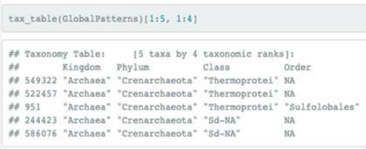
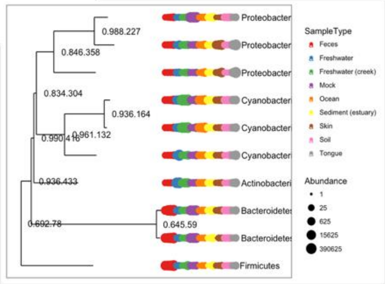
R and Rstudio :Introduction and getting started



```
tax_table(GlobalPatterns)[1:5, 1:4]
```

```
## Taxonomy Table:     [5 taxa by 4 taxonomic ranks]:
##        Kingdom   Phylum          Class          Order
## 549322 "Archaea" "Crenarchaeota" "Thermoprotei" NA
## 522457 "Archaea" "Crenarchaeota" "Thermoprotei" NA
## 951    "Archaea" "Crenarchaeota" "Thermoprotei" "Sulfolobales"
## 244423 "Archaea" "Crenarchaeota" "Sd-NA"        NA
## 586076 "Archaea" "Crenarchaeota" "Sd-NA"        NA
```



What is cran R?

R?

Statistic
Data science

Interpreted
languages

Simple syntax

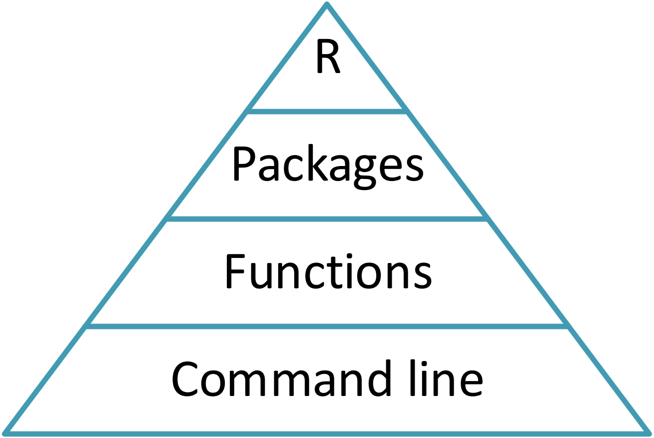
Free
Muti-OS

Gorgeous
graph output

Reproducibility

Large
community

What is cran R?



R

Packages

Functions

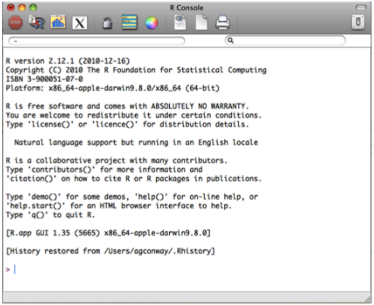
Command line

What is cran R?

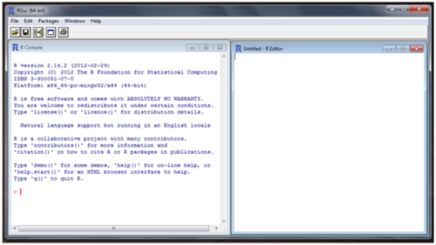
Here are some examples of the core packages that are commonly included with R:

1. **base** - This is the base R package that provides essential functions and data types.
2. **datasets** - This package contains various datasets that are frequently used for examples and testing.
3. **graphics** - Provides functions for creating graphical plots and charts.
4. **stats** - Includes statistical functions and models, such as linear regression and hypothesis testing.
5. **...**

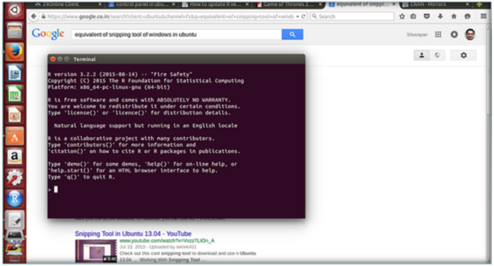
What is cran R?



```
R version 2.12.1 (2010-12-16)
Copyright (C) 2010 The R Foundation for Statistical Computing
ISBN 3-900051-07-0
Platform: x86_64-apple-darwin9.8.0/x86_64 (64-bit)

R is free software and comes with ABSOLUTELY NO WARRANTY.
You are welcome to redistribute it under certain conditions.
Type 'license()' or 'licence()' for distribution details.

  Natural language support but running in an English locale

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.

[R.app GUI 1.35 (5665) x86_64-apple-darwin9.8.0]

[History restored from /Users/agconway/.Rhistory]

>
```



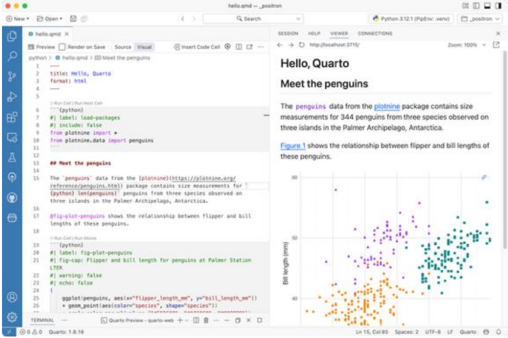
```
R version 2.14.2 (2012-02-29)
Copyright (C) 2012 The R Foundation for Statistical Computing
ISBN 3-900051-07-0
Platform: x86_64-pc-mingw32/x64 (64-bit)

R is free software and comes with ABSOLUTELY NO WARRANTY.
You are welcome to redistribute it under certain conditions.
Type 'license()' or 'licence()' for distribution details.

  Natural language support but running in an English locale

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.

>
```



```
R version 3.2.2 (2015-08-14) -- "Fire Safety"
Copyright (C) 2015 The R Foundation for Statistical Computing
Platform: x86_64-pc-linux-gnu (64-bit)

R is free software and comes with ABSOLUTELY NO WARRANTY.
You are welcome to redistribute it under certain conditions.
Type 'license()' or 'licence()' for distribution details.

  Natural language support but running in an English locale

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.

>
```

Rstudio

Integrated development environment (IDE)



~~~
---
title: Hello, Quarto
format: html
---

```{python}
#| label: load-packages
#| include: false
from plotnine import *
from plotnine.data import penguins
~~~

Meet the penguins

The `penguins` data from the [plotnine](https://plotnine.org/
reference/penguins.html) package contains size measurements for `
{python} len(penguins)` penguins from three species observed on
three islands in the Palmer Archipelago, Antarctica.

@fig-plot-penguins shows the relationship between flipper and bill
lengths of these penguins.

```{python}
#| label: fig-plot-penguins
#| fig-cap: Flipper and bill length for penguins at Palmer Station
LTER
#| warning: false
#| echo: false
(
    ggplot(penguins, aes(x="flipper_length_mm", y="bill_length_mm"))
    + geom_point(aes(color="species", shape="species"))
```

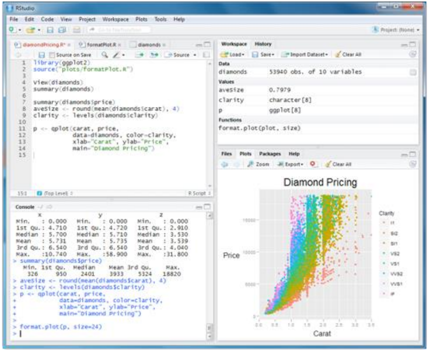
Hello, Quarto

Meet the penguins

The penguins data from the plotnine package contains size measurements for 344 penguins from three species observed on three islands in the Palmer Archipelago, Antarctica.

Figure 1 shows the relationship between flipper and bill lengths of these penguins.

Rstudio



```
library(ggplot2)
source("plots/formatPlot.R")

View(diamonds)
summary(diamonds)

summary(diamonds$price)
aveSize <- round(mean(diamonds$carat), 4)
clarity <- levels(diamonds$clarity)

p <- qplot(carat, price,
           data=diamonds, color=clarity,
           xlab="Carat", ylab="Price",
           main="Diamond Pricing")
```

Data	
diamonds	53940 obs. of 10 variables
Values	
aveSize	0.7979
clarity	character[8]
p	ggplot[8]
Functions	
format.plot(plot, size)	

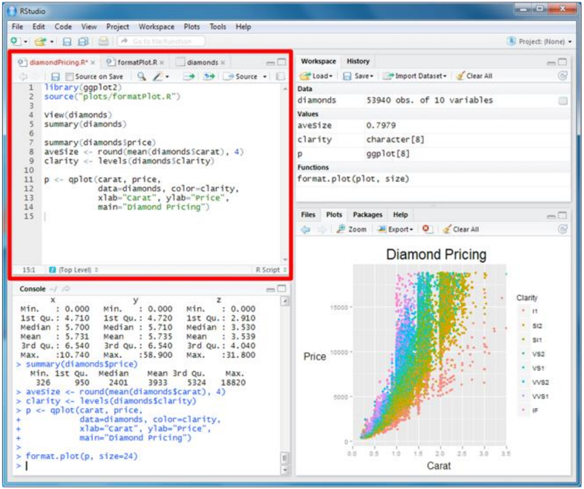
```
       x                y                z
 Min.   : 0.000   Min.   : 0.000   Min.   : 0.000
 1st Qu.: 4.710   1st Qu.: 4.720   1st Qu.: 2.910
 Median : 5.700   Median : 5.710   Median : 3.530
 Mean   : 5.731   Mean   : 5.735   Mean   : 3.539
 3rd Qu.: 6.540   3rd Qu.: 6.540   3rd Qu.: 4.040
 Max.   :10.740   Max.   :58.900   Max.   :31.800
> summary(diamonds$price)
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
    326     950    2401    3933    5324   18820
> aveSize <- round(mean(diamonds$carat), 4)
> clarity <- levels(diamonds$clarity)
> p <- qplot(carat, price,
+           data=diamonds, color=clarity,
+           xlab="Carat", ylab="Price",
+           main="Diamond Pricing")
>
> format.plot(p, size=24)
>
```

Positron

https://positron.posit.co/migrate-rstudio.html

Rstudio

An editing window opens where you can write and save a series of instructions in a file: your script!



```r
library(ggplot2)
source("plots/formatPlot.R")

view(diamonds)
summary(diamonds)

summary(diamonds$price)
aveSize <- round(mean(diamonds$carat), 4)
clarity <- levels(diamonds$clarity)

p <- qplot(carat, price,
           data=diamonds, color=clarity,
           xlab="Carat", ylab="Price",
           main="Diamond Pricing")

```

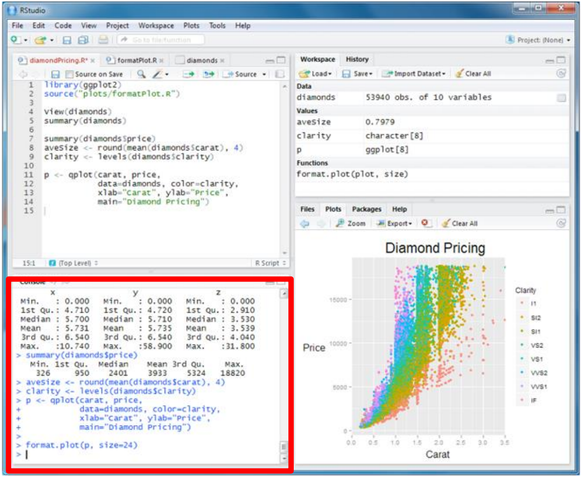
```
       x                y                z         
 Min.   : 0.000   Min.   : 0.000   Min.   : 0.000  
 1st Qu.: 4.710   1st Qu.: 4.720   1st Qu.: 2.910  
 Median : 5.700   Median : 5.710   Median : 3.530  
 Mean   : 5.731   Mean   : 5.735   Mean   : 3.539  
 3rd Qu.: 6.540   3rd Qu.: 6.540   3rd Qu.: 4.040  
 Max.   :10.740   Max.   :58.900   Max.   :31.800  
> summary(diamonds$price)
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max. 
    326     950    2401    3933    5324   18820 
> aveSize <- round(mean(diamonds$carat), 4)
> clarity <- levels(diamonds$clarity)
> p <- qplot(carat, price,
+            data=diamonds, color=clarity,
+            xlab="Carat", ylab="Price",
+            main="Diamond Pricing")
>
> format.plot(p, size=24)
> 
```

Data	
diamonds	53940 obs. of 10 variables
Values	
aveSize	0.7979
clarity	character[8]
p	ggplot[8]
Functions	
format.plot(plot, size)	

Institut Méditerranéen d'Océanologie

Rstudio

A terminal window giving all the output from the script



```r
library(ggplot2)
source("plots/formatPlot.R")

view(diamonds)
summary(diamonds)

summary(diamonds$price)
aveSize <- round(mean(diamonds$carat), 4)
clarity <- levels(diamonds$clarity)

p <- qplot(carat, price,
           data=diamonds, color=clarity,
           xlab="Carat", ylab="Price",
           main="Diamond Pricing")

```

```
       x               y                z
 Min.   : 0.000   Min.   : 0.000   Min.   : 0.000
 1st Qu.: 4.710   1st Qu.: 4.720   1st Qu.: 2.910
 Median : 5.700   Median : 5.710   Median : 3.530
 Mean   : 5.731   Mean   : 5.735   Mean   : 3.539
 3rd Qu.: 6.540   3rd Qu.: 6.540   3rd Qu.: 4.040
 Max.   :10.740   Max.   :58.900   Max.   :31.800
> summary(diamonds$price)
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
    326     950    2401    3933    5324   18820
> aveSize <- round(mean(diamonds$carat), 4)
> clarity <- levels(diamonds$clarity)
> p <- qplot(carat, price,
+            data=diamonds, color=clarity,
+            xlab="Carat", ylab="Price",
+            main="Diamond Pricing")
>
> format.plot(p, size=24)
> 
```

Data

diamonds 53940 obs. of 10 variables

Values

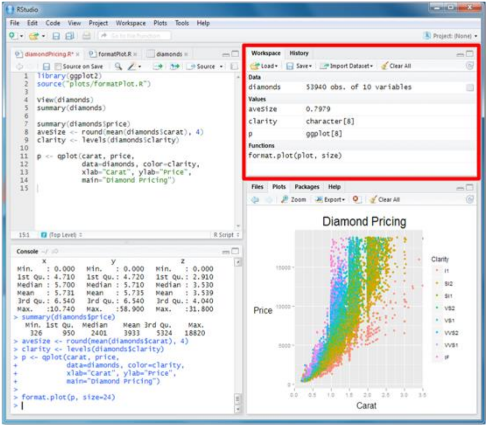
aveSize	0.7979
clarity	character[8]
p	ggplot[8]

Functions

format.plot(plot, size)

Rstudio

A **Workspace** in which all the objects created/generated by the code are displayed (*i.e.* Matrices, Vectors, Functions, Constants...). It includes a **History** tab which returns all the executed instructions.



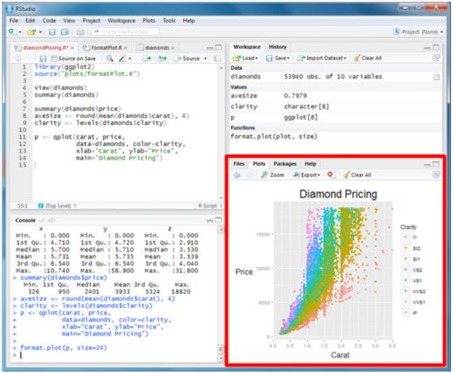
```
library(ggplot2)
source("plots/formatPlot.R")

view(diamonds)
summary(diamonds)

summary(diamonds$price)
aveSize <- round(mean(diamonds$carat), 4)
clarity <- levels(diamonds$clarity)

p <- qplot(carat, price,
           data=diamonds, color=clarity,
           xlab="Carat", ylab="Price",
           main="Diamond Pricing")

```

```
       x               y                z
 Min.   : 0.000   Min.   : 0.000   Min.   : 0.000
 1st Qu.: 4.710   1st Qu.: 4.720   1st Qu.: 2.910
 Median : 5.700   Median : 5.710   Median : 3.530
 Mean   : 5.731   Mean   : 5.735   Mean   : 3.539
 3rd Qu.: 6.540   3rd Qu.: 6.540   3rd Qu.: 4.040
 Max.   :10.740   Max.   :58.900   Max.   :31.800
> summary(diamonds$price)
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
    326     950    2401    3933    5324   18820
> aveSize <- round(mean(diamonds$carat), 4)
> clarity <- levels(diamonds$clarity)
> p <- qplot(carat, price,
+            data=diamonds, color=clarity,
+            xlab="Carat", ylab="Price",
+            main="Diamond Pricing")
>
> format.plot(p, size=24)
>
```

```
Data
diamonds        53940 obs. of 10 variables
Values
aveSize         0.7979
clarity         character[8]
p               ggplot[8]
Functions
format.plot(plot, size)
```

Rstudio

This last window contains 4 tabs :
Files (locates your working directory and allows you to navigate),
Plots (returns all graphic outputs),
Packages (package management)
Help (this is R's 112 or 911, contains all of R's help)



```
library(ggplot2)
source("plots/formatPlot.R")

View(diamonds)
summary(diamonds)

summary(diamonds$price)
aveSize <- round(mean(diamonds$carat), 4)
clarity <- levels(diamonds$clarity)

p <- qplot(carat, price,
           data=diamonds, color=clarity,
           xlab="Carat", ylab="Price",
           main="Diamond Pricing")
```

```
       x                y                z
 Min.   : 0.000   Min.   : 0.000   Min.   : 0.000
 1st Qu.: 4.710   1st Qu.: 4.720   1st Qu.: 2.910
 Median : 5.700   Median : 5.710   Median : 3.530
 Mean   : 5.731   Mean   : 5.735   Mean   : 3.539
 3rd Qu.: 6.540   3rd Qu.: 6.540   3rd Qu.: 4.040
 Max.   :10.740   Max.   :58.900   Max.   :31.800
> summary(diamonds$price)
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
    326     950    2401    3933    5324   18820
> aveSize <- round(mean(diamonds$carat), 4)
> clarity <- levels(diamonds$clarity)
> p <- qplot(carat, price,
+            data=diamonds, color=clarity,
+            xlab="Carat", ylab="Price",
+            main="Diamond Pricing")
>
> format.plot(p, size=24)
> |
```

```
Data
diamonds      53940 obs. of 10 variables
Values
aveSize       0.7979
clarity       character[8]
p             ggplot[8]
Functions
format.plot(plot, size)
```

The commandes on R:

An R command is either an expression or an assignment :

- Expression is directly evaluated, and the result is displayed on terminal

  Exemple :

```
>2+3
[1]5
>sqrt(25)
[1]5
```

Mathematical calcul

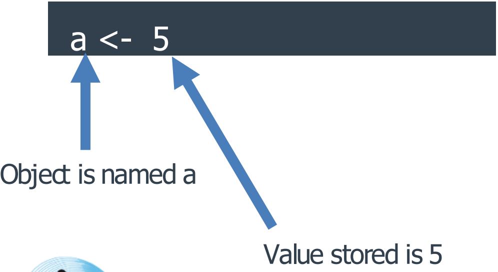
The commandes on R:

An R command is either an expression or an assignment :

- An assignement is an expression stored in object or variable

Exemple :

```
>a<-5
>a
[1]5
```



```
a <-  5
```

Object is named a

Value stored is 5

Sign used for d'affectation

<- ou =

Why store?

To manipulate later

The commandes on R:

An R command is either an expression or an assignment :

- An assignement is an expression stored in object or variable

Exemple :

```
>a<-5
>a
[1]5
```

```
a <-  5
class(a)
[1] Num
```

It is very important to know what type of object

The commandes on R:

An R command is either an expression or an assignment :

- Object can be :
  Character : (”A”, ”B’’, “C”, ..)
  Numeric : (1,2, 5.5, 6.7, …)
  Logic : (TRUE, FALSE)

- But also more complex:
  Vector
  Matrix
  Data frame
  List

The commandes on R:

Object Vector

- Vector contain only element with same mode (ex : numeric or character or logic ..)

```
Age = c(24, 26, 32, 54)
```

C() is the function to build a vector

Object type = Vector

There are the affected elements

The commandes on R:

What is the structure of my object ?

Function class() and str()

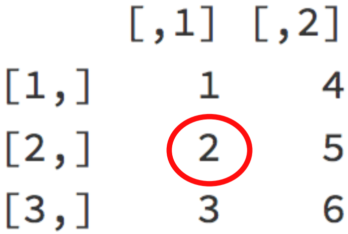
```
Age = c(24, 26, 32, 54)
str(Age)
Num[1:4]
class(Age)
Numeric
```

During the training we will see, how can we choose element inside a Vector or filter a vector

The commandes on R:

Object Matrix

- Matrix contains several vectors that build a table with columns and rows. The vectors belong to the same mode.



```
     [,1] [,2]
[1,]    1    4
[2,]    2    5
[3,]    3    6
```

The commandes on R:

Object data frame

- Data frame contains several vectors that build a table with columns and rows. The vectors can belong to different modes

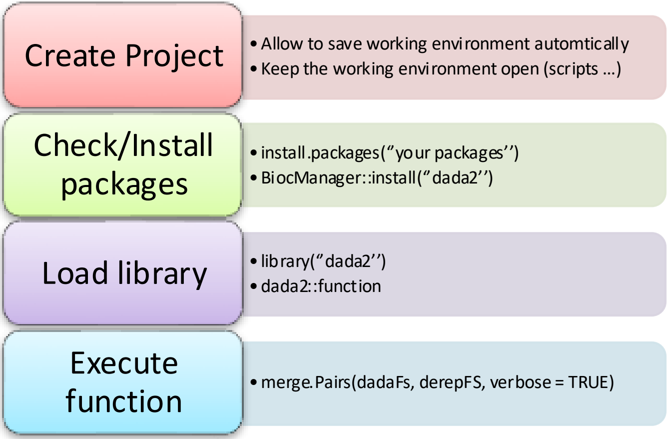
colnames

rownames

	Nom	Prenom	Age	Region_naissance
A	Keïta	Modibo	34	1
B	Traoré	Moussa	23	2
C	Konaré	Oumar	21	2
D	Touré	Toumani	43	4
E	Keïta	Boubacar	54	3

row.names() : row name (i.e. sample name)
colnames() : header(Nom, Prenom, etc)

The Workflow



Create Project

- Allow to save working environment automtically
- Keep the working environment open (scripts …)

Check/Install packages

- install.packages(“your packages’’)
- BiocManager::install(“dada2’’)

Load library

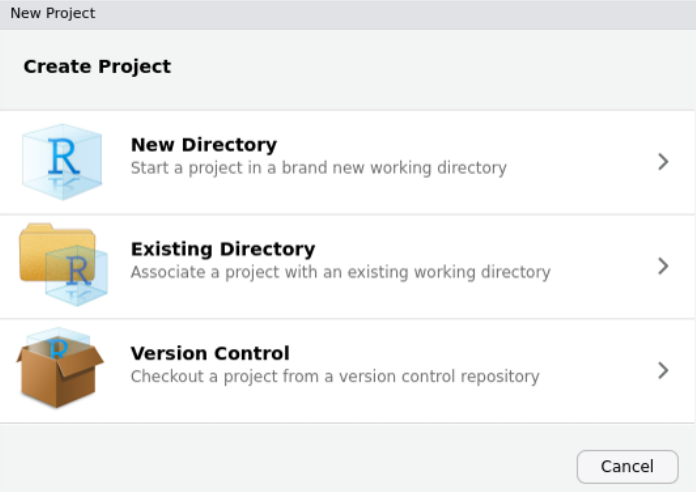
- library(“dada2’’)
- dada2::function

Execute function

- merge.Pairs(dadaFs, derepFS, verbose = TRUE)

Work with projects with RStudio

Work with projects with RStudio



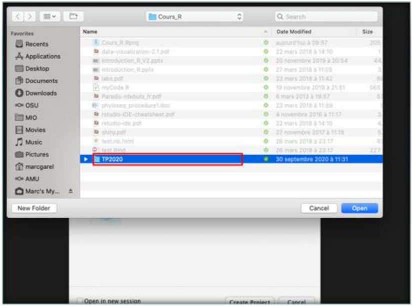
Create Project

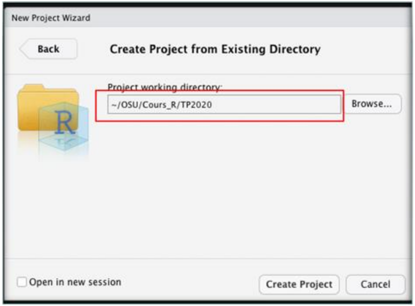
New Directory
Start a project in a brand new working directory

Existing Directory
Associate a project with an existing working directory

Version Control
Checkout a project from a version control repository

Work with projects with RStudio

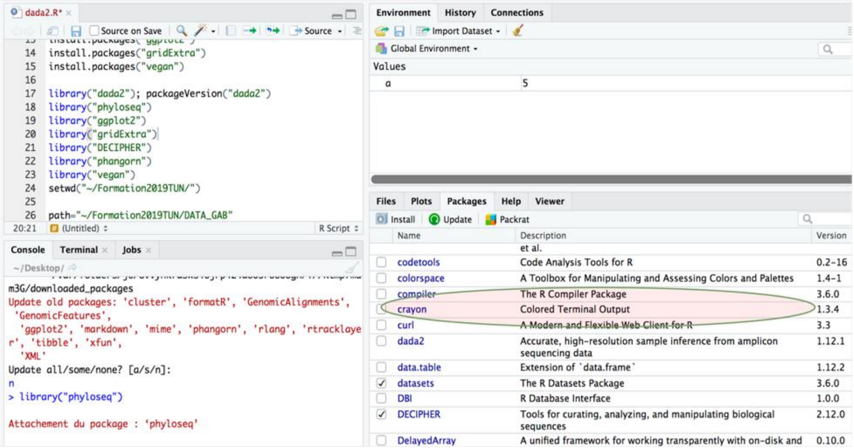




Create Project from Existing Directory

Project working directory: ~/OSU/Cours_R/TP2020

Check install packages



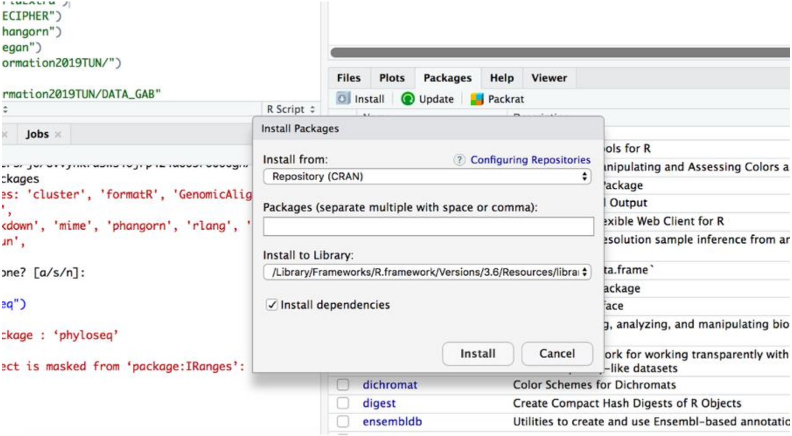
```r
install.packages("gridExtra")
install.packages("vegan")

library("dada2"); packageVersion("dada2")
library("phyloseq")
library("ggplot2")
library("gridExtra")
library("DECIPHER")
library("phangorn")
library("vegan")
setwd("~/Formation2019TUN/")

path="~/Formation2019TUN/DATA_GAB"
```

Values	
a	5

```
m3G/downloaded_packages
Update old packages: 'cluster', 'formatR', 'GenomicAlignments',
 'GenomicFeatures',
  'ggplot2', 'markdown', 'mime', 'phangorn', 'rlang', 'rtracklaye
r', 'tibble', 'xfun',
  'XML'
Update all/some/none? [a/s/n]:
n
> library("phyloseq")

Attachement du package : ‘phyloseq’
```

	Name	Description	Version
		et al.	
☐	codetools	Code Analysis Tools for R	0.2-16
☐	colorspace	A Toolbox for Manipulating and Assessing Colors and Palettes	1.4-1
☐	compiler	The R Compiler Package	3.6.0
☐	crayon	Colored Terminal Output	1.3.4
☐	curl	A Modern and Flexible Web Client for R	3.3
☐	dada2	Accurate, high-resolution sample inference from amplicon sequencing data	1.12.1
☐	data.table	Extension of `data.frame`	1.12.2
☑	datasets	The R Datasets Package	3.6.0
☐	DBI	R Database Interface	1.0.0
☑	DECIPHER	Tools for curating, analyzing, and manipulating biological sequences	2.12.0
☐	DelayedArray	A unified framework for working transparently with on-disk and	0.10.0

Check install packages :CRAN package



```
ECIPHER")
hangorn")
egan")
ormation2019TUN/")

rmation2019TUN/DATA_GAB"
```

```
ckages
es: 'cluster', 'formatR', 'GenomicAlig
',
kdown', 'mime', 'phangorn', 'rlang', '
un',

one? [a/s/n]:

eq")

ckage : 'phyloseq'

ect is masked from 'package:IRanges':
```

Install Packages

Install from: Configuring Repositories

Repository (CRAN)

Packages (separate multiple with space or comma):

Install to Library:

/Library/Frameworks/R.framework/Versions/3.6/Resources/librar

☑ Install dependencies

		ols for R
		nipulating and Assessing Colors a
		ackage
		Output
		xible Web Client for R
		solution sample inference from ar
		ta.frame`
		ackage
		ace
		g, analyzing, and manipulating bio
		ork for working transparently with -like datasets
☐	dichromat	Color Schemes for Dichromats
☐	digest	Create Compact Hash Digests of R Objects
☐	ensembldb	Utilities to create and use Ensembl-based annotatio

Check install packages :Bioconductor package

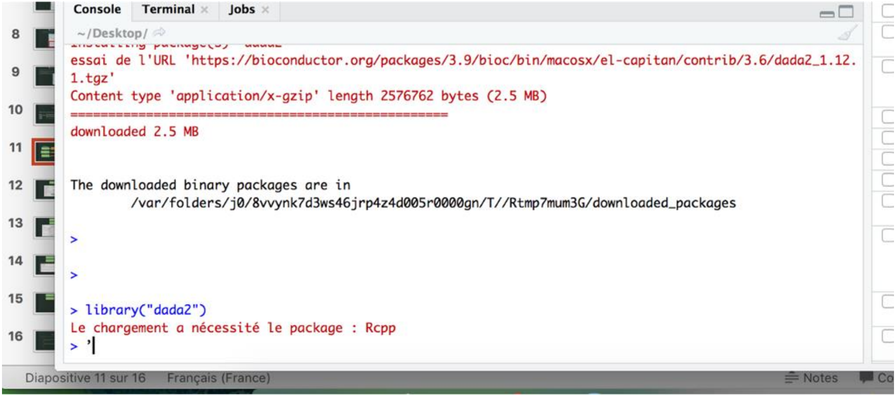
Command line is mandatory

```
library("ggplot2")
library("gridExtra")
library("DECIPHER")
library("phangorn")
library("vegan")
setwd("~/Formation2019TUN/")

path="~/Formation2019TUN/DATA_GAB"
```

```
>
>
> BiocManager::install("dada2", version = "3.9")
Bioconductor version 3.9 (BiocManager 1.30.4), R 3.6.0 (2019-04-26)
Installing package(s) 'dada2'
essai de l'URL 'https://bioconductor.org/packages/3.9/bioc/bin/macosx/el-capitan/contrib/3.6/dada2_1.12.
1.tgz'
Content type 'application/x-gzip' length 2576762 bytes (2.5 MB)
==================================================
downloaded 2.5 MB
```

	Name	Description
☐	fastmatch	Fast match() fu
☐	foreach	Provides Foreac for R
☐	foreign	Read Data Store 'SAS', 'SPSS', 'St 'dBase', ...
☐	formatR	Format R Code
☐	Formula	Extended Mode
☐	futile.logger	A Logging Utilit
☐	futile.options	Futile Options
☐	gdata	Various R Progr Data Manipulati
☐	GenomeInf...	Utilities for ma chromosome n

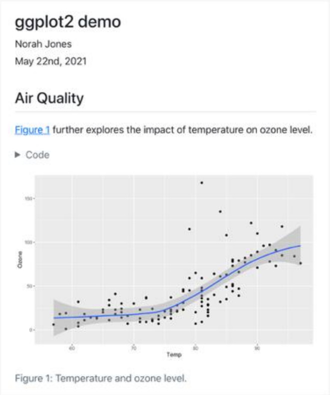
Load library to execute function



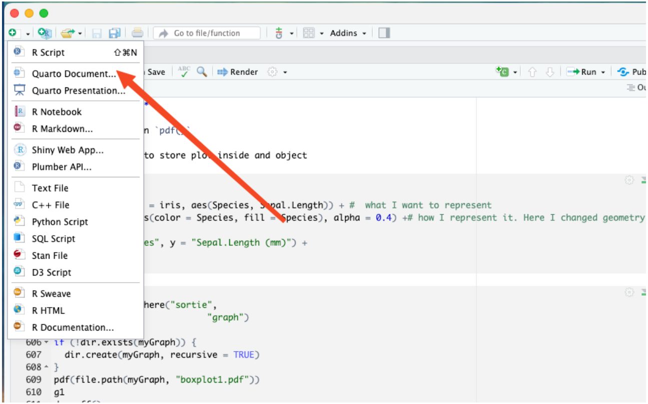
```
essai de l'URL 'https://bioconductor.org/packages/3.9/bioc/bin/macosx/el-capitan/contrib/3.6/dada2_1.12.
1.tgz'
Content type 'application/x-gzip' length 2576762 bytes (2.5 MB)
==================================================
downloaded 2.5 MB


The downloaded binary packages are in
        /var/folders/j0/8vvynk7d3ws46jrp4z4d005r0000gn/T//Rtmp7mum3G/downloaded_packages

>

>

> library("dada2")
Le chargement a nécessité le package : Rcpp
> '
```

Or use command as : package::function()

R Quarto

~~~
---
title: "ggplot2 demo"
author: "Norah Jones"
date: "5/22/2021"
format:
  html:
    code-fold: true
---

## Air Quality

@fig-airquality further explores the impact of temperature on ozone level.

```{r}
#| label: fig-airquality
#| fig-cap: "Temperature and ozone level."
#| warning: false

library(ggplot2)

ggplot(airquality, aes(Temp, Ozone)) +
  geom_point() +
  geom_smooth(method = "loess")
~~~
```



## ggplot2 demo

Norah Jones

May 22nd, 2021

## Air Quality

Figure 1 further explores the impact of temperature on ozone level.

▶ Code

Figure 1: Temperature and ozone level.

An R Quarto, is a document with chunks that can be executed independently and interactively, with output visible immediately beneath the input. It is more flexible than Markdown
```

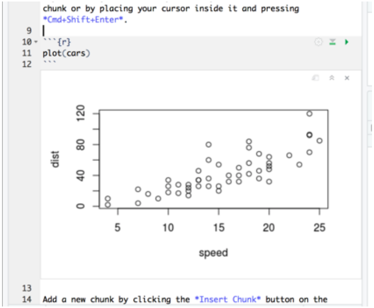
R Quarto



```
n `pdf()`

to store plot inside and object

 = iris, aes(Species, Sepal.Length)) + #  what I want to represent
s(color = Species, fill = Species), alpha = 0.4) +# how I represent it. Here I changed geometry

es", y = "Sepal.Length (mm)") +

here("sortie",
                "graph")
if (!dir.exists(myGraph)) {
  dir.create(myGraph, recursive = TRUE)
}
pdf(file.path(myGraph, "boxplot1.pdf"))
g1
```

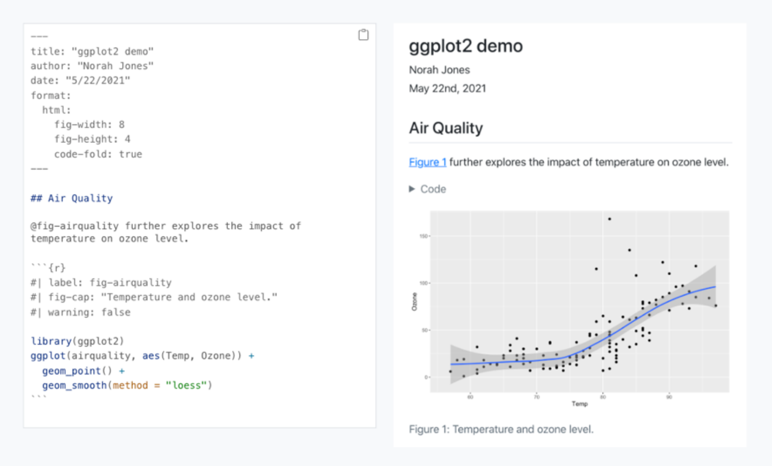
Creating a Quarto instead of the classical script

- You can create a new notebook in RStudio with the menu commandFile → *NewFile* → *Quarto document.*
- Notebook chunks can be inserted quickly using the keyboard shortcut Ctrl + Alt + I (macOS: Cmd + Option + I), or via the Insert menu in the editor toolbar.
- Between chunks, you can comment all your results.

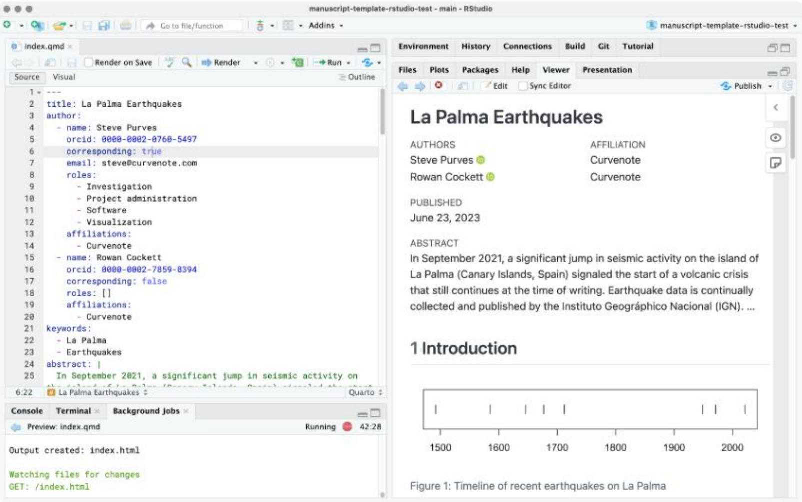


~~~
chunk or by placing your cursor inside it and pressing
*Cmd+Shift+Enter*.

```{r}
plot(cars)
~~~
```

dist
0
40
80
120
5
10
15
20
25
speed

```
Add a new chunk by clicking the *Insert Chunk* button on the
```
```

# R Quarto



~~~
---
title: "ggplot2 demo"
author: "Norah Jones"
date: "5/22/2021"
format:
  html:
    fig-width: 8
    fig-height: 4
    code-fold: true
---

## Air Quality

@fig-airquality further explores the impact of
temperature on ozone level.

```{r}
#| label: fig-airquality
#| fig-cap: "Temperature and ozone level."
#| warning: false

library(ggplot2)
ggplot(airquality, aes(Temp, Ozone)) +
  geom_point() +
  geom_smooth(method = "loess")
~~~
```

ggplot2 demo

Norah Jones

May 22nd, 2021

Air Quality

Figure 1 further explores the impact of temperature on ozone level.

▶ Code

Figure 1: Temperature and ozone level.

Institut Méditerranéen d'Océanologie
```



```
---
title: La Palma Earthquakes
author:
  - name: Steve Purves
    orcid: 0000-0002-0760-5497
    corresponding: true
    email: steve@curvenote.com
    roles:
      - Investigation
      - Project administration
      - Software
      - Visualization
    affiliations:
      - Curvenote
  - name: Rowan Cockett
    orcid: 0000-0002-7859-8394
    corresponding: false
    roles: []
    affiliations:
      - Curvenote
keywords:
  - La Palma
  - Earthquakes
abstract: |
  In September 2021, a significant jump in seismic activity on
```

```
Output created: index.html

Watching files for changes
GET: /index.html
```

La Palma Earthquakes

AUTHORS	AFFILIATION
Steve Purves	Curvenote
Rowan Cockett	Curvenote

PUBLISHED
June 23, 2023

ABSTRACT
In September 2021, a significant jump in seismic activity on the island of La Palma (Canary Islands, Spain) signaled the start of a volcanic crisis that still continues at the time of writing. Earthquake data is continually collected and published by the Instituto Geográphico Nacional (IGN). ...

1 Introduction

Figure 1: Timeline of recent earthquakes on La Palma

https://quarto.org/docs/manuscripts/authoring/rstudio.html

A demo before practice. . .