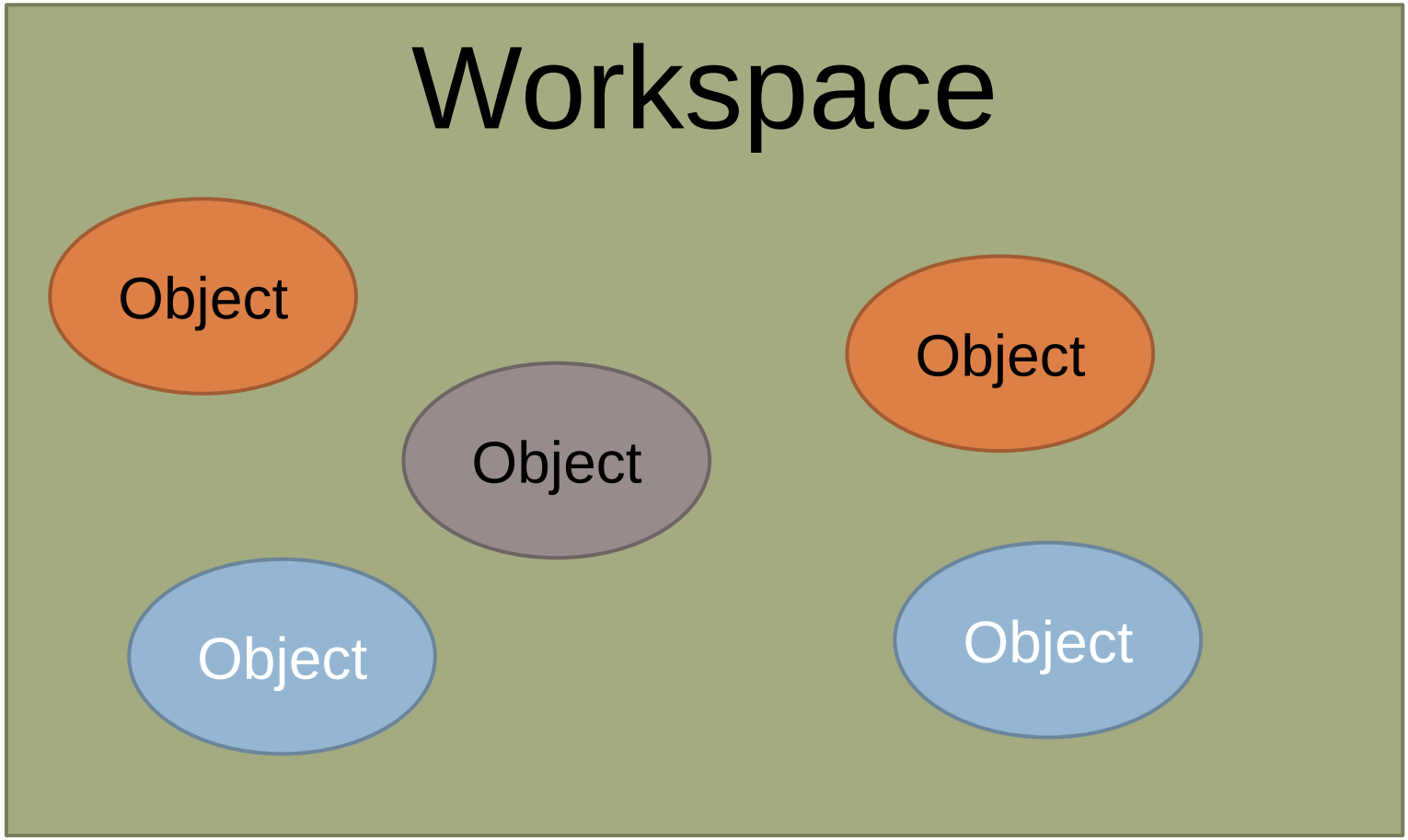
DATA STRUCTURES IN R

Arne Hendrik Schulz

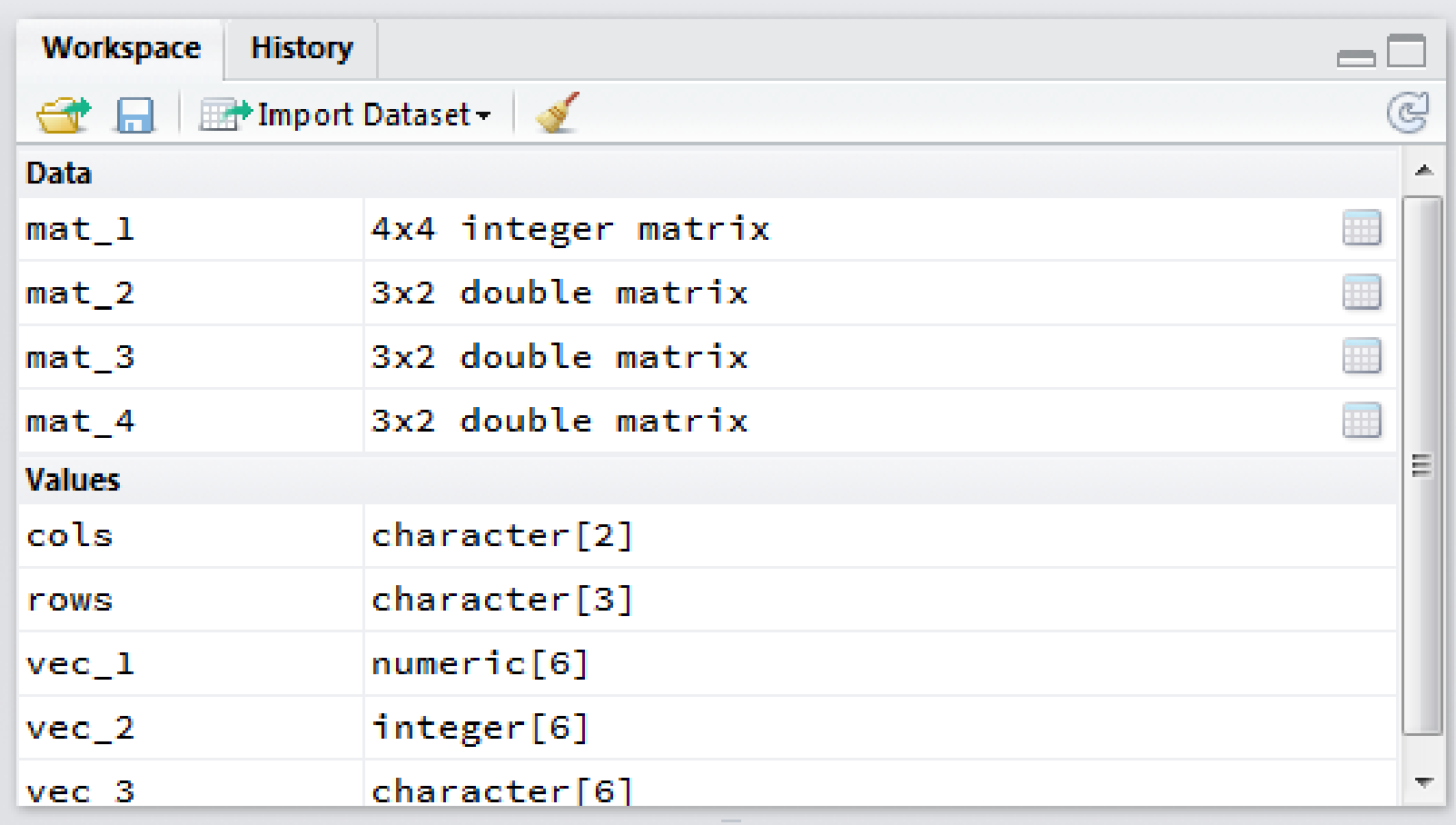
Overview

- General structure
- Data structures
  - Vector
  - Matrix
  - Array
  - Factor
  - Data Frame
  - List
- Accessing data

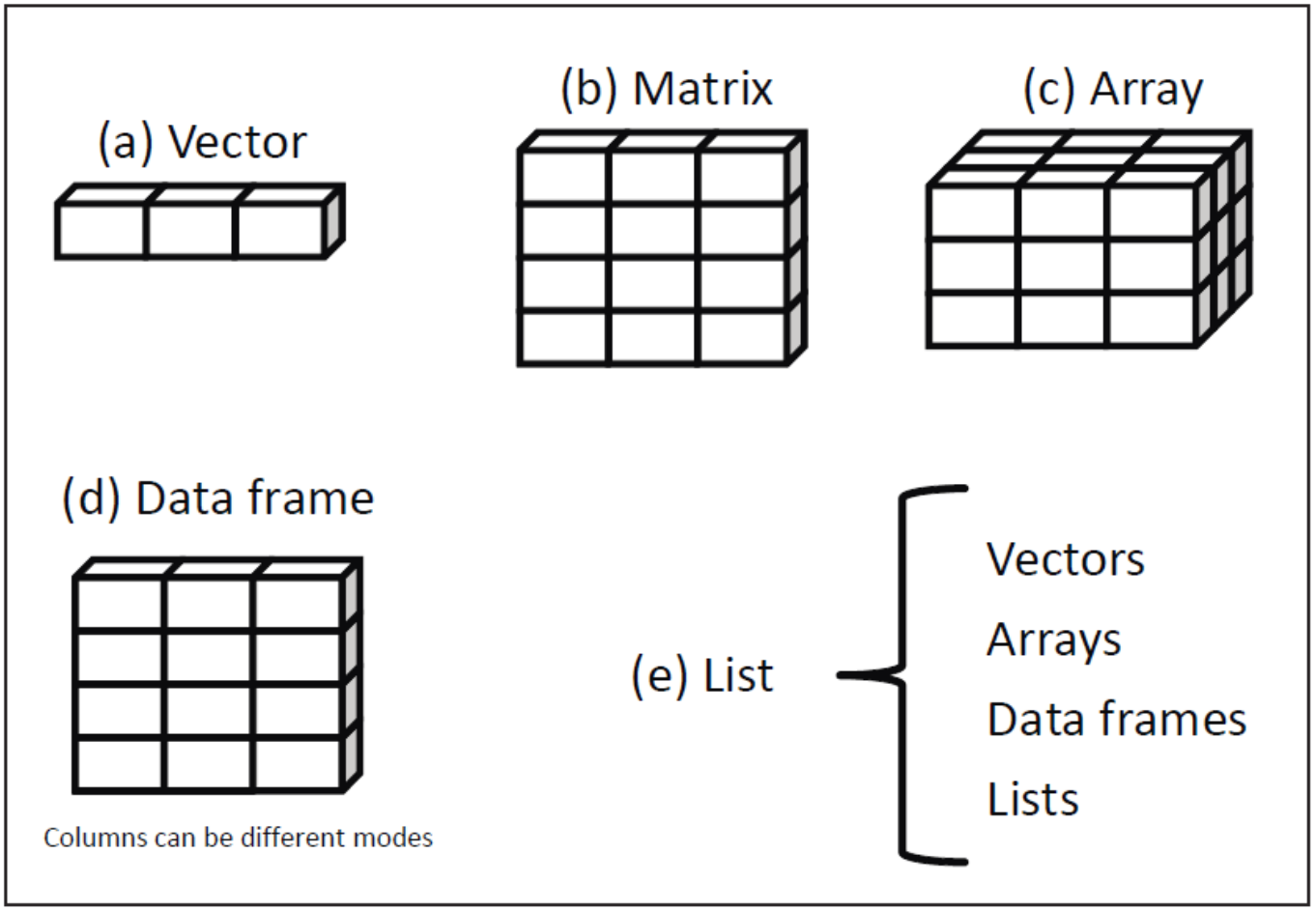
General structure



Workspace


Workspace
History
Import Dataset
Data
mat_1 4x4 integer matrix
mat_2 3x2 double matrix
mat_3 3x2 double matrix
mat_4 3x2 double matrix
Values
cols character[2]
rows character[3]
vec_1 numeric[6]
vec_2 integer[6]
vec_3 character[6]

Data structures



Kabacoff, R. (2011): R in action. Manning, Shelter Island: p. 23

Vector

- One-dimensional values
- Numeric, character or logical values

Matrix

- Two-dimensional Vector
- Same values for each element
  - Numeric, character or logical values
  - Main difference to Data Frames

Array

- Three-dimensional Vector/Matrix
- Same Values for each element
  - Numeric, character or logical values

Factor

- One-dimensional values
- Numeric, character or logical values
- But only useful for character values
  - Internal storage via levels → memory efficient
  - Levels can be ordered →ordinal variables

Data Frame

- Very similiar to Matrices
- Culoums can contain different data types
  - Numerical, character, Logical **and** factorical

List

- Considered as most complex data structure
- But also most flexible
- List can contain all other types of data structures
  - Vectors, Matrices, Arrays, Data Frames and Lists
- No dimensional limitations

Data types

- Numeric/Integer
- Factor
- Character
- Logical
- Date

Accessing data

- Brackets […]
  - Input depends on dimensions
    - Vector/Factor [x]
    - Matrix/Data Frame [x,y]
    - Array [x,y,z]
    - List should be accessed through [[x]]
      - List[x] works, but keeps output as list

Accessing data (II)

- Special forms
  - Data Frames/Matrices/Arrays/Lists can be accessed through variable/element name
    - Matrix/Data Frame: object[c(„rownames“), c(“colnames“)]
    - Array: array[c(„rownames“), c(“colnames“), c(„dimension“)]
    - Data Frame/List: object$variable
    - List: list[[„element“]] or list$element

Accessing data (III)

- Nesting is possible
  - List$data_frame$variable
- Different ways to access data
  - Array[,,z][x,y] ←→ array[x,y,z]