Université Paris Cité

Data analysis and Exploration using R

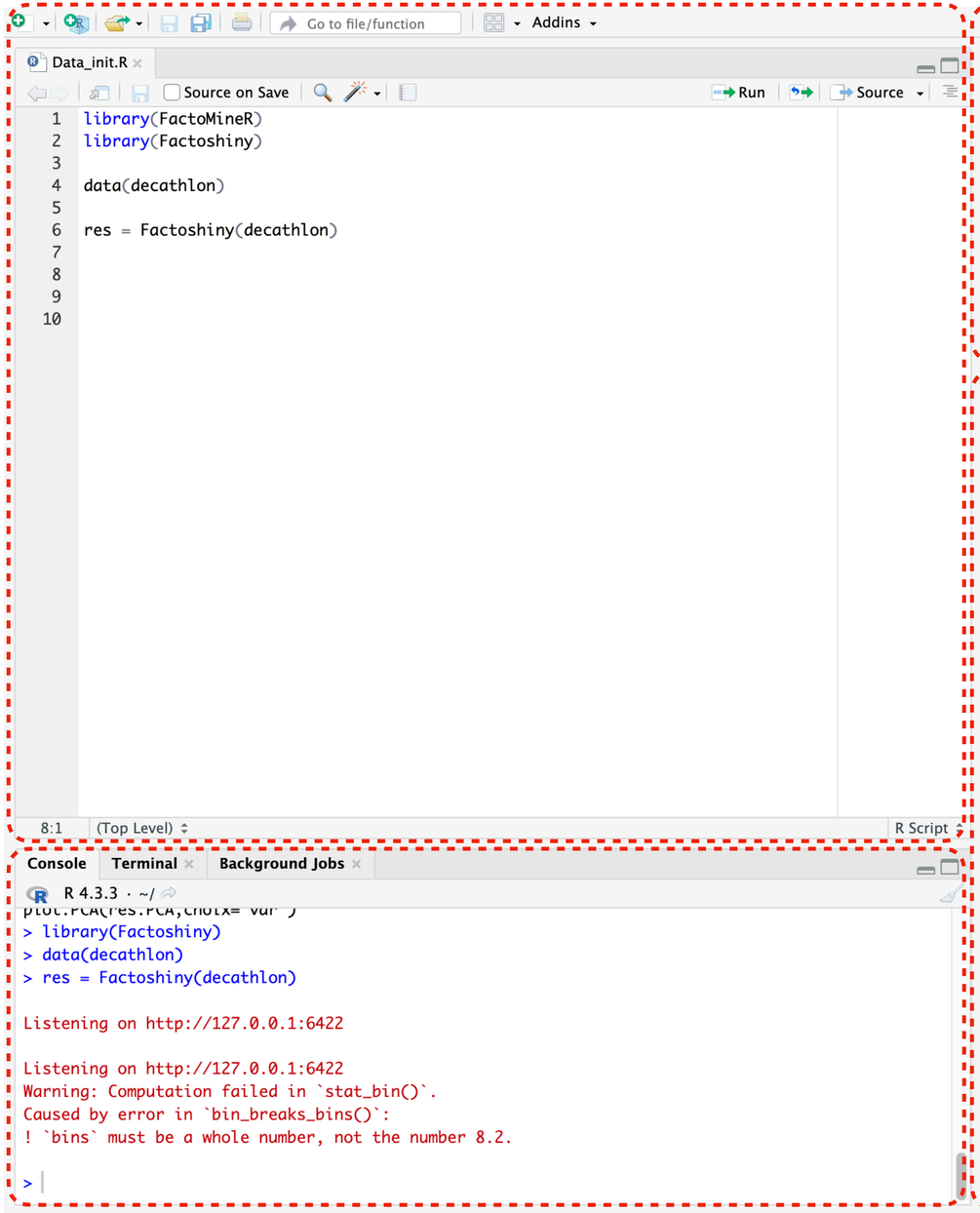
Dimensionality reduction

Amine FERDJAOUI

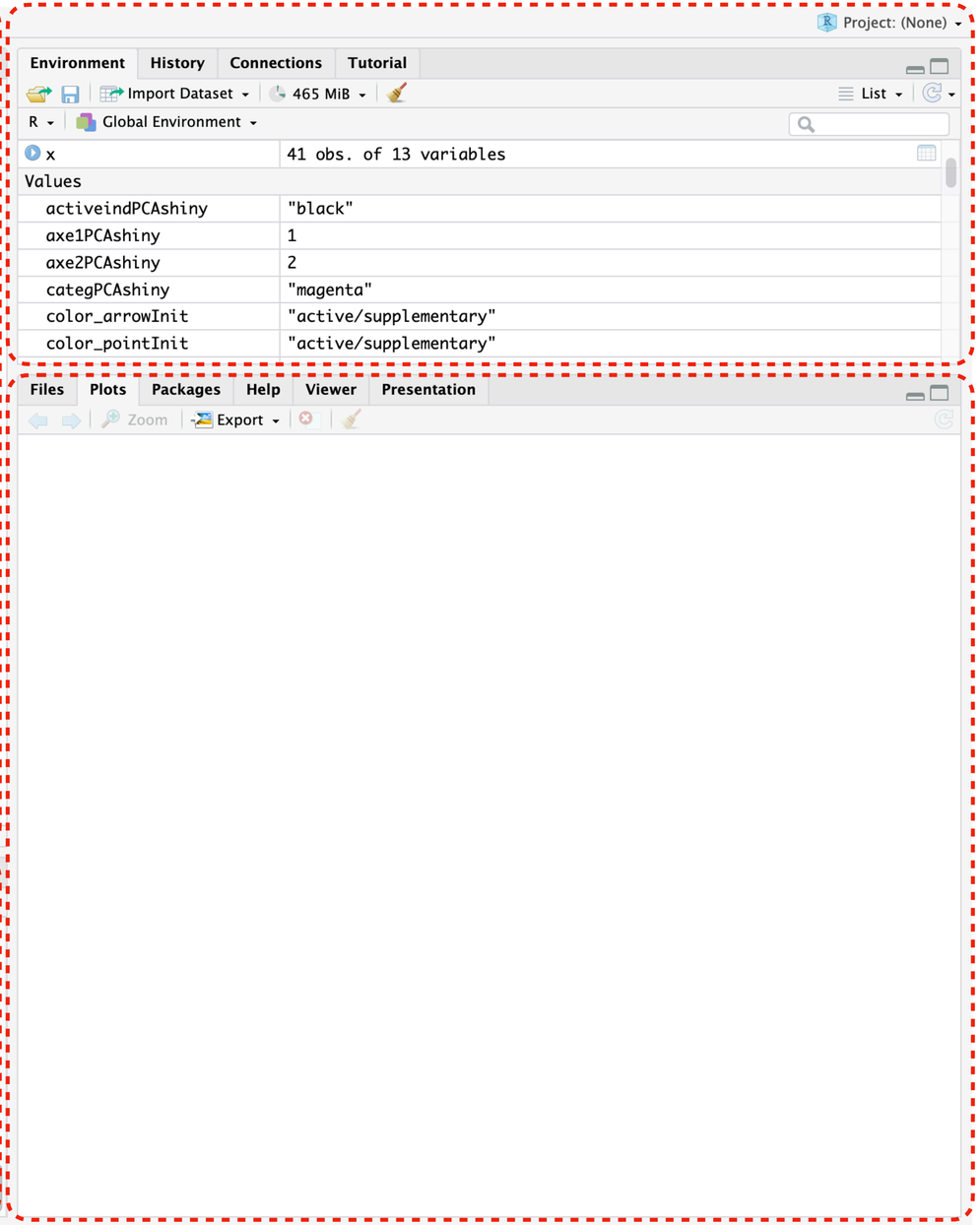
Master I Machine Learning pour la Science des données - Année universitaire 2025/2026

Introduction to 

Code Editor



R console
Terminal



Environment
Variables
History

Plots
Help
Files

R: Introduction

- Install and load library

```
> install.packages("ggplot2") # Install new library
> library(ggplot2) # Load library
```

- Visualise documentation for a function or library

```
> ?mean # or help(mean)
> help("PCA", package = "FactoMineR")
> example(mean)
```

- Load preloaded datasets

```
> data(cars)
> library(help = "datasets")
> view(cars)
```

R: Data Frames

A data frame is a table of data in R:

- each row = one individual (or observation),
- each column = one variable (or attribute),
- columns can be of different types (numeric, text, factor, etc.).

```
> class(cars)
> is.data.frame(cars)
```

```
> df = data.frame(
      Nom = c("Alice", "Bob", "Clara"),
      Age = c(23, 25, 22),
      Sexe = c("F", "M", "F")
   )
```

R: Manipulate Data Frames

```
> head(cars)
> summary(cars)
> head(mtcars)
> names(mtcars) # Show column names
> mtcars$hp <- NULL # Delete a column
> mtcars <- mtcars[-2, ]  # Delete a row
> mtcars$new_var <- 1:nrow(mtcars) # Add a column
```

R: Read csv

- Example of csv file

```
Name; City; Sallary; Year
Alpha; Paris; 22000; 2023
Beta; Lyon; 69500; 2023
Gamma; Marseille; 33400; 2023
Delta; Paris; 12000; 2024
```

- Read csv file

```r
data <- read.csv("ventes.csv", sep = ";", dec = ".", header = TRUE, row.names = 1)
```

Delimiter (eg. , or \t or ;)

Decimal separator (. or ,)

Whether the first like is columns names

Use first column as index

TP1: R basics

A envoyer à : mlds.git@gmail.com

1. Charger un dataset intégré
    1. Chargez le jeu de données **mtcars**.
    2. Visualiser les données sous forme de table.
    3. Affichez les 5 premières lignes et le résumé des variables.
2. Modifier le dataset
    1. Supprimez la colonne drat.
    2. Ajoutez une colonne prix avec des valeurs aléatoires entre 10000 et 40000.
    3. Supprimez la première ligne du tableau.
3. Ajouter une nouvelle ligne
    1. Créez une ligne avec vos propres valeurs et ajoutez-la à la fin.
4. Sauvegarder et réimporter
    1. Sauvegardez votre table CSV en utilisant write.csv (utiliser ?write.csv pour afficher l'aide.)
    2. Réimportez-la avec read.csv() et vérifiez les données.