Final report

Yoon jung min

2019년 6월 11일

## 목적

이 보고서는 광주 광역시의 범죄 발생 데이터와 CCTV 데이터로 CCTV 설치가 범죄 예방 및 검거에 효과가 있는지 파악하는 것을 목적으로 한다. 분석 데이터는 Google maps api를 이용한 시각화를 통해 결론을 냈다. 지역을 광주 광역시로 정한 기준은, 연도별로 자세한 범죄 발생지 데이터가 있는 지역이 광주 광역시 뿐이었기 때문이다.

## 데이터 전처리

CCTV 파일데이터와 연도별 범죄 데이터의 csv 파일을 불러온다. CCTV 데이터는 2017년 설치 기준이다.

folder <- "E:/R\_Projects/Final\_proj/dataset/"  
  
CCTV\_raw <- read.csv(paste(folder, "CCTV\_20180910.csv", sep = ""), stringsAsFactors = F)  
regional\_crim\_2015\_raw <- read.csv(paste(folder, "2015\_Gwangju\_regional\_crime.csv", sep = ""), stringsAsFactors = F)  
regional\_crim\_2016\_raw <- read.csv(paste(folder, "2016\_Gwangju\_regional\_crime.csv", sep = ""), stringsAsFactors = F)  
regional\_crim\_2017\_raw <- read.csv(paste(folder, "2017\_Gwangju\_regional\_crime.csv", sep = ""), stringsAsFactors = F)  
regional\_crim\_2018\_raw <- read.csv(paste(folder, "2018\_Gwangju\_regional\_crime.csv", sep = ""), stringsAsFactors = F)

### CCTV 데이터

summary(CCTV\_raw)

## 관리기관명 소재지도로명주소 소재지지번주소   
## Length:1726 Length:1726 Length:1726   
## Class :character Class :character Class :character   
## Mode :character Mode :character Mode :character   
##   
##   
## 설치목적구분 카메라대수 카메라화소수 촬영방면정보   
## Length:1726 Min. :1.000 Min. : 41.0 Length:1726   
## Class :character 1st Qu.:2.000 1st Qu.:200.0 Class :character   
## Mode :character Median :2.000 Median :200.0 Mode :character   
## Mean :1.938 Mean :189.2   
## 3rd Qu.:2.000 3rd Qu.:200.0   
## Max. :6.000 Max. :330.0   
## NA's :70   
## 보관일수 설치년월 관리기관전화번호   
## Mode:logical Length:1726 Length:1726   
## NA's:1726 Class :character Class :character   
## Mode :character Mode :character   
##

CCTV 데이터 내에서 소재지지번주소(범주형), 설치목적구분(범주형), 카메라 대수(연속형)만 사용한다. 그 데이터에 대한 전처리는 아래와 같다.

CCTV <- CCTV\_raw %>% select(소재지지번주소, 설치목적구분, 카메라대수)

### 연도별 범죄 데이터

# 각 과정은 연도별로 반복한다.

regional\_crim\_2015 <- regional\_crim\_2015\_raw

summary(regional\_crim\_2015)

## 관서명 구분 살인

## Length:36 Length:36 Min. : 0.000

## Class :character Class :character 1st Qu.: 1.000

## Mode :character Mode :character Median : 3.000

## Mean : 3.889

## 3rd Qu.: 4.250

## Max. :18.000

##

## 강도 강간.강제추행 절도 폭력

## Min. : 0.00 Min. : 3.00 Min. : 12 Min. : 3.0

## 1st Qu.: 3.00 1st Qu.: 42.75 1st Qu.: 137 1st Qu.: 753.2

## Median : 7.00 Median : 88.50 Median : 611 Median : 1440.5

## Mean :12.06 Mean :160.36 Mean :1150 Mean : 2290.6

## 3rd Qu.:12.50 3rd Qu.:168.25 3rd Qu.:1193 3rd Qu.: 2508.5

## Max. :66.00 Max. :776.00 Max. :8425 Max. :11774.0

##

occur\_2015 <- regional\_crim\_2015 %>%

filter(구분 == "발 생 건 수" & 관서명 == "광주지방경찰청계") %>%

mutate(total = 살인 + 강도 + 강간.강제추행 + 절도 + 폭력) %>%

mutate(year = 2015)

arrest\_2015 <- regional\_crim\_2015 %>%

filter(구분 == "검 거 건 수" & 관서명 == "광주지방경찰청계") %>%

mutate(total = 살인 + 강도 + 강간.강제추행 + 절도 + 폭력) %>%

mutate(year = 2015)

raw 데이터는 기본적으로 복사한 후 사용한다. 관서명(범주형), 구분(범주형), 범죄 종류에 따른 횟수(연속형)를 전부 이용할 것이며, 합계 및 연도 정보를 추가로 저장한다. 각 발생 건수와 검거 건수에 따른 범죄 통계만 이용할 것이기 때문에 occur(발생건수)와 arrest(검거건수)로 나눈다.

## 2. 데이터 기초분석

### CCTV 데이터

CCTV$지역 <- as.factor(strtrim(CCTV$소재지지번주소, 16))

CCTV\_regional\_purpose <- CCTV %>%

group\_by(설치목적구분, 지역) %>%

summarise(total = sum(카메라대수))

# CCTV 기본 데이터 확인, 기초분석

ggplot(data = CCTV\_regional\_purpose,

aes(x = 지역, y = total, fill = 설치목적구분)) + geom\_col()

## 

기본적으로 광산구와 북구에 설치가 많이 되어있으며 동구는 상대적으로 적음을 알 수 있다. CCTV는 생활 방범용이 가장 많이 설치되어 있다.

### 연도별 범죄 데이터

occur\_arrest\_per\_year <- data.frame(year = c(2015, 2016, 2017, 2018),

dist = c('occur','occur','occur','occur',

'arrest','arrest','arrest','arrest'),

total = c(occur\_2015$total,

occur\_2016$total, occur\_2017$total,

occur\_2018$total, arrest\_2015$total,

arrest\_2016$total, arrest\_2017$total,

arrest\_2018$total))

ggplot(data = occur\_arrest\_per\_year, aes(x = year, y = total, col = dist)) + geom\_line()

지도, 텍스트이(가) 표시된 사진

자동 생성된 설명

2017년에 CCTV가 대량으로 설치되면서, 범죄율과 검거율 사이의 간격이 좁혀 졌음을 볼 수 있다. 범죄율이 꾸준히 하락함 또한 확인할 수 있다.

## 데이터 시각화

# 주소 값을 이용, 위도와 경도를 받아서 새로운 데이터에 넣는다.

cctv\_pos <- CCTV$소재지지번주소

gc <- geocode(enc2utf8(cctv\_pos))

df <- data.frame(name=CCTV$설치목적구분,

lon=gc$lon,

lat=gc$lat)

# 데이터 클린징

df\_nomiss <- df %>% filter(!is.na(df$lon) & !is.na(df$lat))

# 각 범위별로 나누어 따로 저장 및 위도 경도 지정

df\_nomiss\_child <- df\_nomiss %>% filter(df\_nomiss$name == "어린이보호")

df\_nomiss\_security <- df\_nomiss %>% filter(df\_nomiss$name == "생활방범")

df\_nomiss\_disaster <- df\_nomiss %>% filter(df\_nomiss$name == "재난재해")

gc\_child <- df\_nomiss\_child %>% select(lon, lat)

gc\_security <- df\_nomiss\_security %>% select(lon, lat)

gc\_disaster <- df\_nomiss\_disaster %>% select(lon, lat)

# 지도에 표시될 범위의 중앙 지정

cen <- c(mean(df\_nomiss$lon),mean(df\_nomiss$lat))

# google map

qmap(location = cen,

zoom = 12,

maptype = 'roadmap',

source = 'google') +

geom\_point(data = gc\_child,

aes(x = lon, y = lat),

shape = '▼',

color = 'yellow',

size = 3) +

geom\_point(data = gc\_security,

aes(x = lon, y = lat),

shape = '▼',

color = 'red',

size = 2) +

geom\_point(data = gc\_disaster,

aes(x = lon, y = lat),

shape = '▼',

color = 'blue',

size = 3) +

geom\_point(data = region\_crim\_occur\_nomiss,

aes(x = lon, y = lat, size = total, stroke = 5),

color = 'darkblue',

alpha = 0.5)

텍스트, 지도이(가) 표시된 사진

자동 생성된 설명

위의 과정은 google map api를 이용하여 데이터의 위치 정보를 받아오고, 지도에 표시하는 것을 나타냈다. 삼각형은 각각 어린이 보호(노란색), 생활 방범(빨간색), 재난 재해(파란색)역할을 하는 CCTV를 의미한다. 파란색 동그라미는 2018년의 범죄율을 의미하며 크기가 클수록 범죄가 많이 일어났음을 의미한다.

## 의사결정 결론

위의 결과를 통해 CCTV가 비교적 적은 지역은 범죄가 많이 일어남을 확인할 수 있다. 2018년도의 검거율은 큰 의미가 없기 때문에 생략하였다.

## 차별화 분석 포인트

1. CCTV데이터와 범죄 데이터 2가지를 이용하여 분석하였다. 데이터 상에서의 연관 관계는 크지 않았기 때문에 시각화를 통해서 의사결정 결론을 내릴 수 있었다.
2. Google map api 사용을 통해 지도를 직접 불러와서 표시하였다. 실제로 Google cloud service로부터 키 값을 받고, 직접 로드 하는데 있어서 시간을 많이 할애하였다. 데이터 시각화에 주력하는 부분에서 차별화를 진행하였다.
3. 클린징은 따로 할 필요가 거의 없는 데이터이기 때문에 개별 구역에서 진행 후 주석을 달아주었다.