**Data Types**

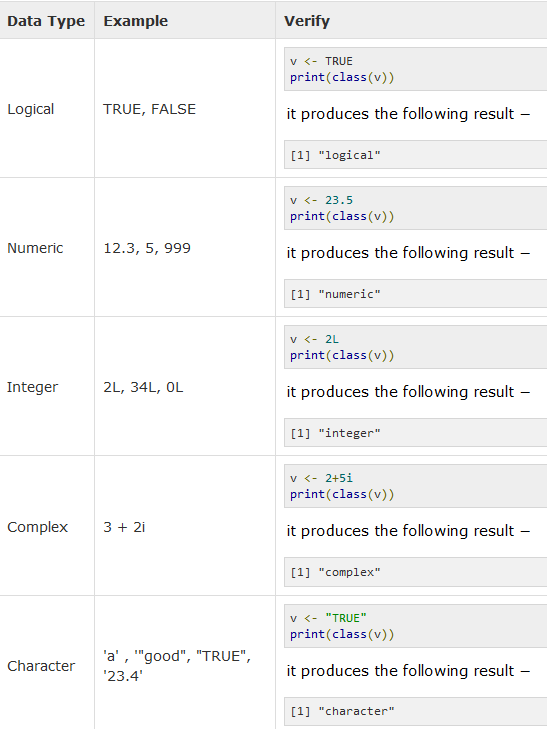
**Introduction:**

Generally, while doing programming in any programming language, you need to use various data types (variables) to store various information. Variables are nothing but reserved memory locations to store values. This means that, when you create a variable you reserve some space in memory.

You may like to store information of various data types like character, wide character, integer, floating point, double floating point, Boolean etc. Based on the data type of a variable, the operating system allocates memory and decides what can be stored in the reserved memory.

In contrast to other programming languages like C and java in R, the variables are not declared as some data type. The variables are assigned with R-Objects and the data type of the R-object becomes the data type of the variable. There are many types of R-objects. The frequently used ones are −

* Vectors
* Lists
* Matrices
* Arrays
* Factors
* Data Frames



In R programming, the very basic data types are the R-objects called **vectors** which hold elements of different classes as shown above. Please note in R the number of classes is not confined to only the above six types.

**Data Frames:**

R works on data frames. Data frames are tabular data objects. Unlike a matrix in data frame each column can contain different modes of data. This is similar to SAS and SPSS datasets. The first column can be numeric while the second column can be character and third column can be logical. It is a list of vectors of equal length.

Data Frames are created using the data.frame () function.

#Create the data frame.

BMI = data.frame(gender = c("Male", "Male", "Female"),

height = c(152, 171.5, 165),

weight = c(81,93, 78),

Age = c(42,38,26))

print(BMI)

At the broader level there are two types of variables

1. Categorical variables
   1. Ordinal Ex: Low, medium, high
   2. Nominal Ex: Tiger, Lion, Elephant
2. Continuous variables
   1. Numeric variables Ex: 4, 5, 3.2

Sometime we need to convert categorical variable into numerical variable for model building. Ex: neural network algorithm takes input as numeric variables

**Convert Numeric variables into categorical variables:**

There are two methods to convert numeric variables into categorical

1. Equal frequency (Number of samples in each bin is equal)
   * Let us say, we have the data for price: 4, 8, 9, 15, 21, 21, 24, 25, 26, 28, 29, 34
   * If we partition into equal-frequency (equi-depth) bins we get the following three bins
     + Bin 1: 4, 8, 9, 15
     + Bin 2: 21, 21, 24, 25
     + Bin 3: 26, 28, 29, 34
2. Equal width (Interval is same (good for uniform distributions))
   * If we partition into equal width then we get
     + Bin 1: 4, 8, 9
     + Bin 2: 15, 21, 21, 24
     + Bin 3: 25, 26, 28, 29, 34

**Additional Information:**

While performing ETL job for data extraction and loading data into R, data types may change. First thing we need to do after loading data into Analytical tool is to check the variable types and convert it accordingly.

**Interview Questions:**

1. Which function can be used to create vectors of objects by concatenating things together?
2. What would be the result of following code ?

x = vector("numeric", length = 10)

This will be more like programming questions. Please refer to R codes