

Εισαγωγή και βασικές εντολές στην R

1. Setting Up the workspace

Working Directory

```
getwd() # Check current working directory
setwd("C:/Users/kgi062/ML_2025") # Set your working directory to your project folder
```

2. Basic R Commands and Data Types

Printing and Arithmetic

```
print("Hello, world!") # Prints message to console
2 + 2                  # Performs simple addition
```

Variables and Data Types

Examples:

- **Numeric:** Store numbers.
- **Character:** Store text.
- **Logical:** Store TRUE or FALSE.

```
x <- 10 # Numeric variable
name <- "Alice" # Character variable
flag <- TRUE # Logical variable
```

Creating Vectors

```
numbers <- c(1, 2, 3, 4, 5) # Numeric vector
mean(numbers) # Calculate average
```

```
sum(numbers)           # Calculate sum
```

3. Exploring built-in datasets

Access and View (the example of CO2 dataset)

```
?CO2                # Help page for the dataset
View(CO2)           # Opens dataset in RStudio viewer
head(CO2)           # First few rows
tail(CO2)           # Last few rows
colnames(CO2)       # Column names
dim(CO2)            # Dimensions (rows, columns)
str(CO2)            # Structure of the dataset
summary(CO2)        # Summary Statistics
summary(CO2$uptake) # Summary for a specific column
median(CO2$uptake)  # Median of uptake
class(CO2$uptake)   # Data type of the column
unique(CO2$Treatment) # Unique treatment types
subset(CO2, conc == min(CO2$conc)) # Subset for minimum concentration
```

4. Basic Data Visualization

Histogram

```
hist(CO2$uptake) # Distribution of 'uptake'
```

Scatter Plot

```
plot(Orange$age, Orange$circumference) # Plot of two variables
```

Customized Plot

```
plot(Orange$age, Orange$circumference,  
     xlab = "Age", ylab = "Circumference",  
     main = "Circumference vs. Age",  
     col = "blue", pch = 16)
```

5. Importing external data

Reading CSV Files

```
data <-  
read.csv("https://www.dropbox.com/scl/fi/ewm9hn0gjtwhrx1v29qpj/external_file.csv?  
rlkey=00qcf2grifbd88t1ouwub6pqz&dl=0")
```

- **Tip:** Ensure your file path is correct; use forward slashes '/' or double backslashes '\\' on Windows.

Visualize external data

```
plot(data)
```

6. Generating random data

Normal distribution

```
d <- rnorm(100, mean = 5, sd = 1) # Generate 100 normally distributed data  
points  
plot(d) # Plot the generated data
```

7. Additional Tips

- **Documentation:** Use `?function_name` to access help.
- **Code Practices:** Comment your code (`# Comment`) for clarity.
- **Visualization:** Always label axes and add titles for interpretable plots.
- **Data Exploration:** Use `str()`, `summary()`, `head()`, `tail()` frequently to understand data structures.

Δοκιμαστικό script

```
# -----  
# 1. Setting up our workspace  
# -----  
  
# Check current working directory  
getwd()  
  
# Set your working directory to your project folder  
setwd("C:/Users/kgi062/ML_2025")  
  
# -----  
# 2. Basic R Commands and Data Types  
# -----  
  
# Printing  
print("Hello, world!")  
  
# Simple arithmetic  
2 + 2  
  
# Variables and Data Types  
x <- 10           # Numeric Variable  
name <- "Alice"   # Character Variable  
flag <- TRUE      # Logical Variable
```

```

# Creating a numeric vector
numbers <- c(1, 2, 3, 4, 5)

# Calculating mean and sum of the vector
mean_numbers <- mean(numbers)
sum_numbers <- sum(numbers)

# Display results
print(paste("Mean:", mean_numbers))
print(paste("Sum:", sum_numbers))

# -----
# 3. Exploring Built-in Datasets
# -----

# Accessing dataset help
?C02

# Viewing the dataset in RStudio's viewer
View(C02)

# Exploring dataset structure
head(C02)           # First few rows
tail(C02)           # Last few rows
colnames(C02)       # Column names
dim(C02)            # Dimensions (rows, columns)
str(C02)            # Structure of the dataset
summary(C02)        # Summary statistics
summary(C02$uptake) # Summary of a specific column
median(C02$uptake)  # Median value of 'uptake'
class(C02$uptake)   # Data type of 'uptake'
unique(C02$Treatment) # Unique treatment groups
subset(C02, conc == min(C02$conc)) # Subset for minimum concentration

# -----
# 4. Basic Data Visualization
# -----

```

```

# Histogram of 'uptake'
hist(CO2$uptake, main="Histogram of CO2 Uptake", xlab="Uptake", col="lightblue")

# Scatter plot: Orange dataset
plot(Orange$age, Orange$circumference,
     main="Orange: Age vs. Circumference",
     xlab="Age (days)", ylab="Circumference (mm)")

# Customized scatter plot
plot(Orange$age, Orange$circumference,
     xlab="Age", ylab="Circumference",
     main="Circumference vs. Age in Orange Trees",
     col="blue", pch=16)

# -----
# 5. Importing External Data
# -----

# Reading CSV file (update the path accordingly)
# Make sure 'external_file.csv' exists in your working directory
data <- read.csv("external_file.csv") # download it via opencourses
# Visualize the imported data
plot(data[,1]) #bug correction: data is read as a data.frame, not a vector

# -----
# 6. Generating Random Data
# -----

# Generate 100 normally distributed random points with mean=5 and sd=1
d <- rnorm(100, mean = 5, sd = 1)

# Plot the generated data
plot(d, main="Random Normal Data", xlab="Index", ylab="Value")

# Access help for functions

```

```
# e.g., ?mean, ?plot
```

Πώς τρέχουμε τον κώδικα:

1. Ανοίξτε το RStudio.
2. Δημιουργήστε ένα νέο script (File > New File > R Script).
3. Επικολλήστε τον παραπάνω κώδικα μέσα στο script.
4. Αποθηκεύστε το με ένα ενδεικτικό όνομα, π.χ., r_course_2.R.
5. Εκτελέστε γραμμή προς γραμμή (επιλέξτε και πατήστε Ctrl+Enter ή χρησιμοποιήστε το κουμπί "Run") ώστε να παρακολουθείτε τη διαδικασία. Διαφορετικά, μπορείτε να επιλέξετε και να τρέξετε ολόκληρο τον κώδικα απευθείας.