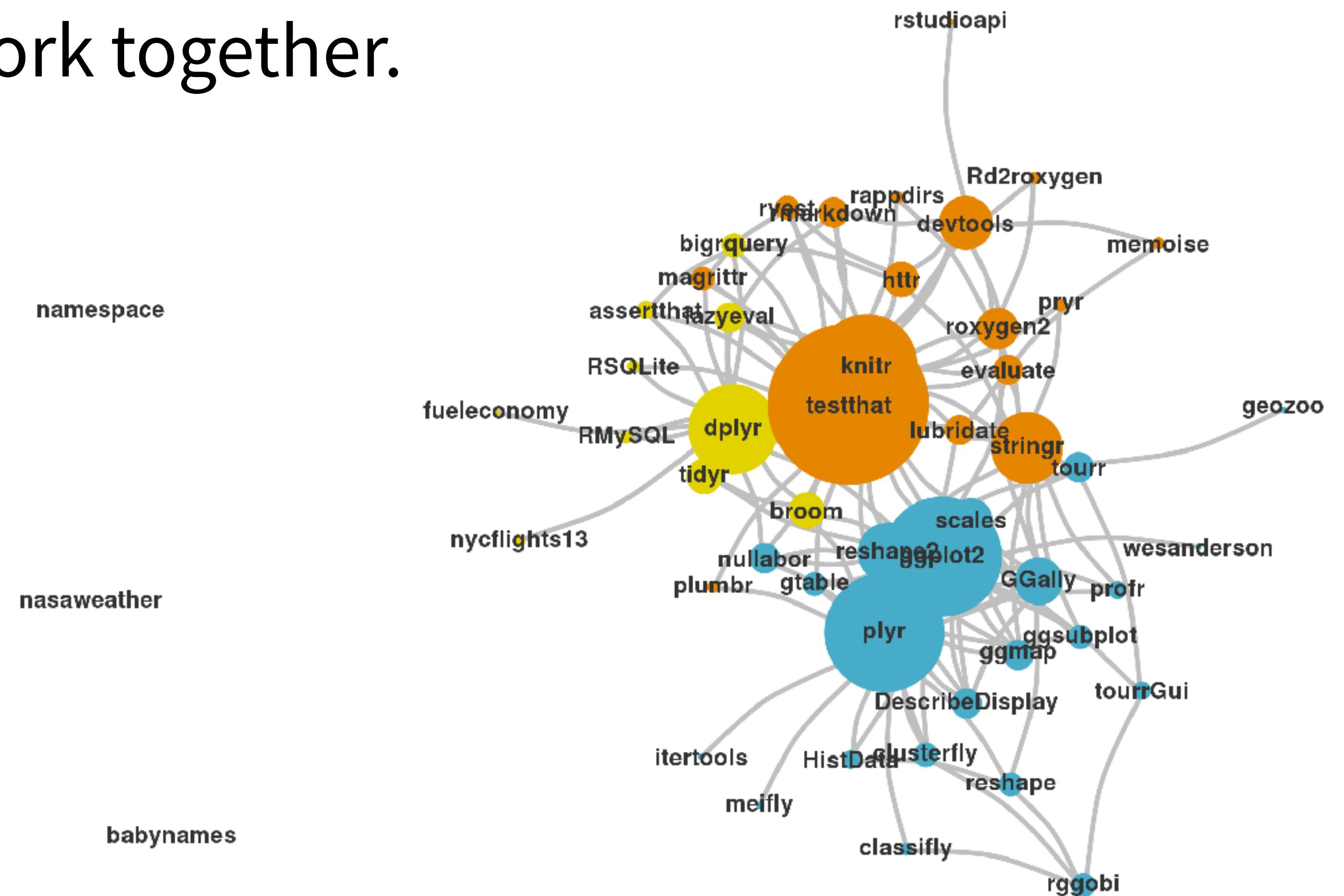






The Tidyverse

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 title bar "tidyverse website • tidyweb" and the user "Garrett". The address bar shows "tidyverse.org". The main content area displays the "The tidyverse" homepage, featuring a grid of hexagonal icons representing various R packages: dplyr, forcats, ggplot2, haven, lubridate, purrr, readr, readxl, stringr, tibble, tidy, and tidyverse. Below the grid, a text block 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." To the right of the main content, there is a sidebar titled "#tidyverse tweets" containing three recent tweets from users PJ B, Martin Monkman, and Hilary Robbins, each with a small profile picture and a link to their Twitter account.

The tidyverse

Components

dplyr
forcats
ggplot2
haven
lubridate
purrr
readr
readxl
stringr
tibble
tidy
tidyverse

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.

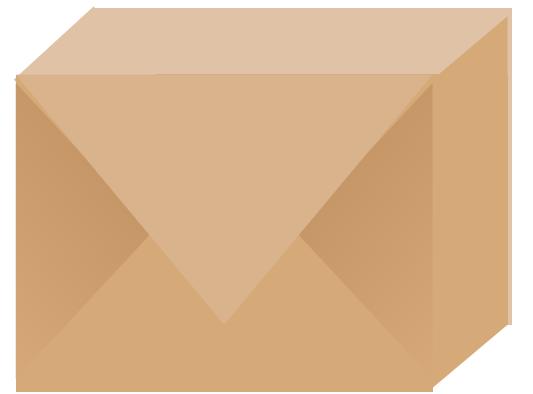
#tidyverse tweets

PJ B @ProfBforBargain I may be excommunicated for it, but readxl and lubridate are heretical packages that should not belong to the tidyverse. #rstats #tidyverse 27m

Martin Monkman @monkmanmh Essential reading as-is. I think I'll leave a paper copy on the lunch room table. #tidyverse #datascience Thanks @kwbroman & @kara_woo 2h

Hilary Robbins @hilaryarobbins Raise your hand if you constantly write broken #tidyverse code because of British vs. American spellings 🤪 #rstats #summarise 2h

tidyverse

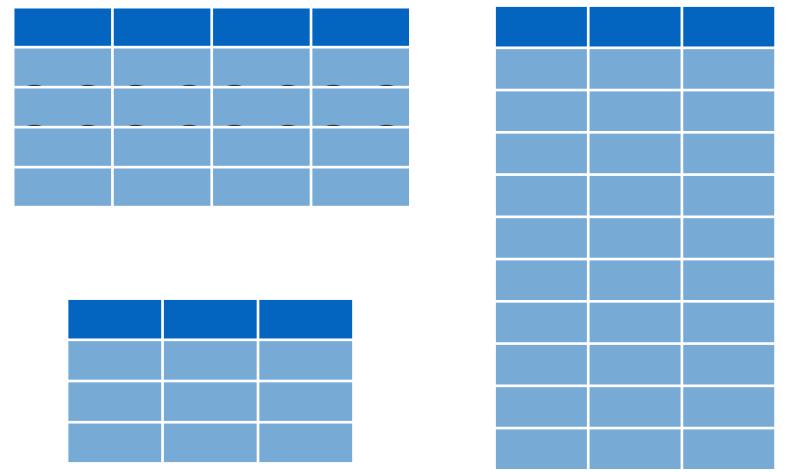


An R package that serves as a short cut for installing and loading the components of the tidyverse.

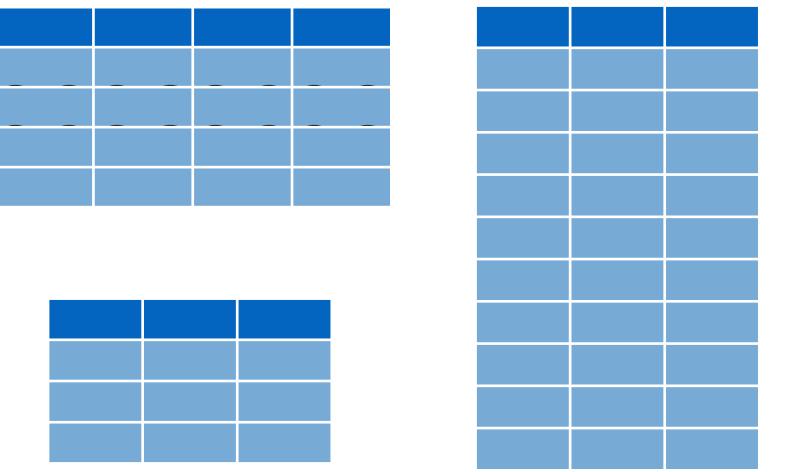
```
library("tidyverse")
```

R Packages

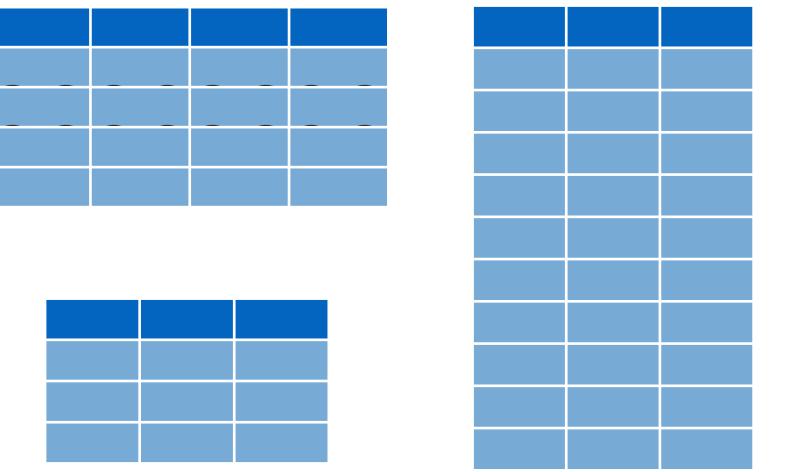
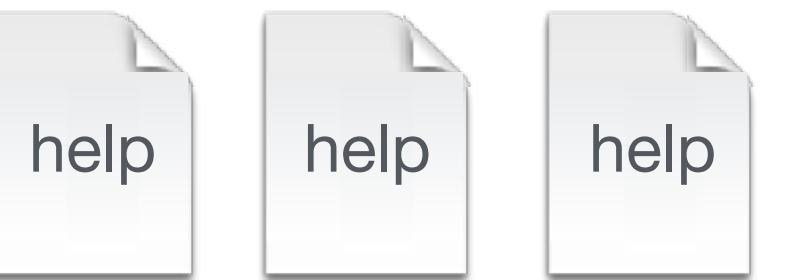




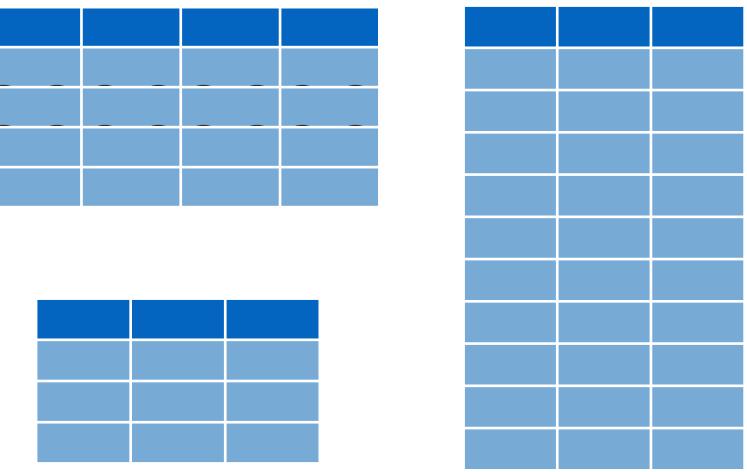
function1()
function2()
function3()
function4()



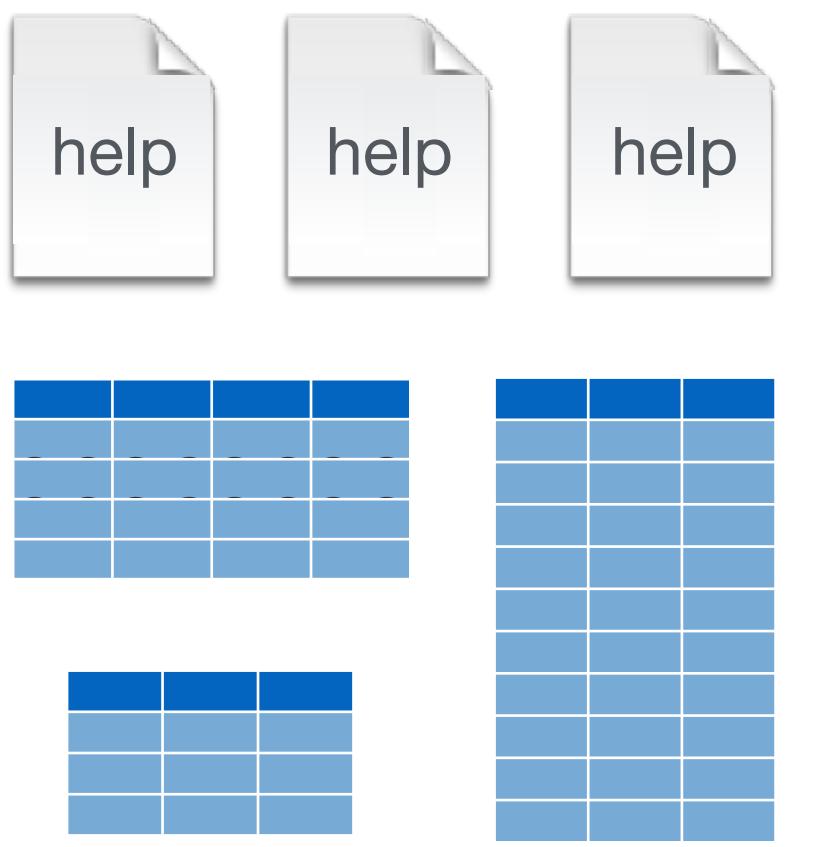
function5()
function6()
function7()
function8()



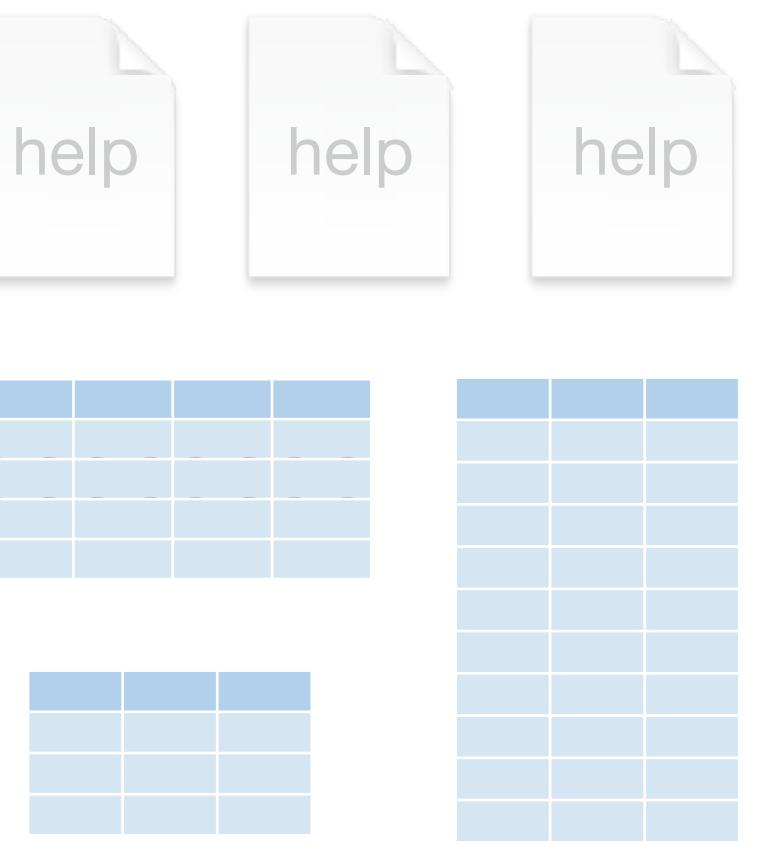
function9()
functionA()
functionB()
functionC()



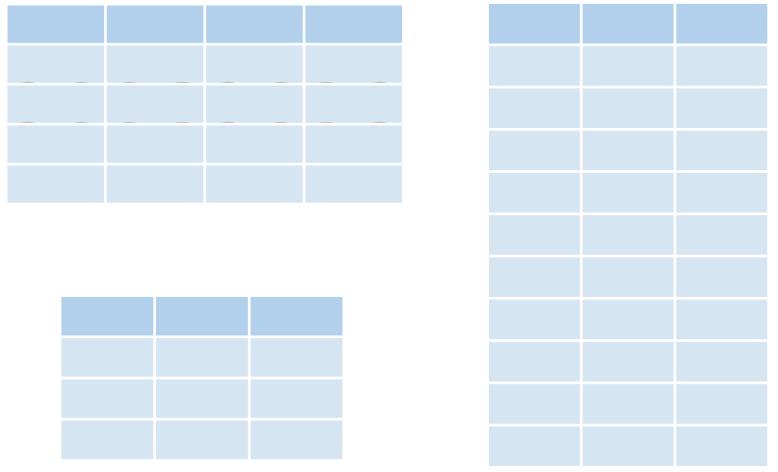
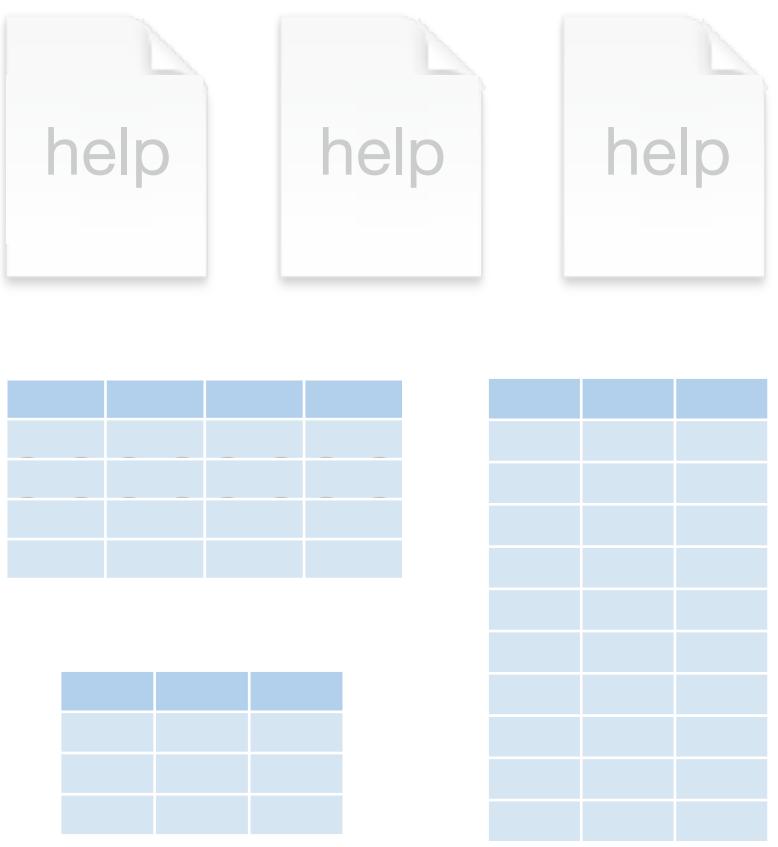
functionD()
functionE()
functionF()
functionG()



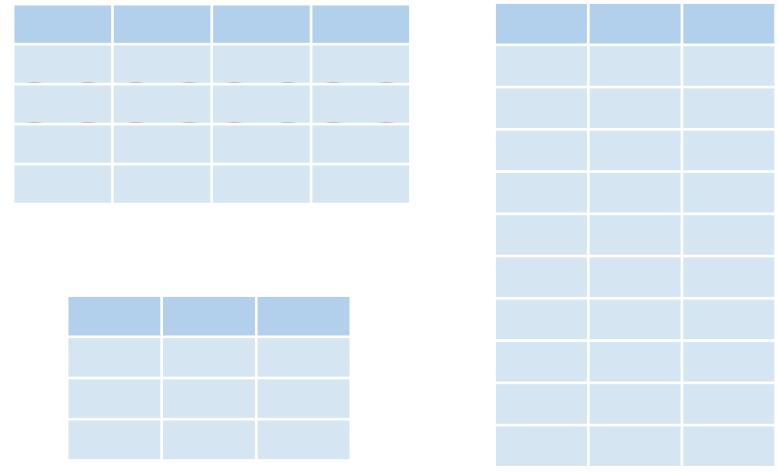
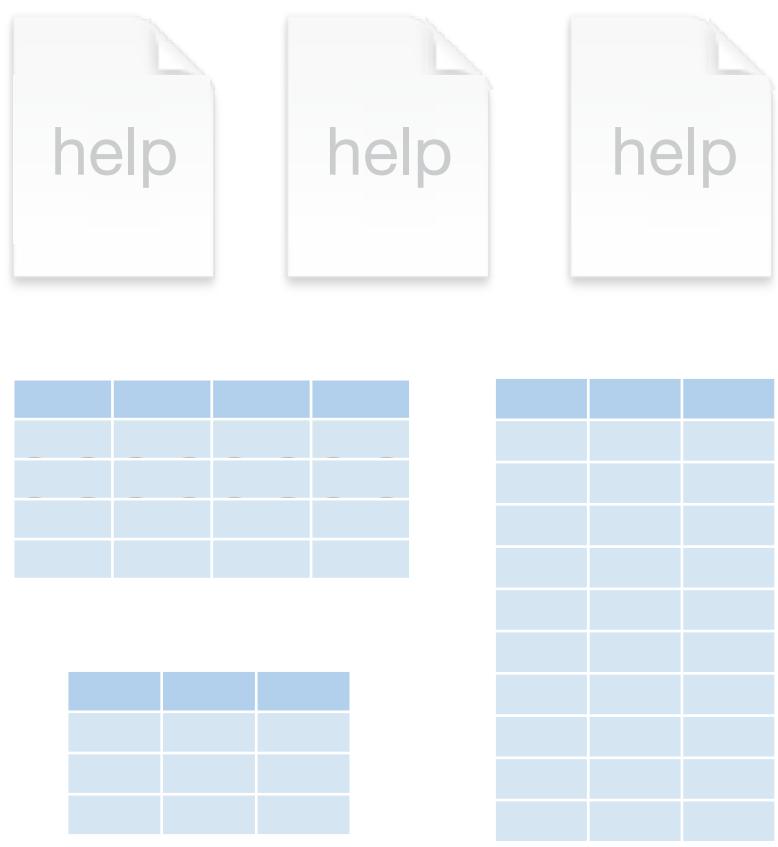
function1()
function2()
function3()
function4()



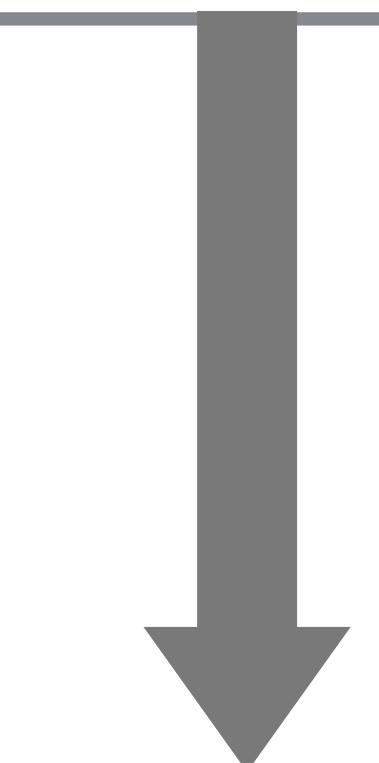
function5()
function6()
function7()
function8()



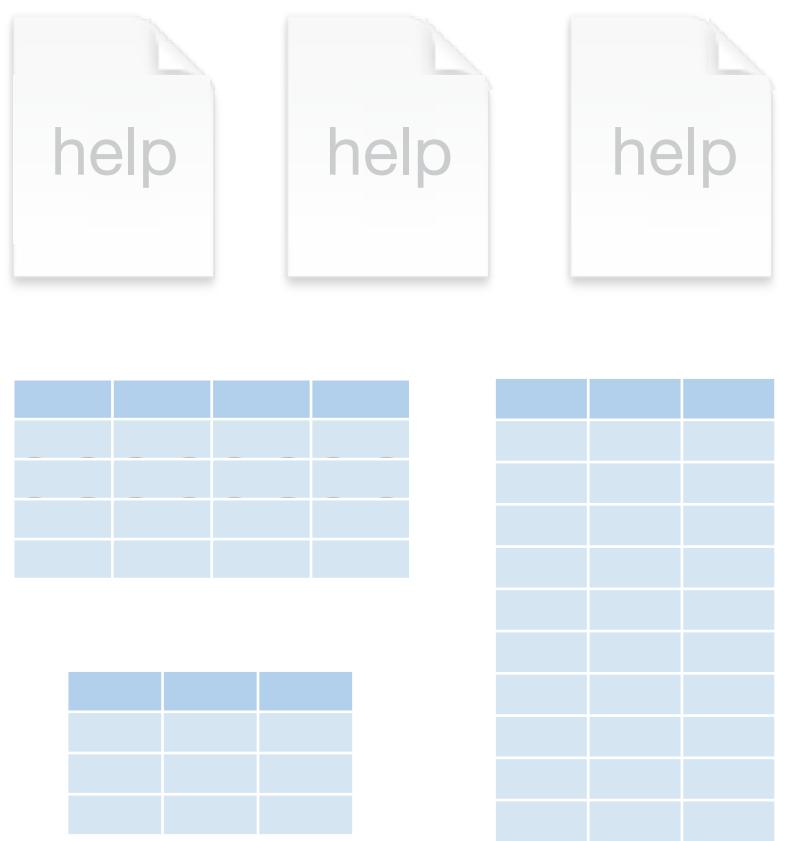
function9()
functionA()
functionB()
functionC()



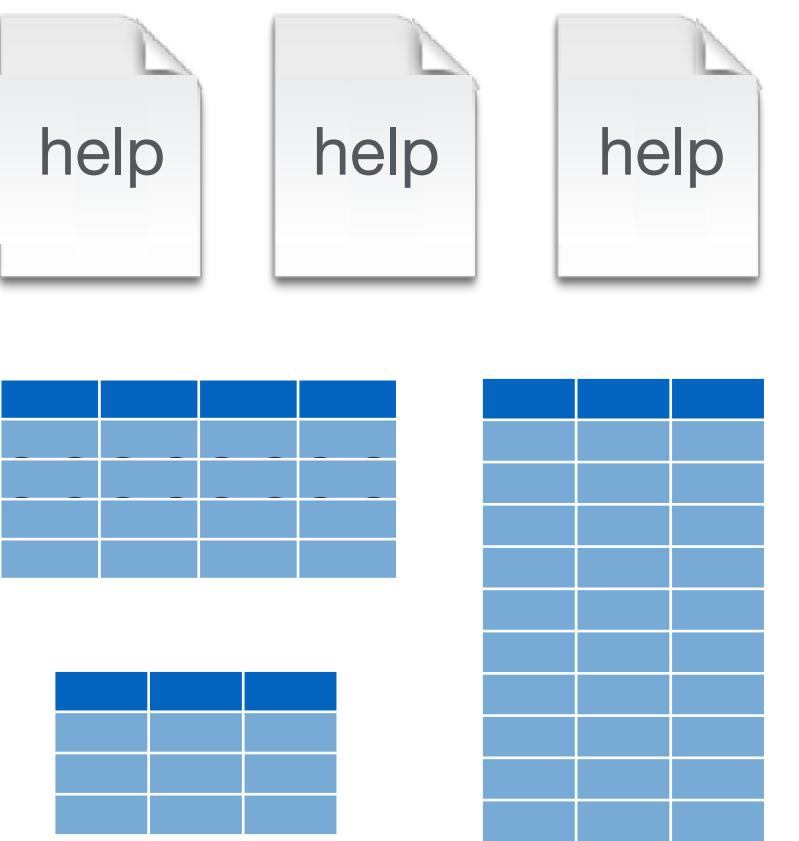
functionD()
functionE()
functionF()
functionG()



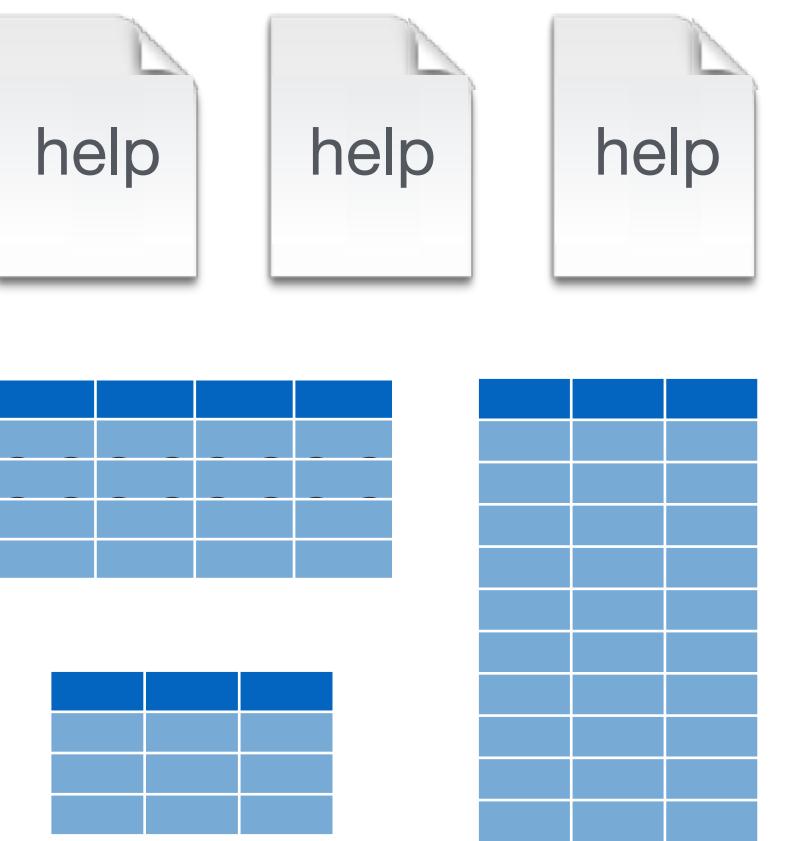
Base R



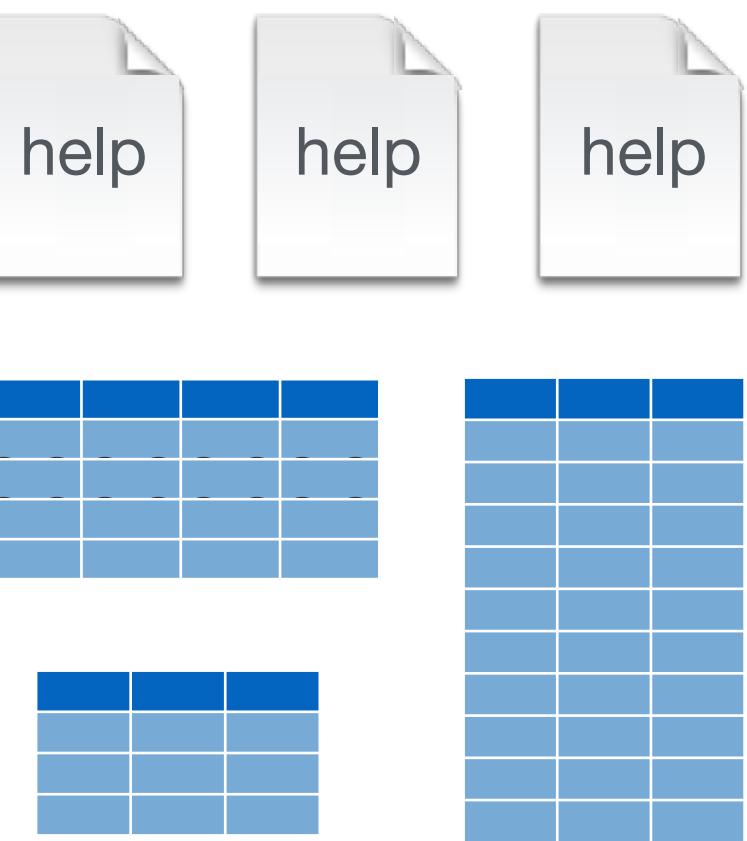
function1()
function2()
function3()
function4()



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



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



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

Base R

R Packages

Using packages

1

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

Downloads files to computer

1 x per computer

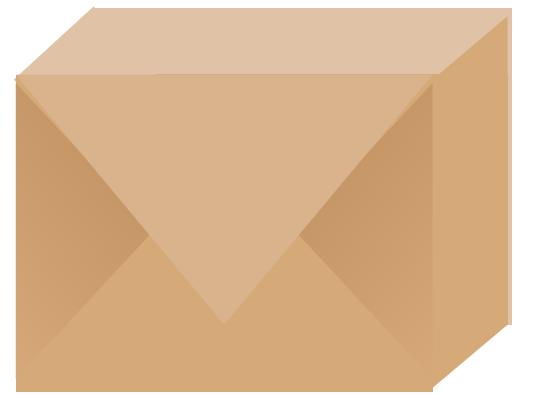
2

```
library("foo")
```

Loads package

1 x per R Session

tidyverse



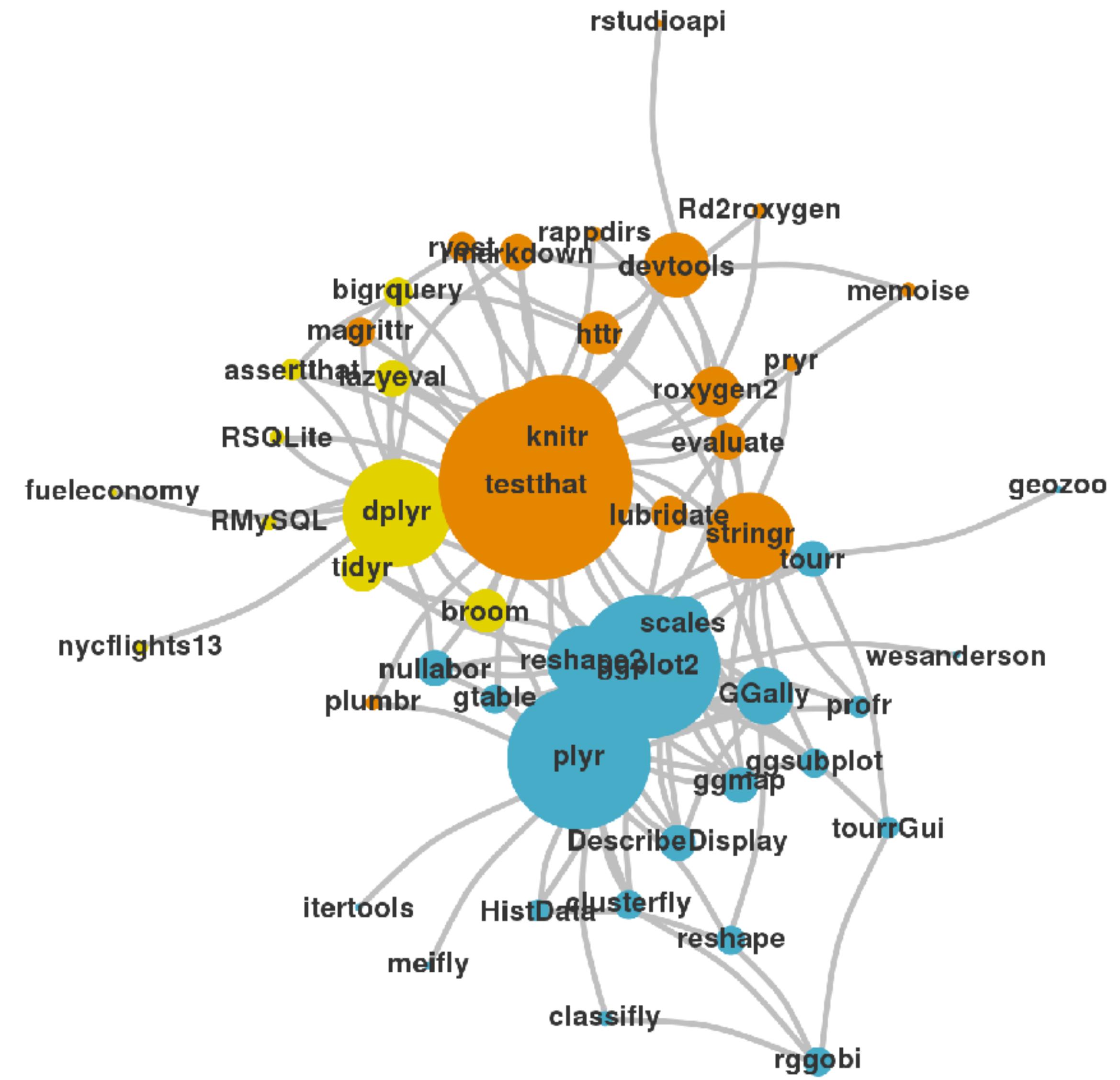
An R package that serves as a short cut for installing and loading the components of the tidyverse.

```
library("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")
```

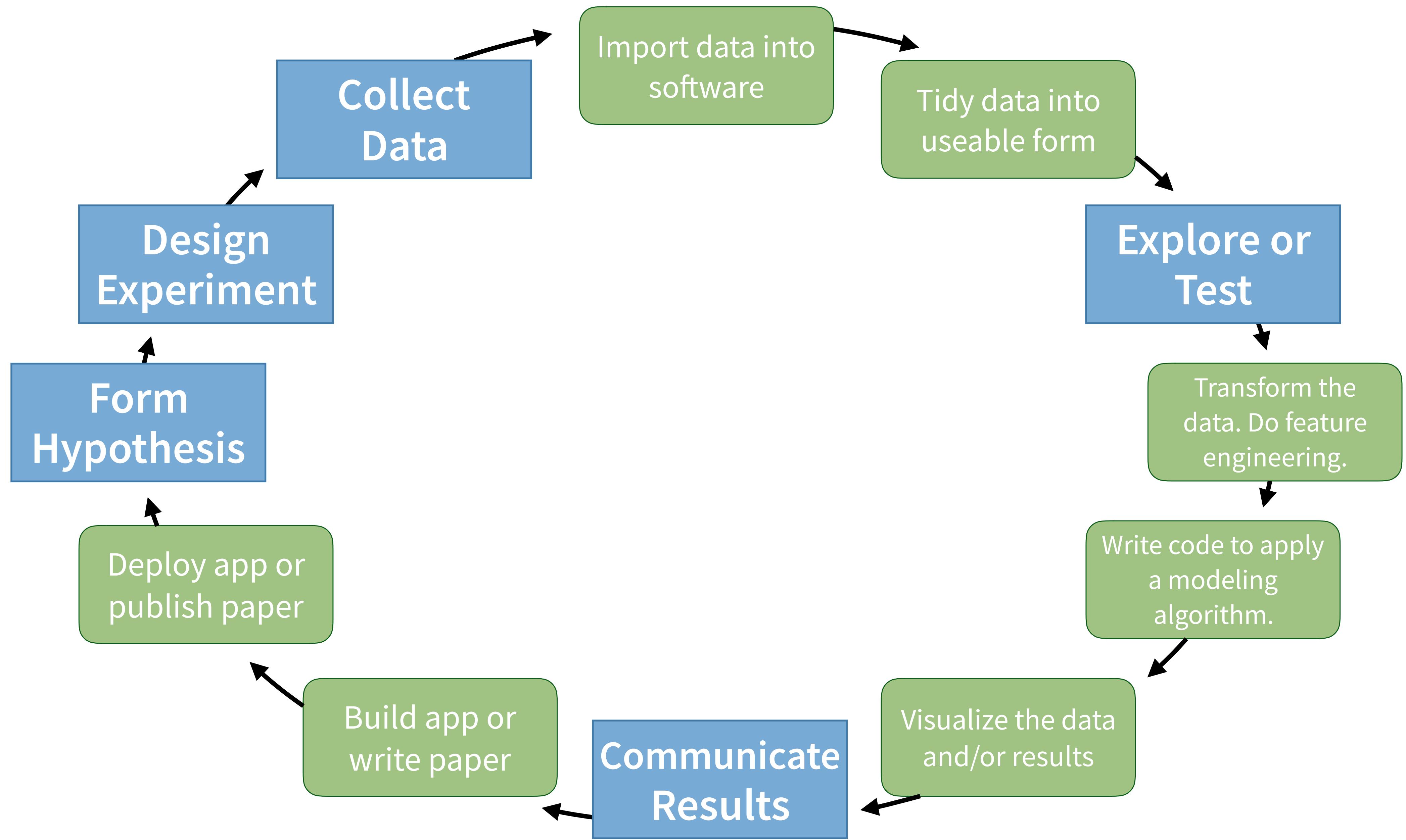
```
library("tidyverse")
```

does the equivalent of

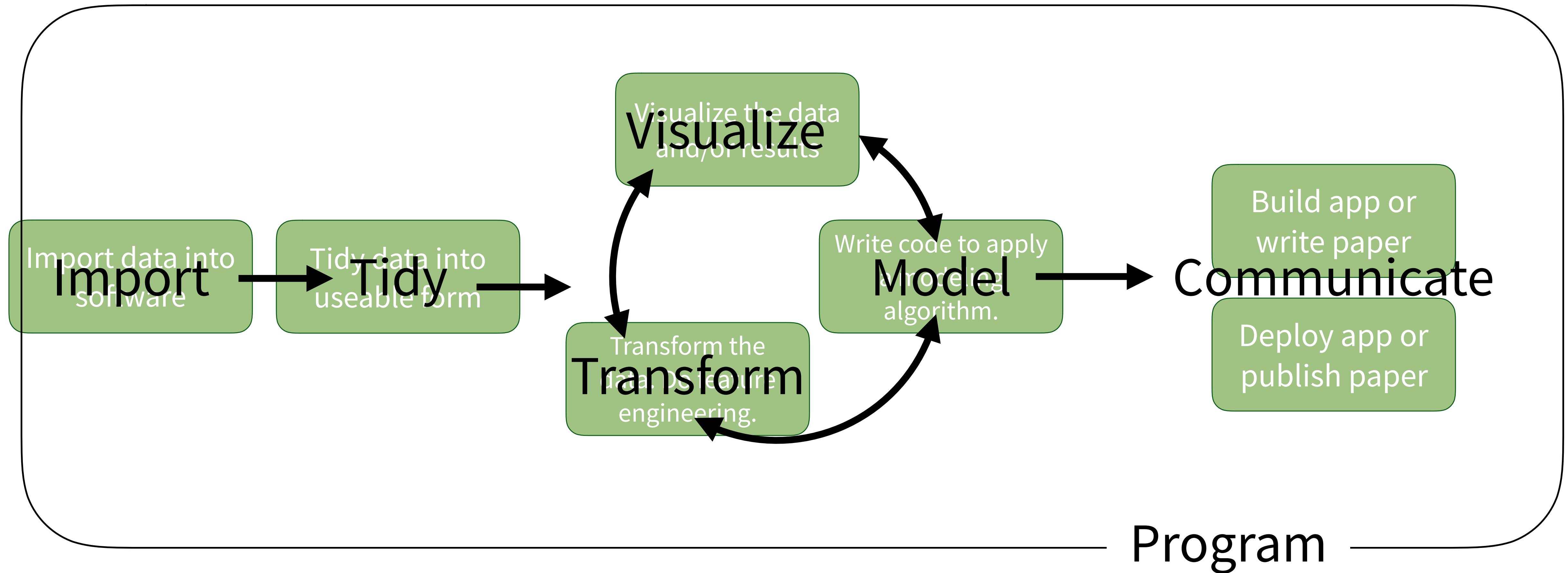
```
library("ggplot2")
library("dplyr")
library("tidyr")
library("readr")
library("purrr")
library("tibble")
```

Outro

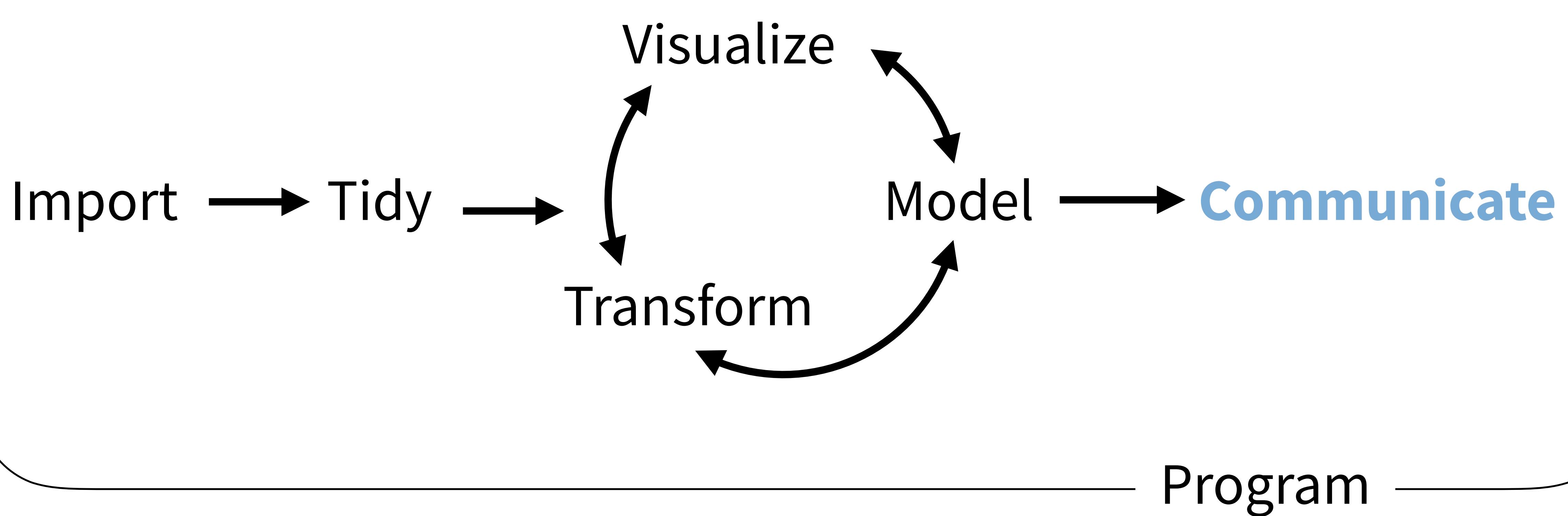




(Applied) Data Science



(Applied) Data Science



R Notebooks

A faint watermark of the R logo is visible in the bottom right corner, consisting of a circular arrow and the letters "R".

Your Turn

Open **00-Introduction.Rmd**. Read through the notebook and do everything it tells you to do.



R Notebooks

An authoring format for Data Science.

The screenshot shows the RStudio interface with an R Notebook open. The notebook file is titled "R-Notebook.Rmd". 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 code editor has three callout boxes with arrows pointing to specific elements:

- A grey callout box points to the green "Run All" button at the top right of the code editor area, containing the text: "Click to run all code chunks above".
- A grey callout box points to the green "Run" button next to the code chunk 9, containing the text: "Click to run code in chunk".
- A dark grey callout box points to the output pane below the code editor, containing the text: "Code result".

The output pane displays the result of running the code in chunk 10:

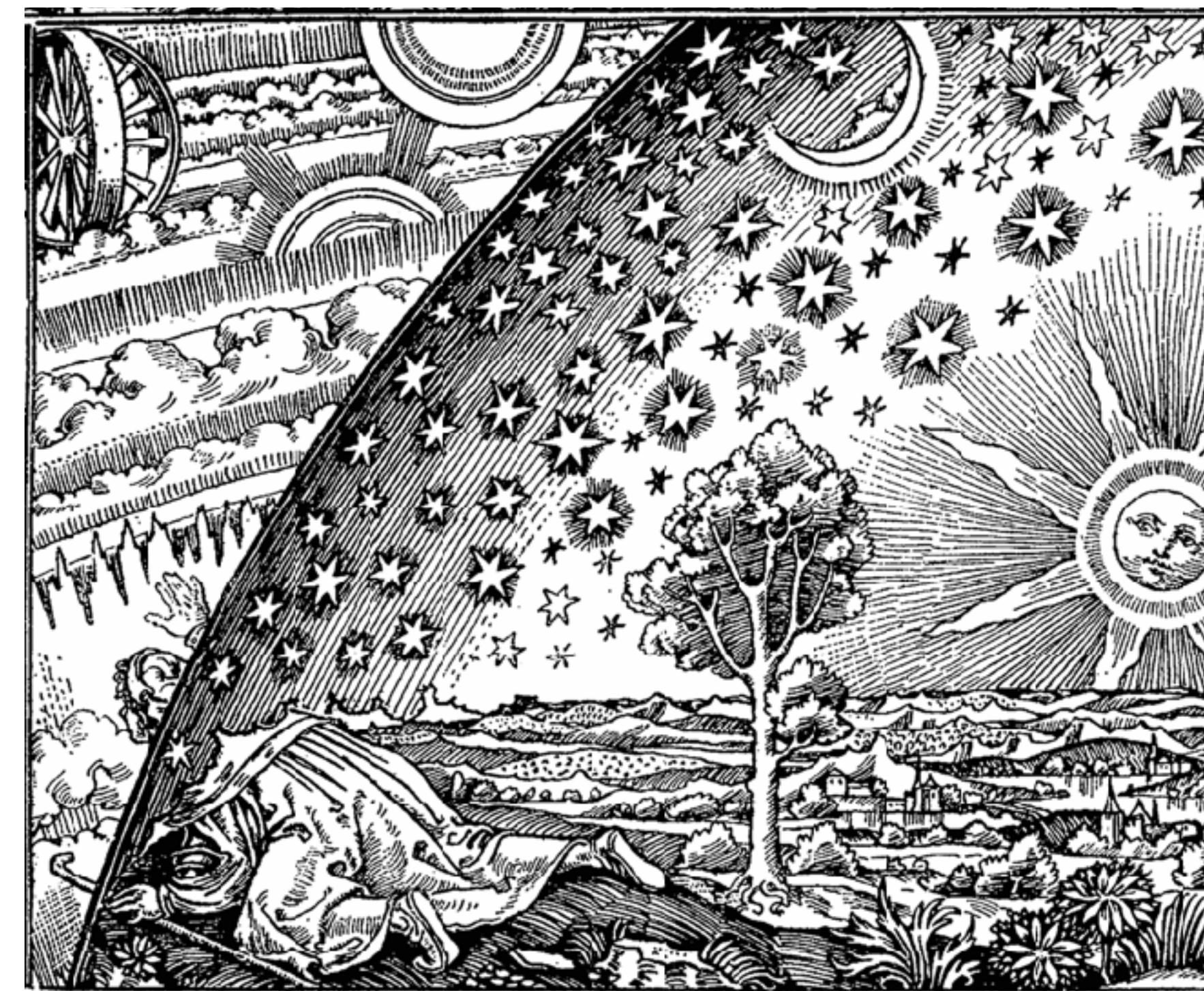
```
[1] -1.2 1.0 -0.5 0.9 -0.6 -1.1 -1.5
```

Your Turn

Open 01-Visualize-Data.Rmd.



Master the Tidyverse



Garrett Grolemund

Data Scientist, Educator

January 2017

RStudio