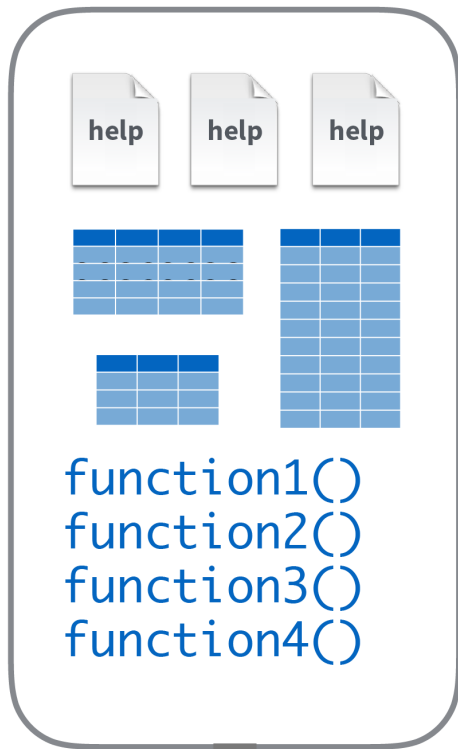


Data basics

Packages

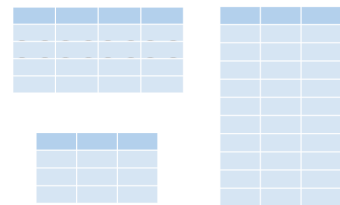
A rounded rectangular icon representing an R package. It contains three 'help' file icons at the top, two data frame icons (represented as blue grids) in the middle, and a list of function names at the bottom.

help help help

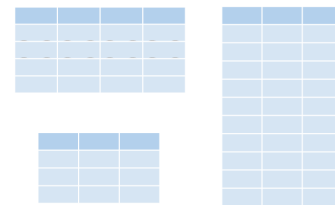
function1()
function2()
function3()
function4()



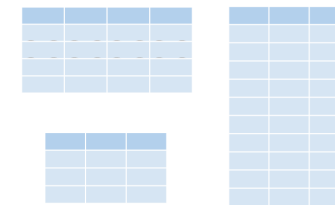
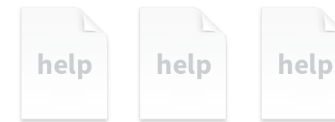
Base R



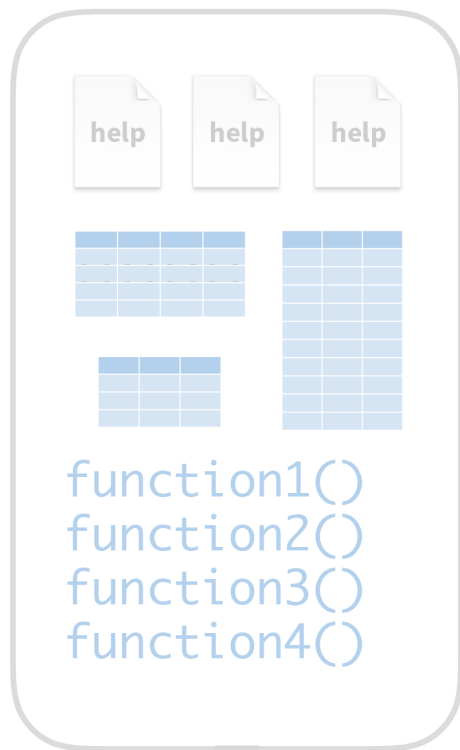
function5()
function6()
function7()
function8()



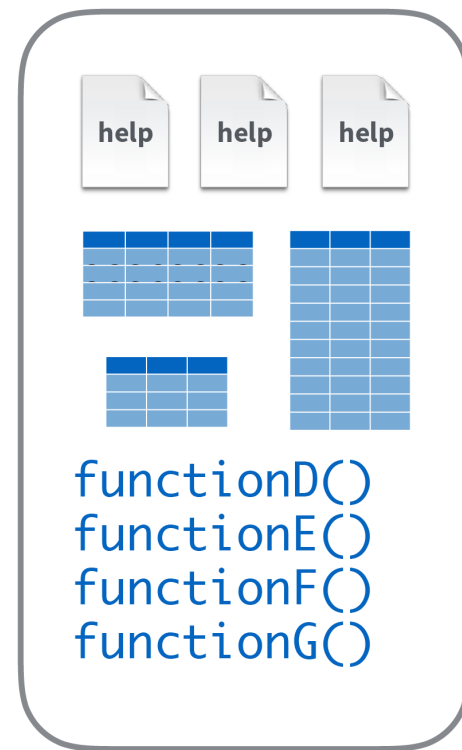
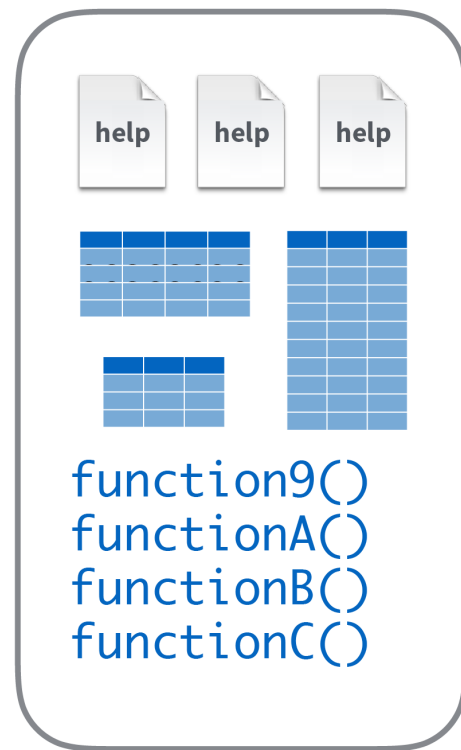
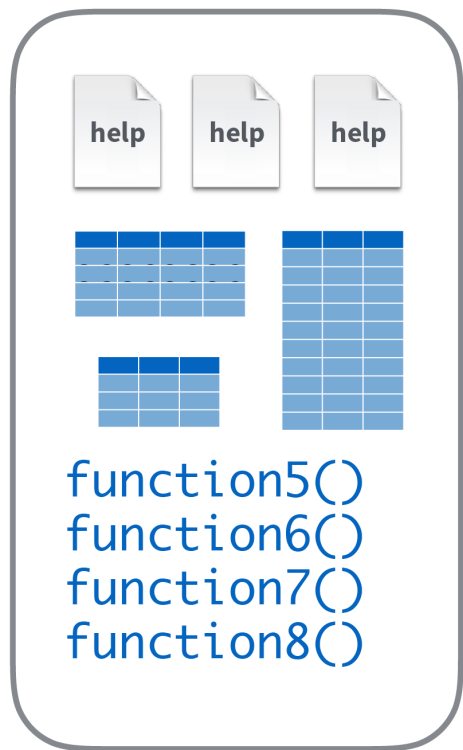
function9()
functionA()
functionB()
functionC()



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



Base R



R Packages

Using packages

```
install.packages("name")
```

**Downloads files
to your computer**

Do this once per computer

```
library("name")
```

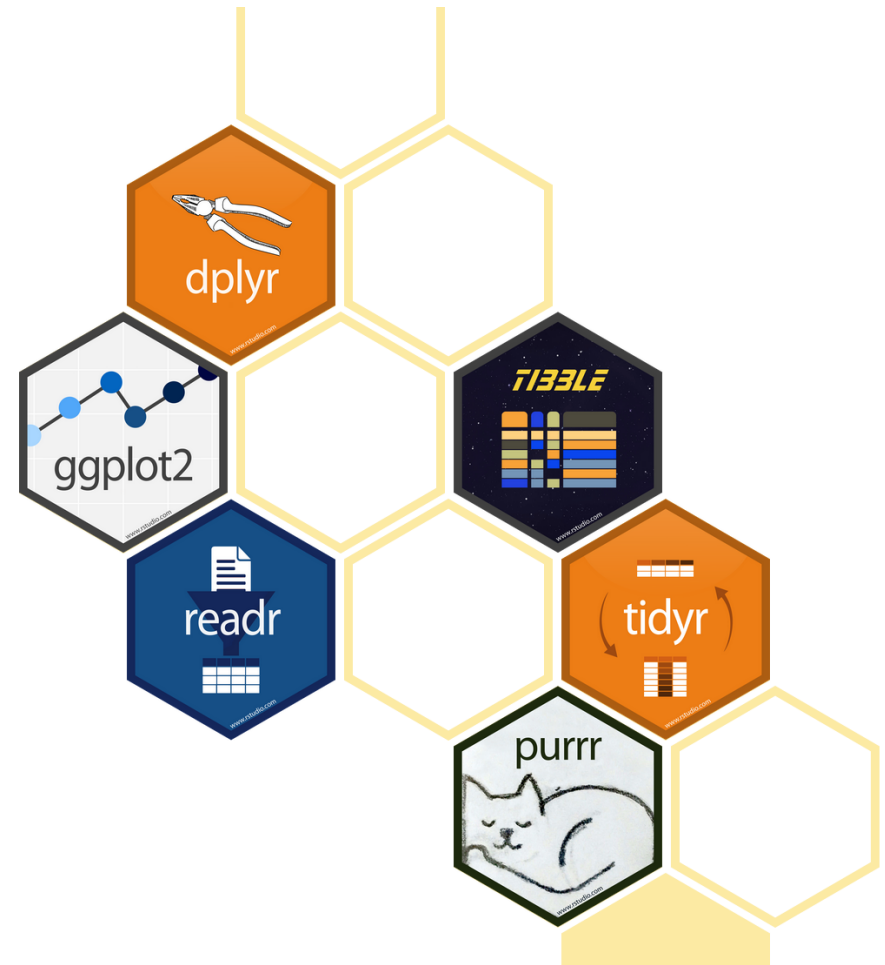
Loads the package

Do this once per R session

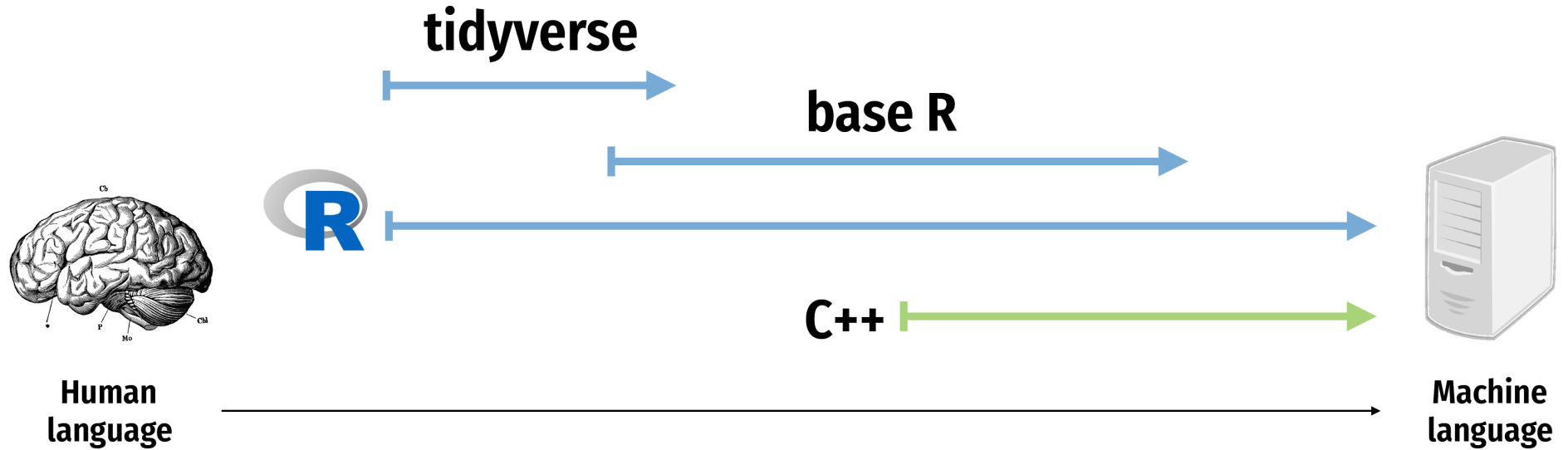
The tidyverse

"The tidyverse is an opinionated collection of R packages designed for data science. All packages share an underlying design philosophy, grammar, and data structures."

... the tidyverse makes data science faster, easier and more fun...



The tidyverse



The tidyverse package

```
library(tidyverse)
```

The tidyverse package is a shortcut for installing and loading all the key tidyverse packages


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

```
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)
```

```
library(ggplot2)  
library(dplyr)  
library(tidyr)  
library(readr)  
library(purrr)  
library(tibble)  
library(stringr)  
library(forcats)  
library(lubridate)
```

Rectangular data

Data frames and tibbles

Data frames are the most common kind of data objects; used for rectangular data (like spreadsheets)

Data frames: R's native data object

Tibbles (tbl): a fancier enhanced kind of data frame

(You really won't notice a difference today)

Vectors and lists

Vectors

Vectors are a list of values of the same type
(all text, or all numbers, etc.)

Make them with `c()`:

```
c(1, 4, 2, 5, 7)
```

You'll usually want to assign them to something:

```
neat_numbers <- c(1, 4, 2, 5, 7)
```

Basic data types

Integer	Whole numbers	<code>c(1, 2, 3, 4)</code>
Double	Numbers	<code>c(1, 2.4, 3.14, 4)</code>
Character	Text	<code>c("1", "blue", "fun", "monster")</code>
Logical	True or false	<code>c(TRUE, FALSE, TRUE, FALSE)</code>
Factor	Category	<code>c("Strongly disagree", "Agree", "Neutral")</code>

Importing data

Packages for importing data



Work with plain text data

```
my_data <-  
read_csv("file.csv")
```



Work with Excel files

```
my_data <-  
read_excel("file.xlsx")
```



Work with Stata, SPSS, and SAS data

```
my_data <-  
read_stata("file.dta")
```


Other types of data

Package	Type of data
jsonlite	JSON data
xml2	XML data
httr	Web APIs
rvest	Web pages (web scraping)
DBI	Databases (SQL)

Next up

Data visualization