Factors

Here we are going to learn about R factors: how to create them and where they are used.

We'll cover the following ∧

- Creating Factors
- Accessing and Manipulating Factors

A **Factor** is an interesting data structure in R language used to categorize data. By categorizing data, we mean fields that take only predefined, a limited, or finite number of values **(categorical variables)**.

For example, the marital status of a person can be one of the following:
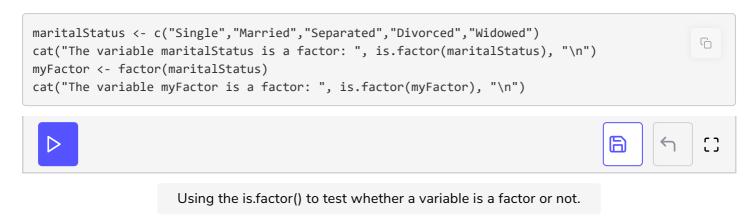
- Single
- Married
- Separated
- Divorced
- Widowed

Here we know that the possible values for marital status are 5. These values are predefined and distinct and are called `levels`.

Creating Factors

Factors can be created using the `factor()` function. This function takes all the `levels` in the form of a **vector**. Let's dive right into the code.
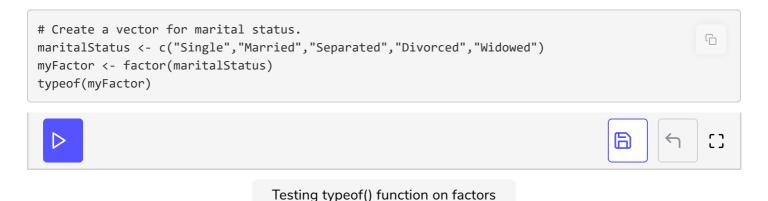
```
# Create a vector for marital status.
maritalStatus <- c("Single","Married","Separated","Divorced","Widowed")
myFactor <- factor(maritalStatus)
print(myFactor)
```

Creating a factor

We can check whether a variable is a factor or not by the function `is.factor()`.

```
maritalStatus <- c("Single","Married","Separated","Divorced","Widowed")
cat("The variable maritalStatus is a factor: ", is.factor(maritalStatus), "\n")
myFactor <- factor(maritalStatus)
cat("The variable myFactor is a factor: ", is.factor(myFactor), "\n")
```

▷

Using the is.factor() to test whether a variable is a factor or not.

Factors are closely related to vectors, i.e., factors are stored as integer vectors.

> R recodes the data in the vector as **integers** and stores the result in an **integer vector**.

We can test this using the `typeof()` function.

```
# Create a vector for marital status.
maritalStatus <- c("Single","Married","Separated","Divorced","Widowed")
myFactor <- factor(maritalStatus)
typeof(myFactor)
```

▷

Testing typeof() function on factors

Accessing and Manipulating Factors

Factors are accessed and manipulated the same way **vectors** are.

```
# Create a vector for marital status.
maritalStatus <- c("Single", "Married", "Separated", "Divorced", "Widowed")
myFactor <- factor(maritalStatus)
print(myFactor[1])
```

▷

Accessing a factor

Let's take a small quiz on **R Objects** in the next lesson.