RNA-seq analysis using R

Atefeh Taherian Fard & Huiwen Zheng

November 2020

Australian Institute for Bioengineering and Nanotechnology

R!

- Is a language
  - Free and open source
  - Can be integrated with other programming languages
  (C/C++, Java and Python)

- Applications
  - One of the most popular language
  - Great tool to explore and investigate data
  - Statistical testing
  - Data visualisation
  - Facebook, Google, Ford Motor Company and millions of people Worldwide from different fields are using R language
  - *Finance, Bio Science, Supply chain, Sports, Retail, Marketing, and Manufacturing.*

RStudio

- It's an Integrated Development Environment (IDE) for R

- A platform to use R

- Makes R easier to use

- Provides many features for authoring, modifying, compiling, deploying and debugging software

- You can use R without RStudio but you can't use RStudio without R

Getting help

- To seek help for special operator use double quote
    - ?"<-"
- Vignette() show all install packages
- Vignette(package="packages-name"); will list all available vignettes for packages
- Vignette("packages-name"); will open the specified vignette
- help("package-name"); if no vignette
- ?function_name; if you know the exact name
- ??function_name; fuzzy search
- CRAN Task Views; specially maintained list of packages grouped into fields.
- ?dput; Will dump the data you're working with into a format so that it can be copy and pasted by anyone else into their R session.
- sessionInfo()

Data types in R

- character: "a" , "AIBN"

- double: 2020, 2.1

} numeric

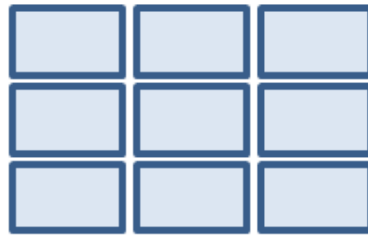- integer: 2L (the L tells R to store this as an integer)

- logical: TURE, FALSE

Basic data structures in R

Vector	Matrix	Data Frame

- 1 column or row of data
- 1 type (numeric or text)

- multiple columns and/or rows of data
- 1 type (numeric or text)

- multiple columns and/or rows of data
- multiple types

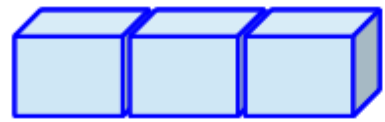- aka *Atomic vectors* (cause they can't be further simplified)

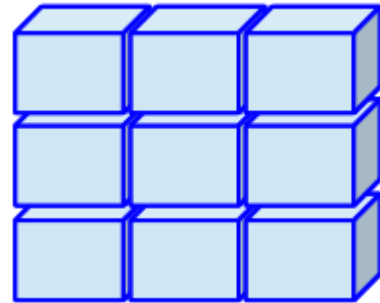Numeric Vectors (mode: *numeric*)
Logical Vectors (model: *logical*)
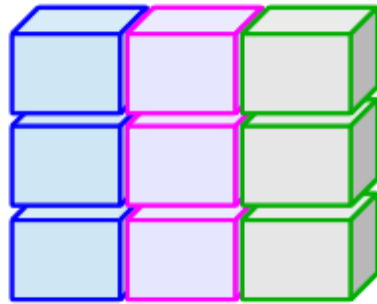Character Vector or text strings (mode: *character*)
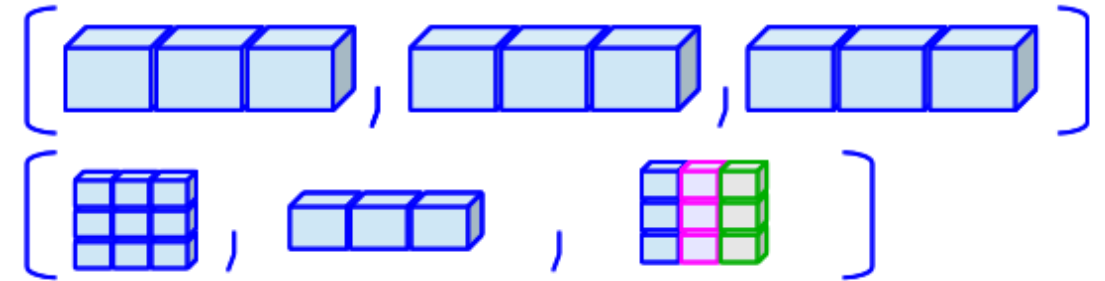
Basic data structures in R (a bit more detail)

vector

One element → 12.3

Four element → 12.3 / 3 / 5.01 / 2.3

matrix

12.3	0.1
3.0	5.2
5.01	3.0
2.3	0.1

dataframe

x	y
12.3	ace
3	tea
5.01	oil
2.3	tree

list

x	y
12.3	ace
3	tea
5.01	oil
2.3	tree

3

Y ~ x −1

some text

Basic data structures in R (cont…)

- Vector
- Matrix
- Dataframe
- List
- Factor: similar to vector, can be used to store *categorical* variable in a data set

age
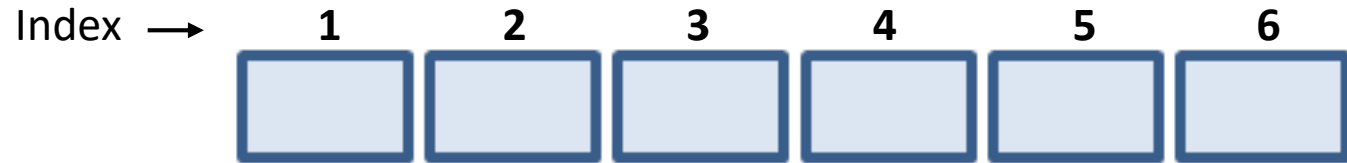| young	old	midage	old	old	young	midage	midage	young	young

Factor: age
Levels: "young", "old", "midage"

Functions to explore the features of vectors and other objects

- Create different atomic vectors with different data types:
  - c()
- Other functions:
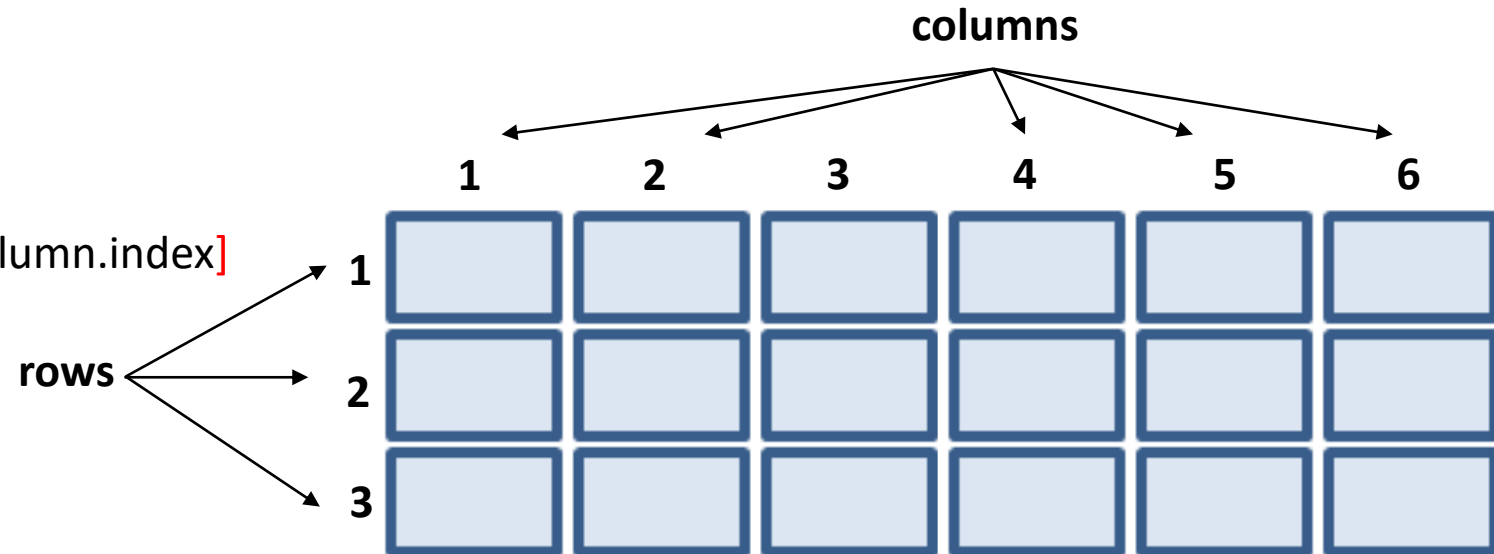 - typeof()
  - length()
  - data.frame()

Indexing

Vector (and factor) (1D)

- Square bracket: [index]

Index →

| 1 | 2 | 3 | 4 | 5 | 6 |

columns

data.frame and matrix (2D)

- Square bracket: [row.index , column.index]

rows

	1	2	3	4	5	6
1						
2						
3						

Indexing

vector (and factor) (1D)

- Square bracket: [index]

Index ⟶ | **ab** | **cd** | **yt** | **ghj** | **apple** | **city** |
|---|---|---|---|---|---|

columns

data.frame and matrix (2D)

- Square bracket: [row.index , column.index]

	ab	**cd**	**age**	**sex**	**ID**	**city**
Ati						
Sas						
Huiwen						

rows
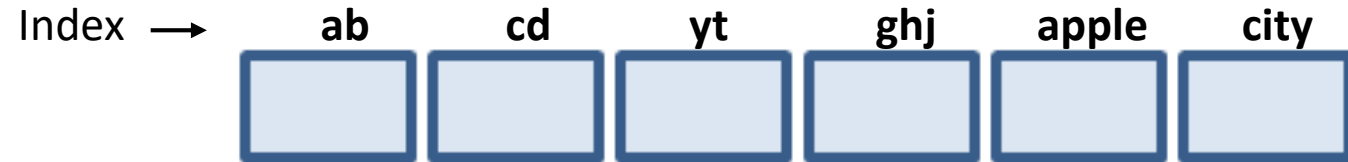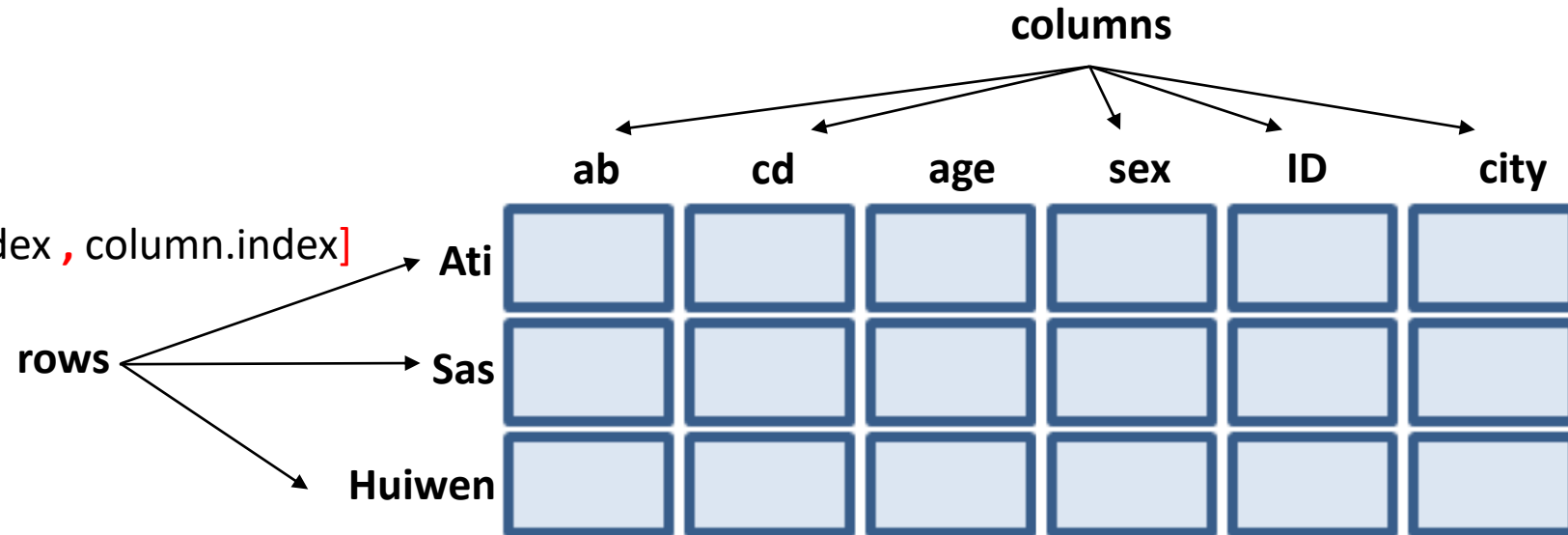
Subsetting

- Six different ways we can subset any kind of object (one example each)
- Three different subsetting operators for the different data structures

  - Access individual values by location using [] (get me the nth number in square bracket)
  - Access slices of data using [low:high]
  - Access arbitrary sets of data using [c(...)]
  - [[to extract individual elements of a list
    - You can't extract more than one element at once
  - $ function to extract elements by name

More help!

- [CRAN Task Views](#) is a specially maintained list of packages grouped into fields.

- [Quick R](#)

- [RStudio cheat sheets](#)

- [Cookbook for R](#)