

Basic Syntax and Data Types

R is a language and environment for statistical computing and graphics.

- Assignment: `←` (preferred) or `=`. `x ← 10`.
- Printing: `print(x)`, `cat("Value:", x)`.
- Help: `?function_name` or `help(function_name)`.
- Basic Types: `numeric`, `integer`, `logical` (TRUE/FALSE), `character`, `complex`, `raw`.

Data Structures

Vectors (Atomic)

One-dimensional arrays of the same type.

- `v ← c(1, 2, 3)` (c stands for combine).
- `seq(1, 10, by=2)`: 1, 3, 5, 7, 9.
- `rep(1, 5)`: 1, 1, 1, 1, 1.

Lists

Can contain elements of different types.

- `L ← list(name="Alice", age=25, scores=c(90, 80))`.
- Access: `L$name` or `L[[1]]`.

Matrices

Two-dimensional arrays of the same type.

- `M ← matrix(1:6, nrow=2, ncol=3)`.
- Access: `M[row, col]`.

Data Frames

Tabular data structure where columns can have different types (Essential for data analysis).

- `df ← data.frame(id=1:3, name=c("A", "B", "C"))`.
- `head(df)`, `summary(df)`, `str(df)`.

Factors

Used to store categorical data.

- `f ← factor(c("low", "medium", "high"))`.

Operators

- Arithmetic: `+`, `-`, `*`, `/`, `^`, `%` (modulo), `%/%` (integer division).
- Relational: `==`, `≠`, `<`, `>`, `≤`, `≥`.
- Logical: `&`, `|`, `!` (Element-wise), `&&`, `||` (Scalar).

Control Flow

- **if-else:**

```
if (condition) {
  # Logic
} else {
  # Logic
}
```
- **Loops:**
 - `for (i in 1:10) { ... }`
 - `while (condition) { ... }`
- **apply Family:** Preferred over loops in R.
 - `lapply(list, function)`: Apply to list, return list.
 - `sapply(list, function)`: Simplified version, returns vector/matrix.

Functions

```
my_function ← function(arg1, arg2 =
default_value) {
  result ← arg1 + arg2
  return(result)
}
```

Packages

- Install: `install.packages("package_name")`.
- Load: `library(package_name)`.

Basic Plotting (Base R)

- `plot(x, y)`: Scatter plot.
- `hist(x)`: Histogram.
- `boxplot(x ~ group)`: Box plot.

Statistical Functions

- `mean(x)`, `median(x)`, `sd(x)`, `var(x)`.
- `cor(x, y)`: Correlation.
- `lm(y ~ x, data=df)`: Linear model.
- `t.test(x, y)`: t-test.

Google Style Guide

File Naming

- End files in `.R`. Use lowercase and hyphens/underscores. `data-cleaning.R`.

Identifiers

- Variable/Function: `variable_name` or `variableName`. Be consistent.
- Constants: `kConstantName` (starts with k).

Formatting

- Indentation: 2 spaces.
- Line Length: Max 80 characters.
- Spacing: Spaces around all binary operators (`+`, `-`, `←`, etc.). No space before a comma, but always one after.

Assignment

- Always use `←` for assignment, not `=`.

Documentation (roxygen2)

- Use `#'` for documentation comments above functions.
- Include `@param`, `@return`, `@export`.

Misc

- Avoid `attach()`.
- Explicitly use `return()` at the end of functions for clarity.
- Use `T` and `F` sparingly; prefer `TRUE` and `FALSE`.