

# FINAL PROJECT

5조

김민정, 김윤전, 박대한,  
이재상, 최은성

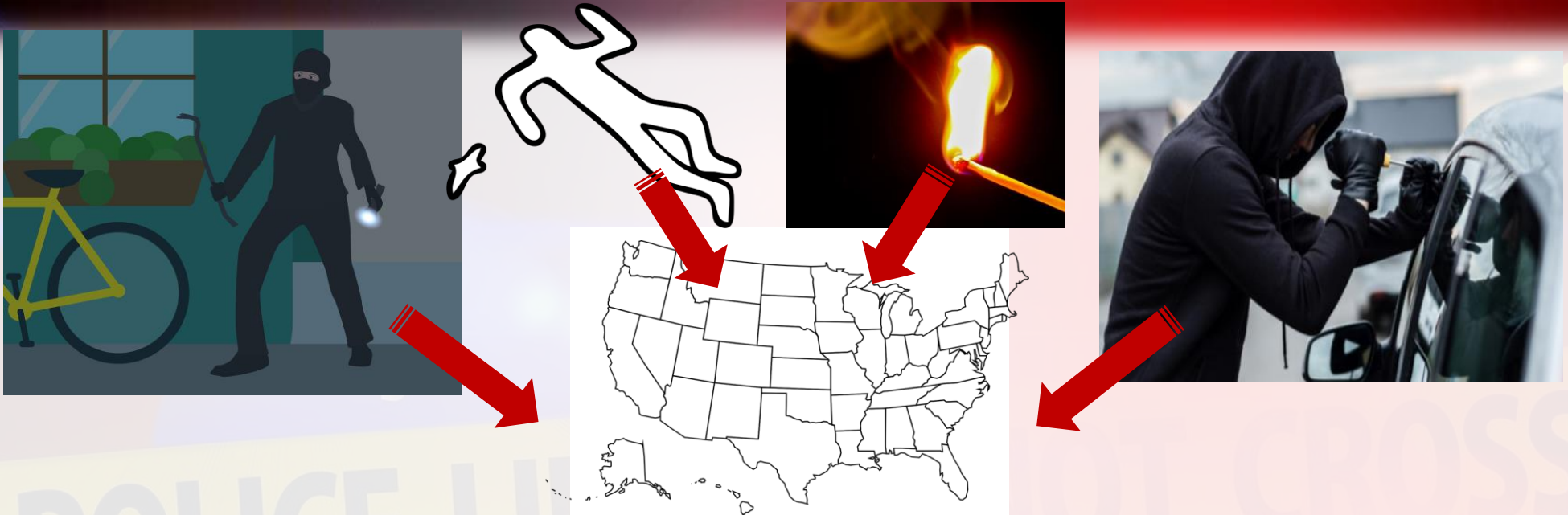
# CONTENTS

- ROLES & RESPONSIBILITIES
- PURPOSE OF THE PROJECT
- DESCRIPTION OF THE DATA
- EDA
- FURTHER PLANS

# ROLES & RESPONSIBILITIES

- **김민정(조장)**
  - 발표 자료 준비 및 발표
- **김윤전**
  - 인터랙티브 시각화
- **박대한**
  - EDA
- **이재상**
  - EDA
- **최은성**
  - 발표 자료 준비

# PURPOSE OF THE PROJECT



- **베이지안 선형회귀 모형으로 미국 범죄지도를 완성하자**
- **인터랙티브 시각화: Missing value가 채워진 모드의 지역별 시각화와 없는 모드의 시각화를 그리자**
- **베이지안 method를 사용하여 결측치를 채우자**



# DESCRIPTION OF THE DATA

- 2215 observations and 147 variables in crimedata.csv

```
'data.frame': 2215 obs. of 147 variables:
 $ communityname : Factor w/ 2018 levels "Aberdeencity",...: 151 1035 1781 665 141 1700 1272 41 566 1860 ...
 $ state          : Factor w/ 48 levels "AK","AL","AR",...: 29 36 35 32 23 24 19 15 27 41 ...
 $ countyCode     : Factor w/ 115 levels "?","1","101",...: 57 60 1 55 84 1 46 1 40 1 ...
 $ communityCode  : Factor w/ 960 levels "?","100","1000",...: 511 426 1 215 473 1 468 1 177 1 ...
 $ fold           : int 1 1 1 1 1 1 1 1 1 1 ...
 $ population     : int 11980 23123 29344 16656 11245 140494 28700 59459 74111 103590 ...
 $ householdsSize : num 3.1 2.82 2.43 2.4 2.76 2.45 2.6 2.45 2.46 2.62 ...
 $ racePctBlack   : num 1.37 0.8 0.74 1.7 0.53 ...
 $ racePctWhite   : num 91.8 95.6 94.3 97.3 89.2 ...
 $ racePctAsian   : num 6.5 3.44 3.43 0.5 1.17 0.9 1.47 0.4 1.25 0.92 ...
 $ racePctHispan : num 1.88 0.85 2.35 0.7 0.52 ...
 $ agePct12t21    : num 12.5 11 11.4 12.6 24.5 ...
 $ agePct12t29    : num 21.4 21.3 25.9 25.2 40.5 ...
 $ agePct16t24    : num 10.9 10.5 11 12.2 28.7 ...
 $ agePct65up     : num 11.3 17.2 10.3 17.6 12.6 ...
 $ numUrban       : int 11980 23123 29344 0 0 140494 28700 59449 74115 103590 ...
 $ pctUrban       : num 100 100 100 0 0 100 100 100 100 100 ...
 $ medIncome      : int 75122 47917 35669 20580 17390 21577 42805 23221 25326 17852 ...
 $ pctWage        : num 89.2 79 82 68.2 69.3 ...
 $ pctFarmSelf    : num 1.55 1.11 1.15 0.24 0.55 1 0.39 0.67 2.93 0.86 ...
 $ pctWInvInc     : num 70.2 64.1 55.7 39 42.8 ...
 $ pctWSocSec     : num 23.6 35.5 22.2 39.5 32.2 ...
 $ pctWPubAsst    : num 1.03 2.75 2.94 11.71 11.21 ...
 $ pctWRetire     : num 18.4 22.9 14.6 18.3 14.4 ...
 $ medFamInc      : int 79584 55323 42112 26501 24018 27705 50394 28901 34269 24058 ...
 $ perCapInc      : int 29711 20148 16946 10810 8483 11878 18193 12161 13554 10195 ...
 $ whitePerCap    : int 30233 20191 17103 10909 9009 12029 18276 12599 13727 12126 ...
 $ blackPerCap    : int 13600 18137 16644 9984 887 7382 17342 9820 8852 5715 ...
 $ indianPerCap   : int 5725 0 21606 4941 4425 10264 21482 6634 5344 11313 ...
 $ AsianPerCap    : int 27101 20074 15528 3541 3352 10753 12639 8802 8011 5770 ...
 $ OtherPerCap    : Factor w/ 1918 levels "?","0","10000",...: 1022 1049 1174 717 784 1418 681 1460 1068 1445 ...
```

# DESCRIPTION OF THE DATA

## • Major Variable description (1/3)

### <지역코드>

communityname  
state  
countyCode  
communityCode

### <총인구수>

population

### <가구별크기>

householdsize

### <인종비율>

racepctblack  
racePctWhite  
racePctAsian  
racePctHisp

### <연령대>

agePct12t21  
agePct12t29  
agePct16t24  
agePct65up

### <도시화>

numbUrban  
pctUrban

### <소득>

medIncome  
pctWWage  
pctWFarmSelf  
pctWInvInc  
pctWSocSec  
pctWPubAsst  
pctWRetire  
medFamInc  
perCapInc

### <인종별 소득>

whitePerCap  
blackPerCap  
indianPerCap  
AsianPerCap  
OtherPerCap  
HispPerCap

### <빈곤>

NumUnderPov  
PctPopUnderPov

# DESCRIPTION OF THE DATA

## • Major Variable description (2/3)

### <교육>

PctLess9thGrade  
PctNotHSGrad  
PctBSorMore

### <고용>

PctUnemployed  
PctEmploy  
PctEmplManu  
PctEmplProfServ  
PctOccupManu  
PctOccupMgmtProf

### <결혼/가정>

MalePctDivorce  
MalePctNevMarr  
FemalePctDiv  
TotalPctDiv  
PersPerFam  
PctFam2Par  
PctKids2Par  
PctYoungKids2Par  
PctTeen2Par  
PctWorkMomYoungKids  
PctWorkMom  
NumKidsBornNeverMar  
PctKidsBornNeverMar

### <이민자>

NumImmig  
PctImmigRecent  
PctImmigRec5  
PctImmigRec8  
PctImmigRec10  
PctRecentImmig  
PctReclImmig5  
PctReclImmig8  
PctReclImmig10  
PctSpeakEnglOnly  
PctNotSpeakEnglWell

### <부동산관련>

PctLargHouseFam  
PctLargHouseOccup  
PersPerOccupHous  
PersPerOwnOccHous  
PersPerRentOccHous  
PctPersOwnOccup  
PctPersDenseHous  
PctHousLess3BR  
MedNumBR  
HousVacant  
PctHousOccup  
PctHousOwnOcc  
PctVacantBoarded

...

# DESCRIPTION OF THE DATA

## • Major Variable description (3/3)

### <지역토박이>

PctForeignBorn  
PctBornSameState  
PctSameHouse85  
PctSameCity85  
PctSameState85

### <기타>

LemasSwornFT  
LemasSwFTPerPop  
LemasSwFTFieldOps  
LemasSwFTFieldPerPop  
LemasTotalReq  
LemasTotReqPerPop

### <경찰>

PolicReqPerOffic  
PolicPerPop  
RacialMatchCommPol  
PctPolicWhite  
PctPolicBlack  
PctPolicHisp  
PctPolicAsian  
PctPolicMinor  
OfficAssgnDrugUnits  
NumKindsDrugsSeiz  
PolicAveOTWorked

...

### <범죄관련>

murders  
murdPerPop  
rapes  
rapesPerPop  
robberies  
robberPerPop  
assaults  
assaultPerPop  
burglaries  
burglPerPop  
larcenies  
larcPerPop  
autoTheft

autoTheftPerPop  
arsons  
arsonsPerPop  
ViolentCrimesPerPop  
nonViolPerPop



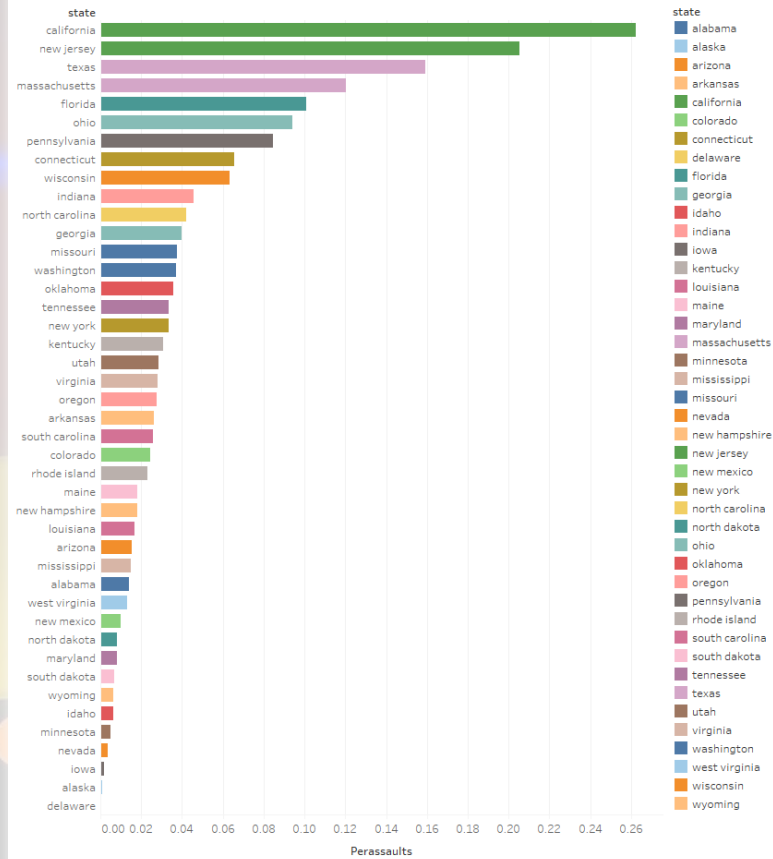
# EDA: MISSING VALUE IMPUTATION

- Total 2215 observations
- ✕ 변수 중 22개 변수에서 결측치 약 84%로 확인
  - 해당 항목의 사람수가 0이거나
  - %의 경우는 너무 적어 0으로 나오는 경우도 발생함
  - 따라서 22개 변수는 삭제함
- 그 외에는 Y 변수임

LemasSwornFT	LemasSwFTPerPop	LemasSwFTFieldOps
0.8451467	0.8451467	0.8451467
LemasSwFTFieldPerPop	LemasTotalReq	LemasTotReqPerPop
0.8451467	0.8451467	0.8451467
PolicReqPerOffic	PolicPerPop	RacialMatchCommPol
0.8451467	0.8451467	0.8451467
PctPolicWhite	PctPolicBlack	PctPolicHisp
0.8451467	0.8451467	0.8451467
PctPolicAsian	PctPolicMinor	officAssgnDrugUnits
0.8451467	0.8451467	0.8451467
NumKindsDrugsSeiz	PolicAveOTworked	PolicCars
0.8451467	0.8451467	0.8451467
PolicOperBudg	LemasPctPolicOnPatr	LemasGangUnitDeploy
0.8451467	0.8451467	0.8451467
PolicBudgPerPop		
0.8451467		

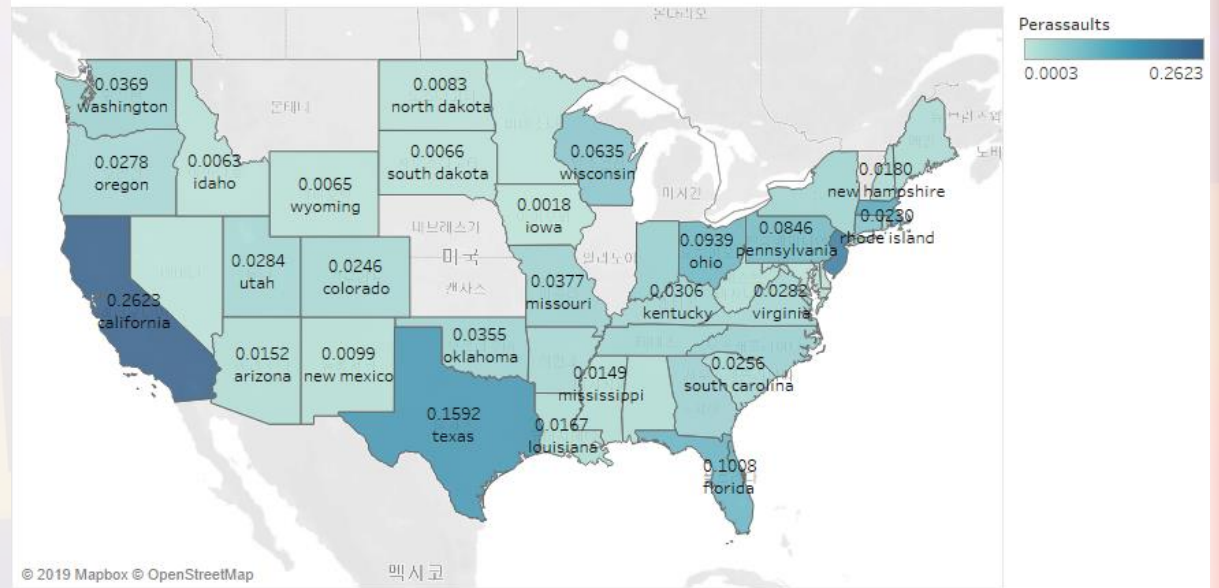
# EDA: VISUALIZATION

<Assault>



각 state(으)로 Perassaults의 합계입니다. 색상이 state의 세부 정보를 표시합니다. NA(들) 제외하는 state에 대한 수가 필터링되었습니다.

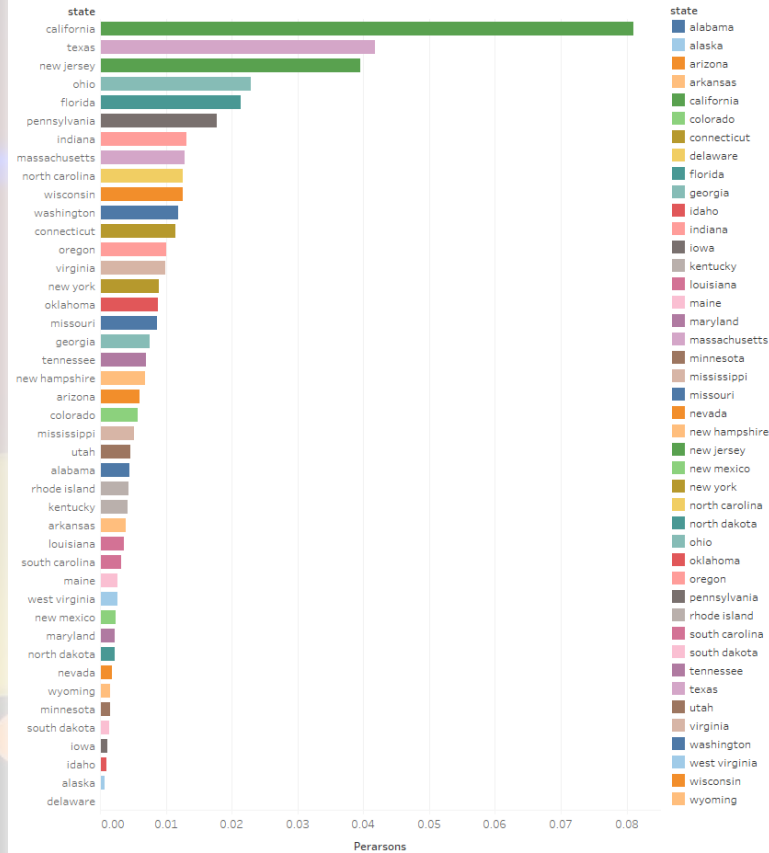
<Assault: 폭행>



경도(생성됨) 및 위도(생성됨)를 기반으로 하는 맵입니다. 색상이 Perassaults의 합계를(들) 표시합니다. 마크에 레이블이 Perassaults의 합계 및 state에 의해 지정되었습니다. state의 세부 정보가 표시되었습니다. alaska(들) 제외하는 state에 대한 수가 필터링되었습니다.

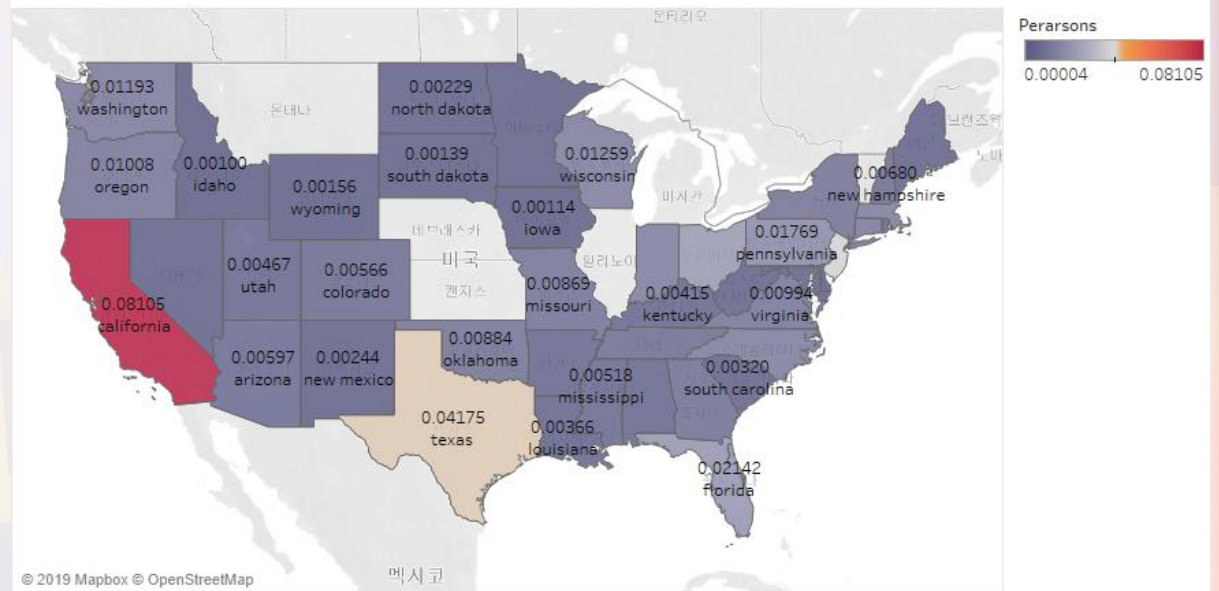
# EDA: VISUALIZATION

<방화>



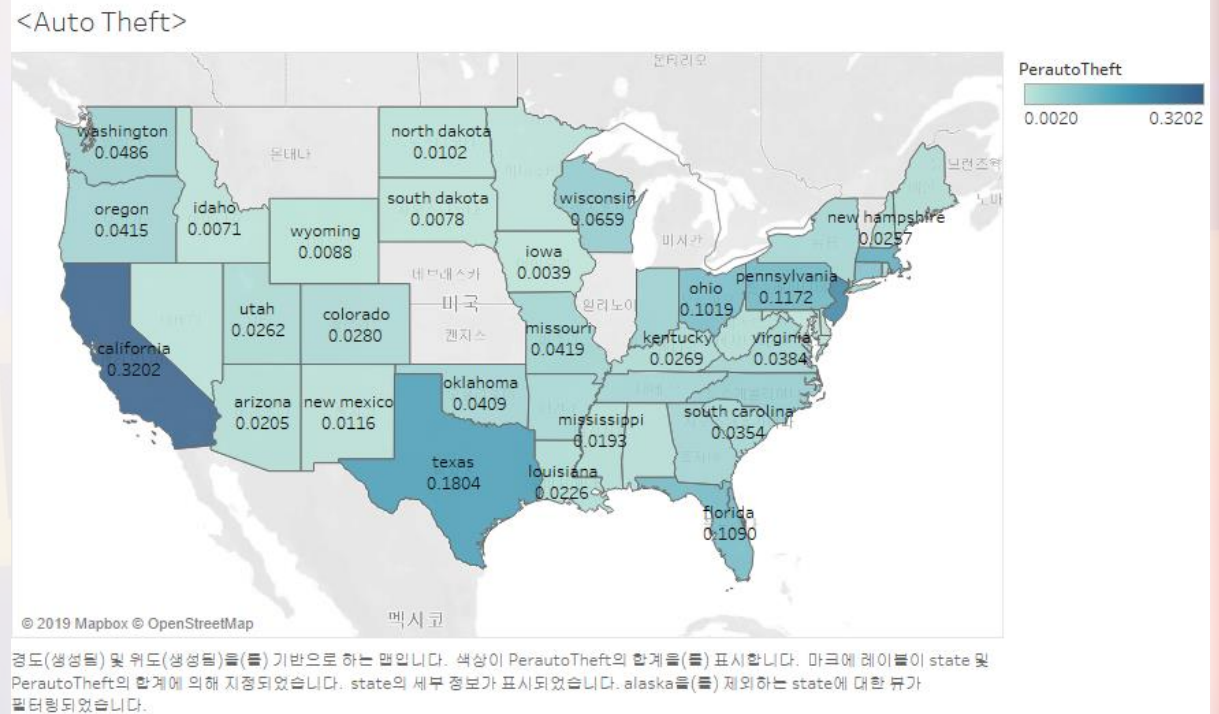
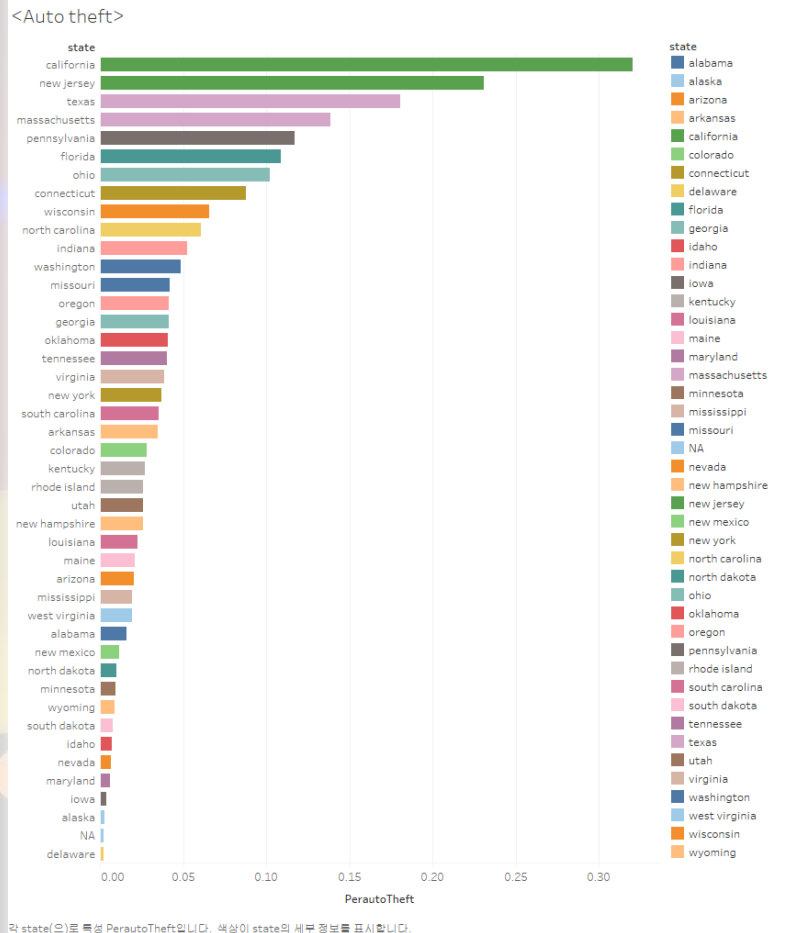
각 state(으)로 특정 Perarsons입니다. 색상이 state의 세부 정보를 표시합니다. NA(들) 제외하는 state에 대한 평가 필터링되었습니다.

<Arson>



경도(생성됨) 및 위도(생성됨)를 기반으로 하는 맵입니다. 색상이 Perarsons의 합계(들) 표시합니다. 마크에 레이블이 Perarsons의 합계 및 state에 의해 지정되었습니다. state의 세부 정보가 표시되었습니다. alaska(들) 제외하는 state에 대한 평가 필터링되었습니다.

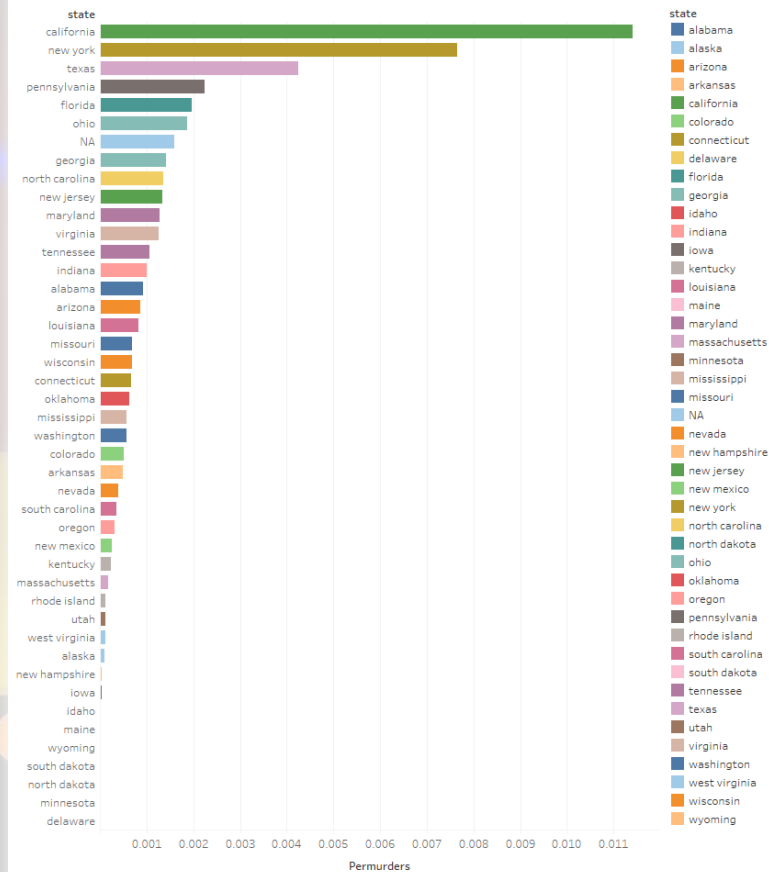
# EDA: VISUALIZATION





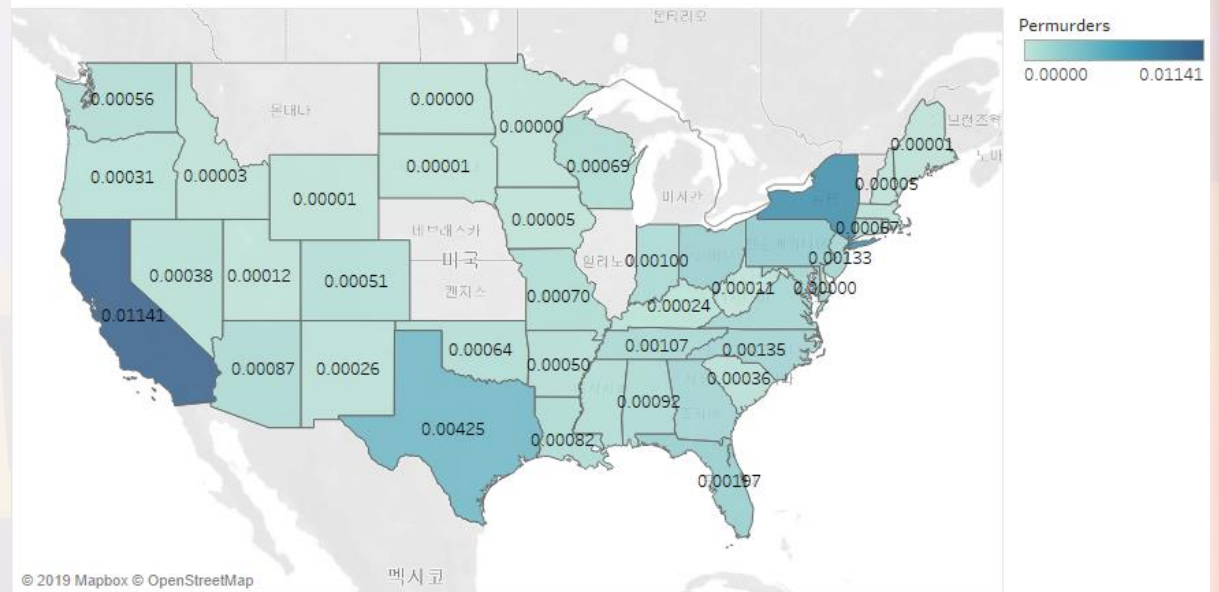
# EDA: VISUALIZATION

<살인>



각 state(으)로 특정 Permurders입니다. 색상이 state의 세부 정보를 표시합니다.

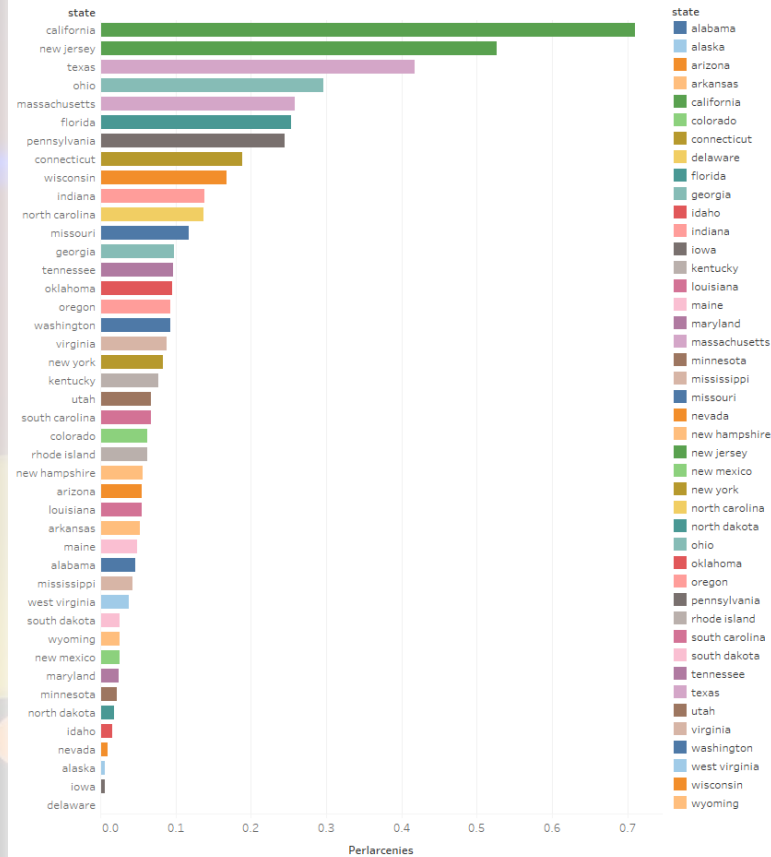
<Murder>



경도(생성됨) 및 위도(생성됨)를 기반으로 하는 맵입니다. 색상이 Permurders의 합계를(=) 표시합니다. 마크에 레이블이 Permurders의 합계에 의해 지정되었습니다. state의 세부 정보가 표시되었습니다. alaska를(=) 제외하는 state에 대한 수가 필터링되었습니다.

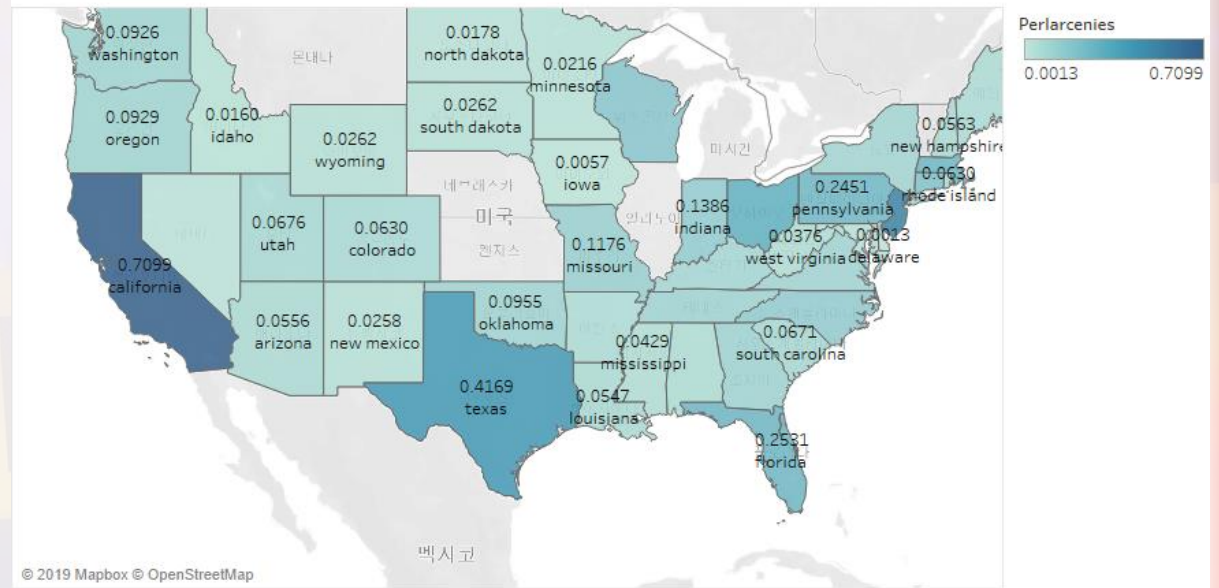
# EDA: VISUALIZATION

<Larcenies>



각 state(으)로 Perlarcenies의 합계입니다. 색상이 state의 세부 정보를 표시합니다. NA를 제외하는 state에 대한 부가 필터링되었습니다.

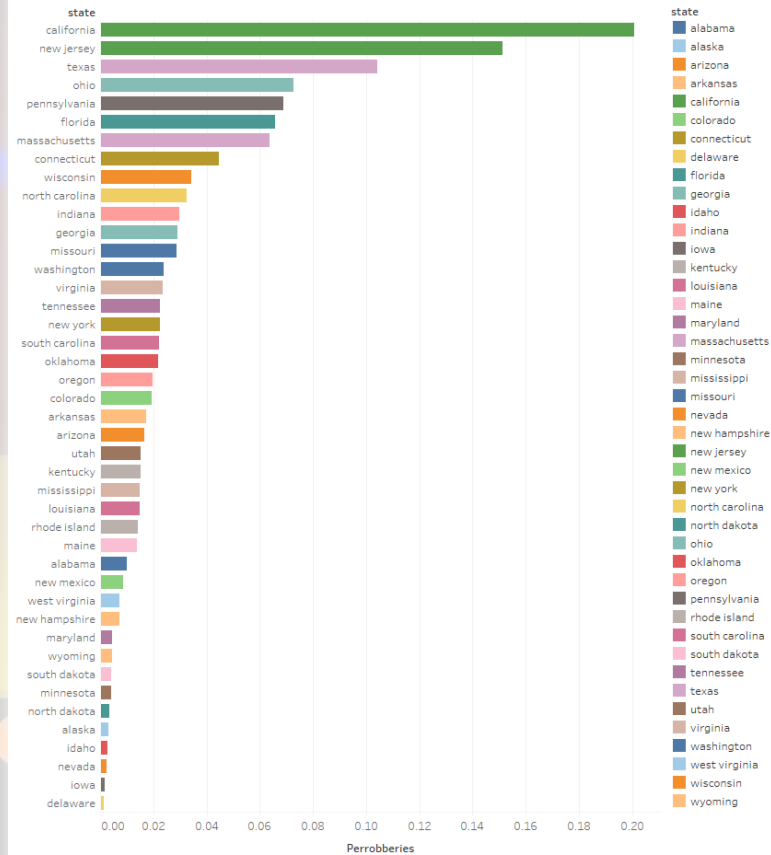
<Larcenies:절도>



경도(생성됨) 및 위도(생성됨)를 기반으로 하는 맵입니다. 색상이 Perlarcenies의 합계를 표시합니다. 마크에 레이블이 Perlarcenies의 합계 및 state에 의해 지정되었습니다. state의 세부 정보가 표시되었습니다. alaska를 제외하는 state에 대한 부가 필터링되었습니다.

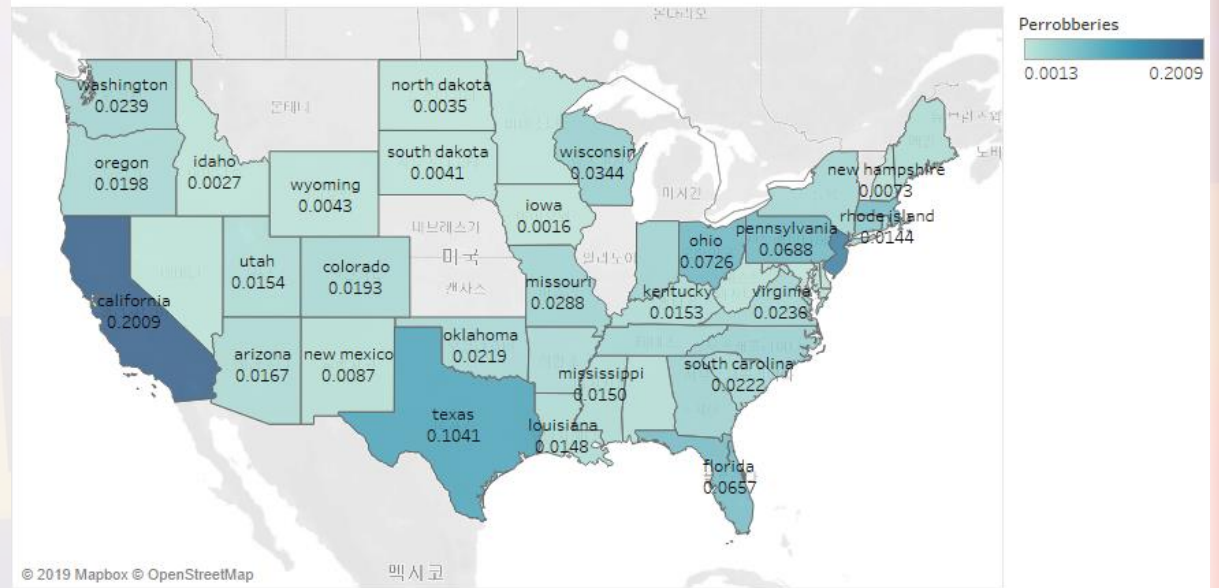
# EDA: VISUALIZATION

<Robberies:강도>



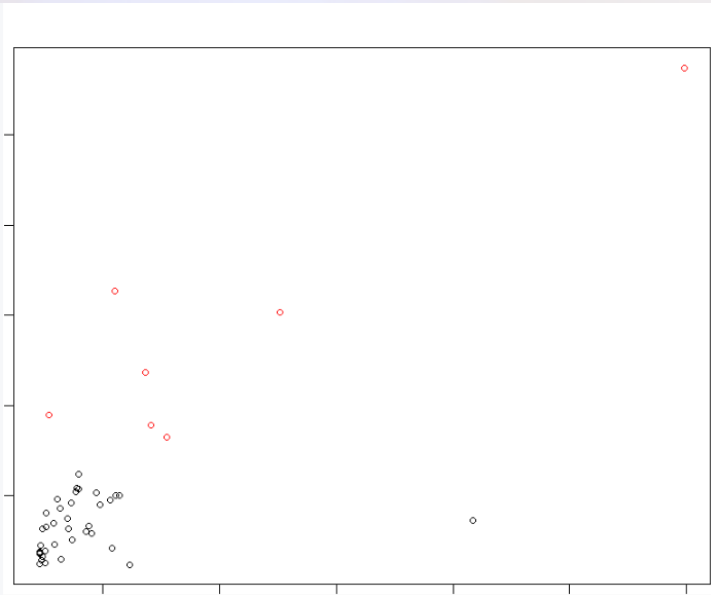
각 state(은)로 Perrobberies의 한계입니다. 색상이 state의 세부 정보를 표시합니다. NA(들) 제외하는 state에 대한 수가 필터링되었습니다.

<Robberies:강도>



경도(생성됨) 및 위도(생성됨)를 기반으로 하는 맵입니다. 색상이 Perrobberies의 한계를(들) 표시합니다. 마크에 레이블이 state 및 Perrobberies의 한계에 의해 지정되었습니다. state의 세부 정보가 표시되었습니다. alaska(들) 제외하는 state에 대한 수가 필터링되었습니다.

# EDA



- K-means clustering을 통해 오른쪽과 같이 결과가 도출됨
- 시각화를 통해 범칙율이 높게 나온 도시들의 결과가 비슷함

```
# A tibble: 2 x 9
  kmeans Permurders Perrapes Perrobberies Perassaults Perburglaries Perlarcenies PerautoTheft Perarsons
  <dbl>   <dbl>   <dbl>   <dbl>   <dbl>   <dbl>   <dbl>   <dbl>   <dbl>
1     1 0.000710 0.00577 0.0152 0.0221 0.0358 0.0625 0.0270 0.00531
2     2 0.00332 0.0344 0.104 0.147 0.243 0.386 0.171 0.0339
```



# FURTHER PLANS

- 다른 추가 정보를 이용한 변수 생성
- 파생변수 추가 고려:
  - State별 뿐 아니라 county별 분석
- 사용할 모델
  - Linear regression
  - Generalized linear model
  - Hierarchical model 등

THANK YOU

POLICE LINE DO NOT CROSS