Computational Biology

BT1013 Section 829

March 23

Data types and data structures

> Source: https://swcarpentry.github.io/r-novice-inflammation/13-supp-data-structures/

Data types:

- character
- numeric
- integer
- logical
- complex

Data structures:

- vector
    - atomic: single type data
    - list: multiple types
- matrix
- data frame
- factors

Object attributes:

- `class()` : kind of object
- `typeof()` : object's data type
- `length()` : vector/list length
- `nchar()` : string size
- `dim()` : array dimensions
- `attributes()` : metadata
- `names()` : list elements names

Missing data: NA

- `is.na()` : returns logical vector with TRUE for each missing value
- `anyNA()` : TRUE if there's any NA

Infinity: Inf (can be + or -)

Not a Number: NaN (undefined)

Vector

- `vector()` : empty logical vector

- `logical(1)`: same as above
- `vector("character", length=sz)`: mode "character" vector of length `sz`
- `character(sz)`: same as above
- `numeric(sz)`: mode "numeric" vector
- `c()`: combine function (implicitly coerces to single type)
- `c(1L, 2L, 3L)`: mode "integer" vector containing 1, 2, 3 as `int`
- `c(TRUE, FALSE)`: mode "logical" vector
- sequence of numbers:
  - `1:10`: vector from 1 to 10 in increments of 1
  - `seq(10)`: same as above
  - `seq(from=1, to=10, by=1)`: same as above
- `as.<class_name>()`: explicitly coerces vectors to class

Matrix

- filled column-wise

- `byrow = TRUE`: forces row-wise fill

- `matrix(nrow=x, ncol=y)`: x-by-y matrix

- `matrix(1:6, nrow=2, ncol=3)`: 2-by-3 matrix containing `[1 3 5 ; 2 4 6]`

- alternate method 1:

  ```
  m <- 1:6
  dim(m) <- c(2,3)
  ```

- alternate method 2 (column bind):

  ```
  c1 <- c(1, 2)
  c2 <- c(3, 4)
  c3 <- c(5, 6)
  m <- cbind(c1, c2, c3)
  ```

- alternate method 3 (row bind):

  ```
  r1 <- c(1, 3, 5)
  r2 <- c(2, 4, 6)
  m <- rbind(r1, r2)
  ```

- `m[x, y]`: access element in x-th row and y-th column

List

- `list(1, "a", TRUE, 3+4i)`: list containing four elements of different data types

- `as.list()` : coerces other objects
- `vector("list", length=sz)` : empty list of length `sz`
- `list(alpha="Bravo", charlie=1:10)` : list containing to named elements
- `li[[x]]` : access x-th element in list
- `li$alpha` : access element named alpha in list

Data frame

- rectangular list (every element has the same length)
- can be accessed like lists
- attributes:
  - `rownames()` : to annotate `subject_id` or `sample_id`
- usually created by `read.csv()` and `read.table()`
- `data.matrix()` : converts data frame into table
- `data.frame()` : new data frame
- `data.frame(id=letters[1:5], x=6:10)` : data frame with letters a-e and numbers 6-10

Data frame functions:

- `head()` : shows first 6 rows
- `tail()` : shows last 6 rows
- `str()` : structure of data frame (name, type, preview of data in each column)
- `sapply(dataframe, class)` : shows class of each column

Summary

Dimensions	Homogenous	Heterogenous
1D	atomic vector	list
2D	matrix	data frame

Functions

```
name <- function(arguments) {operation}
add <- function(x, y) {x + y}
```