HW #7

1678226 Chanmi Yoo(유찬미)

# Load library   
library(tidyverse)

## -- Attaching packages ------------------------------- tidyverse 1.2.1 --

## √ ggplot2 3.2.1 √ purrr 0.3.3  
## √ tibble 2.1.3 √ dplyr 0.8.3  
## √ tidyr 1.0.0 √ stringr 1.4.0  
## √ readr 1.3.1 √ forcats 0.4.0

## -- Conflicts ---------------------------------- tidyverse\_conflicts() --  
## x dplyr::filter() masks stats::filter()  
## x dplyr::lag() masks stats::lag()

1. 각 자료를 tibble data로 만드시오.

T1 <- tibble(  
 school = rep(LETTERS[1:3], each=2),  
 grade = rep(c(1,2), times=3),  
 Female = cumsum(rep(10,6)),  
 Male= cumsum(rep(90,6))  
 )  
  
T2 <- tibble(  
 school = rep(LETTERS[1:3], each=4),  
 grade = rep(c(1,2), each=2, times=3),  
 gender = rep(c('Female', 'Male'), 6),  
 n = c(rbind(seq(10, 60, by=10), seq(90,540, by=90)))  
)  
  
T3 <- tibble(  
 school = rep(LETTERS[1:3], each=2),  
 grade = rep(1:2, 3),  
 FM = paste(seq(10, 60, by=10), seq(90,540, by=90), sep='\_')  
)  
  
T41 <- tibble(  
 school = LETTERS[1:3],  
 '1' = seq(10, 50, by=20),  
 '2' = seq(20, 60, by=20)  
)  
  
T42 <- tibble(  
 school = LETTERS[1:3],  
 '1' = seq(90, 450, length.out = 3),  
 '2' = seq(180, 540, length.out = 3)  
)

1. T1은 tidy data인가? (답에 대한 이유도 함께 쓰시오)

#[답] tidy data이다. tidy data는 3가지 규칙이 있다.  
# 1) 각 변수는 각자의 column을 가지고 있어야 하고,   
# 2)각 observation은 각자의 row를 가지고 있어야 하며,  
# 3)각 값은 각자의 cell을 가지고 있어야 한다.   
# 그런데 T1의 경우 위 3가지 조건을 모두 만족하므로 tidy data이다.

1. T41과 T42를 이용하여 T1의 형태를 만드시오.

T41tidy <- T41 %>%   
 gather('1','2', key = 'grade', value='Female')  
T42tidy <- T42 %>%   
 gather('1','2', key = 'grade', value='Male')  
left\_join(T41tidy, T42tidy)

## Joining, by = c("school", "grade")

## # A tibble: 6 x 4  
## school grade Female Male  
## <chr> <chr> <dbl> <dbl>  
## 1 A 1 10 90  
## 2 B 1 30 270  
## 3 C 1 50 450  
## 4 A 2 20 180  
## 5 B 2 40 360  
## 6 C 2 60 540

1. T2를 T1의 형태로 만드시오.

T2 %>%   
 spread(key='gender', value='n')

## # A tibble: 6 x 4  
## school grade Female Male  
## <chr> <dbl> <dbl> <dbl>  
## 1 A 1 10 90  
## 2 A 2 20 180  
## 3 B 1 30 270  
## 4 B 2 40 360  
## 5 C 1 50 450  
## 6 C 2 60 540

1. T3을 T1의 형태로 만드시오.

T3 %>%   
 separate(FM, into = c('Female', 'Male'), sep='\_', convert = T)

## # A tibble: 6 x 4  
## school grade Female Male  
## <chr> <int> <int> <int>  
## 1 A 1 10 90  
## 2 A 2 20 180  
## 3 B 1 30 270  
## 4 B 2 40 360  
## 5 C 1 50 450  
## 6 C 2 60 540

1. 각 학교별, 학년별로 총 학생수를 구하여 T41과 같은 형태로 만드시오.

T1 %>%   
 group\_by(school, grade) %>%   
 summarise(total=Female+Male) %>%   
 spread(key = 'grade', value='total', convert = T)

## # A tibble: 3 x 3  
## # Groups: school [3]  
## school `1` `2`  
## <chr> <int> <int>  
## 1 A 100 200  
## 2 B 300 400  
## 3 C 500 600