



Initiation aux statistiques avec RStudio

Travaux pratiques

ALEXIS.RUFFAULT@INSEP.FR

Alexis RUFFAULT, Chercheur en psychologie du sport et de la santé
Laboratoire Sport, Expertise et Performance (EA 7370)



1. S'initier à RMarkdown

2. Manier les données dans RStudio

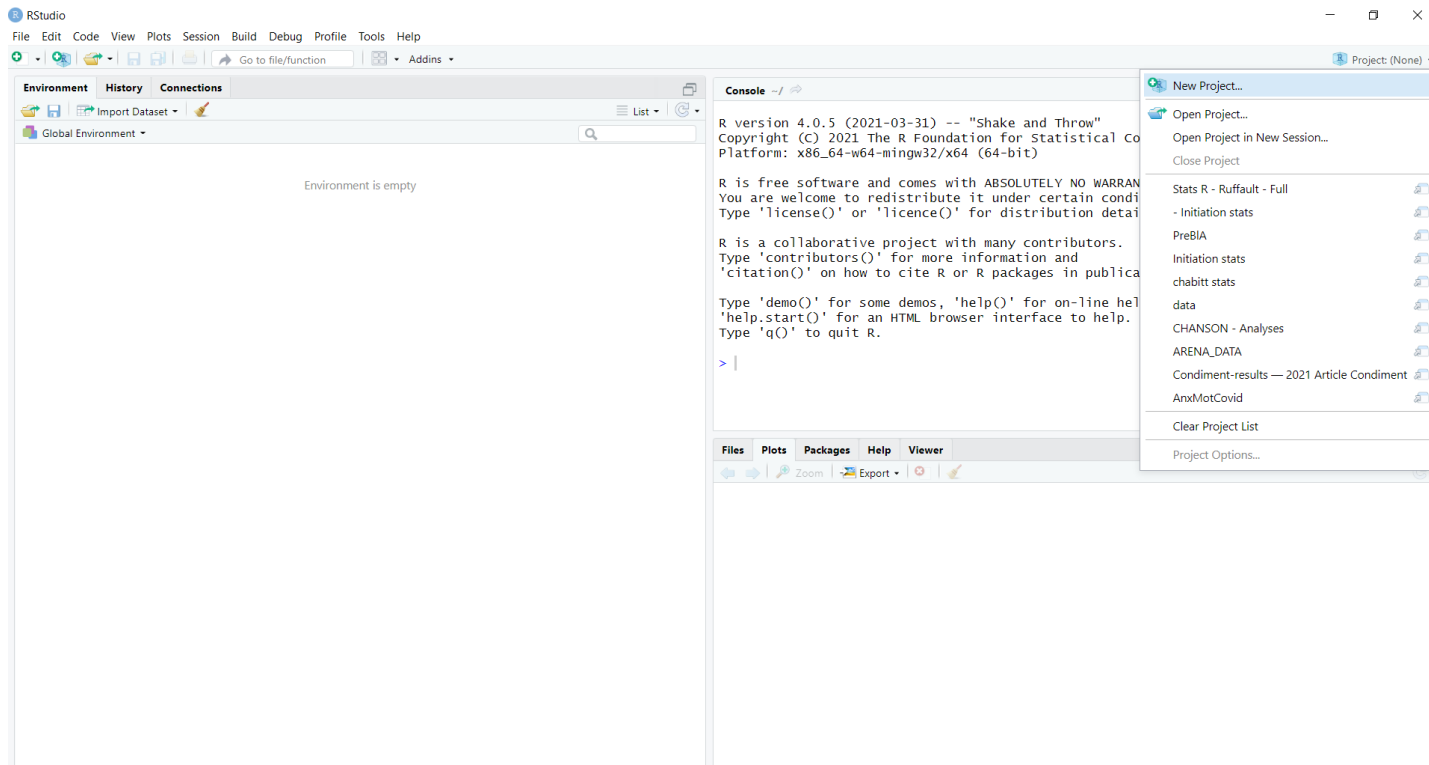
3. Savoir quand appliquer quel test statistique

1. Comparer deux groupes indépendants sur une variable numérique
2. Comparer deux groupes appariés sur une variable numérique
3. Voir l'association entre deux variables numériques
4. Associations entre les niveaux de 2 facteurs
5. Régression linéaire simple, multiples, hiérarchiques
6. Régression logistique
7. ANOVA
8. Modèles linéaires mixtes



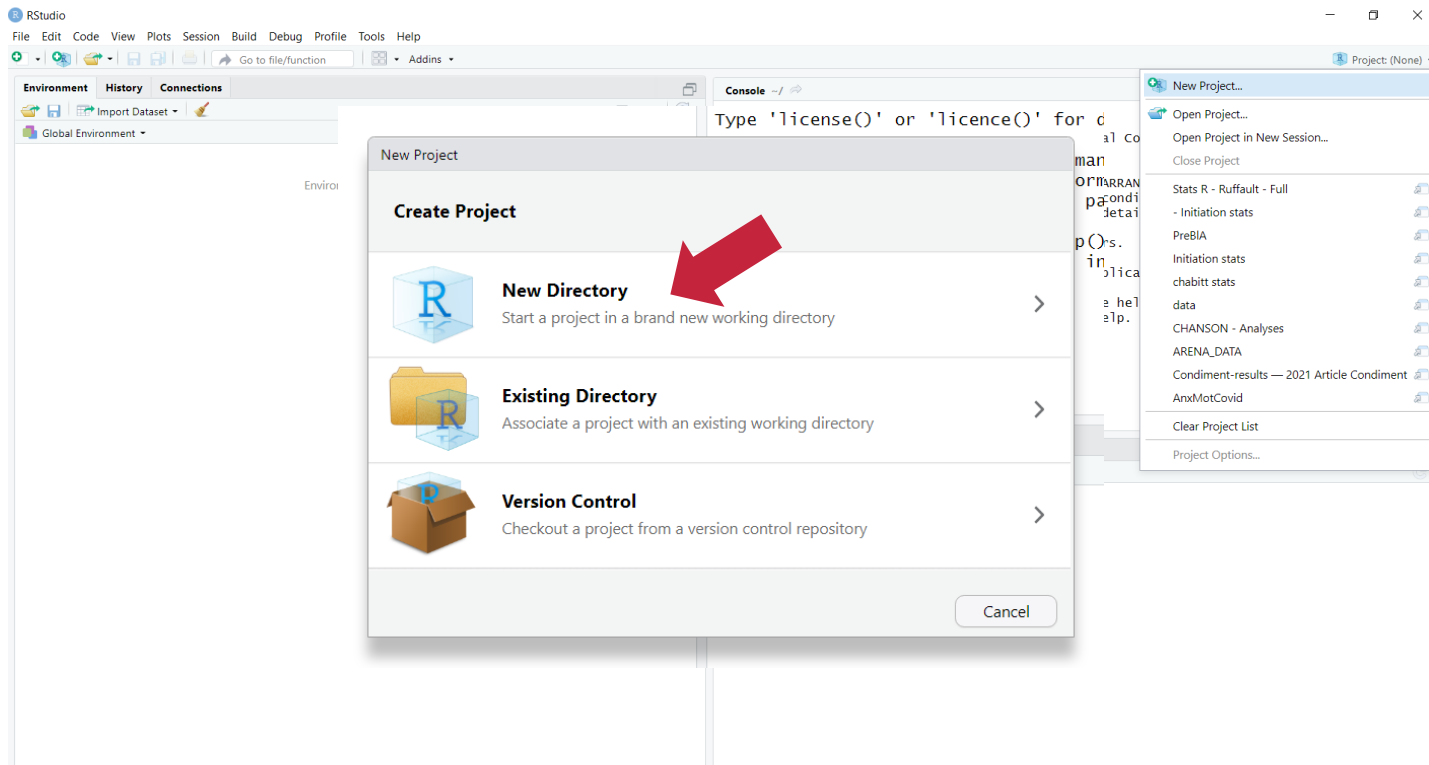
POUR COMMENCER

CREER SON PROJET RStudio



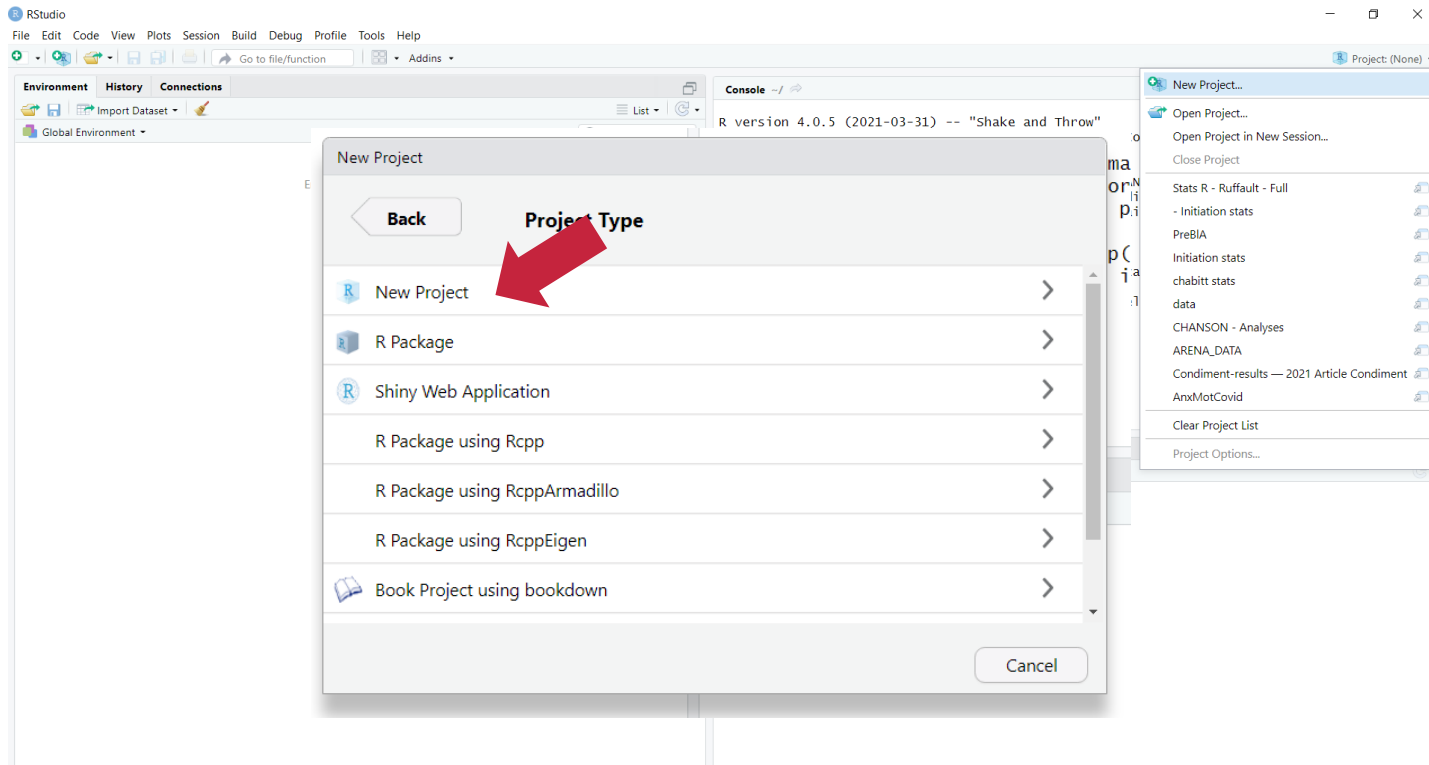
POUR COMMENCER

CREER SON PROJET RStudio



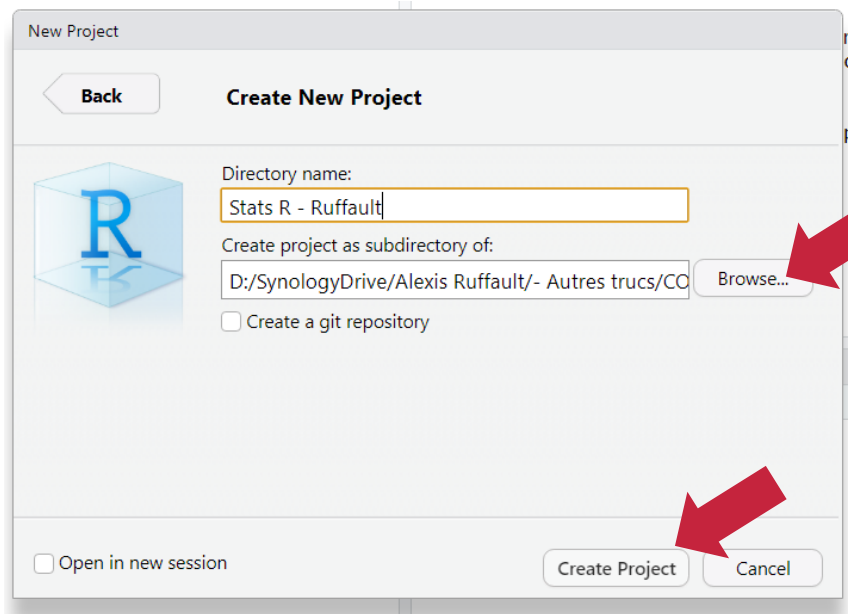
POUR COMMENCER

CREER SON PROJET RStudio



POUR COMMENCER

CREER SON PROJET RStudio



Nommer le projet « Stats R – Ruffault »

Rangez-le dans un dossier de votre ordinateur (par exemple, là où vous rangez vos cours de stats)

Cliquez sur « Create Project »

Attendez que RStudio crée votre projet



POUR COMMENCER

PREMIERS PAS DANS RStudio

The screenshot shows the RStudio interface with the following components:

- Top Menu Bar:** File, Edit, Code, View, Plots, Session, Build, Debug, Profile, Tools, Help.
- Top Bar:** D:/SynologyDrive/Alexis Ruffault/- Autres trucs/COURS/UCO-Angers/Stats R - Ruffault - RStudio
- Left Panel:** Environment, History, Connections. The Environment pane shows "Global Environment" and "Environment is empty".
- Right Panel:** Console, Files, Plots, Packages, Help, Viewer. The Console shows the R startup message. The Files pane shows the directory structure: D:/SynologyDrive/Alexis Ruffault/- Autres trucs/COURS/UCO-Angers/Stats R - Ruffault, containing a file named "Stats R - Ruffault.Rproj".

Annotations:

- A red arrow points to the "Stats R - Ruffault" tab in the top right corner.
- A red arrow points to the console output area.
- A yellow box contains the text: "Copier-coller les fichiers de données et .Rmd dans votre dossier « Stats R – Ruffault »".
- A yellow arrow points from the yellow box to the Files pane.



POUR COMMENCER

PREMIERS PAS DANS RStudio

The screenshot shows the RStudio interface with the following components:

- Environment:** Shows "Global Environment" and "Environment is empty".
- Console:** Displays the R welcome message and a prompt "> |".
- Files:** Shows a file explorer view of the directory "D:\SynologyDrive\Alexis Ruffault\ - Autres trucs\COURS\UCO-Angers\Stats R - Ruffault\". The files listed are:

Name	Size	Modified
..		
ChaBITT-BDD_determinants-Basket.csv	4.6 KB	Jul 21, 2021, 3:48 PM
evirap-dataset_HEC1.rda	2.5 KB	Jun 4, 2018, 5:05 PM
PAPPOdata.xlsx	31.3 KB	Jul 21, 2021, 9:00 AM
prebla-dataset.rda	445.7 KB	Jul 22, 2021, 4:23 PM
Stats R - Ruffault.Rproj	218 B	Aug 2, 2021, 10:31 AM
TD1 - Les variables catégorielles et numériques.Rmd	10.2 KB	Aug 2, 2021, 10:00 AM
TD2 - Les premiers tests statistiques.Rmd	11.5 KB	Aug 2, 2021, 9:55 AM
TD3 - Les régressions.Rmd	21.4 KB	Aug 2, 2021, 9:55 AM
TD4 - Les modèles mixtes.Rmd	8.8 KB	Aug 2, 2021, 9:56 AM

A yellow arrow points to the file "TD1 - Les variables catégorielles et numériques.Rmd".

Cliquez sur le TD1



POUR COMMENCER

PREMIERS PAS DANS RStudio

D:/SynologyDrive/Alexis Ruffault/- Autres trucs/COURS/UCO-Angers/Stats R - Ruffault - RStudio

File Edit Code View Plots Session Build Debug Profile Tools Help

Go to file/function Addins

TD1 - Les variables catégorielles et nu...

```
1 |---
2 title: "TD1 - Les variables catégorielles et numériques"
3 author: "Alexis Ruffault"
4 date: "21/07/2021"
5 output:
6   html_document:
7     number_sections: yes
8     toc: yes
9     toc_depth: 2
10    toc_float: true
11  editor_options:
12    chunk_output_type: console
13  ---
14
15  ```{r setup, include=FALSE}
16
17  #packages
18  library(knitr)
19
20  ## Global options
21  options(max.print="75")
22  knitr::opts_chunk$set(
23    echo = TRUE,
24  )
25  ---
26  1:1 TD1 - Les variables catégorielles et numériques R Markdown
```

Environment History Connections

Global Environment

Environment is empty

Console D:/SynologyDrive/Alexis Ruffault/- Autres trucs/COURS/UCO-Angers/Stats R - Ruffault/

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.

with many contributors. information and for R packages in publications. 'help()' for on-line help, or user interface to help.

More

Autres trucs > COURS > UCO-Angers > Stats R - Ruffault

Name	Size	Modified
ChaBITT-BDD_determinants-Basket.csv	4.6 KB	Jul 21, 2021, 3:48 PM
evirap-dataset_HEC1.rda	2.5 KB	Jun 4, 2018, 5:05 PM
PAPPOdata.xlsx	31.3 KB	Jul 21, 2021, 9:00 AM
prebla-dataset.rda	445.7 KB	Jul 22, 2021, 4:23 PM
Stats R - Ruffault.Rproj	218 B	Aug 2, 2021, 10:31 AM
TD1 - Les variables catégorielles et numériques.Rmd	10.2 KB	Aug 2, 2021, 10:00 AM
TD2 - Les premiers tests statistiques.Rmd	11.5 KB	Aug 2, 2021, 9:55 AM
TD3 - Les régressions.Rmd	21.4 KB	Aug 2, 2021, 9:55 AM
TD4 - Les modèles mixtes.Rmd	8.8 KB	Aug 2, 2021, 9:56 AM

Ici, ce sont les options du fichier Rmarkdown, qui permettent de créer un fichier HTML du cours et de vos manipulations de données



POUR COMMENCER

PREMIERS PAS DANS RStudio

D:/SynologyDrive/Alexis Ruffault/- Autres trucs/COURS/UCO-Angers/Stats R - Ruffault - RStudio

File Edit Code View Plots Session Build Debug Profile Tools Help

Go to file/function Addins

TD1 - Les variables catégorielles et nu...

```
1 |<---
2 title: "TD1 - Les variables catégorielles et numériques"
3 author: "Alexis Ruffault"
4 date: "21/07/2021"
5 output:
6   html_document:
7     number_sections: yes
8     toc: yes
9     toc_depth: 2
10    toc_float: true
11  editor_options:
12    chunk_output_type: console
13  ---
14
15  {r setup, include=FALSE}
16
17  #packages
18  library(knitr)
19
20  ## Global options
21  options(max.print="75")
22  knitr::opts_chunk$set(
23    echo = TRUE,
24  )
25  |>
```

Environment History Connections

Global Environment

Environment is empty

Console D:/SynologyDrive/Alexis Ruffault/- Autres trucs/COURS/UCO-Angers/Stats R - Ruffault/

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.

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.

> |

Files Plots Packages Help Viewer

New Folder Delete Rename More

	Size	Modified
4.6 KB	Jul 21, 2021, 3:48 PM	
2.5 KB	Jun 4, 2018, 5:05 PM	
31.3 KB	Jul 21, 2021, 9:00 AM	
445.7 KB	Jul 22, 2021, 4:23 PM	
218 B	Aug 2, 2021, 10:31 AM	
10.2 KB	Aug 2, 2021, 10:00 AM	
11.5 KB	Aug 2, 2021, 9:55 AM	
21.4 KB	Aug 2, 2021, 9:55 AM	
8.8 KB	Aug 2, 2021, 9:56 AM	

TD3 - Les regressions.Rmd

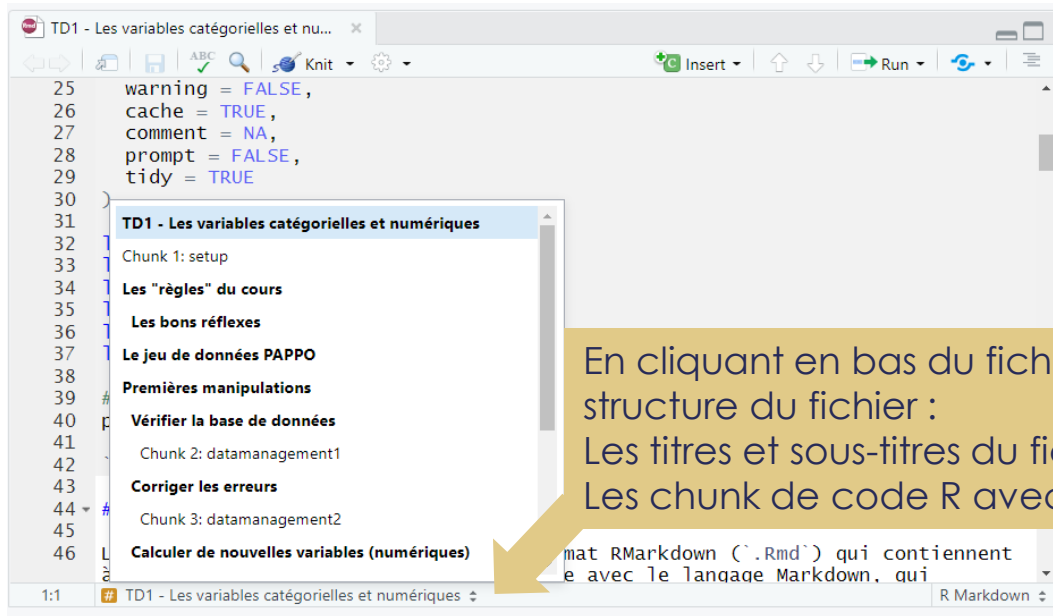
TD4 - Les modèles mixtes.Rmd

Ici, c'est un premier « chunk » de script R, qu'on nomme « setup », et qui permet de paramétrer notre session R qui sera incluse dans le fichier HTML



POUR COMMENCER

PREMIERS PAS DANS RStudio



En cliquant en bas du fichier .Rmd, vous voyez la structure du fichier :

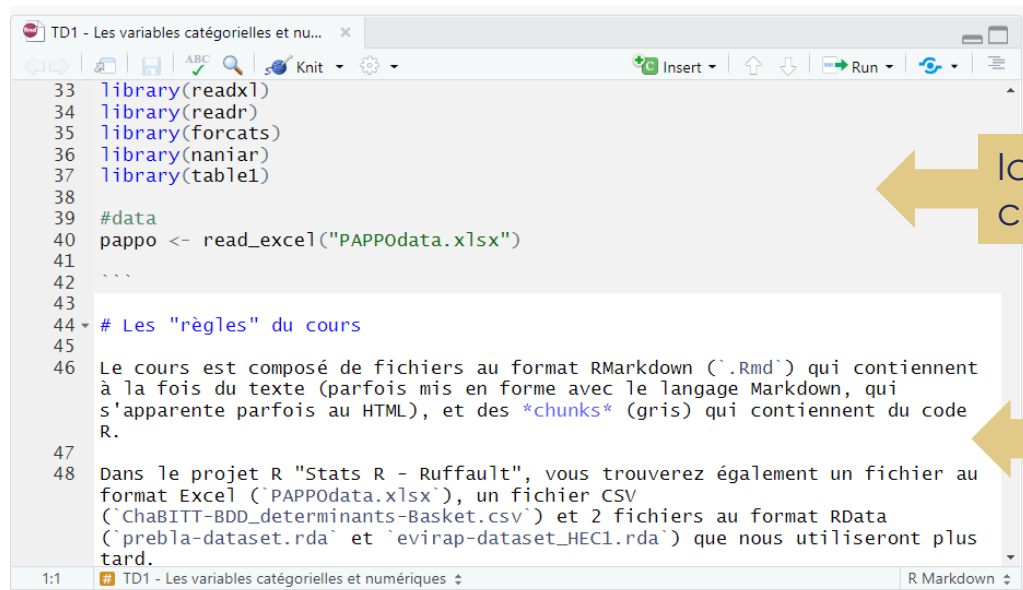
Les titres et sous-titres du fichier HTML en gras

Les chunk de code R avec leur titre en « normal »

mat RMarkdown (".Rmd") qui contiennent
e avec le langage Markdown, qui

POUR COMMENCER

PREMIERS PAS DANS RStudio



```
33 library(readxl)
34 library(readr)
35 library(forcats)
36 library(nanian)
37 library(table1)
38
39 #data
40 pappo <- read_excel("PAPP0data.xlsx")
41
42
43
44 # Les "règles" du cours
45
46 Le cours est composé de fichiers au format RMarkdown (`.Rmd`) qui contiennent
47 à la fois du texte (parfois mis en forme avec le langage Markdown, qui
48 s'apparente parfois au HTML), et des *chunks* (gris) qui contiennent du code
49 R.
50
51 Dans le projet R "Stats R - Ruffault", vous trouverez également un fichier au
52 format Excel (`.xlsx`), un fichier CSV
53 (`.csv`), un fichier RData (`.rda`) et 2 fichiers au format RData
54 (`.rda`) que nous utiliseront plus
55 tard.
```

Ici c'est la suite du premier chunk R « setup »

Ici c'est du texte en format MarkDown (un « langage » qui ressemble à LaTeX et HTML)

A FAIRE EN PREMIER

LANCER LE CODE R
CREER LE FICHIER HTML

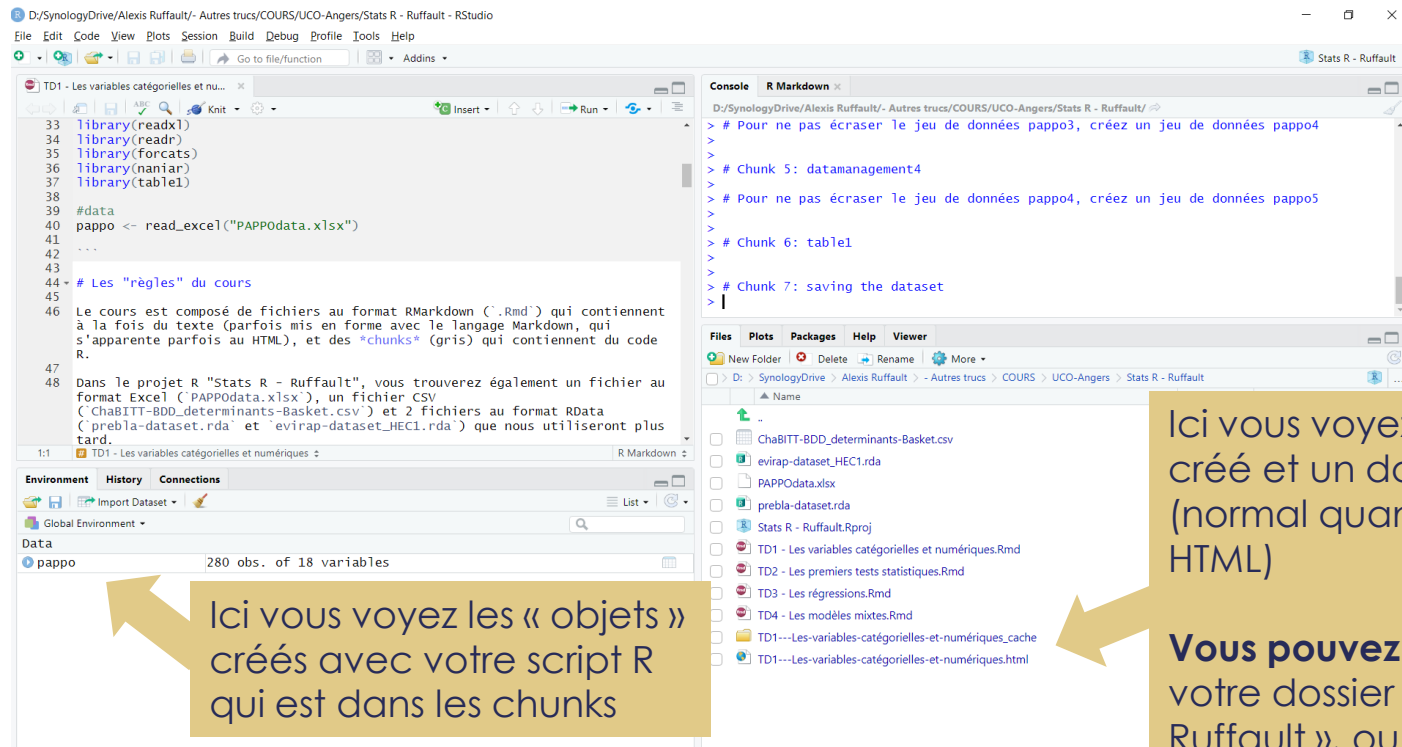


The screenshot shows the RStudio interface with a script editor containing R code. Two yellow callout boxes provide instructions:

- A box on the right points to the 'Run' button in the toolbar: **En premier, cliquez sur « Run »** puis « Run all »
- A box in the center points to the 'Knit' button in the toolbar: **Ensuite, cliquez sur la flèche à côté de « Knit »** puis « Knit to HTML »

The R code in the script editor is as follows:

```
33 library(readxl)
34 library(readr)
35 library(forcats)
36 library(naniar)
37 library(table1)
38
39 #data
40 pappo <- read_excel("PAPPodat
41 ...
42
43
44 # Les "règles" du cours
45
46 Le cours est composé de fichiers au format RMarkdown (`.Rmd`) qui contiennent
47 à la fois du texte (parfois mis en forme avec le langage Markdown, qui
48 s'apparente parfois au HTML), et des *chunks* (gris) qui contiennent du code
49 R.
50
51 Dans le projet R "Stats R - Ruffault", vous trouverez également un fichier au
52 format Excel (`.xlsx`), un fichier CSV
53 (`.ChabITT-BDD_determinants-Basket.csv`) et 2 fichiers au format RData
54 (`.prebla-dataset.rda` et `.evirap-dataset_HEC1.rda`) que nous utiliseront plus
55 tard.
```



The screenshot shows the RStudio interface with the following components:

- Source Editor:** Contains R code for loading libraries (readxl, readr, forcats, nantar, table1), reading an Excel file (PAPPOdata.xlsx), and a comment about the course rules.
- Environment:** Shows the 'pappo' object with 280 observations and 18 variables.
- Console:** Displays the output of the R code, including messages about creating datasets and saving the dataset.
- Files:** Lists the files in the project, including CSV files, R datasets, and R Markdown files.

Annotations on the screenshot:

- Left Arrow:** Points to the Environment pane with the text: "Ici vous voyez les « objets » créés avec votre script R qui est dans les chunks".
- Right Arrow:** Points to the Files pane with the text: "Ici vous voyez le fichier HTML créé et un dossier « cache » (normal quand on crée un HTML)".
- Bottom Right:** A large text box states: "Vous pouvez l'ouvrir depuis votre dossier « Stats R – Ruffault », ou directement ici".



TADA !

ON PEUT COMMENCER A TRAVAILLER

1 Les "règles" du cours

- 1.1 Les bons réflexes
- 2 Le jeu de données PAPPO
- 3 Premières manipulations
- 4 Créer un premier tableau pour voir les tendances
- 5 Et pour terminer...

Ici une table des matières flottante et dynamique

TD1 - Les variables catégorielles

Alexis Ruffault

21/07/2021

1 Les "règles" du cours

Le cours est composé de fichiers au format RMarkdown (`.Rmd`) qui contiennent à la fois du texte (parfois mis en forme avec le langage Markdown, qui s'apparente parfois au HTML), et des *chunks* (gris) qui contiennent du code R.

Dans le projet R "Stats R - Ruffault", vous trouverez également un fichier au format Excel (`PAPPOdata.xlsx`), un fichier CSV (`ChaBITT-BDD_determinants-Basket.csv`) et 2 fichiers au format RData (`prebla-dataset.rda` et `evirap-dataset_HEC1.rda`) que nous utiliserons plus tard.

Dans ce cours, vous apprendrez d'abord à importer un jeu de données et à faire quelques manipulations dessus, pour ensuite les exporter au format `.rda`.

Ensuite, vous apprendrez à mener différentes analyses statistiques, à appliquer en fonction de la question à laquelle vous voulez répondre, et en fonction de la nature de vos données.

Dans tous les cas, faites les exercices les uns après les autres, dans l'ordre, pour ne rien louper.

1.1 Les bons réflexes

Il faut toujours que vous travailliez dans le projet "Stats R - Ruffault" (vous pouvez vérifier que vous êtes bien dedans en regardant tout en haut à droite de votre fenêtre RStudio).

Il ne faut surtout pas que le dossier du projet R sur lequel vous travaillez soit dans le bureau de votre ordinateur : mettez-le dans un dossier de vos documents.

Si vous avez des erreurs

- relisez attentivement votre code
- vérifiez que les packages sont bien installés (`install.packages`) et appelés (`library`)
- lisez l'aide du package ou de la fonction que vous utilisez
- demandez à google

Ici le texte du fichier RMarkdown qui n'est pas dans les chunks



TADA !

ON PEUT COMMENCER A TRAVAILLER

Faites les exercices

1 Les "règles" du cours

3.1 Vérifier la base de données

3.2 Corriger les erreurs

3.3 Calculer de nouvelles variables
(numériques)

3.4 Créer des facteurs (variables
catégorielles)

4 Créer un premier tableau pour voir

Ca, c'est un chunk R, et vous verrez dans votre fichier .Rmd (RStudio) que le code que vous écrivez se retrouvera ici

3.1 Vérifier la base de données

Exercice : Vérifiez la structure de chaque variable. Assurez-vous que les variables numériques soient bien numériques, et que les variables catégorielles soient bien des facteurs. Faites les modifications nécessaires si vous remarquez quelque chose.

TIPS

- la fonction `summary` peut être utile pour vérifier que les variables catégorielles sont bien des facteurs (`SEX` et `SPO`)
- la fonction `class` également
- pour faire des modifications dans le jeu de données `pappo`, la fonction `mutate` (et `mutate_if`) du `tidyverse` peut être utile
- le package `forcats` est également utilisé pour travailler sur les paramètres des facteurs

```
# Pour ne pas écraser le jeu de données pappo, créez un jeu de données pappo2
```

Vous trouverez quelques conseils pour vous aider (il faut chercher vous-même)

3.2 Corriger les erreurs

Exercice : La variable `IPAQ_S` devrait être une variables numérique, et les valeurs `NSP` devraient être considérées manquantes. Faites les modifications nécessaires.

TIPS

- c'est pas super de modifier le fichier Excel, même si c'est très simple
- le package `nanian` peut être très utile
- pensez à faire en sorte que la variable `IPAQ_S` soit bien numérique

```
# Pour ne pas écraser le jeu de données pappo2, créez un jeu de données pappo3
```




A VOUS DE JOUER !

LE TD COMMENCE MAINTENANT

1. Retournez dans RStudio
2. Lisez les consignes du fichier HTML du TD1
3. Faites les exercices dans l'ordre

Si vous avez des questions, levez la main et prenez la parole
Si vous avez des erreurs, appelez-moi

Si vous avez terminé le premier exercice, levez la main
Nous ferons la correction ensemble

Attendez que tout le monde soit au même niveau pour passer à la suite

Initiation aux statistiques avec RStudio

ALEXIS.RUFFAULT@INSEP.FR



LABORATOIRE **SPORT, EXPERTISE ET PERFORMANCE** (EA 7370)

**INSTITUT NATIONAL DU SPORT,
DE L'EXPERTISE ET DE LA PERFORMANCE**

11, avenue du Tremblay - 75012 PARIS

Tél. : 01 41 74 41 00

www.insep.fr

