

# DAY 1 ASSIGNMENT

## Introduction to R and Data Manipulation

### Instruction on the Day 1 assignment

---

DAY 1 과제는 크게 두 개 파트로 나누어져 있습니다. 첫 번째 파트는 Base R에 대한 이해를 돕기 위한 것이며, 두 번째 파트는 {tidyverse} 패키지를 활용하여 데이터의 기본적인 전처리를 수행할 수 있는지 확인하기 위한 것입니다. 각각의 문제는 RMarkdown 또는 Quarto를 이용하여 PDF 형태로 제출하여 주시기 바랍니다. 혹은 PDF 컴파일에 성공하지 못하였다면, .Rmd나 .qmd 파일의 형태로 제출하셔도 좋습니다.

### Part I. Base R

---

bfi는 {psych} 패키지에 내장된 데이터셋으로 International Personality Item Pool 에서 25명의 성격에 대한 자기보고 문항에 관한 정보를 제공하고 있습니다. 이 데이터셋은 다음과 같은 변수들을 포함하고 있습니다:

1. gender는 응답자의 성별을 나타냅니다(Males = 1, Females = 2).
2. education은 응답자의 학업성취도를 나타냅니다(1 = HS, 2 = finished HS, 3 = some college, 4 = college graduate 5 = graduate degree).
3. age는 응답자의 연령을 연 단위로 나타냅니다.
4. E3는 외향성(extraversion)을 측정하는 5가지 항목 중 하나입니다(문항: “Know how to captivate people”).

필요한 패키지를 설치하고, 사용할 데이터셋을 불러오세요.

bfi 데이터셋으로부터 gender, education, age, E3 변수를 각각의 벡터로 저장하세요.

각각의 벡터들은 어떤 자료 유형을 가지고 있습니까?

주어진 bfi에서 여성은 대략 어느 정도 존재합니까? 대략적인 비율로 보여주세요.

26세 미만(<) 남성(gender = 1)과 26세 미만(<) 여성(gender = 2) 사이의 E3의 표본 평균의 차이는 얼마입니까? (Hint: 26세 미만 남성의 E3 평균 - 26세 미만 여성의 E3 평균)

bfi 데이터셋에서 16세 남성 중 고등학교 졸업자는 몇 명입니까?

bfi 데이터셋에서 42세 여성은 총 몇 명입니까?

## Part II. {tidyverse}: Mutate, Select, and Filter

---

미국의 비행에 관련된 정보를 담고 있는 {nycflights13} 패키지를 설치하고 flights 데이터를 df라는 객체로 저장하세요.

3월에 이륙한 비행만을 나타내보세요. month 변수는 정수형(integer)으로 1 = 1월 (January), 2 = 2월(February) 등을 나타냅니다. 새로운 객체로 저장하지 말고 RMarkdown 또는 Quarto 문서 작업을 통해 출력하세요.

겨울 시즌(11월부터 2월)에 출발하는 공항이 'JFK'거나 'LGA'인 경우만을 선택하세요.

이 데이터는 모두 2013년에 관한 것입니다. 따라서 불필요한 year 변수를 삭제해보세요.

언더스코어(\_)를 변수명에 포함한 경우, 가장 맨 앞쪽에 오도록 변수들의 순서를 바꾸어보세요.

변수의 유형이 정수형(Integers)인 변수들만을 선택해보세요.

변수의 이름이 arr 또는 dep 로 시작하는 변수들만을 선택해보세요.

도착 지연(arrival delay)과 출발 지연(departure delay)를 더한 변수, total\_delay를 만들고 total delay, arrival delay, departure delay, 그리고 나머지 변수 순으로 재정렬해보세요.

모든 문자형(character) 변수들을 요인형 변수(factor)로 만들어보세요.

비행 거리가 1000보다 짧을 경우를 'short', 1000-2000 사이일 경우 'medium', 2000보다 클 경우를 'long'이라고 하는 값을 갖는 distance\_bins이라는 변수를 만들어보세요, if\_else() 또는 case\_when() 함수를 이용하여 만들어보세요. 만약 그 변수를 만들었다면, short-medium-long 순으로 순서를 갖게끔 요인변수를 조작하세요.

매 달마다 각각의 출발 공항에서의 평균 출발 지연(average departure delay)는 얼마입니까?

월(month) 변수를 이용하여 사계절을 나타내는 변수, season을 만드세요. 겨울은 12월부터 2월, 봄은 3월부터 5월, 여름은 6월부터 8월, 가을은 9월부터 11월입니다.