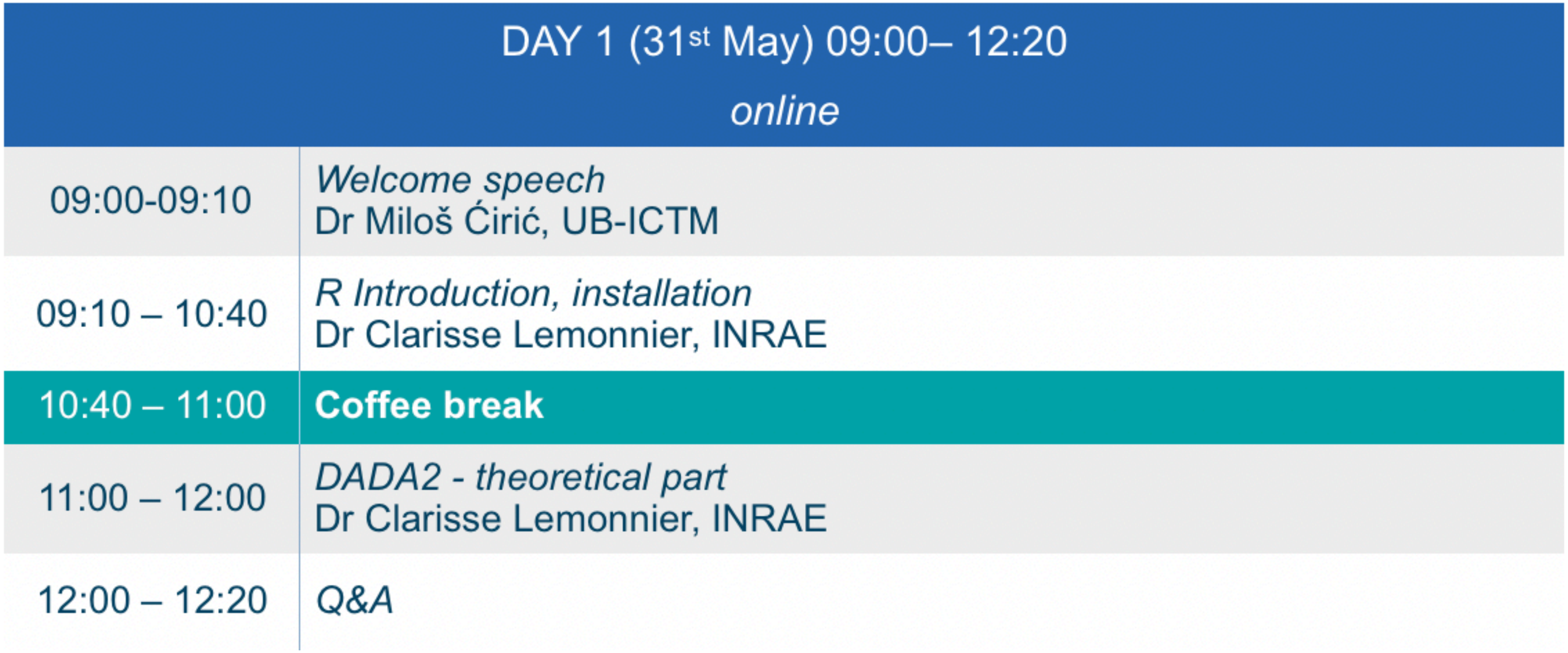
BIOLAWEB

Bioinformatic analysis of metabarcoding data using DADA2

Introduction to R

Clarisse Lemonnier

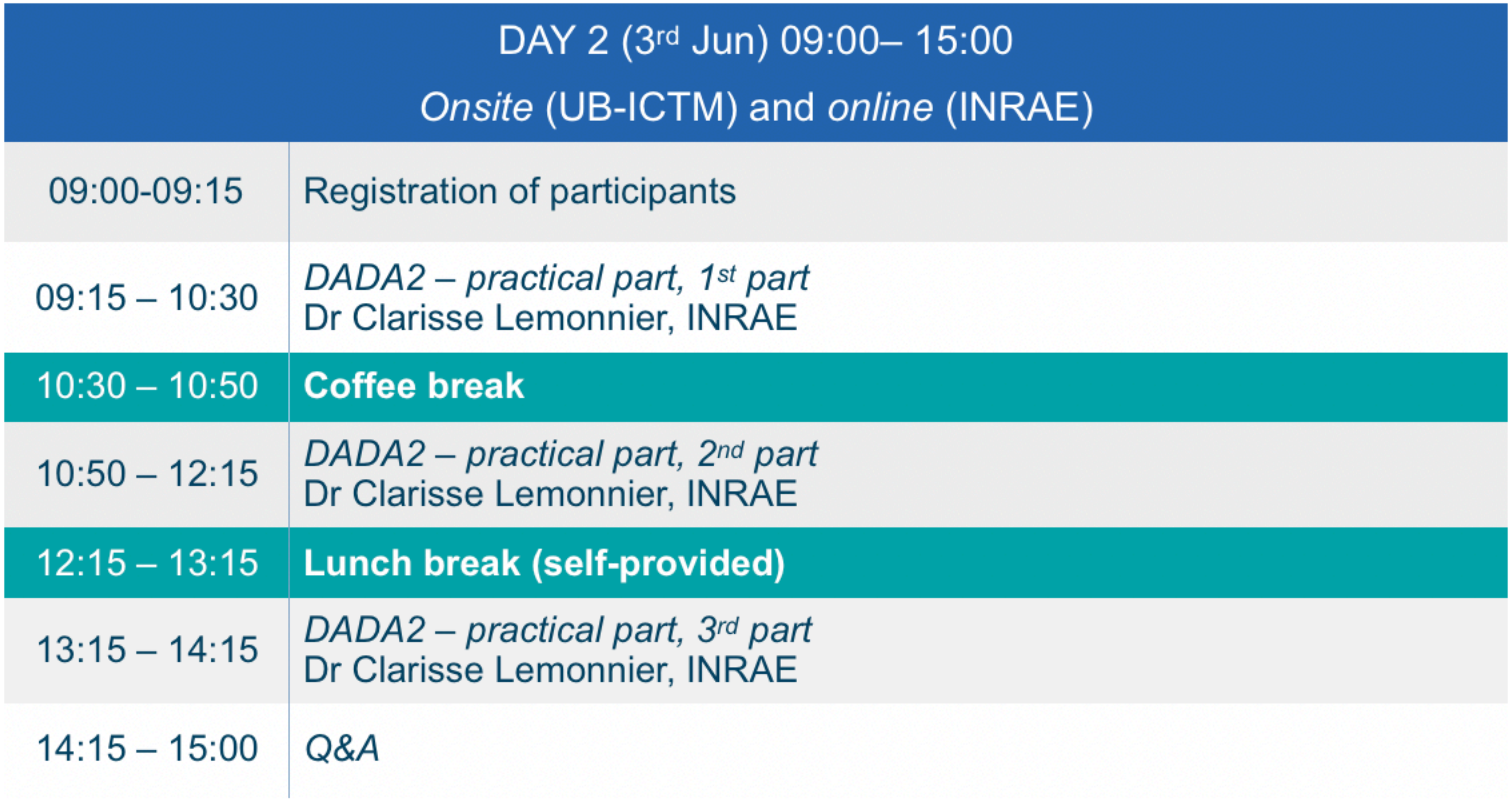
Program



DAY 1 (31st May) 09:00– 12:20 *online*	
09:00-09:10	*Welcome speech* Dr Miloš Ćirić, UB-ICTM
09:10 – 10:40	*R Introduction, installation* Dr Clarisse Lemonnier, INRAE
10:40 – 11:00	**Coffee break**
11:00 – 12:00	*DADA2 - theoretical part* Dr Clarisse Lemonnier, INRAE
12:00 – 12:20	*Q&A*

- See if everything is ready for the practical part
- Theory on metabarcoding pipeline

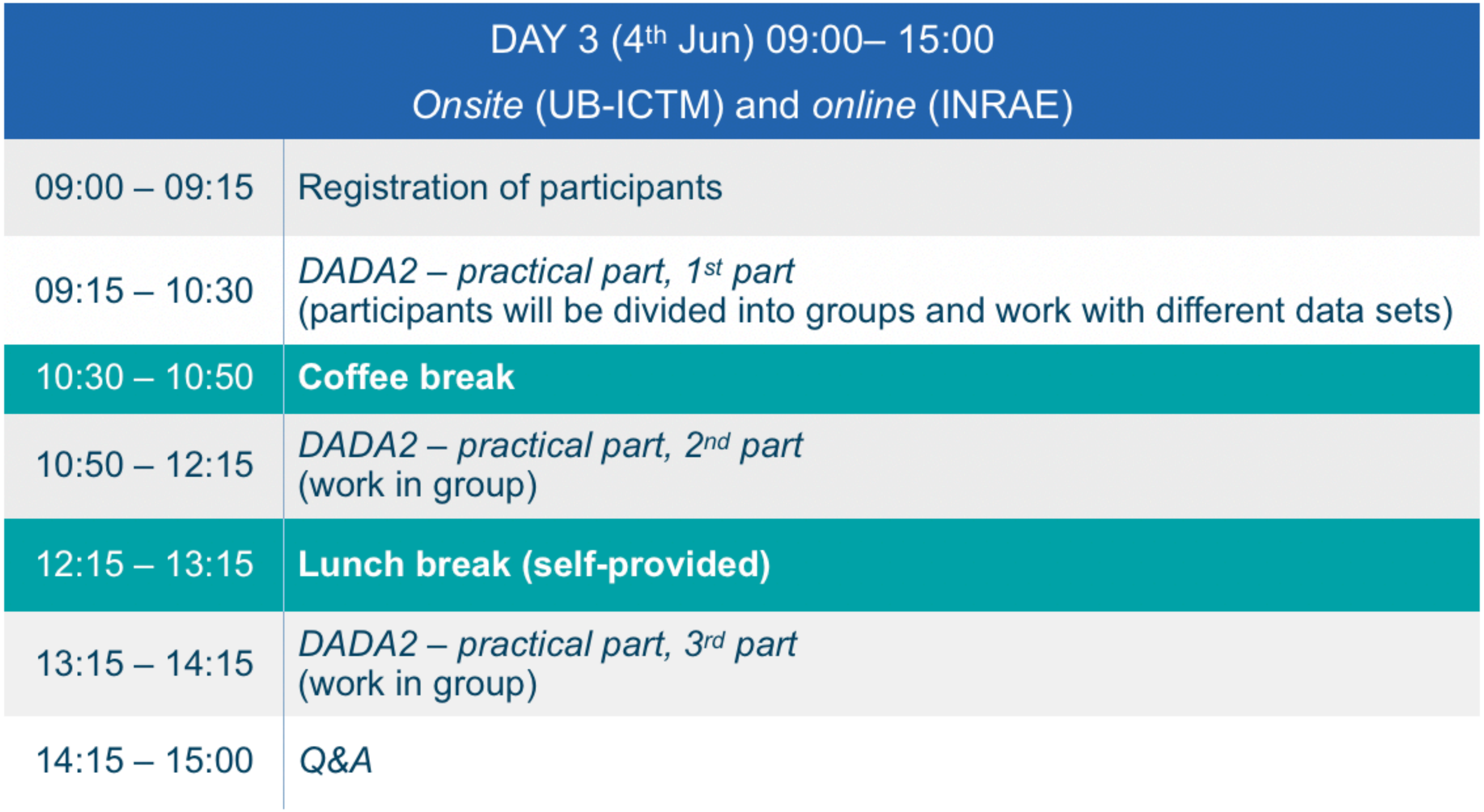
Program



DAY 2 (3rd Jun) 09:00– 15:00 *Onsite* (UB-ICTM) and *online* (INRAE)	
09:00-09:15	Registration of participants
09:15 – 10:30	*DADA2 – practical part, 1st part* Dr Clarisse Lemonnier, INRAE
10:30 – 10:50	**Coffee break**
10:50 – 12:15	*DADA2 – practical part, 2nd part* Dr Clarisse Lemonnier, INRAE
12:15 – 13:15	**Lunch break (self-provided)**
13:15 – 14:15	*DADA2 – practical part, 3rd part* Dr Clarisse Lemonnier, INRAE
14:15 – 15:00	*Q&A*

We will go through all the pipeline step by step
Listen carefully, because you will do it on real data on Tuesday !

Program



DAY 3 (4th Jun) 09:00– 15:00 *Onsite* (UB-ICTM) and *online* (INRAE)	
09:00 – 09:15	Registration of participants
09:15 – 10:30	*DADA2 – practical part, 1st part* (participants will be divided into groups and work with different data sets)
10:30 – 10:50	**Coffee break**
10:50 – 12:15	*DADA2 – practical part, 2nd part* (work in group)
12:15 – 13:15	**Lunch break (self-provided)**
13:15 – 14:15	*DADA2 – practical part, 3rd part* (work in group)
14:15 – 15:00	*Q&A*

You will be divided into 3 groups - one with a specific dataset
Diatom (rbcl) - Phytoplankton 16S - Phytoplankton - 23S

Program

DAY 4 (5th Jun) 09:00– 11:00 *Onsite* (UB-ICTM) and *online* (INRAE)	
09:00 – 09:15	Registration of participants
09:15 – 09:45	*Presentation of results* (Group 1)
09:45 – 10:05	**Coffee break**
10:05 – 10:35	*Presentation of results* (Group 2)
10:35 – 11:00	*Q&A*

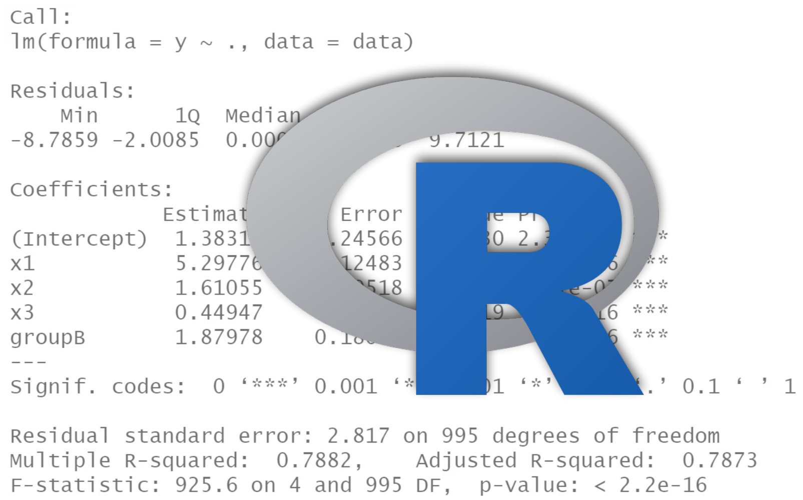
DAY 5 (6th Jun) 09:00– 11:00 *Onsite* (UB-ICTM) and *online* (INRAE)	
09:00 – 09:15	Registration of participants
09:15 – 09:45	*Presentation of results* (Group 3)
09:45 – 10:05	**Coffee break**
10:05 – 10:35	*Summarizing* Dr Clarisse Lemonnier, INRAE
10:35 – 11:00	*Q&A*

Each group will present the results they got at the end, the difficulties they had to run the dada2 pipeline

So don't forget to save figures, or take notes to make a little presentation

At the end there will be a little questionary :)

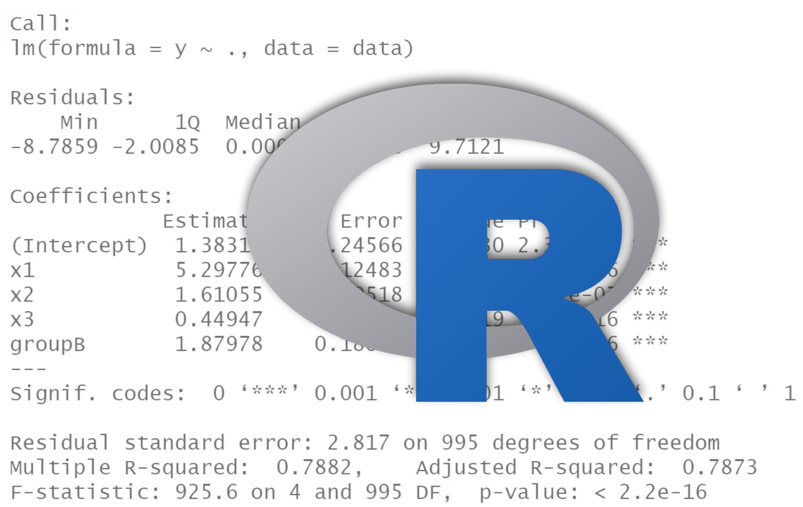
R use - basics



```
Call:
lm(formula = y ~ ., data = data)

Residuals:
    Min      1Q  Median         9.7121
-8.7859 -2.0085  0.00

Coefficients:
            Estimat      Error
(Intercept)  1.3831      .24566      30 2.          *
x1           5.29776      12483              6     **
x2           1.61055       518            e-0    ***
x3           0.44947                 19     16 ***
groupB       1.87978    0.18               6 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘*     01 ‘*’      .’ 0.1 ‘ ’ 1

Residual standard error: 2.817 on 995 degrees of freedom
Multiple R-squared:  0.7882,    Adjusted R-squared:  0.7873
F-statistic: 925.6 on 4 and 995 DF,  p-value: < 2.2e-16
```

What is R ?

- R is an **open source** programming **language**
- It was designed for **statistical computing**, **data analysis** and **graph display**
- It works in all operating system
- It is widely used in academic research



```
Call:
lm(formula = y ~ ., data = data)

Residuals:
    Min      1Q  Median                9.7121
-8.7859 -2.0085  0.00

Coefficients:
            Estimat       Error
(Intercept)  1.3831        .24566       30 2.          *
x1           5.29776        12483                    6  **
x2           1.61055         518                e-0   ***
x3           0.44947                  19         16  ***
groupB       1.87978    0.18                      6  ***
---
Signif. codes:  0 ‘***’ 0.001 ‘*      01 ‘*’       .’ 0.1 ‘ ’ 1

Residual standard error: 2.817 on 995 degrees of freedom
Multiple R-squared:  0.7882,    Adjusted R-squared:  0.7873
F-statistic: 925.6 on 4 and 995 DF,  p-value: < 2.2e-16
```

What can we do with R?

Analytics

Basic mathematics

Statistical tests

Big data analysis

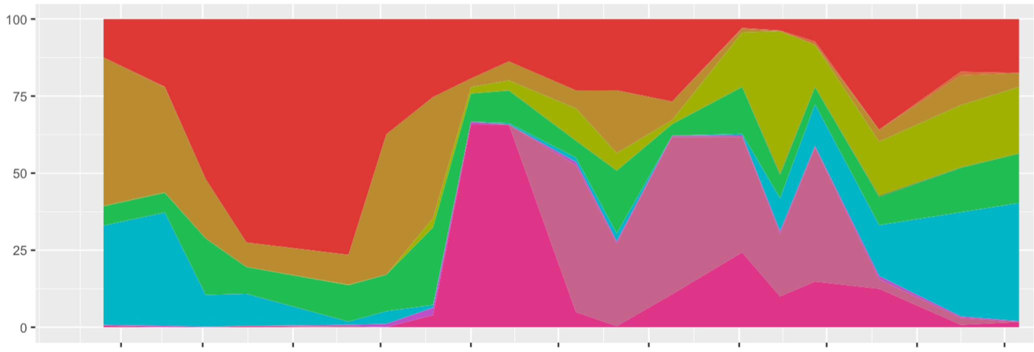
Statistical modeling

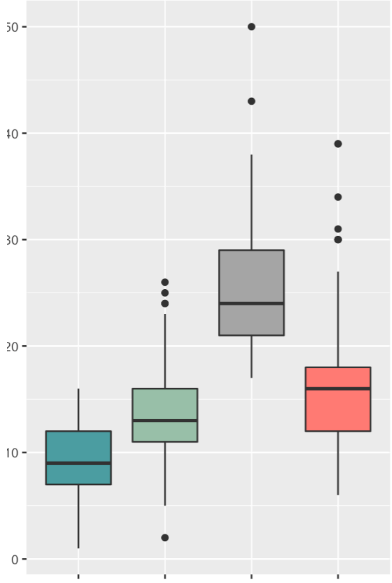
Machine learning

Graphics and visualisation

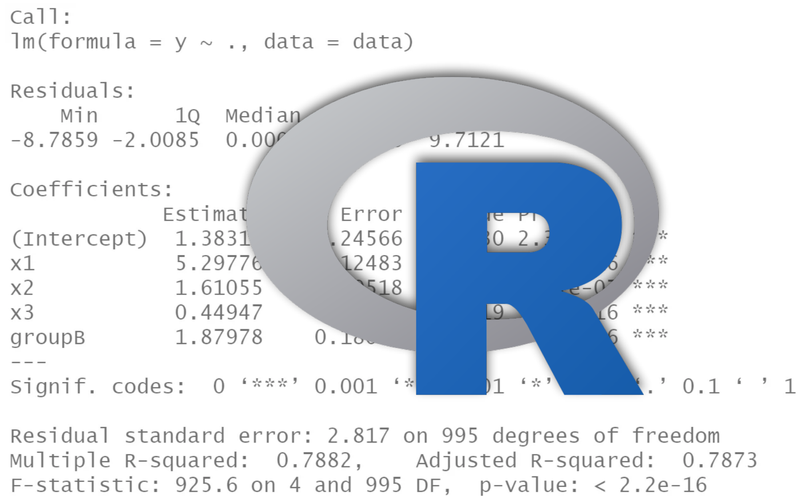
Static graphics

Cartography





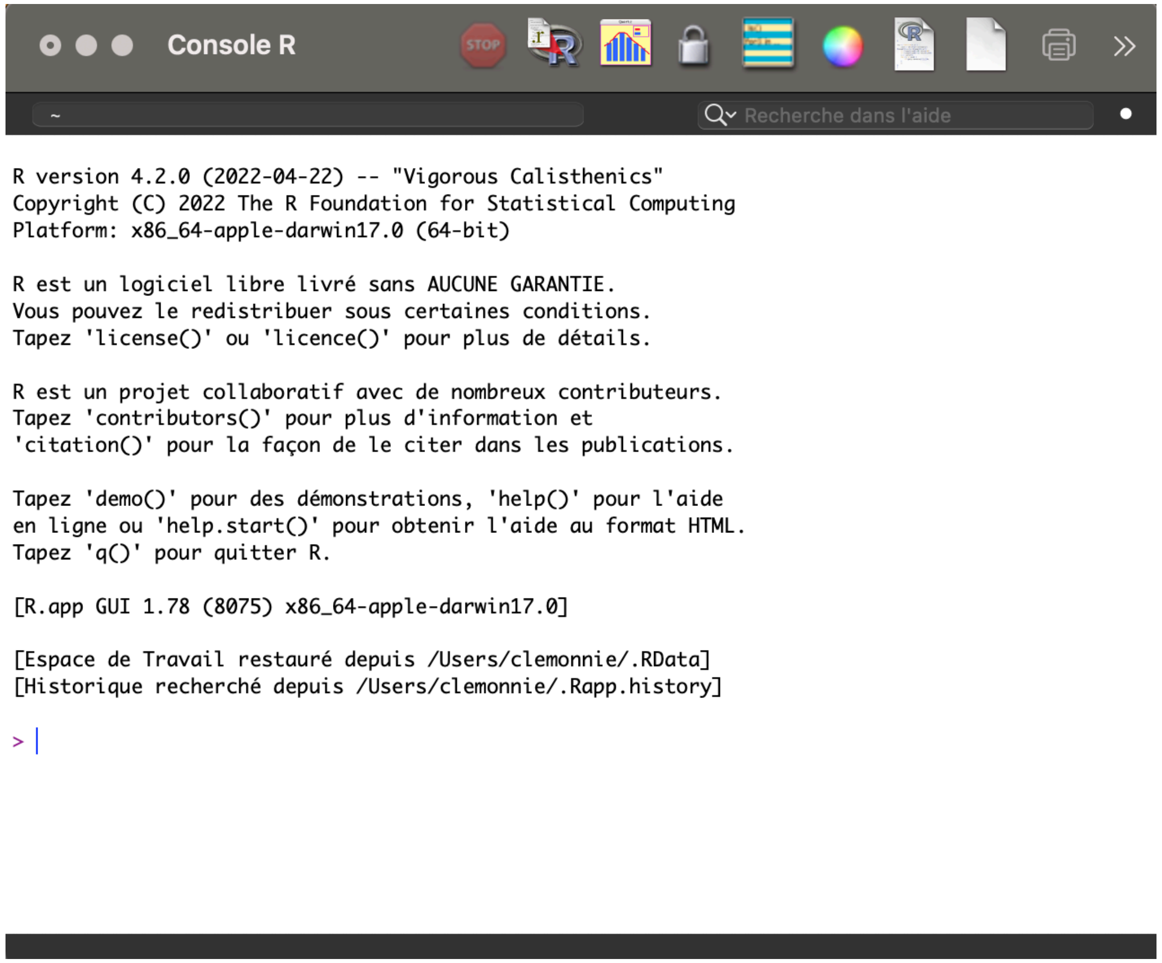
A little introduction to R



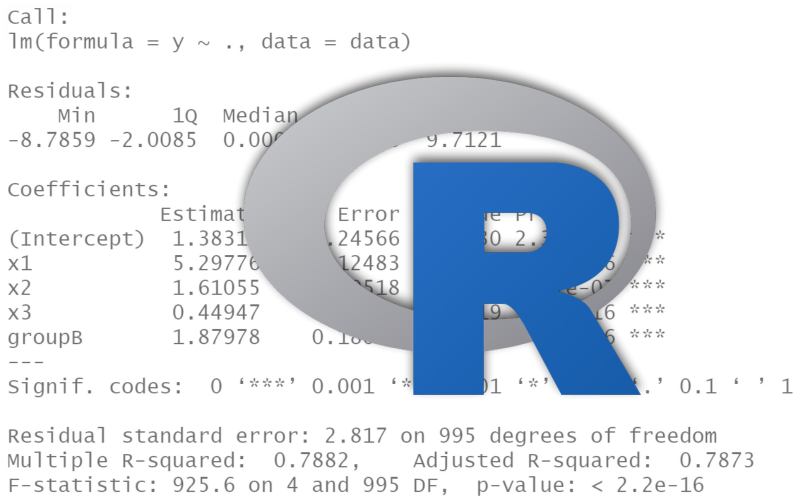
```
Call:
lm(formula = y ~ ., data = data)

Residuals:
    Min      1Q  Median              9.7121
-8.7859 -2.0085  0.00

Coefficients:
            Estimat        Error      e Pr
(Intercept)  1.3831        .24566     30 2.           *
x1           5.29776       12483                   6  **
x2           1.61055        518           e-0     ***
x3           0.44947                  19       16 ***
groupB       1.87978    0.18                    6 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘*       01 ‘*’        .’ 0.1 ‘ ’ 1

Residual standard error: 2.817 on 995 degrees of freedom
Multiple R-squared:  0.7882,	Adjusted R-squared:  0.7873 
F-statistic: 925.6 on 4 and 995 DF,  p-value: < 2.2e-16
```

R console

R is used through a command line console



```
R version 4.2.0 (2022-04-22) -- "Vigorous Calisthenics"
Copyright (C) 2022 The R Foundation for Statistical Computing
Platform: x86_64-apple-darwin17.0 (64-bit)

R est un logiciel libre livré sans AUCUNE GARANTIE.
Vous pouvez le redistribuer sous certaines conditions.
Tapez 'license()' ou 'licence()' pour plus de détails.

R est un projet collaboratif avec de nombreux contributeurs.
Tapez 'contributors()' pour plus d'information et
'citation()' pour la façon de le citer dans les publications.

Tapez 'demo()' pour des démonstrations, 'help()' pour l'aide
en ligne ou 'help.start()' pour obtenir l'aide au format HTML.
Tapez 'q()' pour quitter R.

[R.app GUI 1.78 (8075) x86_64-apple-darwin17.0]

[Espace de Travail restauré depuis /Users/clemonnie/.RData]
[Historique recherché depuis /Users/clemonnie/.Rapp.history]

> |
```



```
Call:
lm(formula = y ~ ., data = data)

Residuals:
    Min      1Q  Median
-8.7859 -2.0085  0.00            9.7121

Coefficients:
            Estimat       Error
(Intercept)  1.3831       .24566       30 2.          *
x1           5.29776      12483                   6  **
x2           1.61055       518              e-0   ***
x3           0.44947                  19        16 ***
groupB       1.87978    0.18                    6 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘*      01 ‘*’      .’ 0.1 ‘ ’ 1

Residual standard error: 2.817 on 995 degrees of freedom
Multiple R-squared:  0.7882,    Adjusted R-squared:  0.7873
F-statistic: 925.6 on 4 and 995 DF,  p-value: < 2.2e-16
```

R console

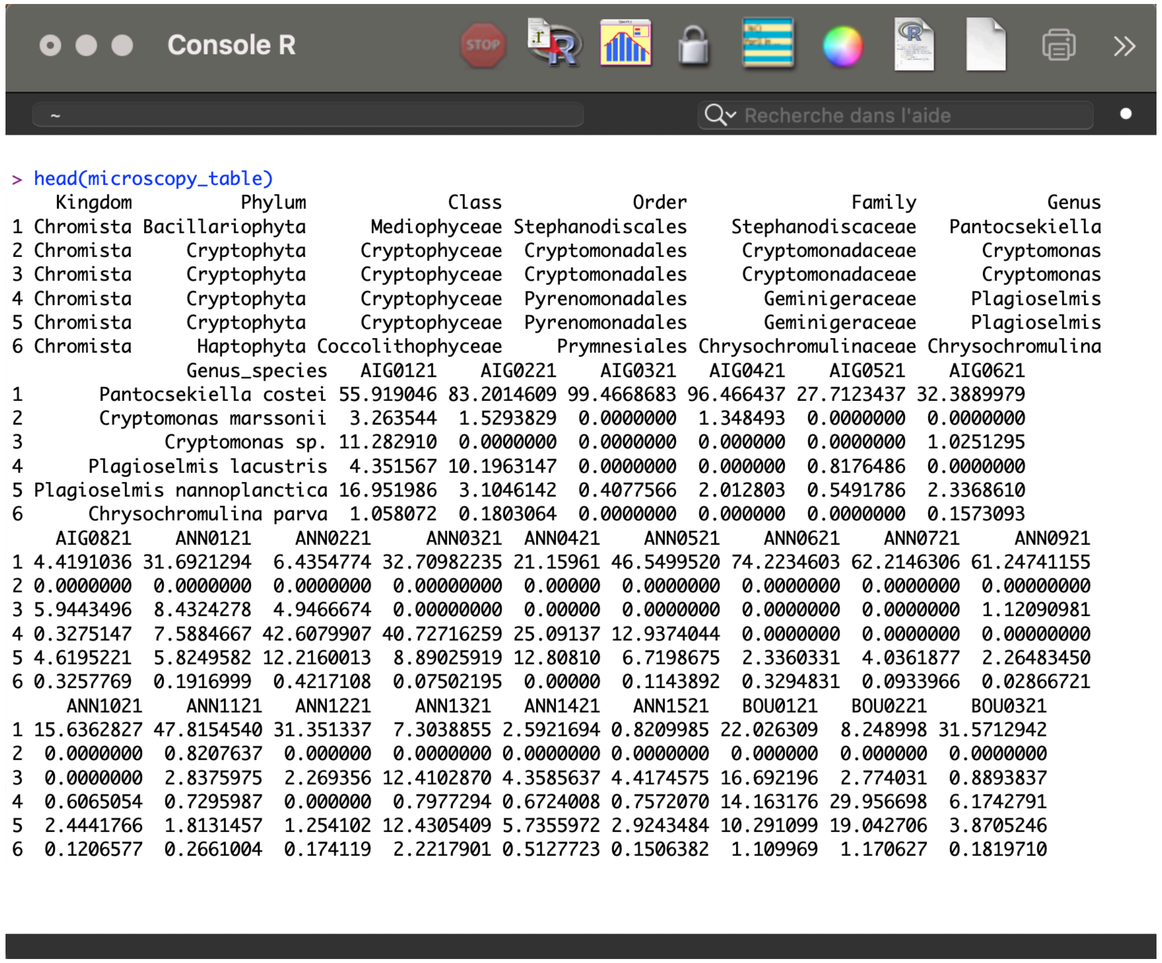
R is used through a command line console

Load a table

```
> Microscopy_table <- read.table("Microscopy_table.csv", header=T, sep=";")
```

Visualize the table

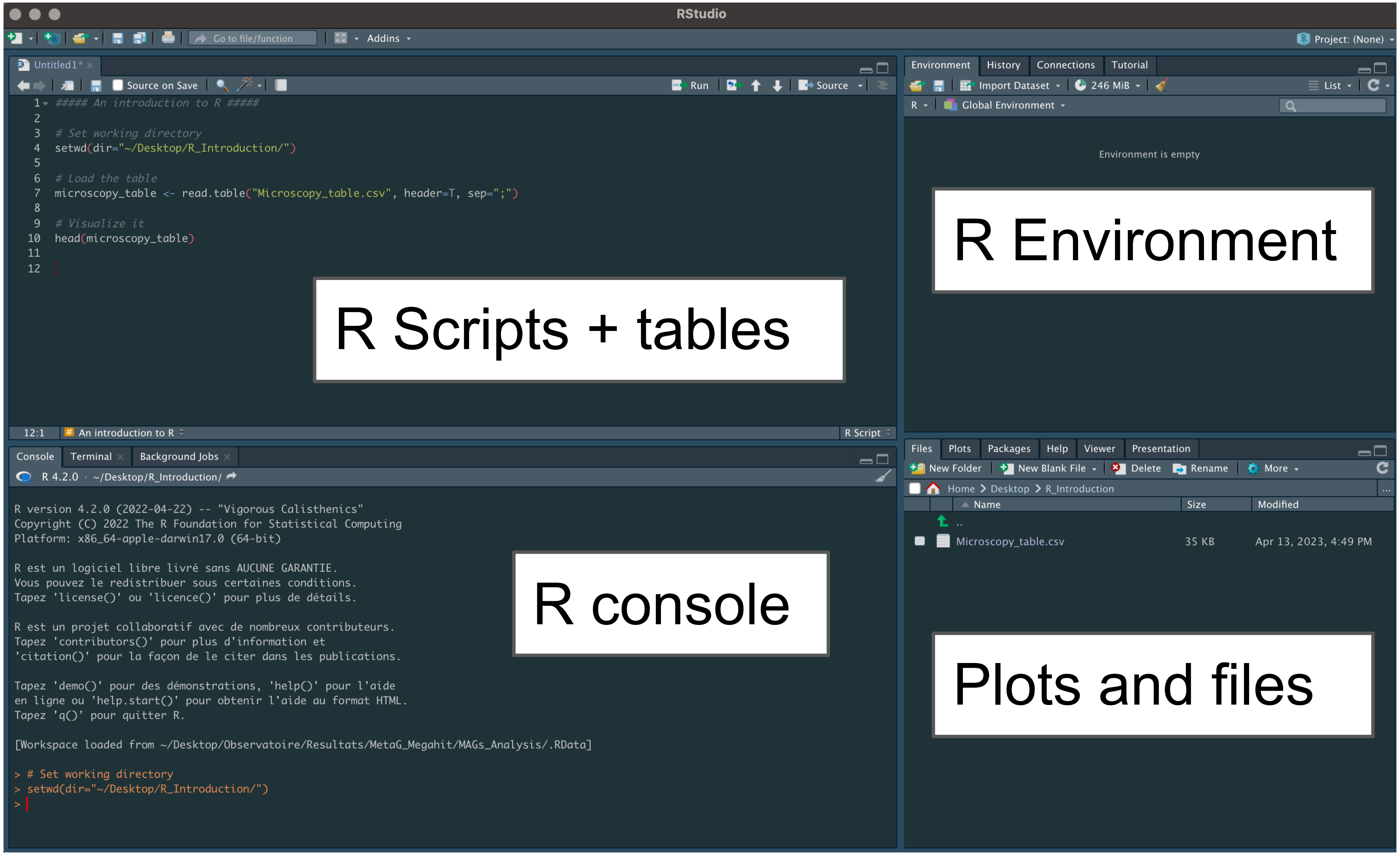
```
> head(microscopy_table)
```



```
> head(microscopy_table)
    Kingdom          Phylum              Class            Order              Family            Genus
1 Chromista Bacillariophyta      Mediophyceae Stephanodiscales    Stephanodiscaceae   Pantocsekiella
2 Chromista     Cryptophyta     Cryptophyceae  Cryptomonadales     Cryptomonadaceae      Cryptomonas
3 Chromista     Cryptophyta     Cryptophyceae  Cryptomonadales     Cryptomonadaceae      Cryptomonas
4 Chromista     Cryptophyta     Cryptophyceae  Pyrenomonadales       Geminigeraceae      Plagioselmis
5 Chromista     Cryptophyta     Cryptophyceae  Pyrenomonadales       Geminigeraceae      Plagioselmis
6 Chromista      Haptophyta Coccolithophyceae      Prymnesiales Chrysochromulinaceae Chrysochromulina
               Genus_species   AIG0121    AIG0221    AIG0321   AIG0421    AIG0521    AIG0621
1        Pantocsekiella costei 55.919046 83.2014609 99.4668683 96.466437 27.7123437 32.3889979
2        Cryptomonas marssonii  3.263544  1.5293829  0.0000000  1.348493  0.0000000  0.0000000
3              Cryptomonas sp. 11.282910  0.0000000  0.0000000  0.000000  0.0000000  1.0251295
4       Plagioselmis lacustris  4.351567 10.1963147  0.0000000  0.000000  0.8176486  0.0000000
5 Plagioselmis nannoplanctica 16.951986  3.1046142  0.4077566  2.012803  0.5491786  2.3368610
6       Chrysochromulina parva  1.058072  0.1803064  0.0000000  0.000000  0.0000000  0.1573093
    AIG0821    ANN0121    ANN0221     ANN0321  ANN0421    ANN0521    ANN0621    ANN0721     ANN0921
1 4.4191036 31.6921294  6.4354774 32.70982235 21.15961 46.5499520 74.2234603 62.2146306 61.24741155
2 0.0000000  0.0000000  0.0000000  0.00000000  0.00000  0.0000000  0.0000000  0.0000000  0.00000000
3 5.9443496  8.4324278  4.9466674  0.00000000  0.00000  0.0000000  0.0000000  0.0000000  1.12090981
4 0.3275147  7.5884667 42.6079907 40.72716259 25.09137 12.9374044  0.0000000  0.0000000  0.00000000
5 4.6195221  5.8249582 12.2160013  8.89025919 12.80810  6.7198675  2.3360331  4.0361877  2.26483450
6 0.3257769  0.1916999  0.4217108  0.07502195  0.00000  0.1143892  0.3294831  0.0933966  0.02866721
     ANN1021    ANN1121   ANN1221    ANN1321   ANN1421   ANN1521   BOU0121   BOU0221    BOU0321
1 15.6362827 47.8154540 31.351337  7.3038855 2.5921694 0.8209985 22.026309  8.248998 31.5712942
2  0.0000000  0.8207637  0.000000  0.0000000 0.0000000 0.0000000  0.000000  0.000000  0.0000000
3  0.0000000  2.8375975  2.269356 12.4102870 4.3585637 4.4174575 16.692196  2.774031  0.8893837
4  0.6065054  0.7295987  0.000000  0.7977294 0.6724008 0.7572070 14.163176 29.956698  6.1742791
5  2.4441766  1.8131457  1.254102 12.4305409 5.7355972 2.9243484 10.291099 19.042706  3.8705246
6  0.1206577  0.2661004  0.174119  2.2217901 0.5127723 0.1506382  1.109969  1.170627  0.1819710
```

A little introduction to R

RStudio® Is an integrated development environment (IDE)



```
##### An introduction to R #####

# Set working directory
setwd(dir="~/Desktop/R_Introduction/")

# Load the table
microscopy_table <- read.table("Microscopy_table.csv", header=T, sep=";")

# Visualize it
head(microscopy_table)
```

R Scripts + tables

Environment is empty

R Environment

```
R version 4.2.0 (2022-04-22) -- "Vigorous Calisthenics"
Copyright (C) 2022 The R Foundation for Statistical Computing
Platform: x86_64-apple-darwin17.0 (64-bit)

R est un logiciel libre livré sans AUCUNE GARANTIE.
Vous pouvez le redistribuer sous certaines conditions.
Tapez 'license()' ou 'licence()' pour plus de détails.

R est un projet collaboratif avec de nombreux contributeurs.
Tapez 'contributors()' pour plus d'information et
'citation()' pour la façon de le citer dans les publications.

Tapez 'demo()' pour des démonstrations, 'help()' pour l'aide
en ligne ou 'help.start()' pour obtenir l'aide au format HTML.
Tapez 'q()' pour quitter R.

[Workspace loaded from ~/Desktop/Observatoire/Resultats/MetaG_Megahit/MAGs_Analysis/.RData]

> # Set working directory
> setwd(dir="~/Desktop/R_Introduction/")
> 
```

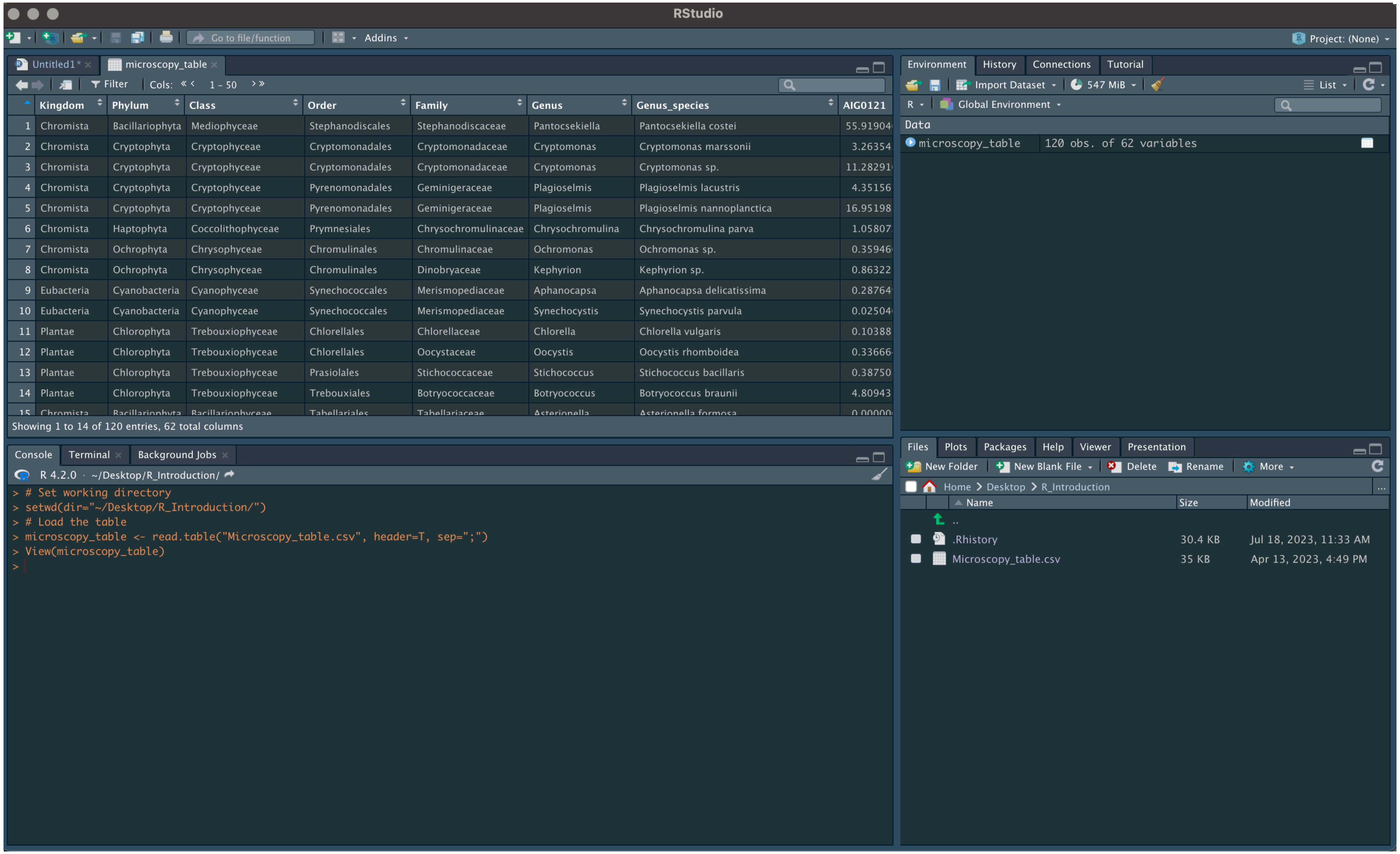
R console

Name	Size	Modified
..		
Microscopy_table.csv	35 KB	Apr 13, 2023, 4:49 PM

Plots and files

A little introduction to R

RStudio® Is an integrated development environment (IDE)



	Kingdom	Phylum	Class	Order	Family	Genus	Genus_species	AIG0121
1	Chromista	Bacillariophyta	Mediophyceae	Stephanodiscales	Stephanodiscaceae	Pantocsekiella	Pantocsekiella costei	55.91904
2	Chromista	Cryptophyta	Cryptophyceae	Cryptomonadales	Cryptomonadaceae	Cryptomonas	Cryptomonas marssonii	3.26354
3	Chromista	Cryptophyta	Cryptophyceae	Cryptomonadales	Cryptomonadaceae	Cryptomonas	Cryptomonas sp.	11.28291
4	Chromista	Cryptophyta	Cryptophyceae	Pyrenomonadales	Geminigeraceae	Plagioselmis	Plagioselmis lacustris	4.35156
5	Chromista	Cryptophyta	Cryptophyceae	Pyrenomonadales	Geminigeraceae	Plagioselmis	Plagioselmis nannoplanctica	16.95198
6	Chromista	Haptophyta	Coccolithophyceae	Prymnesiales	Chrysochromulinaceae	Chrysochromulina	Chrysochromulina parva	1.05807
7	Chromista	Ochrophyta	Chrysophyceae	Chromulinales	Chromulinaceae	Ochromonas	Ochromonas sp.	0.35946
8	Chromista	Ochrophyta	Chrysophyceae	Chromulinales	Dinobryaceae	Kephyrion	Kephyrion sp.	0.86322
9	Eubacteria	Cyanobacteria	Cyanophyceae	Synechococcales	Merismopediaceae	Aphanocapsa	Aphanocapsa delicatissima	0.28764
10	Eubacteria	Cyanobacteria	Cyanophyceae	Synechococcales	Merismopediaceae	Synechocystis	Synechocystis parvula	0.02504
11	Plantae	Chlorophyta	Trebouxiophyceae	Chlorellales	Chlorellaceae	Chlorella	Chlorella vulgaris	0.10388
12	Plantae	Chlorophyta	Trebouxiophyceae	Chlorellales	Oocystaceae	Oocystis	Oocystis rhomboidea	0.33666
13	Plantae	Chlorophyta	Trebouxiophyceae	Prasiolales	Stichococcaceae	Stichococcus	Stichococcus bacillaris	0.38750
14	Plantae	Chlorophyta	Trebouxiophyceae	Trebouxiales	Botryococcaceae	Botryococcus	Botryococcus braunii	4.80943
15	Chromista	Bacillariophyta	Bacillariophyceae	Tabellariales	Tabellariaceae	Asterionella	Asterionella formosa	0.00000

Showing 1 to 14 of 120 entries, 62 total columns

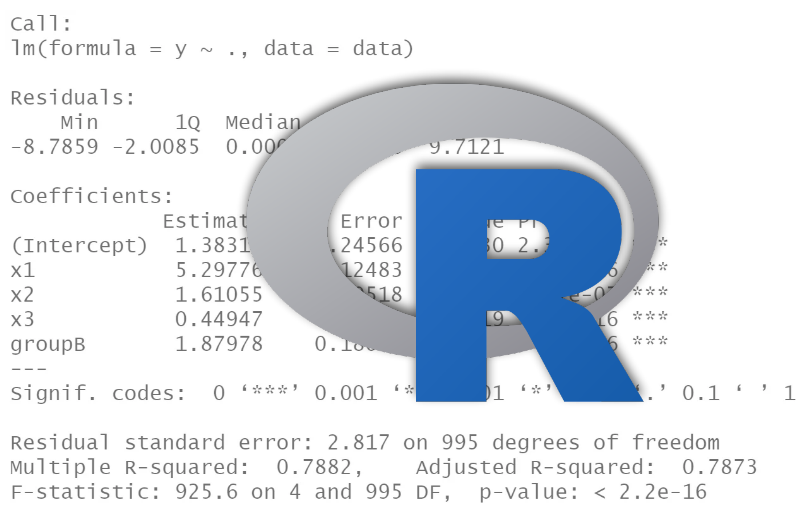
```
> # Set working directory
> setwd(dir="~/Desktop/R_Introduction/")
> # Load the table
> microscopy_table <- read.table("Microscopy_table.csv", header=T, sep=";")
> View(microscopy_table)
>
```

Data

microscopy_table   120 obs. of 62 variables

Name	Size	Modified
..		
.Rhistory	30.4 KB	Jul 18, 2023, 11:33 AM
Microscopy_table.csv	35 KB	Apr 13, 2023, 4:49 PM

A little introduction to R



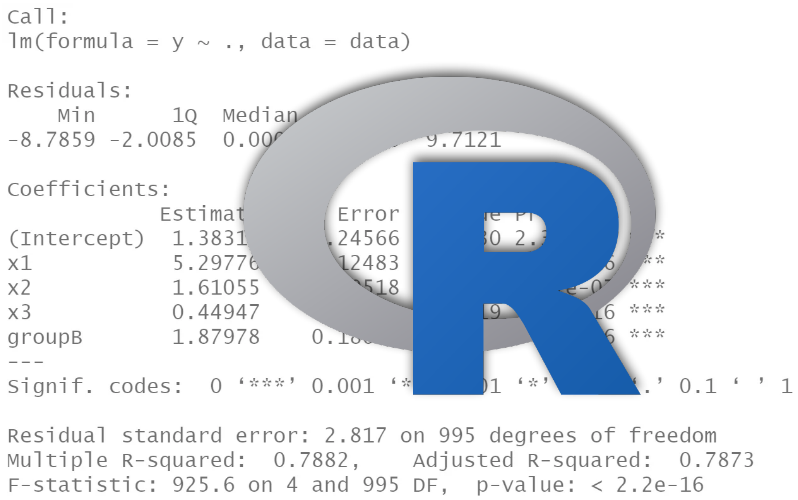
```
Call:
lm(formula = y ~ ., data = data)

Residuals:
    Min      1Q  Median
-8.7859 -2.0085  0.000            9.7121

Coefficients:
            Estimat        Error
(Intercept)  1.3831        .24566       30 2.
x1           5.29776        12483             6  **
x2           1.61055         518           e-07 ***
x3           0.44947                 19      16 ***
groupB       1.87978    0.18             6 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘*      01 ‘*’      ‘.’ 0.1 ‘ ’ 1

Residual standard error: 2.817 on 995 degrees of freedom
Multiple R-squared:  0.7882,    Adjusted R-squared:  0.7873
F-statistic: 925.6 on 4 and 995 DF,  p-value: < 2.2e-16
```

R Language

Command line example

```
> ASV_table <- read.table("ASV_table.txt", header=T, row.names=1, sep="\t")
```

A little introduction to R



```
Call:
lm(formula = y ~ ., data = data)

Residuals:
    Min      1Q  Median
-8.7859 -2.0085  0.00             9.7121

Coefficients:
            Estimat        Error     e Pr
(Intercept)  1.3831       .24566     30 2.           *
x1           5.29776       12483               6    **
x2           1.61055         518         e-0     ***
x3           0.44947               19        16  ***
groupB       1.87978    0.18                  6  ***
---
Signif. codes:  0 ‘***’ 0.001 ‘*     01 ‘*’      ‘.’ 0.1 ‘ ’ 1

Residual standard error: 2.817 on 995 degrees of freedom
Multiple R-squared:  0.7882,    Adjusted R-squared:  0.7873
F-statistic: 925.6 on 4 and 995 DF,  p-value: < 2.2e-16
```

R Language

Command line example

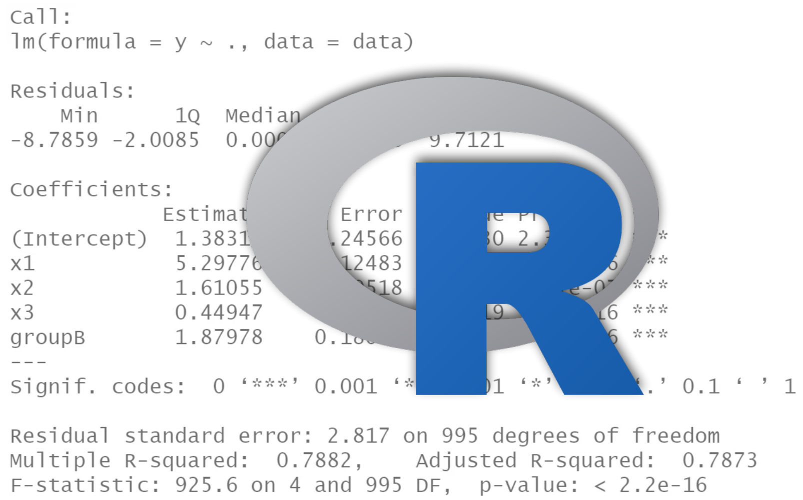
```
> ASV_table <- read.table("ASV_table.txt", header=T, row.names=1, sep="\t")
```

function

arguments

A function() is made to perform a specific task.

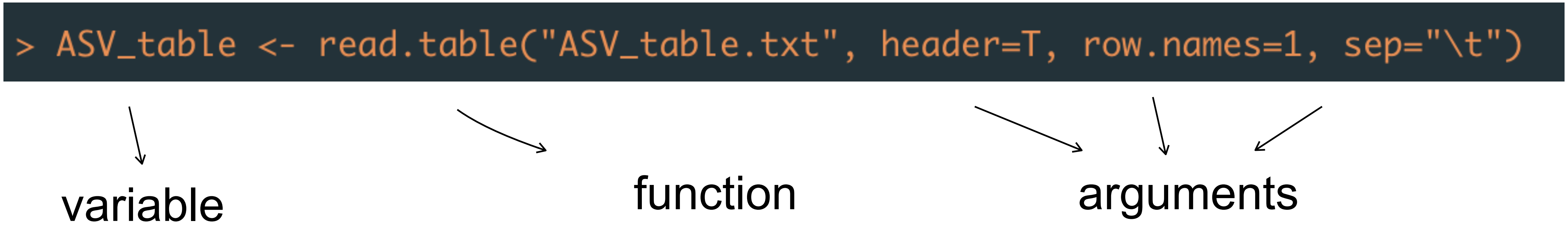
It might works with arguments



```
Call:
lm(formula = y ~ ., data = data)

Residuals:
    Min      1Q  Median
-8.7859 -2.0085  0.00             9.7121

Coefficients:
            Estimat      Error
(Intercept)  1.3831      .24566        30 2.           *
x1           5.29776      12483                   6   **
x2           1.61055        518                e-0    ***
x3           0.44947                  19          16 ***
groupB       1.87978    0.18                       6 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘*      01 ‘*’       .’ 0.1 ‘ ’ 1

Residual standard error: 2.817 on 995 degrees of freedom
Multiple R-squared:  0.7882,    Adjusted R-squared:  0.7873
F-statistic: 925.6 on 4 and 995 DF,  p-value: < 2.2e-16
```

R Language

Command line example



```
> ASV_table <- read.table("ASV_table.txt", header=T, row.names=1, sep="\t")
```

variable

function

arguments

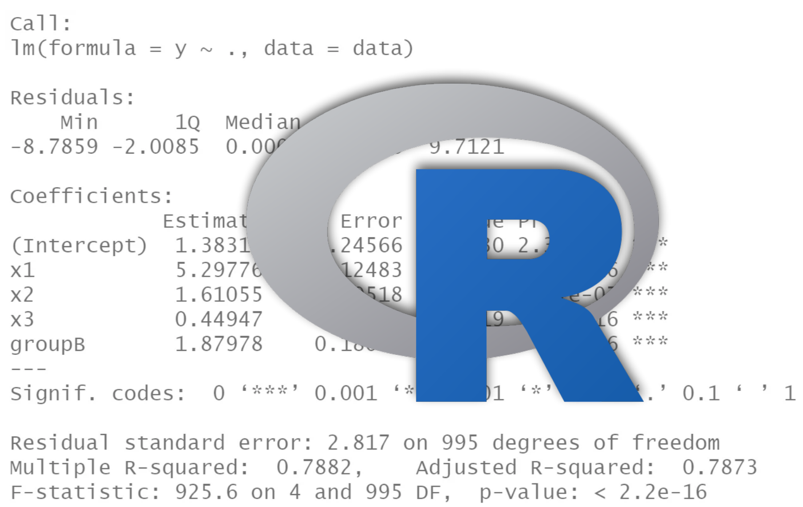
A variable in R is the memory allocated to the stockage of a specific **object**

It is assigned to an object using either `=` or `<-`

A function() is made to perform a specific task.

It might works with arguments

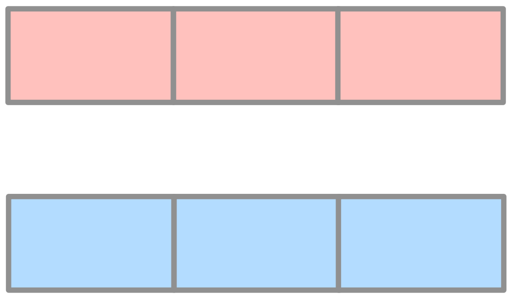
A little introduction to R



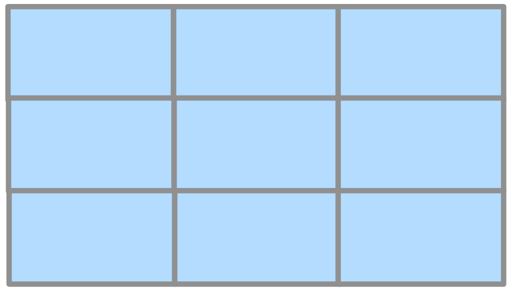
```
Call:
lm(formula = y ~ ., data = data)

Residuals:
    Min      1Q  Median
-8.7859 -2.0085  0.00            9.7121

Coefficients:
            Estimat       Error
(Intercept)  1.3831       .24566       30 2.          *
x1           5.29776      12483                  6   **
x2           1.61055       518               e-0    ***
x3           0.44947                  19        16 ***
groupB       1.87978    0.18                     6 ***
---
Signif. codes:  0 '***' 0.001 '*      01 '*'      '.' 0.1 ' ' 1

Residual standard error: 2.817 on 995 degrees of freedom
Multiple R-squared:  0.7882,    Adjusted R-squared:  0.7873
F-statistic: 925.6 on 4 and 995 DF,  p-value: < 2.2e-16
```

R Language

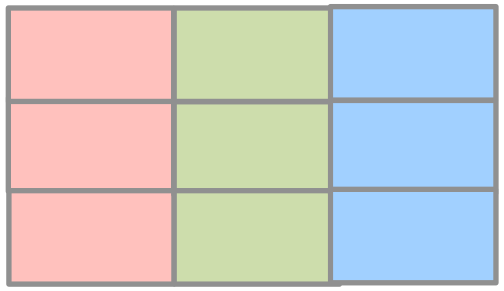
Different type of objects in R



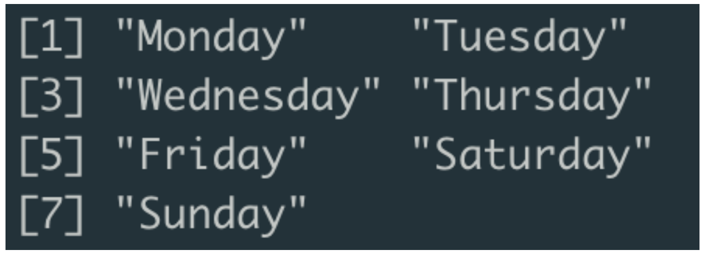
Vector



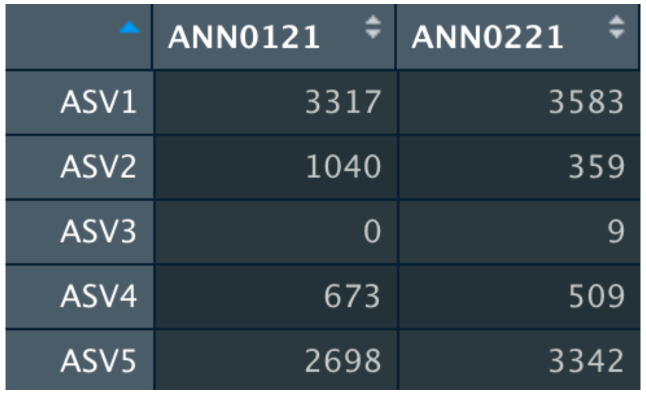
Matrix



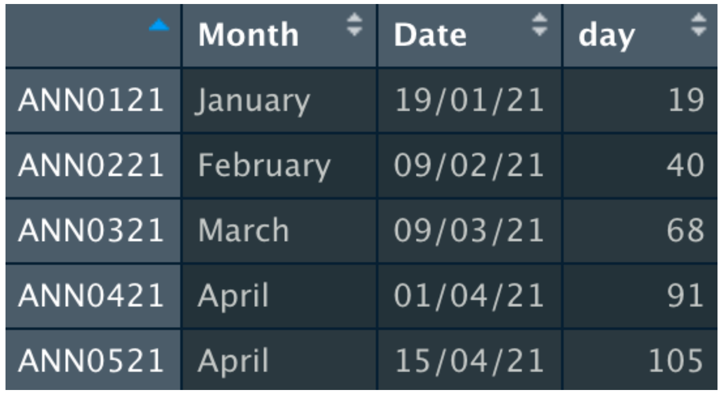
Data frame



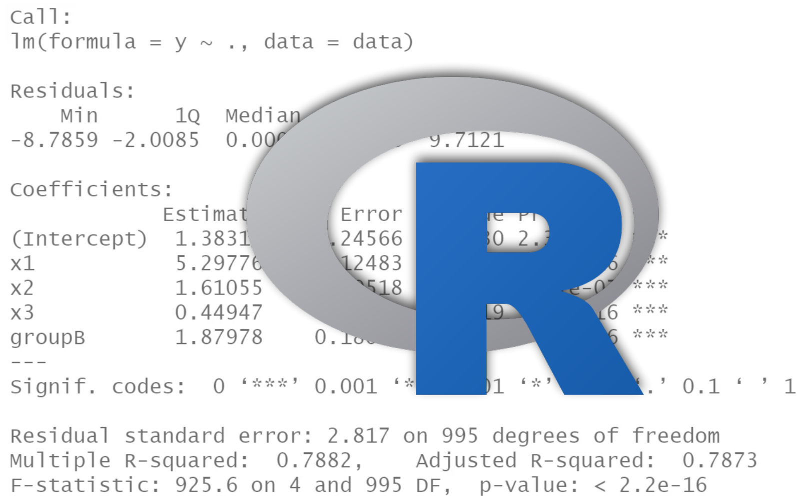
```
[1] "Monday"    "Tuesday"
[3] "Wednesday" "Thursday"
[5] "Friday"    "Saturday"
[7] "Sunday"
```



	ANN0121	ANN0221
ASV1	3317	3583
ASV2	1040	359
ASV3	0	9
ASV4	673	509
ASV5	2698	3342



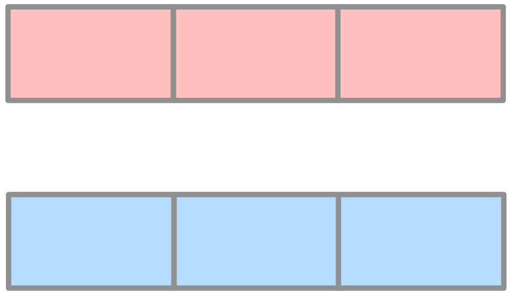
	Month	Date	day
ANN0121	January	19/01/21	19
ANN0221	February	09/02/21	40
ANN0321	March	09/03/21	68
ANN0421	April	01/04/21	91
ANN0521	April	15/04/21	105



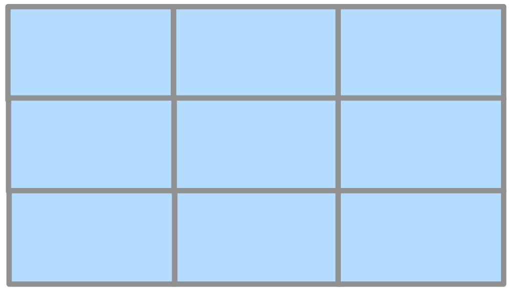
```
Call:
lm(formula = y ~ ., data = data)

Residuals:
    Min      1Q  Median
-8.7859 -2.0085  0.00          9.7121

Coefficients:
            Estimat       Error
(Intercept)  1.3831      .24566
x1           5.29776     12483
x2           1.61055     518
x3           0.44947
groupB       1.87978     0.18
---
Signif. codes:  0 ‘***’ 0.001 ‘*          01 ‘*’        .’ 0.1 ‘ ’ 1

Residual standard error: 2.817 on 995 degrees of freedom
Multiple R-squared:  0.7882,    Adjusted R-squared:  0.7873
F-statistic: 925.6 on 4 and 995 DF,  p-value: < 2.2e-16
```

R Language

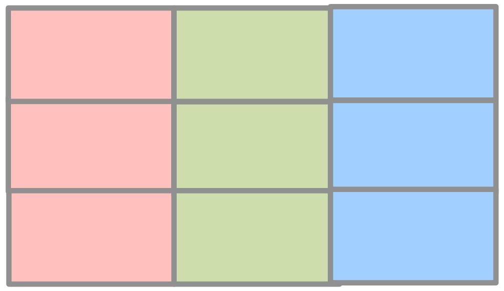
Different type of objects in R



Vector



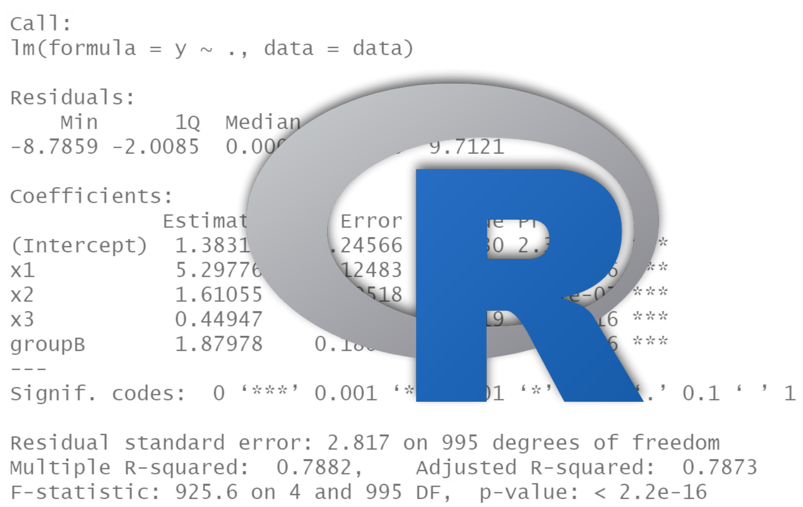
Matrix



Data frame

List {
- Vector
- Matrix
- Data frame
- Array

Source: R in Action, p.23



```
Call:
lm(formula = y ~ ., data = data)

Residuals:
    Min      1Q  Median
-8.7859 -2.0085  0.00           9.7121

Coefficients:
            Estimat       Error     e Pr
(Intercept)  1.3831       .24566    30 2.          *
x1           5.29776      12483               6   **
x2           1.61055       518             e-0   ***
x3           0.44947               19         16 ***
groupB       1.87978    0.18                  6 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘*     01 ‘*’      ‘.’ 0.1 ‘ ’ 1

Residual standard error: 2.817 on 995 degrees of freedom
Multiple R-squared:  0.7882,	Adjusted R-squared:  0.7873 
F-statistic: 925.6 on 4 and 995 DF,  p-value: < 2.2e-16
```

R Language

Different classes of objects in R

Character

« word »

Numeric

« 4.356 »

Integer

« 4 » or « 4L »

Date

« 01/10/2023»

Logical

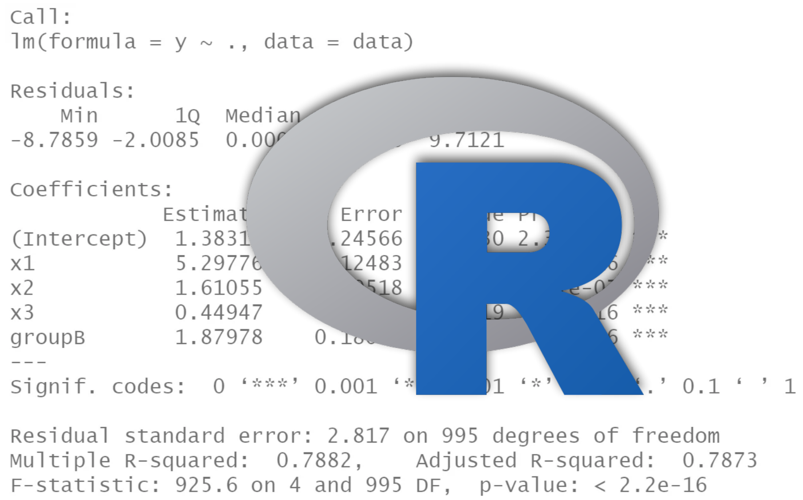
«TRUE» «FALSE»

Get the class of the object A

class(A)

Change A to a specific class

as.numeric(A)



```
Call:
lm(formula = y ~ ., data = data)

Residuals:
    Min      1Q  Median              9.7121
-8.7859 -2.0085  0.00

Coefficients:
            Estimat        Error       e Pr
(Intercept)  1.3831        .24566      30 2.          *
x1           5.29776       12483              6     **
x2           1.61055        518            e-0    ***
x3           0.44947                  19       16 ***
groupB       1.87978    0.18                    6 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘*     01 ‘*’      .’ 0.1 ‘ ’ 1

Residual standard error: 2.817 on 995 degrees of freedom
Multiple R-squared:  0.7882,    Adjusted R-squared:  0.7873
F-statistic: 925.6 on 4 and 995 DF,  p-value: < 2.2e-16
```

R Language

Basic operations and functions

Different operators

R operators	
+	Addition
*	Multiplication
/	Division
-	Soustraction

Relational operators	
>	More than
<	Less than
>=	More or equal
!=	Not equal to

Logical operators	
&	And
|	Or
!	Not

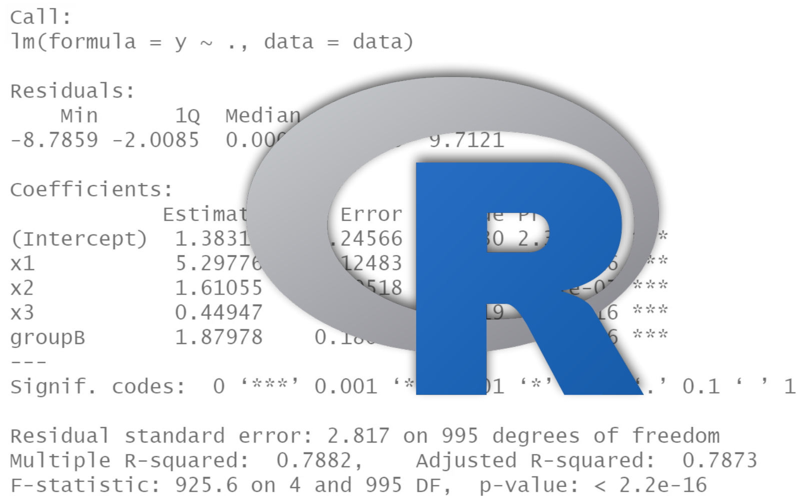
Miscellaneous oper.	
%in%	in

To do a specific mathematical or logical operation

Different functions

setwd()
sort()
order()
factor()
str()
print()
plot()
merge()
na.omit()
…

To do a specific task



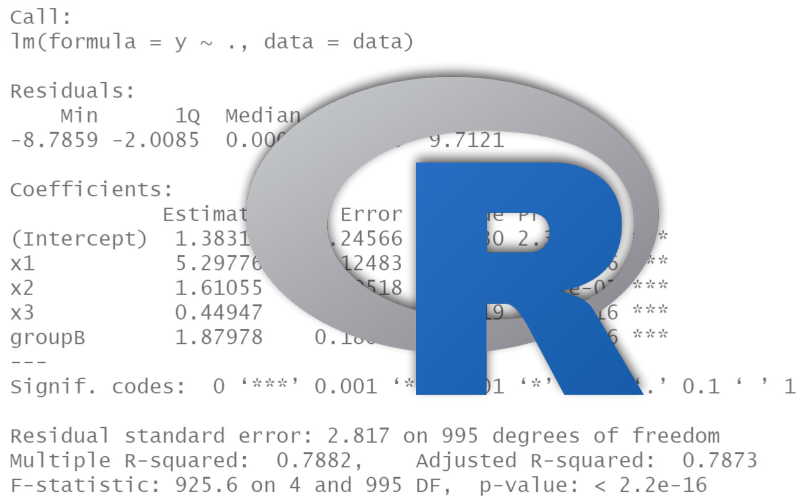
```
Call:
lm(formula = y ~ ., data = data)

Residuals:
    Min      1Q  Median
-8.7859 -2.0085  0.00          9.7121

Coefficients:
            Estimat       Error
(Intercept)  1.3831      .24566       0 2.              *
x1           5.29776      12483                     6   **
x2           1.61055        518                 e-0   ***
x3           0.44947                   19          16 ***
groupB       1.87978    0.18                        6 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘*        01 ‘*’        .’ 0.1 ‘ ’ 1

Residual standard error: 2.817 on 995 degrees of freedom
Multiple R-squared:  0.7882,    Adjusted R-squared:  0.7873
F-statistic: 925.6 on 4 and 995 DF,  p-value: < 2.2e-16
```

R Language

Basic operations and functions

Load a file

```
> ASV_table <- read.table("ASV_table.txt", header=T, row.names=1, sep="\t")
```



```
Call:
lm(formula = y ~ ., data = data)

Residuals:
    Min      1Q  Median
-8.7859 -2.0085  0.000          9.7121

Coefficients:
            Estimat        Error       e Pr
(Intercept)  1.3831        .24566      30 2.       *
x1           5.29776       12483               6  **
x2           1.61055        518              e-07 ***
x3           0.44947                 19       16 ***
groupB       1.87978    0.18                   6 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘*      01 ‘*’       .’ 0.1 ‘ ’ 1

Residual standard error: 2.817 on 995 degrees of freedom
Multiple R-squared:  0.7882,    Adjusted R-squared:  0.7873
F-statistic: 925.6 on 4 and 995 DF,  p-value: < 2.2e-16
```

R Packages

Combination of different functions for a specific purpose

Most of the time they are available in CRAN

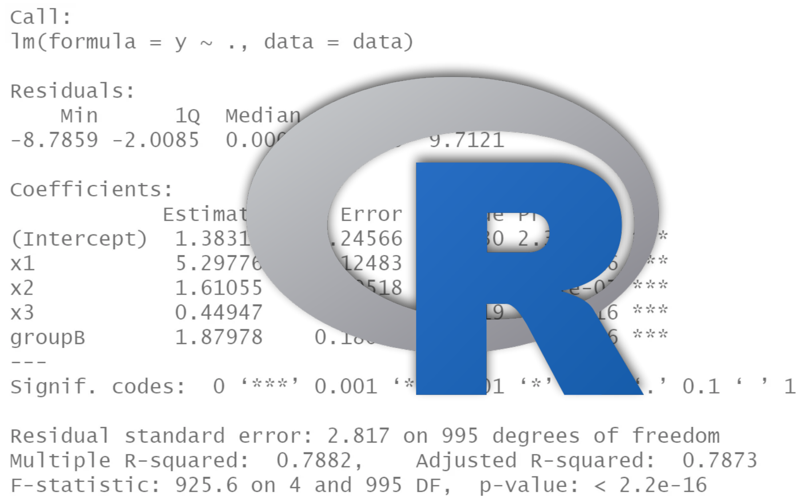
CRAN R

The Comprehensive R Archive Network

But sometimes (particularly if they are recent) you can directly install them from Bioconductor or Github

Bioconductor

OPEN SOURCE SOFTWARE FOR BIOINFORMATICS



```
Call:
lm(formula = y ~ ., data = data)

Residuals:
    Min      1Q  Median
-8.7859 -2.0085  0.000          9.7121

Coefficients:
            Estimat       Error        e Pr
(Intercept)  1.3831       .24566      30 2.            *
x1           5.29776      12483                  6    **
x2           1.61055       518           e-07  ***
x3           0.44947              19          16 ***
groupB       1.87978    0.18                  6 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘*    01 ‘*’      .’ 0.1 ‘ ’ 1

Residual standard error: 2.817 on 995 degrees of freedom
Multiple R-squared:  0.7882,    Adjusted R-squared:  0.7873
F-statistic: 925.6 on 4 and 995 DF,  p-value: < 2.2e-16
```

R Packages

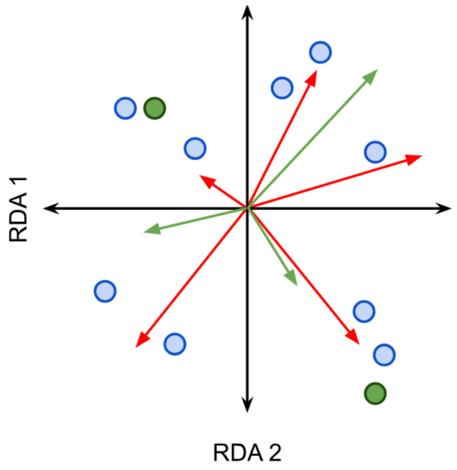
Community ecology : vegan package / ade4 package

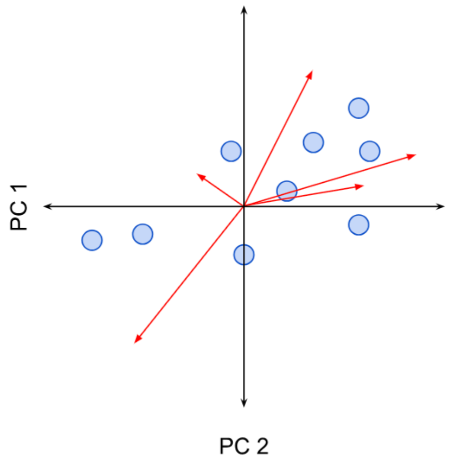
Test hypothesis

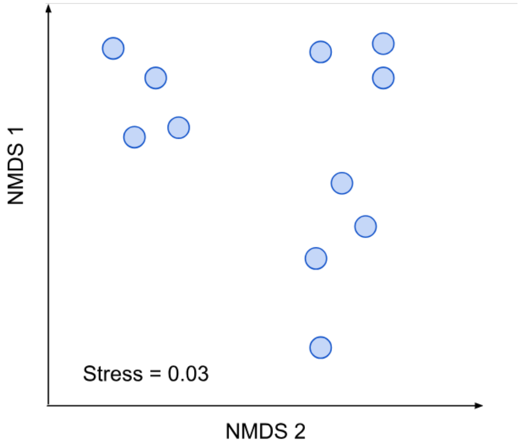
Multivariate analysis

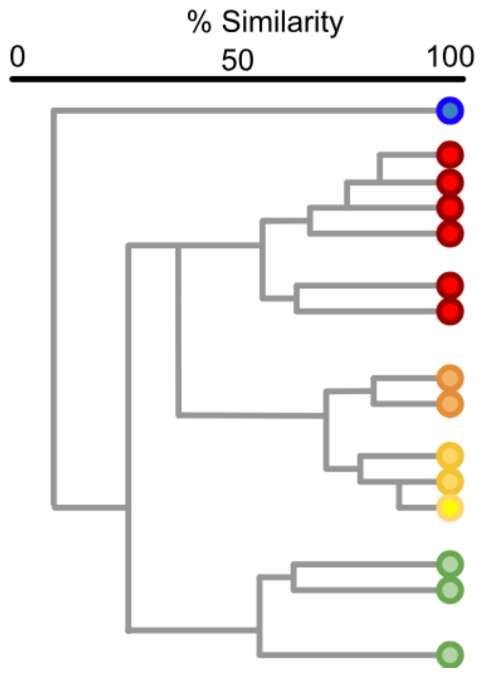
Constrained analysis

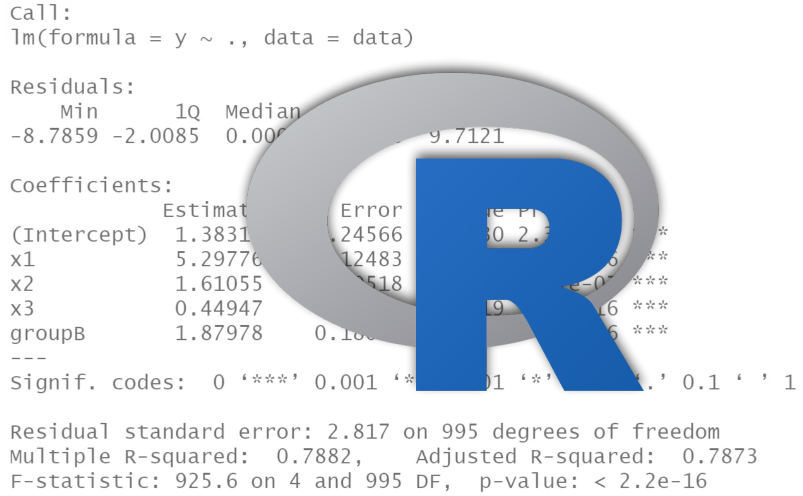
Composant analysis, PCA







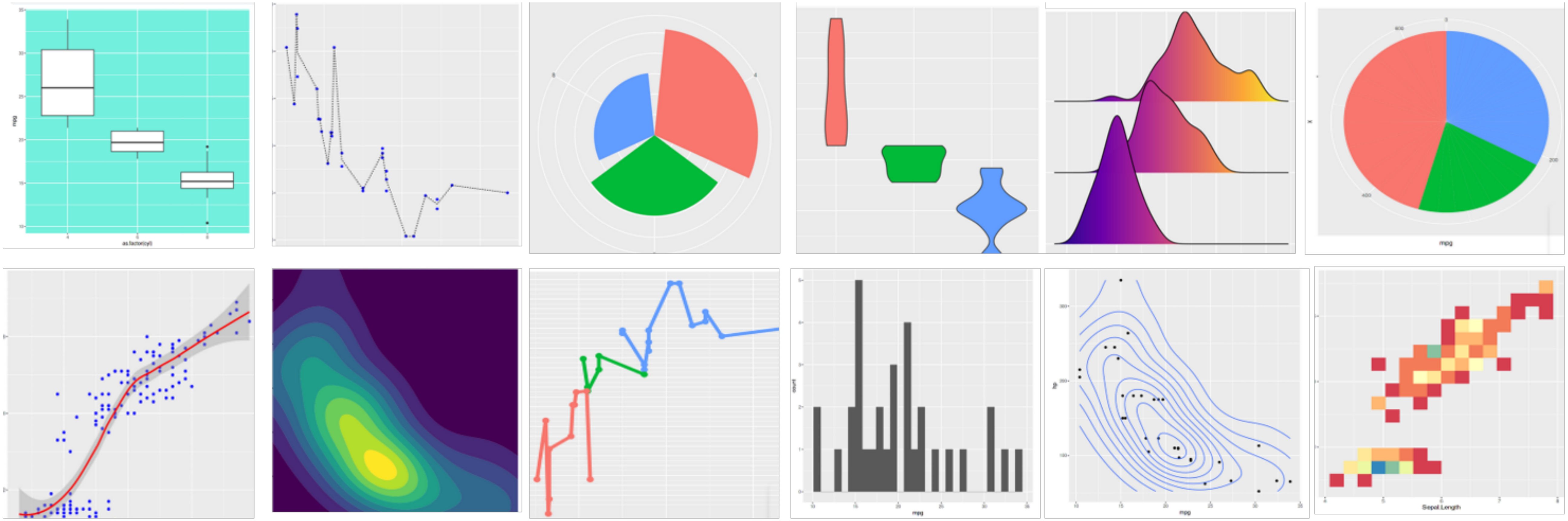


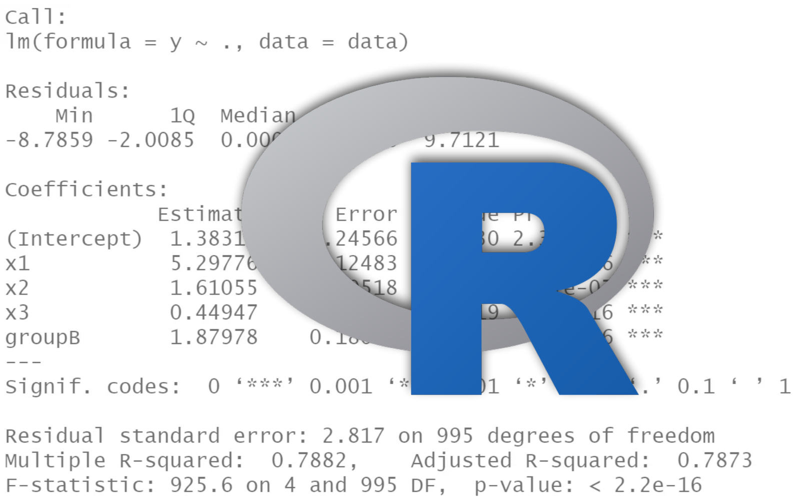


```
Call:
lm(formula = y ~ ., data = data)

Residuals:
    Min      1Q  Median
-8.7859 -2.0085  0.00          9.7121

Coefficients:
            Estimat       Error      e Pr
(Intercept)  1.3831       .24566     30 2.        *
x1           5.29776      12483            6    **
x2           1.61055        518          e-07  ***
x3           0.44947                 19      16 ***
groupB       1.87978    0.18                 6 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘*      01 ‘*’      ‘.’ 0.1 ‘ ’ 1

Residual standard error: 2.817 on 995 degrees of freedom
Multiple R-squared:  0.7882,    Adjusted R-squared:  0.7873
F-statistic: 925.6 on 4 and 995 DF,  p-value: < 2.2e-16
```

R Packages

Graphic display : ggplot2

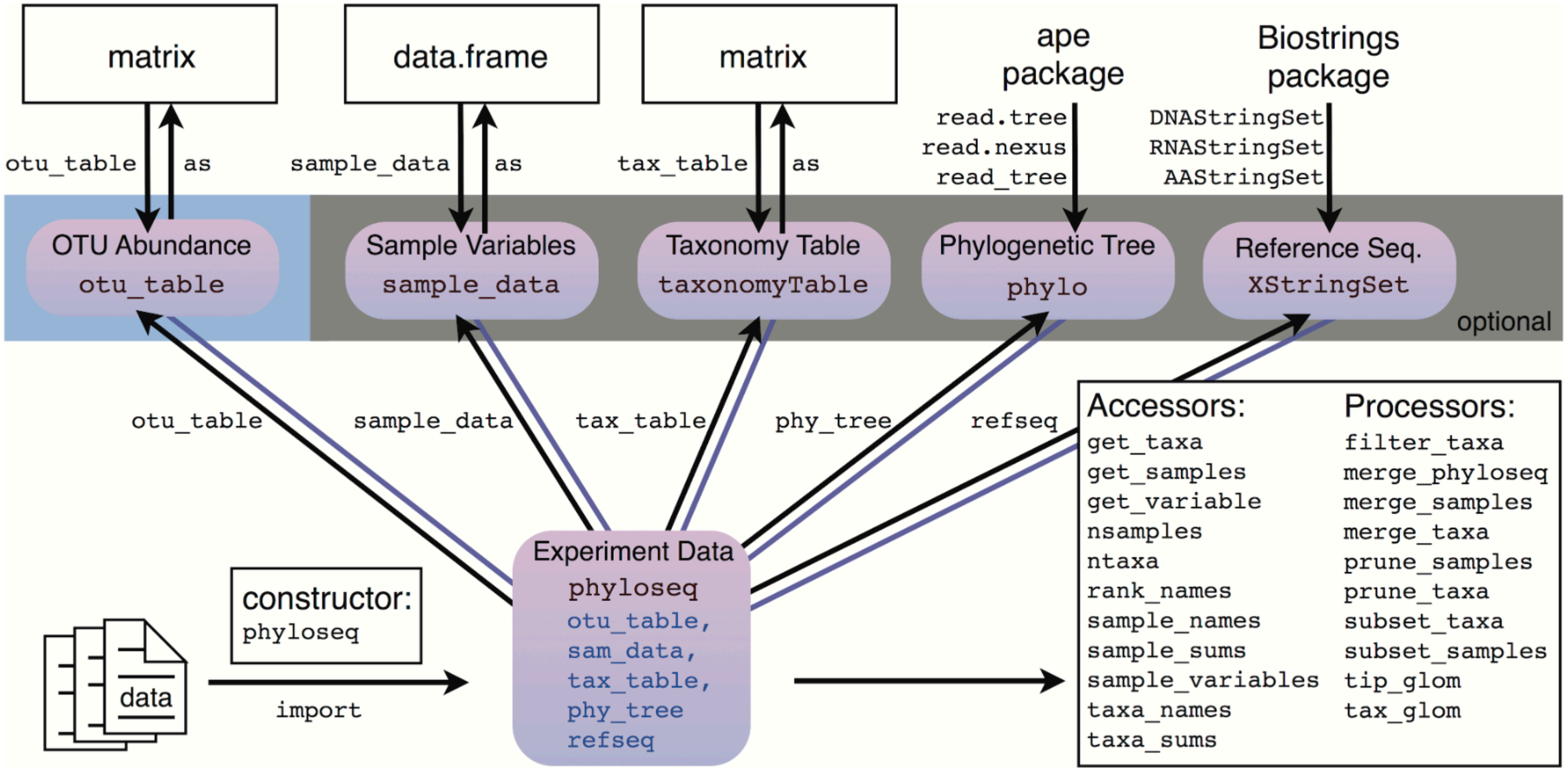


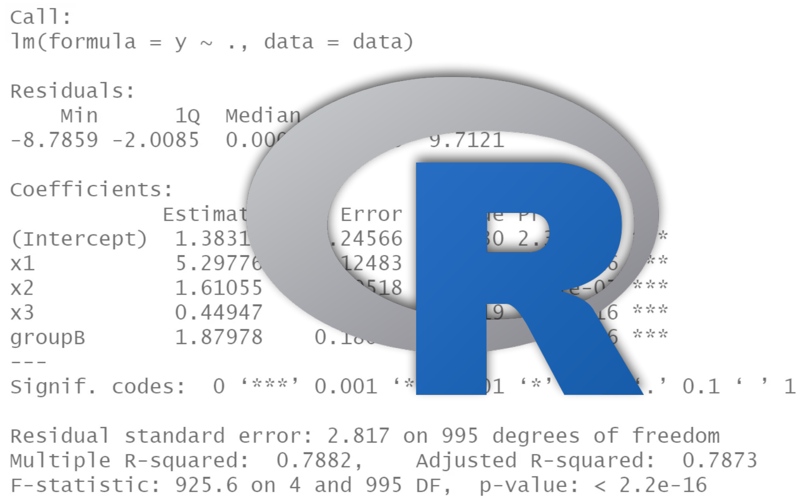


```
Call:
lm(formula = y ~ ., data = data)

Residuals:
    Min      1Q  Median
-8.7859 -2.0085  0.00              9.7121

Coefficients:
            Estimat        Error
(Intercept)  1.3831       .24566           0 2.                *
x1           5.29776       12483                      6        **
x2           1.61055        518                        e-0    ***
x3           0.44947                      19            16    ***
groupB       1.87978     0.18                            6    ***
---
Signif. codes:  0 ‘***’ 0.001 ‘*           01 ‘*’          .’ 0.1 ‘ ’ 1

Residual standard error: 2.817 on 995 degrees of freedom
Multiple R-squared:  0.7882,    Adjusted R-squared:  0.7873
F-statistic: 925.6 on 4 and 995 DF,  p-value: < 2.2e-16
```

R Packages

A package dedicated to metabarcoding data analysis : phyloseq





```
Call:
lm(formula = y ~ ., data = data)

Residuals:
    Min      1Q  Median
-8.7859 -2.0085  0.00          9.7121

Coefficients:
            Estimat      Error
(Intercept)  1.3831      .24566      30 2.          *
x1           5.29776      12483               6  **
x2           1.61055       518             e-07 ***
x3           0.44947              19         16 ***
groupB       1.87978    0.18                 6 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘*      01 ‘*’      .’ 0.1 ‘ ’ 1

Residual standard error: 2.817 on 995 degrees of freedom
Multiple R-squared:  0.7882,    Adjusted R-squared:  0.7873
F-statistic: 925.6 on 4 and 995 DF,  p-value: < 2.2e-16
```

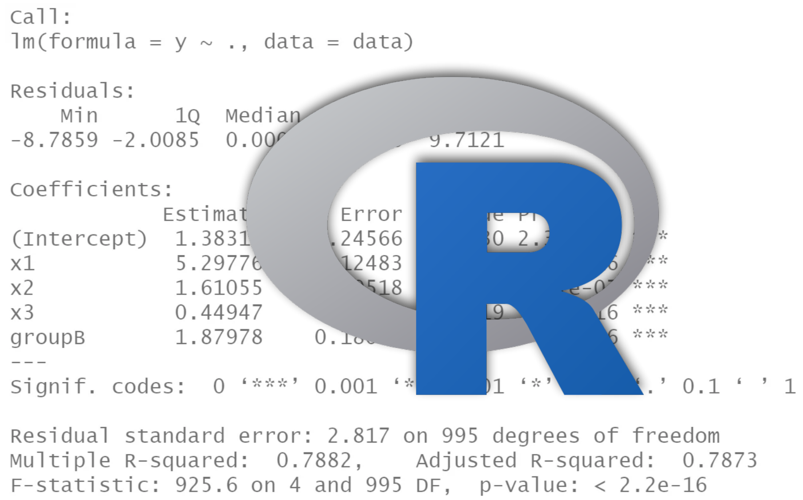
R Help

Learning how to code in R language could be the subject of an entire workshop

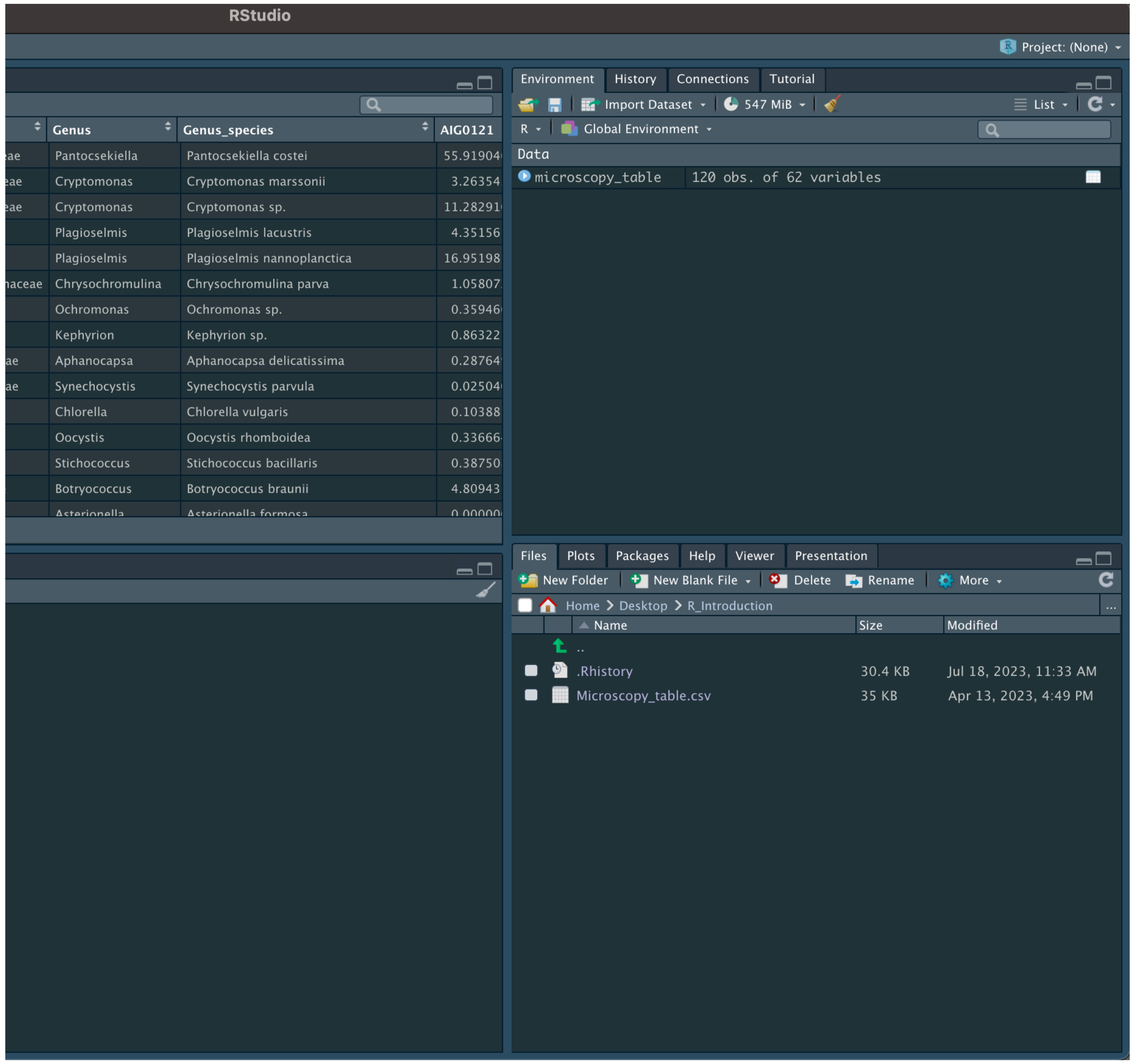
It takes time to understand how the R function works, how to have the good code that will do what we want…

An important reflex : search for help

A little introduction to R



```
Call:
lm(formula = y ~ ., data = data)

Residuals:
    Min      1Q  Median
-8.7859 -2.0085  0.00          9.7121

Coefficients:
            Estimat       Error      e Pr
(Intercept)  1.3831      .24566     30 2.        *
x1           5.29776      12483            6    **
x2           1.61055        518        e-07   ***
x3           0.44947               19      16  ***
groupB       1.87978    0.18               6  ***
---
Signif. codes:  0 ‘***’ 0.001 ‘*       01 ‘*’       ‘.’ 0.1 ‘ ’ 1

Residual standard error: 2.817 on 995 degrees of freedom
Multiple R-squared:  0.7882,	Adjusted R-squared:  0.7873
F-statistic: 925.6 on 4 and 995 DF,  p-value: < 2.2e-16
```



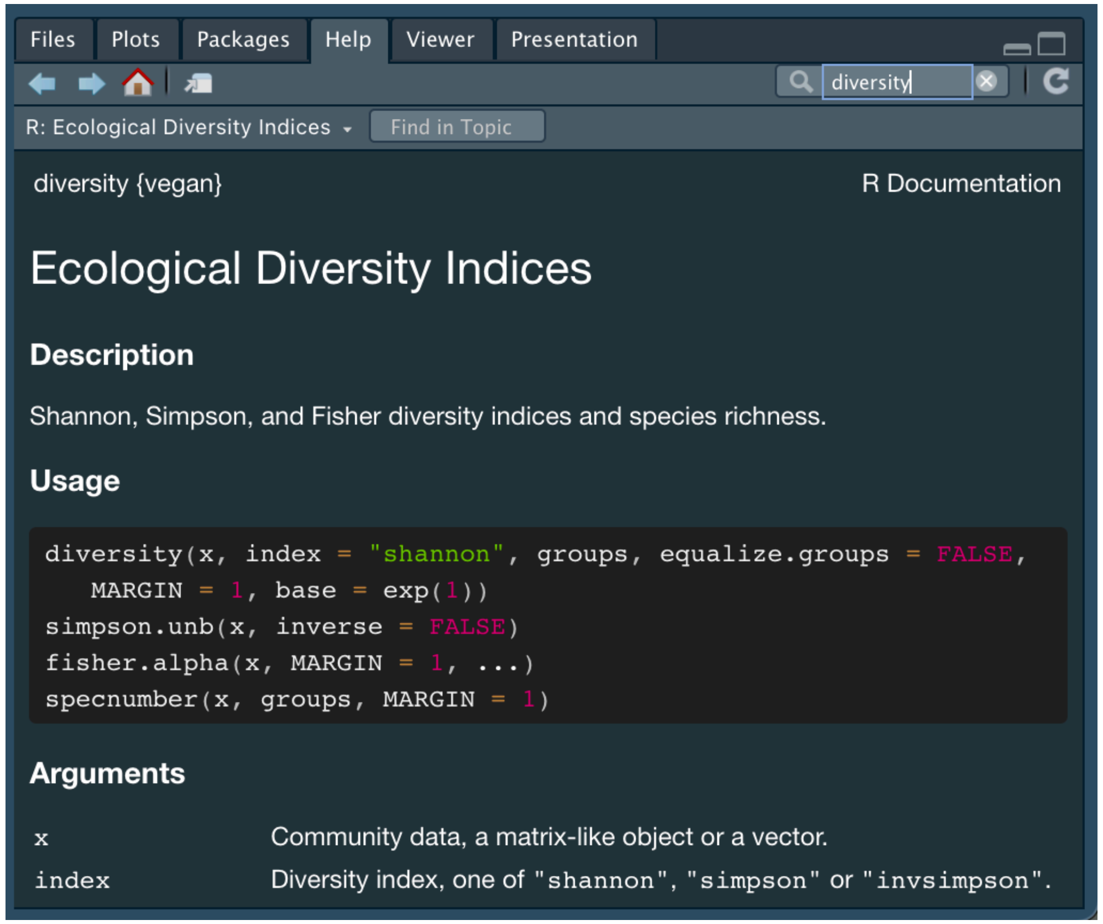
	Genus	Genus_species	AIG0121
ae	Pantocsekiella	Pantocsekiella costei	55.91904
eae	Cryptomonas	Cryptomonas marssonii	3.26354
eae	Cryptomonas	Cryptomonas sp.	11.28291
	Plagioselmis	Plagioselmis lacustris	4.35156
	Plagioselmis	Plagioselmis nannoplanctica	16.95198
naceae	Chrysochromulina	Chrysochromulina parva	1.05807
	Ochromonas	Ochromonas sp.	0.35946
	Kephyrion	Kephyrion sp.	0.86322
ae	Aphanocapsa	Aphanocapsa delicatissima	0.28764
ae	Synechocystis	Synechocystis parvula	0.02504
	Chlorella	Chlorella vulgaris	0.10388
	Oocystis	Oocystis rhomboidea	0.33666
	Stichococcus	Stichococcus bacillaris	0.38750
	Botryococcus	Botryococcus braunii	4.80943
	Asterionella	Asterionella formosa	0.00000

Data

microscopy_table	120 obs. of 62 variables

Name	Size	Modified
..		
.Rhistory	30.4 KB	Jul 18, 2023, 11:33 AM
Microscopy_table.csv	35 KB	Apr 13, 2023, 4:49 PM

R Help



diversity {vegan}	R Documentation

Ecological Diversity Indices

Description

Shannon, Simpson, and Fisher diversity indices and species richness.

Usage

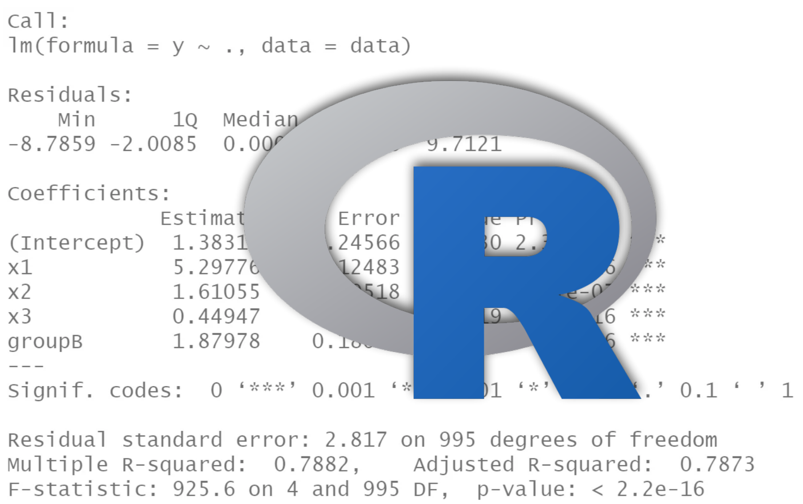
```
diversity(x, index = "shannon", groups, equalize.groups = FALSE,
   MARGIN = 1, base = exp(1))
simpson.unb(x, inverse = FALSE)
fisher.alpha(x, MARGIN = 1, ...)
specnumber(x, groups, MARGIN = 1)
```

Arguments

`x`	Community data, a matrix-like object or a vector.

`index`	Diversity index, one of "shannon", "simpson" or "invsimpson".

A little introduction to R

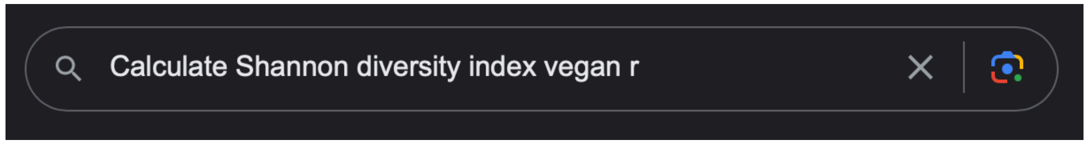


```
Call:
lm(formula = y ~ ., data = data)

Residuals:
    Min      1Q  Median
-8.7859 -2.0085  0.00          9.7121

Coefficients:
            Estimat       Error        e Pr
(Intercept)  1.3831       .24566      30 2.           *
x1           5.29776      12483               6      **
x2           1.61055        518             e-0      ***
x3           0.44947                  9          16  ***
groupB       1.87978    0.18                     6  ***
---
Signif. codes:  0 ‘***’ 0.001 ‘*      01 ‘*’      .’ 0.1 ‘ ’ 1

Residual standard error: 2.817 on 995 degrees of freedom
Multiple R-squared:  0.7882,    Adjusted R-squared:  0.7873
F-statistic: 925.6 on 4 and 995 DF,  p-value: < 2.2e-16
```

R Help

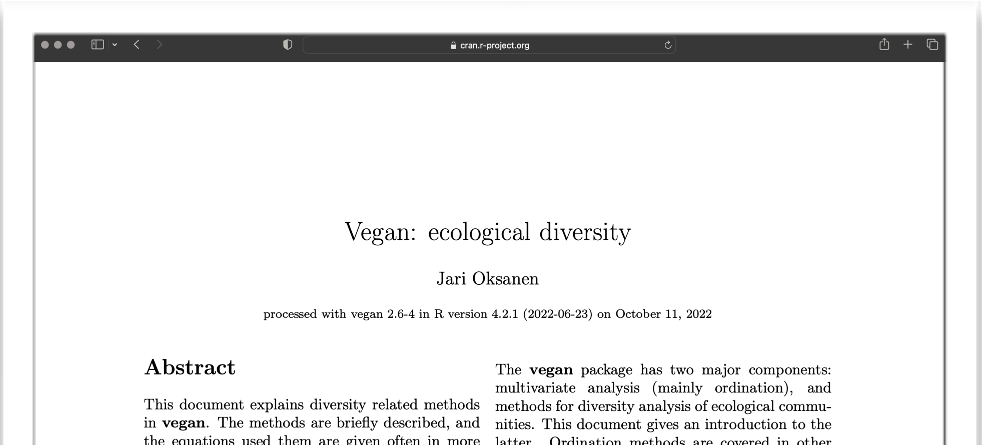
Any question has an answer on internet !!

Official documentation of Vegan package



Calculate Shannon diversity index vegan r

Tutorial made by other scientists



cran.r-project.org

Vegan: ecological diversity

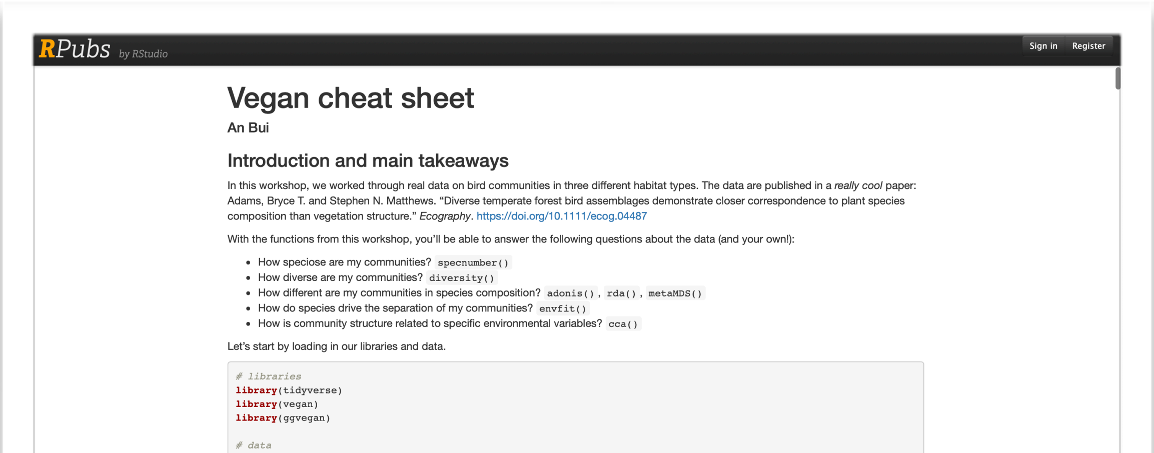
Jari Oksanen

processed with vegan 2.6-4 in R version 4.2.1 (2022-06-23) on October 11, 2022

Abstract

This document explains diversity related methods in **vegan**. The methods are briefly described, and the equations used them are given often in more detail than in their help pages. The methods d cussed include common diversity indices and r efaction, families of diversity indices, species ab dance models, species accumulation models a beta diversity, extrapolated richness and probab ity of being a member of the species pool. T document is still incomplete and does not cover

The **vegan** package has two major components: multivariate analysis (mainly ordination), and methods for diversity analysis of ecological communities. This document gives an introduction to the latter. Ordination methods are covered in other



RPubs by RStudio

Sign in Register

Vegan cheat sheet

An Bui

Introduction and main takeaways

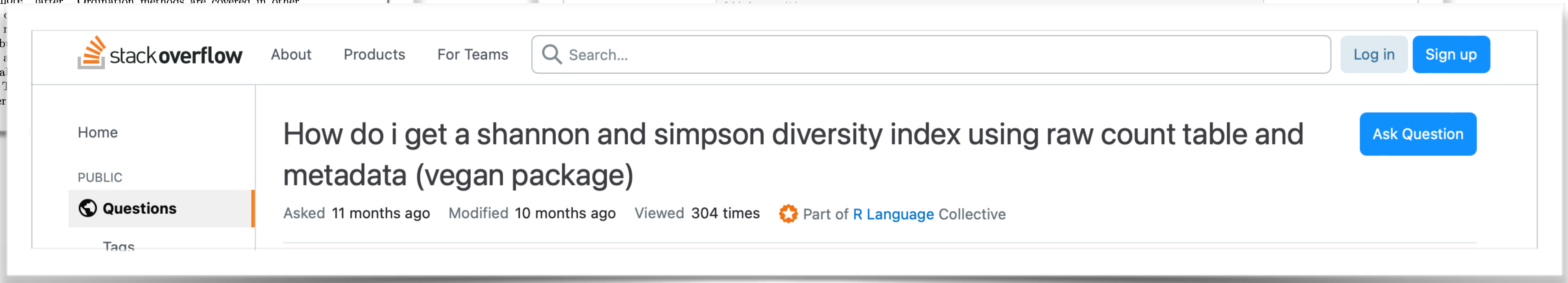
In this workshop, we worked through real data on bird communities in three different habitat types. The data are published in a *really cool* paper: Adams, Bryce T. and Stephen N. Matthews. “Diverse temperate forest bird assemblages demonstrate closer correspondence to plant species composition than vegetation structure.” *Ecography*. https://doi.org/10.1111/ecog.04487

With the functions from this workshop, you’ll be able to answer the following questions about the data (and your own!):

- How speciose are my communities? `specnumber()`
- How diverse are my communities? `diversity()`
- How different are my communities in species composition? `adonis()`, `rda()`, `metaMDS()`
- How do species drive the separation of my communities? `envfit()`
- How is community structure related to specific environmental variables? `cca()`

Let’s start by loading in our libraries and data.

```
# libraries
library(tidyverse)
library(vegan)
library(ggvegan)

# data
```



stackoverflow About Products For Teams Search... Log in Sign up

Home

PUBLIC

Questions

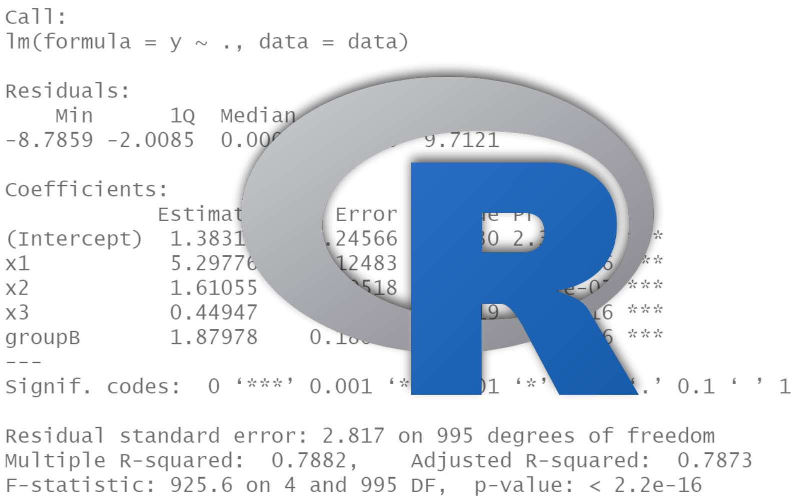
Tags

How do i get a shannon and simpson diversity index using raw count table and metadata (vegan package)

Ask Question

Asked 11 months ago Modified 10 months ago Viewed 304 times Part of R Language Collective

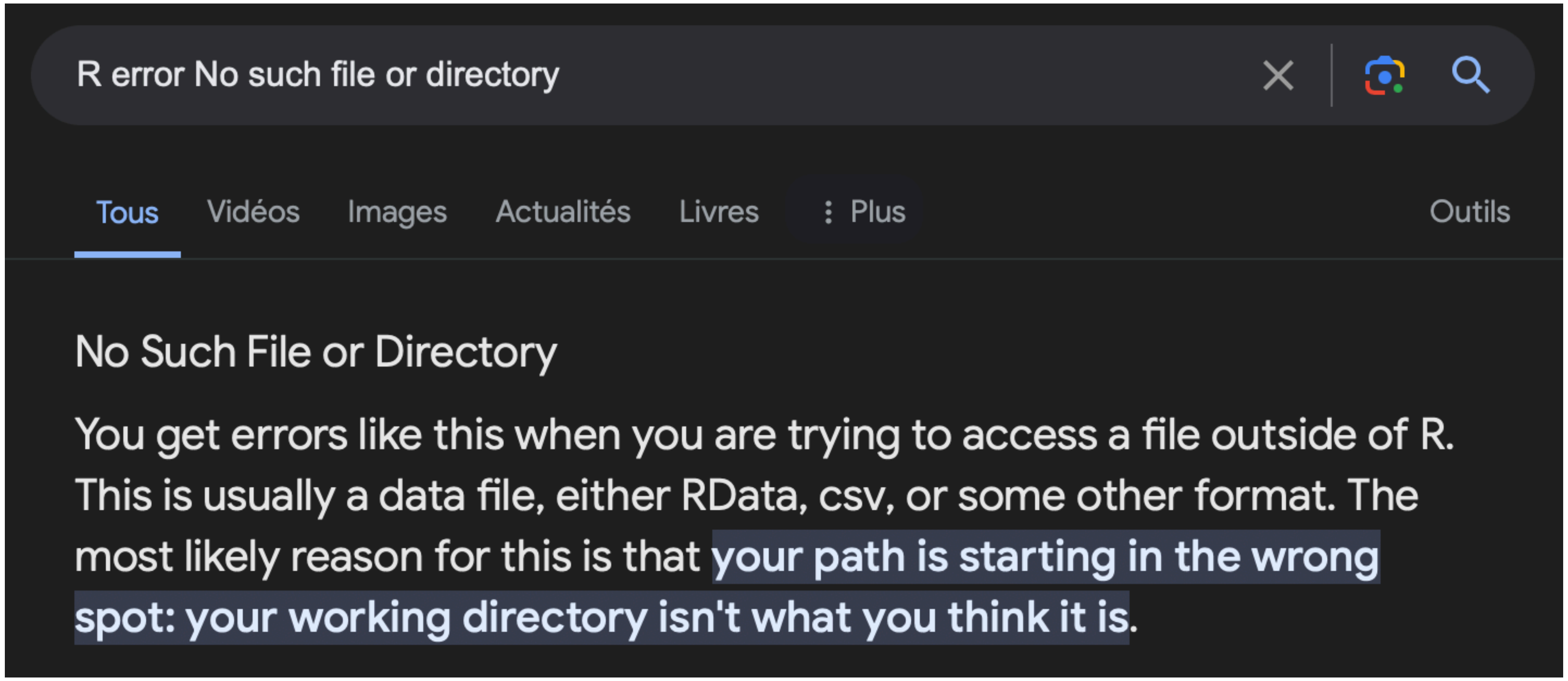
Forums



```
Call:
lm(formula = y ~ ., data = data)

Residuals:
    Min      1Q  Median
-8.7859 -2.0085  0.00          9.7121

Coefficients:
            Estimat       Error         e Pr
(Intercept)  1.3831       .24566       30 2.        *
x1           5.29776      12483              6    **
x2           1.61055        518           e-0  ***
x3           0.44947                 19      16 ***
groupB       1.87978     0.18              6 ***
---
Signif. codes:  0 '***' 0.001 '*       01 '*'      '.' 0.1 ' ' 1

Residual standard error: 2.817 on 995 degrees of freedom
Multiple R-squared:  0.7882,    Adjusted R-squared:  0.7873
F-statistic: 925.6 on 4 and 995 DF,  p-value: < 2.2e-16
```

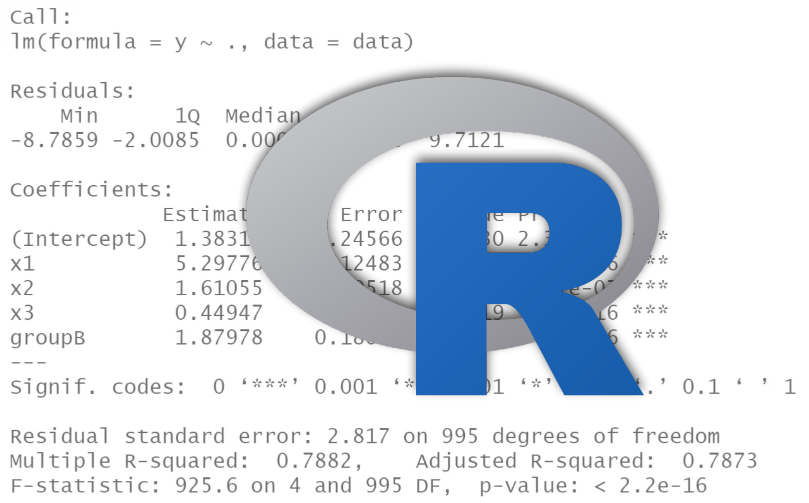
R Help

Any question has an answer on internet !!



No Such File or Directory

You get errors like this when you are trying to access a file outside of R. This is usually a data file, either RData, csv, or some other format. The most likely reason for this is that your path is starting in the wrong spot: your working directory isn't what you think it is.



```
Call:
lm(formula = y ~ ., data = data)

Residuals:
    Min      1Q  Median
-8.7859 -2.0085  0.00        9.7121

Coefficients:
            Estimat       Error         e Pr
(Intercept)  1.3831       .24566        30 2.       *
x1           5.29776      12483              6    **
x2           1.61055       518            e-0    ***
x3           0.44947                 19      16  ***
groupB       1.87978    0.18              6  ***
---
Signif. codes:  0 ‘***’ 0.001 ‘*      01 ‘*’      ‘.’ 0.1 ‘ ’ 1

Residual standard error: 2.817 on 995 degrees of freedom
Multiple R-squared:  0.7882,    Adjusted R-squared:  0.7873
F-statistic: 925.6 on 4 and 995 DF,  p-value: < 2.2e-16
```

TYPICAL ANALYSIS WITH R

1. OPEN OR START AN R SCRIPT

2. SET WORKING DIRECTORY - FILE PATHS

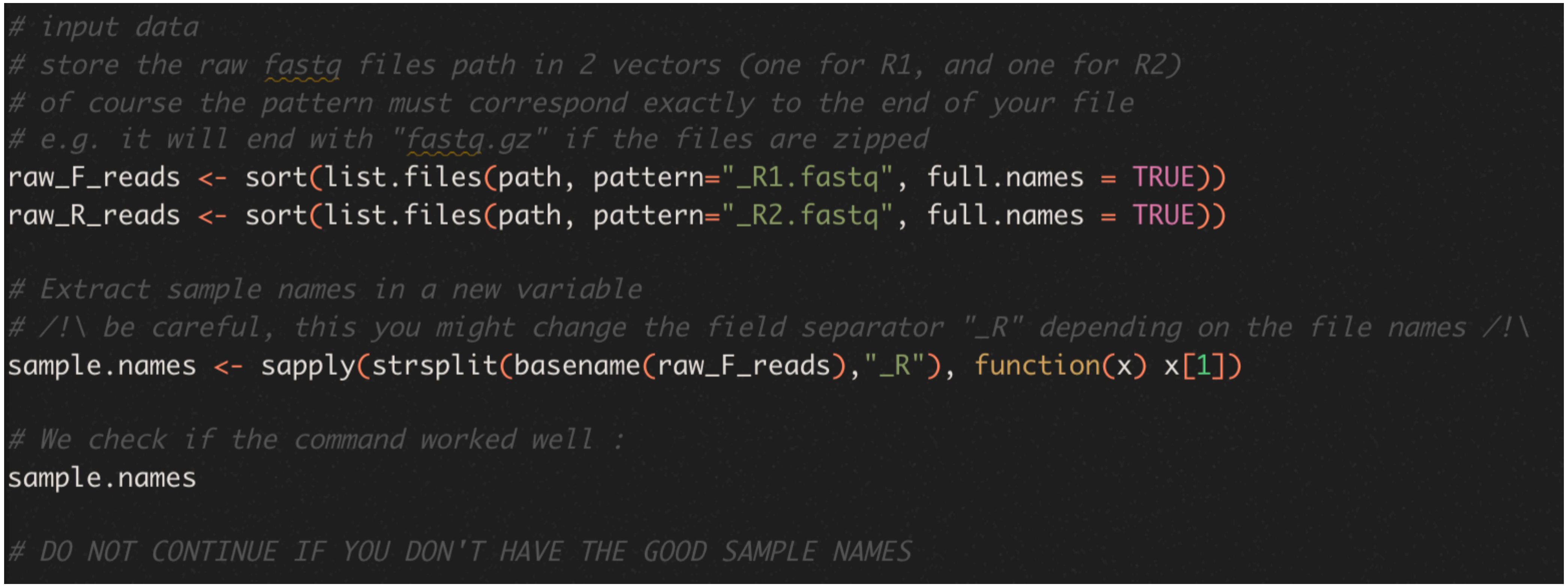
3. LOAD R PACKAGES

4. BE CAREFUL OF ERRORS

5. ONCE A SCRIPT WORKED, SAVE IT

TYPICAL ANALYSIS WITH R

1. OPEN OR START AN R SCRIPT



```
# input data
# store the raw fastq files path in 2 vectors (one for R1, and one for R2)
# of course the pattern must correspond exactly to the end of your file
# e.g. it will end with "fastq.gz" if the files are zipped
raw_F_reads <- sort(list.files(path, pattern="_R1.fastq", full.names = TRUE))
raw_R_reads <- sort(list.files(path, pattern="_R2.fastq", full.names = TRUE))

# Extract sample names in a new variable
# /!\ be careful, this you might change the field separator "_R" depending on the file names /!\
sample.names <- sapply(strsplit(basename(raw_F_reads),"_R"), function(x) x[1])

# We check if the command worked well :
sample.names

# DO NOT CONTINUE IF YOU DON'T HAVE THE GOOD SAMPLE NAMES
```

An R script is a succession of command lines (functions) to perform an analysis

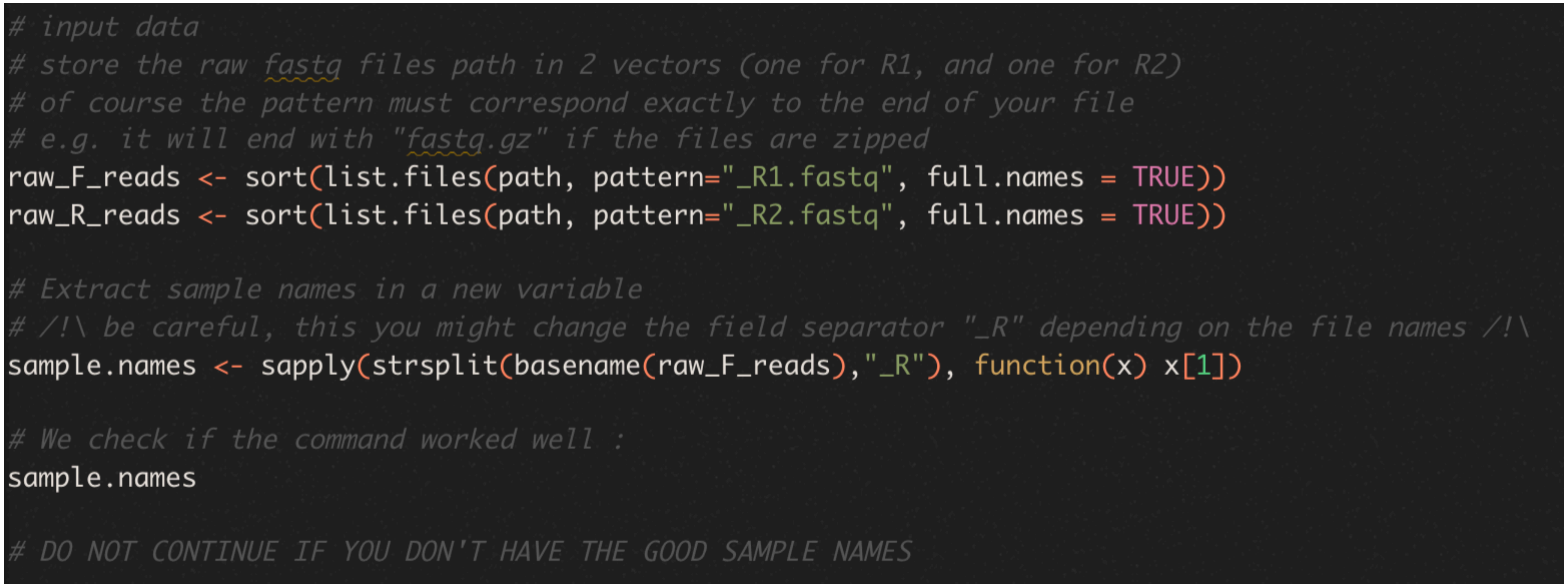
It needs to be adapted to your data / computer

It is important to carefully do it in order and regularly check if everything worked

If one step was not successful, the rest of the script will not work

TYPICAL ANALYSIS WITH R

1. OPEN OR START AN R SCRIPT



```
# input data
# store the raw fastq files path in 2 vectors (one for R1, and one for R2)
# of course the pattern must correspond exactly to the end of your file
# e.g. it will end with "fastq.gz" if the files are zipped
raw_F_reads <- sort(list.files(path, pattern="_R1.fastq", full.names = TRUE))
raw_R_reads <- sort(list.files(path, pattern="_R2.fastq", full.names = TRUE))

# Extract sample names in a new variable
# /!\ be careful, this you might change the field separator "_R" depending on the file names /!\
sample.names <- sapply(strsplit(basename(raw_F_reads),"_R"), function(x) x[1])

# We check if the command worked well :
sample.names

# DO NOT CONTINUE IF YOU DON'T HAVE THE GOOD SAMPLE NAMES
```

<u>When you want to run a command line in a script :</u>

Click somewhere in the line

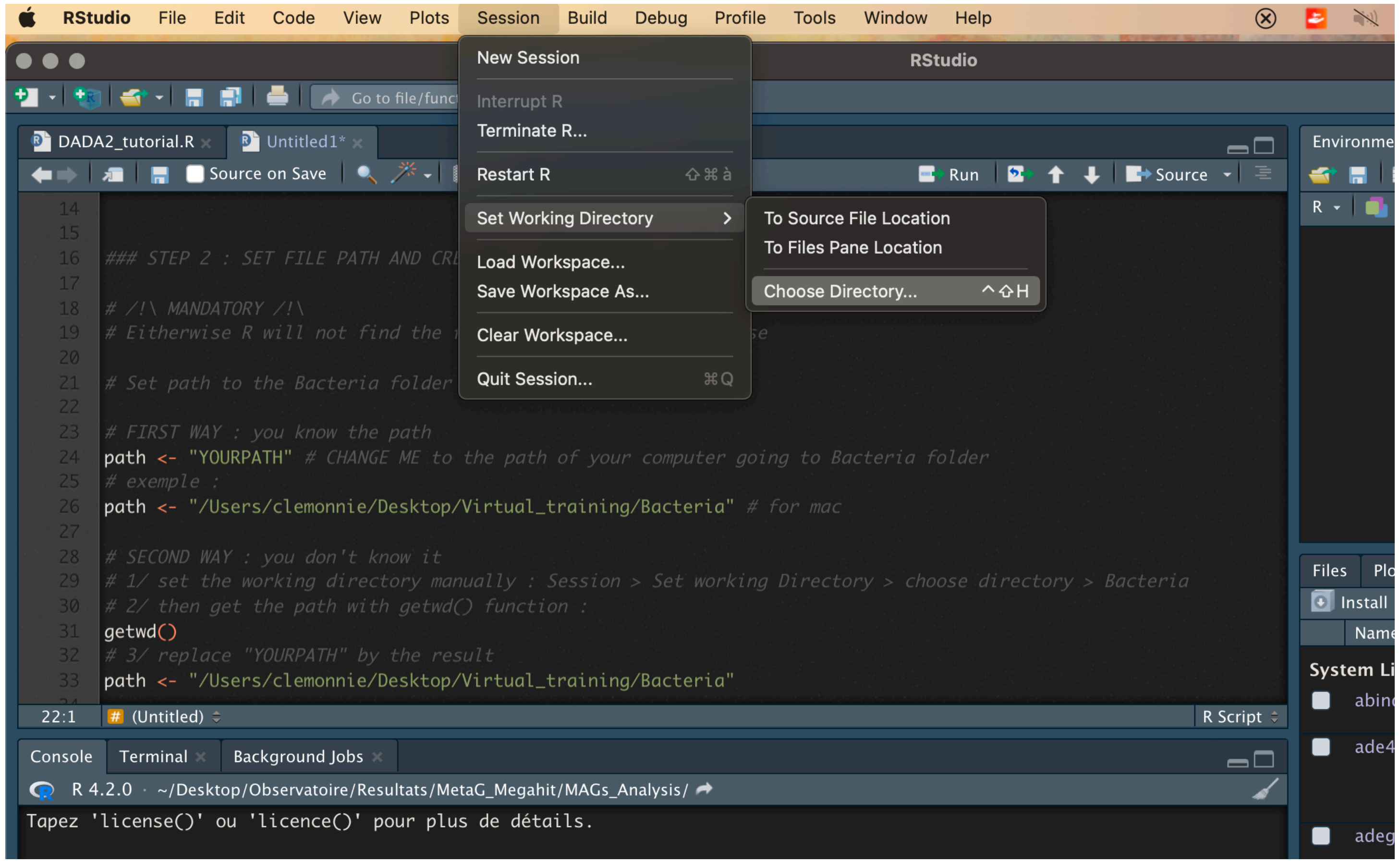
And then do : ctrl + enter (for Mac it is command + enter)

TYPICAL ANALYSIS WITH R

1. OPEN OR START AN R SCRIPT

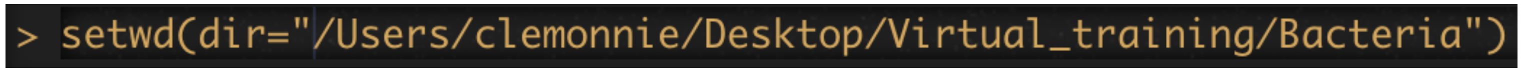
2. SET WORKING DIRECTORY - FILE PATHS

R does not know where are the files you want to analyse. You need to specify it.

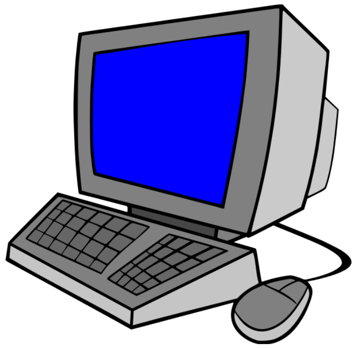


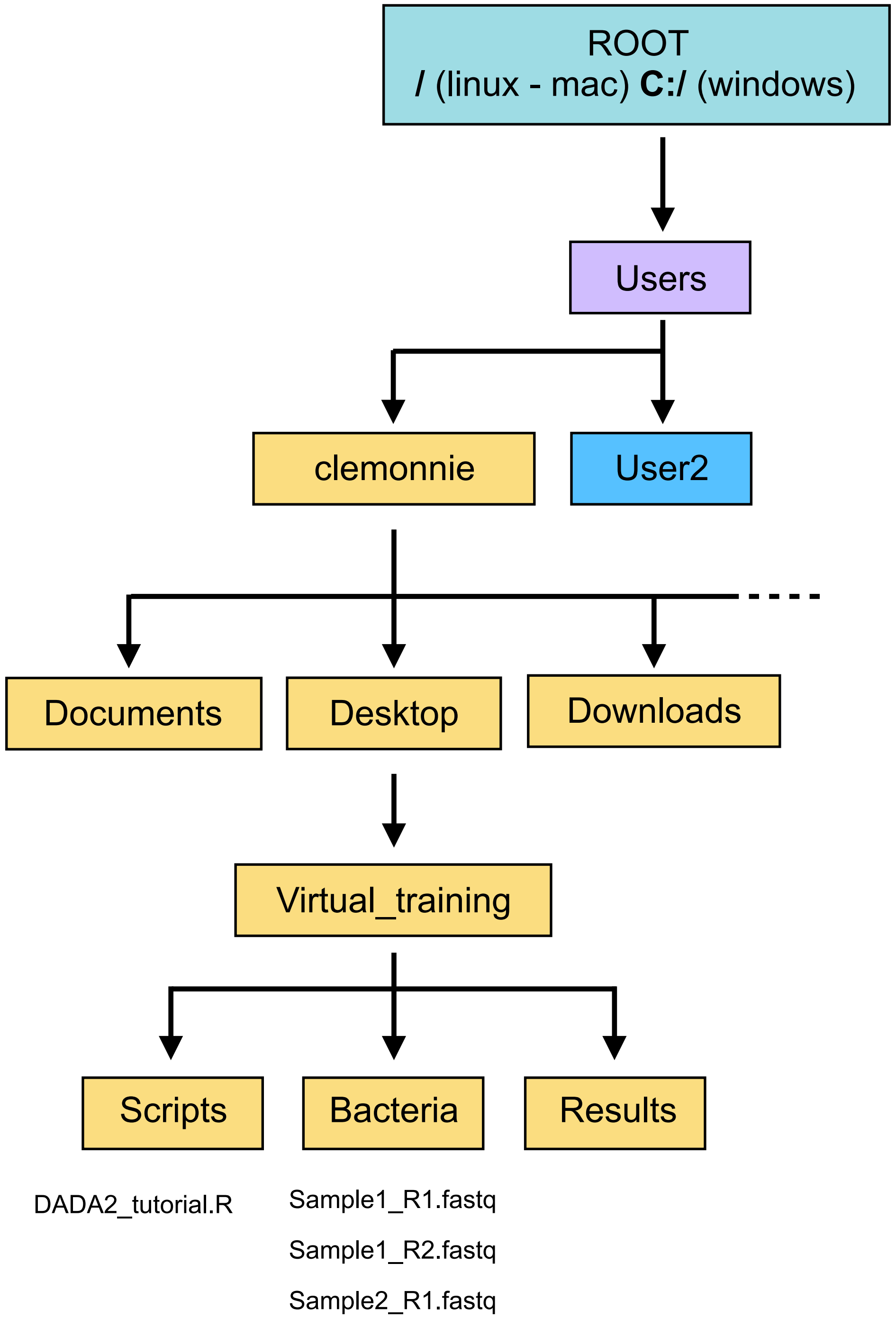
```


### STEP 2 : SET FILE PATH AND CR

# /!\ MANDATORY /!\
# Eitherwise R will not find the                                  se

# Set path to the Bacteria folder

# FIRST WAY : you know the path
path <- "YOURPATH" # CHANGE ME to the path of your computer going to Bacteria folder
# exemple :
path <- "/Users/clemonnie/Desktop/Virtual_training/Bacteria" # for mac

# SECOND WAY : you don't know it
# 1/ set the working directory manually : Session > Set working Directory > choose directory > Bacteria
# 2/ then get the path with getwd() function :
getwd()
# 3/ replace "YOURPATH" by the result
path <- "/Users/clemonnie/Desktop/Virtual_training/Bacteria"
```

```
R 4.2.0 · ~/Desktop/Observatoire/Resultats/MetaG_Megahit/MAGs_Analysis/
Tapez 'license()' ou 'licence()' pour plus de détails.
```

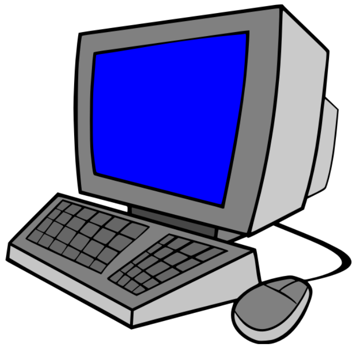


```
> setwd(dir="/Users/clemonnie/Desktop/Virtual_training/Bacteria")
```





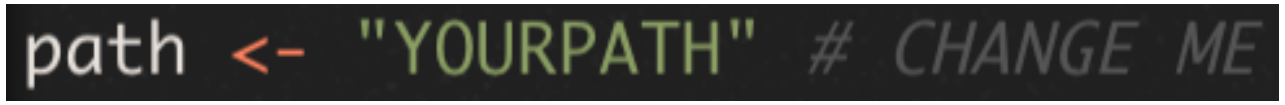
TYPICAL ANALYSIS WITH R



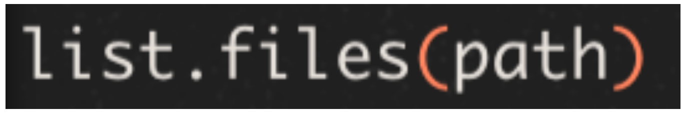
1. OPEN OR START AN R SCRIPT

2. SET WORKING DIRECTORY - FILE PATHS

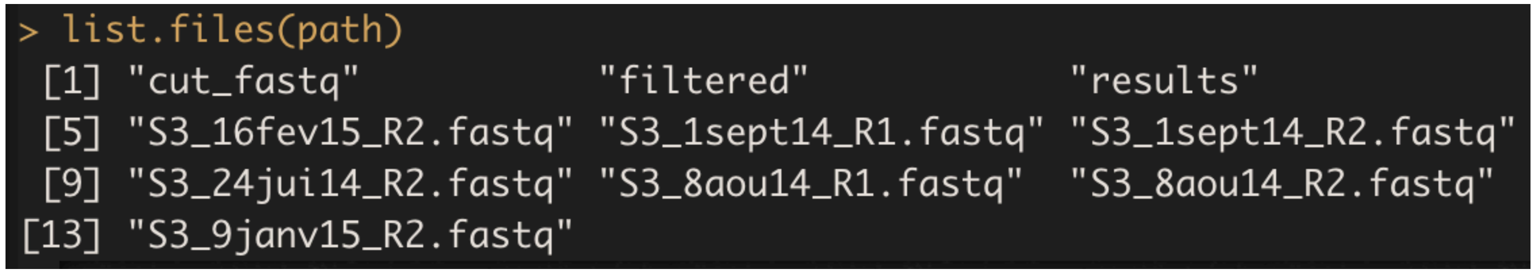
In the DADA2 pipeline, we will tell the path



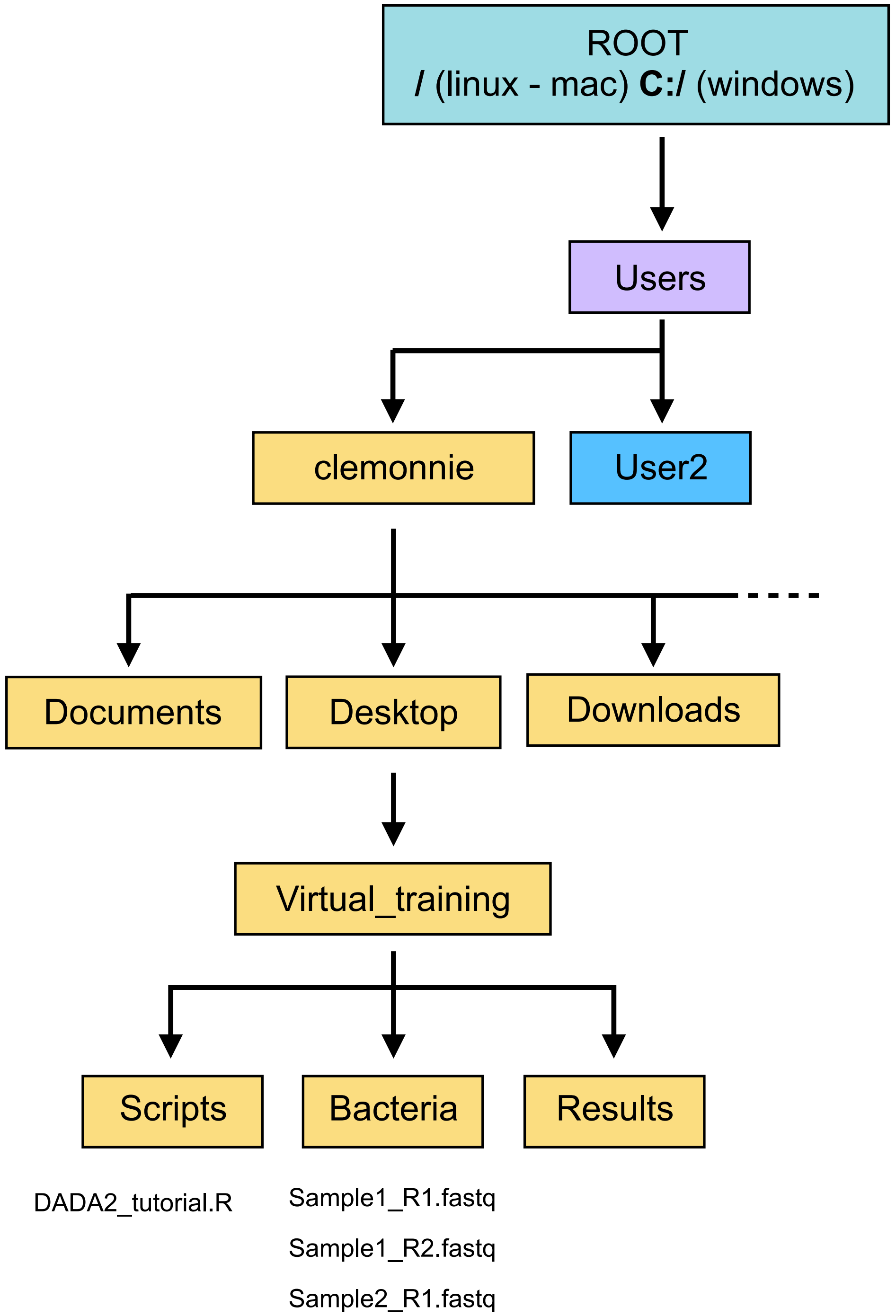
```
path <- "YOURPATH" # CHANGE ME
```



```
list.files(path)
```



```
> list.files(path)
 [1] "cut_fastq"            "filtered"             "results"
 [5] "S3_16fev15_R2.fastq" "S3_1sept14_R1.fastq" "S3_1sept14_R2.fastq"
 [9] "S3_24jui14_R2.fastq" "S3_8aou14_R1.fastq"   "S3_8aou14_R2.fastq"
[13] "S3_9janv15_R2.fastq"
```



ROOT
/ (linux - mac) **C:/** (windows)

Users

clemonnie

User2

Documents

Desktop

Downloads

Virtual_training

Scripts

Bacteria

Results

DADA2_tutorial.R

Sample1_R1.fastq
Sample1_R2.fastq
Sample2_R1.fastq

TYPICAL ANALYSIS WITH R

1. OPEN OR START AN R SCRIPT

2. SET WORKING DIRECTORY - FILE PATHS

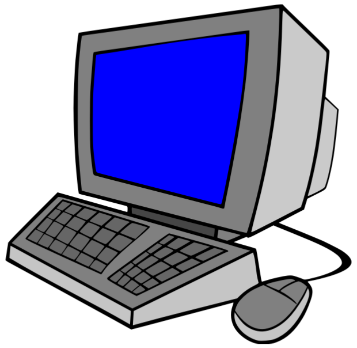
In the DADA2 pipeline, we will tell the path

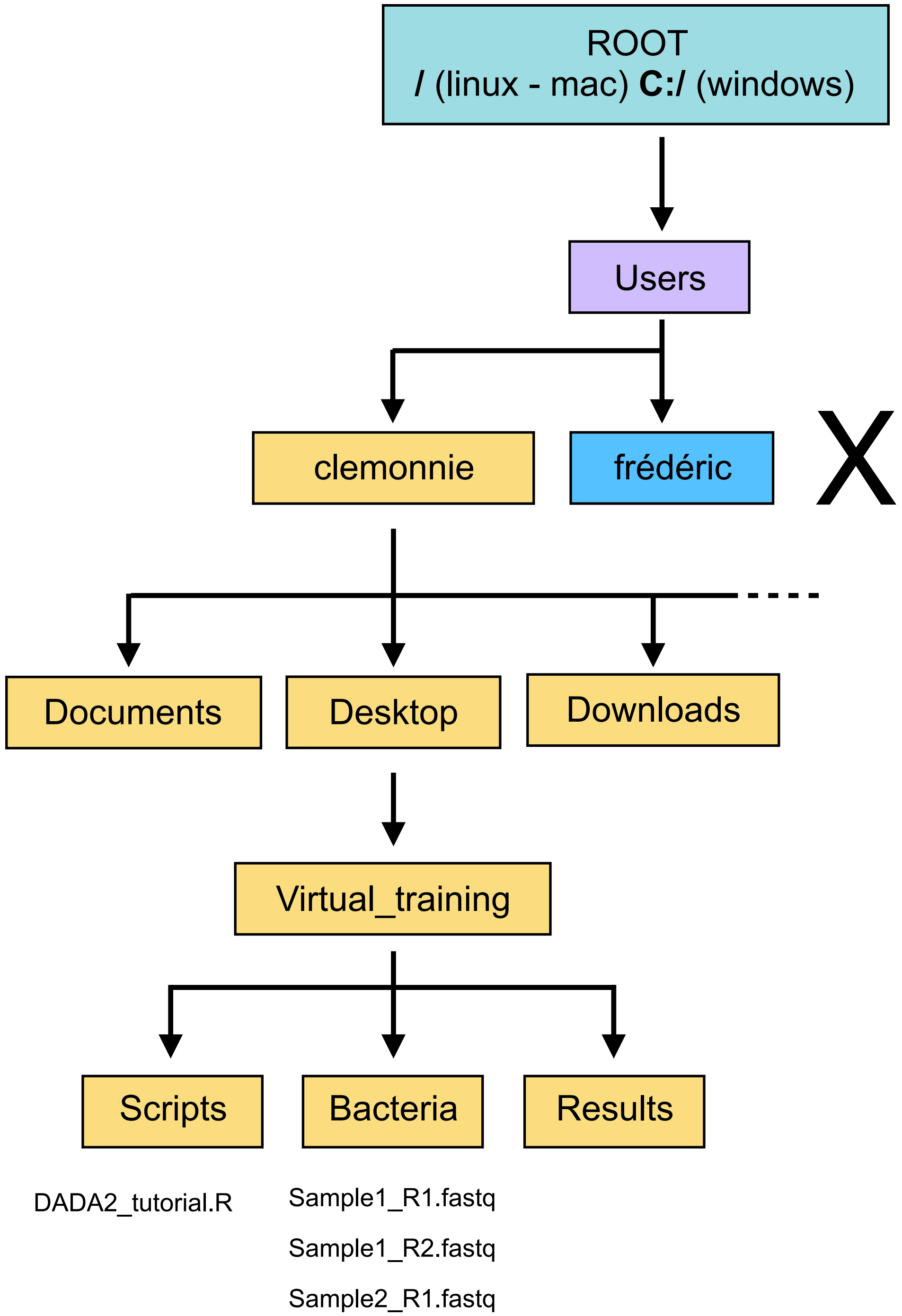
BE CAREFUL :

To space in the folder names

Virtual_training   Virtual training X

To special characters (e.g. accent)





ROOT
/ (linux - mac) **C:/** (windows)

Users

clemonnie

frédéric X

Documents

Desktop

Downloads

Virtual_training

Scripts

Bacteria

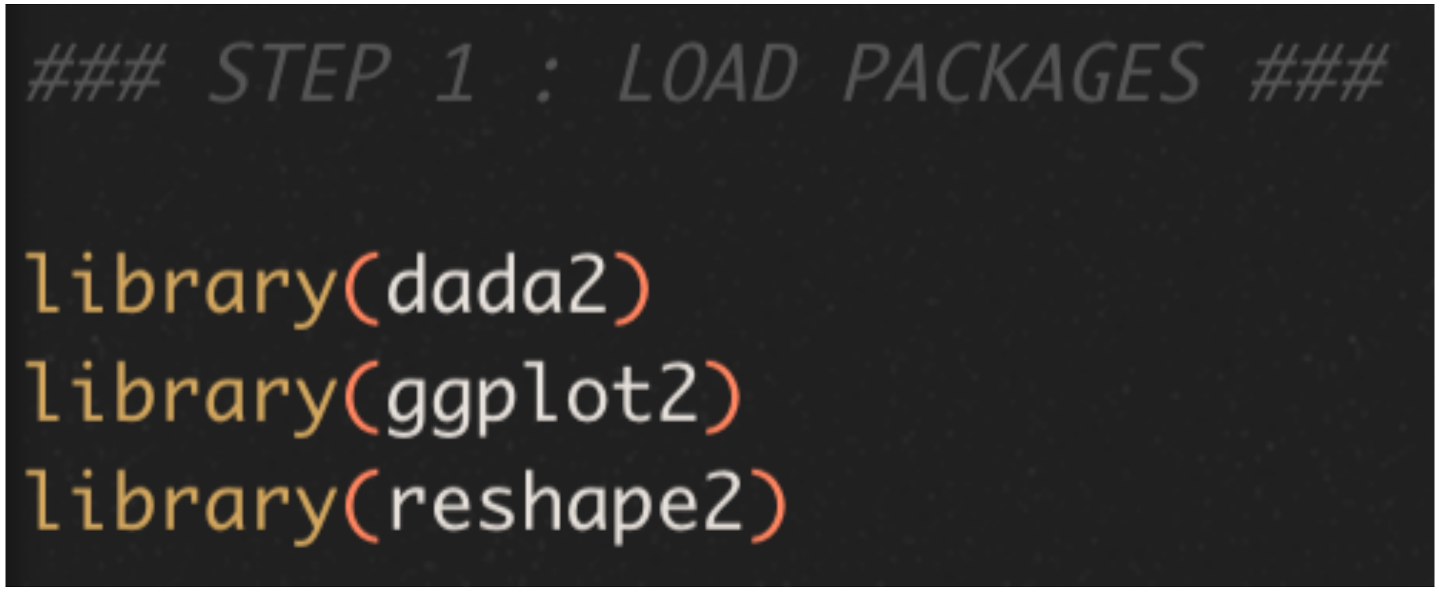
Results

DADA2_tutorial.R

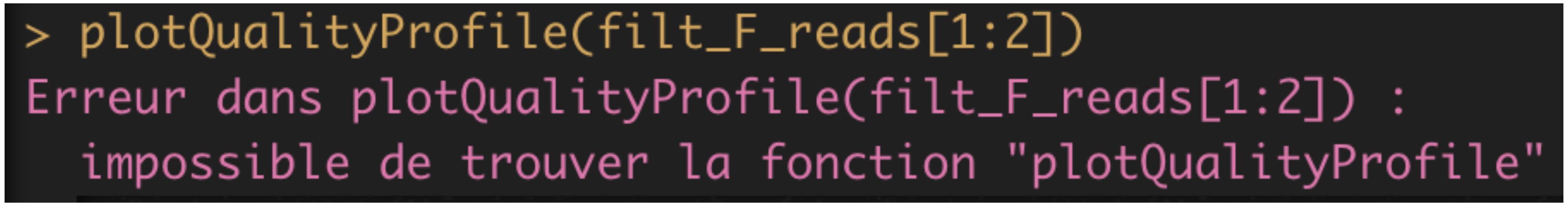
Sample1_R1.fastq
Sample1_R2.fastq
Sample2_R1.fastq

TYPICAL ANALYSIS WITH R

1. OPEN OR START AN R SCRIPT
2. SET WORKING DIRECTORY - FILE PATHS
3. LOAD R PACKAGES



```
### STEP 1 : LOAD PACKAGES ###

library(dada2)
library(ggplot2)
library(reshape2)
```

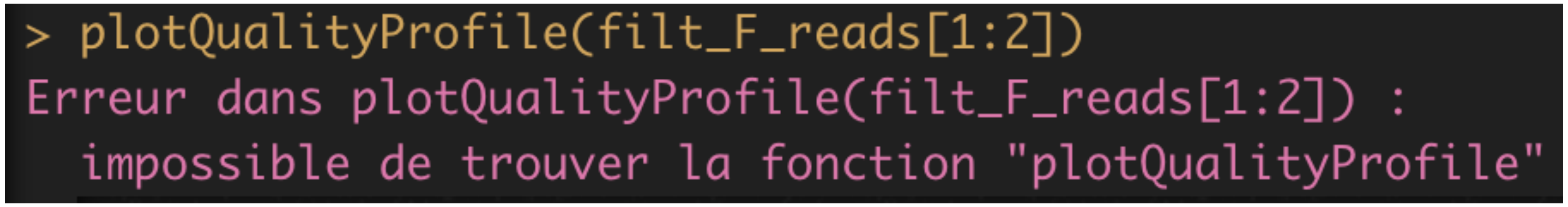


```
> plotQualityProfile(filt_F_reads[1:2])
Erreur dans plotQualityProfile(filt_F_reads[1:2]) :
  impossible de trouver la fonction "plotQualityProfile"
```

TYPICAL ANALYSIS WITH R

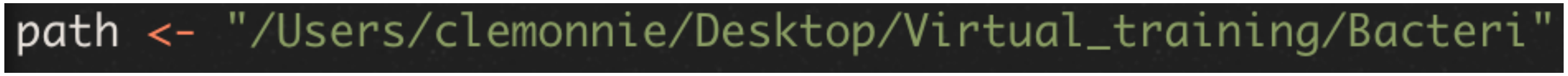
1. OPEN OR START AN R SCRIPT
2. SET WORKING DIRECTORY - FILE PATHS
3. LOAD R PACKAGES
4. BE CAREFUL OF ERRORS

Sometimes R write down the error directly

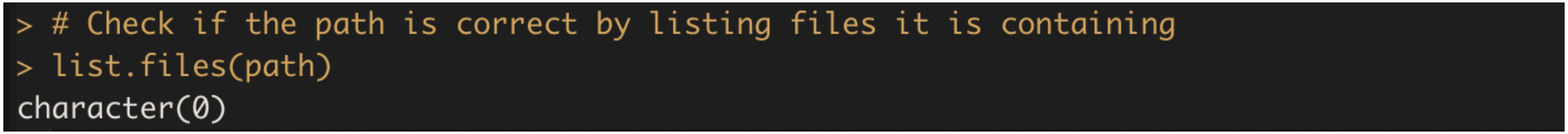


```
> plotQualityProfile(filt_F_reads[1:2])
Erreur dans plotQualityProfile(filt_F_reads[1:2]) :
  impossible de trouver la fonction "plotQualityProfile"
```

But sometime it does not know that you made a mistake :



```
path <- "/Users/clemonnie/Desktop/Virtual_training/Bacteri"
```



```
> # Check if the path is correct by listing files it is containing
> list.files(path)
character(0)
```

TYPICAL ANALYSIS WITH R

1. OPEN OR START AN R SCRIPT

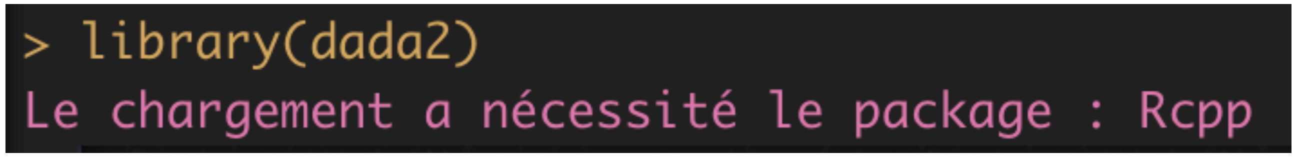
2. SET WORKING DIRECTORY - FILE PATHS

3. LOAD R PACKAGES

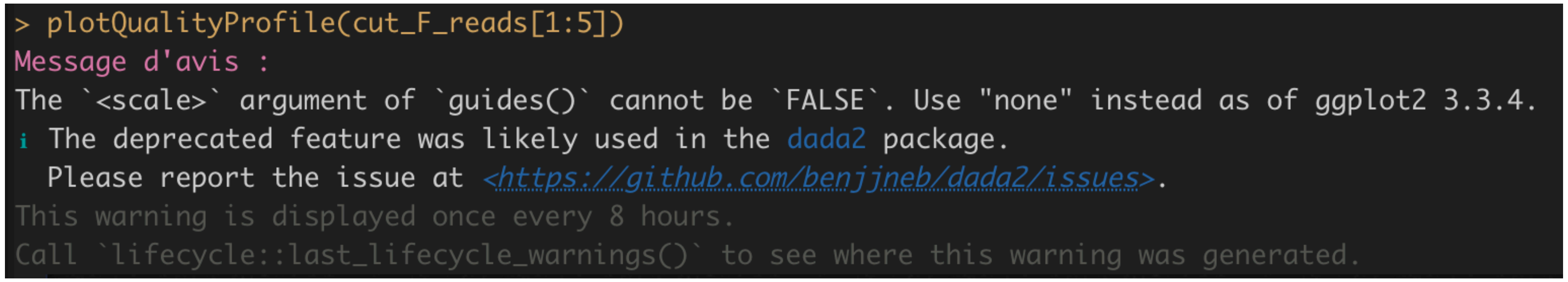
4. BE CAREFUL OF ERRORS

All things written in red does not mean an error

Sometimes it is just that R wants to speak to you



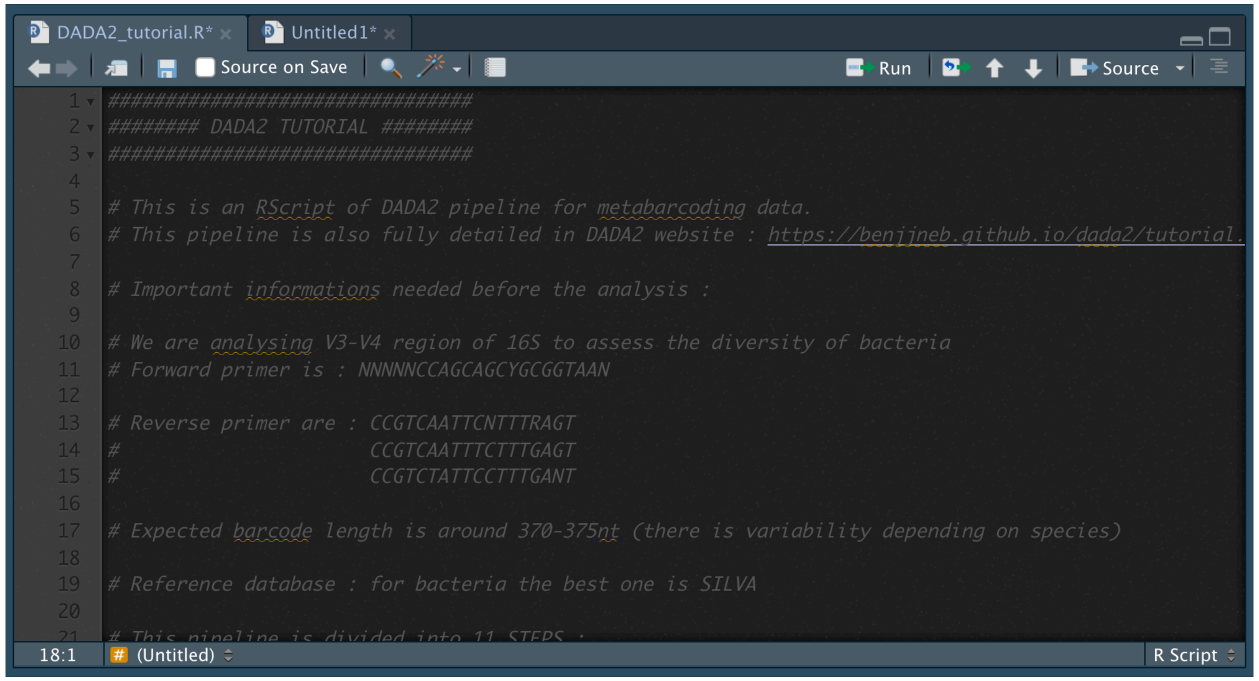
```
> library(dada2)
Le chargement a nécessité le package : Rcpp
```



```
> plotQualityProfile(cut_F_reads[1:5])
Message d'avis :
The `<scale>` argument of `guides()` cannot be `FALSE`. Use "none" instead as of ggplot2 3.3.4.
ℹ The deprecated feature was likely used in the dada2 package.
  Please report the issue at <https://github.com/benjjneb/dada2/issues>.
This warning is displayed once every 8 hours.
Call `lifecycle::last_lifecycle_warnings()` to see where this warning was generated.
```

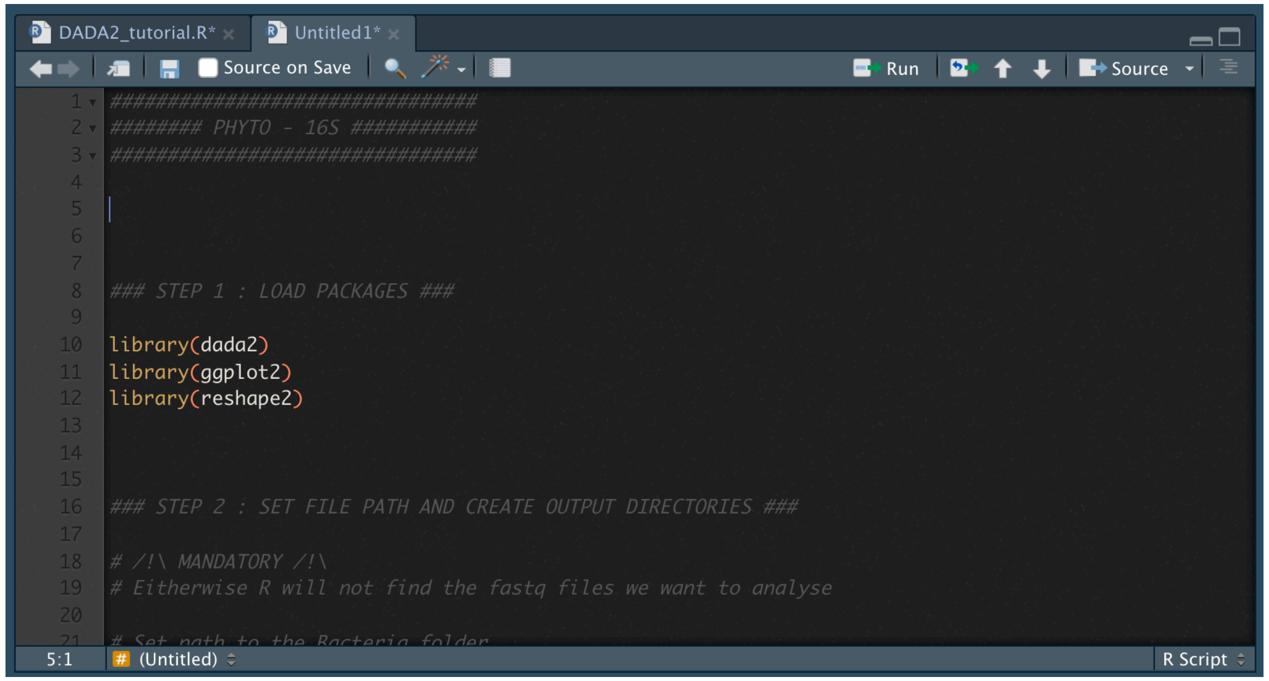
TYPICAL ANALYSIS WITH R

1. OPEN OR START AN R SCRIPT
2. SET WORKING DIRECTORY - FILE PATHS
3. LOAD R PACKAGES
4. BE CAREFUL OF ERRORS
5. ONCE A SCRIPT WORKED, SAVE IT - so you can re-use it later



```
################################
######## DADA2 TUTORIAL ########
################################

# This is an RScript of DADA2 pipeline for metabarcoding data.
# This pipeline is also fully detailed in DADA2 website : https://benjjneb.github.io/dada2/tutorial.

# Important informations needed before the analysis :

# We are analysing V3-V4 region of 16S to assess the diversity of bacteria
# Forward primer is : NNNNNCCAGCAGCYGCGGTAAN

# Reverse primer are : CCGTCAATTCNTTTRAGT
#                      CCGTCAATTTCTTTGAGT
#                      CCGTCTATTCCTTTGANT

# Expected barcode length is around 370-375nt (there is variability depending on species)

# Reference database : for bacteria the best one is SILVA

# This pipeline is divided into 11 STEPS :
```

DADA2_tutorial.R



```
################################
######## PHYTO - 16S ###########
################################




### STEP 1 : LOAD PACKAGES ###

library(dada2)
library(ggplot2)
library(reshape2)



### STEP 2 : SET FILE PATH AND CREATE OUTPUT DIRECTORIES ###

# /!\ MANDATORY /!\
# Eitherwise R will not find the fastq files we want to analyse

# Set path to the Bacteria folder
```

Phyto_16S.R

ANY QUESTIONS ?

PREPARATION OF THE PRACTICAL PART

Does everyone has RStudio and all the following packages ?

RStudio®

ggplot2

DADA²

RESHAPE2

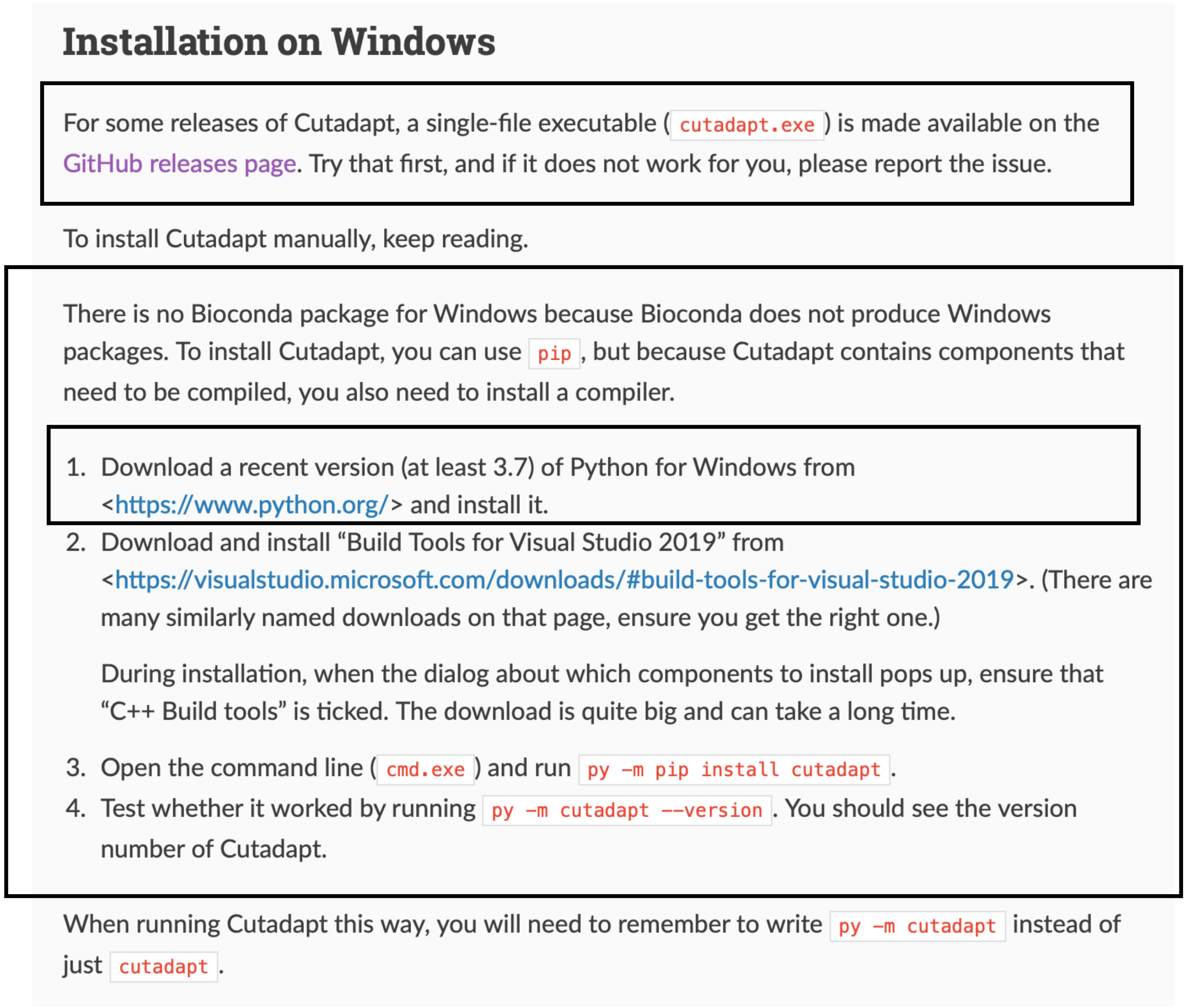
library(ggplot2)

library(dada2)

library(reshape2)

PREPARATION OF THE PRACTICAL PART

Does everyone has cutadapt and python ?



Installation on Windows

For some releases of Cutadapt, a single-file executable (`cutadapt.exe`) is made available on the GitHub releases page. Try that first, and if it does not work for you, please report the issue.

To install Cutadapt manually, keep reading.

There is no Bioconda package for Windows because Bioconda does not produce Windows packages. To install Cutadapt, you can use `pip`, but because Cutadapt contains components that need to be compiled, you also need to install a compiler.

1. Download a recent version (at least 3.7) of Python for Windows from <https://www.python.org/> and install it.
2. Download and install "Build Tools for Visual Studio 2019" from <https://visualstudio.microsoft.com/downloads/#build-tools-for-visual-studio-2019>. (There are many similarly named downloads on that page, ensure you get the right one.)

   During installation, when the dialog about which components to install pops up, ensure that "C++ Build tools" is ticked. The download is quite big and can take a long time.
3. Open the command line (`cmd.exe`) and run `py -m pip install cutadapt`.
4. Test whether it worked by running `py -m cutadapt --version`. You should see the version number of Cutadapt.

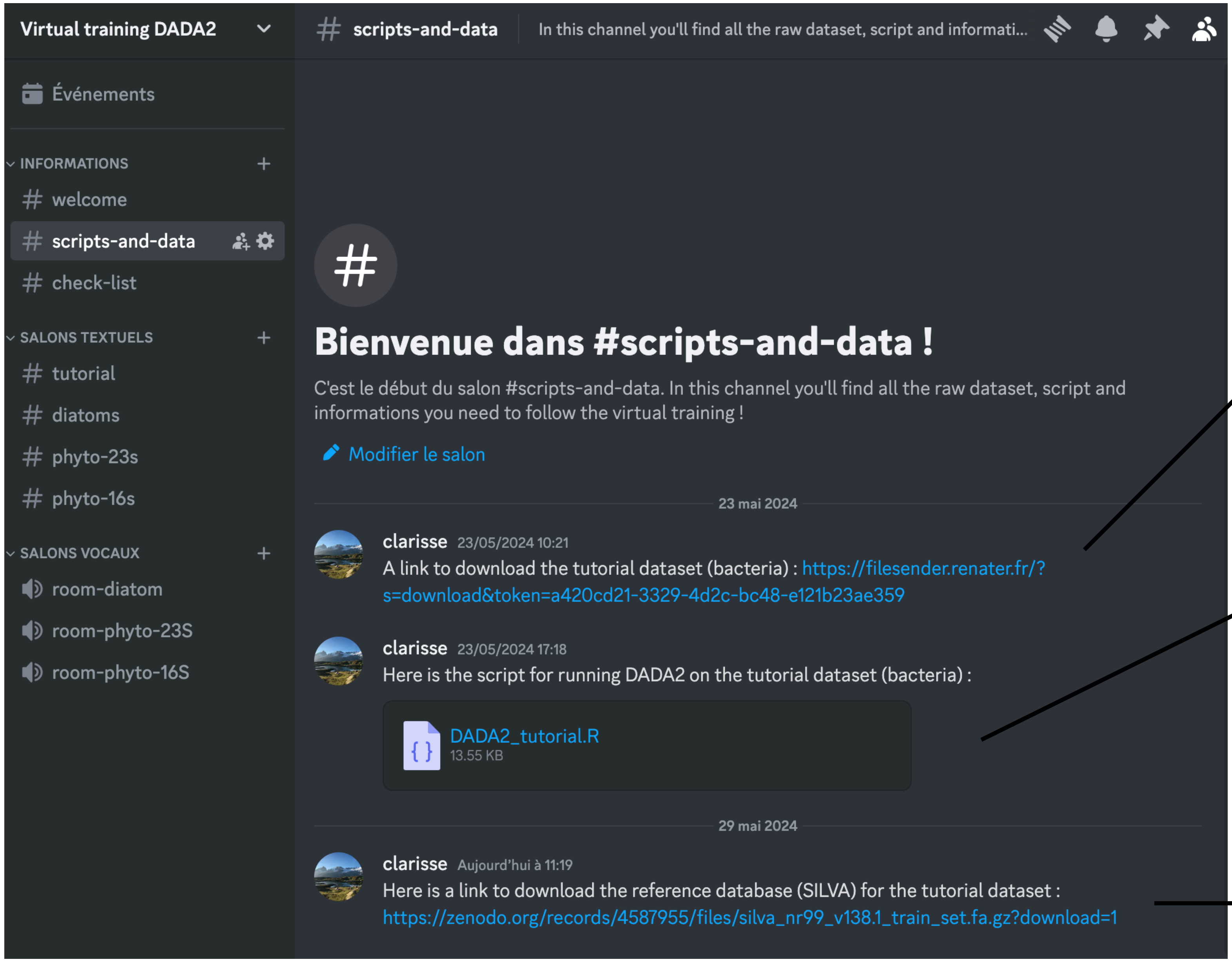
When running Cutadapt this way, you will need to remember to write `py -m cutadapt` instead of just `cutadapt`.

PREPARATION OF THE PRACTICAL PART

Does everyone has discord?

PREPARATION OF THE PRACTICAL PART

FOR MONDAY : download tutorial dataset



Bienvenue dans #scripts-and-data !

C'est le début du salon #scripts-and-data. In this channel you'll find all the raw dataset, script and informations you need to follow the virtual training !

23 mai 2024

clarisse 23/05/2024 10:21
A link to download the tutorial dataset (bacteria) : https://filesender.renater.fr/?s=download&token=a420cd21-3329-4d2c-bc48-e121b23ae359

clarisse 23/05/2024 17:18
Here is the script for running DADA2 on the tutorial dataset (bacteria) :

DADA2_tutorial.R
13.55 KB

29 mai 2024

clarisse Aujourd'hui à 11:19
Here is a link to download the reference database (SILVA) for the tutorial dataset :
https://zenodo.org/records/4587955/files/silva_nr99_v138.1_train_set.fa.gz?download=1

On your desktop, on a folder « **virtual_training** »

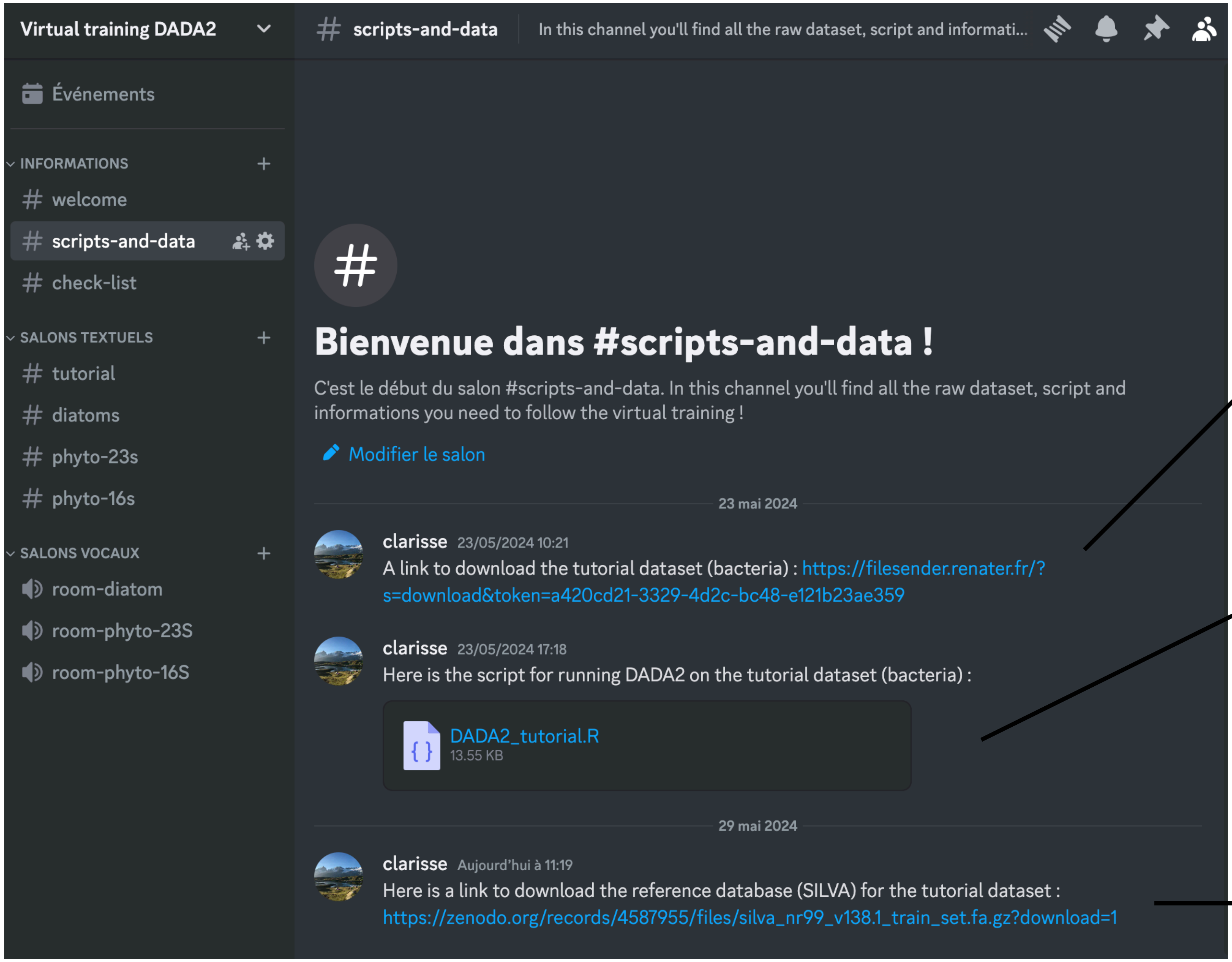
Put fastq files in a subfolder named « **bacteria** »

Put the script DADA2_tutorial.R in a subfolder named « **scripts** »

Put the script DADA2_tutorial.R in a subfolder named « **database** »

PREPARATION OF THE PRACTICAL PART

FOR MONDAY : download tutorial dataset



Bienvenue dans #scripts-and-data !

C'est le début du salon #scripts-and-data. In this channel you'll find all the raw dataset, script and informations you need to follow the virtual training !

23 mai 2024

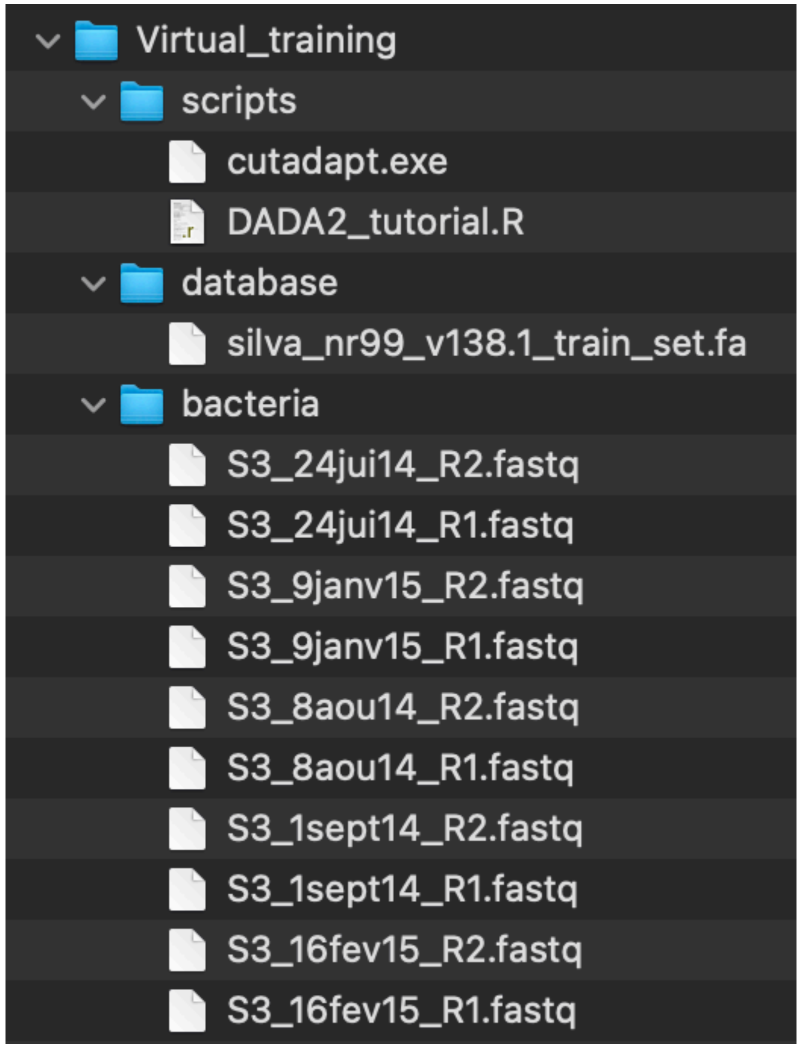
clarisse 23/05/2024 10:21
A link to download the tutorial dataset (bacteria) : https://filesender.renater.fr/?s=download&token=a420cd21-3329-4d2c-bc48-e121b23ae359

clarisse 23/05/2024 17:18
Here is the script for running DADA2 on the tutorial dataset (bacteria) :

DADA2_tutorial.R
13.55 KB

29 mai 2024

clarisse Aujourd'hui à 11:19
Here is a link to download the reference database (SILVA) for the tutorial dataset :
https://zenodo.org/records/4587955/files/silva_nr99_v138.1_train_set.fa.gz?download=1



```
Virtual_training
  scripts
    cutadapt.exe
    DADA2_tutorial.R
  database
    silva_nr99_v138.1_train_set.fa
  bacteria
    S3_24jui14_R2.fastq
    S3_24jui14_R1.fastq
    S3_9janv15_R2.fastq
    S3_9janv15_R1.fastq
    S3_8aou14_R2.fastq
    S3_8aou14_R1.fastq
    S3_1sept14_R2.fastq
    S3_1sept14_R1.fastq
    S3_16fev15_R2.fastq
    S3_16fev15_R1.fastq
```

ANY QUESTIONS ?