1. **Introduction to Factors**

Factor is a data structure used for fields that takes only a predefined, finite number of values (categorical data). For example, a data field such as marital status may contain only values from single, married, separated, divorced, or widowed. In such a case, we know the possible values beforehand and these predefined, distinct values are called levels. They are useful in data analysis for statistical modelling.

Factors are created using the factor () function by taking a vector as input.

***x <-factor (c("single", "married", "married", "single"));***

***x***

We can check if a variable is a factor or not using class() function.

***class(x)***

Similarly, levels of a factor can be checked using the levels() function

***levels(x)***

**x = factor(c("single", "married", "married", "single"), levels = c("single", "married", "divorced","seperated"));**

**x**

We can see from the above example that levels may be predefined even if not used. Factors are closely related with vectors. In fact, factors are stored as integer vectors. This is clearly seen from its structure

***x <-factor(c("single","married","married","single"))***

***str(x)***

Accesing Factor Components

To access the third element

***x[3]***

***we can access the factors by integer vvector orlogical vector***

***x[c(2,4)]***

***x[-1]***

**Modifying Factors**

***x[2]="divorced"***

***x***

***x[1]="Not\_married"***

**1.1 Factors in Data Frame**

On creating any data frame with a column of text data, R treats the text column as categorical data and creates factors on it.

height <- c(132,151,162,139,166,147,122)

weight <- c(48,49,66,53,67,52,40)

gender <- c("male","male","female","female","male","female","male")

input\_data <- data.frame(height,weight,gender)

print(input\_data)

print(is.factor(input\_data$gender))

print(input\_data$gender)

**GETTING DATA IN AND OUT OF R**

**CSV**

**To create a csv, use notepad/excel and use the following command**

id,name,salary,start\_date,dept

1,Rick,623.3,2012-01-01,IT

2,Dan,515.2,2013-09-23,Operations

3,Michelle,611,2014-11-15,IT

4,Ryan,729,2014-05-11,HR

5,Gary,843.25,2015-03-27,Finance

6,Nina,578,2013-05-21,IT

7,Simon,632.8,2013-07-30,Operations

8,Guru,722.5,2014-06-17,Finance

**Data=read.csv(“employee.csv)**

To get the max salary from the employee dataset

**max\_salary=max(data$salary)**

To get the details of person having max salary

**detail = subset(data,salary == max(salary))**