

R Basics

Factors

Example 1

```
v <- c(1,3,5,8,2,1,3,5,3,5)
is.factor(v)
```

```
## [1] FALSE
```

```
factor(v)
```

```
## [1] 1 3 5 8 2 1 3 5 3 5
## Levels: 1 2 3 5 8
```

```
x <- factor(v)
```

Example 2

```
data <-
c("East","West","East","North","North","East","West","West","West","Ea
st","North")
# Then print out this vector
data
```

```
## [1] "East" "West" "East" "North" "North" "East" "West" "West"
## [9] "West" "Ea\nst" "North"
```

```
# Now, check whether this is a factor
is.factor(data)
```

```
## [1] FALSE
```

```
# Then, apply the factor function to create a factor from the vector
factor_data <- factor(data)
# Then see our newly created factor
factor_data
```

```
## [1] East West East North North East West West West Ea\nst
## [11] North
## Levels: Ea\nst East North West
```

```
# Check whether this is a factor
is.factor(factor_data)
```

```
## [1] TRUE
```

Example 3

```
#sex <- factor(c("male", "female", "female", "male"))
# Determining the levels
#levels(sex)
# Then checking the number of levels using nlevels()
#nlevels()
```

Example 4

```
food <- factor(c("low", "high", "medium", "high", "low", "medium",
"high"))
# then print out levels of food
levels(food)
```

```
## [1] "high" "low" "medium"
```

```
nlevels(food)
```

```
## [1] 3
```

Dataframes

A data frame is used for storing data tables. Unlike a matrix in data frame each column can contain different modes of data.

Example 1

```
BMI <- data.frame(
  gender = c("Male", "Male", "Female"),
  height = c(152, 171.5, 165),
  weight = c(81, 93, 78),
  Age = c(42, 38, 26)
)
# Then print it out below
BMI
```

```
##   gender height weight Age
## 1   Male  152.0     81  42
## 2   Male  171.5     93  38
## 3 Female  165.0     78  26
```

Example 2

```
family_data <- data.frame(
  Names = c("Jane", "Pauline", "Daniel", "Mildred", "Cyprian"),
  Age = c(39, 37, 35, 33, 31),
  Gender = c("Female", "Female", "Male", "Female", "Male"),
  occupation = c("Doctor", "Teacher", "Nurse", "Tutor", "Instructor"),
  Residence = c("Miami", "Shianda", "Nairobi", "Yala", "Kiambu")
)
# Print the output
family_data
```

```
##      Names Age Gender occupation Residence
## 1    Jane  39 Female      Doctor      Miami
## 2 Pauline  37 Female      Teacher    Shianda
## 3 Daniel  35   Male        Nurse    Nairobi
## 4 Mildred 33 Female        Tutor      Yala
## 5 Cyprian 31   Male Instructor    Kiambu
```

Selecting Elements from Data Frame

```
# selecting row 1
family_data[1,]
```

```
##      Names Age Gender occupation Residence
## 1   Jane  39 Female      Doctor      Miami
```

```
# selecting rows 1 to 2
family_data[1:2, ]
```

```
##      Names Age Gender occupation Residence
## 1    Jane  39 Female      Doctor      Miami
## 2 Pauline  37 Female      Teacher    Shianda
```

```
# selecting column 1
family_data[,1]
```

```
## [1] "Jane"      "Pauline" "Daniel"  "Mildred" "Cyprian"
```

```
# selecting column 1 to 2
family_data[,1:2 ]
```

```
##      Names Age
## 1    Jane  39
## 2 Pauline  37
## 3 Daniel  35
## 4 Mildred 33
## 5 Cyprian 31
```

```
# selecting row 1 in column 2
family_data[1,2]
```

```
## [1] 39
```

Sorting

```
# Sort in ascending order by gender
sorted_by_gender <- BMI[order(BMI$height),]
# Print out sorted_by_gender below
# ---
#
sorted_by_gender
```

```
##   gender height weight Age
## 1   Male  152.0     81  42
## 3 Female  165.0     78  26
## 2   Male  171.5     93  38
```

```
# Sort in descending order by weight
# ---
#
sorted_by_weight <- BMI[order(-BMI$weight),]
# Print out sorted_by_weight below
# ---
#
sorted_by_weight
```

```
##   gender height weight Age
## 2   Male  171.5     93  38
## 1   Male  152.0     81  42
## 3 Female  165.0     78  26
```

Data Table

Creating Data Table

```
# Create a data table DT
# ---
library(data.table)
#
DT = data.table(
  ID = c("b", "b", "b", "a", "a", "c"),
  a = 1:6,
  b = 7:12,
  c = 13:18
)
DT
```

```
##   ID a  b  c
## 1:  b 1  7 13
## 2:  b 2  8 14
## 3:  b 3  9 15
## 4:  a 4 10 16
## 5:  a 5 11 17
## 6:  c 6 12 18
```

```
# selecting
# Selecting Row 1
DT[1,]
```

```
##   ID a  b  c
## 1:  b 1  7 13
```

```
# Selecting Rows 1 to 2  
DT[1:2,]
```

```
##      ID a  b  c  
## 1:   b 1  7 13  
## 2:   b 2  8 14
```

```
# Find out what happens when we print out the following statement  
DT[,1]
```

```
##      ID  
## 1:   b  
## 2:   b  
## 3:   b  
## 4:   a  
## 5:   a  
## 6:   c
```

```
# Find out what happens when we print out the following statement  
DT[,1:2]
```

```
##      ID a  
## 1:   b 1  
## 2:   b 2  
## 3:   b 3  
## 4:   a 4  
## 5:   a 5  
## 6:   c 6
```

```
# And lastly find out what happens when we print out the following  
#statement  
DT[1,2]
```

```
##      a  
## 1:  1
```

Sorting

```
# Performing the sort  
#  
sorted_by_c <- DT[order(DT$c),]  
# Printing out sorted_by_c  
# ---  
sorted_by_c
```

```
##      ID a  b  c  
## 1:   b 1  7 13  
## 2:   b 2  8 14  
## 3:   b 3  9 15  
## 4:   a 4 10 16  
## 5:   a 5 11 17  
## 6:   c 6 12 18
```

```
#
# Sort in descending order by b, uncommenting the line below
# ---
#
sorted_by_b <- DT[order(-DT$b),]
# Finally printing out sorted_by_b below
sorted_by_b
```

```
##      ID a  b  c
## 1:   c 6 12 18
## 2:   a 5 11 17
## 3:   a 4 10 16
## 4:   b 3  9 15
## 5:   b 2  8 14
## 6:   b 1  7 13
```

Tibble

Creating Tibble

```
# First, we load the tibble package
#library(tibble)
# Then create our tibble tb
#tb <- tibble(
#  x = 1:5,
#  y = 1,
#  z = x ^ 2 + y
#)
# And finally print the created tibble
#tb
```