타이디 데이터 만들기
중복 변수는 제거하자

데이터 분석을 하기 전에 데이터가 어떻게 생겼는지 확인하는 작업은 필수적이다. 내가 분석할 데이터가 타이디(tidy)한 데이터가 아니라면 그에 맞는 처리를 수행해 데이터를 가공해야 한다. 다음의 who 데이터는 중복 변수를 가지고 있는데 하나씩 살펴보자.

country	**iso2**	**iso3**	**year**	**new_sp_m014**	**new_sp_m1524**	**…**
Afghanistan	AF	AFG	1980	NA	NA	…

who 데이터 세트에는 여러가지 문제가 포함 되어있다. 중복 변수, 이상한 변수, 여러개의 결측값이 그 문제이다.

1) 먼저 결측값 제거 후 이상한 변수부터 처리해보자.[1]

country	iso2	iso3	year	key	cases
Afghanistan	AF	AFG	1997	new_sp_m014	0
Afghanistan	AF	AFG	1998	new_sp_m014	30
Afghanistan	AF	AFG	1999	new_sp_m014	8

2) 여러 값이 혼합되어 있는 상태를 분리한다.[2]

country	iso2	iso3	year	type	sex	age	cases
Afghanistan	AF	AFG	1997	sp	m	014	0
Afghanistan	AF	AFG	1998	sp	m	014	30
Afghanistan	AF	AFG	1999	sp	m	014	8

중복 문제만 남겨두고 타이디한 데이터를 만들었다. 데이터를 보면 **country**와 **iso2**[3], **iso3**[4] 모두 국가를 나타내는 변수라는 것을 확인할 수 있다. 따라서 세 변수가 중복 변수라고 할 수 있으므로 국가의 풀네임인 country를 남겨두고 iso2와 iso3를 제거한다.

country	year	type	sex	age	cases
Afghanistan	1997	sp	m	014	0
Afghanistan	1998	sp	m	014	30
Afghanistan	1999	sp	m	014	8

[1] 상위 3개 행만 출력

[2] 상위 3개 행만 출력

[3] 국가 ISO 코드, ISO 3166-1 alpha-2

[4] 국가 ISO 코드, ISO 3166-1 alpha-3

```
"Written by Lee, Da-In"
```