

Débuter avec R et RStudio





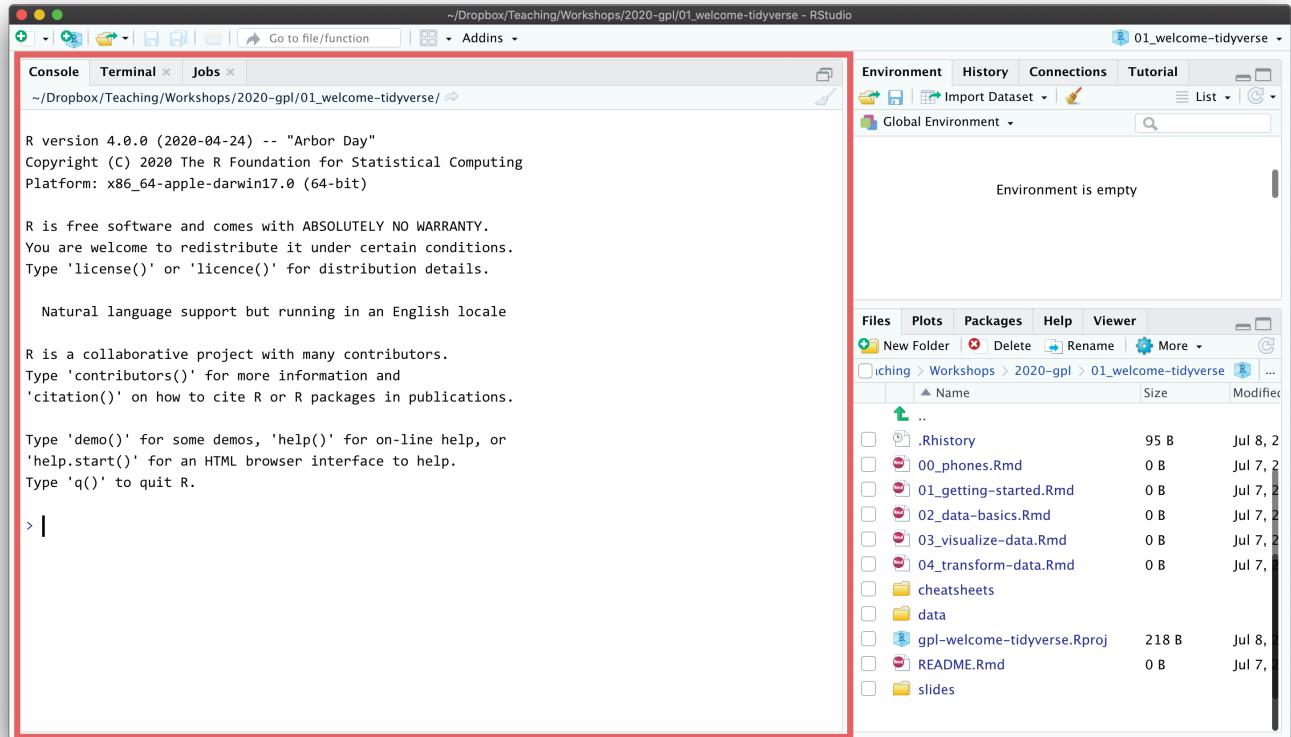
langage de programmation



environnement de développement
intégré

Un panorama de RStudio

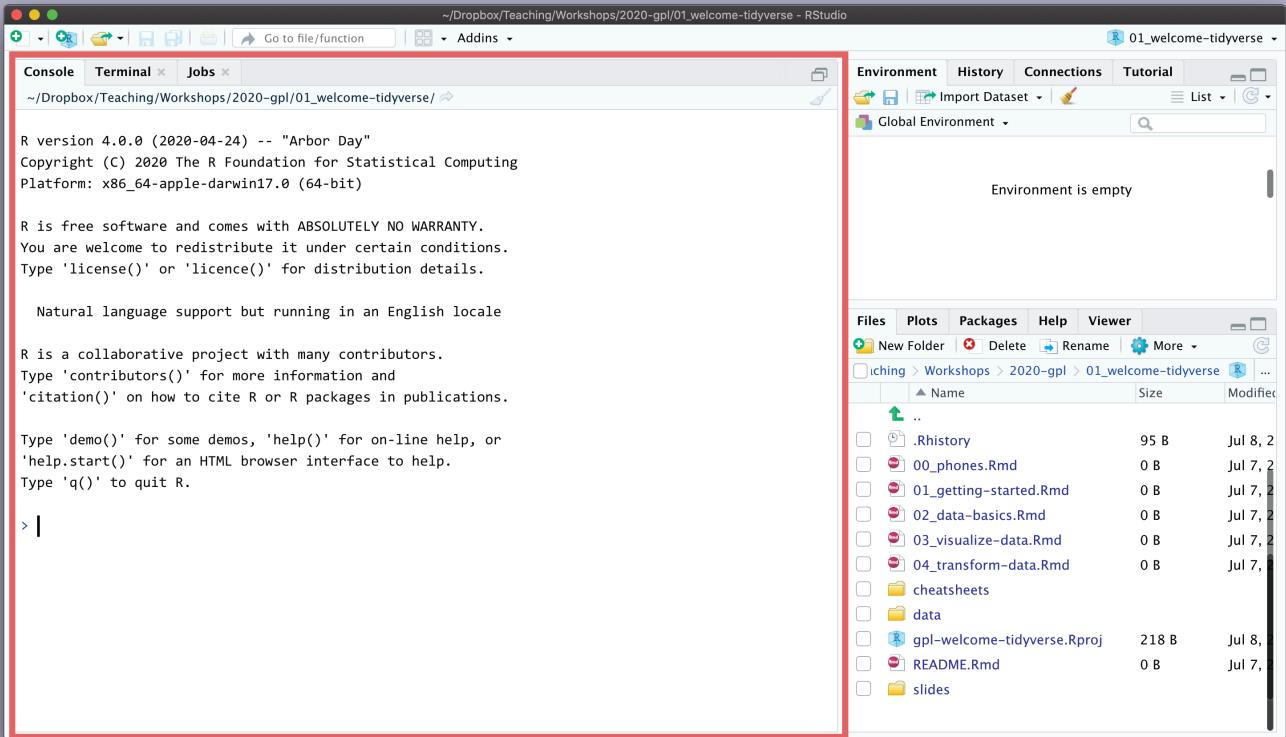
Console



R est en attente de vos instructions

Tapez votre code, et appuyez sur Retour pour le compiler

À votre tour



Tapez $2 + 2$ dans la console

Appuyez sur Retour

2 + 2

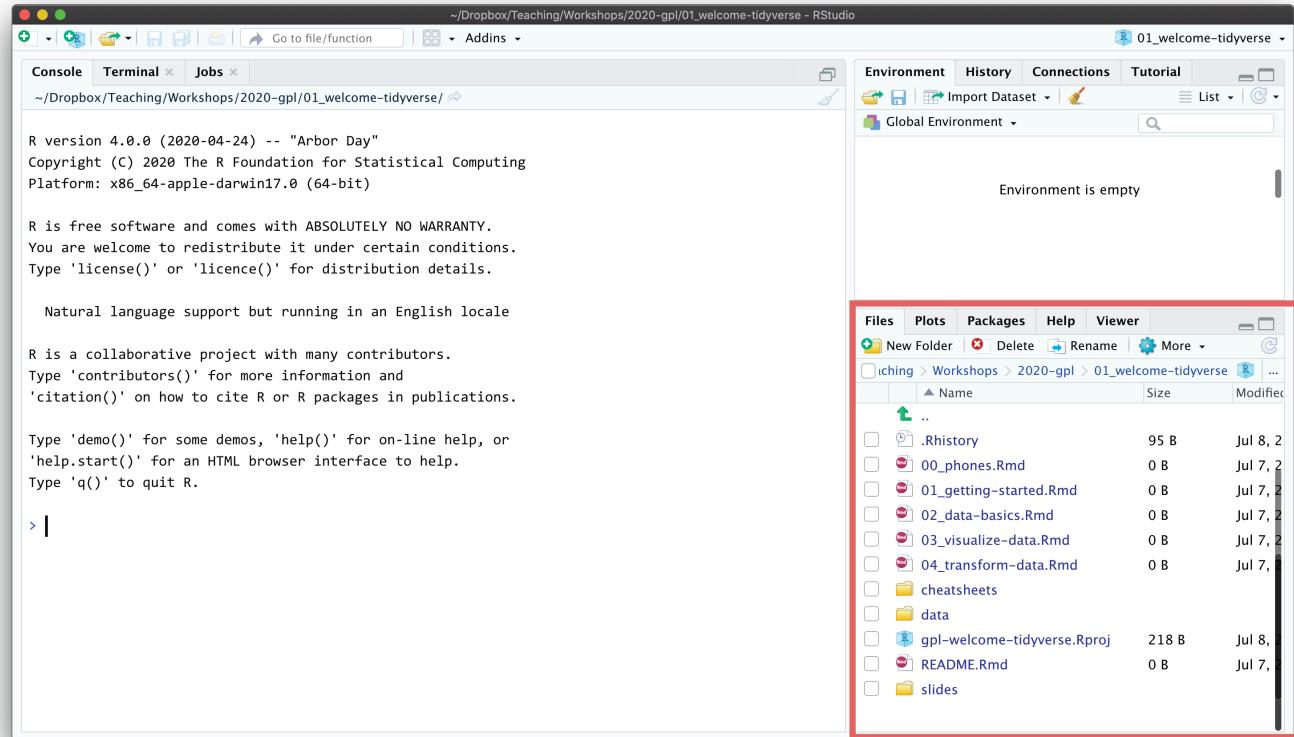
```
## [1] 4
```

**Ce calcul est éphémère.
Pour recompiler, il faudra retaper les instructions.**

On va plutôt enregistrer le code dans un document R

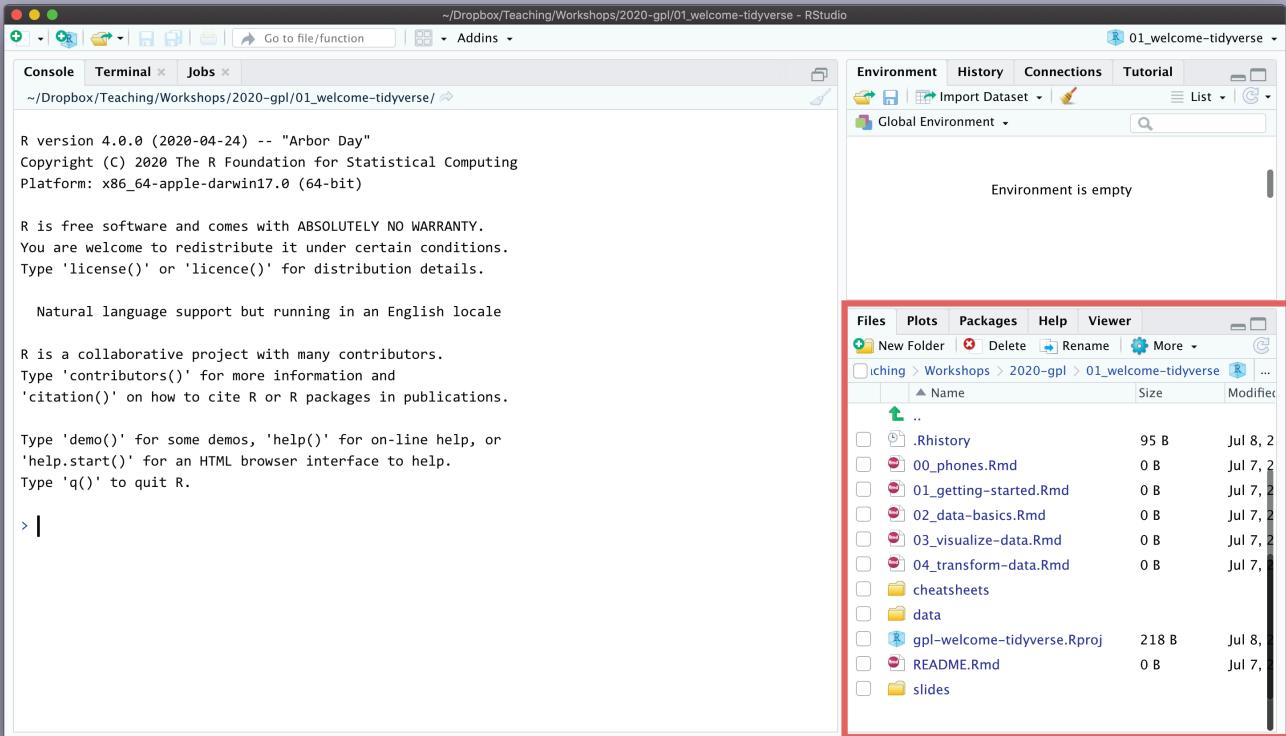
**On peut aussi assigner le résultat à une variable
pour l'utiliser subséquemment**

Panneau avec fichiers



Tous les fichiers dans
votre répertoire de
travail

À votre tour



Trouvez `a-intro.Rmd`

Cliquez sur le nom pour ouvrir le fichier

Panneau source

The screenshot shows the RStudio interface. On the left, the code editor displays the contents of the file '01_getting-started.Rmd'. The code includes R Markdown syntax like backticks for code chunks and a library call to 'tidyverse'. The right side of the interface shows the 'Console' window, which displays the initial R startup message and help text.

```
1: ---  
2: title: "Getting Started with R and RStudio"  
3: output: html_document  
4: ---  
5:  
6: `r setup`  
7: library(tidyverse)  
8:  
9:  
10: ## R Markdown  
11:  
12: This is an [R Markdown](http://rmarkdown.rstudio.com) file (it has a .Rmd file extension).  
When you execute code within the file, the results appear beneath the code.  
13:  
14: R code goes in **code chunks**, denoted by three backticks. Try executing this chunk by  
1:4: # Getting Started with R and RStudio : R Markdown :  
Console Terminal Jobs  
~/Dropbox/Teaching/Workshops/2020-gpl/01_welcome-tidyverse/ ~  
  
R is a collaborative project with many contributors.  
Type 'contributors()' for more information and  
'citation()' on how to cite R or R packages in publications.  
  
Type 'demo()' for some demos, 'help()' for on-line help, or  
'help.start()' for an HTML browser interface to help.  
Type 'q()' to quit R.  
> |
```

Les documents
s'ouvrent ici

Paquets

Utiliser des paquets

```
install.packages("nom")
```

```
library(nom)
```

Télécharge les fonctionnalités
sur votre ordinateur.

Une fois par ordinateur.

Fichiers téléchargés via le CRAN.

Charge le paquet

Effectuer une fois par session

À votre tour

Installez le paquet remotes

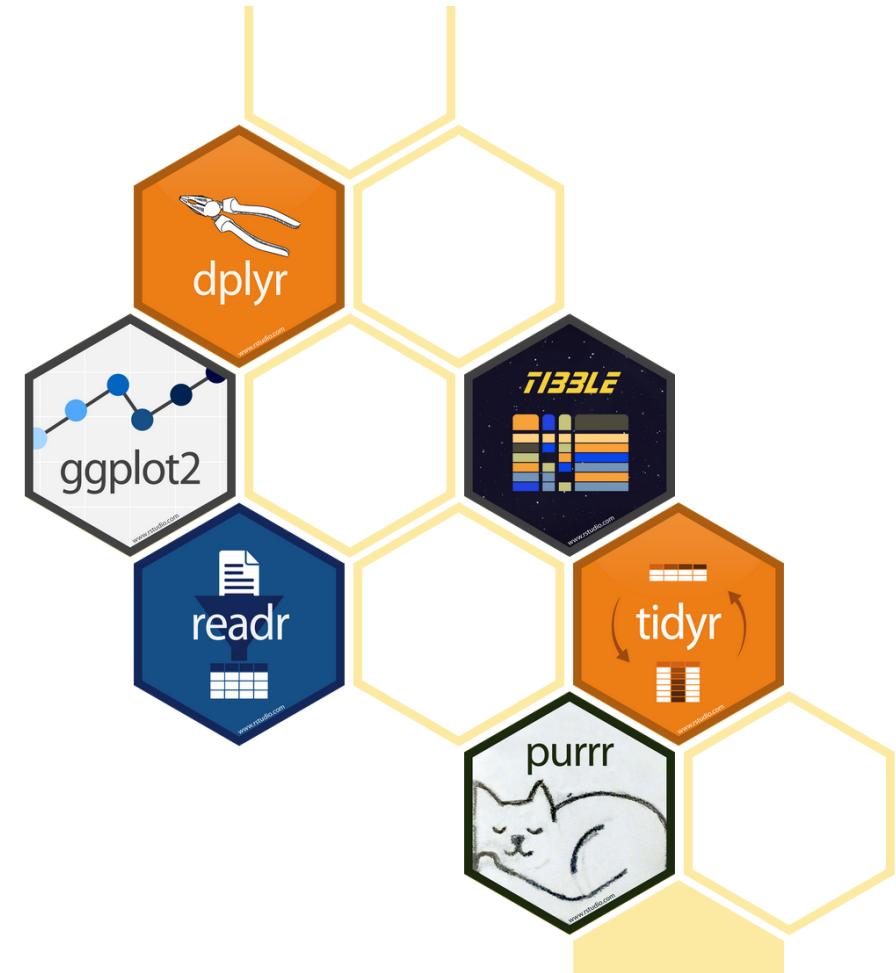
Utilisez les commandes suivantes pour installer hecmulti:

```
remotes::install_github("lbelzile/hecmulti")
```

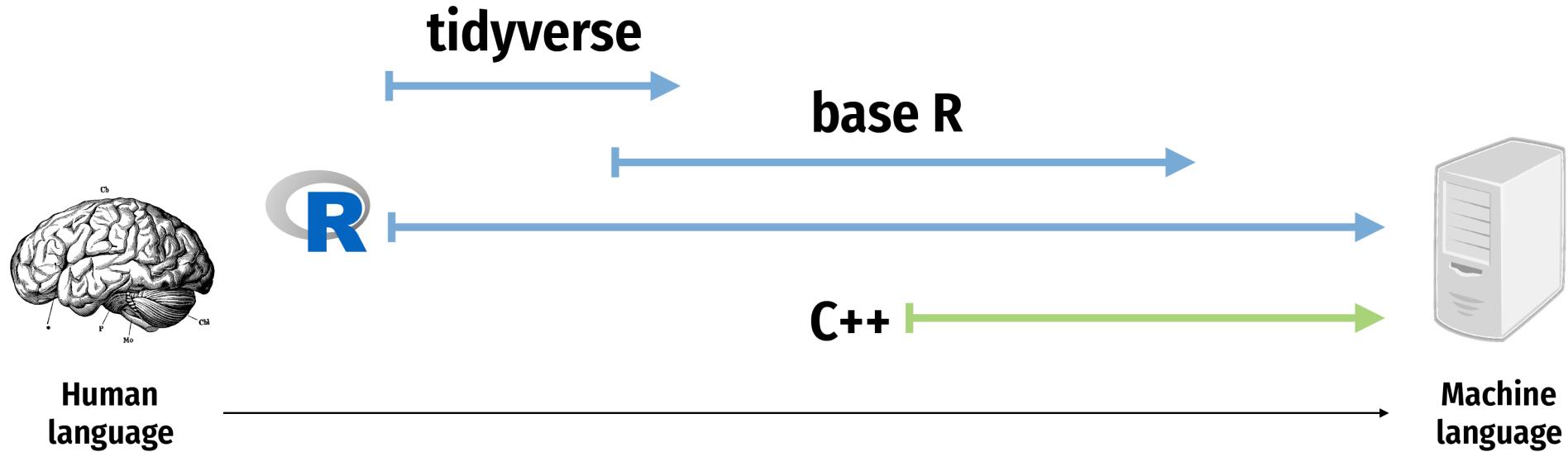
02:00

tidyverse

« Le tidyverse est une collection de paquets R conçus par des programmeurs opiniâtre pour la science des données. Tous les paquets partagent la même philosophie de conception, et ont une grammaire et une structure commune. »



tidyverse



Paquet tidyverse

```
library(tidyverse)
```

**Le paquet `tidyverse` est un raccourci
pour installer et charger les paquets clés du `tidyverse`**

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

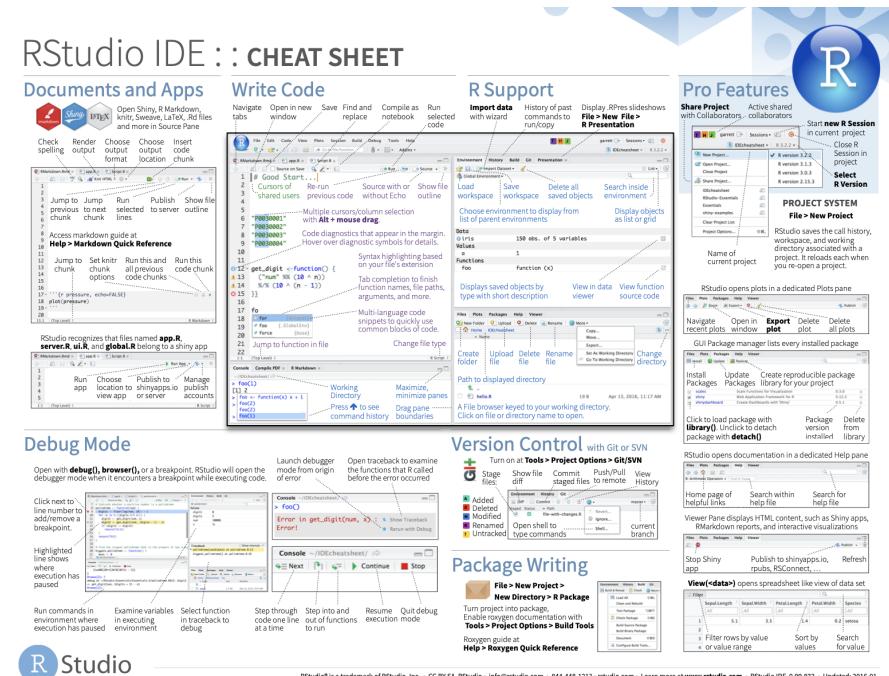
```
install.packages("ggplot2")
install.packages("dplyr")
install.packages("tidyverse")
install.packages("readr")
install.packages("purrr")
install.packages("tibble")
install.packages("stringr")
install.packages("forcats")
install.packages("lubridate")
install.packages("hms")
install.packages("DBI")
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("tidyverse")
library("readr")
library("purrr")
library("tibble")
library("stringr")
library("forcats")
```

Aides-mémoires

Allez à Help > Cheatsheets pour trouver des guides de référence sur certains différents paquets



The RStudio IDE Cheat Sheet is a comprehensive guide to various features of the RStudio interface. It is organized into several sections:

- Documents and Apps**: Shows icons for Open Shiny, R Markdown, knitr, Sweave, LaTeX, Rd files, and more. It includes a "Help > Markdown Quick Reference" section.
- Write Code**: A screenshot of the RStudio code editor showing various keyboard shortcuts and navigation features.
- R Support**: A screenshot of the RStudio Help menu showing options like Import data with wizard, History of past commands, and Run selected code.
- Pro Features**: A screenshot of the RStudio Project System showing features like Share Project with Collaborators, Start new R Session in current project, and Select R Version.
- Debug Mode**: A screenshot of the RStudio debugger interface showing a stack trace and error details.
- Version Control**: A screenshot of the RStudio Git interface showing stages like Add, Commit, Push/Pull, and View.
- Package Writing**: A screenshot of the RStudio Build Tools interface showing options like File > New Project > New Directory > R Package.
- View(<data>)**: A screenshot of the RStudio View pane showing a data frame with columns like Sepal.Length, Sepal.Width, Petal.Length, Petal.Width, and Species.

The bottom of the sheet features the RStudio logo and the text: "RStudio® is a trademark of RStudio, Inc. • CC BY SA RStudio • info@rstudio.com • 844-448-1222 • rstudio.com • Learn more at www.rstudio.com • RStudio IDE 0.99.832 • Updated: 2016-01".

Paquets et fonctionnalités

Charger un paquet ajoute à l'environnement toutes ses fonctions et données.

Peut masquer des fonctions homonymes!

```
> library(tidyverse)
— Attaching packages ━━━━━━━━━━ tidyverse 1.3.2 ━━━━━
✓ ggplot2 3.3.6      ✓ purrr    0.3.4
✓ tibble   3.1.7      ✓ stringr  1.4.0
✓ tidyr    1.2.0      ✓forcats  0.5.1
✓ readr    2.1.2
— Conflicts ━━━━━━━━━━ tidyverse_conflicts() ━━━
✖ dplyr::filter() masks stats::filter()
✖ dplyr::lag()    masks stats::lag()
```

Fonctionalités sans chargement

Charger une base de données

```
data(bd, package = "nomdupaquet")
```

Utiliser directement une fonction

```
paquet::fonction(...)
```

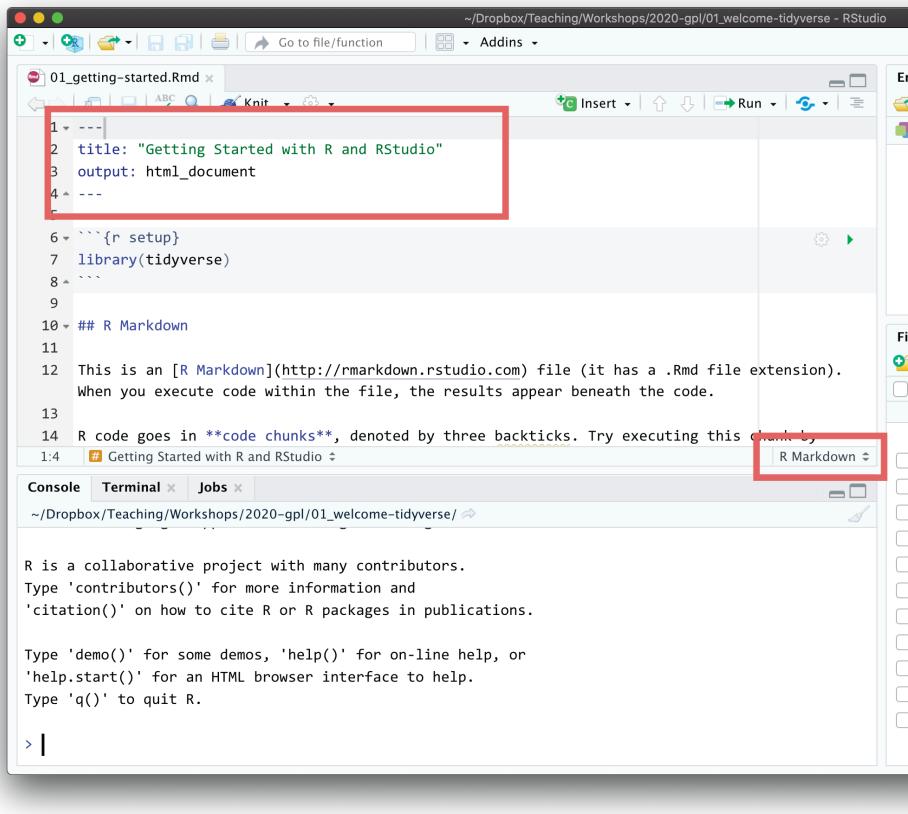
À votre tour

Chargez les données renfe du paquet hecmulti

```
data(renfe, package = 'hecmulti')  
renfe
```

01:00

R Markdown



```
1 ---  
2 title: "Getting Started with R and RStudio"  
3 output: html_document  
4 ---  
5   
6 ```{r setup}  
7 library(tidyverse)  
8 ---  
9   
10 ## R Markdown  
11   
12 This is an [R Markdown](http://rmarkdown.rstudio.com) file (it has a .Rmd file extension).  
When you execute code within the file, the results appear beneath the code.  
13   
14 R code goes in **code chunks**, denoted by three backticks. Try executing this chunk!  
1:4 # Getting Started with R and RStudio
```

R is a collaborative project with many contributors.
Type 'contributors()' for more information and
'citation()' on how to cite R or R packages in publications.

Type 'demo()' for some demos, 'help()' for on-line help, or
'help.start()' for an HTML browser interface to help.
Type 'q()' to quit R.

Format de document
qui combine texte et code

Un livret
pour votre analyse

R Markdown

The screenshot shows the RStudio interface. The top bar includes tabs for 'File', 'Edit', 'View', 'Tools', 'Help', and 'Knit'. Below the tabs are buttons for 'Insert', 'Run', and other functions. The main area displays an R Markdown file named '01_getting-started.Rmd'. A red box highlights the following text:

```
10 ## R Markdown
11
12 This is an [R Markdown](http://rmarkdown.rstudio.com) file (it has a .Rmd file extension).
13 When you execute code within the file, the results appear beneath the code.
14 R code goes in **code chunks**, denoted by three backticks. Try executing this chunk by
15 clicking the *Run* button (a small green triangle) within the chunk or by placing your cursor
16 inside it and pressing *Ctrl+Shift+Enter* (or *Cmd+Shift+Enter* on Mac).
17
18 ````{r}
19 ggplot(data = mpg) +
20   geom_point(mapping = aes(x = cty, y = hwy), alpha = 0.2)
21 ````
```

Below the code, a ggplot2 scatter plot is displayed. The x-axis is labeled 'cty' and the y-axis is labeled 'hwy'. The plot shows a positive correlation between fuel economy in the city (cty) and on the highway (hwy). Most data points are clustered between 10 and 45 on both axes, with a few outliers at higher values.

Texte

R Markdown

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

```
## R Markdown
This is an [R Markdown](http://rmarkdown.rstudio.com) file (it has a .Rmd file extension). When you execute code within the file, the results appear beneath the code.
R code goes in **code chunks**, denoted by three backticks. Try executing this chunk by clicking the *Run* button (a small green triangle) within the chunk or by placing your cursor inside it and pressing *Ctrl+Shift+Enter* (or *Cmd+Shift+Enter* on Mac).
```{r}
ggplot(data = mpg +
 geom_point(mapping = aes(x = cty, y = hwy), alpha = 0.2)
```

```

A red box highlights the code chunk from line 16 to 19. Below the code editor is a ggplot2 scatter plot showing the relationship between city fuel economy (cty) on the x-axis and highway fuel economy (hwy) on the y-axis. The plot includes a light gray grid and semi-transparent points.

Texte
Code

R Markdown

The screenshot shows the RStudio interface. The top bar displays the file name "01_getting-started.Rmd". The code editor contains the following R Markdown code:

```
10 ## R Markdown
11
12 This is an [R Markdown](http://rmarkdown.rstudio.com) file (it has a .Rmd file extension).
13 When you execute code within the file, the results appear beneath the code.
14 R code goes in **code chunks**, denoted by three backticks. Try executing this chunk by
15 clicking the *Run* button (a small green triangle) within the chunk or by placing your cursor
16 inside it and pressing *Ctrl+Shift+Enter* (or *Cmd+Shift+Enter* on Mac).
17
18 ```{r}
19 ggplot(data = mpg +
20   geom_point(mapping = aes(x = cty, y = hwy), alpha = 0.2)
21 ```

A red box highlights the ggplot2 scatter plot output. The plot shows the relationship between city fuel economy (cty) on the x-axis and highway fuel economy (hwy) on the y-axis. The data points are semi-transparent grey circles, with a few black ones at the bottom left.
```

Texte

Code

Sortie

À votre tour

The screenshot shows the RStudio interface with an R Markdown file open. The code chunk at line 16 is highlighted with a red box:

```
10 ## R Markdown
11
12 This is an [R Markdown](http://rmarkdown.rstudio.com) file (it has a .Rmd file extension).
When you execute code within the file, the results appear beneath the code.
13
14 R code goes in **code chunks**, denoted by three backticks. Try executing this chunk by
clicking the *Run* button (a small green triangle) within the chunk or by placing your cursor
inside it and pressing *Ctrl+Shift+Enter* (or *Cmd+Shift+Enter* on Mac).
15
16 ```{r}
17 ggplot(data = mpg +
18   geom_point(mapping = aes(x = cty, y = hwy), alpha = 0.2)
19 ```



Below the code, a scatter plot is displayed with 'hwy' on the y-axis and 'cty' on the x-axis. The plot shows a positive correlation between fuel efficiency.


```

Lisez les instructions

Compilez le bloc de code
avec le bouton démarrer

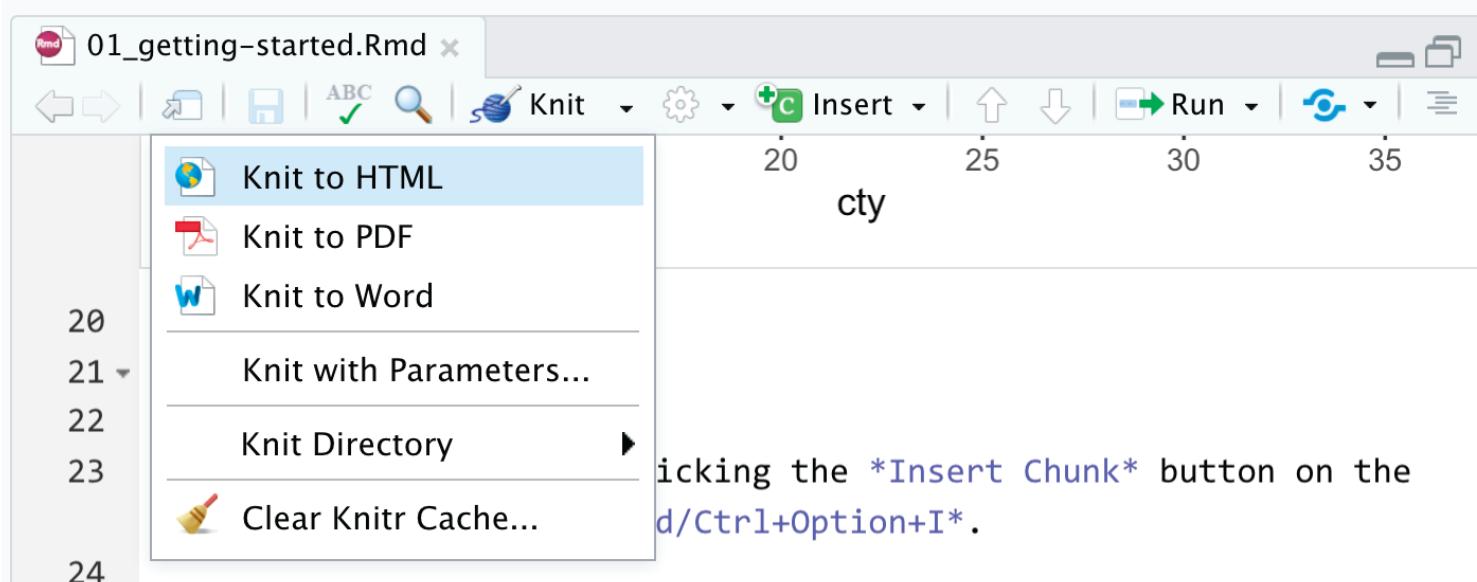
À votre tour

Ajoutez un nouveau bloc

Inscrivez $2 + 2$ et compilez

Tricoter

"Knit" un document R Markdown en un fichier portable



R Markdown

Combiner code R et texte

Nous l'utiliserons en classe:

Je fournis un script

Vous complétez les sections « À votre tour »

À la fin, vous aurez une copie annotée

À votre tour

Quelle est la différence entre:

```
filter(mtcars, cyl == 4)
```

```
four_cyls <- filter(mtcars, cyl == 4)
```

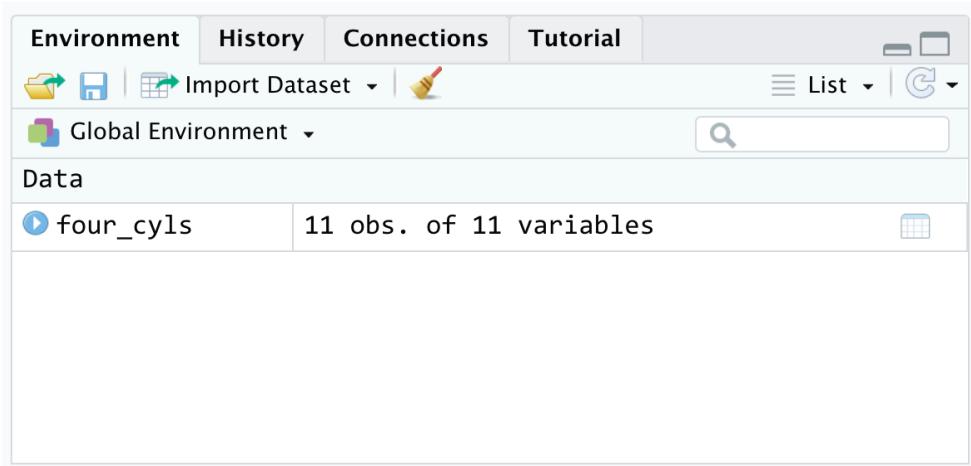
**Trouvez ces instructions dans votre carnet et compilez.
Quelle est la différence en sortie?**

Assignation

<- assigne la sortie à droite à la variable de gauche

```
four_cyls <- filter(mtcars, cyl == 4)
```

Panneau environnement



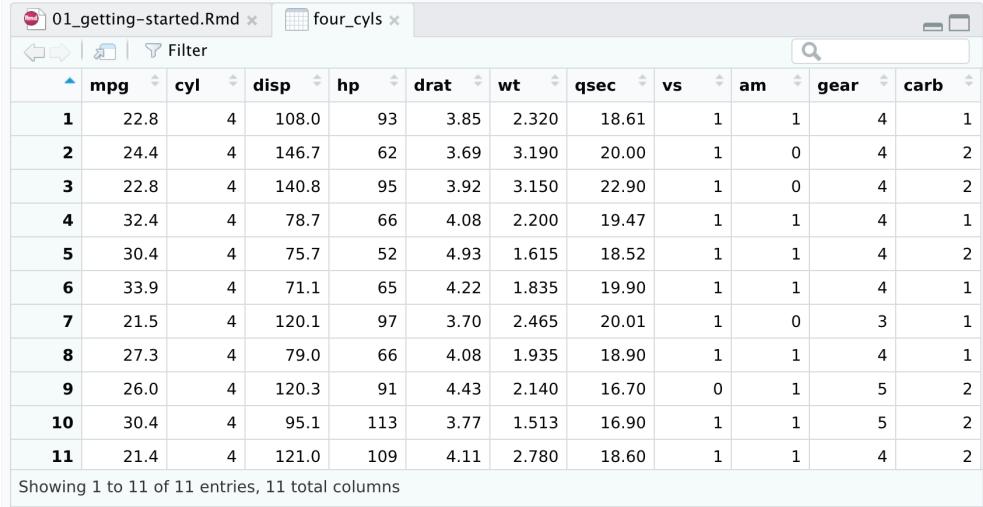
Liste de toutes les variables existantes

À votre tour

**Trouvez `four_cyls` dans le panneau environnement.
Cliquez sur le nom `four_cyls`**

Qu'est-ce qui arrive?

Visualisation



The screenshot shows the RStudio interface with two tabs open: '01_getting-started.Rmd' and 'four_cyls'. The 'four_cyls' tab is active, displaying a data viewer with the 'mtcars' dataset. The data is presented as a table with 11 rows and 11 columns. The columns are labeled: mpg, cyl, disp, hp, drat, wt, qsec, vs, am, gear, carb. The rows are numbered 1 through 11. The data includes various car specifications like miles per gallon, engine displacement, horsepower, and number of cylinders.

| | mpg | cyl | disp | hp | drat | wt | qsec | vs | am | gear | carb |
|----|------|-----|-------|-----|------|-------|-------|----|----|------|------|
| 1 | 22.8 | 4 | 108.0 | 93 | 3.85 | 2.320 | 18.61 | 1 | 1 | 4 | 1 |
| 2 | 24.4 | 4 | 146.7 | 62 | 3.69 | 3.190 | 20.00 | 1 | 0 | 4 | 2 |
| 3 | 22.8 | 4 | 140.8 | 95 | 3.92 | 3.150 | 22.90 | 1 | 0 | 4 | 2 |
| 4 | 32.4 | 4 | 78.7 | 66 | 4.08 | 2.200 | 19.47 | 1 | 1 | 4 | 1 |
| 5 | 30.4 | 4 | 75.7 | 52 | 4.93 | 1.615 | 18.52 | 1 | 1 | 4 | 2 |
| 6 | 33.9 | 4 | 71.1 | 65 | 4.22 | 1.835 | 19.90 | 1 | 1 | 4 | 1 |
| 7 | 21.5 | 4 | 120.1 | 97 | 3.70 | 2.465 | 20.01 | 1 | 0 | 3 | 1 |
| 8 | 27.3 | 4 | 79.0 | 66 | 4.08 | 1.935 | 18.90 | 1 | 1 | 4 | 1 |
| 9 | 26.0 | 4 | 120.3 | 91 | 4.43 | 2.140 | 16.70 | 0 | 1 | 5 | 2 |
| 10 | 30.4 | 4 | 95.1 | 113 | 3.77 | 1.513 | 16.90 | 1 | 1 | 5 | 2 |
| 11 | 21.4 | 4 | 121.0 | 109 | 4.11 | 2.780 | 18.60 | 1 | 1 | 4 | 2 |

Showing 1 to 11 of 11 entries, 11 total columns

Cliquer sur un objet dans le panneau environnement l'ouvre dans un onglet de visualisation interactive.

Objets, classes et types

Vecteurs

Objet de base dans R

Un vecteur est une liste ordonnée de valeurs de même type
(texte, nombres, etc.)

Concaténez avec `c()`:

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

Types de données de base

| | | |
|------------------------------|---------------|---|
| Entier (integer) | Nombre entier | c(1L, 2L, 3L, 4L); 1:4 |
| Double | Nombres | c(1, 2.4, 3.14, 4) |
| Caractère (character) | Texte | c("1", "bleu", "plaisir", "monstre") |
| Logique (logical) | Vrai ou faux | c(TRUE, FALSE, TRUE, FALSE) |
| Facteur (factor) | Catégorielle | c("En désaccord", "D'accord", "Neutre") |

Utiliser la fonction `typeof()`
pour déterminer le type d'un vecteur.

Fonctions

Fonctions égal actions

Arguments en intrant, résultats en sortie

Spécifier les arguments avec nomarg =, sinon en ordre

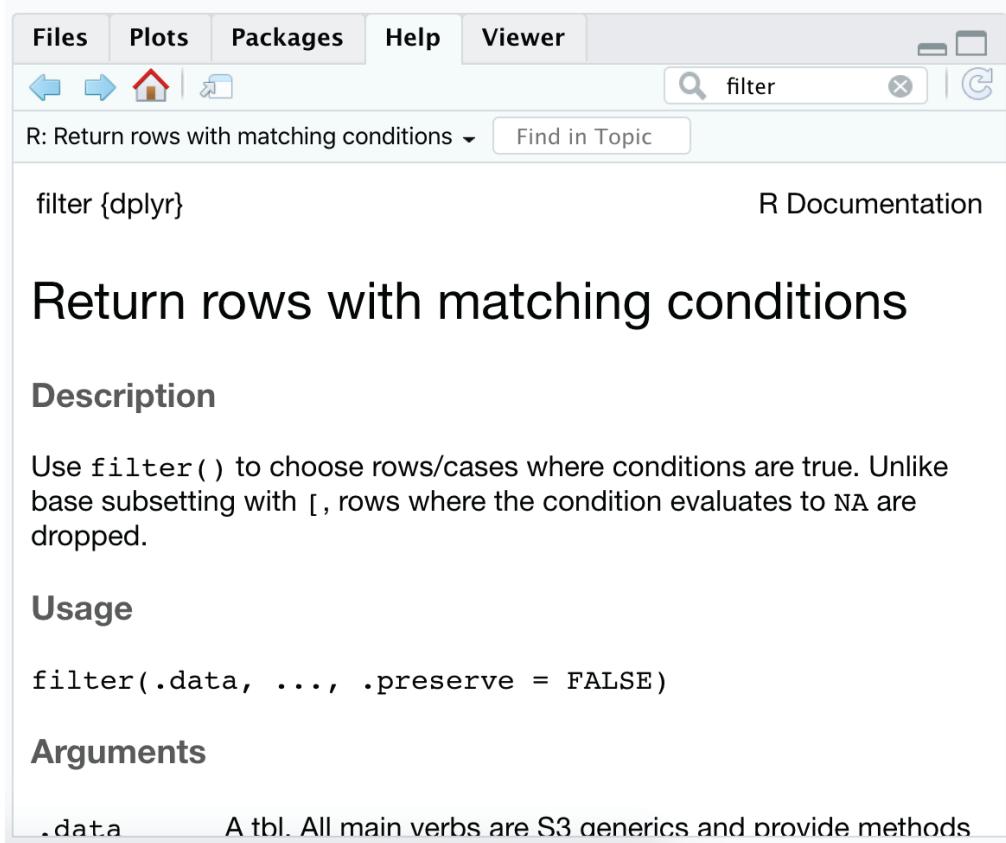
Aide

**Pour voir la page d'aide sur une fonction R,
tapez dans la console:**

```
?nom_fonction  
# Fonction pas dans l'environnement ou chaîne de caractères  
??"normal distribution"
```

(ou utilisez un moteur de recherche!)

Panneau d'aide



The screenshot shows the R Help Viewer interface. The title bar includes 'Files', 'Plots', 'Packages', 'Help', and 'Viewer' tabs. A search bar at the top right contains the word 'filter'. Below the search bar, a dropdown menu says 'R: Return rows with matching conditions' and a 'Find in Topic' button. The main content area is titled 'filter {dplyr}' and 'R Documentation'. It starts with the text 'Return rows with matching conditions'. Below this are sections for 'Description' and 'Usage'. The 'Usage' section contains the code 'filter(.data, ..., .preserve = FALSE)'. Under 'Arguments', there is a table with one row. The first column is '.data' and the second column is 'A tbl. All main verbs are S3 generics and provide methods'. The entire screenshot is framed by a light gray border.

filter {dplyr} R Documentation

Return rows with matching conditions

Description

Use `filter()` to choose rows/cases where conditions are true. Unlike base subsetting with `[`, rows where the condition evaluates to NA are dropped.

Usage

```
filter(.data, ..., .preserve = FALSE)
```

Arguments

| | |
|-------|---|
| .data | A <code>tbl</code> . All main verbs are S3 generics and provide methods |
|-------|---|

Ces fichiers contiennent notamment des détails sur les arguments des fonctions

Il y a souvent des exemples à la fin

À votre tour

Lisez l'aide pour la fonction seq

Ajouter un bloc qui utilise seq() pour créer un vecteur de nombres de 5 à 30, espacés par sauts de 5
(5, 10, 15, 20, ...)

02 : 00

```
seq(from = 5, to = 30, by = 5)
```

```
## [1] 5 10 15 20 25 30
```

Problèmes fréquents de syntaxe #1

Parenthèses ou accolades de fermeture manquante

```
mean(mtcars
```

```
"Oops problème
```

Problèmes fréquents de syntaxe #2

Ajouter des accolades quand ce n'est pas adéquat (et vice versa)

```
mean("mtcars")
```

```
## Warning in mean.default("mtcars"): l'argument n'est ni  
## numérique, ni logique : renvoi de NA
```

```
## [1] NA
```

À votre tour

Il y a trois blocs sous "Syntaxe fautive"

Compilez chacune, lisez le message d'erreur et essayez de corriger le code

Listes

Une liste est une collection d'objets

```
maliste <- list(elementun = 2:4, FALSE)  
maliste
```

```
## $elementun  
## [1] 2 3 4  
##  
## [[2]]  
## [1] FALSE
```

Format de sortie fréquent pour les fonctions!

Éléments d'une liste

**Accéder à un élément
par nom avec \$**

```
maliste$elementun
```

**Accéder à un élément
par position avec [[]]**

```
maliste[[1]]
```

Classes et méthodes

La classe d'un objet permet de définir des fonctions génériques (`methods`)

```
methods(class = "lm")
```

Autrement dit, le résultat dépend de la classe.

Méthodes fréquentes

impression
(`print`)

récapitulatif
(`summary`)

graphique
(`plot`)

Données rectangulaires

Bases de données et tableaux

Les bases de données `data.frame` sont utilisées pour stocker les données (comme un chiffrier)

`data.frame`: une liste de variables de type potentiellement différents

`tibble (tbl)`: version `tidyverse`, plus sophistiquée

Importer des données

Paquets pour l'importation de données



Données texte brutes

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



Chiffriers Excel

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



Données Stata, SPSS, et SAS

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