3. { Data 전처리(2) }

(R Study group 자료)

+) 지난 내용 요약

filter : 원하는 값(조건에 부합하는 값)을 추려냄
select : 원하는 변수만 선택

%>% : 파이프 연산자 “and than” 의 개념
코드를 더 간결하게 사용 가능

%in% : 조건들 중 하나라도 부합되면 (or의 개념)

+) 지난 내용 요약

데이터 안에서 원하는 변수와, 원하는 값들만 추려는 방법
(select, filter)

```
titanic_2 <- titanic %>%
                select(Name, Survived, Fare) %>%
                filter(Survived == 1 & Fare >= 100 )

mpg_2 <- mpg %>%
          select(year, manufacturer, model, cty) %>%
          filter(manufacturer %in% c( “audi”, “chevrolet”,
                                      “ford”) & cty >= 20)
```

+) 복습

titanic 데이터에서 Name, Survived, Embarked 만 추려내고,
그 중 생존자(Survived 가 1)만 뽑아 titanic_2에 저장하시요.

+) 복습

titanic 데이터에서 Name, Survived, Fare 만 추려내고,
그 중 Fare가 100 이상이면서 사망자인(Survived 가 0)값을 뽑아
titanic_3에 저장하시요.

+) 복습

titanic 데이터에서 Name, Survived, Sex 만 추려내고,
그 중 남성 사망자를 출력하시오.

+) 복습

titanic 데이터에서 Name, Survived, Sex, Age 만 추려내고,
그 중 사망자의 나이가 30세 미만인 경우를 출력하시요.

1) 특정 기준으로 요약하기

특정한 기준으로 데이터를 요약해서 보고 싶은 경우가 대부분.
ex) titanic data에서 생존자(특정기준)가 몇 명인지(요약) ?

```
titanic  %>%
    group_by(Survived) %>%  # Survived 기준으로
    summarise(freq = n())  # 빈도로 요약해달라
                           # n() 은 빈도계산
```

기본구조

```
data %>%
    group_by(변수, 변수, …) %>%
    summarise(변수명 = function(),
              변수명 = function(), …)
```

1) 특정 기준으로 요약하기

select, filter 함수들을 활용하면 훨씬 더 효과적이다 !

titanic 데이터에서 Name, Survived, Embarked 만 추려내고,
그 중 생존자(Survived 가 1)의 수를 구하시요.

```
titanic %>%
        select(Name, Survived, Embarked) %>%
        group_by(Embarked) %>%
        filter(Survived == 1) %>%
        summarise(n = n())
```

1) 특정 기준으로 요약하기

select, filter 함수들을 활용하면 훨씬 더 효과적이다 !

titanic 데이터에서 Name, Survived, Pclass 만 추려내고,
그 중 생존자(Survived 가 1)의 수를 구하시요.

```
titanic %>%
        select(Name, Survived, Pclass) %>%
        group_by(Pclass) %>%
        filter(Survived == 1) %>%
        summarise(n = n())
```

1) 특정 기준으로 요약하기

select, filter 함수들을 활용하면 훨씬 더 효과적이다 !
특정 조건은 여러 개 가능하다.

titanic 데이터에서 Name, Survived, Pclass 만 추려내고,
그 중 생존 여부에 따라 빈도를 구하시요.

```
titanic %>%
        select(Name, Survived, Pclass) %>%
        group_by(Pclass, Survived) %>%
        summarise(n = n())
```