

**EXAMPLE 29 : WHAT IS A FACTOR?**

LET'S TAKE THE PROBLEM  
OF **SALES DATA ANALYSIS**  
YOU WORK FOR **A CHAIN**  
OF **RETAIL STORES**

YOU WOULD LIKE TO ANSWER  
A FEW QUESTIONS ABOUT  
HOW THE STORES ARE DOING

WHAT ARE THE **TOP**  
**SELLING PRODUCT**  
**CATEGORIES?**

WHAT ARE THE  
**SALES IN EACH**  
**CITY?**

**FACTORS** ARE REALLY USEFUL FOR ANY  
SUCH DATA ANALYSIS IN R

# EXAMPLE 29 : WHAT IS A FACTOR?

HERE IS SOME SALES DATA:

Bangalore	Clothing	INR 5000
New Delhi	Footwear	INR 4500
Mumbai	Cosmetics	INR 3500
Bangalore	Cosmetics	INR 2500
Bangalore	Footwear	INR 1000
Mumbai	Clothing	INR 2000
New Delhi	Clothing	INR 5500

WE CAN REPRESENT THIS DATA USING 3 VECTORS

```
city <- c("Bangalore", "New Delhi", "Mumbai", "Bangalore",  
          "Bangalore", "Mumbai", "New Delhi")
```

```
category <- c("Clothing", "Footwear", "Cosmetics",  
              "Cosmetics", "Footwear", "Clothing", "Clothing")
```

```
saleAmount <- c(5000, 4500, 3500, 2500, 1000, 2000, 5500)
```

THE CITY AND CATEGORY VARIABLES  
HERE HAVE AN INTERESTING  
CHARACTERISTIC

THEY TAKE ONE OF A LIMITED  
SET OF VALUES

# EXAMPLE 29 : WHAT IS A FACTOR?

HERE IS SOME SALES DATA:

Bangalore	Clothing	INR 5000
New Delhi	Footwear	INR 4500
Mumbai	Cosmetics	INR 3500
Bangalore	Cosmetics	INR 2500
Bangalore	Footwear	INR 1000
Mumbai	Clothing	INR 2000
New Delhi	Clothing	INR 5500

WE CAN REPRESENT THIS DATA USING 3 VECTORS

```
city <- c("Bangalore", "New Delhi", "Mumbai", "Bangalore",  
         "Bangalore", "Mumbai", "New Delhi")
```

```
category <- c("Clothing", "Footwear", "Cosmetics",  
             "Cosmetics", "Footwear", "Clothing", "Clothing")
```

```
saleAmount <- c(5000, 4500, 3500, 2500, 1000, 2000, 5500)
```

THE CITY AND CATEGORY VARIABLES HERE HAVE AN  
INTERESTING CHARACTERISTIC

THEY TAKE ONE OF A LIMITED SET OF VALUES

SUCH VARIABLES ARE CALLED  
**CATEGORICAL VARIABLES**



# EXAMPLE 29 : WHAT IS A FACTOR?

HERE IS SOME SALES DATA:

Bangalore
New Delhi
Mumbai
Bangalore
Bangalore
Mumbai
New Delhi

```
city <- c("Bangalore", "New Delhi", "Mumbai", "Bangalore",  
          "Bangalore", "Mumbai", "New Delhi")
```

THE CITY VARIABLE TAKES ONE  
OF **A LIMITED SET OF VALUES**  
**LEVELS**

**BANGALORE**  
**NEW DELHI**  
**MUMBAI**

**A FACTOR VECTOR IS A SPECIAL  
KIND OF VECTOR THAT IS**

**1) AWARE THAT THE VARIABLE  
TAKES A LIMITED SET OF VALUES**

**2) KNOWS WHAT THAT SET OF  
VALUES CONTAINS**

# EXAMPLE 29 : WHAT IS A FACTOR?

A **FACTOR VECTOR** IS A SPECIAL  
KIND OF VECTOR THAT IS

1) **AWARE** THAT THE VARIABLE  
TAKES A LIMITED SET OF VALUES

2) KNOWS **WHAT** THAT SET OF  
VALUES CONTAINS

# EXAMPLE 29 : WHAT IS A FACTOR?

HERE IS SOME SALES DATA:

Bangalore
New Delhi
Mumbai
Bangalore
Bangalore
Mumbai
New Delhi

```
city <- c("Bangalore", "New Delhi", "Mumbai", "Bangalore",  
          "Bangalore", "Mumbai", "New Delhi")
```

```
cityAsFactor <- factor(city)  
print(cityAsFactor)
```

```
[1] Bangalore New Delhi Mumbai Bangalore Bangalore Mumbai New Delhi  
Levels: Bangalore Mumbai New Delhi
```

THE **FACTOR()** FUNCTION CONVERTS  
THE CITY VECTOR TO A **FACTOR VECTOR**

THIS NEW VECTOR IS THE SAME AS THE OLD  
ONE, EXCEPT IT IS **AWARE** THAT IT HAS **LEVELS**

# EXAMPLE 29 : WHAT IS A FACTOR?

HERE IS SOME SALES DATA:

1	Bangalore
3	New Delhi
2	Mumbai
1	Bangalore
1	Bangalore
2	Mumbai
3	New Delhi

```
city <- c("Bangalore", "New Delhi", "Mumbai", "Bangalore",  
          "Bangalore", "Mumbai", "New Delhi")
```

```
cityAsFactor <- factor(city)  
print(cityAsFactor)
```

```
[1] Bangalore New Delhi Mumbai    Bangalore Bangalore Mumbai    New Delhi  
Levels: Bangalore Mumbai New Delhi
```

**INTERNALLY A FACTOR VECTOR STORES THE DATA BY  
MAPPING EACH LEVEL TO AN INTEGER**

**MAPPING EACH VALUE OF THE VECTOR TO THE  
CORRESPONDING INTEGER**



Bangalore	1
Mumbai	2
New Delhi	3



# EXAMPLE 29 : WHAT IS A FACTOR?

HERE IS SOME SALES DATA:

1	Bangalore
3	New Delhi
2	Mumbai
1	Bangalore
1	Bangalore
2	Mumbai
3	New Delhi

**LEVELS**

Bangalore	1
Mumbai	2
New Delhi	3

```
city <- c("Bangalore", "New Delhi", "Mumbai", "Bangalore",  
          "Bangalore", "Mumbai", "New Delhi")  
cityAsFactor <- factor(city)  
print(cityAsFactor)
```

```
[1] Bangalore New Delhi Mumbai Bangalore Bangalore Mumbai New Delhi  
Levels: Bangalore Mumbai New Delhi
```

```
as.numeric(cityAsFactor)
```

```
[1] 1 3 2 1 1 2 3
```

**IF YOU CONVERT A FACTOR TO NUMERIC YOU'LL BE ABLE TO SEE THE INTERNAL INTEGER MAPPING**

**EXAMPLE 30 : FIND THE DISTINCT  
SET OF VALUES IN A FACTOR**

FIND THE **DISTINCT SET OF CITIES**  
IN WHICH SALES OCCUR

**SALES DATA**

Bangalore	Clothing	INR 5000
New Delhi	Footwear	INR 4500
Mumbai	Cosmetics	INR 3500
Bangalore	Cosmetics	INR 2500
Bangalore	Footwear	INR 1000
Mumbai	Clothing	INR 2000
New Delhi	Clothing	INR 5500

```
city <- c("Bangalore", "New Delhi", "Mumbai", "Bangalore",  
         "Bangalore", "Mumbai", "New Delhi")  
cityAsFactor <- factor(city)  
print(cityAsFactor)
```

```
[1] Bangalore New Delhi Mumbai    Bangalore Bangalore Mumbai    New Delhi  
Levels: Bangalore Mumbai New Delhi
```

```
levels(cityAsFactor)
```

```
[1] "Bangalore" "Mumbai"     "New Delhi"
```

**1. FIND THE DISTINCT SET OF  
CITIES IN WHICH SALES OCCUR**



# EXAMPLE 31: REPLACE THE LEVELS OF A FACTOR

REPLACE ALL THE CITY NAMES  
WITH SHORT NAMES

SALES DATA

Bangalore	Clothing	INR 5000
New Delhi	Footwear	INR 4500
Mumbai	Cosmetics	INR 3500
Bangalore	Cosmetics	INR 2500
Bangalore	Footwear	INR 1000
Mumbai	Clothing	INR 2000
New Delhi	Clothing	INR 5500

```
city <- c("Bangalore", "New Delhi", "Mumbai", "Bangalore",  
         "Bangalore", "Mumbai", "New Delhi")  
cityAsFactor <- factor(city)  
print(cityAsFactor)  
[1] Bangalore New Delhi Mumbai Bangalore Bangalore Mumbai New Delhi  
Levels: Bangalore Mumbai New Delhi  
[1] "Bangalore" "Mumbai" "New Delhi"
```

```
levels(cityAsFactor) <- c("BLR", "MUM", "DEL")  
print(cityAsFactor)  
[1] BLR DEL MUM BLR BLR MUM DEL  
Levels: BLR MUM DEL
```

**YOU CAN JUST REPLACE THE  
LEVELS OF A FACTOR VECTOR**

```
cityCodeFactor <- factor(cityAsFactor, labels=c("B", "M", "D"))  
print(cityCodeFactor)
```

```
[1] B D M B B M D  
Levels: B M D
```

**AN ALTERNATIVE IS TO  
CREATE A NEW FACTOR  
FROM THE OLD ONE**



## EXAMPLE 32 : TABLE() FUNCTION

- FIND THE COUNTS FOR EACH LEVEL IN A FACTOR VECTOR

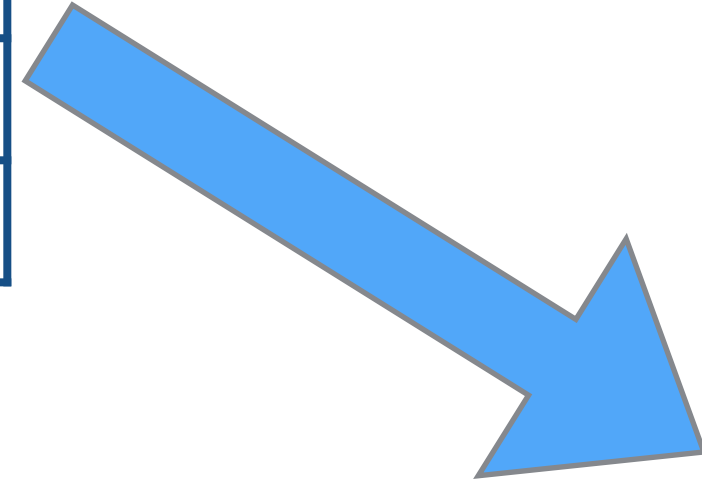
HERE IS SOME SALES DATA:

Bangalore	Clothing	INR 5000
New Delhi	Footwear	INR 4500
Mumbai	Cosmetics	INR 3500
Bangalore	Cosmetics	INR 2500
Bangalore	Footwear	INR 1000
Mumbai	Clothing	INR 2000
New Delhi	Clothing	INR 5500

WHAT ARE THE NUMBER OF  
TRANSACTIONS IN EACH CITY?

Bangalore	Clothing	INR 5000
New Delhi	Footwear	INR 4500
Mumbai	Cosmetics	INR 3500
Bangalore	Cosmetics	INR 2500
Bangalore	Footwear	INR 1000
Mumbai	Clothing	INR 2000
New Delhi	Clothing	INR 5500

WE CAN REPRESENT THIS DATA  
USING 3 VECTORS



```
city <- factor(c("Bangalore", "New Delhi", "Mumbai",  
                 "Bangalore", "Bangalore", "Mumbai", "New Delhi"))
```

Bangalore	Clothing	INR 5000
New Delhi	Footwear	INR 4500
Mumbai	Cosmetics	INR 3500
Bangalore	Cosmetics	INR 2500
Bangalore	Footwear	INR 1000
Mumbai	Clothing	INR 2000
New Delhi	Clothing	INR 5500

WE CAN REPRESENT THIS DATA  
USING 3 VECTORS

```
city <- factor(c("Bangalore", "New Delhi", "Mumbai",  
                "Bangalore", "Bangalore", "Mumbai", "New Delhi"))
```

```
category <- factor(c("Clothing", "Footwear", "Cosmetics",  
                    "Cosmetics", "Footwear", "Clothing", "Clothing"))
```



Bangalore	Clothing	INR 5000
New Delhi	Footwear	INR 4500
Mumbai	Cosmetics	INR 3500
Bangalore	Cosmetics	INR 2500
Bangalore	Footwear	INR 1000
Mumbai	Clothing	INR 2000
New Delhi	Clothing	INR 5500

WE CAN REPRESENT THIS DATA  
USING **3 VECTORS**

```
city <- factor(c("Bangalore", "New Delhi", "Mumbai",
                 "Bangalore", "Bangalore", "Mumbai", "New Delhi"))
category <- factor(c("Clothing", "Footwear", "Cosmetics",
                     "Cosmetics", "Footwear", "Clothing", "Clothing"))
```

```
saleAmount <- c(5000, 4500, 3500, 2500, 1000, 2000, 5500)
```

```
city <- factor(c("Bangalore", "New Delhi", "Mumbai",  
               "Bangalore", "Bangalore", "Mumbai", "New Delhi"))  
category <- factor(c("Clothing", "Footwear", "Cosmetics",  
                    "Cosmetics", "Footwear", "Clothing", "Clothing"))  
saleAmount <- c(5000, 4500, 3500, 2500, 1000, 2000, 5500)
```

**NOTE: THE CITY AND CATEGORY  
VECTORS ARE FACTORS**

```
city <- factor(c("Bangalore", "New Delhi", "Mumbai",  
               "Bangalore", "Bangalore", "Mumbai", "New Delhi"))  
category <- factor(c("Clothing", "Footwear", "Cosmetics",  
                    "Cosmetics", "Footwear", "Clothing", "Clothing"))  
saleAmount <- c(5000, 4500, 3500, 2500, 1000, 2000, 5500)
```

1) WHAT ARE THE NUMBER OF TRANSACTIONS IN EACH CITY?

```
table(city)
```

city	
Bangalore	3
Mumbai	2
New Delhi	2

**TABLE() WILL TAKE A FACTOR AND GIVE US THE COUNT FOR EACH LEVEL IN THE FACTOR**

# EXAMPLE 33 : TAPPLY() FUNCTION



# `lapply()`

APPLY A FUNCTION ON ANOTHER  
VECTOR BASED ON VALUES IN THE  
FACTOR VECTOR

THIS IS SIMILAR TO **A PIVOT**  
**TABLE IN EXCEL** OR TO **A GROUPING IN SQL**

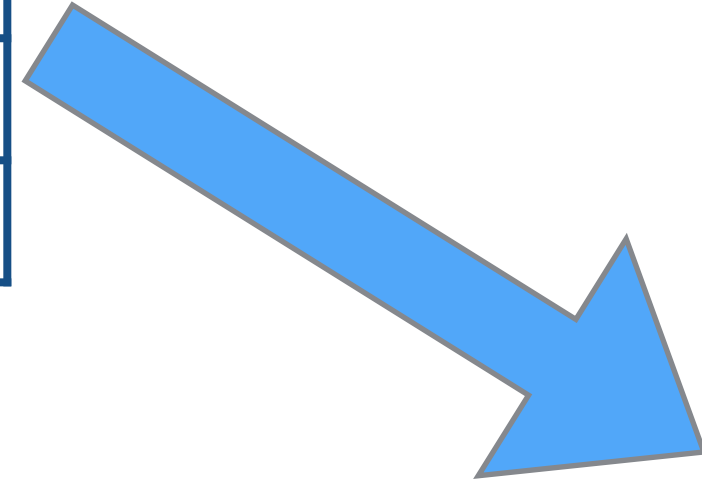
HERE IS **SOME SALES DATA:**

Bangalore	Clothing	INR 5000
New Delhi	Footwear	INR 4500
Mumbai	Cosmetics	INR 3500
Bangalore	Cosmetics	INR 2500
Bangalore	Footwear	INR 1000
Mumbai	Clothing	INR 2000
New Delhi	Clothing	INR 5500

WHAT IS THE **SUM OF SALES** IN EACH  
**CATEGORY?**

Bangalore	Clothing	INR 5000
New Delhi	Footwear	INR 4500
Mumbai	Cosmetics	INR 3500
Bangalore	Cosmetics	INR 2500
Bangalore	Footwear	INR 1000
Mumbai	Clothing	INR 2000
New Delhi	Clothing	INR 5500

WE CAN REPRESENT THIS DATA  
USING 3 VECTORS



```
city <- factor(c("Bangalore", "New Delhi", "Mumbai",  
                 "Bangalore", "Bangalore", "Mumbai", "New Delhi"))
```

Bangalore	Clothing	INR 5000
New Delhi	Footwear	INR 4500
Mumbai	Cosmetics	INR 3500
Bangalore	Cosmetics	INR 2500
Bangalore	Footwear	INR 1000
Mumbai	Clothing	INR 2000
New Delhi	Clothing	INR 5500

WE CAN REPRESENT THIS DATA  
USING 3 VECTORS

```
city <- factor(c("Bangalore", "New Delhi", "Mumbai",  
                 "Bangalore", "Bangalore", "Mumbai", "New Delhi"))
```

```
category <- factor(c("Clothing", "Footwear", "Cosmetics",  
                     "Cosmetics", "Footwear", "Clothing", "Clothing"))
```

Bangalore	Clothing	INR 5000
New Delhi	Footwear	INR 4500
Mumbai	Cosmetics	INR 3500
Bangalore	Cosmetics	INR 2500
Bangalore	Footwear	INR 1000
Mumbai	Clothing	INR 2000
New Delhi	Clothing	INR 5500

WE CAN REPRESENT THIS DATA  
USING 3 VECTORS

```
city <- factor(c("Bangalore", "New Delhi", "Mumbai",  
                "Bangalore", "Bangalore", "Mumbai", "New Delhi"))  
category <- factor(c("Clothing", "Footwear", "Cosmetics",  
                    "Cosmetics", "Footwear", "Clothing", "Clothing"))
```

```
saleAmount <- c(5000, 4500, 3500, 2500, 1000, 2000, 5500)
```

```
city <- factor(c("Bangalore", "New Delhi", "Mumbai",  
               "Bangalore", "Bangalore", "Mumbai", "New Delhi"))  
category <- factor(c("Clothing", "Footwear", "Cosmetics",  
                    "Cosmetics", "Footwear", "Clothing", "Clothing"))  
saleAmount <- c(5000, 4500, 3500, 2500, 1000, 2000, 5500)
```

**NOTE: THE CITY AND CATEGORY  
VECTORS ARE FACTORS**



```
city <- factor(c("Bangalore", "New Delhi", "Mumbai",  
               "Bangalore", "Bangalore", "Mumbai", "New Delhi"))  
category <- factor(c("Clothing", "Footwear", "Cosmetics",  
                    "Cosmetics", "Footwear", "Clothing", "Clothing"))  
saleAmount <- c(5000, 4500, 3500, 2500, 1000, 2000, 5500)
```

```
tapply(saleAmount, category, sum)
```

Clothing	Cosmetics	Footwear
12500	6000	5500

## HOW DOES THIS WORK?

```
tdapply(saleAmount, category, sum)
```

**TAKE EACH LEVEL IN THE CATEGORY FACTOR AND  
FIND THE SUM OF CORRESPONDING VALUES IN  
SALEAMOUNT**

`apply(saleAmount, category, sum)`

TAKE EACH LEVEL IN THE CATEGORY FACTOR AND FIND THE SUM OF CORRESPONDING VALUES IN SALEAMOUNT

INR 5000
INR 4500
INR 3500
INR 2500
INR 1000
INR 2000
INR 5500

Clothing
Footwear
Cosmetics
Cosmetics
Footwear
Clothing
Clothing

LEVELS

Clothing
Cosmetics
Footwear

`tapply(saleAmount, category, sum)`

+	INR 5000		Clothing
	INR 4500		Footwear
	INR 3500		Cosmetics
	INR 2500		Cosmetics
	INR 1000		Footwear
+	INR 2000		Clothing
+	INR 5500		Clothing

**LEVELS**

Clothing
Cosmetics
Footwear

```
tapply(saleAmount, category, sum)
```

	INR 12500		Clothing
	INR 4500		Footwear
+	INR 3500		Cosmetics
+	INR 2500		Cosmetics
	INR 1000		Footwear

**LEVELS**

Clothing
Cosmetics
Footwear

```
tapply(saleAmount, category, sum)
```

	INR 12500		Clothing
	INR 6000		Cosmetics
+	INR 4500		Footwear
+	INR 1000		Footwear

**LEVELS**

Clothing
Cosmetics
Footwear

EXAMPLE 18 : SUMMARIZING DATA WITH FACTORS



```
tapply(saleAmount, category, sum)
```

INR 12500	Clothing
INR 6000	Cosmetics
INR 5500	Footwear

## LEVELS

Clothing
Cosmetics
Footwear

EXAMPLE 18 : SUMMARIZING DATA WITH FACTORS

```
tapply(saleAmount, category, sum)
```

Clothing	Cosmetics	Footwear
12500	6000	5500

INR 12500	Clothing
INR 6000	Cosmetics
INR 5500	Footwear

EXAMPLE 18 : SUMMARIZING DATA WITH FACTORS

```
tapply(saleAmount, category, sum)
```

Clothing	Cosmetics	Footwear
12500	6000	5500

**YOU CAN USE  
ANY FUNCTION  
YOU LIKE  
INSTEAD OF SUM  
  
EVEN FUNCTIONS  
YOU HAVE DEFINED**

EXAMPLE 18 : SUMMARIZING DATA WITH FACTORS