

Introduction to the Tidyverse

Empirical Economics Course – Tutorials first week

The content of these slides is taken from the course and slides by

Garett Grolemund

<https://github.com/rstudio-education/welcome-to-the-tidyverse>

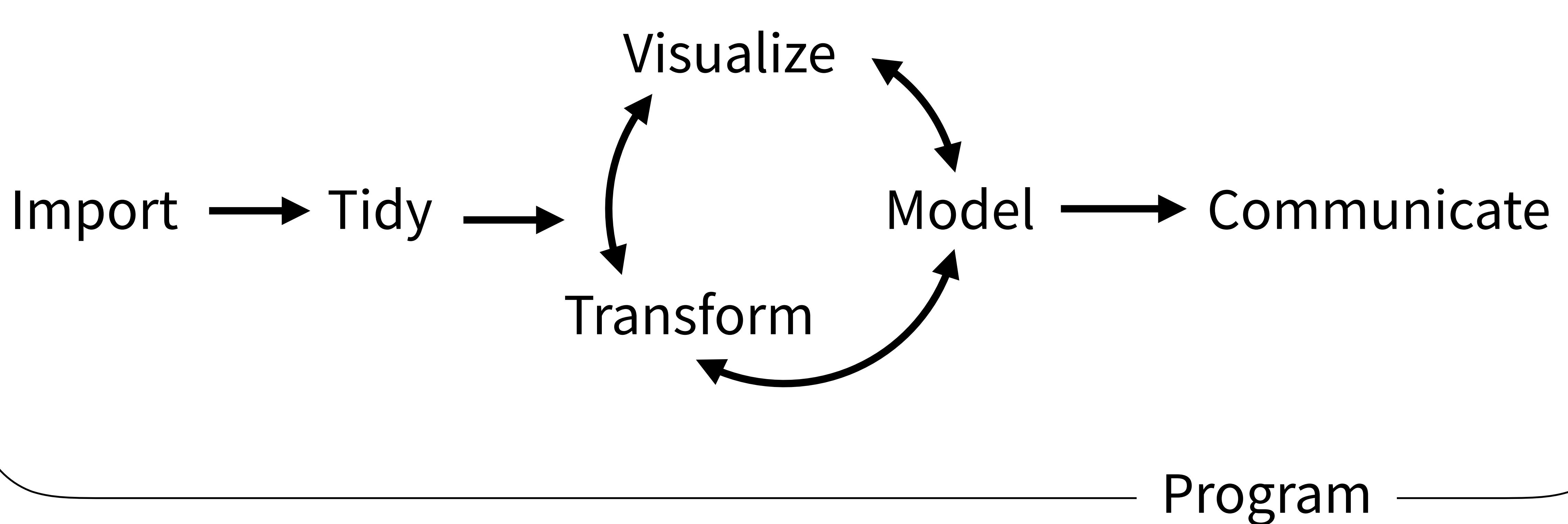
Table of contents

- Basics of R and R packages
- Visualization with “ggplot2”
- Transforming data with “dplyr”

Welcome to the Tidyverse



(Applied) Data Science



R the language

Values - 1, "Florida", "2010-01-25"

R the language

Values - 1, "Florida", "2010-01-25"

Objects - x <- 22/7

A name
without quotes

< followed by -
(it looks like an arrow)

A value,
object, or
function result

R the language

Values - 1, "Florida", "2010-01-25"

Objects - x <- c(22/7, 0.99, 3)

To put multiple values in an object,
combine the values with c()

R the language

Values - 1, "Florida", "2010-01-25"

Objects - x <- c(22/7, 0.99, 3)

Functions - round(x, digits = 3)

A name
without
quotes

followed by
() to run the
function

Arguments:
values, objects, or
function results

Warm Up

Which of these are numbers?

1

"1"

"one"

one

Warm Up

Which of these are numbers?

1

"1"

"one"

one

number

Warm Up

Which of these are numbers?

1

number

"1"

"one"

one

words (strings)

Warm Up

Which of these are numbers?

1

number

"1"

words (strings)

"one"

one

object

Warm Up

Which of these will work? Suppose `one <- 1.`

`log(1)` `log("1")` `log("one")` `log(one)`

Warm Up

Which of these will work? Suppose `one <- 1.`

log(1)

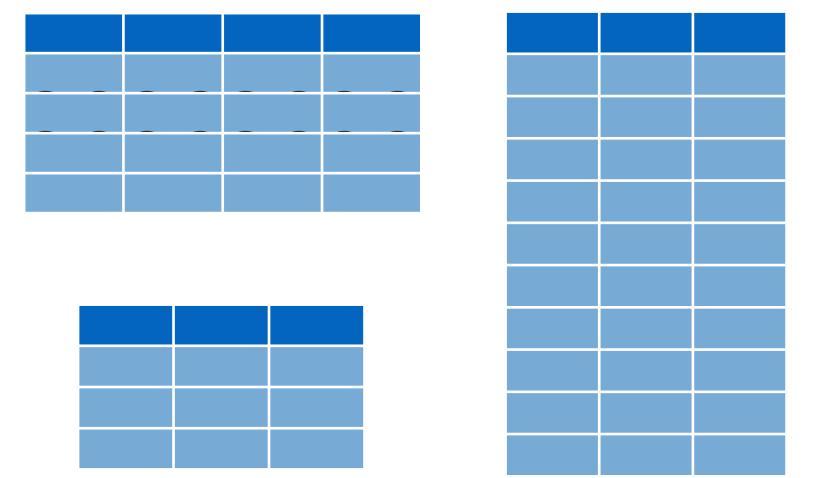
log("1")

log("one")

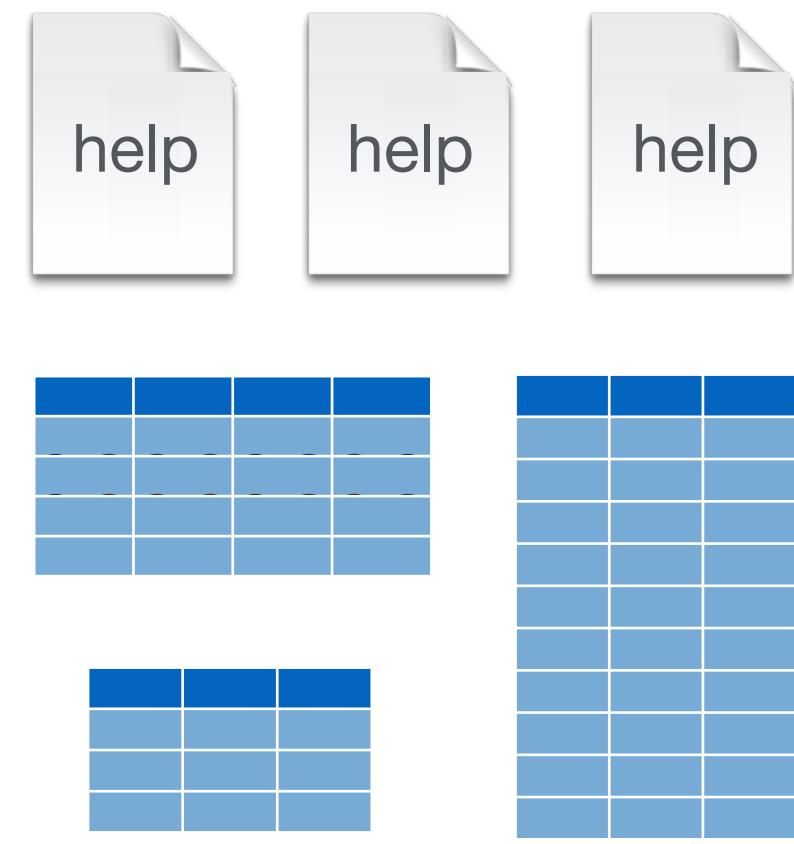
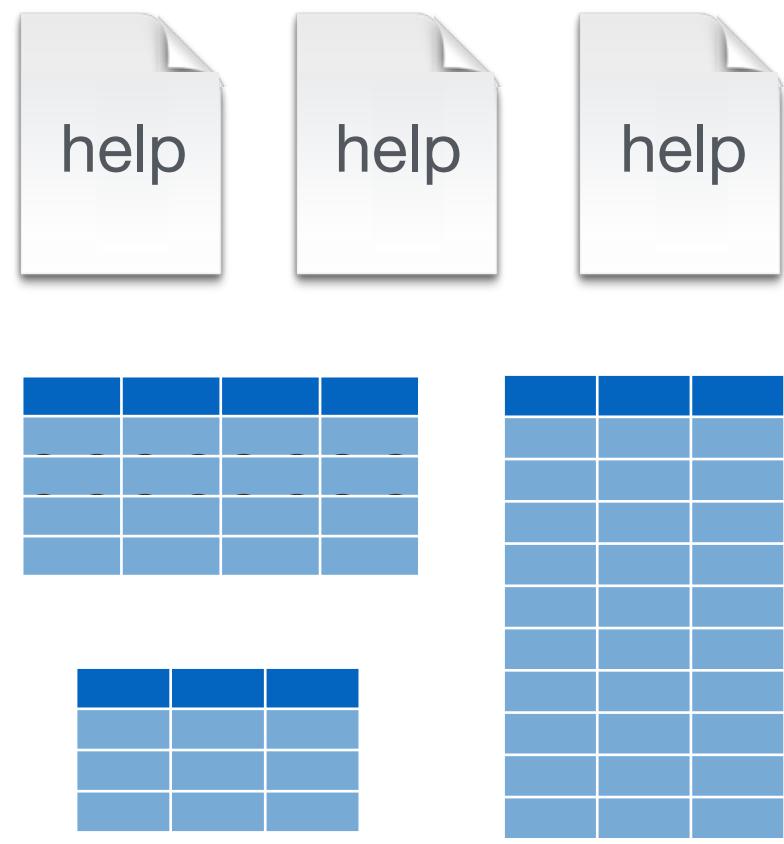
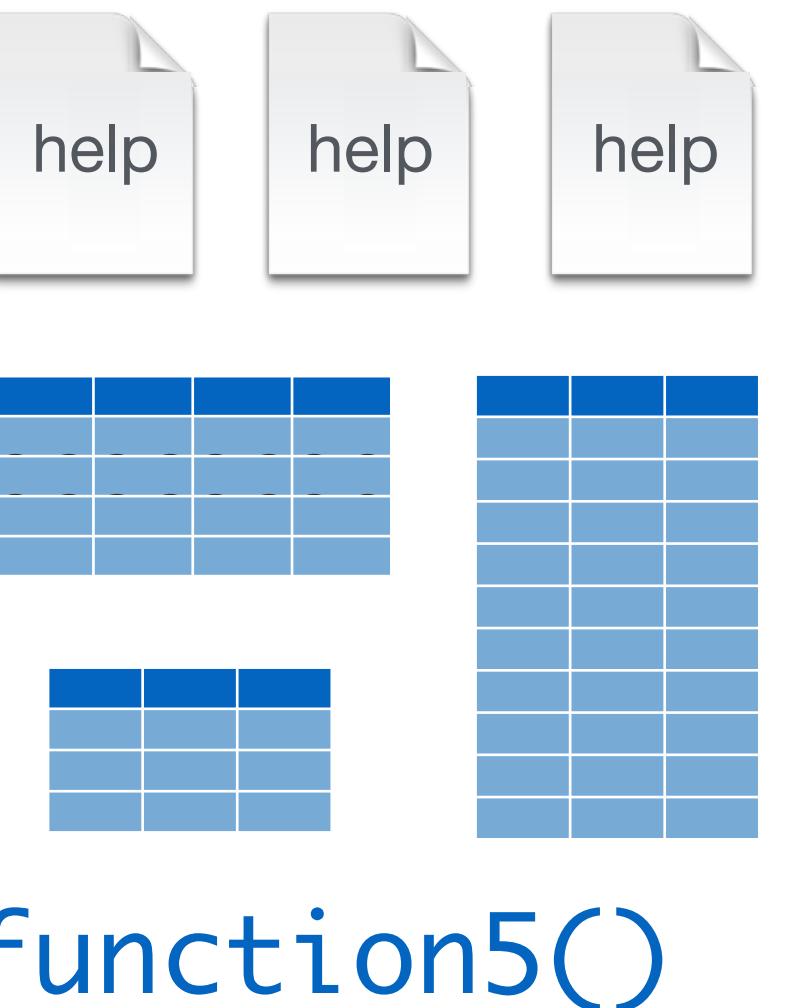
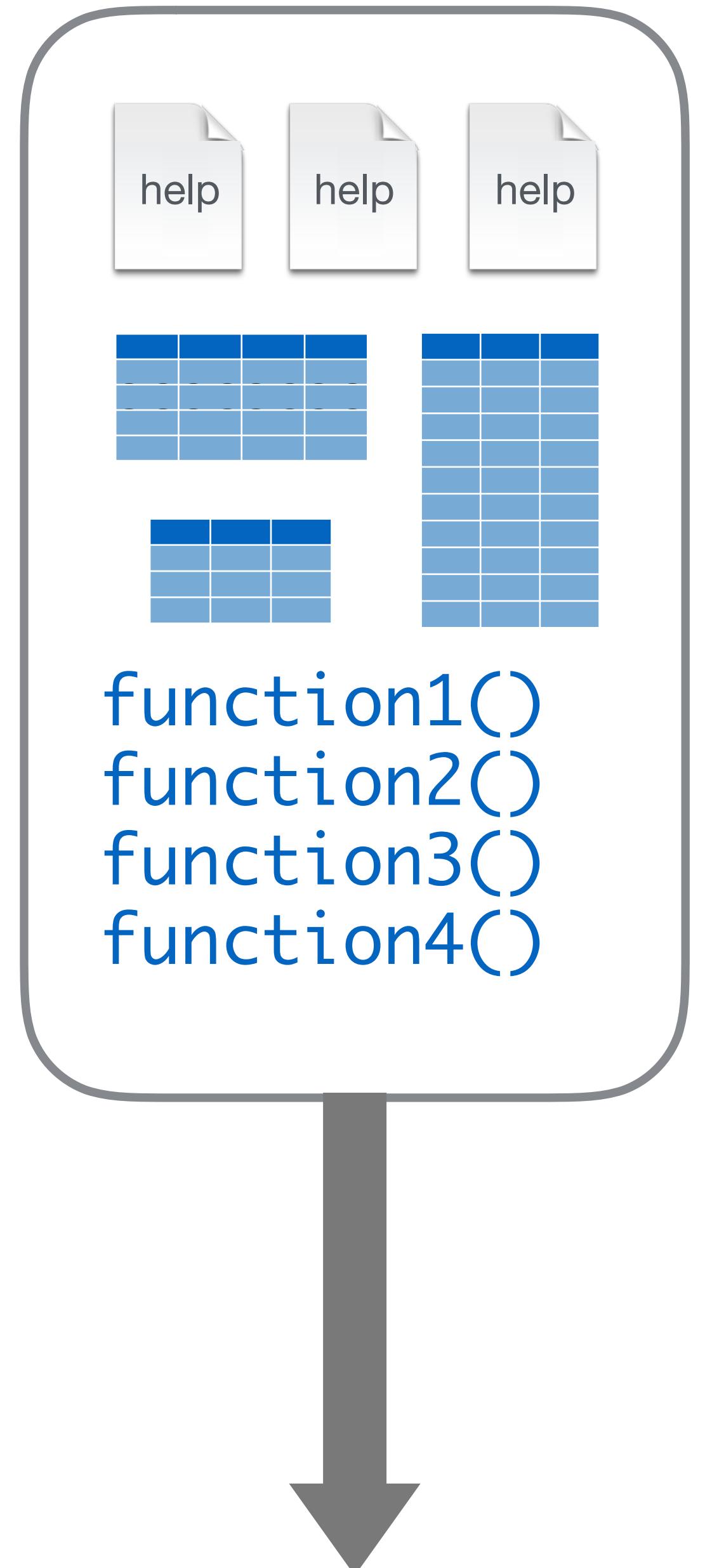
log(one)

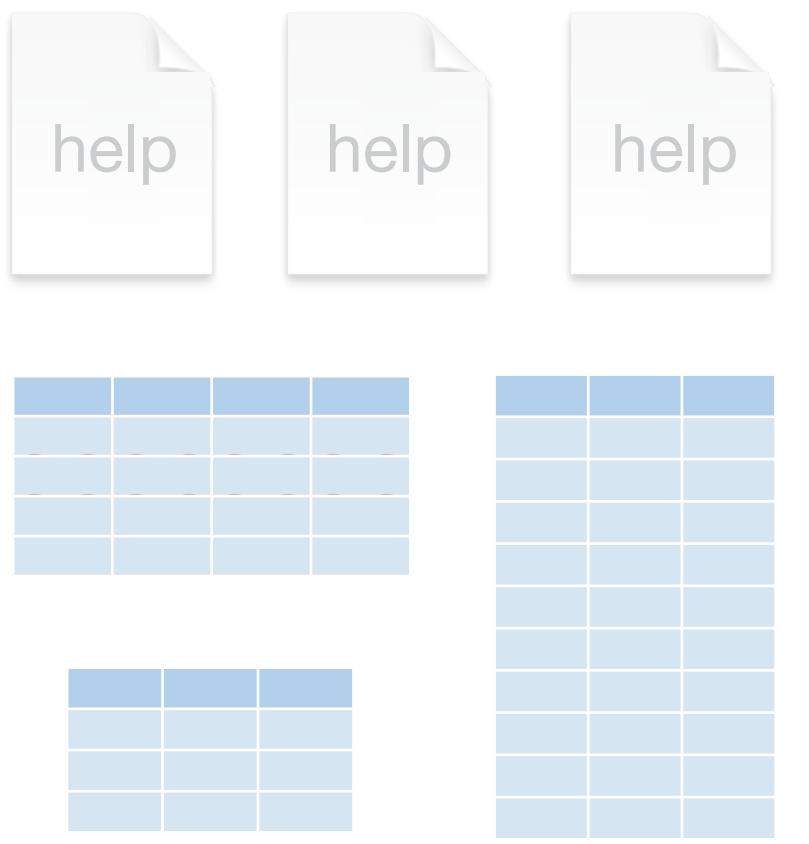
R Packages



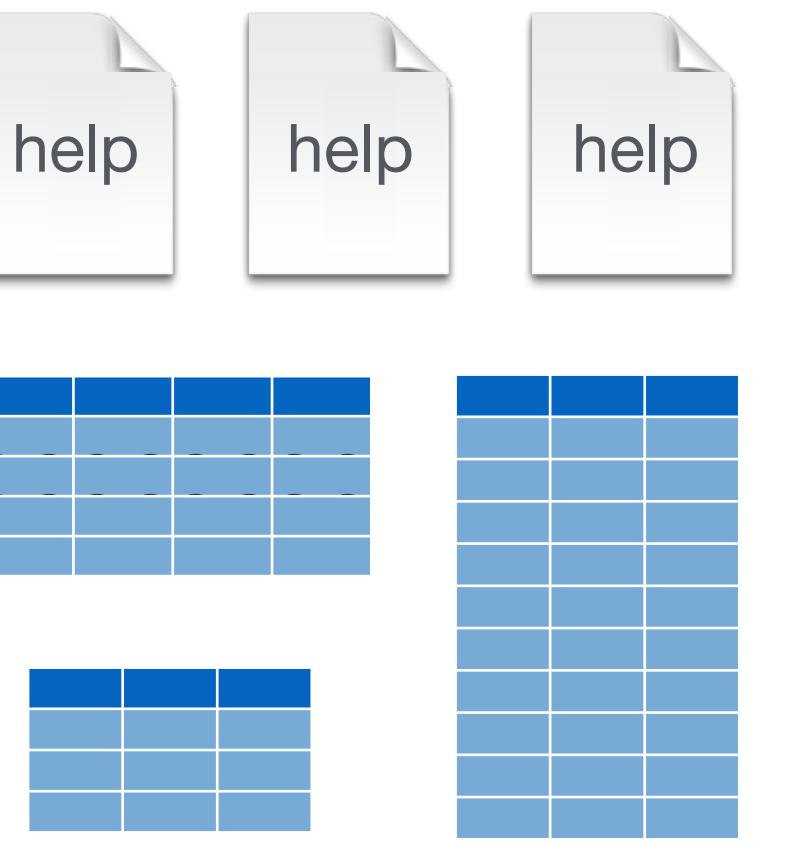


function1()
function2()
function3()
function4()

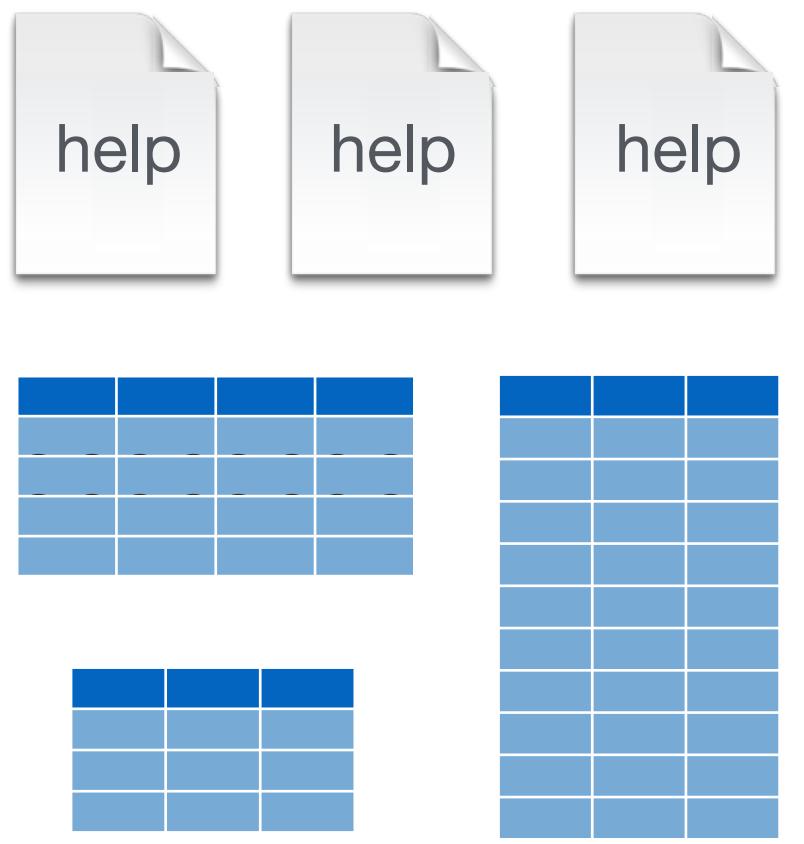




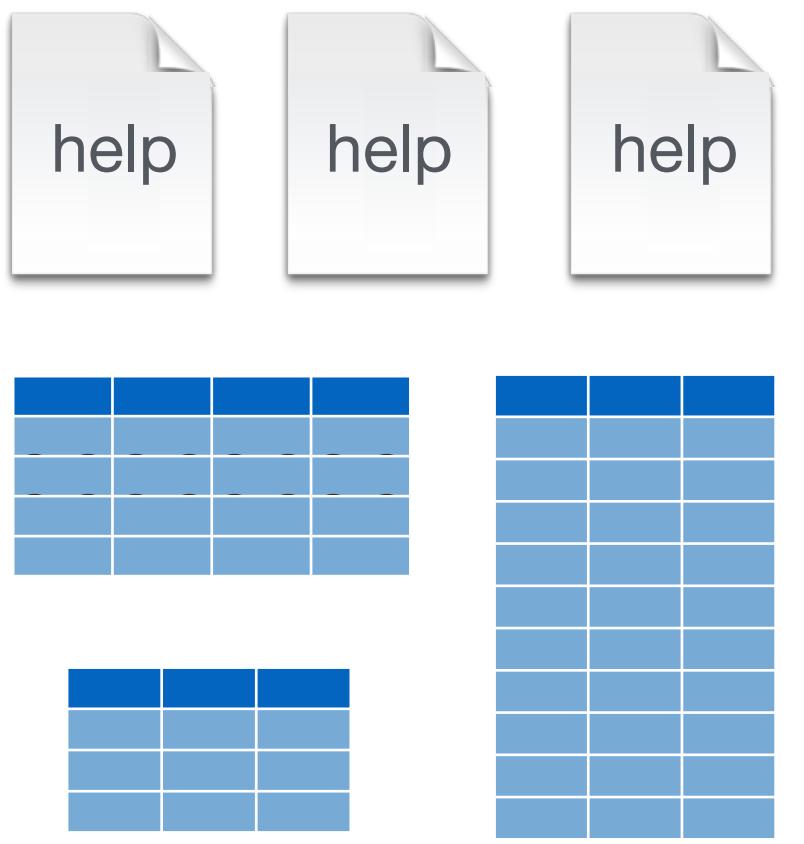
function1()
function2()
function3()
function4()



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

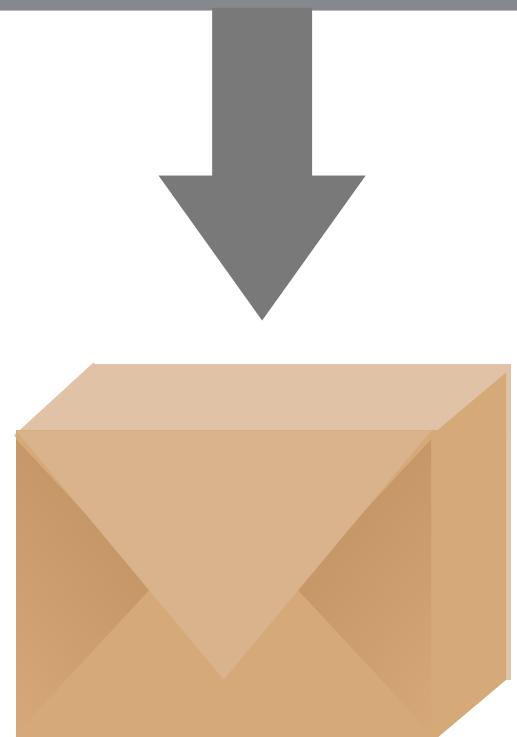
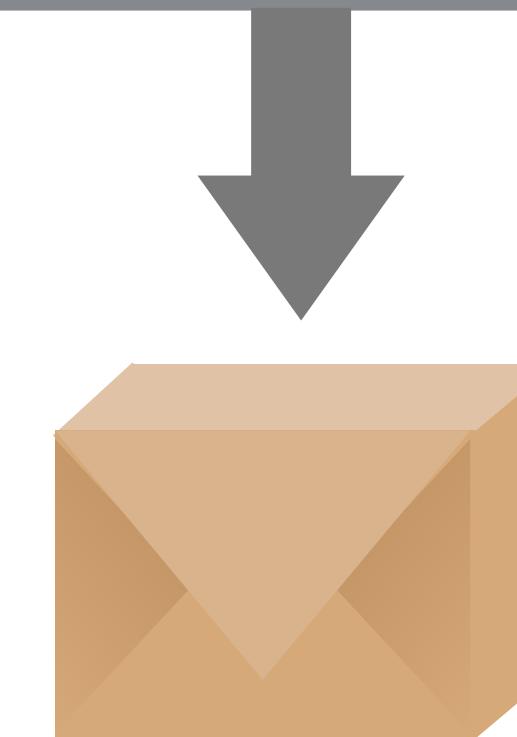
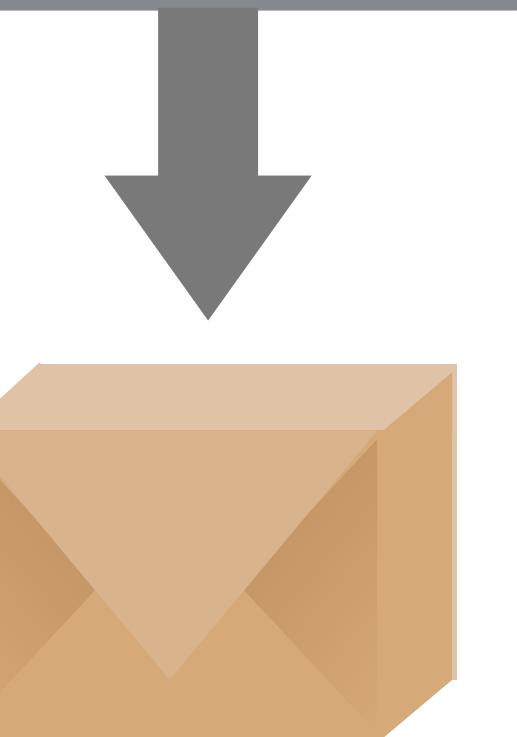


function9()
functionA()
functionB()
functionC()

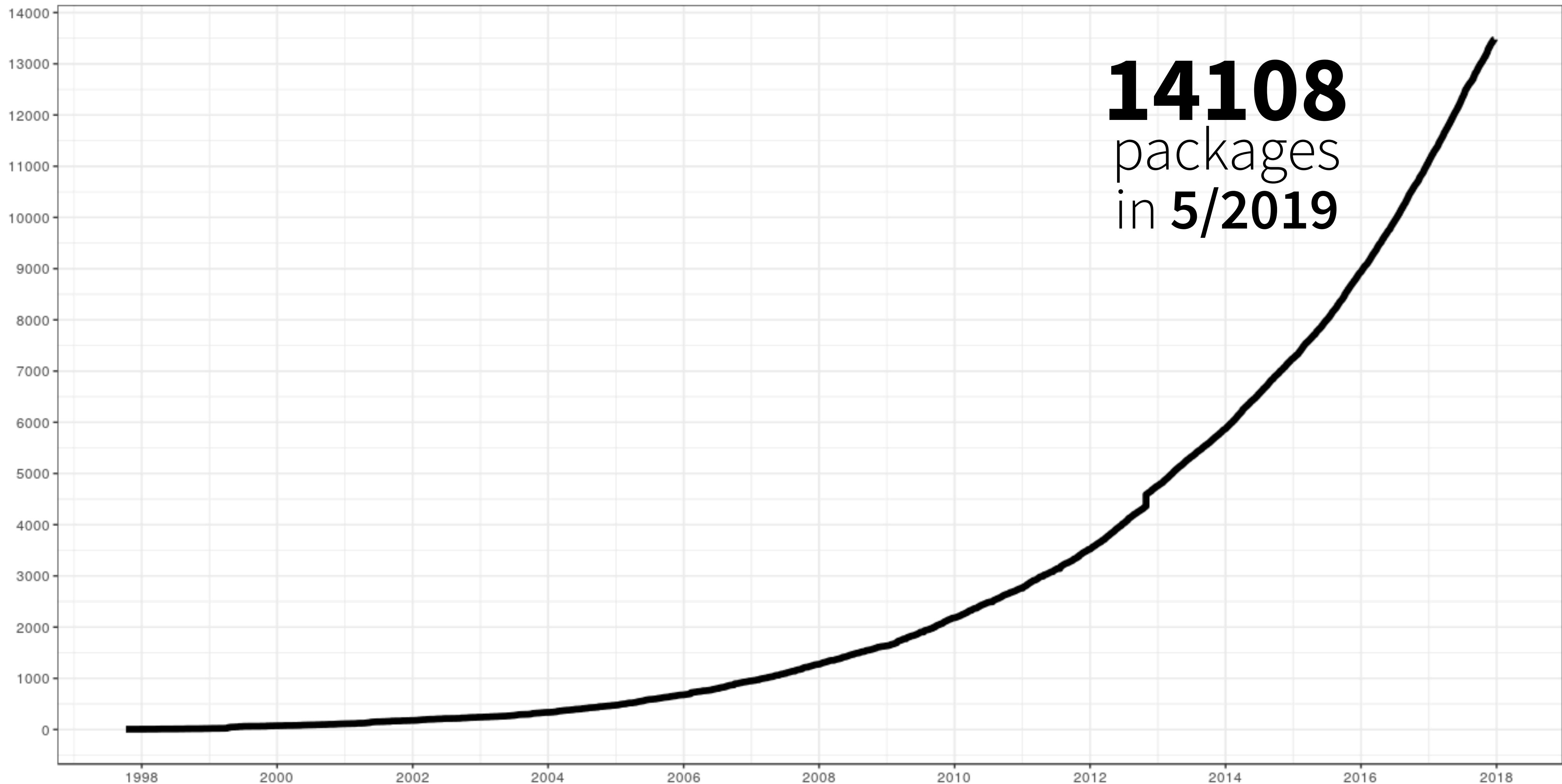


functionD()
functionE()
functionF()
functionG()

Base R



Number of R packages ever published on CRAN



The screenshot shows a web browser window with the title "The Comprehensive R Archive". The address bar indicates a secure connection to "https://cran.r-project.org". In the top right corner, there is a user profile icon with the name "Garrett". The main content area displays the "Available CRAN Packages By Name" page, featuring a large blue "R" logo on the left. Below the logo, there is a sidebar with various links: "CRAN Mirrors", "What's new?", "Task Views", "Search", "About R", "R Homepage", "The R Journal", "Software", "R Sources", "R Binaries", "Packages", and "Other". The main content area lists packages starting with "A", such as "A3", "abbyyR", "abc", "ABCAnalysis", "abc.data", "abcdeFBA", "ABCOptim", "ABCp2", "ABC.RAP", "abcrf", "abctools", "abd", "abf2", "ABHgenotypeR", "abind", "abjutils", "abn", and "abodOutlier", each followed by a brief description.



[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](#)

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

[abbyyR](#) 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 Jurimetrical Analysis Used by the Brazilian Jurimetrics Association

[abn](#) Modelling Multivariate Data with Additive Bayesian Networks

[abodOutlier](#) Angle-Based Outlier Detection

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 of what the tidyverse is: "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 code block showing the command to install the tidyverse: "install.packages("tidyverse")". On the left side of the page, there's a graphic featuring six hexagonal icons representing different tidyverse 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 (dark blue, with a grid icon), and tidyr (orange, with a circular arrow icon).

Tidyverse

Packages Articles Learn Help Contribute

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.

Install the complete tidyverse with:

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

Using packages

1

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

Downloads files to computer

1 x per computer

Pop Quiz!

The tidyverse contains the following packages.
What function would you use to install them?

ggplot2

dplyr

tidyr

readr

purrr

tibble

hms

stringr

lubridate

forcats

DBI

haven

httr

jsonlite

readxl

rvest

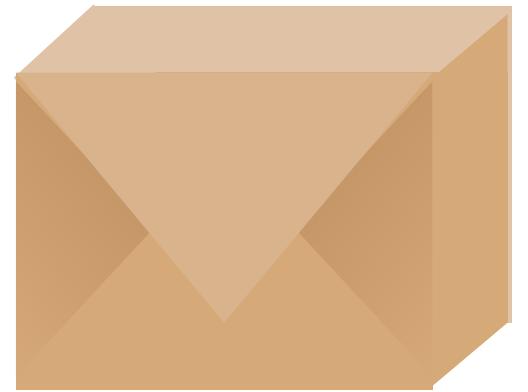
xml2

modelr

tidyverse

```
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")
```

tidyverse



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

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

A shortcut!

```
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")
```

I've already installed all of
the packages you need

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("stringr")
install.packages("forcats")
install.packages("lubridate")
install.packages("hms")
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")
library("stringr")
library("forcats")
```

R Markdown



Your Turn

Go to **01-Introduction.**

Open **01-RMarkdown-Exercises.Rmd**. Read through the file and do everything it tells you to do.



R Markdown

An authoring format for Data Science.

The screenshot shows the RStudio interface with an R Markdown file open. The file contains the following content:

```
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 ````  
19  
20 (Top Level) ◊
```

The code chunk at line 8-11 is highlighted with a green background. The resulting output, a vector of random numbers, is displayed in the console at line 15:

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

Annotations with arrows point to specific parts of the interface:

- A callout bubble points to the code chunk in the editor with the text "Code goes in a chunk".
- A callout bubble points to the green play button icon next to the code chunk in the editor with the text "Click to run code in chunk".
- A callout bubble points to the console output with the text "Code result".

Code chunks

Insert a chunk of R code with

```
```{r}  
some code
```
```

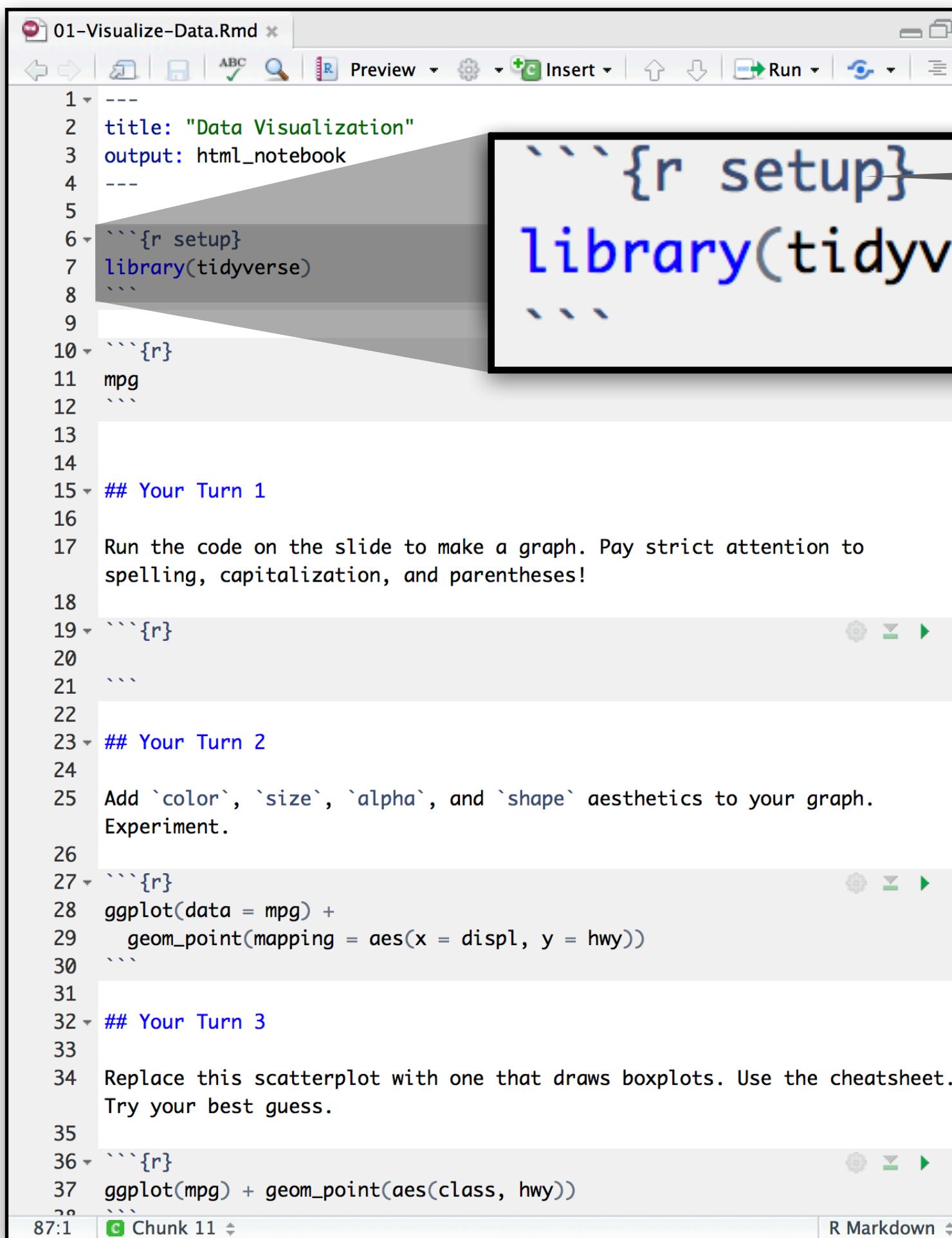
 + Opt + i (Mac)

Ctrl + Alt + i (PC)



Setup

The setup chunk is always run once before anything else

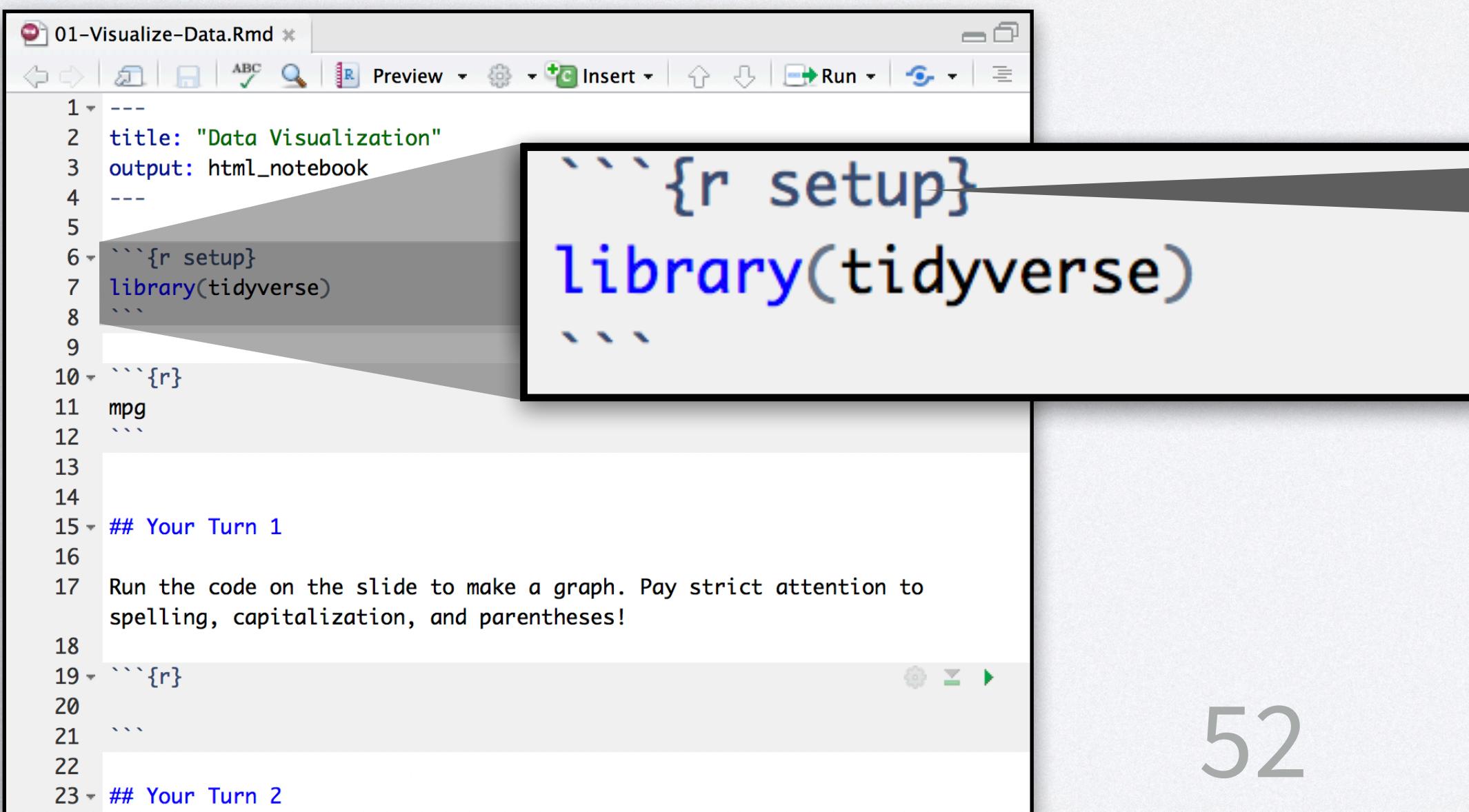


```
1 ---  
2 title: "Data Visualization"  
3 output: html_notebook  
4 ---  
5  
6 ```{r setup}  
7 library(tidyverse)  
8 ```  
9  
10 ```{r}  
11 mpg  
12  
13  
14  
15 ## Your Turn 1  
16  
17 Run the code on the slide to make a graph. Pay strict attention to  
spelling, capitalization, and parentheses!  
18  
19 ```{r}  
20  
21  
22  
23 ## Your Turn 2  
24  
25 Add `color`, `size`, `alpha`, and `shape` aesthetics to your graph.  
Experiment.  
26  
27 ```{r}  
28 ggplot(data = mpg) +  
29   geom_point(mapping = aes(x = displ, y = hwy))  
30  
31  
32 ## Your Turn 3  
33  
34 Replace this scatterplot with one that draws boxplots. Use the cheatsheet.  
Try your best guess.  
35  
36 ```{r}  
37 ggplot(mpg) + geom_point(aes(class, hwy))  
38  
87:1 | [green] Chunk 11 | R Markdown
```

chunk labels are optional,
the setup label is special

Your Turn

Add a setup chunk to the top of **01-RMarkdown-Exercises.Rmd**. Use it to load the tidyverse package, then uncomment and run the ggplot2 chunk at the bottom of your file.



```
1 ---  
2 title: "Data Visualization"  
3 output: html_notebook  
4 ---  
5  
6 ```{r setup}  
7 library(tidyverse)  
8 ```  
9  
10 ```{r}  
11 mpg  
12 ```  
13  
14  
15 ## Your Turn 1  
16  
17 Run the code on the slide to make a graph. Pay strict attention to  
spelling, capitalization, and parentheses!  
18  
19 ```{r}  
20 ```  
21  
22  
23 ## Your Turn 2
```

chunk labels are optional,
the setup label is special

Your Turn 0

Navigate up to the **02-Visualize** folder.

Open 02-Visualize-Exercises.Rmd

Add a setup chunk that loads the
tidyverse packages.

