

TP1

Initiation à la Programmation en R

Prise en main de R et RStudio

Écrire des scripts simples • Manipuler des données • Visualisations de base

Table des matières

Table des matières.....	2
Objectifs.....	3
Contenu du TP	3
1. Introduction à RStudio.....	3
1.1 Découverte de l'interface	3
1.2 Configuration de l'espace de travail.....	3
1.3 Installation des packages.....	3
2. Écriture de Scripts R Simples	4
2.1 Variables et Types de Données.....	4
2.2 Opérations de Base	4
2.3 Création et Manipulation de Vecteurs	4
3. Importation de Fichiers CSV	5
3.1 Lecture des données.....	5
3.2 Exploration Rapide	5
3.3 Détection des Valeurs Manquantes.....	5
4. Manipulation Simple des Données	6
4.1 Sélection de Colonnes et Lignes	6
4.2 Filtrage de Données	6
4.3 Calculs sur Colonnes (Statistiques Descriptives)	6
4.4 Création de Nouvelles Variables	6
5. Création de Graphiques de Distribution	7
5.1 Histogrammes (Distribution)	7
5.2 Boxplots (Comparaison).....	7
5.3 Identification Visuelle du Type de Distribution	8
6. Synthèse du TP1 - Commandes Essentielles	8
Annexe : Description des Variables du Dataset.....	8

Note : Clic droit sur la table → "Mettre à jour le champ" pour actualiser les numéros de page.

Objectifs

Prendre en main l'environnement R et RStudio, écrire des scripts simples, manipuler des données et produire des visualisations de base.

Contenu du TP

- Introduction à RStudio : Interface, Configuration, Installation de packages
- Écriture de scripts R simples : Variables, Types, Vecteurs
- Importation de fichiers CSV : Lecture, Exploration, Valeurs manquantes
- Manipulation simple des données : Sélection, Filtrage, Calculs
- Création de graphiques de distribution : Histogrammes, Boxplots

1. Introduction à RStudio

1.1 Découverte de l'interface

Une fois RStudio ouvert, vous découvrez une interface moderne et organisée, conçue pour optimiser votre workflow d'analyse de données. L'interface RStudio est divisée en quatre panneaux principaux que nous allons explorer en détail.

L'interface RStudio se compose de quatre zones principales disposées de manière logique :

1. Panneau Source/Éditeur (en haut à gauche) : Zone de rédaction des scripts R
2. Console (en bas à gauche) : Exécution interactive des commandes
3. Environnement/Historique (en haut à droite) : Variables et objets en mémoire
4. Fichiers/Graphiques/Packages/Aide (en bas à droite) : Navigation et visualisation

1.2 Configuration de l'espace de travail

```
# Définir le répertoire de travail (où est votre fichier CSV)
setwd("C:/Users/VotreNom/Documents/TP_Ecologie")

# Vérifier le répertoire actuel
getwd()

# Lister les fichiers du dossier
list.files()
```

1.3 Installation des packages

```
# À faire une seule fois
install.packages("tidyverse")
install.packages("readr")

# À faire à chaque ouverture de R
library(tidyverse)
```

```
library(readr)
```

Exercice 1 : Créer un nouveau script TP1_chene.R et sauvegarder-le dans votre dossier de travail.

2. Écriture de Scripts R Simples

2.1 Variables et Types de Données

```
# NUMERIC : Nombres (altitudes, températures, mesures)
altitude_zouagha <- 450
temperature <- 17.8
ph_sol <- 7.2
class(altitude_zouagha) # "numeric"

# CHARACTER : Texte (noms des forêts, IDs)
foret <- "Zouagha"
arbre_id <- "CL01"
class(foret) # "character"

# LOGICAL : Vrai/Faux (présence de dégâts)
degat_feu <- TRUE # ou FALSE
class(degat_feu) # "logical"
```

2.2 Opérations de Base

```
# Calculs sur les mesures de liège
epaisseur_2022 <- 18
epaisseur_2024 <- 24
accroissement <- epaisseur_2024 - epaisseur_2022 # 6 mm

# Calcul du volume approximatif (cylindre)
diametre <- 42 # cm
rayon <- diametre / 2
surface <- pi * rayon^2 # pi = 3.14159...
hauteur <- 8.5
volume <- surface * hauteur / 10000 # Conversion en m³
```

2.3 Création et Manipulation de Vecteurs

```
# Vecteur des épaisseurs de liège (Zouagha)
epaisseur_zouagha <- c(18, 20, 15, 22, 19, 21, 17, 23, 16, 24, 20, 18, 25,
14, 21, 22, 19, 17, 23, 20)

# Vecteur des températures
temperatures <- c(17.8, 17.6, 18.0, 17.9, 17.7, 17.8, 18.1, 17.5, 17.9,
17.6,
18.0, 17.8, 17.7, 18.1, 17.9, 17.5, 17.8, 18.0, 17.6,
17.9)

# Séquences régulières
altitudes <- seq(450, 920, by = 10) # 450, 460, 470...
```

```
# Répétitions
forets <- rep("Zouagha", 20) # Répète "Zouagha" 20 fois

# Calculs sur vecteurs
moyenne_epaisseur <- mean(epaisseur_zouagha)
max_epaisseur <- max(epaisseur_zouagha)
min_epaisseur <- min(epaisseur_zouagha)
```

Exercice 2 : Créer un vecteur contenant les 10 premières valeurs de production_glands et calculer la moyenne.

3. Importation de Fichiers CSV

3.1 Lecture des données

```
# Méthode 1 (Base R)
chene <- read.csv2("Mila.csv", header = TRUE, sep = ";")

# Méthode 2 (Package readr - plus rapide et robuste)
library(readr)
chene <- read_csv2("Mila.csv")

# Vérification du chargement
dim(chene) # Doit afficher : 60 13 (60 lignes, 13 colonnes)
```

3.2 Exploration Rapide

```
# Visualiser les premières lignes
head(chene) # Affiche les 6 premières lignes
head(chene, 10) # Affiche les 10 premières

# Visualiser les dernières lignes
tail(chene)

# Structure des données (types de variables)
str(chene)

# Résumé statistique rapide
summary(chene)
```

3.3 Détection des Valeurs Manquantes

```
# Compter les NA (Not Available) par colonne
colSums(is.na(chene))

# Vérifier s'il y a des manquants dans une colonne spécifique
sum(is.na(chene$epaisseur_liege_2024)) # Devrait être 0
sum(is.na(chene$degats_feu)) # Devrait être 0

# Afficher les lignes avec des valeurs manquantes (s'il y en a)
chene[!complete.cases(chene), ]
```

4. Manipulation Simple des Données

4.1 Sélection de Colonnes et Lignes

```
# Sélection d'une colonne avec $
liste_altitudes <- chene$altitude
liste_ph <- chene$ph_sol

# Sélection par position [ligne, colonne]
chene[1, ] # Première ligne (toutes les colonnes)
chene[, 2] # Deuxième colonne (toutes les lignes)
chene[1:5, c("arbre_id", "foret", "diametre_chene")] # 5 premiers arbres, 3 colonnes

# Sélection par condition
arbres_zouagha <- chene[chene$foret == "Zouagha", ] # 20 arbres
arbres_tassadane <- chene[chene$foret == "Tassadane", ]
arbres_chelghoum <- chene[chene$foret == "Chelghoum_Laid", ]
```

4.2 Filtrage de Données

```
# Arbres avec dégâts de feu
feu <- chene[chene$degats_feu == TRUE, ]
nrow(feau) # Nombre d'arbres touchés : 23

# Arbres sans dégâts de feu
sans_feu <- chene[chene$degats_feu == FALSE, ]

# Arbres de Tassadane avec pH < 6.5
tassadane_acide <- chene[chene$foret == "Tassadane" & chene$ph_sol < 6.5, ]

# Arbres avec gros diamètre (> 50 cm)
gros_chenes <- chene[chene$diametre_chene > 50, ]
```

4.3 Calculs sur Colonnes (Statistiques Descriptives)

```
# Calculs simples
moyenne_diametre <- mean(chene$diametre_chene)
ecart_type_diametre <- sd(chene$diametre_chene)
mediane_hauteur <- median(chene$hauteur_totale)

# Calculs par groupe (fonction tapply)
# Tapply : "table apply" - applique une fonction par groupe
moyenne_epaisseur_2024 <- tapply(chene$epaisseur_liege_2024, chene$foret,
mean)
# Résultat :
#   Chelghoum Laid      Tassadane      Zouagha
#         31.60         40.20         25.95

# Statistiques par forêt
stats_ph <- tapply(chene$ph_sol, chene$foret, summary)
```

4.4 Création de Nouvelles Variables

```
# Calculer l'épaisseur moyenne des deux années
chene$epaisseur_moyenne <- (chene$epaisseur_liege_2022 +
chene$epaisseur_liege_2024) / 2

# Créer une variable catégorielle (gros vs petit diamètre)
chene$categorie <- ifelse(chene$diametre_chene > 50, "Gros", "Petit")
table(chene$categorie) # Compte : Gros vs Petit

# Calculer le ratio production/diamètre
chene$ratio_production <- chene$production_glands / chene$diametre_chene
```

Exercice 4 : Calculer la production moyenne de glands pour chaque forêt. Quelle forêt produit le plus de glands en moyenne ?

5. Création de Graphiques de Distribution

5.1 Histogrammes (Distribution)

```
# Distribution de l'épaisseur du liège 2024
hist(chene$epaisseur_liege_2024,
     main = "Distribution de l'épaisseur du liège (2024)",
     xlab = "Épaisseur (mm)",
     ylab = "Nombre d'arbres",
     col = "lightblue",
     breaks = 10)

# Superposition des 3 forêts (3 histogrammes côte à côte)
par(mfrow = c(1, 3)) # 1 ligne, 3 colonnes
hist(chene$epaisseur_liege_2024[chene$foret == "Zouagha"],
     main = "Zouagha (450m)", xlab = "mm", col = "green")
hist(chene$epaisseur_liege_2024[chene$foret == "Tassadane"],
     main = "Tassadane (680m)", xlab = "mm", col = "blue")
hist(chene$epaisseur_liege_2024[chene$foret == "Chelghoum_Laid"],
     main = "Chelghoum Laid (920m)", xlab = "mm", col = "orange")
par(mfrow = c(1, 1)) # Remettre à 1 graphique par fenêtre
```

5.2 Boxplots (Comparaison)

```
# Comparaison des diamètres entre forêts
boxplot(diametre_chene ~ foret,
       data = chene,
       main = "Diamètre des chênes par forêt",
       xlab = "Forêt",
       ylab = "Diamètre (cm)",
       col = c("forestgreen", "steelblue", "darkorange"))

# Comparaison du pH du sol entre forêts
boxplot(ph_sol ~ foret,
       data = chene,
       main = "pH du sol par station",
       ylab = "pH",
       col = "lightgray")
```

5.3 Identification Visuelle du Type de Distribution

```
# Vérification de la normalité (QQ-plot)
qqnorm(chene$epaisseur_liege_2024)
qqline(chene$epaisseur_liege_2024, col = "red")

# Interprétation :
# - Si points alignés sur la ligne rouge = distribution normale (Gaussienne)
# - Si points courbés = distribution asymétrique (skewed)
```

6. Synthèse du TP1 - Commandes Essentielles

Action	Commande R	Exemple avec chêne
Importer	<code>read.csv()</code>	<code>chene <- read.csv("Mila.csv")</code>
Explorer	<code>head()</code> , <code>summary()</code>	<code>head(chene)</code> , <code>summary(chene\$altitude)</code>
Sélectionner	<code>\$</code> ou <code>[]</code>	<code>chene\$ph_sol</code> , <code>chene[chene\$foret=="Zouagha",]</code>
Calculer	<code>mean()</code> , <code>sd()</code> , <code>median()</code>	<code>mean(chene\$accroissement_liege)</code>
Par groupe	<code>tapply()</code>	<code>tapply(chene\$diametre, chene\$foret, mean)</code>
Histogramme	<code>hist()</code>	<code>hist(chene\$epaisseur_liege_2024)</code>
Boxplot	<code>boxplot()</code>	<code>boxplot(epaisseur_2024 ~ foret, data=chene)</code>

Tableau 1 : Commandes R essentielles du TP1

Annexe : Description des Variables du Dataset

Variable	Type	Description	Unité	Exemple
arbre_id	Character	Identifiant unique de l'arbre	Code	CL01, CL02...
foret	Character	Nom de la forêt/station	Nom	Zouagha, Tassadane...
altitude	Numeric	Altitude du site	mètres (m)	450, 680, 920
pluie_annuelle	Numeric	Précipitations annuelles	mm	550, 720, 890
temperature_moy	Numeric	Température moyenne annuelle	°C	17.8, 16.2, 14.5
ph_sol	Numeric	Acidité du sol	pH (0-14)	7.8, 6.5, 5.8
epaisseur_liege_2022	Numeric	Épaisseur du liège en 2022	mm	18, 28, 22
epaisseur_liege_2024	Numeric	Épaisseur du liège en 2024	mm	24, 38, 30
accroissement_liege	Numeric	Croissance du liège	mm	6, 10, 8
diametre_chene	Numeric	Diamètre du tronc	cm	42, 58, 48
hauteur_totale	Numeric	Hauteur de l'arbre	mètres	8.5, 12.5, 10.5
production_glands	Numeric	Production de glands	nombre	12, 28, 18
degats_feu	Logical	Dommages par feu	VRAI/Faux	TRUE, FALSE

Tableau 2 : Description des variables du dataset Mila.csv