Red and Black 170303 : id Masked

coop711

2018-03-19

Data

```
class_roll <- read.table("../data/class_roll_masked.txt",
                         header = TRUE,
                         stringsAsFactors = FALSE,
                         encoding = "UTF-8")
str(class_roll)
```

```
## 'data.frame':    160 obs. of  6 variables:
## $ dept   : chr  "○○학과" "○○학과" "○○학과" "○○학과" ...
## $ id     : int  20119999 20119999 20179999 20149999 20169999 20129999 20149999 20169999 20179999 2
0129999 ...
## $ name   : chr  "강○○" "강○○" "강○○" "강○○" ...
## $ year   : int  4 4 1 4 2 3 4 2 1 3 ...
## $ email  : chr  "user_name@naver.com" "user_name@hanmail.net" "user_name@naver.com" "user_name@han
mail.net" ...
## $ cell_no: chr  "010-9164-xxxx" "010-8574-xxxx" "010-6435-xxxx" "010-2066-xxxx" ...
```

Randomization

```
# set.seed(107)
N <- nrow(class_roll)
class_roll$group <- sample(1:N) %% 2 + 1
class_roll$group <- factor(class_roll$group,
                           labels = c("Red", "Black"))
red_id <- which(class_roll$group == "Red")
black_id <- which(class_roll$group == "Black")
```

학번

```
ID_16 <- factor(ifelse(substr(class_roll$id, 1, 4) >= 2016,
                       "younger_16", "older_16"),
                levels = c("younger_16", "older_16"))
kable(table("그룹" = class_roll$group,
            "16학번 기준" = ID_16))
```

	younger_16	older_16
Red	40	40
Black	47	33

```
ID_15 <- factor(ifelse(substr(class_roll$id, 1, 4) >= 2015,
                       "younger_15", "older_15"),
                levels = c("younger_15", "older_15"))
kable(table("그룹" = class_roll$group,
            "15학번 기준" = ID_15))
```

	younger_15	older_15
Red	48	32
Black	52	28

```
ID_14 <- factor(ifelse(substr(class_roll$id, 1, 4) >= 2014,
                       "younger_14", "older_14"),
                levels = c("younger_14", "older_14"))
kable(table("그룹" = class_roll$group,
            "14학번 기준" = ID_14))
```

	younger_14	older_14
Red	58	22
Black	62	18

```
ID_13 <- factor(ifelse(substr(class_roll$id, 1, 4) >= 2013,
                       "younger_13", "older_13"),
                levels = c("younger_13", "older_13"))
kable(table("그룹" = class_roll$group,
            "13학번 기준" = ID_13))
```

	younger_13	older_13
Red	74	6
Black	72	8

e-mail 서비스업체

```
email_list <- strsplit(class_roll$email, "@", fixed = TRUE)
mail_com <- sapply(email_list, `[`, 2)
kable(table("그룹" = class_roll$group,
            "e-mail" = mail_com))
```

	daum.net	gmail.com	hanmail.net	nate.com	naver.com
Red	2	2	3	3	69
Black	0	4	6	4	66

성씨 분포

```
f_name <- substring(class_roll$name,
                    first = 1, last = 1)
kable(table("Group" = class_roll$group,
            "Family Name" = f_name))
```

	강	고	구	권	김	나	명	문	박	반	방	배	서	성	손	송	신	심	안	양	우	유	윤	이	임	장	전	정	조	차	최	하	한	황
Red	4	1	1	2	17	0	0	0	7	0	0	2	2	0	1	1	2	1	2	0	0	2	3	9	1	2	2	3	6	1	4	1	1	2
Black	2	0	0	2	19	1	1	1	6	1	1	0	2	1	1	2	1	0	3	1	1	3	2	12	2	1	0	5	1	1	5	0	1	1

많이 나오는 성씨

```
f_name_f <- factor(ifelse(f_name %in% c("김", "이", "박"),
                          f_name, "기타"),
                   levels = c("김", "이", "박", "기타"))
kable(table("Group" = class_roll$group,
            "Family Name" = f_name_f))
```

	김	이	박	기타
Red	17	9	7	47
Black	19	12	6	43

출석부에서 20명 비복원 랜덤 표집

```
kable(class_roll[sample(1:nrow(class_roll), size = 20), ])
```

	dept	id	name	year	email	cell_no	group
121	○○학과	20159999	임○○	3	user_name@gmail.com (mailto:user_name@gmail.com)	010-6413-xxxx	Black
108	○○학과	20179999	이○○	1	user_name@gmail.com (mailto:user_name@gmail.com)	010-9632-xxxx	Black
64	○○학과	20159999	박○○	3	user_name@naver.com (mailto:user_name@naver.com)	010-5023-xxxx	Red
151	○○학과	20149999	최○○	3	user_name@gmail.com (mailto:user_name@gmail.com)	010-9529-xxxx	Black
90	○○학과	20139999	유○○	3	user_name@naver.com (mailto:user_name@naver.com)	010-9984-xxxx	Black
14	○○학과	20119999	김○○	3	user_name@naver.com (mailto:user_name@naver.com)	010-8009-xxxx	Black
117	○○학과	20169999	이○○	2	user_name@naver.com (mailto:user_name@naver.com)	010-4650-xxxx	Red
130	○○학과	20129999	정○○	4	user_name@naver.com (mailto:user_name@naver.com)	010-9411-xxxx	Black
6	○○학과	20129999	강○○	3	user_name@gmail.com (mailto:user_name@gmail.com)	010-7499-xxxx	Red
8	○○학과	20169999	구○○	2	user_name@naver.com (mailto:user_name@naver.com)	010-8228-xxxx	Red
140	○○학과	20169999	조○○	2	user_name@naver.com (mailto:user_name@naver.com)	010-2425-xxxx	Red
29	○○학과	20149999	김○○	3	user_name@daum.net (mailto:user_name@daum.net)	010-4039-xxxx	Red
36	○○학과	20119999	김○○	4	user_name@naver.com (mailto:user_name@naver.com)	010-8799-xxxx	Red
85	○○학과	20139999	안○○	3	user_name@naver.com (mailto:user_name@naver.com)	010-5030-xxxx	Black
59	○○학과	20169999	박○○	2	user_name@naver.com (mailto:user_name@naver.com)	010-7728-xxxx	Black
66	○○학과	20119999	방○○	3	user_name@naver.com (mailto:user_name@naver.com)	010-8975-xxxx	Black
136	○○학과	20139999	정○○	2	user_name@hanmail.net (mailto:user_name@hanmail.net)	010-7231-xxxx	Red
25	○○학과	20179999	김○○	1	user_name@naver.com (mailto:user_name@naver.com)	010-8810-xxxx	Red
142	○○학과	20139999	조○○	3	user_name@naver.com (mailto:user_name@naver.com)	010-4879-xxxx	Red
86	○○학과	20169999	안○○	2	user_name@naver.com (mailto:user_name@naver.com)	010-9310-xxxx	Black