

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

TA

Day 1

Introduction and
Visualize Data

9:00 - 10:45

Morning Break

10:45 - 11:00

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

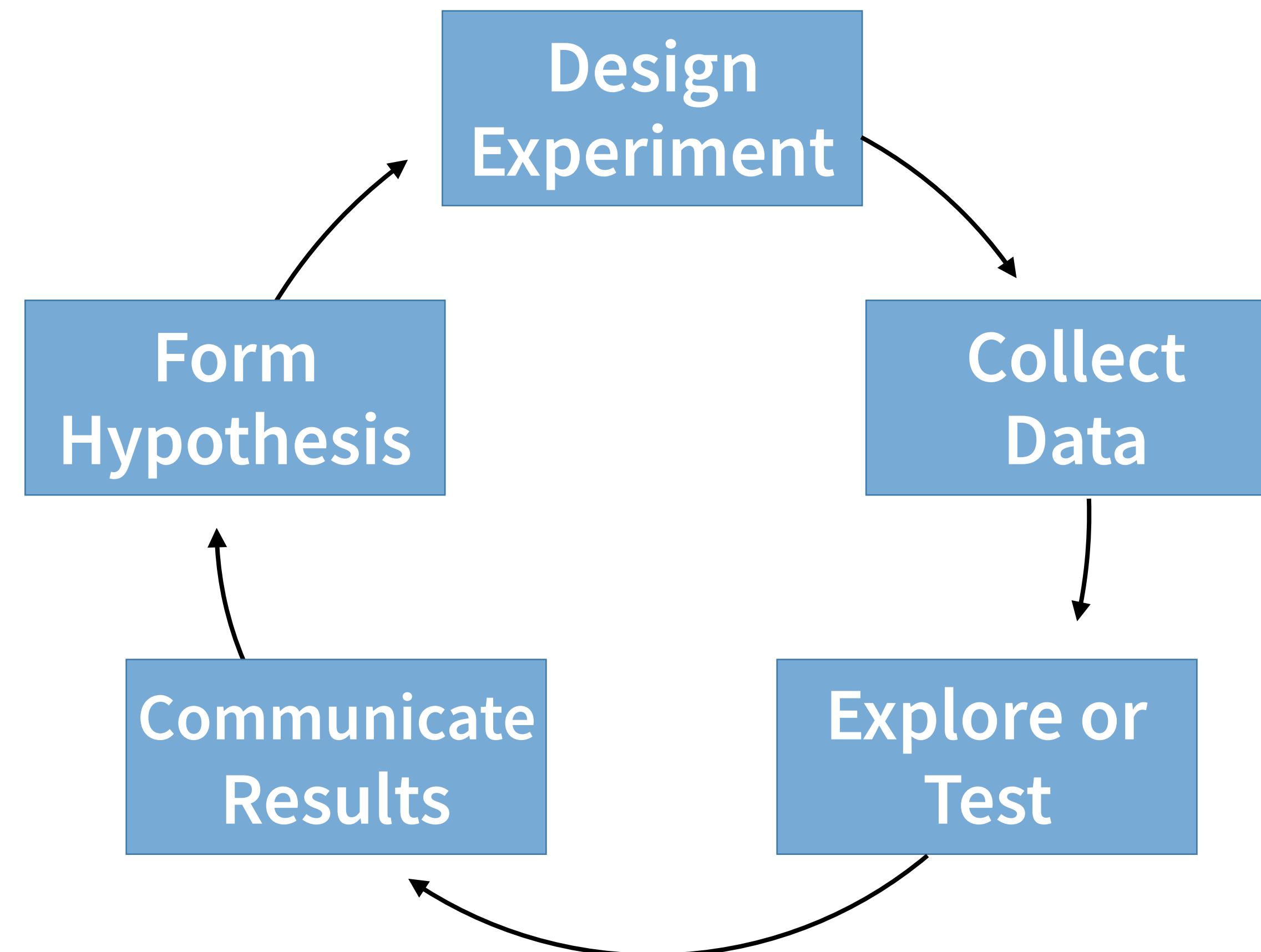
Your Turn

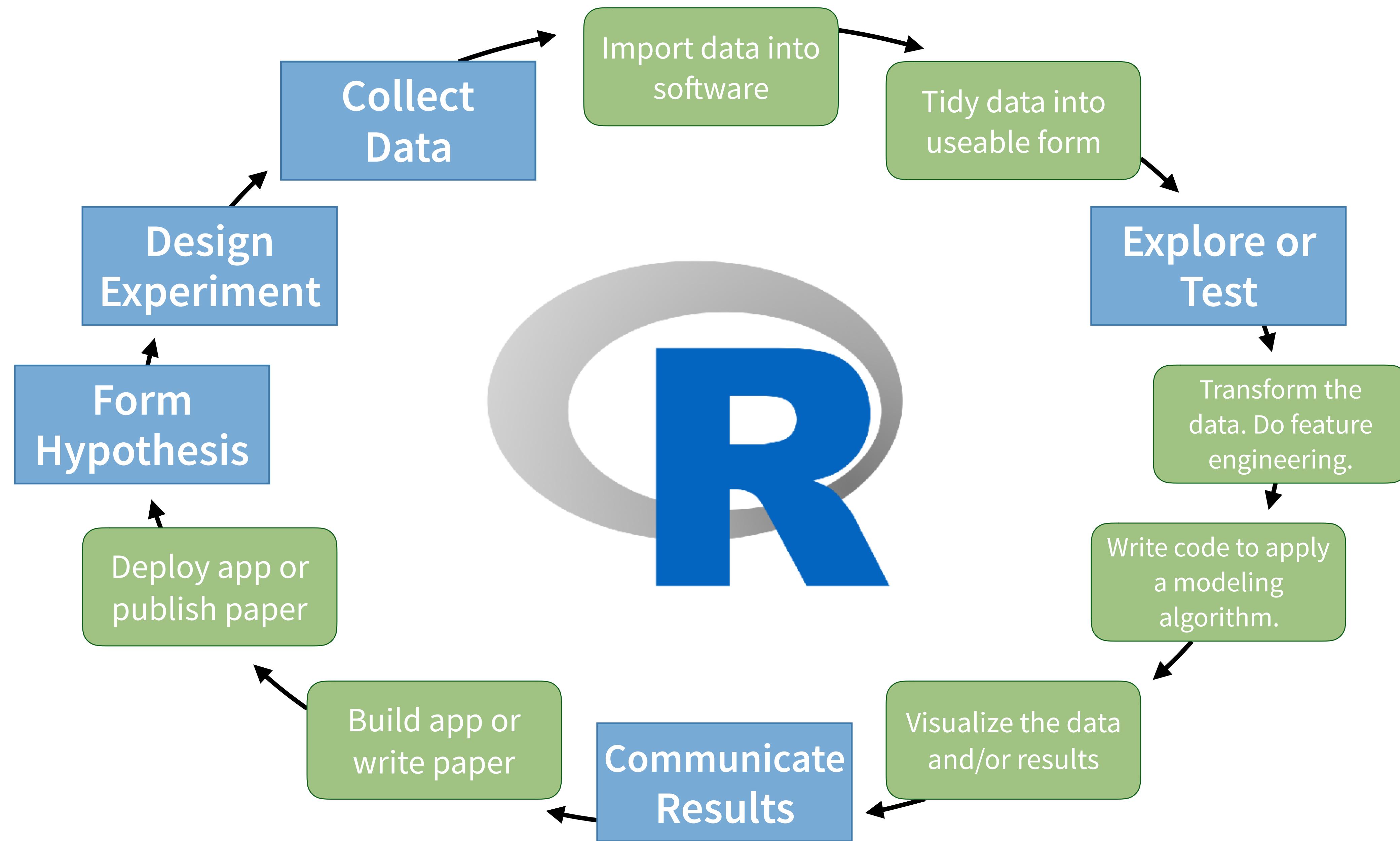
Form groups of 2-4 people. Introduce yourself to your group members. Tell them:

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

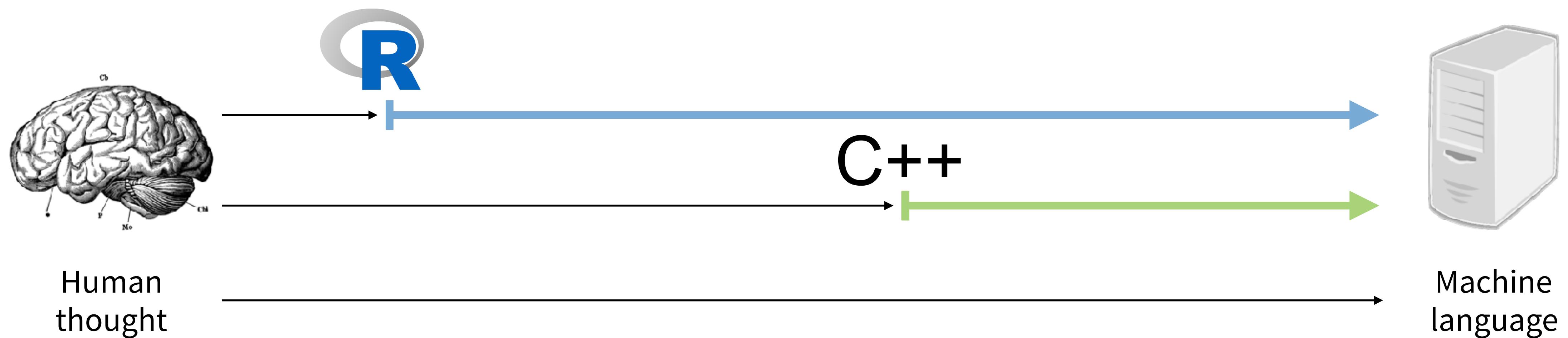


"Data Science"

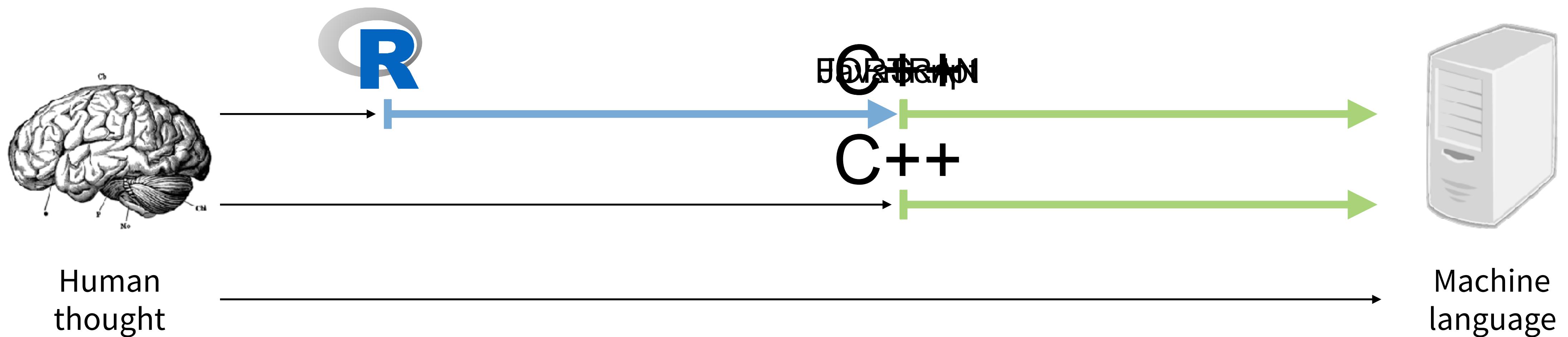




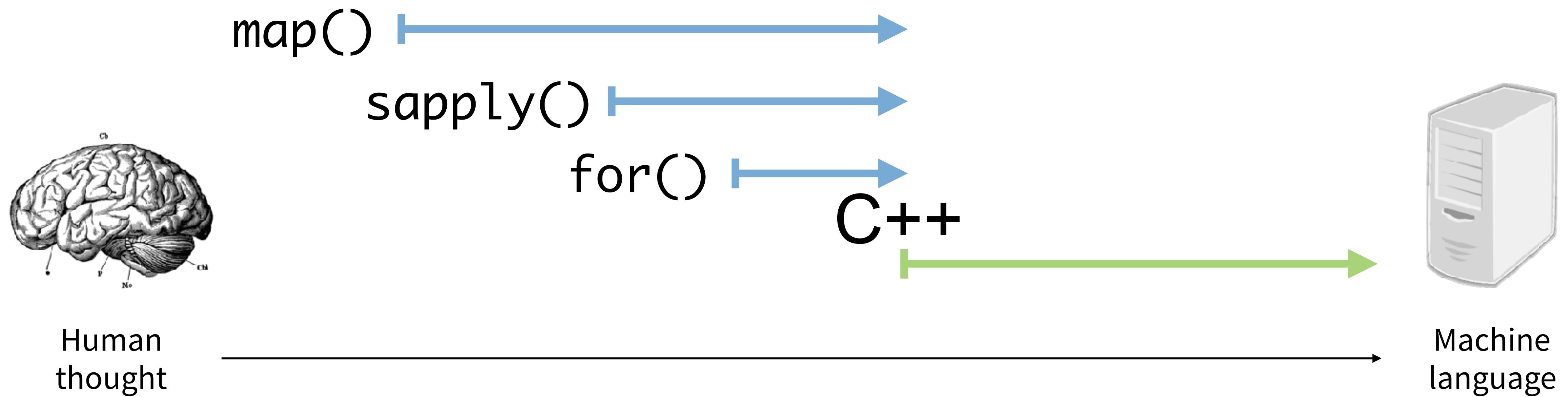
R - A computer language for scientists

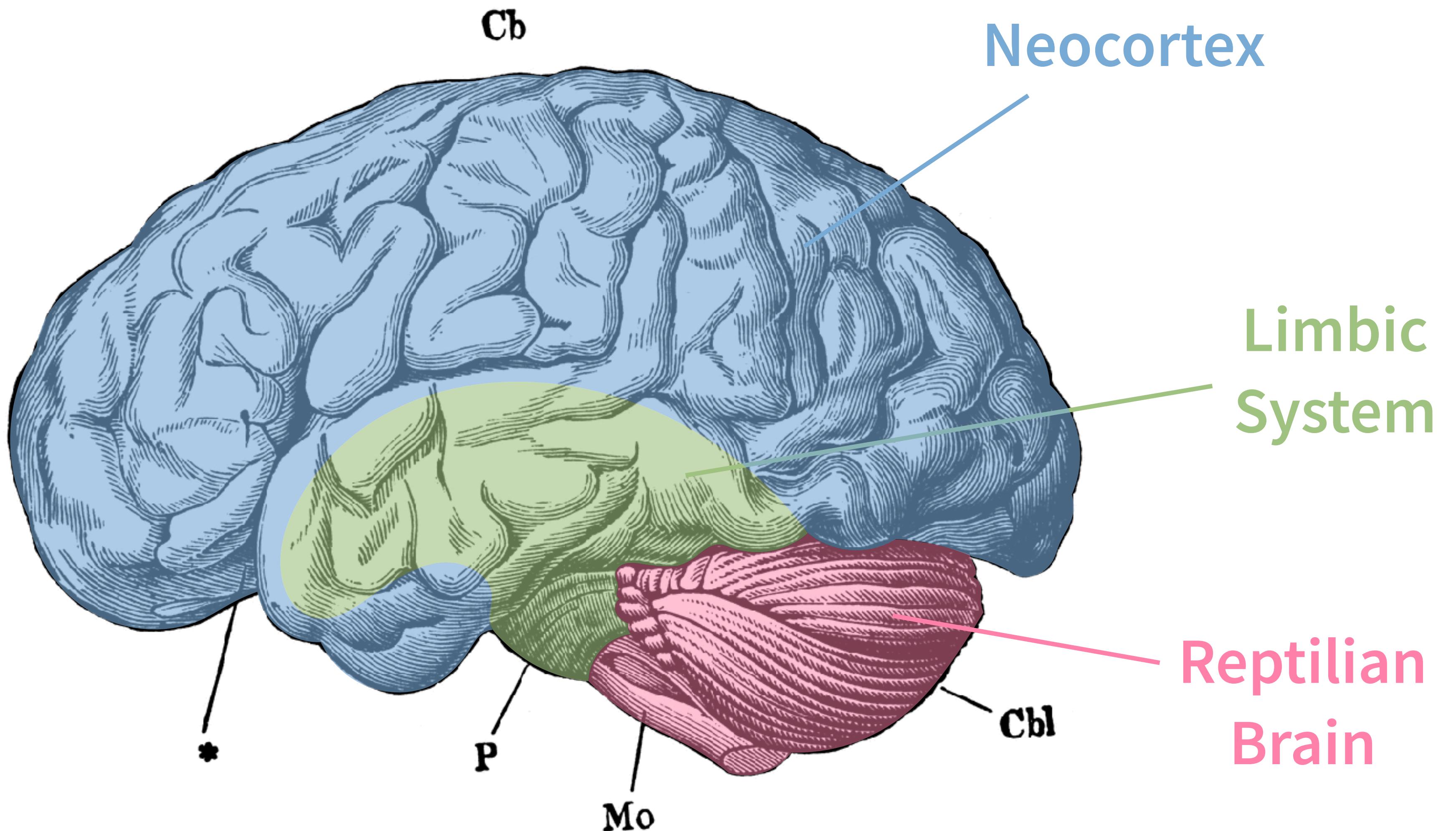


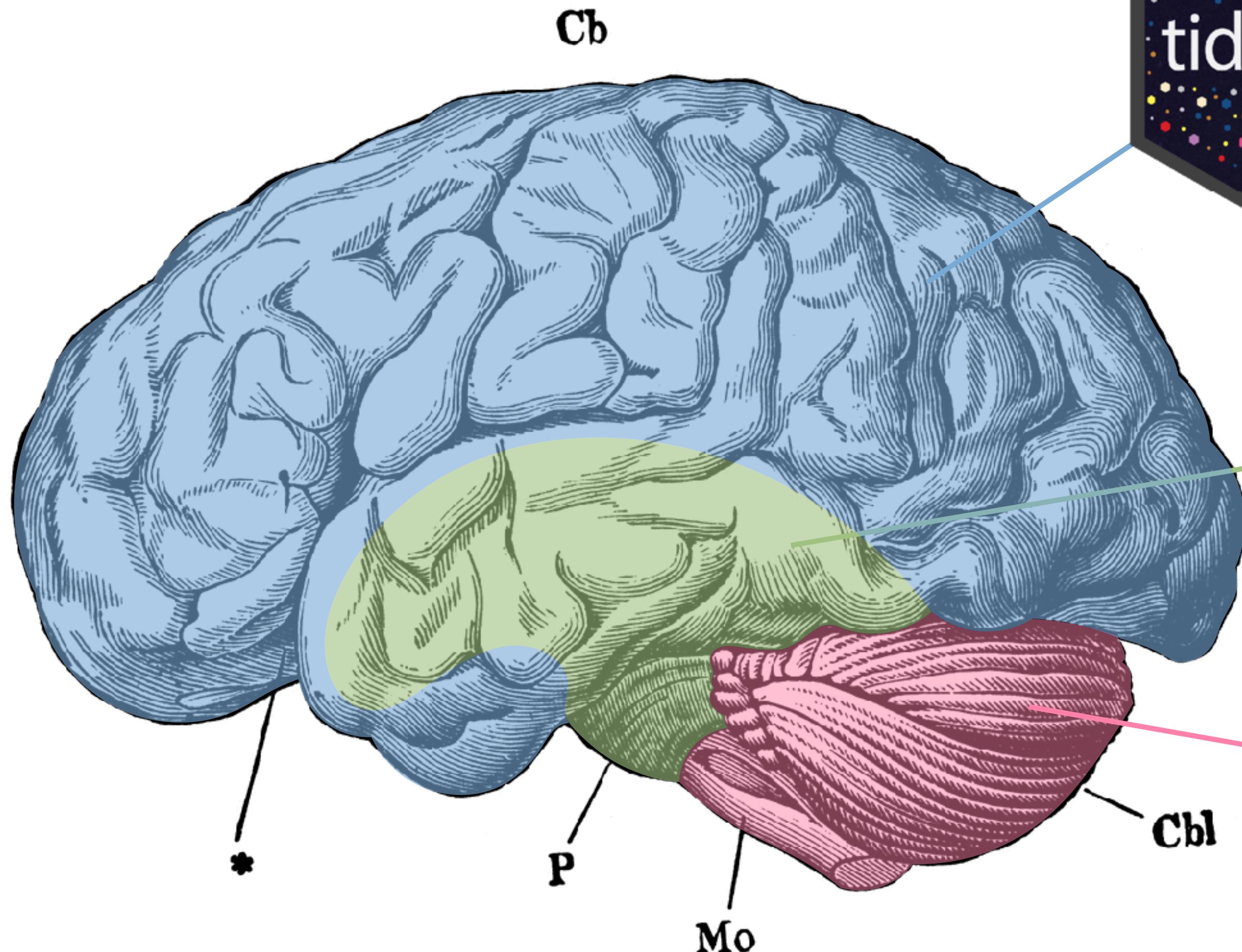
R - A computer language for scientists



R - A computer language for scientists





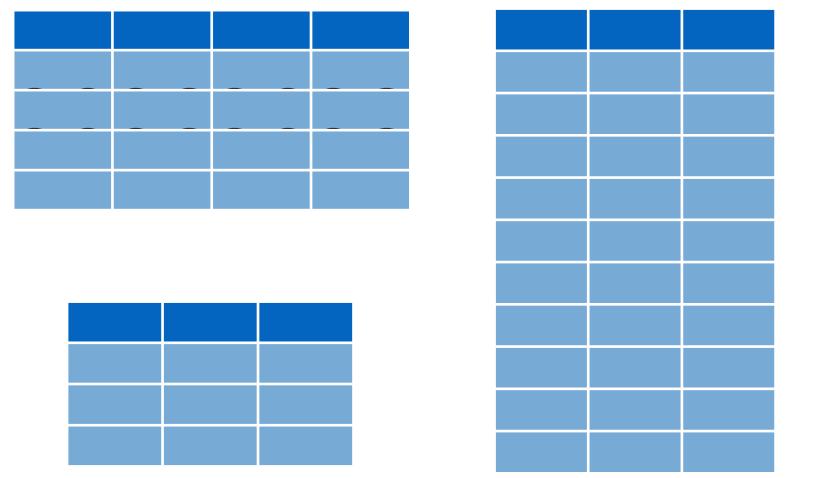


Limbic
System

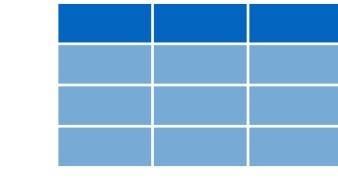
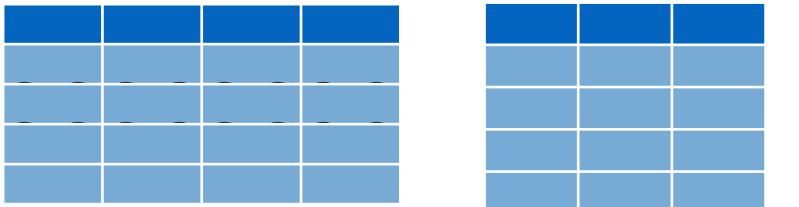
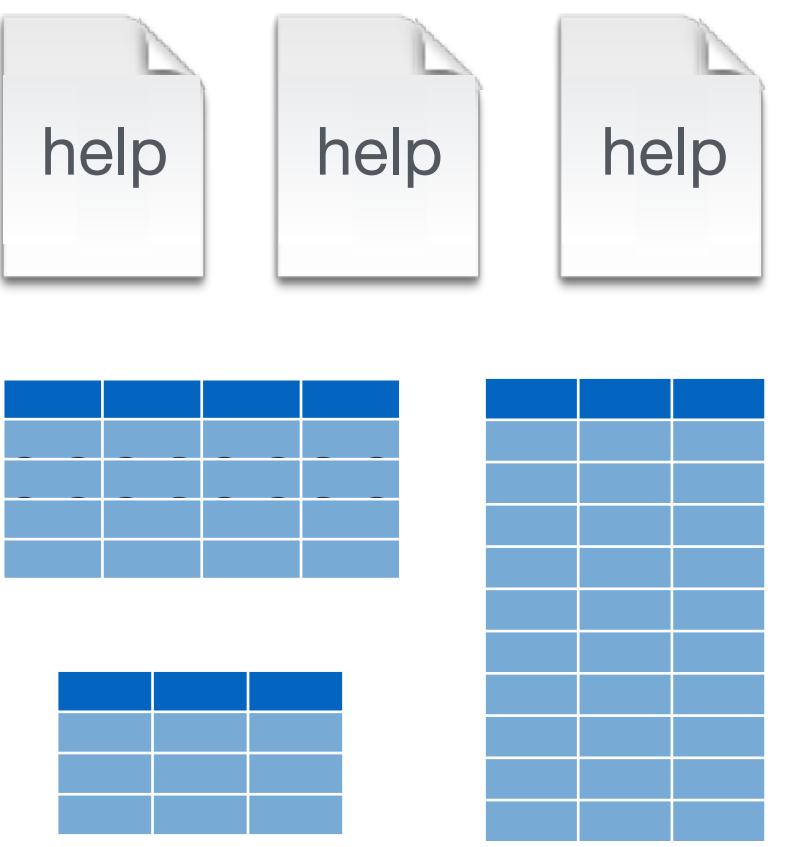
Reptilian
Brain

R Packages

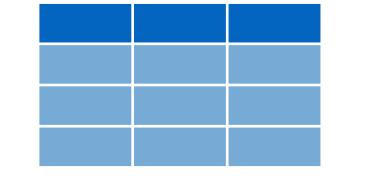
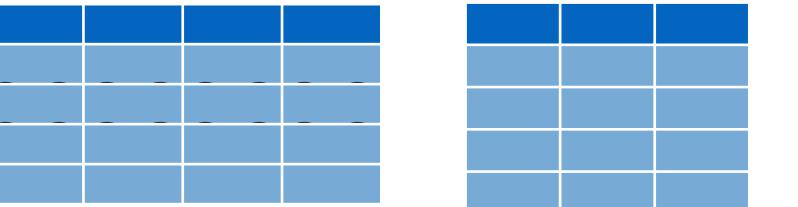
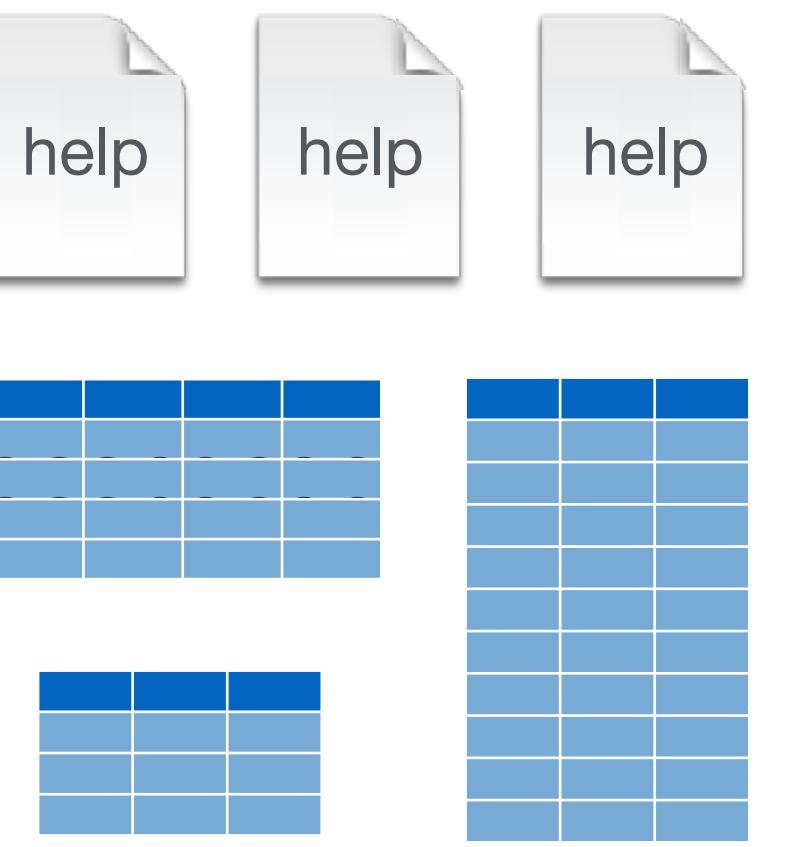




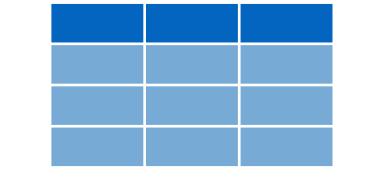
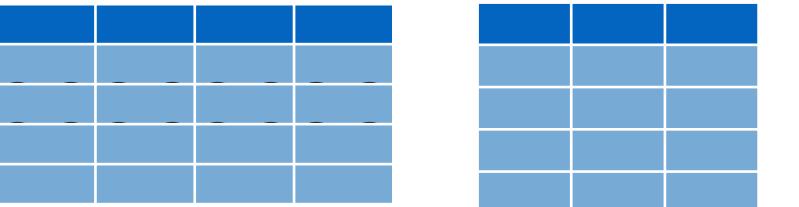
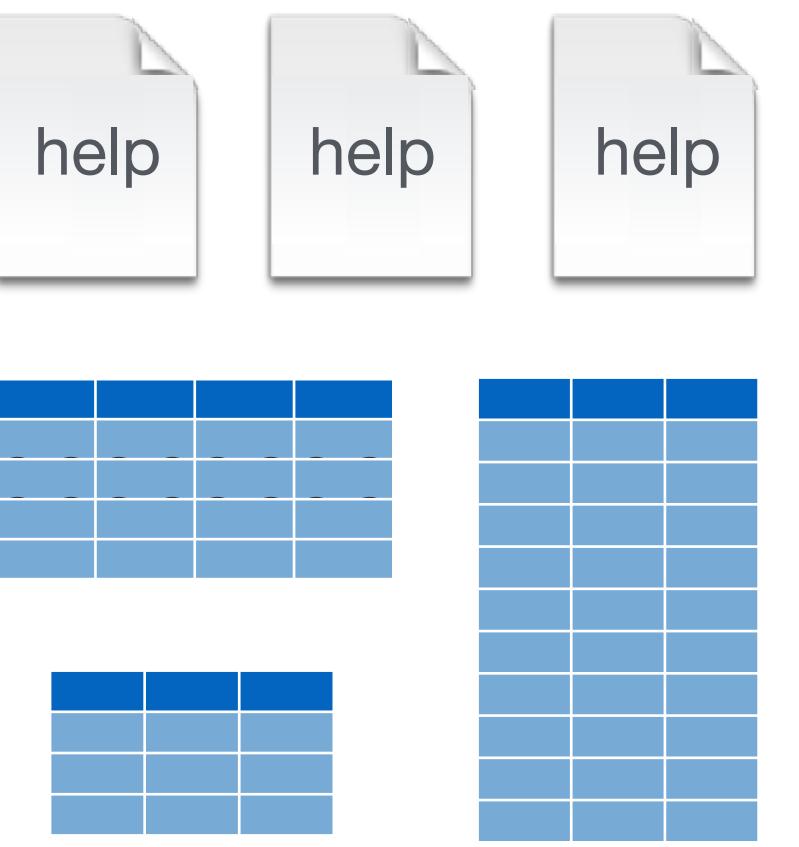
function1()
function2()
function3()
function4()



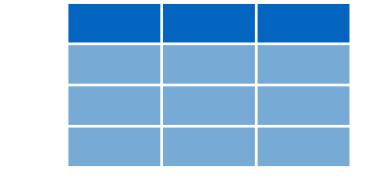
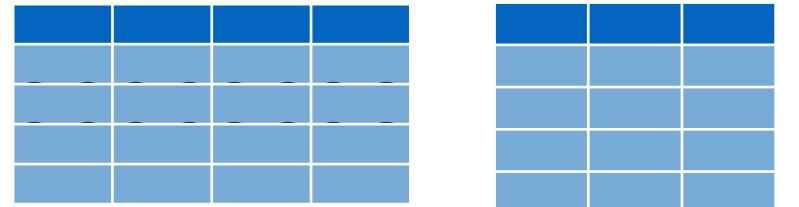
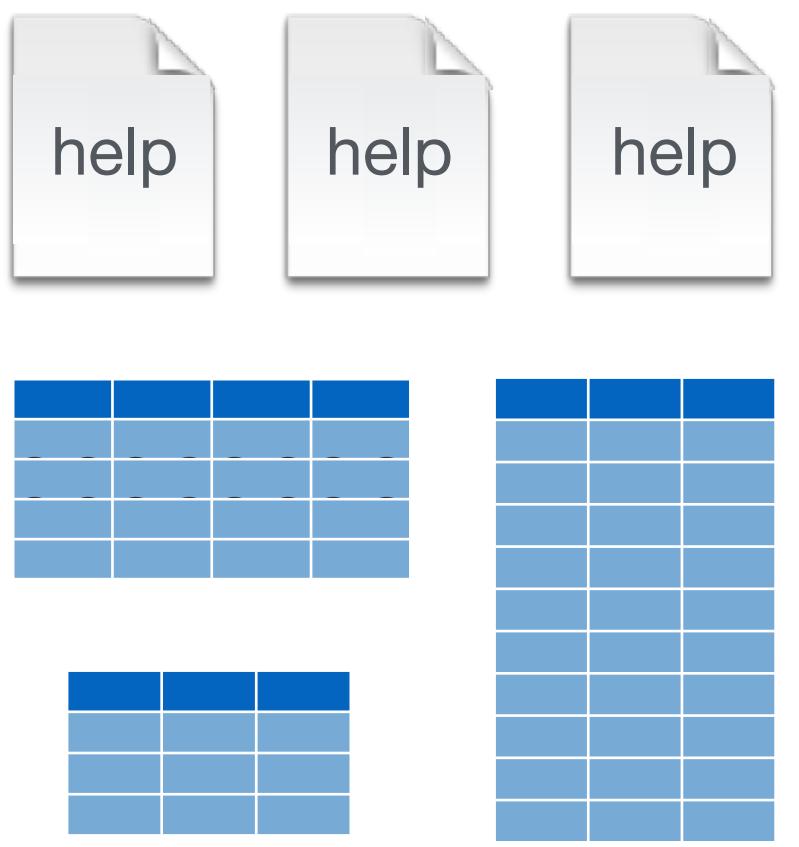
function1()
function2()
function3()
function4()



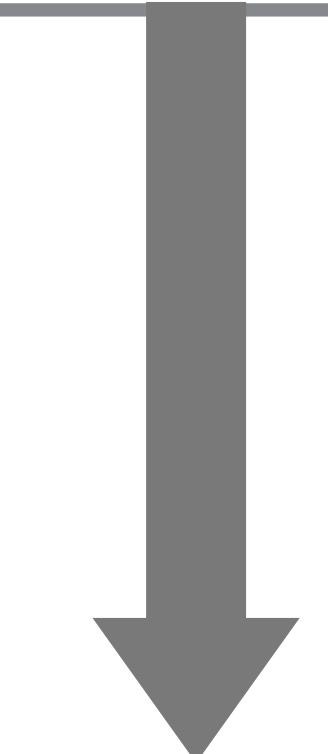
function5()
function6()
function7()
function8()



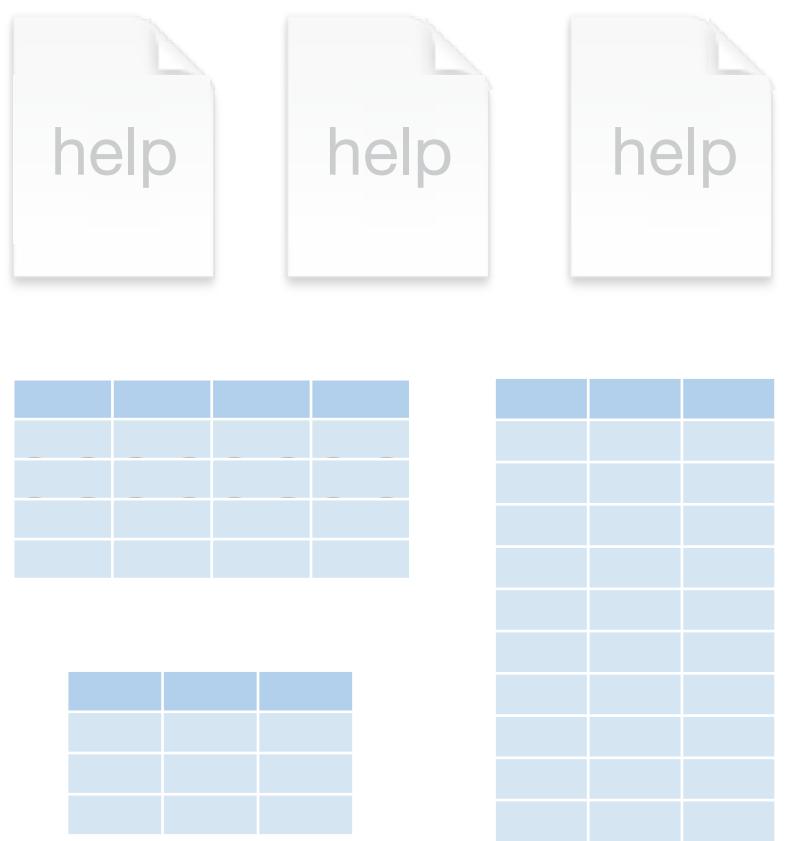
function9()
functionA()
functionB()
functionC()



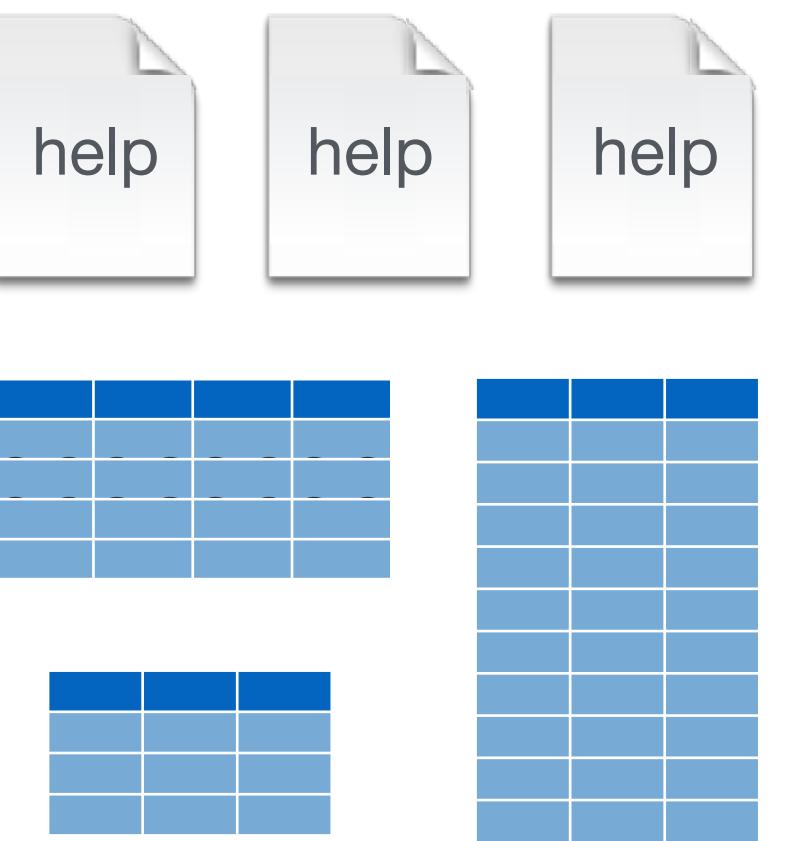
functionD()
functionE()
functionF()
functionG()



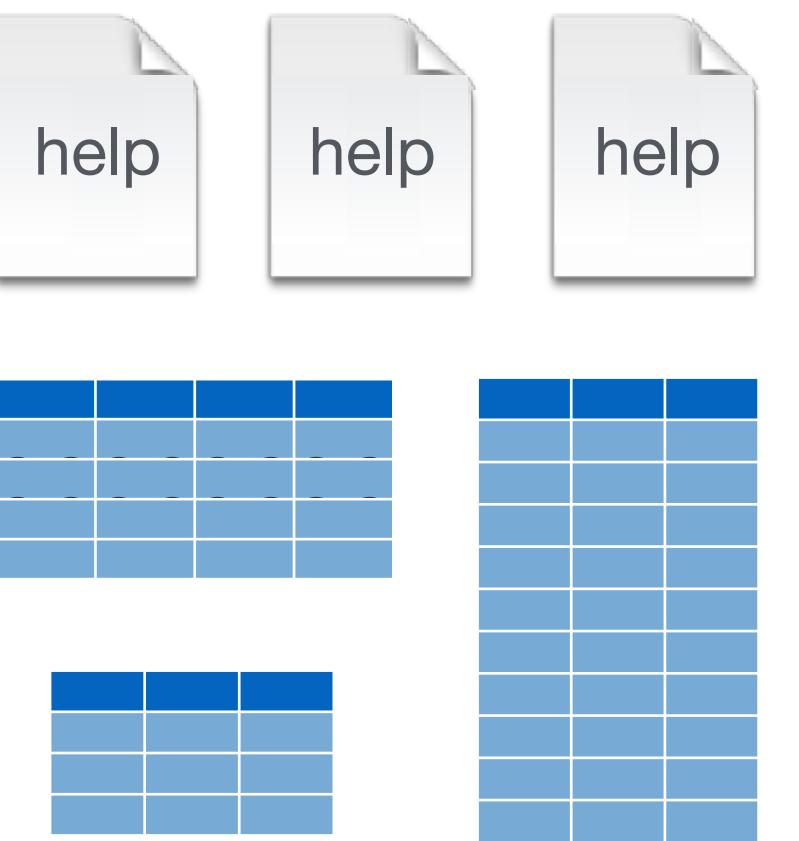
Base R



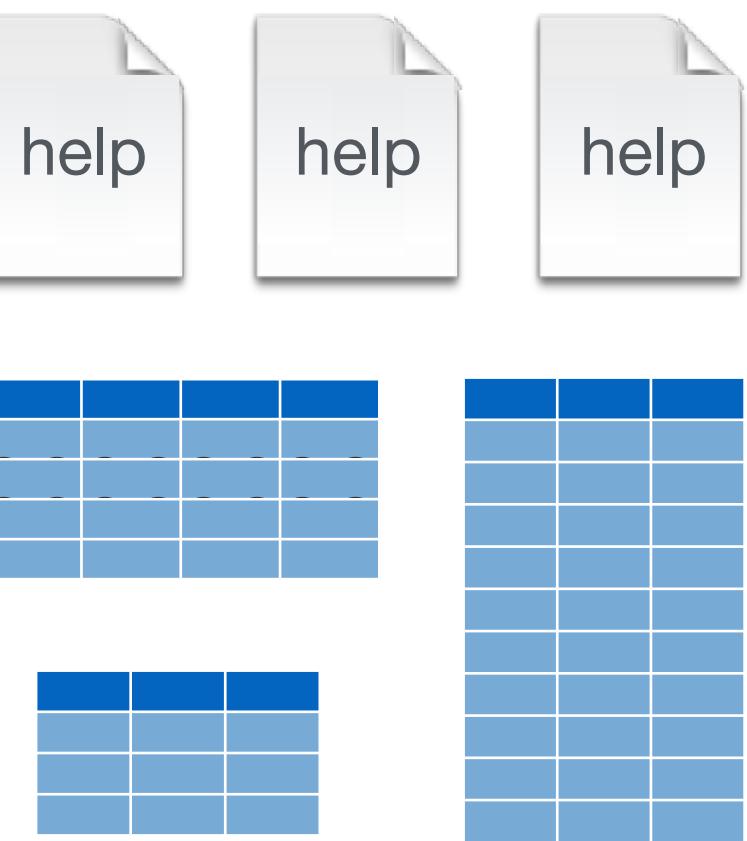
function1()
function2()
function3()
function4()



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



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



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

Base R

R Packages



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](#)

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

CRAN Mirrors

[What's new?](#)

[Task Views](#)

[Search](#)

About R

[R Homepage](#)

[The R Journal](#)

Software

[R Sources](#)

[R Binaries](#)

[Packages](#)

[Other](#)

Documentation

[Manuals](#)

[FAQs](#)

[Contributed](#)

Using packages

1

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

Downloads files to computer

1 x per computer

2

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

Loads package

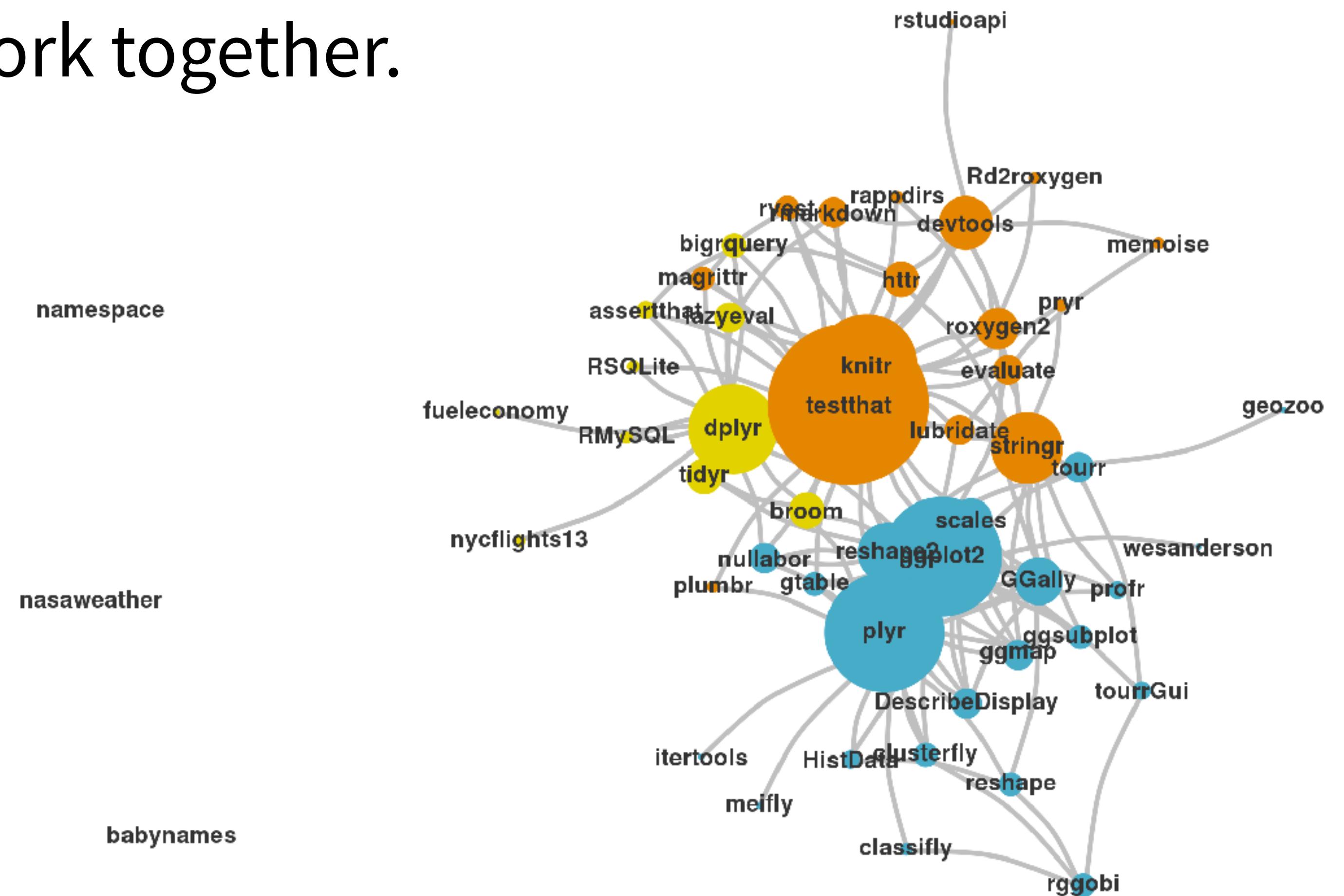
1 x per R Session

The Tidyverse



The Tidyverse

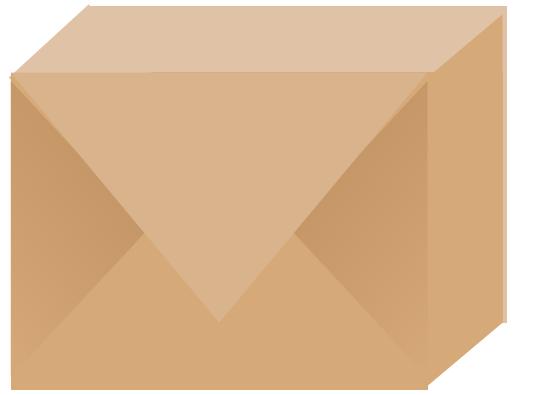
A collection of modern R packages that share common philosophies, embed best practices, and are designed to work together.



tidyverse.org

The screenshot shows 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 heading "Tidyverse". To the right of the heading, there's a brief description: "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 text, there's a section titled "Install the complete tidyverse with:" followed by a code block: `install.packages("tidyverse")`. On the left side of the page, there's a decorative graphic consisting of several hexagonal icons representing different R packages: dplyr (orange, with a pliers icon), ggplot2 (grey, with a line plot icon), readr (blue, with a document icon), purrr (white with a cat icon), tibble (black, with a grid icon), and tidyr (orange, with a circular arrow icon).

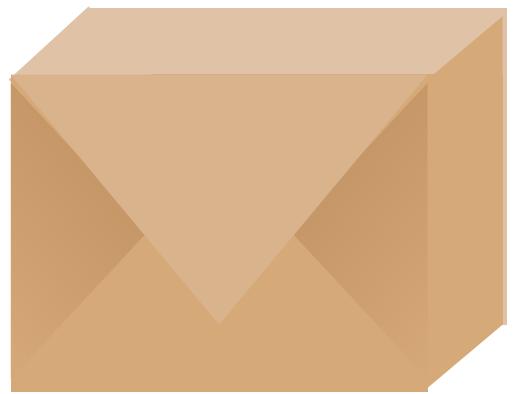
tidyverse



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

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

tidyverse



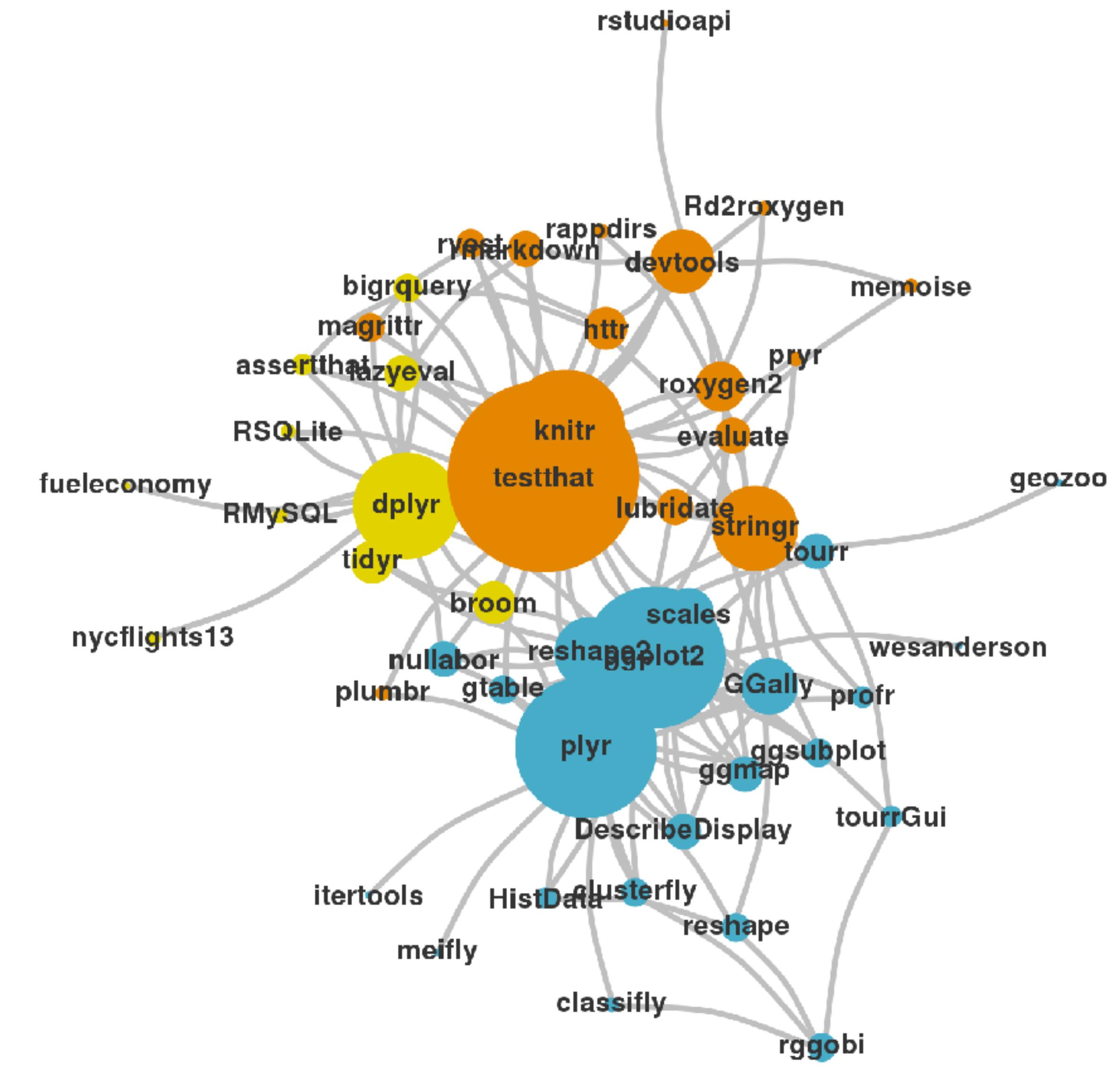
An R package that serves as a short cut for installing and loading the components of the tidyverse.

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

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

does the equivalent of

```
install.packages("ggplot2")
install.packages("dplyr")
install.packages("tidyverse")
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")
```

R Notebooks

(Let's start!)



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

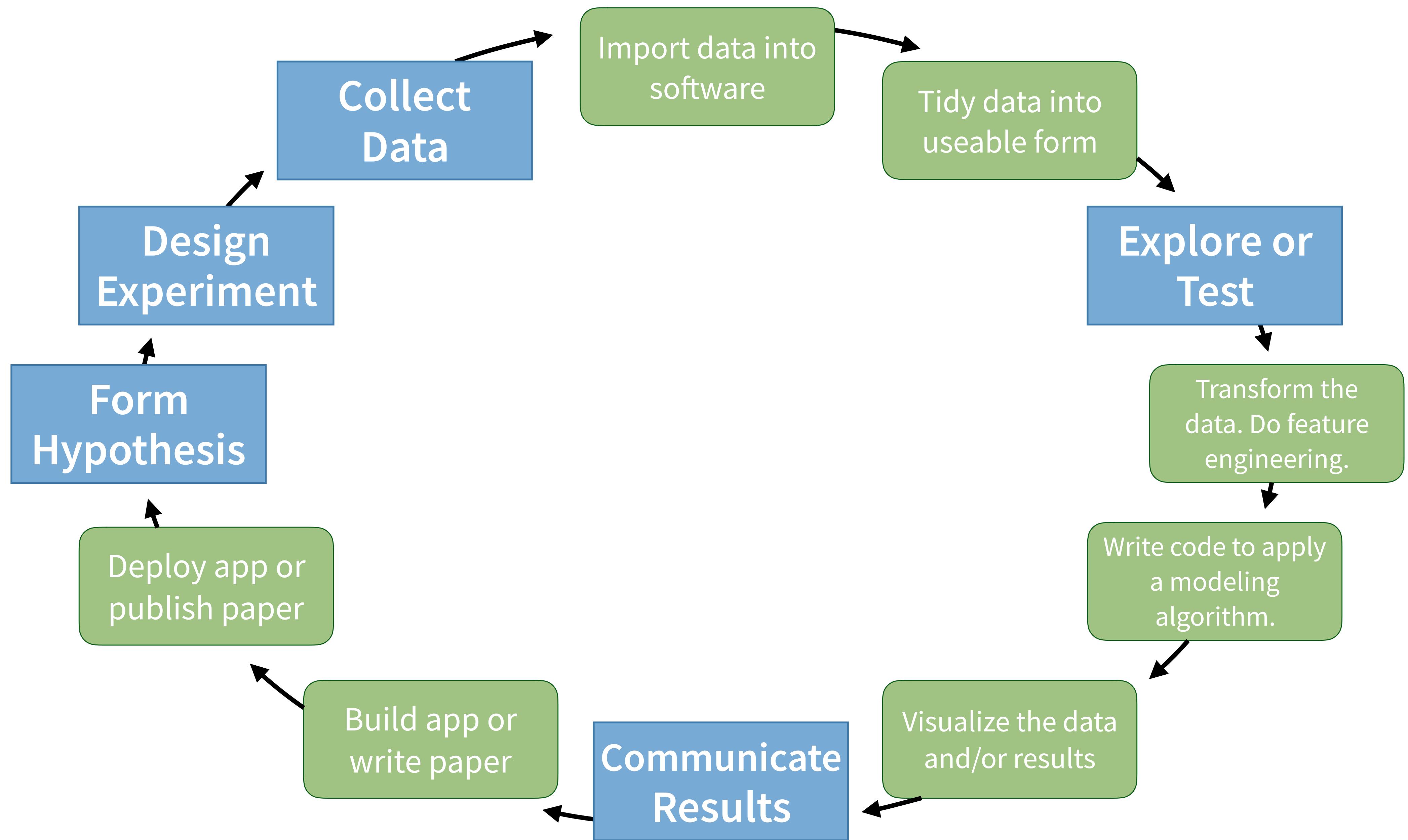
```
[1] -1.2  1.0 -0.5  0.9 -0.6 -1.1 -1.5
```

Three callout boxes highlight features of the interface:

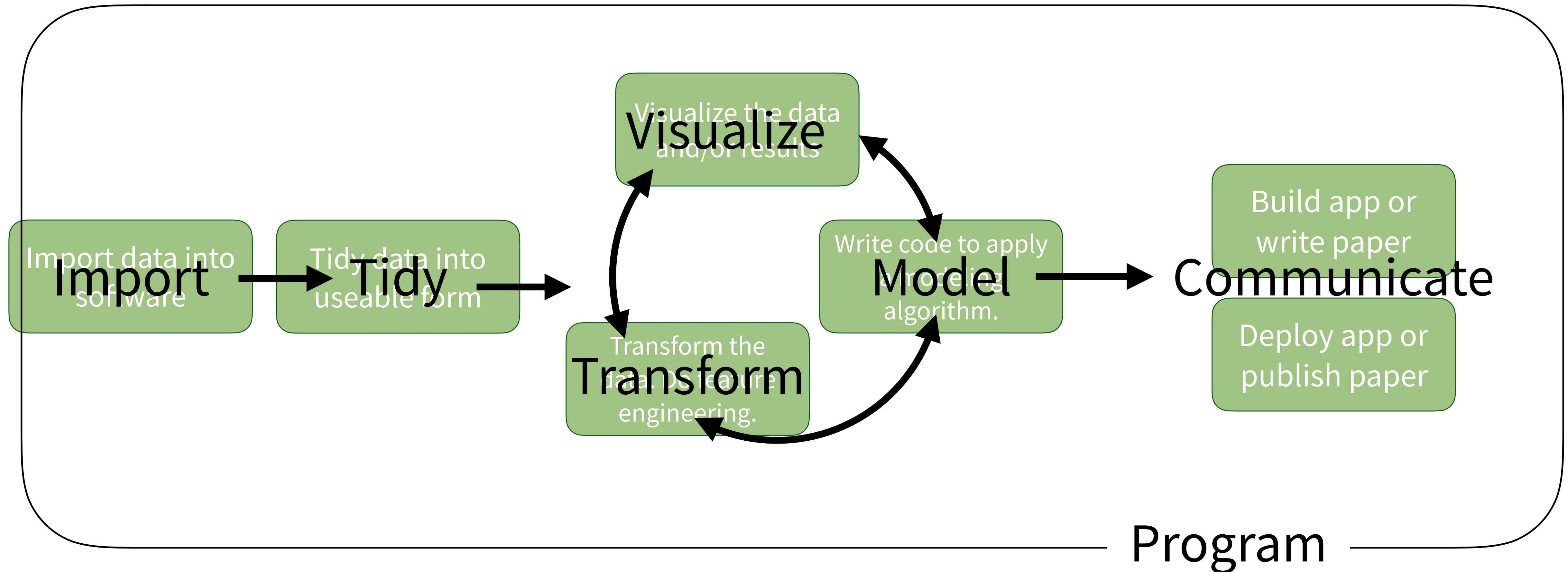
- A grey callout box points to the green "Run All" button in the toolbar, containing the text: "Click to run all code chunks above".
- A grey callout box points to the green "Run Current Chunk" button in the toolbar, containing the text: "Click to run code in chunk".
- A dark grey callout box points to the R console output, containing the text: "Code result".

Outro

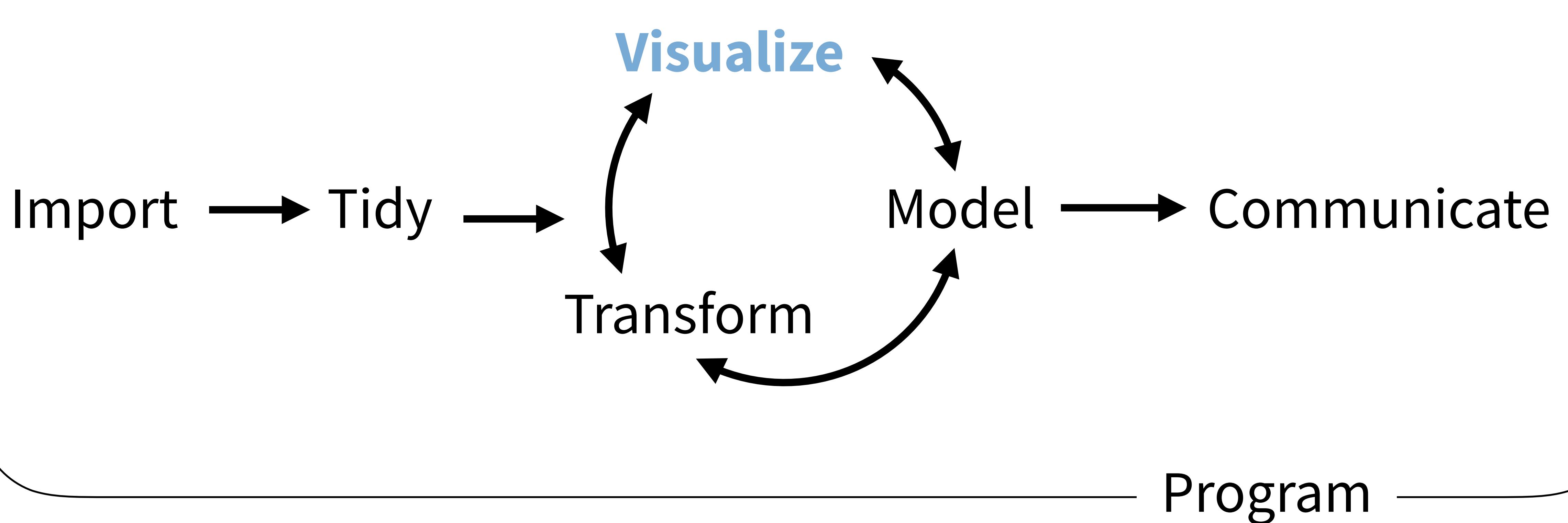




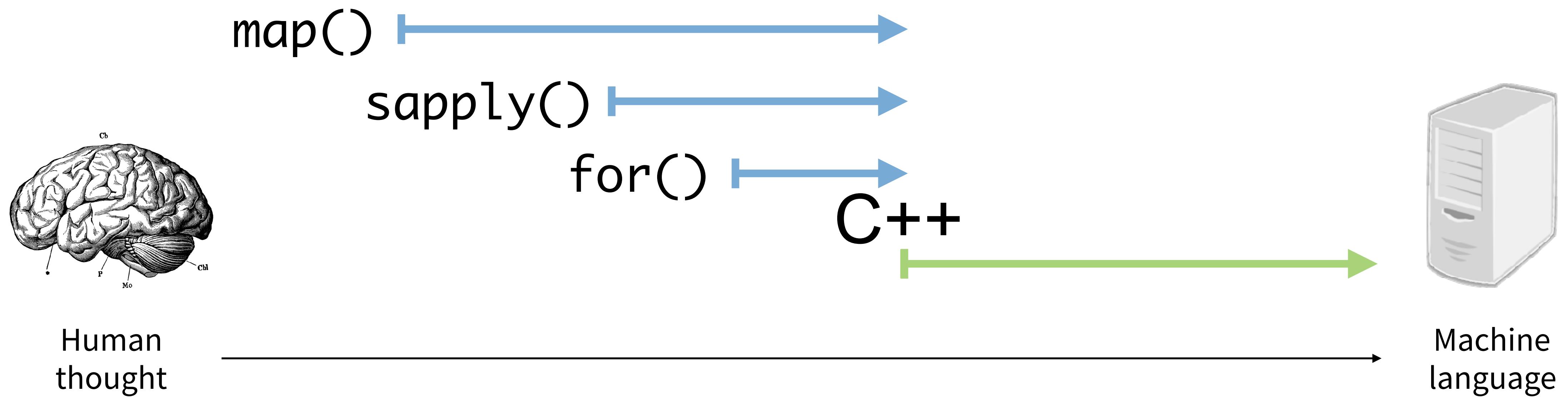
(Applied) Data Science



(Applied) Data Science



R - A computer language for scientists





Your Turn

Open 01-Visualize-Data.Rmd.



Master the Tidyverse



Garrett Grolemund

Data Scientist, Educator

January 2017

RStudio