Data Frame in R

R_Programming for DataScience

What is a Data Frame?

A **data frame** is a list of vectors which are of equal length. A matrix contains only one type of data, while a data frame accepts different data types (numeric, character, factor, etc.).

How to Create a Data Frame

We can create a data frame by passing the variable a,b,c,d into the data.frame() function. We can name the columns with name() and simply specify the name of the variables.

```
data.frame(df, stringsAsFactors = TRUE)
```

Arguments:

- **df**: It can be a matrix to convert as a data frame or a collection of variables to join
- **stringsAsFactors**: Convert string to factor by default

We can create our first data set by combining four variables of same length.

```
# Create a, b, c, d variables

a <- c(10,20,30,40)

b <- c('book', 'pen', 'textbook', 'pencil_case')

c <- c(TRUE,FALSE,TRUE,FALSE)

d <- c(2.5, 8, 10, 7)

# Join the variables to create a data frame

df <- data.frame(a,b,c,d)

df
```

We can see the column headers have the same name as the variables. We can change the column name with the function names().

```
# Name the data frame

names(df) <- c('ID', 'items', 'store', 'price')

df
```

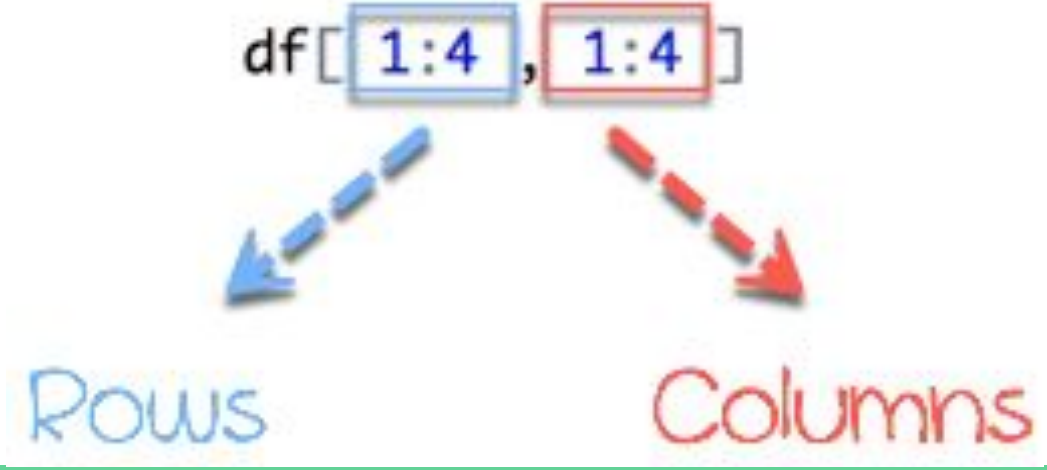
```
# Print the structure

str(df)
```

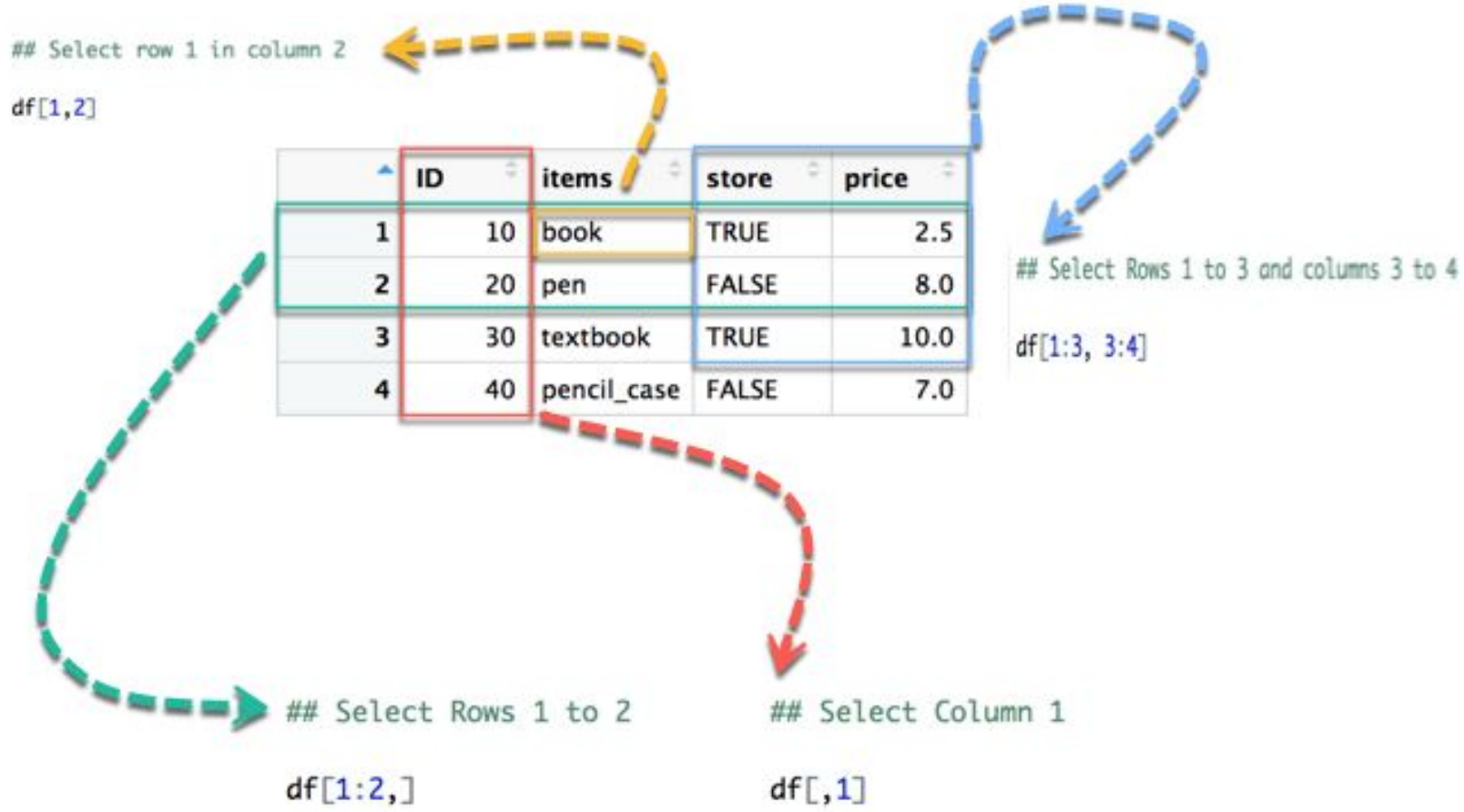
Slice Data Frame

Slice value of Data Frame

It is possible to SLICE values of a Data Frame. We select the rows and columns to return into bracket precede by the name of the data frame.

A data frame is composed of rows and columns, df[A, B]. A represents the rows and B the columns. We can slice either by specifying the rows and/or columns.




Select row 1 in column 2
df[1,2]
ID
items
store
price
1
10
book
TRUE
2.5
2
20
pen
FALSE
8.0
3
30
textbook
TRUE
10.0
4
40
pencil_case
FALSE
7.0
Select Rows 1 to 3 and columns 3 to 4
df[1:3, 3:4]
Select Rows 1 to 2
df[1:2,]
Select Column 1
df[,1]

Append a Column to Data Frame

You can also append a column to a Data Frame. You need to use the symbol $ to append a new variable.

```
# Create a new vector

quantity <- c(10, 35, 40, 5)



# Add `quantity` to the `df` data frame

df$quantity <- quantity

df
```

Select a Column of a Data Frame

Sometimes, we need to store a column of a data frame for future use or perform operation on a column. We can use the $ sign to select the column from a data frame.

```
# Select the column ID

df$ID
```

Subset a Data Frame

n the previous section, we selected an entire column without condition. It is possible to subset based on whether or not a certain condition was true.

We use the subset() function.

```
subset(x, condition)
```

```
arguments:
```

```
- x: data frame used to perform the subset
```

```
- condition: define the conditional statement
```

We want to return only the items with price above 10, we can do:

```
# Select price above 5

subset(df, subset = price > 5)
```