



## 12주차 과제

과 목	데이터사이언스프로그래밍
담당 교수	이 두 호
학 번	201720970
학 과	소프트웨어·미디어·산업공학부
이 름	권 대 한

1. GPA 데이터에서 성적이 4.1 이상이면 "A등급" 장학금, 3.8~4.0 사이이면 "B등급" 장학금, 3.5~3.7 사이이면 "C등급" 장학금을 부여한다. 장학금별 장학금액은 아래의 sch 테이블과 같다. join 함수를 이용하여 학생별 장학금액을 출력하고, 성별로 "A등급", "B등급", "C등급" 이 각각 몇명인지 출력하시오.

<실행 코드>

가) sch <- data.frame(scholarship = as.factor(c("A","B","C","N/A")), amount = as.character(c(200,100,50,0)))

나) GPA <- read.csv("GPA.csv")

다) GPA\$scholarship <- 1

라) GPA\$scholarship %<>% as.factor()

마) GPA\$gpa %<>% as.numeric()

바) GPA\$scholarship <- ifelse(GPA\$gpa > 4.0, "A",

사) ifelse(GPA\$gpa >= 3.8, "B",

아) ifelse(GPA\$gpa >= 3.5, "C", NA)))

자) GPA %<>% left\_join(sch, by="scholarship")

차) GPA %>% group\_by(gender, scholarship) %>% summarise(n = n()) -> result

코드 가)에서 장학금 기준 데이터 프레임을 선언하였으며, 나)에서 GPA.csv를 불러옴으로, GPA 데이터 프레임을 선언하였다. 코드 다)에서 join 사용을 위해 scholarship 필드를 선언해주었으며, 라)에서 sch\$scholarship과 같은 타입으로 변환해주었다.

코드 마)에서는 학점 비교를 위해 숫자 자료형으로 바꿔주었으며, 코드 바)에서 실질적인 학점 비교를 통해 A, B, C, NA를 정의해주었으며, 코드 자)에서 해당 등급을 기반으로 장학금 필드를 합쳐준다.

이후 성별, 등급에 따른 사람 수를 출력하기 위해, 코드 차)에서는 group\_by로 그룹화 해주었으며, 해당 그룹 당 등급에 따른 인원 수를 result를 선언함과 동시에 저장한다.