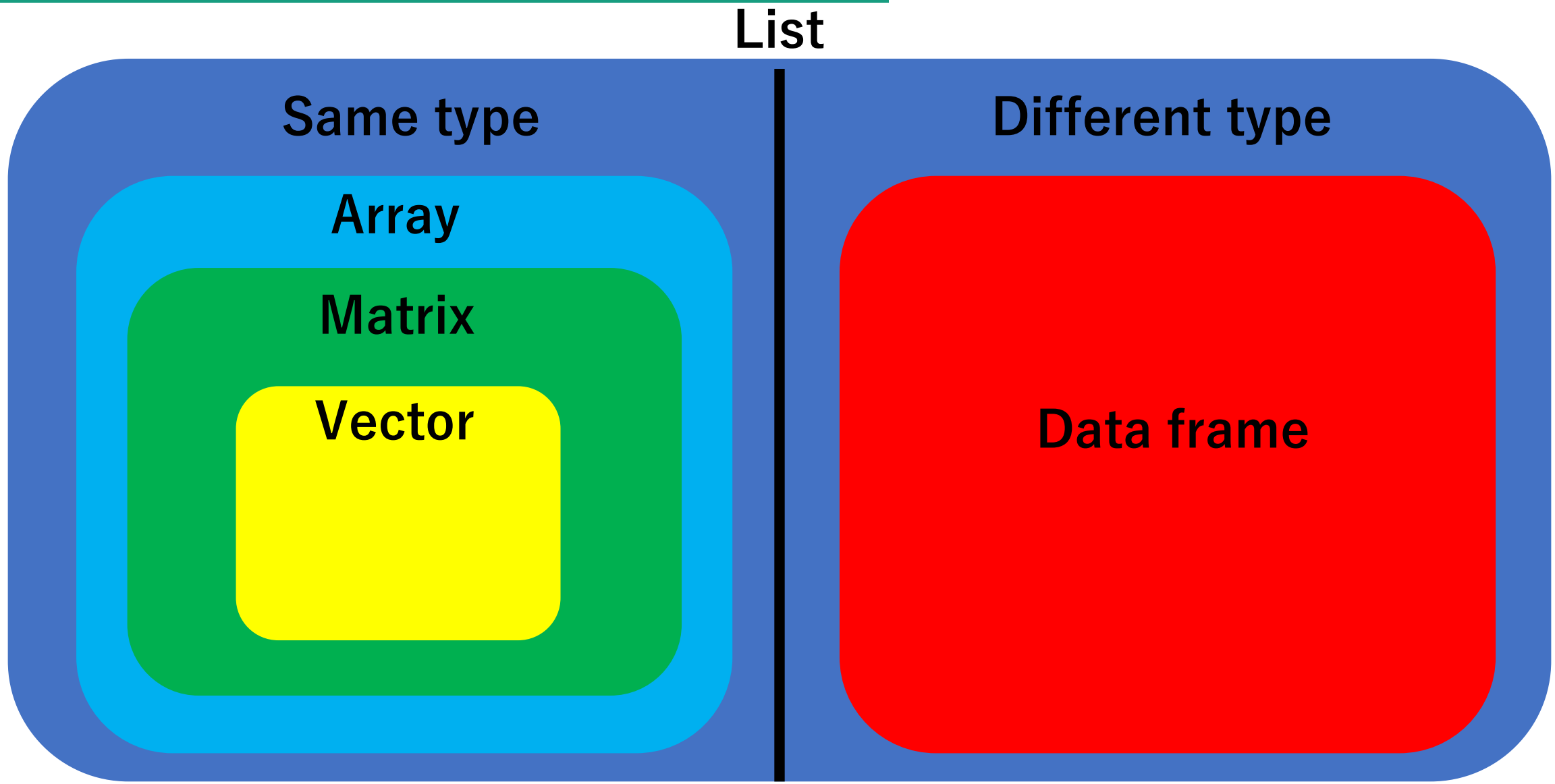
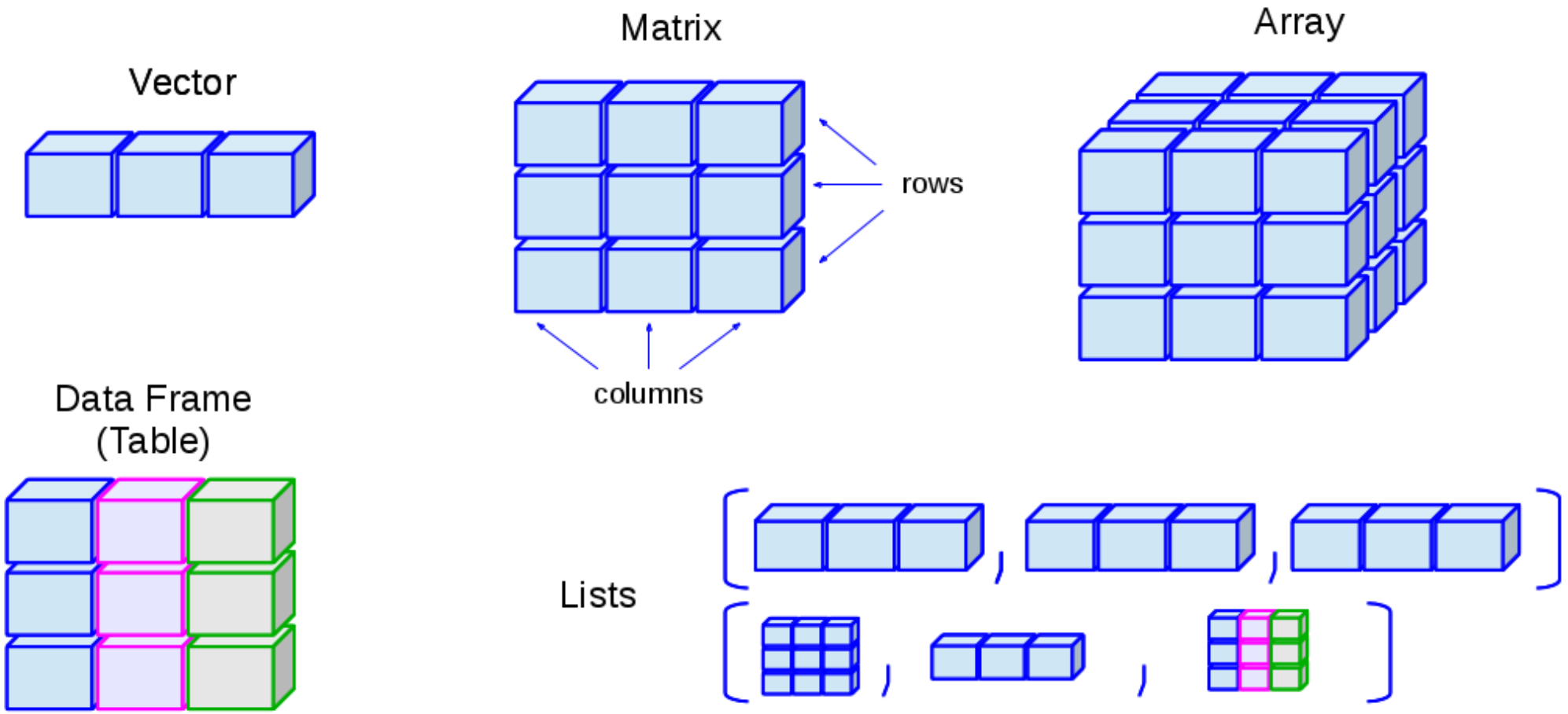
R for toxicology

1. Data structure and Index

Types of data structure in R



Types of data structure in R



Reference URL: http://venus.ifca.unican.es/Rintro/dataStruct.html

Data structure

1. Vector: One or more factors + 1D

```
> b <- c(1,2,3,4,5)
> b
[1] 1 2 3 4 5
```

2. Matrix: same structure + 2D

```
> mt <- matrix(1:6, ncol = 3)
> mt
     [,1] [,2] [,3]
[1,]    1    3    5
[2,]    2    4    6
```

Data structure

3. Array: same structure + 2D or higher

```
> ary <- array(1:6, c(1,2,2))
> ary
, , 1

     [,1] [,2]
[1,]    1    2

, , 2

     [,1] [,2]
[1,]    3    4
```

1x2 Matrix to second structure

Data structure

4. Data frame: 2D + including all types of data

```
> student_name <- c("Jack", "Mark", "John")
> avg_score <- c(90, 85, 80)
> dt <- data.frame(student_name, avg_score)
> dt
  student_name avg_score
1         Jack        90
2         Mark        85
3         John        80
```

Data structure

5. List: Includes all different structures of data

```
> lst <- list(a,ary,dt)
> lst
[[1]]
[1] 1

[[2]]
, , 1

     [,1] [,2]
[1,]    1    2

, , 2

     [,1] [,2]
[1,]    3    4


[[3]]
  student_name avg_score
1         Jack        90
2         Mark        85
3         John        80
```

What is indexing?

- Selecting and searching data

- When you do preprocessing using R, indexing is extremely important

Data indexing

1. Vector: using [a]

```
> b <- c(234,6,58,234,7657,23454245,65878)
> b[3]
[1] 58
> b[c(1,3)]
[1] 234  58
```

3rd one

1st and 3rd one

Data indexing

2． Matrix：[a, b]を使う

```
> mt
     [,1] [,2] [,3]
[1,]    1    3    5
[2,]    2    4    6
> mt[1,2]
[1] 3
> mt[1,]
[1] 1 3 5
```

(1, 2)の結果

１行の結果全部

Data indexing

3. Array: use [a, b, c]

```
> ary
, , 1

     [,1] [,2] [,3]
[1,]    1    3    5
[2,]    2    4    6

, , 2

     [,1] [,2] [,3]
[1,]    7    9   11
[2,]    8   10   12

, , 3

     [,1] [,2] [,3]
[1,]   13   15   17
[2,]   14   16   18

, , 4

     [,1] [,2] [,3]
[1,]   19   21   23
[2,]   20   22   24
```

[row, column, array number]

```
> ary[2,3,4]
[1] 24
```

Data indexing

4. Data frame: use [a], [[a]], and/or $

```
> dt
      names score
1       Kim    95
2 Yamasaki    83
3       Liu   100
4   Jackson    39
```

```
> dt[2]
   score
1     95
2     83
3    100
4     39
```

Data frame

Different numbers of square brackets can show different results

```
> dt[[2]]
[1]  95  83 100  39
```

Vector

Data indexing

```
> dt
      names score
1       Kim    95
2 Yamasaki    83
3       Liu   100
4   Jackson    39
```

Select ‘names’ column from ‘dt’

```
> dt$names
[1] Kim       Yamasaki Liu      Jackson
Levels: Jackson Kim Liu Yamasaki
```

Data indexing

5. List: select table using [[a]]

```
> lst
[[1]]
[1] "Lee"  "Kim"  "Park"

[[2]]
      names score
1       Kim    95
2 Yamasaki    83
3       Liu   100
4   Jackson    39
```

```
> lst[[2]]
      names score
1       Kim    95
2 Yamasaki    83
3       Liu   100
4   Jackson    39
> lst[[2]][1,2]
[1] 95
```

Practice

1. Make a data frame

	Student_Names	Student_Scores
1	James	100
2	Tom	94
3	Jack	85
4	Kim	97
5	Yamaoka	95
6	Liu	85

Practice

2. Index the former data frame and show results as follows

（1）

	Student_Scores
1	100
2	94
3	85
4	97
5	95
6	85

（2）

```
[1] 100
```

Practice

（3）

```
  Student_Names
1         James
2           Tom
3          Jack
4           Kim
5       Yamaoka
6           Liu
```

（4）

```
[1] 100  94  85  97  95  85
```