

HW #5 (2018. 11. 15(금) 9:30am까지 제출)

* tidyverse 에서 제공하는 table3 자료를 이용

1. 각 연도별, 나라별, TB 수를 뽑아내시오.
2. 각 연도별, 나라별로 population 수를 뽑아내시오.
3. 1,2의 결과를 이용하여 비율을 계산하시오.
4. 시간에 따른 TB 수의 변화를 나라별로 나타내는 그림을 그리시오.

5. gather()과 spread()는 대칭적인 함수인가?

아래의 예를 실행시켜보고 이를 이용하여 설명하시오.

```
stocks <- tibble(  
  year = c(2015, 2015, 2016, 2016),  
  half = c( 1,    2,    1,    2),  
  return = c(1.88, 0.59, 0.92, 0.17)  
)  
stocks %>%  
  spread(year, return) %>%  
  gather("year", "return", `2015`:`2016`)
```

* nycflights13 패키지에 있는 flights 자료를 이용하여 각 상황에 맞는 비행을 모두 찾으시오.

6. dep_time과 sched_dep_time은 특수한 형태로 저장된 자료이다. (519는 5시 19분을 나타냄). 이 두 변수의 시간을 자정을 기준으로 몇 분 후인지를 나타내는 dep_time_min과 sched_dep_time_min을 만드시오.

7. 가장 delay 건수가 많은 도착공항은? (여기서 delay는 dep_delay, arr_delay 모두를 고려) 가장 departure delay 가 많은 공항은? 가장 arrival delay 가 많은 공항은?

8. 자료가 잘못 입력되었다고 의심될 만큼 빠르거나 느린 비행이 있는지 살펴보고 한다. 이를 위해 필요한 변수를 만들고 살펴보시오.