

# Identity Masking with Class Roll Data : DS1801

coop711

2018-04-07

## Data

```
class_roll <- read.xlsx("../data/class_roll1801.xlsx",
                        sheetIndex = 1,
                        startRow = 2,
                        endRow = 15,
                        colIndex = c(3:7, 9),
                        colClasses = rep("character", 6),
                        encoding = "UTF-8",
                        stringsAsFactors = FALSE)
names(class_roll) <- c("dept", "id", "name", "year", "email", "cell_no")
```

## 학번 가리기

학번은 입학연도를 나타내는 첫 네자리와 개인 식별번호로 구성되어 있다. 여기서, 개인식별번호를 “9999”로 가려보자. `substr()` 을 이용하면 학번의 개인정보를 가리는 일은 한 줄의 코드로 가능하다.

```
substr(class_roll$id, start = 5, stop = 8) <- "9999"
kable(head(class_roll))
```

dept	id	name	year	email	cell_no
바이오메디컬학과	20179999	김주은	2	juneils@naver.com (mailto:juneils@naver.com)	010-4148-9162
금융정보통계학과	20179999	김현아	2	tkdlfrhrhd@nate.com (mailto:tkdlfrhrhd@nate.com)	010-8417-3094
금융정보통계학과	20149999	박찬영	3	cksdud436@nate.com (mailto:cksdud436@nate.com)	010-6354-6719
경영학과	20139999	서동혁	4	pianomov@hanmail.net (mailto:pianomov@hanmail.net)	010-7474-5999
금융정보통계학과	20179999	손봉호	2	zkvmek963@naver.com (mailto:zkvmek963@naver.com)	010-6604-0727
전자공학과	20139999	송소정	3	ricassong@gmail.com (mailto:ricassong@gmail.com)	010-2172-1735

## 이름 가리기

`substring()` 을 이용하면 각 이름의 2번째 글자 이후를 모두 “○○”으로 대체할 수 있다.

```
substring(class_roll$name, 2) <- "○○"
kable(head(class_roll))
```

dept	id	name	year	email	cell_no
바이오메디컬학과	20179999	김○○	2	juneils@naver.com (mailto:juneils@naver.com)	010-4148-9162
금융정보통계학과	20179999	김○○	2	tkdlfrhrhd@nate.com (mailto:tkdlfrhrhd@nate.com)	010-8417-3094
금융정보통계학과	20149999	박○○	3	cksdud436@nate.com (mailto:cksdud436@nate.com)	010-6354-6719
경영학과	20139999	서○○	4	pianomov@hanmail.net (mailto:pianomov@hanmail.net)	010-7474-5999
금융정보통계학과	20179999	손○○	2	zkvmek963@naver.com (mailto:zkvmek963@naver.com)	010-6604-0727
전자공학과	20139999	송○○	3	ricassong@gmail.com (mailto:ricassong@gmail.com)	010-2172-1735

## 전화번호 가리기

모바일 폰 번호의 끝 네 자리를 “xxxx” 로 대체한다. 정상적으로 번호가 나올 경우 열번째 글자부터 열세번째글자에 해당한다.

```
substring(class_roll$cell_no, 10, 13) <- "xxxx"
kable(head(class_roll))
```

dept	id	name	year	email	cell_no
바이오메디컬학과	20179999	김○○	2	juneils@naver.com (mailto:juneils@naver.com)	010-4148-xxxx
금융정보통계학과	20179999	김○○	2	tkdlfrhrhd@nate.com (mailto:tkdlfrhrhd@nate.com)	010-8417-xxxx
금융정보통계학과	20149999	박○○	3	cksdud436@nate.com (mailto:cksdud436@nate.com)	010-6354-xxxx
경영학과	20139999	서○○	4	pianomov@hanmail.net (mailto:pianomov@hanmail.net)	010-7474-xxxx
금융정보통계학과	20179999	손○○	2	zkvmek963@naver.com (mailto:zkvmek963@naver.com)	010-6604-xxxx
전자공학과	20139999	송○○	3	ricassong@gmail.com (mailto:ricassong@gmail.com)	010-2172-xxxx

## 전공 단위 이름 가리기

전공 단위 이름은 “학과”, “과”, “학”, “전공” 등 매우 다양한 명칭이 있으므로 `gsub()` 함수의 정규표현(regular expression)을 활용하여 “○○학과” 와 같은 방식으로 이름을 가릴 수 있다.

```
class_roll$dept <- sub("^.+$", "○○학과", class_roll$dept)
kable(head(class_roll))
```

dept	id	name	year	email	cell_no
○○학과	20179999	김○○	2	juneils@naver.com (mailto:juneils@naver.com)	010-4148-xxxx
○○학과	20179999	김○○	2	tkdlfrhrhd@nate.com (mailto:tkdlfrhrhd@nate.com)	010-8417-xxxx
○○학과	20149999	박○○	3	cksdud436@nate.com (mailto:cksdud436@nate.com)	010-6354-xxxx
○○학과	20139999	서○○	4	pianomov@hanmail.net (mailto:pianomov@hanmail.net)	010-7474-xxxx
○○학과	20179999	손○○	2	zkvmek963@naver.com (mailto:zkvmek963@naver.com)	010-6604-xxxx
○○학과	20139999	송○○	3	ricassong@gmail.com (mailto:ricassong@gmail.com)	010-2172-xxxx

## e-mail 가리기

email 주소는 @ 를 사이에 두고 나뉘어진다. gsub() 함수와 정규표현(regular expression)을 활용하면 email 주소에서 서  
비스업체명은 그대로 두고 개인 식별이 가능한 이름 부분을 user\_name 으로 대체할 수 있다.

```
class_roll$email <- sub("^.+@", "user_name@", class_roll$email)
kable(class_roll[sample(1:nrow(class_roll)), ])
```

dept	id	name	year	email	cell_no
3   ○○학과	20149999	박○○	3	user_name@nate.com (mailto:user_name@nate.com)	010-6354-xxxx
2   ○○학과	20179999	김○○	2	user_name@nate.com (mailto:user_name@nate.com)	010-8417-xxxx
12   ○○학과	20159999	차○○	4	user_name@naver.com (mailto:user_name@naver.com)	010-5164-xxxx
8   ○○학과	20149999	우○○	3	user_name@naver.com (mailto:user_name@naver.com)	010-6293-xxxx
11   ○○학과	20149999	조○○	3	user_name@naver.com (mailto:user_name@naver.com)	010-4191-xxxx
13   ○○학과	20179999	황○○	2	user_name@naver.com (mailto:user_name@naver.com)	010-7569-xxxx
5   ○○학과	20179999	손○○	2	user_name@naver.com (mailto:user_name@naver.com)	010-6604-xxxx
4   ○○학과	20139999	서○○	4	user_name@hanmail.net (mailto:user_name@hanmail.net)	010-7474-xxxx
9   ○○학과	20169999	이○○	3	user_name@naver.com (mailto:user_name@naver.com)	010-5104-xxxx
7   ○○학과	20179999	신○○	2	user_name@gmail.com (mailto:user_name@gmail.com)	010-3298-xxxx
1   ○○학과	20179999	김○○	2	user_name@naver.com (mailto:user_name@naver.com)	010-4148-xxxx

dept	id	name	year	email	cell_no
10   ○○학과	20159999	정○○	4	user_name@naver.com (mailto:user_name@naver.com)	010-7664-xxxx
6   ○○학과	20139999	송○○	3	user_name@gmail.com (mailto:user_name@gmail.com)	010-2172-xxxx