

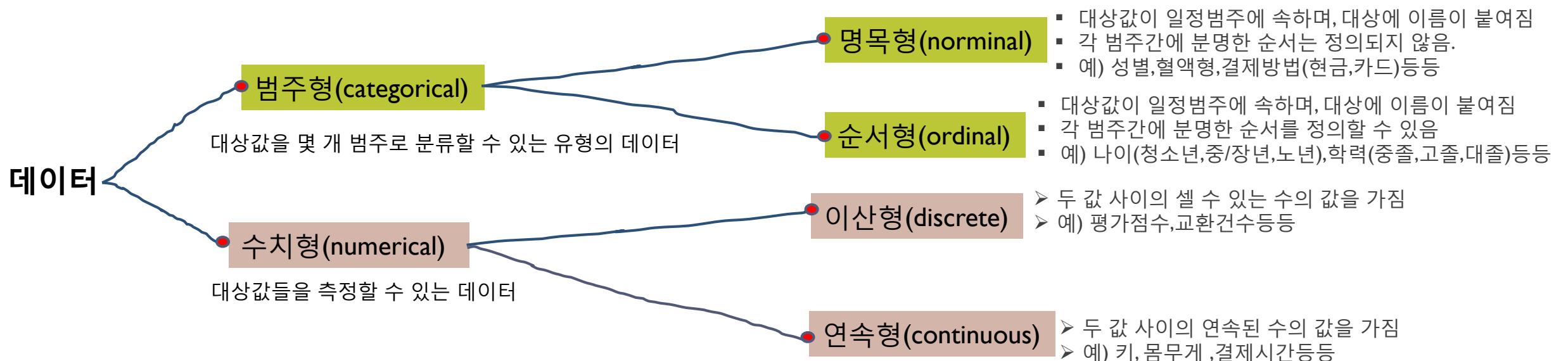
# 1. 데이터 분류



데이터의 분류 및 특징 이해

## I. 데이터 분류

- 데이터 – 어떤 결과를 예측하거나 추론해나갈 때 가장 기본이 되는 재료



## Factor

- 범주형 데이터를 저장하기 위한 특화된 벡터형태 타입
- factor 타입을 생성 → factor() 함수 이용
- 형식

factor(

x, ← 팩터로 표현하고자 하는 값

levels, <- 팩터 레벨의 목록을 보여줌

ordered ← TRUE(순서형 데이터), FALSE(명목형 데이터)

)

```
> work<-c("주간","야간","주간","주간","야간","야간")  
> work  
[1] "주간" "야간" "주간" "주간" "야간" "야간"
```

```
> f_work<-factor(work)
```

```
> f_work
```

```
[1] 주간 야간 주간 주간 야간
```

```
Levels: 야간 주간
```

```
> f_work2<-factor(work, levels=c("주간", "야간"))
```

```
> f_work2
```

```
[1] 주간 야간 주간 주간 야간 야간
```

```
Levels: 주간 야간
```

```
> table(f_work2)
```

```
f_work2
```

```
주간 야간
```

```
3 3
```

```
> |
```

레벨의 목록을 사용자가 원하는대로 설정할 수 있어요.

각 범주에 속한 데이터 값을 카운트하여 보여주는 함수

```
> s_data<-read.csv("d:\\r1\\subway_data.csv",header=T,col.names=c("lines","station","totusg","avgusg"))
> table(s_data$lines)
```

1호선	2호선	3호선	4호선
10	49	33	25
.	.	.	.

수치 데이터 요약 통계- summary() 함수 이용

```
> myage<-c(23,28,31,27,36,37,32,25)
> summary(myage)
  Min. 1st Qu. Median      Mean 3rd Qu.      Max.
 23.00    26.50   29.50    29.88   33.00    37.00
> |
```