

---

# 시카고의 범죄 기록 및 그 외 지역 데이터를 활용한 우범 지역 예측

---

이승용, 유건식

# TABLE OF CONTENTS



## PROJECT MOTIVATION

## DATA & PREPROCESSING



## EXPIORATION

## MINING & ANALYSIS



## PERFORMANCE

## DISCUSSION



# PROJECT MOTIVATION

OTHER STUDIES OF  
CRIME FACTORS

근거리 환경 특성 고려 X

ANALYSIS USING  
SPATIAL UNITS[1]



사회통계자료 사용 X

사회통계와 근거리 환경을  
모두 사용한 범죄 예측 모형

[1] W Bernasco and R Block. Robberies in Chicago: A BlockLevel Analysis of the Influence of Crime Generators, Crime Attractors, and Offender Anchor Points.

# DATA: DATASET

[data.cityofchicago.org](https://data.cityofchicago.org)

## STATISTICAL DATA

## MAIN DATA

## SPATIAL DATA

### Crime Record

- **Case Number**

- Date
- Block
- IUCR(범죄 대분류 + 소분류)
- Location Description
- Arrest(체포여부)
- Domestic
- District(세부지역)
- Ward(순찰지역)
- **Community Area(지역)**
- **Latitude & Longitude**

### SocioEco Indicators

- **Community Area Number**

- Community Area Name
- % Housing Crowded
- % Households Poverty
- % Aged 16+ Unemployed
- % Aged 25+ No High School Diploma
- % Aged -18 or 64+
- % Capital Income
- Hardship Index

### Police Station

- **Office Code**

- Office Name
- **Latitude & Longitude**

### Vacant Building

- **Docket Number**

- Violation Number
- Issued Date
- Last Hearing Date
- **Property Address**
- Violation Type
- Entity or Person(s)
- Disposition Description

### Abandoned Vehicles

- **A.V. ID**

- Creation Date
- Status
- Completion Date
- Service Request Number
- How Many Days
- XY Coordinate
- Community Area
- **Latitude & Longitude**
- Location

### Liquor Shop

- **Shop Number**

- Address
- License Code
- License Description
- License Number
- Application Type
- **Latitude & Longitude**

# DATA: DATASET

[data.cityofchicago.org](http://data.cityofchicago.org)

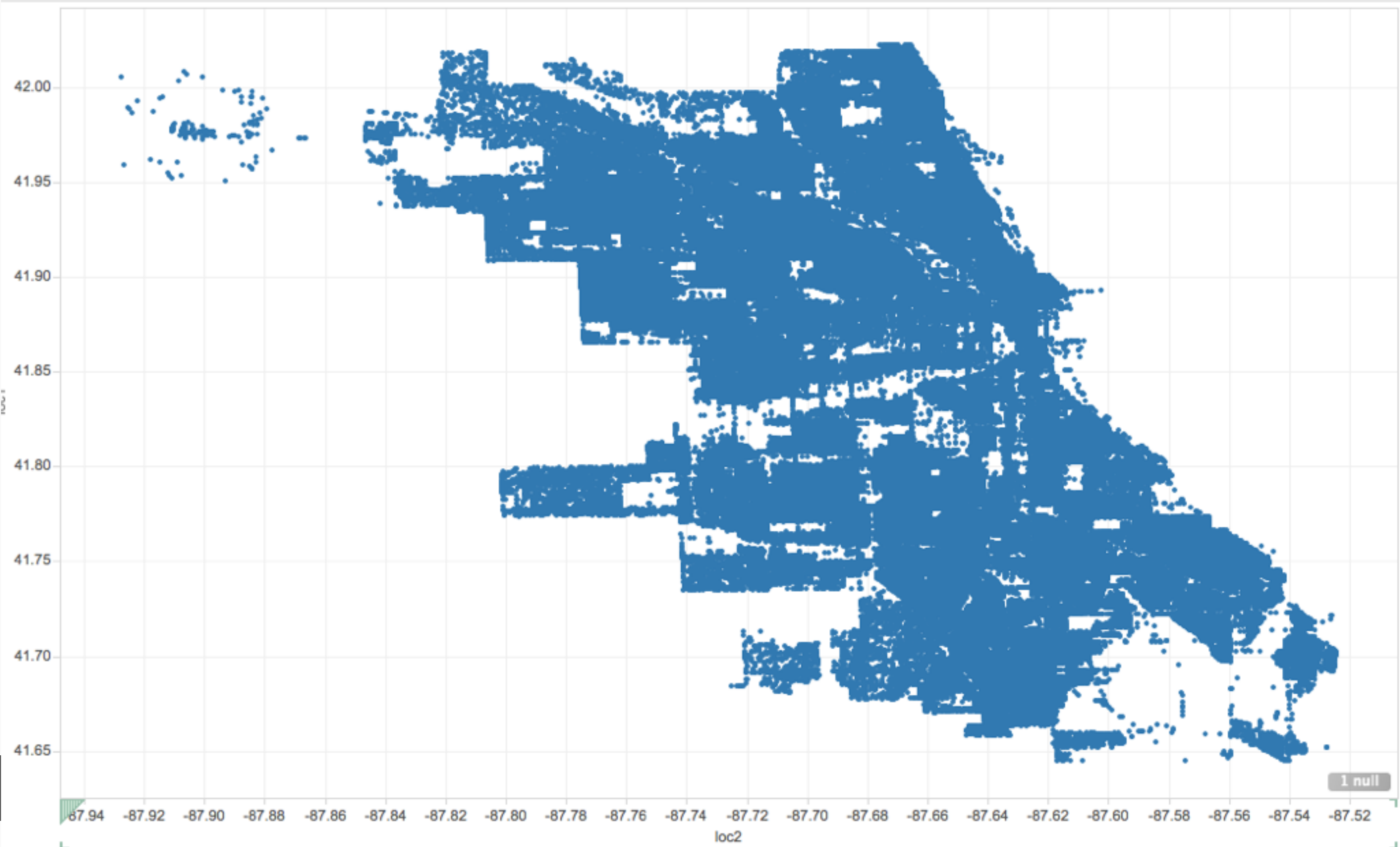
## Crime Record

### • Index

- Date
- CRIME TYPE(대분류)
- Arrest(체포여부)
- **Community Area(지역)**
- **Latitude & Longitude**

2012년 1월 1일 ~ 2012년 12월 31일  
총 335,498 건

	A	B	C	D	E	F
1	Date	Primary Type	Arrest	Community Area	Location	
2	06/01/2012 12:00:00 PM	OTHER OFFENSE	FALSE	23	(41.894535834	-87.709114061)
3	09/13/2012 07:00:00 PM	BATTERY	TRUE	6	(41.933216923	-87.63317991)
4	01/16/2012 10:56:00 AM	ARSON	FALSE	71	(41.733172334	-87.661918574)
5	04/29/2012 02:30:00 AM	ARSON	FALSE	68	(41.763828232	-87.635992227)
6	02/13/2012 07:30:00 AM	CRIM SEXUAL ASSAULT	TRUE	67	(41.779913331	-87.67042849)
7	04/07/2012 10:37:00 PM	BATTERY	FALSE	63	(41.801222426	-87.68674604)
8	12/14/2012 03:10:00 PM	CRIM SEXUAL ASSAULT	TRUE	60	(41.845217963	-87.648288405)

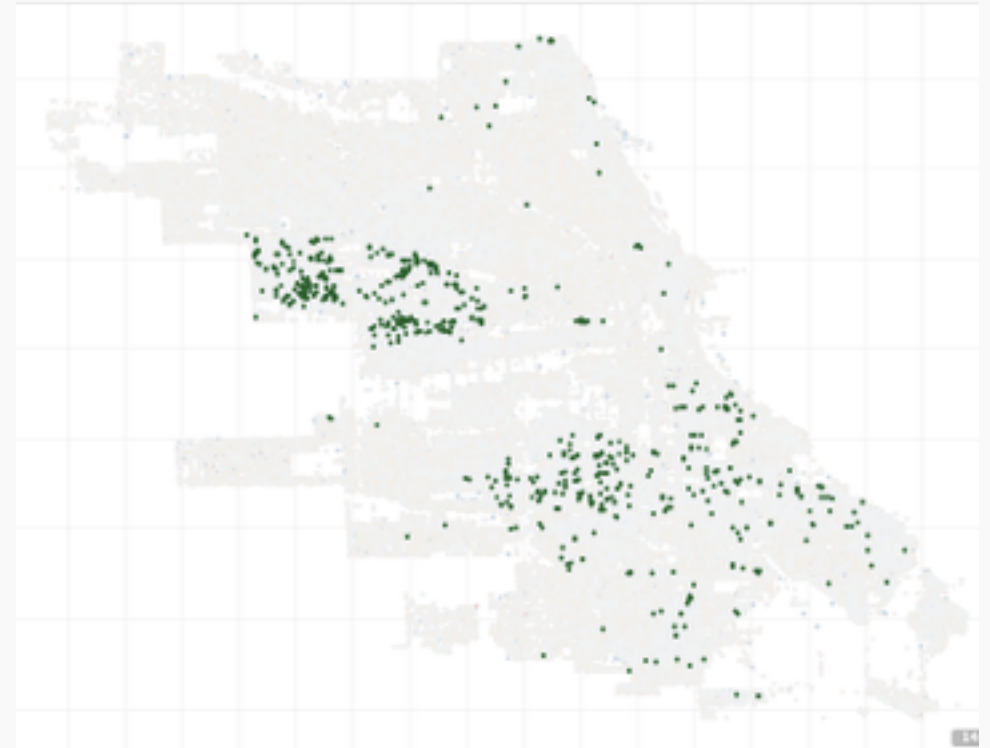




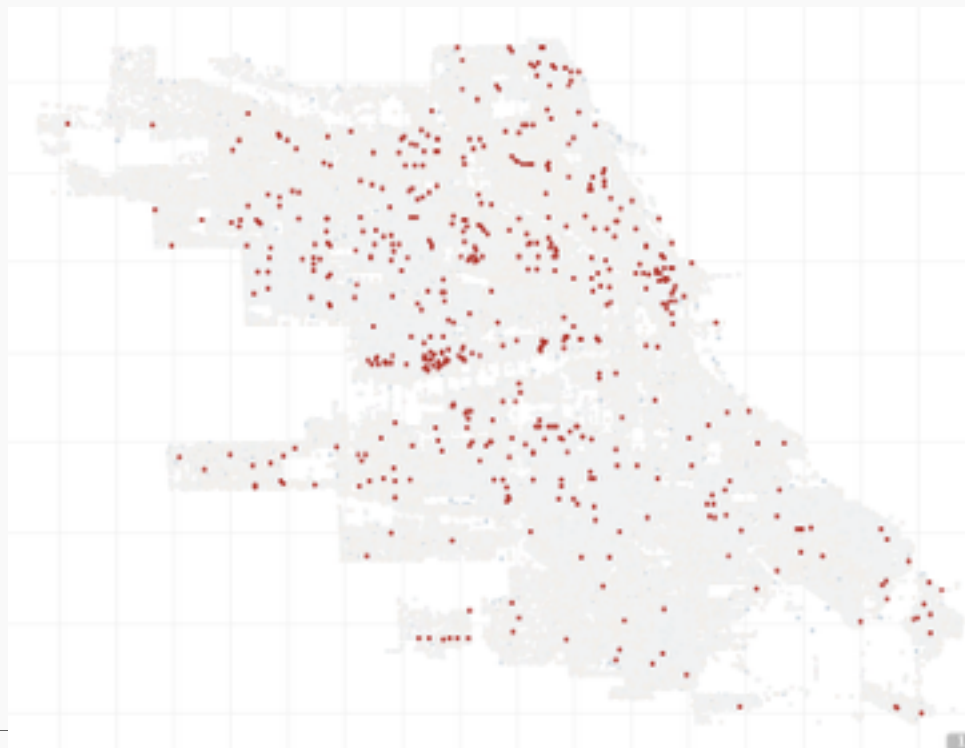
**CRIMINAL DAMAGE: 430,208건**



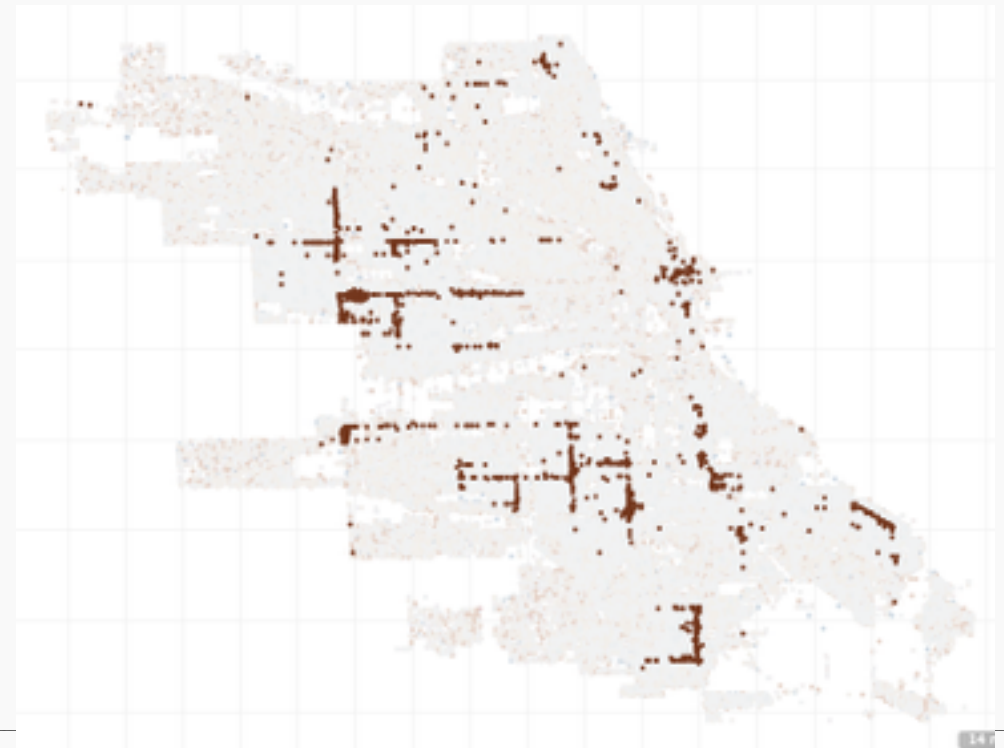
**GAMBLING: 8,688건**



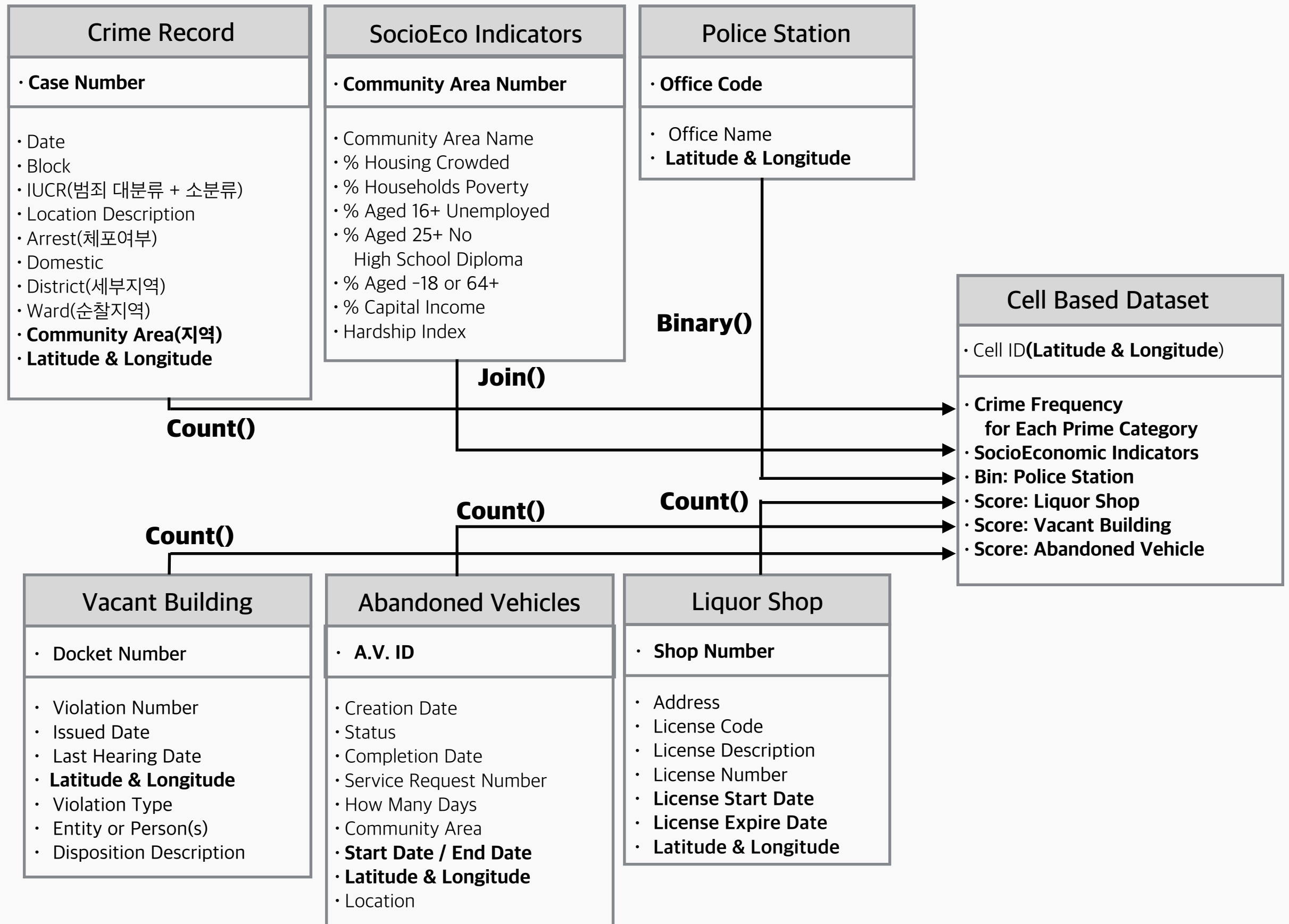
**LIQUOR LAW VIOLATION: 6,876건**



**PROSTITUTION**



# DATA: PREPROCESSING





# DATA: CELL BASED DATASET

{		
	Longitude/Latitude MID POINT	(Float, Float)
	Crime Frequency for Each Prime Category	(Integer Array)
	- Total # of Crime	
	- Prostitution(매춘)	
	- Criminal Damage(재물 손괴)	
	- Gambling(불법 도박)	
	- Liquor Law Violation(주류법 위반)	
	SocioEconomic Indicators	(Float Array)
	- Percent Households below Poverty(빈곤 가구 비율)	
	- Percent of Housing Crowded(밀집 거주 가옥 비율)	
	- Hardship index(곤궁도 지수)	
	- Per Capita Income(1인당 소득)	
	- Percent Aged 18- or 64+(18세 이하 64세 이상 인구 비율)	
	- Percent Aged 25+, No High School Diploma(고등학교 졸업 미만 학력, 25세 이상)	
	- Percent Aged 16+ Unemployed(16세 이상 미취업 인구)	
	Existence of Police Station	(Binary)
	Number of Liquor Shop	(Integer)
	Number of Vacant Building	(Integer)
	Score of Abandoned Vehicle	(Integer)
}		

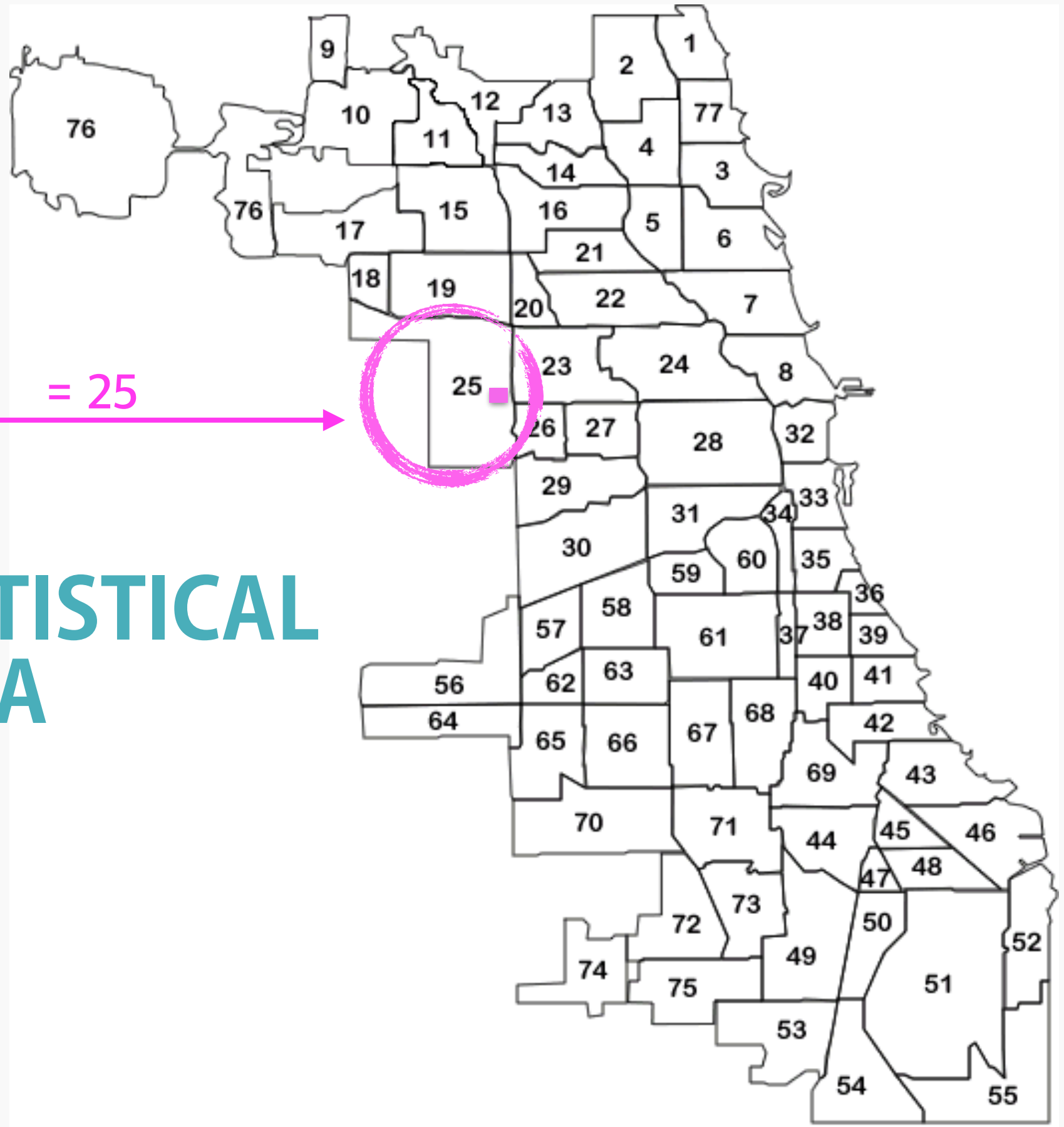
# DATA: PREDICTOR

## SocioEco Indicators

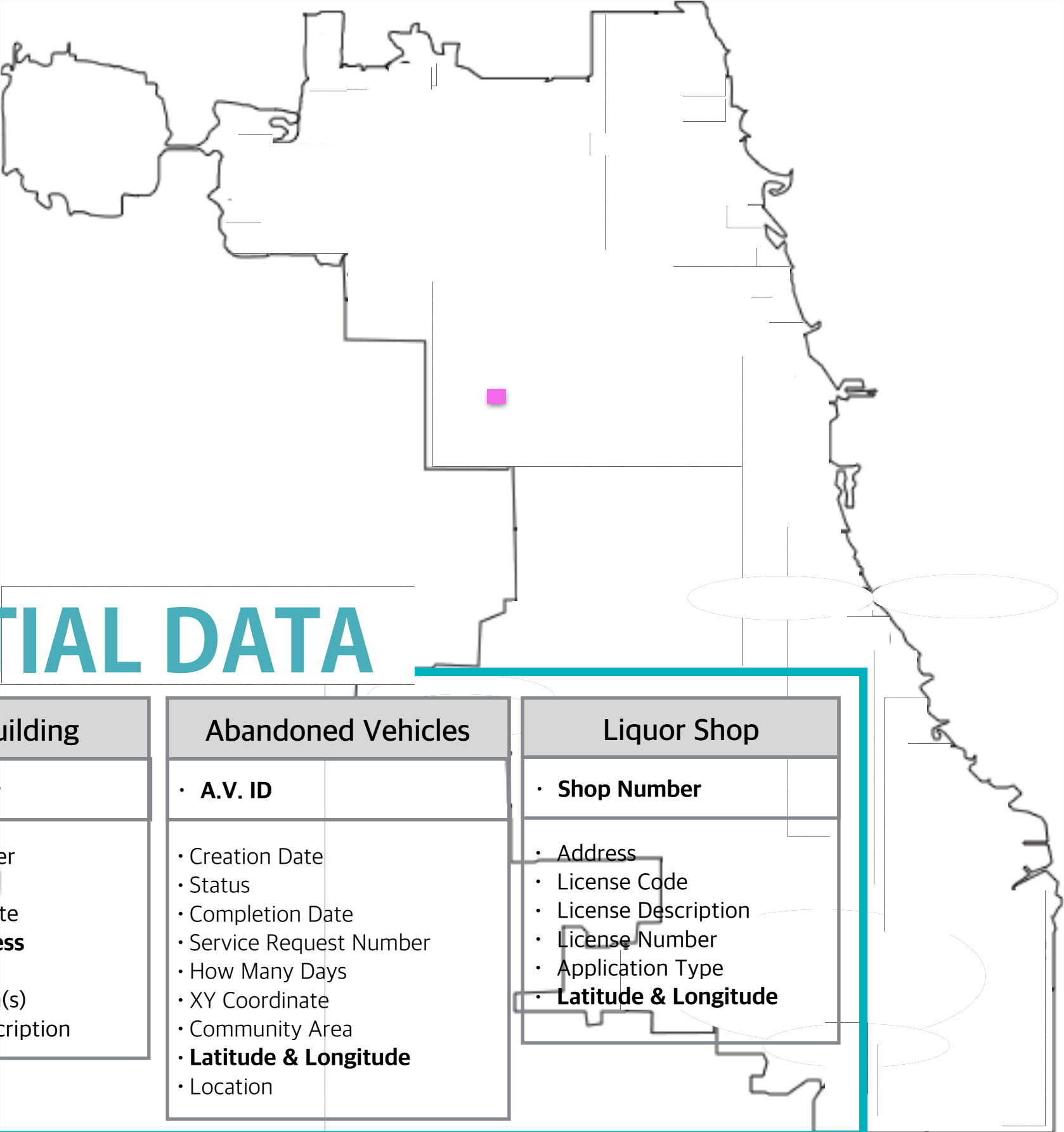
### • Community Area Number

- Community Area Name
- % Housing Crowded
- % Households Poverty
- % Aged 16+ Unemployed
- % Aged 25+ No High School Diploma
- % Aged -18 or 64+
- % Capital Income
- Hardship Index

# STATISTICAL DATA



# DATA: PREDICTOR



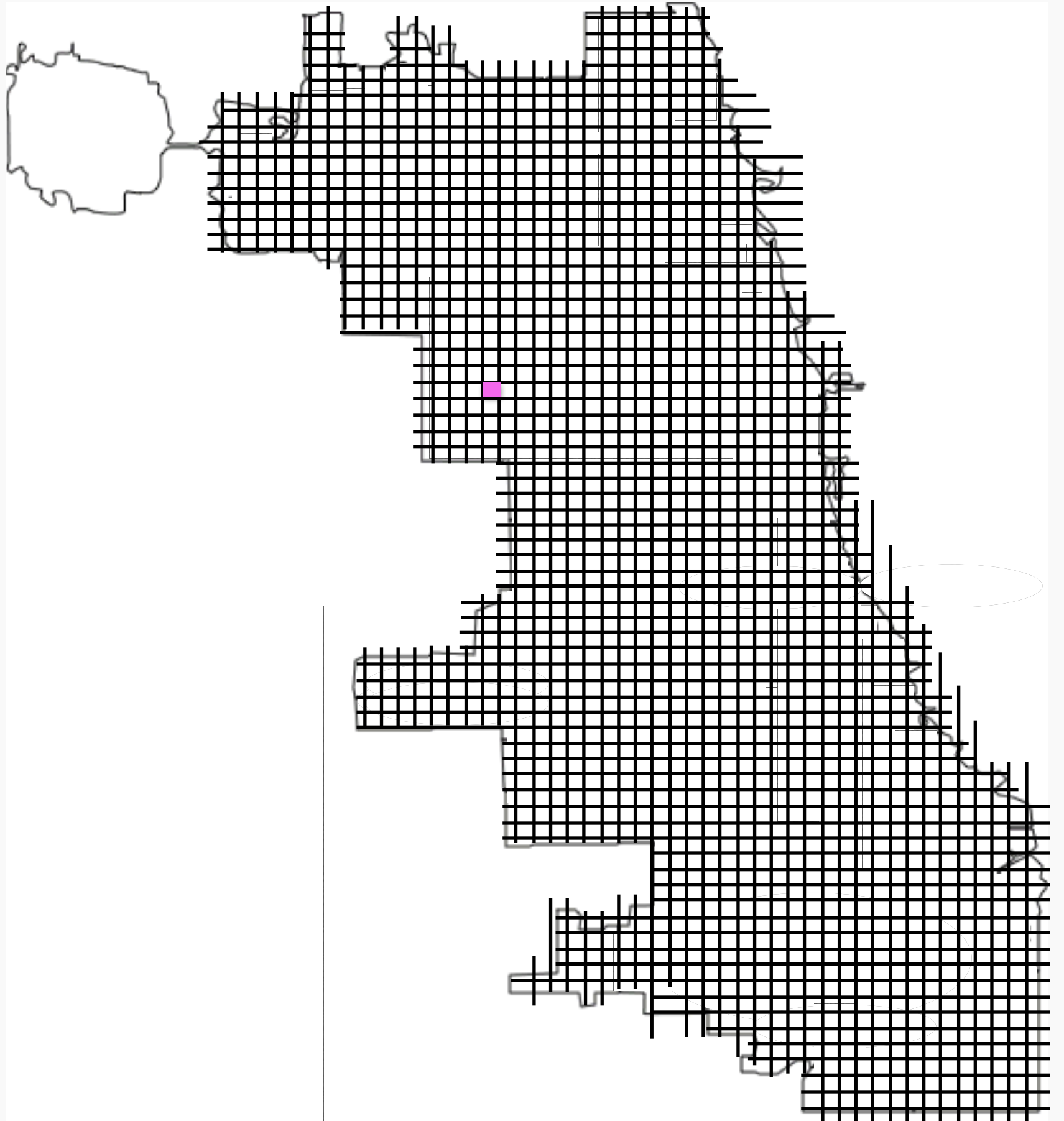
## SPATIAL DATA

Police Station	Vacant Building	Abandoned Vehicles	Liquor Shop
<ul style="list-style-type: none"><li>• Office Code</li></ul>	<ul style="list-style-type: none"><li>• Docket Number</li></ul>	<ul style="list-style-type: none"><li>• A.V. ID</li></ul>	<ul style="list-style-type: none"><li>• Shop Number</li></ul>
<ul style="list-style-type: none"><li>• Office Name</li><li>• <b>Latitude &amp; Longitude</b></li></ul>	<ul style="list-style-type: none"><li>• Violation Number</li><li>• Issued Date</li><li>• Last Hearing Date</li><li>• <b>Property Address</b></li><li>• Violation Type</li><li>• Entity or Person(s)</li><li>• Disposition Description</li></ul>	<ul style="list-style-type: none"><li>• Creation Date</li><li>• Status</li><li>• Completion Date</li><li>• Service Request Number</li><li>• How Many Days</li><li>• XY Coordinate</li><li>• Community Area</li><li>• <b>Latitude &amp; Longitude</b></li><li>• Location</li></ul>	<ul style="list-style-type: none"><li>• Address</li><li>• License Code</li><li>• License Description</li><li>• License Number</li><li>• Application Type</li><li>• <b>Latitude &amp; Longitude</b></li></ul>

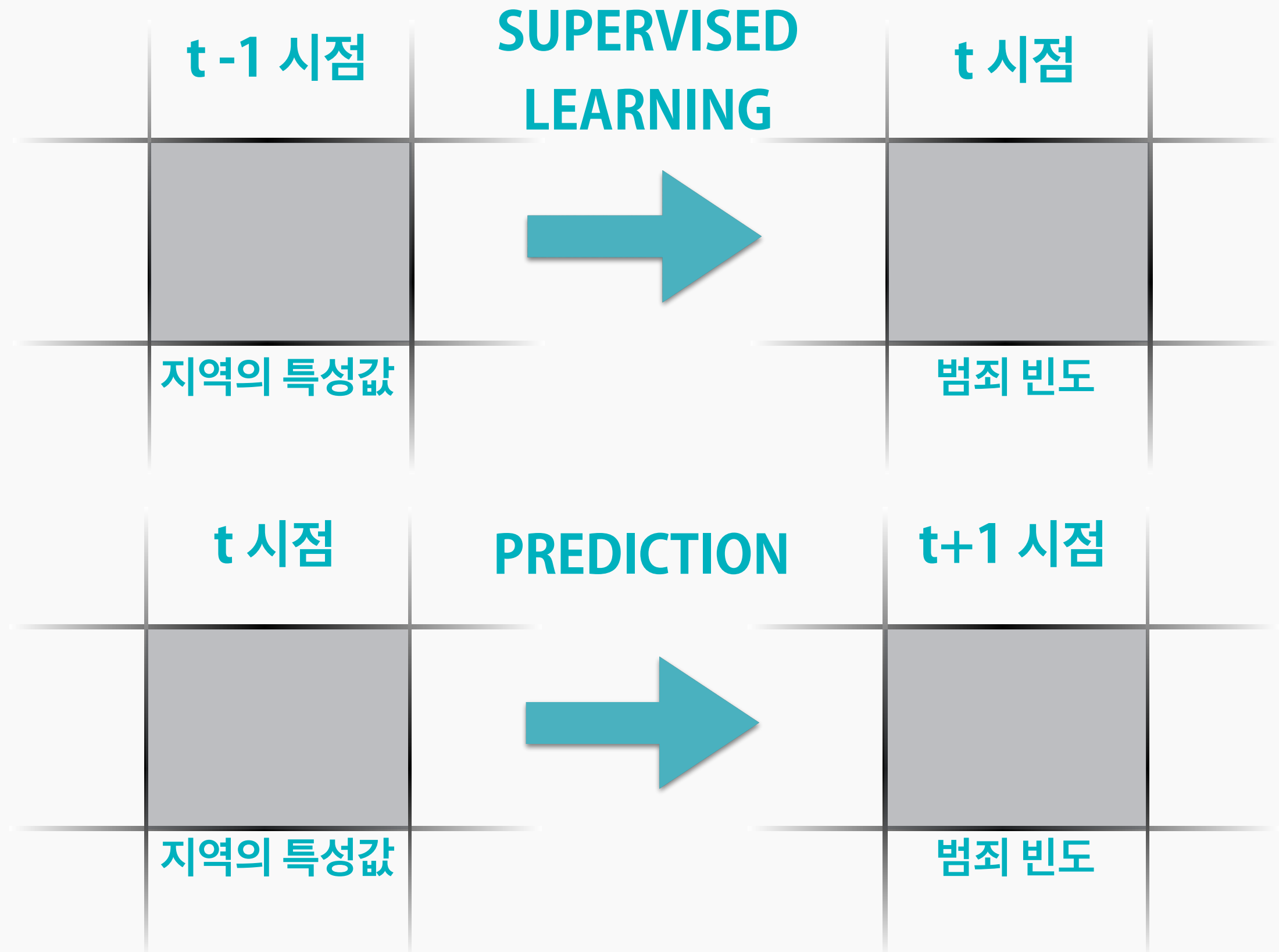
**DATA: MAP**  
**SPATIAL DIVISION:**  
**0.005 x 0.005 CELL**  
**About 2000 CELLS**

해당구역에서는...

1. 어떤 범죄가  
몇건이나 일어났는가?
2. 사회 통계치는  
얼마나 되는가?
3. 공간적인 특성은  
어떻게 되는가?



# MINING PROCESS



# MINING PROCESS

1월 자료 + 2월 범죄 빈도: LEARNING

2월 자료 + 3월 범죄 빈도: PREDICTION TEST

⋮

10월 자료 + 11월 범죄 빈도: LEARNING

11월 자료 + 12월 범죄 빈도: PREDICTION TEST

} 총 10회

( LEARNING METHOD - RESPONSE )

: MLR - Liquor Law Violation(주류법 위반)

: DECISION TREE - Gambling(불법 도박)

: NEURAL NETWORK - Criminal Damage(재물 손괴)



# MLR, LIQUOR LAW VIOLATION

## Regsubsets, “exhaustive”의 결과

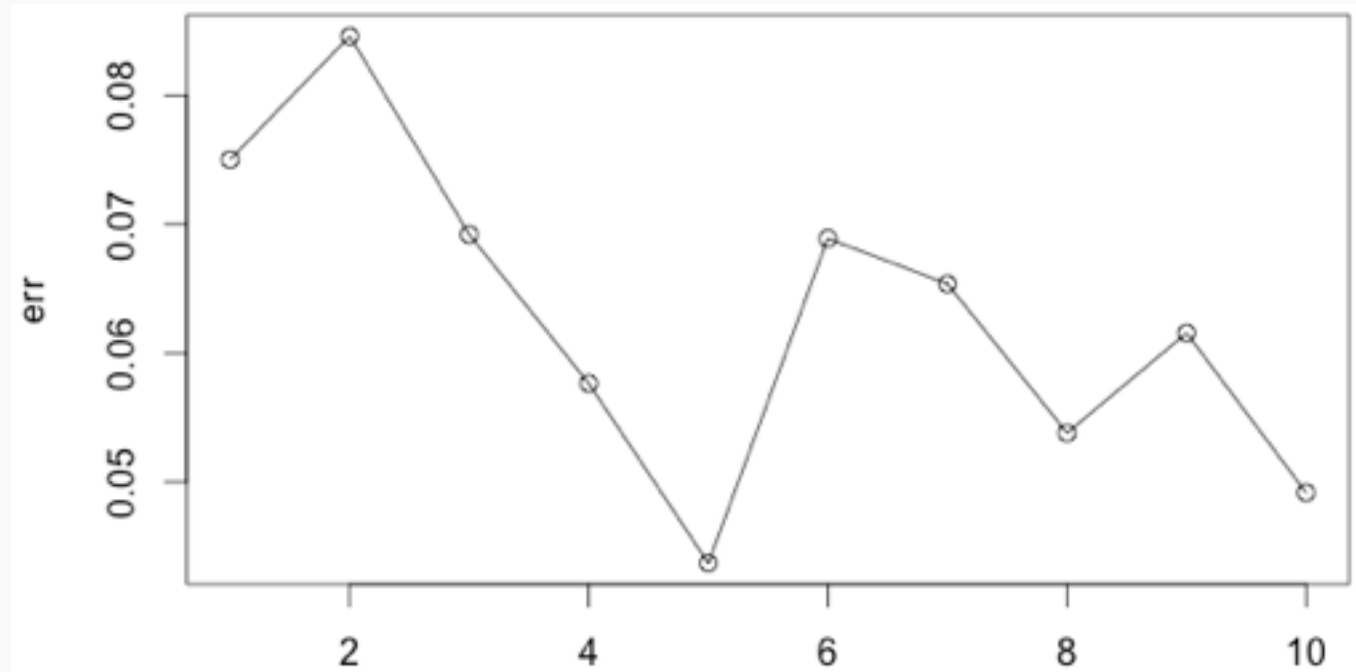
Selection Algorithm: exhaustive

		ALCOHOL	POLICE	BUILDING	HC.	HP.	U16.	W25.	X18.64	PCI	HI	CRIMINAL.DAMAGE	PROSTITUTION
1	( 1 )	"*"	" "	" "	" "	" "	" "	" "	" "	" "	" "	" "	" "
2	( 1 )	"*"	" "	" "	" "	" "	" "	" "	" "	" "	" "	" "	"*"
3	( 1 )	"*"	" "	" "	" "	" "	" "	" "	" "	" "	" "	" "	"*"
4	( 1 )	"*"	" "	" "	" "	" "	" "	" "	"*"	" "	" "	" "	"*"

**ALCOHOL(지역 내 주류 판매점 수): 4번**

**LIQUOR.LAW.VIOLATION(전 기의 지역 내 주류법 위반 건수): 2회**

**RMSE**



**PERIOD**

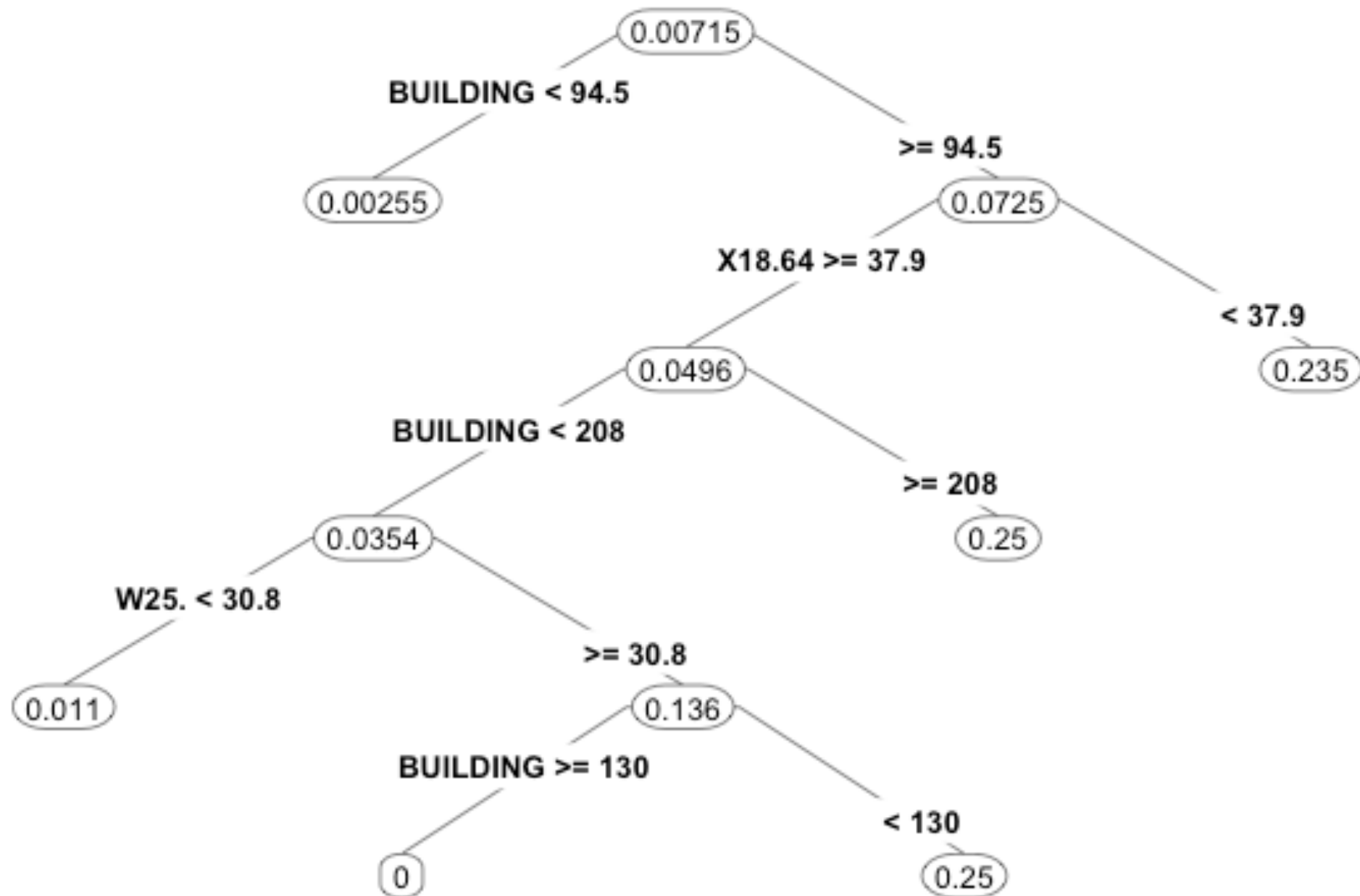
# MLR, LIQUOR LAW VIOLATION

2월 주류법 위반

MLR의 1월 데이터 기반 예측

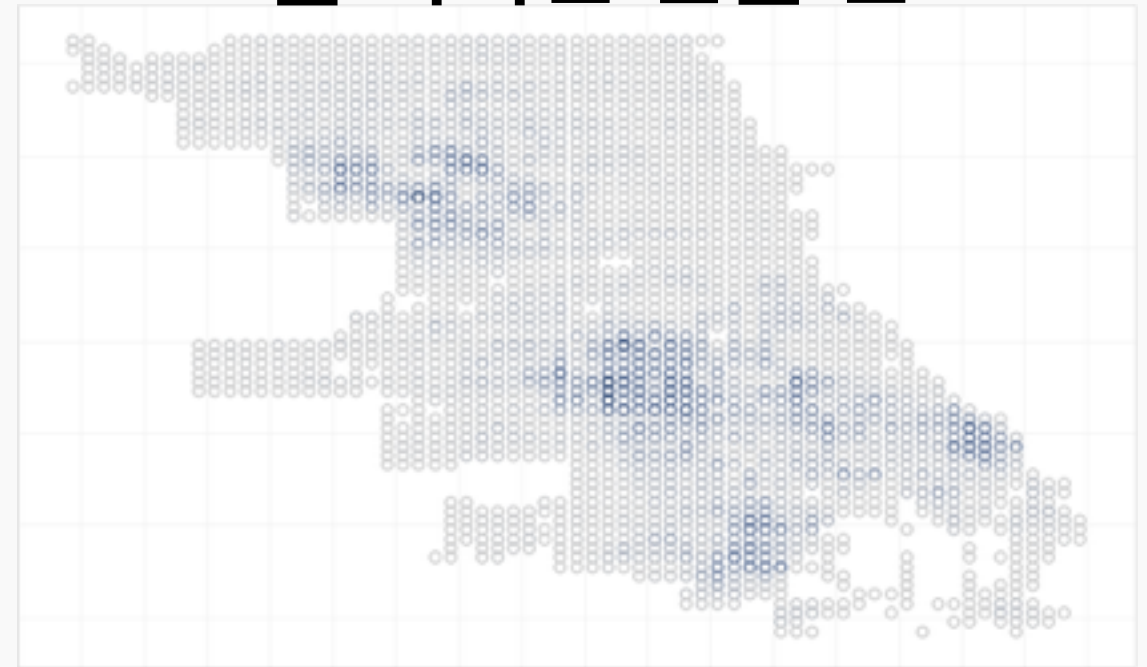
# DECISION TREE, GAMBLING

4월 데이터를 근거로한 RPART

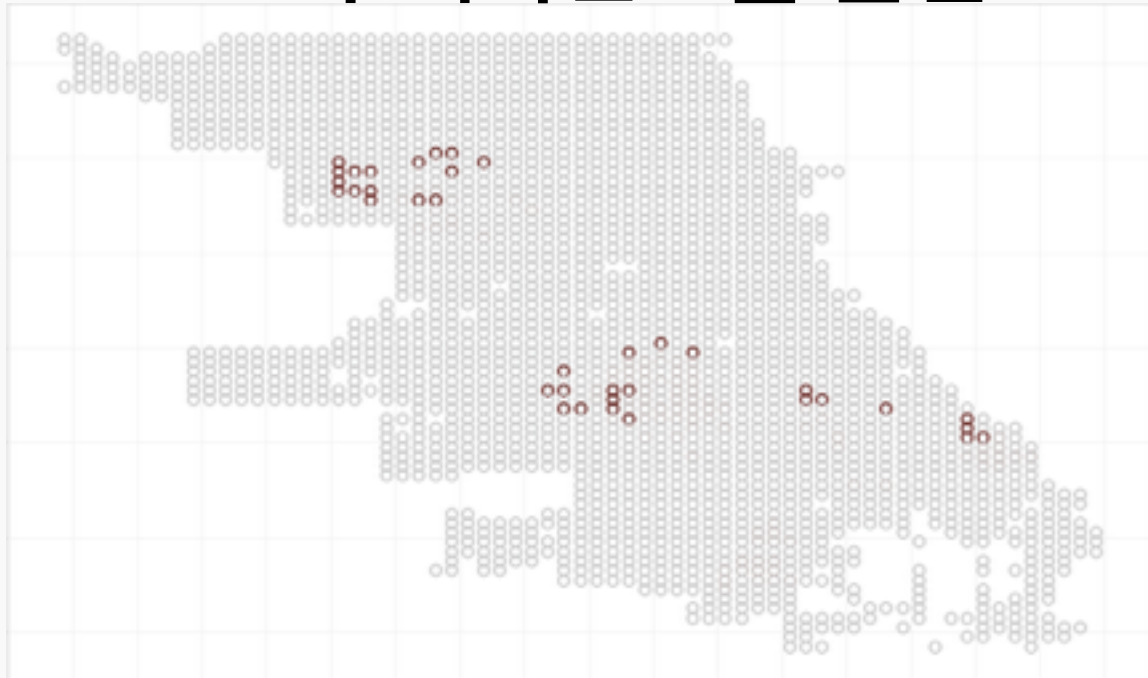


# DECISION TREE, GAMBLING

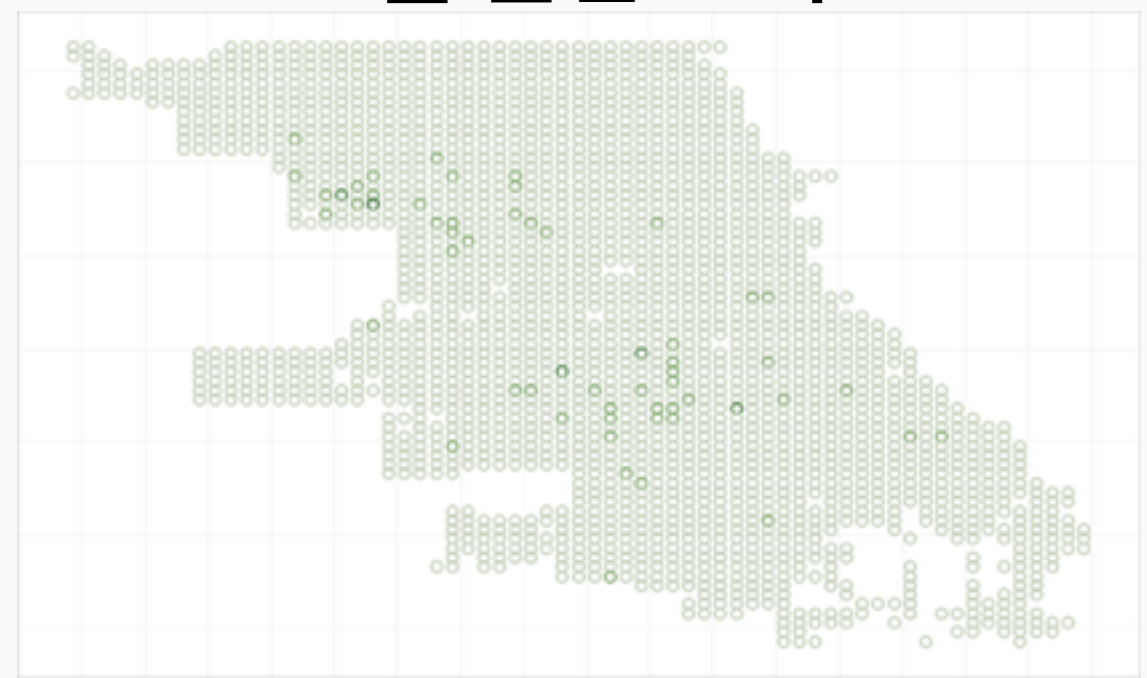
4월 버려진 건물 분포



4월 데이터를 근거로한  
RPART 가 예측한 5월 불법 도박

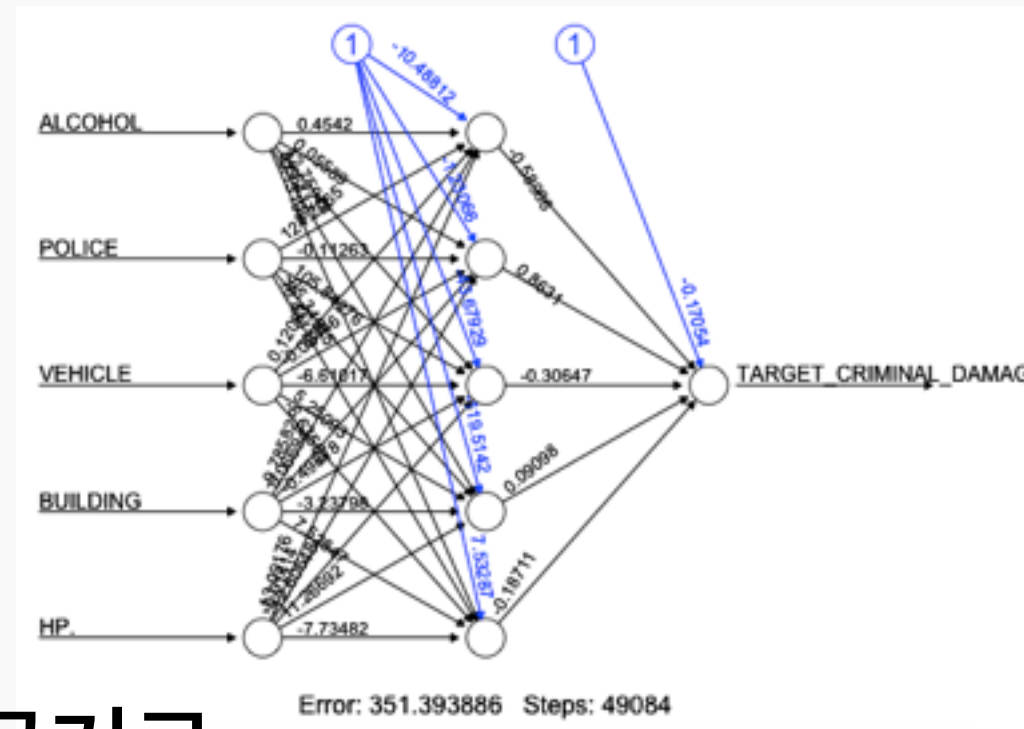


5월 불법 도박





# NEURAL NET, CRIMINAL DAMAGE



2월 데이터를 근거로  
인공신경망이 예측한 3월 불법 도박



3월 불법 도박



# PROBLEMS

1. CELL SIZE 가 작아질수록, 예측이 어려워진다.  
PERIOD LENGTH 가 짧아질수록, 예측이 어려워진다.
2. 기간별 모형이 크게 달라지면서, 모형의 의미를 해석하기 어렵다.  
(MLR, Decision Tree의 경우)
3. 성능 비교를 위한 예측 기법 선정 문제.
4. 예측 성능 측정을 위한 새로운 방법.



---

**THANK YOU**

---