

산보연 Data check 07월 04일 Version

작성자 : 이은경

1. “Leukemia data”에 대해 이전에는 중분류(UP2)를 기준으로 grouping 한 후, 통합 누적 발생률, 간접 SIR등 여러 지표 확인 → UP1에 대해서 똑같은 지표 다시 한 번 확인.

: 대조군 생성 계획은 이전 방식과 동일 (Data check 6월 29일 version 참고)

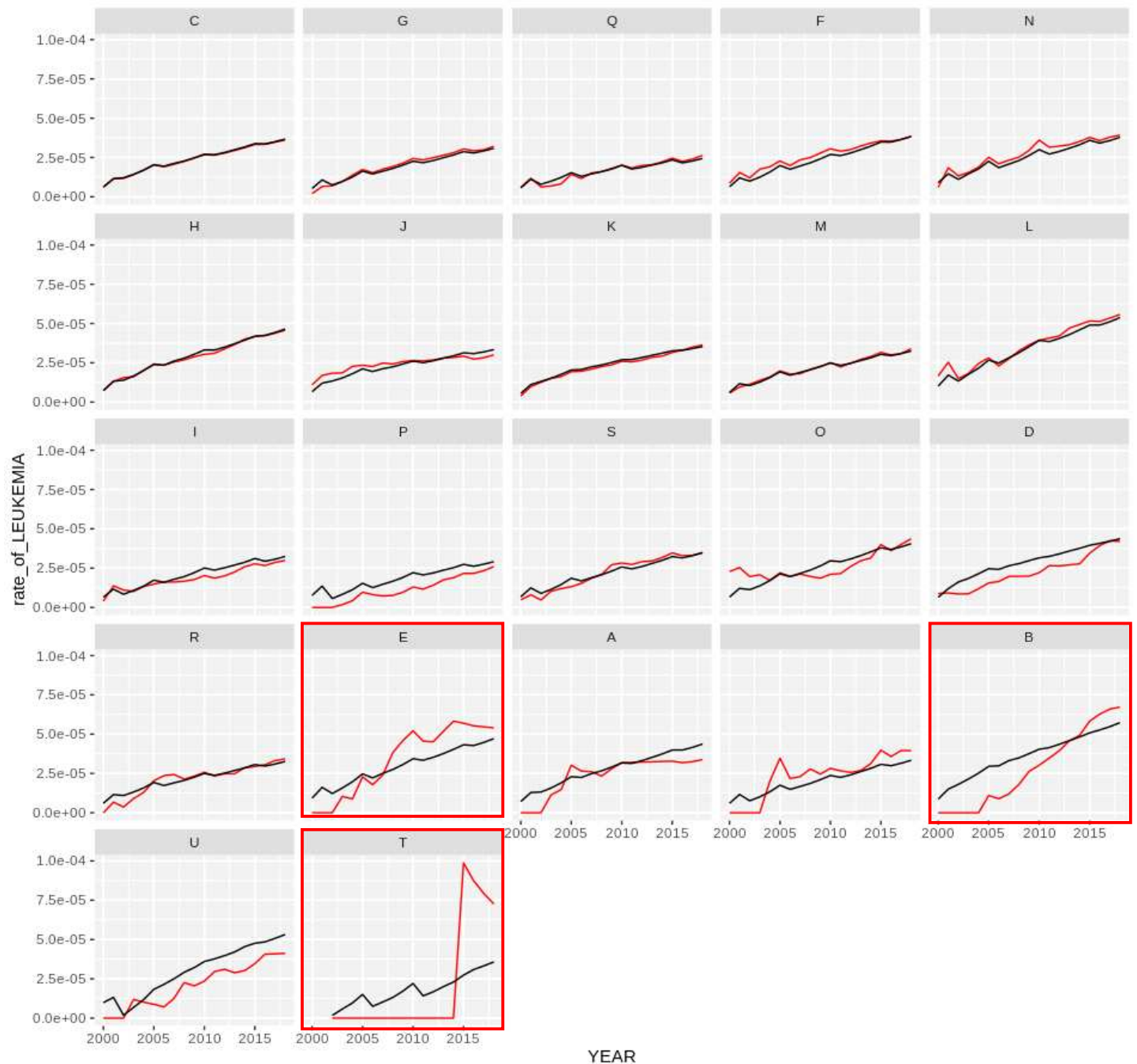
: 대조군 생성 후 통합 누적 발생률 계산. 대분류 기준 각 사업장에 대한 통합 누적 발생률(“origin rate” 변수명)과 대조군의 통합 누적 발생률(“weight_rate” 변수명) 겹쳐서 시각화

1-1. 사업장별 “통합 누적 발생률”과 대조군의 “통합 누적 발생률” 비교

① y축의 scale을 조정하지 않은 경우

: 이전 version과 마찬가지로, 그래프의 순서는 2018년 기준 UP1 기준으로 grouping한 후 계산한 추적 인년 합계의 순위(오름차순 정렬)

: **빨간색** 그래프가 **대분류 기준 사업장 별** 통합 누적 발생률, **검은색** 그래프가 **대조군의 통합 누적 발생률** 의미



⇒ 위 그래프에서 대조군의 통합 누적 발생률 변화 추세와 많은 차이를 보이는 그래프에 빨간색 테두리를 그렸다. 경향을 보면, 이전과 마찬가지로 추적 인년 합계가 적을수록 대조군의 추세와 차이를 보이는 것을 알 수 있다. 해당 사업장명은 아래 표에 제시되어 있다.

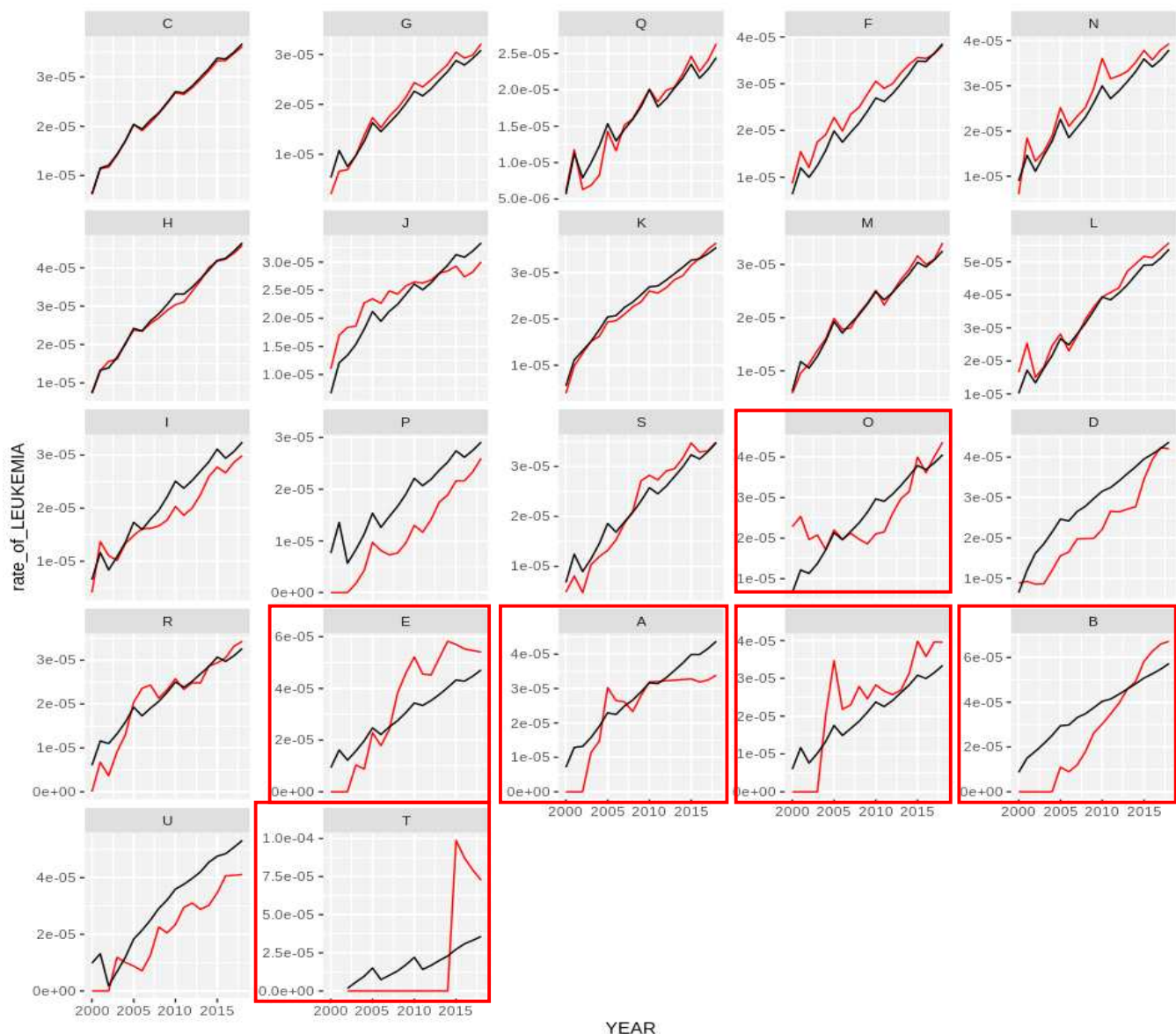
UP1	사업장명
E	수도, 하수 및 폐기물 처리, 원료 재생업
B	광업
T	가구 내 고용 활동 & 자가소비 생산 활동

: 위 그래프를 똑같은 사항에 대해 UP2를 기준으로 그린 그래프(6월 29일 version 2 page 그래프)와 비교해보았을 때, UP2 기준 “36”, “37”, “39”는 UP1 “E”에 속하고, UP2 기준 “5”, “6”, “8”은 UP1 기준 “B”에 속하며, UP2 기준 “97”, “98”은 UP1 기준 “T”에 속한다. (UP2 기준 “36”, “37”, “39”, “5”, “6”, “8”, “97”, “98”은 추세가 불안정하다고 보고 된 사업장이다) 그러나, UP2 기준 결과와 다르게 **제조업과 관련된 사업장(“C”)은 불안정한 추세를 보이지 않는다.**

② y축의 scale을 조정한 경우

: 이전 version과 마찬가지로, 그래프의 순서는 2018년 기준 UP1 기준으로 grouping한 후 계산한 추적 인년 합계의 순위(오름차순 정렬)

: **빨간색** 그래프가 **대분류 기준 사업장 별** 통합 누적 발생률, **검은색** 그래프가 **대조군의 통합 누적 발생률** 의미



⇒ 위 그래프에서 대조군의 통합 누적 발생률 변화 추세와 많은 차이를 보이는 그래프에 빨간색 테두리를 그렸다. 경향을 보면, 이전과 마찬가지로 추적 인년 합계가 적을수록 대조군의 추세와 차이를 보이는 것을 알 수 있다. 해당 사업장명은 아래 표에 제시되어 있다.

UP1	사업장명
O	공공행정, 국방 및 사회보장 행정
E	수도, 하수 및 폐기물 처리, 원료 재생업
A	농업, 어업, 임업
“ ”	(명시되어 있지 않은 코드)
B	광업
T	가구 내 고용 활동 & 자가소비 생산 활동

: ①의 결과와 동일하게 나온 것을 알 수 있다.

: 위 그래프를 똑같은 사항에 대해 UP2를 기준으로 그린 그래프(6월 29일 version 3 page 그래프)와 비교해보았을 때, UP2 기준 “36”, “37”, “38”, “39”는 UP1 “E”에 속하고, UP2 기준 “5”, “6”, “8”은 UP1 기준 “B”에 속하며, UP2 기준 “97”, “98”은 UP1 기준 “T”에 속한다. 또한, UP2 기준 “2”는 UP1 기준 “A”에 속한다. (UP2 기준 “36”, “37”, “38”, “39”, “5”, “6”, “8”, “2”, “97”, “98”은 추세가 불안정하다고 보고 된 사업장이다) 반면, UP1 기준 “O”는 UP2 기준 “84” 사업장과 대응되는데, UP2 기준으로 그린 그래프에서는 안정된 추세를 보였다. 참고로, UP2를 기준으로 그린 그래프에서 UP2 기준 (“21”, “34” : UP1 기준 “C”인 사업장), (“50”, “51” : UP1 기준 “H”인 사업장), (“59” : UP1 기준 “J”인 사업장), (“73” : UP1 기준 “M”인 사업장), (“76” : UP1 기준 “N”인 사업장), (“90” : UP1 기준 “R”인 사업장)인 사업장의 추세가 불안정하다고 판단을 하였는데, UP1 기준에서는 대조군의 추세와 크게 차이 나지 않는다는 점을 확인하였다.

1-2. 각 사업장 별 통합 누적 발생률과 대조군의 통합 누적 발생률의 “비율”(간접 SIR)이 가장 큰 사업장 TOP10은?

UP1	사업장명	간접 SIR
T	가구 내 고용 활동 & 자가소비 생산 활동	2.03
“ ”	(명시되어 있지 않은 코드)	1.18
B	광업	1.17
E	수도, 하수 및 폐기물 처리, 원료 재생업	1.15
Q	보건업 및 사회복지 서비스업	1.08
O	공공행정, 국방 및 사회보장 행정	1.08
R	예술, 스포츠 여가관리 서비스	1.05
M	전문, 과학, 기술 서비스업	1.05
G	도매 및 서비스업	1.04
N	임대 서비스업, 사업 시설 관리 지원 서비스업	1.04

: 위 표 결과를 UP2 기준으로 파악한 결과(6월 29일 version 4 page 3번 표)와 비교해보았을 때, 간접 SIR이 가장 큰 사업장은 동일하다. UP2 기준으로 파악한 결과와 UP1 기준으로 파악한 결과를 비교한 표를 아래에 제시하였다. (겹치는 사업장만 작성하였다.)

UP2 기준 간접 SIR TOP10에 포함된 사업장	UP1 기준 간접 SIR TOP10에 포함된 사업장 중 왼쪽 중분류 사업장에 대응되는 대분류
“97”	T
“90”	R
“37”	E
“5”, “7”	B
“76”	N

→ 결과들이 비슷한 점을 확인할 수 있다. 그런데, UP2 기준 결과와 다르게 제조업과 관련된 사업장(“C”)은 TOP10에 포함되지 않는다.

1-3. 백혈병 통합 누적 발생률 TOP10인 사업장은?

UP1	사업장명	통합 누적 발생률
T	가구 내 고용 활동 & 자가소비 생산 활동	0.0000726
B	광업	0.0000673
L	부동산업	0.0000557
E	수도, 하수 및 폐기물 처리, 원료 재생업	0.0000540
H	운수 및 창고업	0.0000458
O	공공행정, 국방 및 사회보장 행정	0.0000437
D	전기 가스, 증기 및 공기 조절 공급업	0.0000421
U	국제 및 외국기관	0.0000412
“ ”	명시되어 있지 않은 코드	0.0000395
N	임대 서비스업, 사업 시설 관리 지원 서비스업	0.0000393

: 위 표 결과를 UP2 기준으로 파악한 결과(6월 29일 version 4 page 4번 표)와 비교해보았을 때, TOP10에 속하는 사업장은 대부분 비슷한 것을 확인할 수 있었다. UP2 기준으로 파악한 결과와 UP1 기준으로 파악한 결과를 비교한 표를 아래에 제시하였다.
(겹치는 사업장만 작성하였다.)

UP2 기준 통합 누적 발생률 TOP10에 포함된 사업장	UP1 기준 통합 누적 발생률 TOP10에 포함된 사업장 중 왼쪽 중분류 사업장에 대응되는 대분류
“97”	T
“37”, “38”	E
“5”, “7”	B
“74”	N
“68”	L

→ 결과가 비슷한 것을 확인할 수 있다. 그런데, UP2 기준 결과와 다르게 제조업과 관련된 사업장(“C”)은 TOP10에 포함되지 않은 반면, 운수업과 관련된 사업장(“H”)가 포함된 것을 확인하였다.

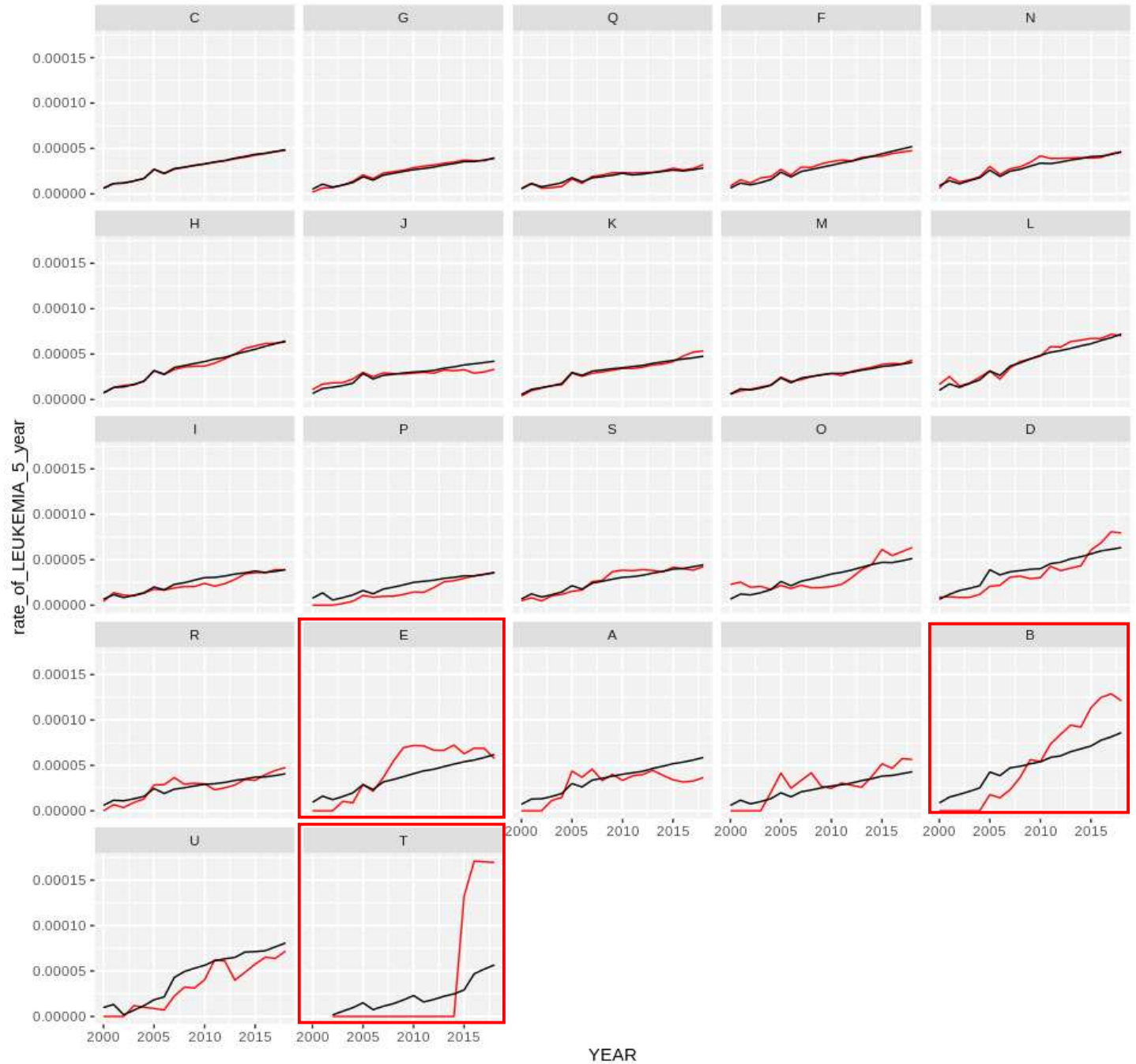
----- 다음 페이지로 이어집니다. -----

1-4. 각 사업장 별, 대조군 별로 5년 단위 통합 누적 발생률 계산 & 비교

: 5년 단위 통합 누적 발생률 계산 후, 그래프 통해 시각화. (그래프의 순서는 2018년 기준 UP1 기준으로 grouping한 후 합산한 추적 인년 합계의 순위(오름차순 정렬))

: **빨간색** 그래프가 **UP1 기준 사업장 별** 5년 단위 통합 누적 발생률, **검은색** 그래프가 **대조군의 5년 단위 통합 누적 발생률** 의미

① y축의 scale을 조정하지 않은 경우



⇒ 위 그래프에서 대조군의 통합 누적 발생률 변화 추세와 많은 차이를 보이는 그래프에 빨간색 테두리를 그렸다. 경향을 보면, 이전과 마찬가지로 추적 인년 합계가 적을수록 대조군의 추세와 차이를 보이는 것을 알 수 있다. 해당 사업장명은 아래 표에 제시되어 있다.

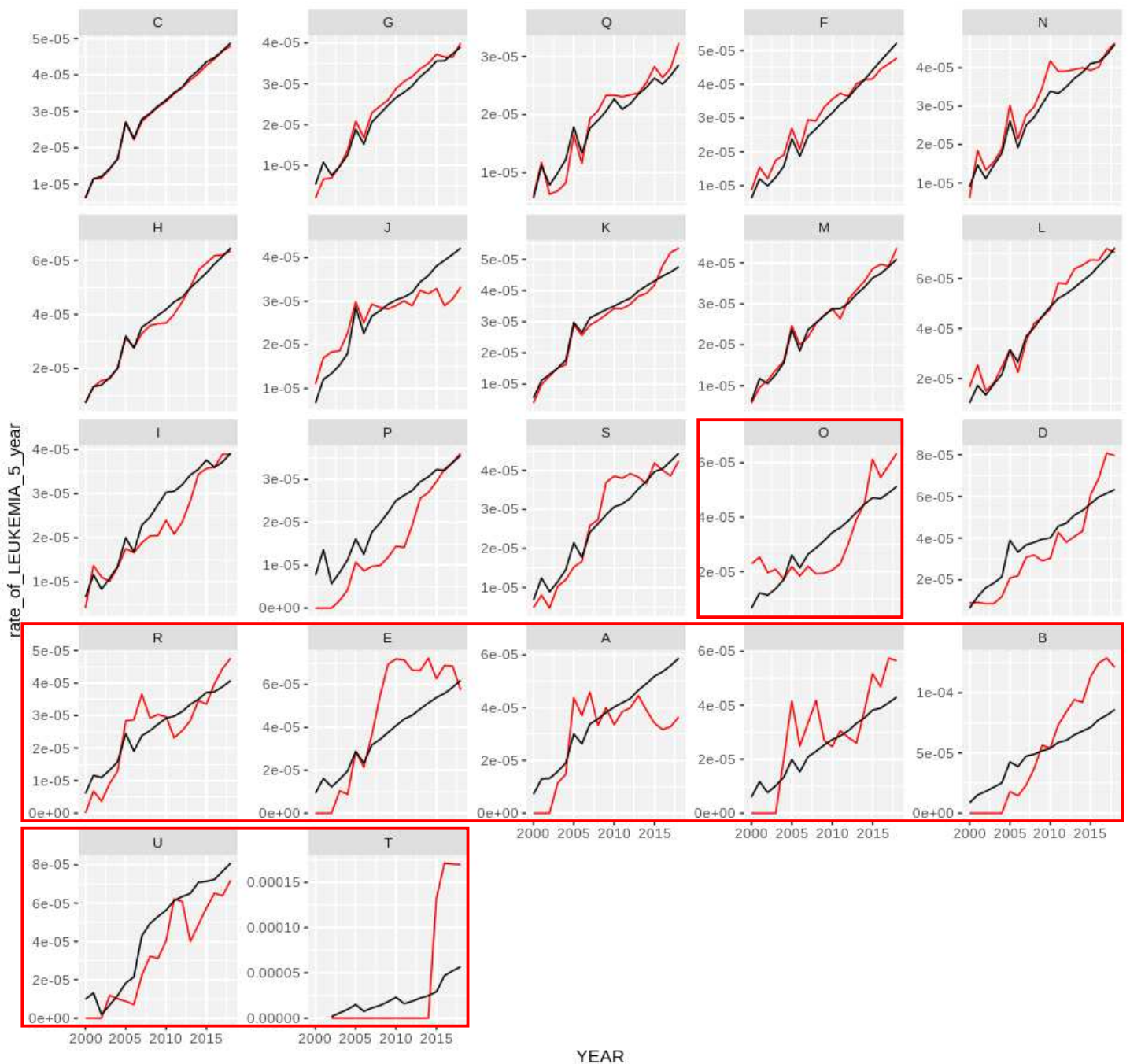
UP1	사업장명
E	수도, 하수 및 폐기물 처리, 원료 재생업
B	광업
T	가구 내 고용 활동 & 자가소비 생산 활동

: 위 그래프를 똑같은 사항에 대해 UP2를 기준으로 그린 그래프(6월 29일 version 6 page 그래프)와 비교해보았을 때, 대조군의 추세와 차이를 보이는 사업장은 비슷한 것을 확인할 수 있었다. 비교한 결과를 아래 표에 제시하였다. (겹치는 사업장만 표기하였다.)

UP2 기준 대조군의 추세와 차이 나는 사업장(y축 scale 조정 안한 version)	UP1 기준 대조군의 추세와 차이를 보이는 사업장 중 왼쪽 중분류 사업장에 대응되는 대분류
"97", "98"	T
"36", "37", "39"	E
"5", "6", "8"	B

→ 결과가 대부분 비슷한 것을 확인할 수 있다. 그런데, UP2 기준으로 시각화한 결과와 다르게 **제조업과 관련된 사업장("C")**가 UP1 기준으로 시각화한 결과에서는 안정된 추세를 보이는 것을 확인하였다.

② y축 scale을 조정한 경우



⇒ 위 그래프에서 대조군의 통합 누적 발생률 변화 추세와 많은 차이를 보이는 그래프에 빨간색 테두리를 그렸다. 경향을 보면, 이전과 마찬가지로 추적 인년 합계가 적을수록 대조군의 추세와 차이를 보이는 것을 알 수 있다. 해당 사업장명은 아래 표에 제시되어 있다.

UP1	사업장명
O	공공행정, 국방 및 사회보장 행정
R	예술, 스포츠, 여가 관련 서비스업
E	수도, 하수 및 폐기물 처리, 원료 재생업
A	농업, 어업, 임업
“ ”	(명시되어 있지 않은 코드)
B	광업
U	국제 및 외국기관
T	가구 내 고용 활동 & 자가소비 생산 활동

: 위 그래프를 똑같은 사항에 대해 UP2를 기준으로 그린 그래프(6월 29일 version 7 page 그래프)와 비교해보았을 때, 대조군의 추세와 차이를 보이는 사업장은 비슷한 것을 확인할 수 있었다. 비교한 결과를 아래 표에 제시하였다. (겹치는 사업장만 표기하였다.)

UP2 기준 대조군의 추세와 차이 나는 사업장(y축 scale 조정한 version)	UP1 기준 대조군의 추세와 차이를 보이는 사업장 중 왼쪽 중분류 사업장에 대응되는 대분류
“1”, “2”, “3”	A
“5”, “6”, “8”	B
“36”, “37”, “38”, “39”	E
“90”	R
“97”, “98”	T

→ 결과가 대부분 비슷한 것을 확인할 수 있다. 그런데, UP2 기준으로 시각화한 결과와 다르게 **제조업과 관련된 사업장(“C”)**가 UP1 기준으로 시각화한 결과에서는 안정된 추세를 보이는 것을 확인하였다.

1-5. 각 사업장 별(UP1 기준) 5년 단위 통합 누적 발생률과 대조군의 5년 단위 통합 누적 발생률의 “비율”(간접 SIR)이 가장 큰 사업장 TOP10은?

UP1	YEAR	사업장명	간접 SIR
T	2015	가구 내 고용 활동 & 자가소비 생산 활동	4.54
T	2016	가구 내 고용 활동 & 자가소비 생산 활동	3.64
O	2000	공공행정, 국방 및 사회보장 행정	3.47
T	2017	가구 내 고용 활동 & 자가소비 생산 활동	3.27
T	2018	가구 내 고용 활동 & 자가소비 생산 활동	2.99
“ ”	2005	(명시되어 있지 않은 코드)	2.10
O	2001	공공행정, 국방 및 사회보장 행정	2.09
E	2009	수도, 하수 및 폐기물 처리, 원료 재생업	1.85
“ ”	2008	(명시되어 있지 않은 코드)	1.82
E	2010	수도, 하수 및 폐기물 처리, 원료 재생업	1.76

: 1-2번(연도를 5년 단위로 나누지 않고 계산한 간접 SIR TOP10 사업장 파악) 결과와 겹치는 사업장을 파란색으로 표시

1-6. 5년 단위 백혈병 통합 누적 발생률 TOP10인 사업장(UP1 기준)은? (2018년 기준으로 파악하지 않음.)

UP1	YEAR	사업장명	통합 누적 발생률
T	2016	가구 내 고용 활동 & 자가소비 생산 활동	0.000171
T	2017	가구 내 고용 활동 & 자가소비 생산 활동	0.000170
T	2018	가구 내 고용 활동 & 자가소비 생산 활동	0.000170
T	2015	가구 내 고용 활동 & 자가소비 생산 활동	0.000132
B	2017	광업	0.000129
B	2016	광업	0.000125
B	2018	광업	0.000121
B	2015	광업	0.000113
B	2013	광업	0.0000943
B	2014	광업	0.0000922

: 1-3번(연도를 5년 단위로 나누지 않고 계산한 통합 누적 발생률 TOP10 사업장 파악) 결과와 겹치는 사업장을 **파란색**으로 표시

→ 1-5번 결과와 1-6번 결과를 UP2 기준으로 파악한 결과(6월 29일 version의 8 page의 6, 7번)들과 비교해보았을 때 비슷한 결과를 보여준다는 것을 확인하였다. UP2 기준으로 조사해보았을 때도, “가구 내 고용활동”, “광업”이 우위를 차지하고 있었다. 다만, UP2 기준으로 파악한 결과와 달리 **제조업 관련 사업장은 UP1 기준으로 파악했을 때 우위를 차지하지 않는다는 것**에 주목할 필요가 있다.

1-7. 2018년 기준 백혈병 5년 단위 통합 누적 발생률 사업장(UP1 기준) TOP10은?

UP1	사업장명	통합 누적 발생률
T	가구 내 고용 활동 & 자가소비 생산 활동	0.000170
B	광업	0.000121
D	전기 가스, 증기 및 공기 조절 공급업	0.0000795
U	국제 및 외국기관	0.0000719
L	부동산업	0.0000703
O	공공행정, 국방 및 사회보장 행정	0.0000635
H	운수 및 창고업	0.0000634
E	수도, 하수 및 폐기물 처리, 원료 재생업	0.0000575
“ ”	(명시되어 있지 않은 코드)	0.0000565
K	금융 및 보험업	0.0000536

: **빨간색**으로 표시한 항목은 1-6번 결과(모든 연도 고려한 5년 단위 통합 누적 발생률 TOP10)와 겹치는 사업장 의미

→ 1-7번 결과를 UP2 기준으로 파악한 결과(6월 29일 version 9 page의 8번)와 비교해보았을 때 비슷한 결과가 나온 것을 확인할 수 있었다. 비교한 결과를 아래 표에 제시하였다. (겹치는 사업장만 표기하였다.)

UP2 기준 2018년도 백혈병 통합 누적 발생률 TOP10에 포함되는 사업장	UP1 기준 2018년도 백혈병 통합 누적 발생률 TOP10에 포함되는 사업장 왼쪽 중분류 사업장에 대응되는 대분류
“97”	T
“5”, “6”, “7”	B
“36”, “37”	E
“35”	D

→ 결과가 대부분 비슷하며, “가구 내 고용활동”과 “광업” 그리고 “수도, 하수 및 폐기물 처리, 원료 재생업”이 중점적으로 나온다는 것에 주목할 필요가 있다. 또한, UP2 기준 파악한 결과와 달리 **제조업 관련 사업장(“C”)이 우위를 차지하지 않는다는 점** 또한 주목할 필요가 있다.

1-8. 2018년 기준 각 사업장 별 5년 단위 통합 누적 발생률과 대조군의 5년 단위 통합 누적 발생률의 “비율”(간접 SIR)이 가장 큰 TOP10 사업장(UP1 기준)은?

UP1	사업장명	간접 SIR
T	가구 내 고용 활동 & 자가소비 생산 활동	2.99
B	광업	1.41
“ ”	(명시되어 있지 않은 업종 코드)	1.31
D	전기 가스, 증기 및 공기 조절 공급업	1.26
O	공공행정, 국방 및 사회보장 행정	1.24
R	예술, 스포츠 여가관리 서비스	1.17
Q	보건업 및 사회복지 서비스업	1.13
K	금융 및 보험업	1.12
M	전문, 과학, 기술 서비스업	1.07
G	도매 및 소매업	1.02

: 빨간색으로 표시한 항목은 1-5번 결과(모든 연도 고려한 5년 단위 간접 SIR TOP10)와 겹치는 사업장 의미

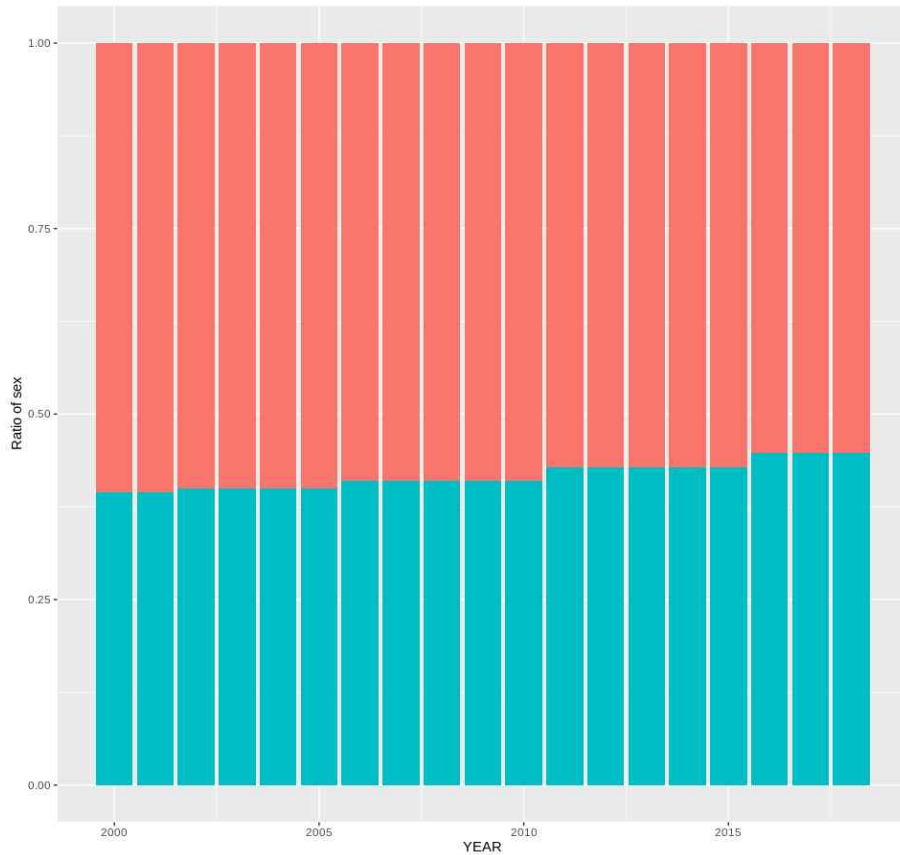
→ 1-8번 결과를 UP2 기준으로 파악한 결과(6월 29일 version 9 page의 9번)와 비교해보았을 때 비슷한 결과가 나온 것을 확인할 수 있었다. 비교한 결과를 아래 표에 제시하였다. (겹치는 사업장만 표기하였다.)

UP2 기준 2018년도 백혈병 간접 SIR TOP10에 포함되는 사업장	UP1 기준 2018년도 백혈병 간접 SIR TOP10에 포함되는 사업장 왼쪽 중분류 사업장에 대응되는 대분류
“97”	T
“5”, “6”	B
“90”	R

→ 결과가 대부분 비슷하며, “가구 내 고용활동”과 “광업”이 중점적으로 나온다는 것에 주목할 필요가 있다. 또한, UP2 기준 파악한 결과와 달리 제조업 관련 사업장(“C”), 수도, 하수 및 폐기물 처리, 원료 재생업(“E”)이 우위를 차지 않는다는 점 또한 주목할 필요가 있다.

1. YEAR에 따른 SEX의 비율 변화 시각화

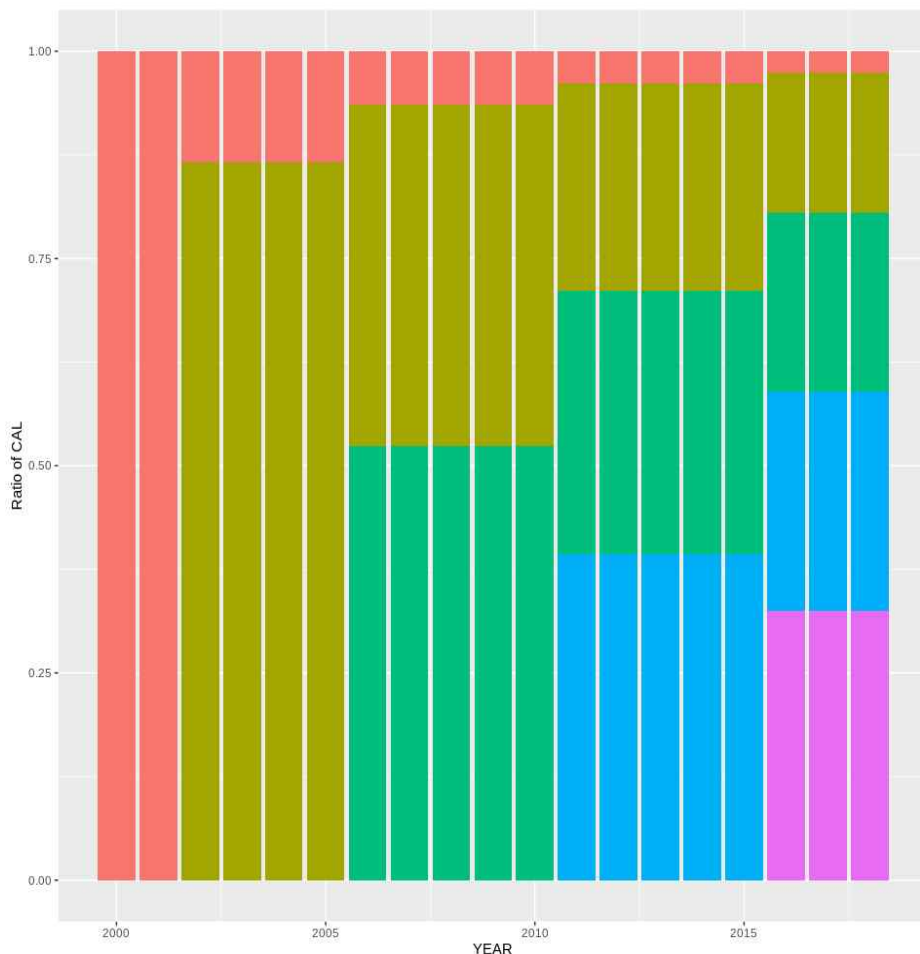
: 업종별로 나누지 않음. / **빨간색**이 “남성”, **파란색**이 “여성” 의미 / 성 비율 상대도수 표시 (전체 합이 1이 되도록)



연도별 성별 비율 변화를 살펴 보았을 때, 연도에 따라 성별 비율이 크게 변하지 않는다. 모든 연도에 대해 남성의 비율이 더 크나, 해가 지날수록 여성의 비율이 점차 커지는 것을 알 수 있다. 또한, 여성 비율에 해당하는 “파란색” 그래프만을 보았을 때, 증가하는 주기가 5년 단위인 것을 알 수 있다.

2. YEAR에 따른 CAL(입사 시기)의 비율 변화 시각화

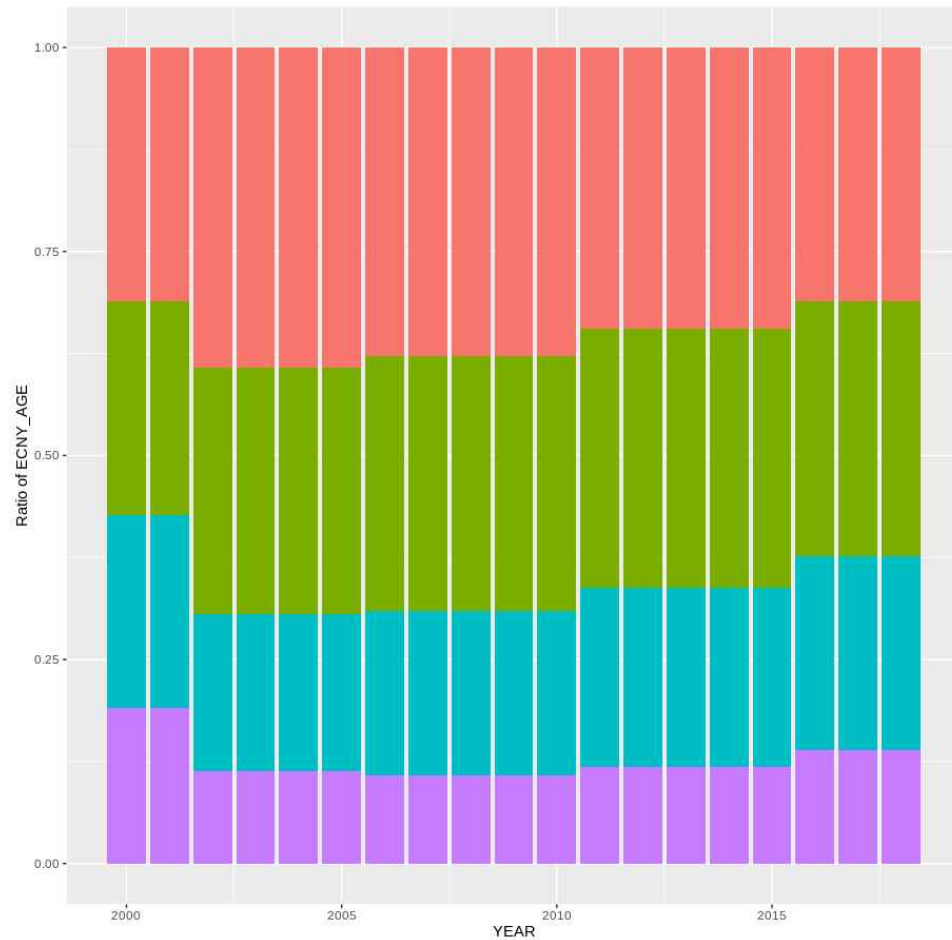
: 업종 별로 나누지 않음. / CAL(입사 시기) 비율 상대도수 표시 (전체 합이 1이 되도록)



2000년 ~ 2001년에는 CAL=0인, 즉 입사 시기가 1995년 이전인 근로자의 비율이 전부였고, 2002년 ~ 2005년까지는 CAL=1인, 입사 시기가 1996년 ~ 2000년인 근로자의 비율이 대다수이며, 2006년 ~ 2010년까지는 CAL=1인 근로자의 비율이 CAL=2(입사 시기가 01년 ~ 05년)인 근로자의 비율이 우위를 점한다. 2011년 ~ 2015년 사이에는 CAL=3(입사 시기가 06년 ~ 10년)인 근로자들의 비율이 급증하였으며, 2016년 이후에는 CAL=4(11년 ~ 15년)인 근로자들의 비율이 급증하였다. 성 비율 변화처럼 입사 시기도 5년 단위로 우위를 점하는 CAL의 범주가 변하는 것을 알 수 있다.

3. YEAR에 따른 ECNY_AGE(입사 시 연령대)의 비율 변화 시각화

: 업종 별로 나누지 않음. / ECNY_AGE(입사 시 연령대) 비율 상대도수 표시 (전체 합이 1이 되도록)



연도별 입사 시기 연령대의 비율을 살펴보았을 때, 연도가 변함에 따라 연령대 비율의 차이가 크게 변하지 않는 것으로 파악된다. 20대와 30대의 비율이 항상 우위를 점하고 있으며, 40대와 50대는 작은 비율을 차지하고 있다.