

# 클래스 롤 데이터를 사용한 아이덴티티 마스킹

문빈

2020-03-26

## 데이터

```
class_roll <- read.xlsx("../data/class_roll10303_deid.xlsx",
                        sheetIndex = 1,
                        startRow = 2,
                        endRow = 162,
                        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 공개 연도를 선택하십시오.

여기서, 개인 식별 번호를 "9999"로 가려 보자. `substr() <-` 을 이용하면 학번의 개인 정보를 숨 깁니다.

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

부서	신분증	이름	년	이메일	cell_no
중국학과	20119999	강가 나	4	a@naver.com (mailto:a@naver.com)	010-1164-5954
전자 공학과	20119999	강 나다	4	b@hanmail.net (mailto:b@hanmail.net)	010-1174-8159
컴퓨터 공학과	20179999	강 다라	1	c@naver.com (mailto:c@naver.com)	010-1135-5735
화학과	20149999	강 라마	4	d@hanmail.net (mailto:d@hanmail.net)	010-1166-8619
경영학과	20169999	강마 바	2	e@naver.com (mailto:e@naver.com)	010-1120-6892
경제학과	20129999	강사 아	삼	f@gmail.com (mailto:f@gmail.com)	010-1199-8710

## 이름 가리기

`substring() <-` 모든 "ㅇㅇ"로 대체 할 수 있습니다.

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

부서	신분증	이름	년	이메일	cell_no
중국학과	20119999	강 ㅇㅇ	4	a@naver.com (mailto:a@naver.com)	010-1164-5954
전자 공학과	20119999	강 ㅇㅇ	4	b@hanmail.net (mailto:b@hanmail.net)	010-1174-8159

부서	신분증	이름	년	이메일	cell_no
컴퓨터 공학과	20179999	강 ○ ○	1	c@naver.com (mailto:c@naver.com)	010-1135-5735
화학과	20149999	강 ○ ○	4	d@hanmail.net (mailto:d@hanmail.net)	010-1166-8619
경영학과	20169999	강 ○ ○	2	e@naver.com (mailto:e@naver.com)	010-1120-6892
경제학과	20129999	강 ○ ○	삼	f@gmail.com (mailto:f@gmail.com)	010-1199-8710

## 전화 번호 가리기

모바일 폰 번호의 끝 네 자리를“xxxx”로 대체한다. 가열 번호가 나올 경우 열 번째입니다.

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

부서	신분증	이름	년	이메일	cell_no
중국학과	20119999	강 ○ ○	4	a@naver.com (mailto:a@naver.com)	010-1164-xxxx
전자 공학과	20119999	강 ○ ○	4	b@hanmail.net (mailto:b@hanmail.net)	010-1174-xxxx
컴퓨터 공학과	20179999	강 ○ ○	1	c@naver.com (mailto:c@naver.com)	010-1135-xxxx
화학과	20149999	강 ○ ○	4	d@hanmail.net (mailto:d@hanmail.net)	010-1166-xxxx
경영학과	20169999	강 ○ ○	2	e@naver.com (mailto:e@naver.com)	010-1120-xxxx
경제학과	20129999	강 ○ ○	삼	f@gmail.com (mailto:f@gmail.com)	010-1199-xxxx

## 전공 단위 이름 가리기

전공 단위 이름은“학과”,“과”,“학”,“전공”등 gsub() 다양하고 다양한 방식으로 정규 표현 (regular expression)을 선택할 수 있습니다.

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

부서	신분증	이름	년	이메일	cell_no
○○학과	20119999	강 ○ ○	4	a@naver.com (mailto:a@naver.com)	010-1164-xxxx
○○학과	20119999	강 ○ ○	4	b@hanmail.net (mailto:b@hanmail.net)	010-1174-xxxx
○○학과	20179999	강 ○ ○	1	c@naver.com (mailto:c@naver.com)	010-1135-xxxx
○○학과	20149999	강 ○ ○	4	d@hanmail.net (mailto:d@hanmail.net)	010-1166-xxxx
○○학과	20169999	강 ○ ○	2	e@naver.com (mailto:e@naver.com)	010-1120-xxxx
○○학과	20129999	강 ○ ○	삼	f@gmail.com (mailto:f@gmail.com)	010-1199-xxxx

## 전자 메일 가리기

email 주소는 @ 를 사이에 두고 나뉘어진다. gsub() 함수와 정규표현(regular expression)을 활용하면 email 주소에서 서비스업체명은 그대로 두고 개인 식별이 가능한 이름 부분을 user\_name 으로 대체할 수 있다. 160명 중 20명만 랜덤하게 표본추출한다.

```
class_roll$email <- sub("^.+@", "user_name@", class_roll$email)
kable(class_roll[sample(1:nrow(class_roll), size = 25), ])
```

	dept	id	name	year	email	cell_no
45	○○학과	20179999	김○○	1	user_name@naver.com (mailto:user_name@naver.com)	010-1178-xxxx
6	○○학과	20129999	강○○	3	user_name@gmail.com (mailto:user_name@gmail.com)	010-1199-xxxx
94	○○학과	20139999	유○○	3	user_name@naver.com (mailto:user_name@naver.com)	010-1121-xxxx
37	○○학과	20179999	김○○	1	user_name@naver.com (mailto:user_name@naver.com)	010-1116-xxxx
144	○○학과	20139999	차○○	3	user_name@naver.com (mailto:user_name@naver.com)	010-1175-xxxx
129	○○학과	20159999	정○○	3	user_name@naver.com (mailto:user_name@naver.com)	010-1138-xxxx
93	○○학과	20179999	유○○	1	user_name@naver.com (mailto:user_name@naver.com)	010-1138-xxxx
13	○○학과	20179999	김○○	1	user_name@naver.com (mailto:user_name@naver.com)	010-1186-xxxx
41	○○학과	20179999	김○○	1	user_name@naver.com (mailto:user_name@naver.com)	010-1180-xxxx
109	○○학과	20179999	이○○	1	user_name@naver.com (mailto:user_name@naver.com)	010-1163-xxxx
100	○○학과	20179999	이○○	1	user_name@naver.com (mailto:user_name@naver.com)	010-1151-xxxx
42	○○학과	20139999	김○○	4	user_name@nate.com (mailto:user_name@nate.com)	010-1191-xxxx
82	○○학과	20179999	심○○	1	user_name@nate.com (mailto:user_name@nate.com)	010-1179-xxxx
127	○○학과	20139999	전○○	3	user_name@naver.com (mailto:user_name@naver.com)	010-1129-xxxx
23	○○학과	20169999	김○○	2	user_name@naver.com (mailto:user_name@naver.com)	010-1148-xxxx
10	○○학과	20129999	권○○	3	user_name@naver.com (mailto:user_name@naver.com)	010-1169-xxxx
26	○○학과	20179999	김○○	1	user_name@naver.com (mailto:user_name@naver.com)	010-1155-xxxx
3	○○학과	20179999	강○○	1	user_name@naver.com (mailto:user_name@naver.com)	010-1135-xxxx

	dept	id	name	year	email	cell_no
97	○○학과	20179999	윤○○	1	user_name@naver.com (mailto:user_name@naver.com)	010-1123-xxxx
15	○○학과	20169999	김○○	2	user_name@naver.com (mailto:user_name@naver.com)	010-1164-xxxx
87	○○학과	20179999	안○○	1	user_name@naver.com (mailto:user_name@naver.com)	010-1174-xxxx
35	○○학과	20169999	김○○	2	user_name@naver.com (mailto:user_name@naver.com)	010-1103-xxxx
142	○○학과	20139999	조○○	3	user_name@naver.com (mailto:user_name@naver.com)	010-1179-xxxx
73	○○학과	20179999	성○○	1	user_name@naver.com (mailto:user_name@naver.com)	010-1132-xxxx
145	○○학과	20179999	차○○	1	user_name@naver.com (mailto:user_name@naver.com)	010-1116-xxxx

```
write.table(class_roll, file = "../data/class_roll_masked.txt")
save.image("../R/class_roll_170303_data_masked.RData")
```

→