

毒性学メンバーのためのR分析

1. データ構造とIndexing



Rのデータ構造の種類

List

同じタイプ

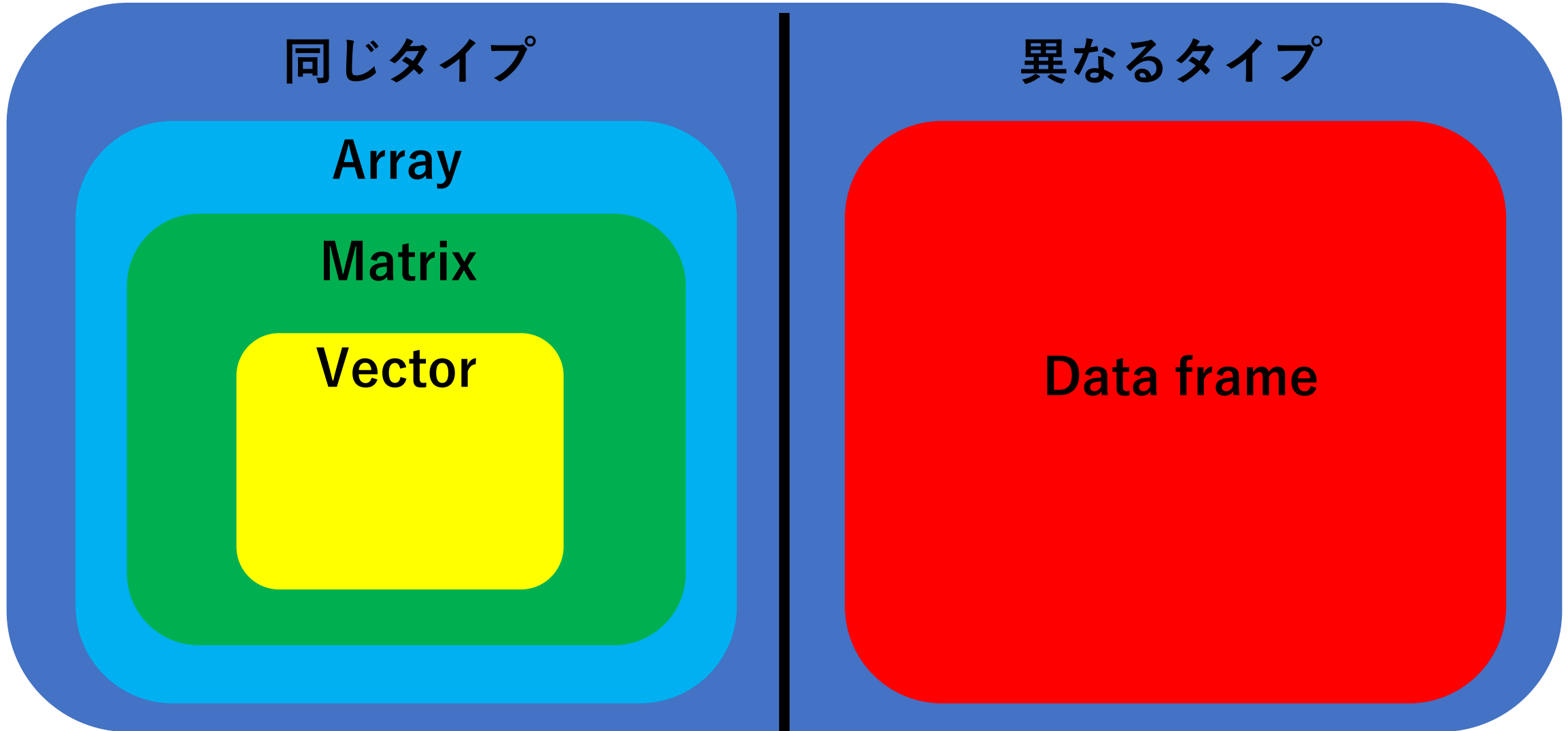
Array

Matrix

Vector

異なるタイプ

Data frame

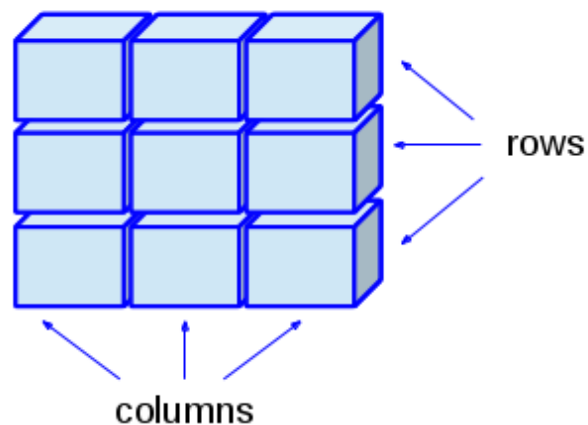


Rのデータ構造の種類

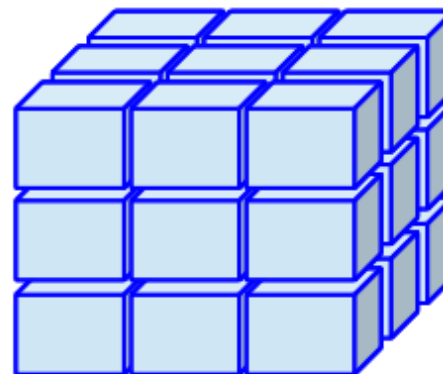
Vector



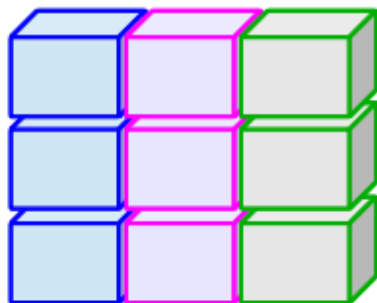
Matrix



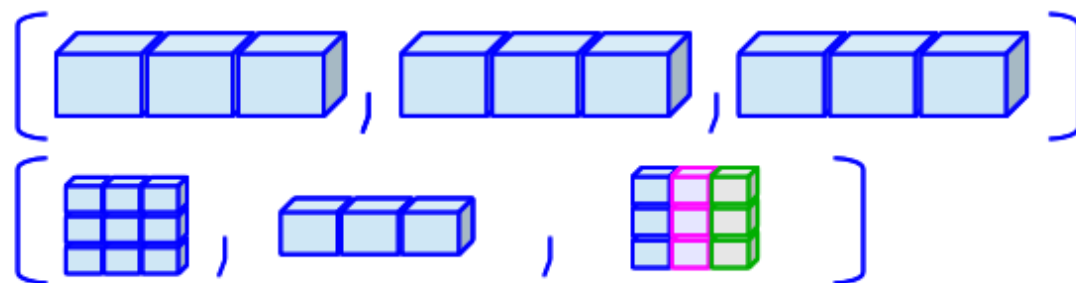
Array



Data Frame
(Table)



Lists



Rのデータ構造

1. Vector : 構成因子が一つ以上 + 1次元

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

2. Matrix : 同じ型 + 2次元

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

Rのデータ構造

3. Array : 同じ型 + 2次元以上

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

```
> ary  
[, , 1
```

1 x 2 Matrixを 2 階構造に

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

```
[, , 2
```

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

Rのデータ構造

4. Data frame : 2次元+データ型は関係ない

```
> 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

Rのデータ構造

5. List：異なる構造のデータを全部含んでいる

```
> 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

Indexingとは？

- データを選別、検索する作業
- Rを使ったデータの前処理、分析に非常に重要

RのデータIndexing

1. Vector : [a]を使う

```
> b <- c(234, 6, 58, 234, 7657, 23454245, 65878)
```

```
> b[3] 3 番目
```

```
[1] 58
```

```
> b[c(1, 3)] 1 番目と 3 番目
```

```
[1] 234 58
```

RのデータIndexing

2. Matrix : [a, b]を使う

```
> mt
```

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

```
> mt[1,2] (1, 2)の結果
```

```
[1] 3
```

```
> mt[1,] 1行の結果全部
```

```
[1] 1 3 5
```

RのデータIndexing

3. Array : [a, b, c]を使う

```
> 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

```
> ary[2,3,4] [行, 列, 層]  
[1] 24
```

RのデータIndexing

4. Data frame : [a], [[a]], \$などをを使う

```
> dt
```

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

```
> dt[2]
```

データフレーム

	score
1	95
2	83
3	100
4	39

括弧の種類によって
結果の種類が違う！

```
> dt[[2]]
```

ベクトル

[1]	95	83	100	39
-----	----	----	-----	----

RのデータIndexing

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

データフレームのnamesの部分！

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

RのデータIndexing

5. List : [[a]]で層を選ぶ

```
> 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
```

練習問題

1. 以下のデータフレームつくりなさい

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

練習問題

2. 作ったデータフレームから次のIndexingの結果が出るようにIndexingをしない

(1)

student_scores

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

(2)

[1] 100

練習問題

2. 次のIndexingの結果が出るようにIndexingしなさい

(3)

Student_Names

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

(4)

[1] 100 94 85 97 95 85