

데이터 분석 보고서

동양어문학과 201531539 박희선

2019년 6월 8일

<들어가며>

많은 사람들은 ‘사람은 쉽게 바뀌지 않는다.’, ‘사람은 고쳐쓰는 게 아니다.’ 라는 말을 자주 하곤 한다. 그래서 필자는 이 말이 범죄에도 적용되는 이야기인지 즉, 전과자들의 수가 초범자들의 수보다 많은지 범죄의 초범자의 비율과 전과자의 비율을 비교하여 정말 이 말이 맞는지를 알아 보고싶어서 이 데이터를 선택하였다.

1. 데이터 전처리

먼저, 공공데이터포털에서 ‘범죄자범행시전과및횡수_2016년_.csv’ 파일을 다운 받고 R로 불러온 후에 이것의 복사본을 만든 후 중범죄라고 할 수 있는 10개의 범죄를 추출하였다. 그리고 초범인지 재범인지 모르는 미상의 데이터는 제외하였다.

```
criminal_record <- read.csv("범죄자범행시전과및횡수_2016년_.csv")
new_record <- criminal_record
```

```
library(dplyr)
```

```
##
## Attaching package: 'dplyr'
```

```
## The following objects are masked from 'package:stats':
##
##   filter, lag
```

```
## The following objects are masked from 'package:base':
##
##   intersect, setdiff, setequal, union
```

```
new_record <- new_record %>%
  filter(범죄대분류 %in% c('강력범죄', '절도범죄', '폭력범죄', '지능범죄'))
new_record <- new_record %>%
  filter(범죄중분류 %in% c('살인기수', '강도', '강간', '강제추행', '절도범죄', '상해', '폭행', '협박', '폭력행위등', '사기')) %>%
  select(-미상)
```

2. 데이터 기초분석

전과없음이나, 전과 소계등의 데이터 형태가 범주형 데이터이기 때문에 이를 연속형 데이터로 바꾸었다. 또한, 초범자와 전과자의 차이를 보기 위해서 전과가 없음과 전과소계를 따로 추출하여 초범자의 비율과 전과자 비율을 생성하였다.

```
str(new_record)
```

```
## 'data.frame':   10 obs. of  14 variables:
## $ 범죄대분류   : Factor w/ 15 levels "강력범죄","교통범죄",...: 1 1 1 1 10 13 13 13 13 11
## $ 범죄중분류   : Factor w/ 38 levels "강간","강도",...: 17 2 1 3 27 19 35 36 34 16
## $ 계           : Factor w/ 38 levels "1,396","1,525",...: 24 4 26 9 5 34 13 10 25 15
## $ 없음        : Factor w/ 38 levels "1,633","116",...: 2 21 1 26 24 7 22 12 5 15
## $ 전과.소계.   : Factor w/ 38 levels "1,243","1,887",...: 9 1 12 32 29 25 33 37 13 35
## $ 전과.1범.    : Factor w/ 35 levels "1,268","1,452",...: 24 12 28 14 7 35 9 3 23 6
## $ 전과.2범.    : Factor w/ 38 levels "1","1,299","1,335",...: 21 12 27 3 35 32 7 2 18 6
## $ 전과.3범.    : Factor w/ 36 levels "1,021","1,197",...: 8 6 20 36 27 23 34 1 12 32
## $ 전과.4범.    : Factor w/ 37 levels "1,150","1,644",...: 5 37 11 29 25 21 30 33 2 28
## $ 전과.5범.    : Factor w/ 37 levels "-","1,302","1,576",...: 11 35 9 29 21 20 27 31 2 26
## $ 전과.6범.    : Factor w/ 36 levels "-","1","1,063",...: 33 32 9 25 17 16 23 30 3 24
## $ 전과.7범.    : Factor w/ 35 levels "-","1","1,081",...: 6 25 7 19 12 11 17 24 30 18
## $ 전과.8범.    : Factor w/ 35 levels "-","1","1,284",...: 35 26 7 19 5 4 15 22 28 20
## $ 전과.9범이상.: Factor w/ 38 levels "-","1","1,247",...: 26 27 30 5 10 9 14 11 24 21
```

```
new_record$계 <- as.character(new_record$계)
new_record$계 <- as.numeric(gsub("-", "", new_record$계))

new_record$없음 <- as.character(new_record$없음)
new_record$없음 <- as.numeric(gsub("-", "", new_record$없음))

new_record$전과.소계. <- as.character(new_record$전과.소계.)
new_record$전과.소계. <- as.numeric(gsub("-", "", new_record$전과.소계.))

new_record <- new_record %>%
  select(범죄대분류, 범죄중분류, 계, 없음, 전과.소계.) %>%
  mutate(초범률 = (없음 / 계) * 100,
         전과자_비율 = (전과.소계. / 계) * 100)

str(new_record)
```

```
## 'data.frame':   10 obs. of  7 variables:
## $ 범죄대분류 : Factor w/ 15 levels "강력범죄","교통범죄",...: 1 1 1 1 10 13 13 13 13 11
## $ 범죄중분류 : Factor w/ 38 levels "강간","강도",...: 17 2 1 3 27 19 35 36 34 16
## $ 계         : num  379 1716 5829 16016 106415 ...
## $ 없음       : num  116 318 1633 5613 40423 ...
## $ 전과.소계. : num  196 1243 2617 8197 57388 ...
## $ 초범률     : num  30.6 18.5 28 35 38 ...
## $ 전과자_비율: num  51.7 72.4 44.9 51.2 53.9 ...
```

3. 데이터 클린징

데이터에 결측치나 이상치가 없었기 때문에 따로 데이터 클린징 작업을 하지 않았다.

```
is.na(new_record)
```

```
##      범죄대분류 범죄중분류   계   없음   전과.소계.   초범률   전과자_비율
## [1,]      FALSE      FALSE FALSE FALSE      FALSE  FALSE      FALSE
## [2,]      FALSE      FALSE FALSE FALSE      FALSE  FALSE      FALSE
## [3,]      FALSE      FALSE FALSE FALSE      FALSE  FALSE      FALSE
## [4,]      FALSE      FALSE FALSE FALSE      FALSE  FALSE      FALSE
## [5,]      FALSE      FALSE FALSE FALSE      FALSE  FALSE      FALSE
## [6,]      FALSE      FALSE FALSE FALSE      FALSE  FALSE      FALSE
## [7,]      FALSE      FALSE FALSE FALSE      FALSE  FALSE      FALSE
## [8,]      FALSE      FALSE FALSE FALSE      FALSE  FALSE      FALSE
## [9,]      FALSE      FALSE FALSE FALSE      FALSE  FALSE      FALSE
## [10,]     FALSE      FALSE FALSE FALSE      FALSE  FALSE      FALSE
```

```
View(new_record)
```

4. 데이터 시각화

데이터를 좀 더 보기 쉽게 하기위하여 초범률 데이터와 전과자 비율 데이터로 그래프를 만들었다.

```
library(ggplot2)
library(plotly)
```

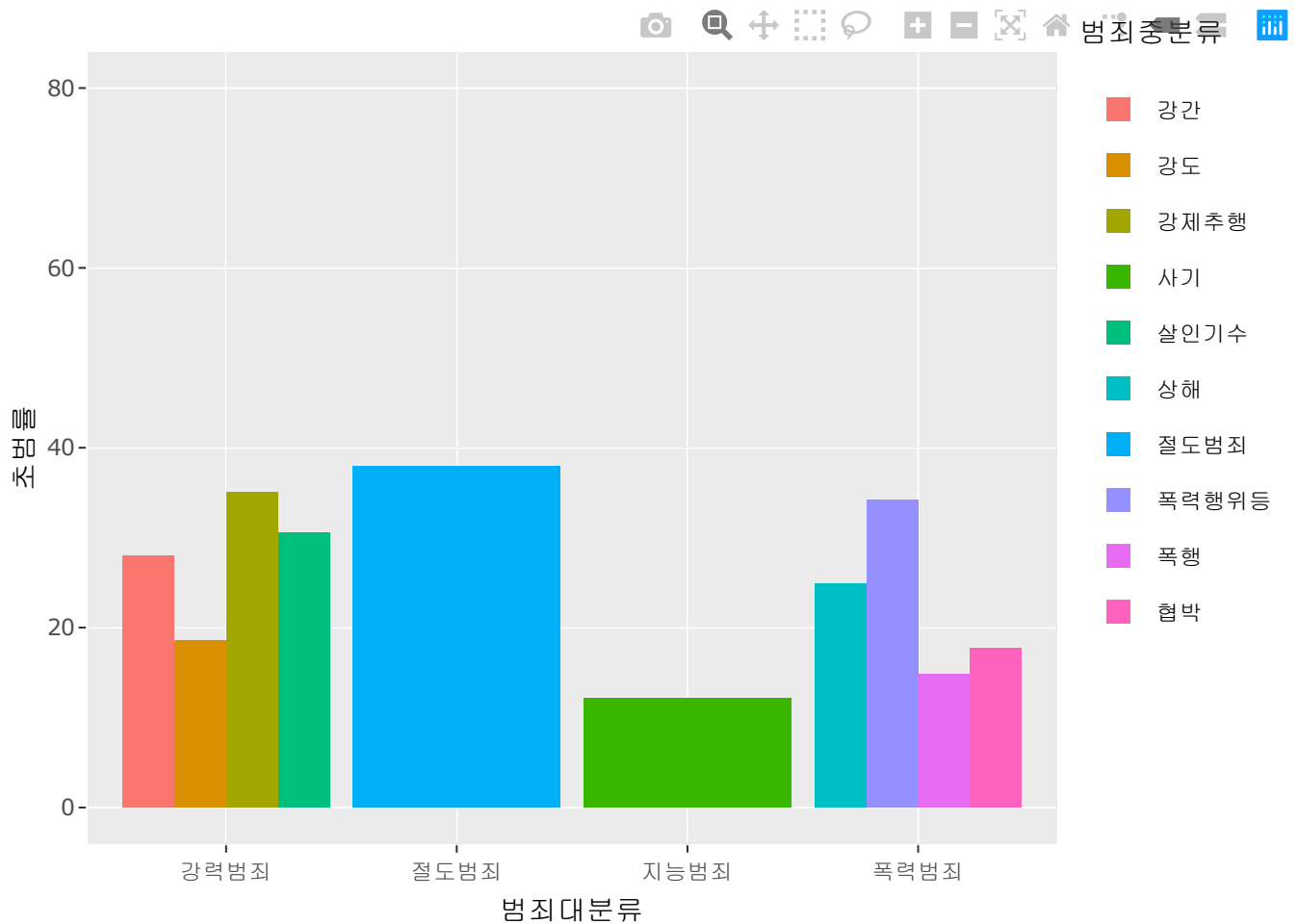
```
##
## Attaching package: 'plotly'
```

```
## The following object is masked from 'package:ggplot2':
##
##   last_plot
```

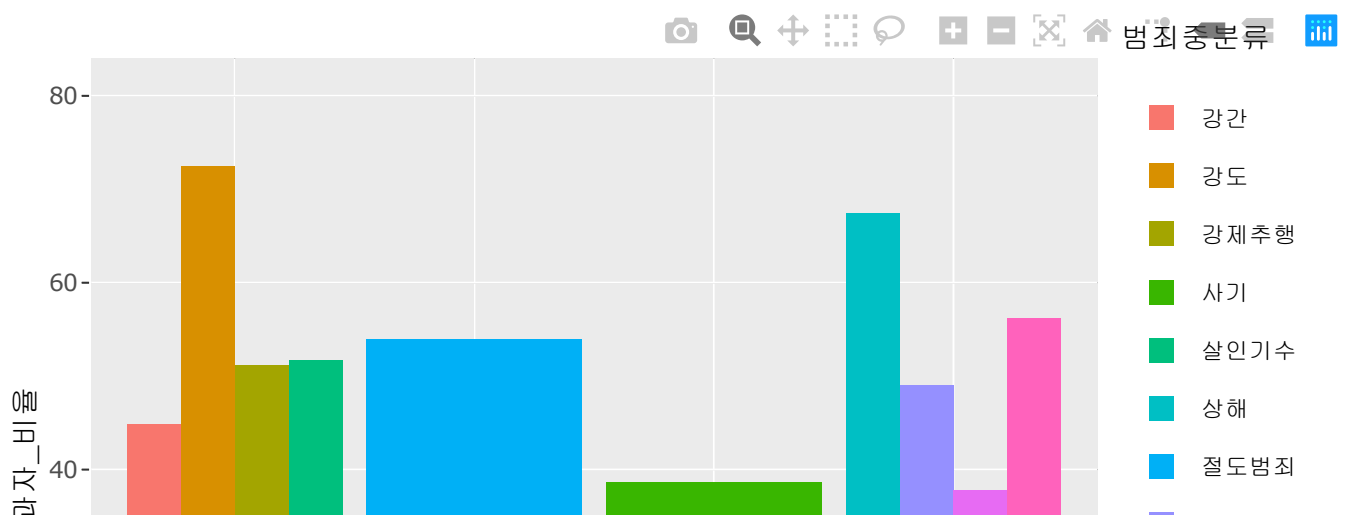
```
## The following object is masked from 'package:stats':
##
##   filter
```

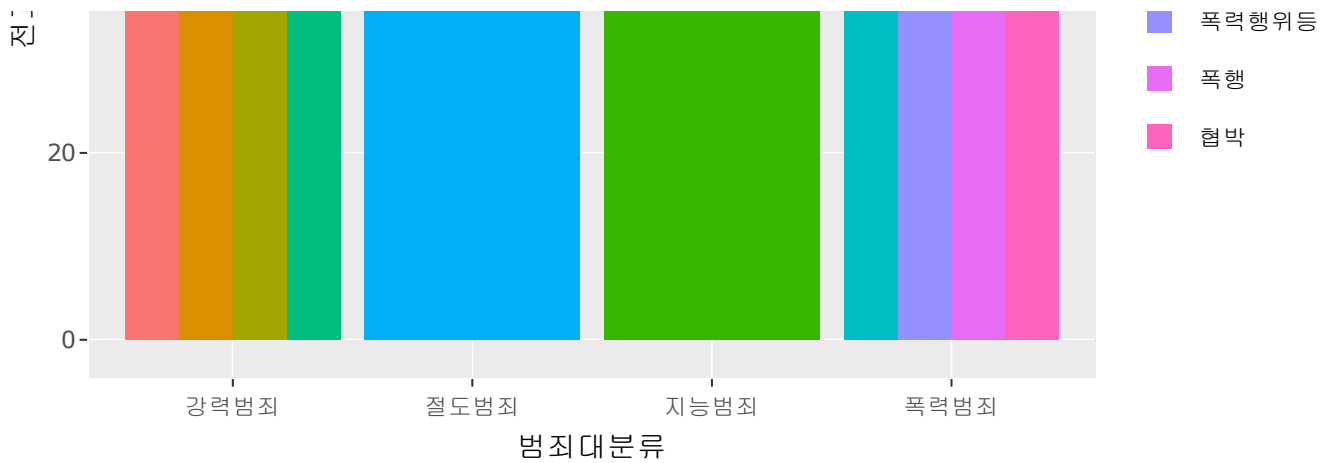
```
## The following object is masked from 'package:graphics':
##
##   layout
```

```
no_cri_record <- ggplot(data = new_record, aes(x = 범죄대분류 , y = 초범률, fill = 범죄중분류))
+ geom_col(position = "dodge") +
  ylim(0, 80)
ggplotly(no_cri_record)
```



```
have_cri_record <- ggplot(data = new_record, aes(x = 범죄대분류 , y = 전과자_비율, fill = 범죄중분류))
+ geom_col(position = "dodge") +
  ylim(0, 80)
ggplotly(have_cri_record)
```





이 그래프를 보면 10개의 범죄항목이 초범자보다는 전과자의 비율이 더 높다는 것을 알 수 있다. 즉, 초범자는 높아봐야 40% 정도이지만, 전과자의 비율은 지능범죄인 사기와 폭력범죄인 폭행을 제외하고 40%가 넘는다.

5. 의사결정 결론

```
new_record %>%
  summarise(초범률_평균 = mean(초범률),
            전과비율_평균 = mean(전과자_비율))
```

```
##   초범률_평균 전과비율_평균
## 1      25.40197      52.32416
```

그래프와 자료를 보면 초범자의 비율은 평균 25%, 전과자의 비율 평균 52%로 전과자들의 비율이 더 높았다. 즉 전과자들의 수가 더 많았다는 것이다. 전체의 경우가 아닌 비록 10개의 범죄에서만 뽑은 것이지만 필자가 뽑은 10개의 범죄는 강력 범죄와 폭행 등 매우 위험한 범죄들 즉, 중범죄 위주로 뽑은 것이므로 여기서 전과자들의 수가 많다는 것은 전과자이면서 중범죄를 저지른 경우가 많다는 것이다. 전에 저지른 범죄들이 경범죄인지 중범죄 인지는 이 데이터에서는 알 수 없지만, 이는 범죄를 저지르고 또 범죄를 저질렀다는 것을 알 수 있다. 따라서 결과에 따르면 전과가 있음에도 다시 범죄를 저지른다는 것이 더 많으므로 '사람은 변하지 않는다'는 말이 범죄에도 적용될 수 있다고 할 수 있다.

6. 본인만의 차별화 분석 포인트

여기서 범죄 10가지는 중범죄라 할 수 있는 것들을 뽑았는데 여기서 협박, 폭력행위, 사기 등은 다른 범죄에 비해서 범죄의 정도가 중한 건 아니지만, 협박은 폭행이나 이러한 중범죄의 원인이 될 수 있고, 그리고 뉴스나 신문기사들을 보면 폭력행위를 심심치 않게 볼 수 있기 때문에 이를 선택하였고, 또한 최근에 '빛투' 즉, 사기죄로 인한 피해가 세간에 많이 알려졌고, 사기로 인해 많은 사람들의 삶이 나락으로 떨어진 것을 많이 봐왔기 때문에 이들을 중범죄로 판단하여 10개의 중범죄 목록에 넣었다. 그리고 중범죄를 선택한 이유는 경범죄 보다는 중범죄의 전과자 수가 더 적을 것이라고 판단하였기 때문에 중 범죄들을 선택하였다.

그래프는 인터랙티브 그래프를 사용하였는데, 비율을 그래프로 나타내기에는 정확한 값을 보기가 어려워서 이 그래프를 선택하였다.