

산보연 Data Quality check summary

작성자 : 이은경

1. 3가지 업종 Category(대분류, 중분류, 소분류)의 unique한 level은 몇 개인가.

① 대분류(UP1)

: 총 22개로, “A”~“U”와 “ ”(빈칸)이 있다. UP1 값이 “ ”인 관측치는 총 722570개이다. Supplementary file “업종코드” 파일을 참고해보면, UP1 값 중 “ ”(빈칸)은 없다.

② 중분류(UP2)

: 총 78개로, “01” ~ “99”와 (“04”, “09”, “40”, “43”, “44”, “48”, “53”, “54”, “57”, “67”, “76”~“83”, “88”, “89”, “92”, “93”은 제외), “ ”(빈칸)중 하나의 값을 가진다. UP2 값이 “ ”인 관측치는 총 5586개이다.

Supplementary file “업종코드” 파일을 참고해보면, UP2 값 중 “ ”(빈칸)은 없다.

그리고, 주어진 데이터의 UP2 값 중 “34”와 “76”이 있는데 이 값은 Supplementary file “업종코드” 파일에 주어진 UP2 값에는 없다.

(참고) Supplementary file “업종코드” 파일에서는, UP2가 가질 수 있는 값 중 “A” ~ “U”도 있다.

③ 소분류(UP3)

: 총 233개로, “01” ~ “990”과 “ ”(빈칸)중 하나의 값을 가진다. UP3 값이 “ ”인 관측치는 총 5586개이다.

Supplementary file “업종코드” 파일을 참고해보면, UP3 값 중 “ ”(빈칸)은 없다.

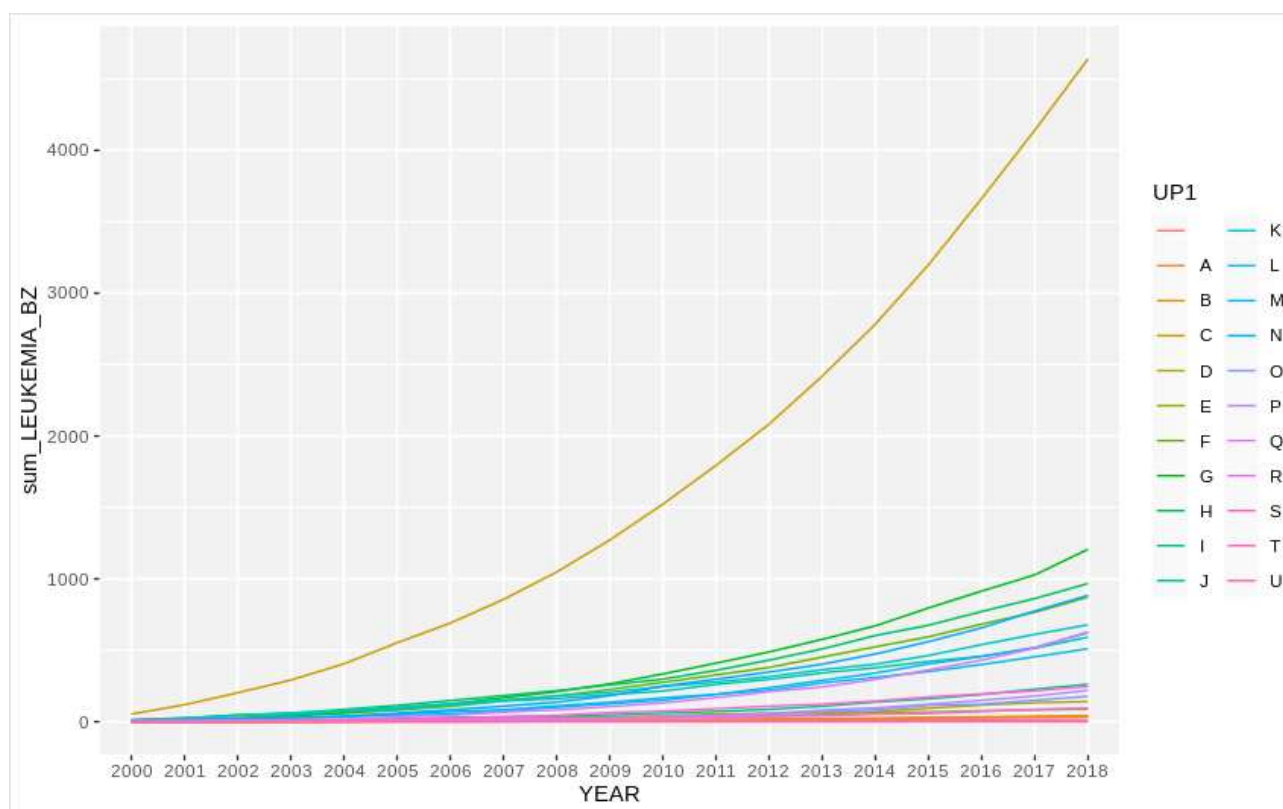
그리고, 주어진 데이터의 UP3 값 중 “304”, “426”, “716”, “761” ~ “764”가 있는데 이 값은 Supplementary file “업종코드” 파일에 주어진 UP3 값에는 없다.

(참고) Supplementary file “업종코드” 파일에서는, UP3가 가질 수 있는 값 중 “A” ~ “U”도 있다.

2. 모든 사업장 분류 기준(대분류, 중분류, 소분류)별, 입사 시기의 나이(ECNY_AGE)별, SEX 별(SEX) 누적 근로자수, 누적발병 건수가 단조 증가한다.

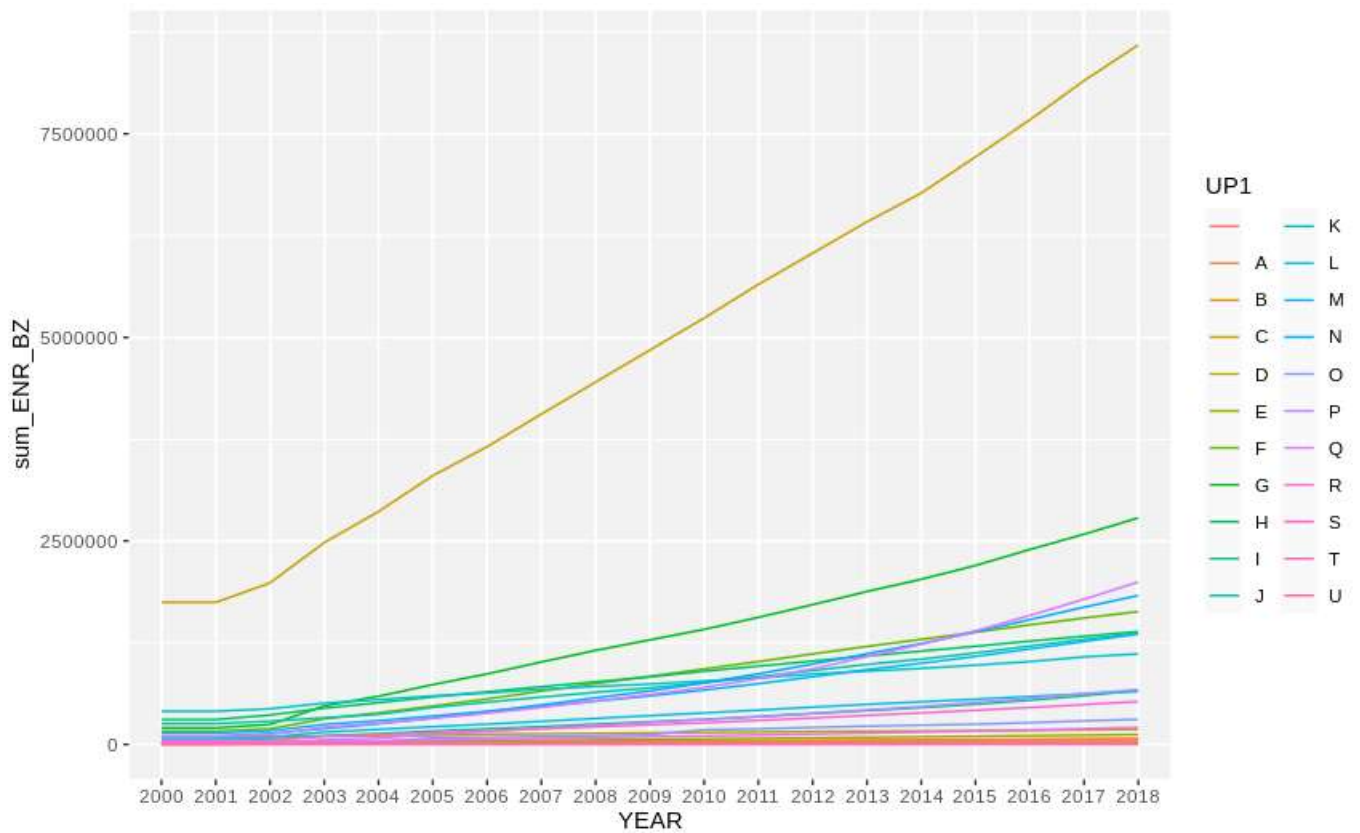
3. 사업장 대분류 category별 year에 따른 누적발병 건수, 누적 근로자 수 plotting

① 사업장 대분류 category별 year에 따른 누적 근로자 수



→ 그래프 파악 결과, 대분류가 “C”(제조업)인 사업장의 누적 근로자 수가 가장 가파르게 증가한다.

② 사업장 대분류 category별 year에 따른 누적발병 건수



→ 그래프 파악 결과, 대분류가 “C”(제조업)인 사업장의 누적발병 건수가 가장 가파르게 증가한다.