

## Basic Data Types (vectors, matrices, data frame, list)

<http://www.r-tutor.com/r-introduction/basic-data-types>

- **Vectors**

A **vector** is a sequence of data elements of the same basic type. Members in a vector are called **components**. Here is a vector containing three numeric values 2, 3 and 5.

```
> x=c(2, 3, 5)
[1] 2 3 5
```

A vector can contain character strings.

```
> c("aa", "bb", "cc", "dd", "ee")
[1] "aa" "bb" "cc" "dd" "ee"
```

A vector of logical values.

```
> c(TRUE, FALSE, TRUE, FALSE, FALSE)
[1] TRUE FALSE TRUE FALSE FALSE
```

Incidentally, the number of members in a vector is given by the length function.

```
> length(c("aa", "bb", "cc", "dd", "ee"))
[1] 5
```

- **Matrices**

A **matrix** is a collection of data elements arranged in a two-dimensional rectangular layout (rows and columns). The following is an example of a matrix with 2 rows and 3 columns.

$$A = \begin{bmatrix} 2 & 4 & 3 \\ 1 & 5 & 7 \end{bmatrix}$$

```
> A = matrix(
+   c(2, 4, 3, 1, 5, 7), # the data elements
+   nrow=2,             # number of rows
+   ncol=3,             # number of columns
+   byrow = TRUE)      # fill matrix by rows

> A                    # print the matrix
     [,1] [,2] [,3]
[1,]    2    4    3
[2,]    1    5    7
```

An element at the  $m^{\text{th}}$  row,  $n^{\text{th}}$  column of A can be accessed by the expression A[m, n].

```
> A[2, 3]             # element at 2nd row, 3rd column
[1] 7
```

The entire  $m^{\text{th}}$  row A can be extracted as A[m, ].

```
> A[2, ]              # the 2nd row
[1] 1 5 7
```

Similarly, the entire  $n^{\text{th}}$  column A can be extracted as A[,n].

```
> A[,3]      # the 3rd column
[1] 3 7
```

- **Data frame**

A **data frame** is used for storing data tables. **It is a list of vectors of equal length.** For example, the following variable df is a data frame containing three vectors n, s, b.

```
> n = c(2, 3, 5)
> s = c("aa", "bb", "cc")
> b = c(TRUE, FALSE, TRUE)
> df = data.frame(n, s, b)      # df is a data frame
```

### Example: TROPHY data frame

Reading TROPHY Data into R:

- A) TROPHY=read.csv("D:/ Data/TROPHY.csv") (D:/Data is the directory where data is saved)
- B) Use Import Dataset Tab under Environment Window (Top Right window in R-Studio)

```
>head(TROPHY)
```

	Sex	Smoke	Age	BMI	Insulin	Gluc_fast	Ins_gluc	Tri glyceride	...
1	Male	4	50	30.4	16.1	88	24	158	...
2	Male	1	57	29.1	12.6	121	13	106	
3	Male	1	52	25.2	5.7	113	7	250	
4	Male	1	47	28.4	4.2	89	6	61	
5	Male	2	46	27.7	4.5	84	7	111	
6	Male	1	43	29.1	8.9	100	12	96	...
...									

The top line of the table, called the **header**, contains the column names. Each horizontal line afterward denotes a **data row**, which begins with the name (or index number) of the row, and then followed by the actual data. Each data member of a row is called a **cell**.

- **List**

A **list** is a generic vector containing other objects.

For example, the following variable x is a list containing copies of three vectors n, s, b, and a numeric value 3.

```
> n = c(2, 3, 5)
> s = c("aa", "bb", "cc", "dd", "ee")
> b = c(TRUE, FALSE, TRUE, FALSE, FALSE)
> x = list(n, s, b, 3)      # x contains copies of n, s, b
```

R uses list to automatically store the output from various analysis. We will see examples of lists later in labs