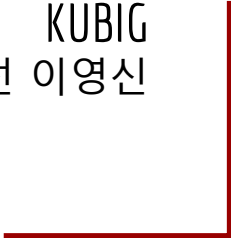




# CHICAGO CRIME DATA

KUBIG  
박소현 김효익 조송현 조규선 이영신



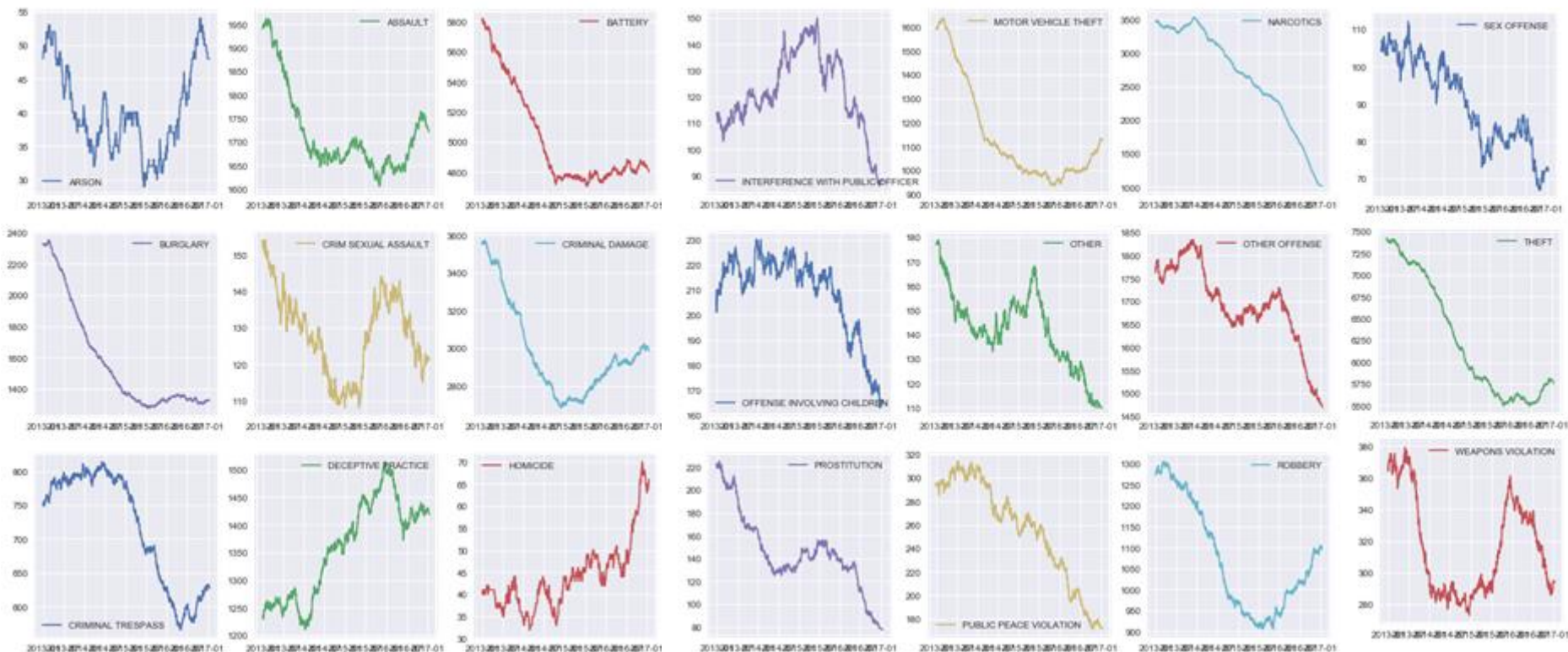
# Visualization(EDA)

# Visualization(EDA)

---

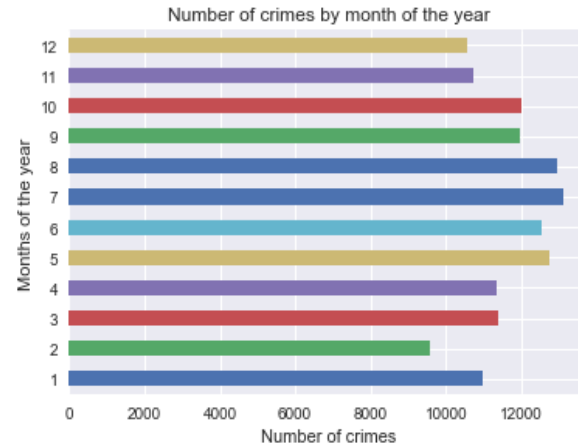
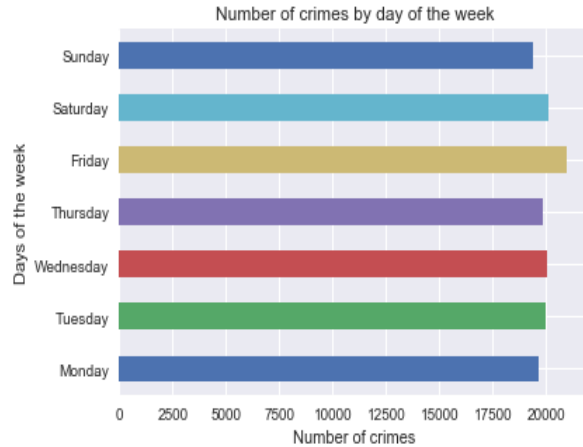


# Visualization(EDA)

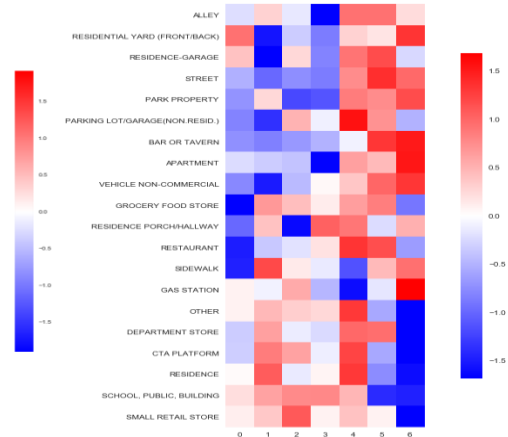
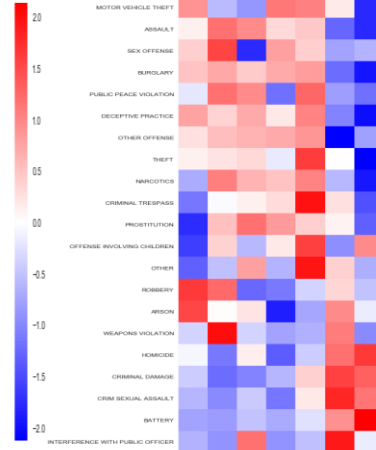
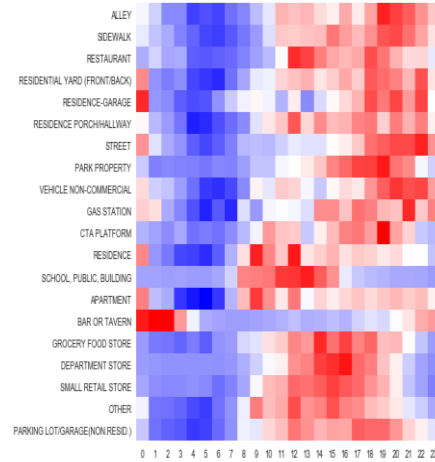
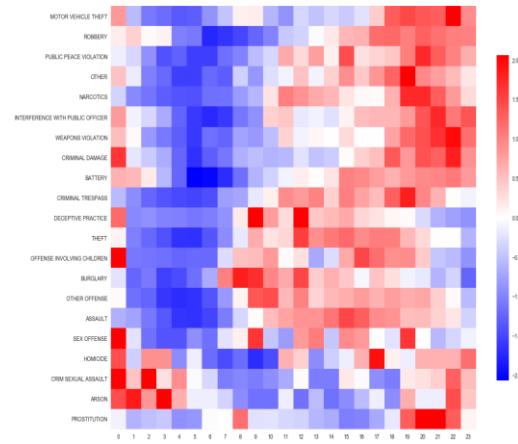


# Visualization(EDA)

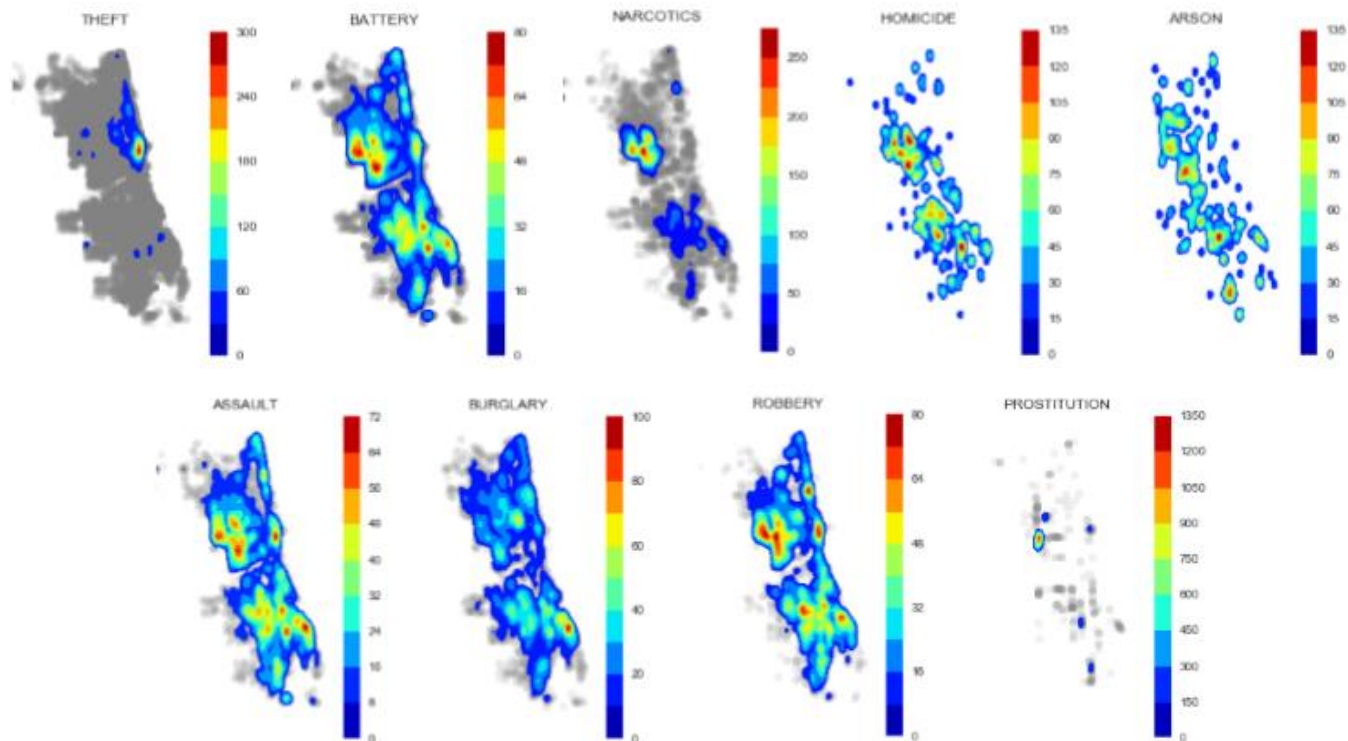
---



# Visualization(EDA)



# Visualization(EDA)



# Modeling



# Modeling

---

```
> table(crime$Primary.Type)
```

ARSON	209	ASSAULT	8729
BATTERY	25476	BURGLARY	8129
CONCEALED CARRY LICENSE VIOLATION	8	CRIM SEXUAL ASSAULT	660
CRIMINAL DAMAGE	15242	CRIMINAL TRESPASS	3590
DECEPTIVE PRACTICE	6770	GAMBLING	206
HOMICIDE	229	INTERFERENCE WITH PUBLIC OFFICER	590
INTIMIDATION	76	KIDNAPPING	104
LIQUOR LAW VIOLATION	193	MOTOR VEHICLE THEFT	5948
NARCOTICS	12962	NON - CRIMINAL	10
OBSCENITY	22	OFFENSE INVOLVING CHILDREN	1027
OTHER NARCOTIC VIOLATION	3	OTHER OFFENSE	8466
PROSTITUTION	739	PUBLIC INDECENCY	8
PUBLIC PEACE VIOLATION	1254	ROBBERY	5503
SEX OFFENSE	454	STALKING	79
THEFT	31719	WEAPONS VIOLATION	1595

# Modeling

---

Category	Type of Crime
Violent	"ASSAULT", "BATTERY", "HOMICIDE", "INTIMIDATION", "KIDNAPPING", "CONCEALED CARRY LICENSE VIOLATION", "WEAPONS VIOLATION"
Property	"ARSON", "BURGLARY", "CRIMINAL DAMAGE", "CRIMINAL TRESPASS", "MOTOR VEHICLE THEFT", "ROBBERY", "THEFT"
Sex	"CRIM SEXUAL ASSAULT", "OFFENSE INVOLVING CHILDREN", "PROSTITUTION", "SEX OFFENSE", "STALKING"
White	"DECEPTIVE PRACTICE", "GAMBLING"
Public	"INTERFERENCE WITH PUBLIC OFFICER", "OBSCENITY", "PUBLIC INDECENCY", "PUBLIC PEACE VIOLATION"
Drug	"LIQUOR LAW VIOLATION", "NARCOTICS", "OTHER NARCOTIC VIOLATION"
Others	"NON - CRIMINAL", "NON - CRIMINAL", "OTHER OFFENSE"

# Modeling

```
> violent <- c("ASSAULT", "BATTERY", "HOMICIDE", "INTIMIDATION", "KIDNAPPING", "CONCEALED CARRY LICENSE VIOLATION", "WEAPONS VIOLATION")
> property <- c("ARSON", "BURGLARY", "CRIMINAL DAMAGE", "CRIMINAL TRESPASS", "MOTOR VEHICLE THEFT", "ROBBERY", "THEFT")
> sex <- c("CRIM SEXUAL ASSAULT", "OFFENSE INVOLVING CHILDREN", "PROSTITUTION", "SEX OFFENSE", "STALKING")
> white <- c("DECEPTIVE PRACTICE", "GAMBLING")
> public <- c("INTERFERENCE WITH PUBLIC OFFICER", "OBSCENITY", "PUBLIC INDECENCY", "PUBLIC PEACE VIOLATION")
> drug <- c("LIQUOR LAW VIOLATION", "NARCOTICS", "OTHER NARCOTIC VIOLATION")
> others <- c("NON - CRIMINAL", "NON - CRIMINAL", "OTHER OFFENSE")
>
> for(i in 1:nrow(crime)){
+ if(crime$Primary.Type[i] %in% violent){
+ crime$Primary.Type[i] <- "violent"
+ }
+ if(crime$Primary.Type[i] %in% property){
+ crime$Primary.Type[i] <- "property"
+ }
+ if(crime$Primary.Type[i] %in% sex){
+ crime$Primary.Type[i] <- "sex"
+ }
+ if(crime$Primary.Type[i] %in% white){
+ crime$Primary.Type[i] <- "white"
+ }
+ if(crime$Primary.Type[i] %in% public){
+ crime$Primary.Type[i] <- "public"
+ }
+ if(crime$Primary.Type[i] %in% drug){
+ crime$Primary.Type[i] <- "drug"
+ }
+ if(crime$Primary.Type[i] %in% others){
+ crime$Primary.Type[i] <- "others"
+ }
+ }
```


> table(crime\$Primary.Type)

drug	others	property	public	sex	violent	white
13158	8476	70297	1874	2959	36216	6906

# Modeling

```
> crime$newdate <- as.POSIXct(crime$Date, format = "%m/%d/%Y %I:%M:%S %p")
> library(chron)
> crime$time <- times(format(crime$newdate, "%H:%M:%S"))
> time.tag <- chron(times=c("00:00:00", "06:00:00", "12:00:00", "18:00:00", "23:59:00"))
> crime$time.tag <- cut(crime$time, breaks=time.tag, labels=c("00-06", "06-12", "12-18",
+ "18-24"), include.lowest=T)

> crime[c("Primary.Type", "Date")]
# A tibble: 139,886 x 2
  Primary.Type Date
  <chr>         <chr>
1 drug        10/04/2013 09:05:00 AM
2 others      02/29/2012 05:33:00 PM
3 property    12/19/2013 04:30:00 PM
4 violent     09/15/2013 08:10:00 PM
5 property    09/22/2012 08:00:00 AM
6 violent     04/15/2013 10:15:00 PM
7 property    02/08/2012 09:30:00 AM
8 property    03/05/2016 12:00:00 PM
9 property    07/12/2015 01:15:00 PM
10 others     03/22/2012 04:19:00 AM
# ... with 139,876 more rows
```



```
> crime[c("Primary.Type", "time", "time.tag")]
# A tibble: 139,886 x 3
  Primary.Type time      time.tag
  <chr>         <times> <fct>
1 drug        09:05:00 06-12
2 others      17:33:00 12-18
3 property    16:30:00 12-18
4 violent     20:10:00 18-24
5 property    08:00:00 06-12
6 violent     22:15:00 18-24
7 property    09:30:00 06-12
8 property    12:00:00 06-12
9 property    13:15:00 12-18
10 others     04:19:00 00-06
# ... with 139,876 more rows
```

# Modeling

```
> crime$newdate1 <- as.POSIXlt(strptime(crime$newdate, format = "%Y-%m-%d"))
> crime$day <- weekdays(crime$newdate1, abbreviate=T)
> crime$newdate1 <- as.POSIXlt(strptime(crime$newdate, format = "%Y-%m-%d"))
> crime$month <- months(crime$newdate1, abbreviate=T)
```

```
> crime[c("Primary.Type", "Date")]
# A tibble: 139,886 x 2
  Primary.Type Date
  <chr>        <chr>
1 drug        10/04/2013 09:05:00 AM
2 others      02/29/2012 05:33:00 PM
3 property    12/19/2013 04:30:00 PM
4 violent     09/15/2013 08:10:00 PM
5 property    09/22/2012 08:00:00 AM
6 violent     04/15/2013 10:15:00 PM
7 property    02/08/2012 09:30:00 AM
8 property    03/05/2016 12:00:00 PM
9 property    07/12/2015 01:15:00 PM
10 others     03/22/2012 04:19:00 AM
# ... with 139,876 more rows
```



```
> crime[c("Primary.Type", "day", "month")]
# A tibble: 139,886 x 3
  Primary.Type day month
  <chr>        <chr> <chr>
1 drug        Fri  Oct
2 others      Wed  Feb
3 property    Thu  Dec
4 violent     Sun  Sep
5 property    Sat  Sep
6 violent     Mon  Apr
7 property    Wed  Feb
8 property    Sat  Mar
9 property    Sun  Jul
10 others     Thu  Mar
# ... with 139,876 more rows
```

# Modeling

---

- 36개의 범죄 종류를 7개로 분류
- 4개의 시간대(00-06, 06-12, 12-18, 18-24)로 분류하는 time.tag 변수 추가
- 요일을 구분한 day 변수 추가
- 달을 구분한 month 변수 추가

# Modeling - Decision Tree

---

```
for(i in 1:nrow(crime)){  
  if(crime$time.tag[i] == "00-06"){  
    crime$time.tag[i] <- 1  
  }  
  if(crime$time.tag[i] == "06-12"){  
    crime$time.tag[i] <- 2  
  }  
  if(crime$time.tag[i] == "12-18"){  
    crime$time.tag[i] <- 3  
  }  
  if(crime$time.tag[i] == "18-24"){  
    crime$time.tag[i] <- 4  
  }  
}  
crime$time.tag<-as.numeric(crime$time.tag)
```

```
set.seed(100)  
train.index<-sample(nrow(crime),139886*0.7)  
train<-crime[train.index,]  
test<-crime[-train.index,]  
crime_ctree1<-ctree(Arrest~Domestic+Year+time.tag,data=train)  
crime_ctree2<-ctree(Arrest~Year+time.tag,data=train)  
crime_ctree3<-ctree(Domestic~Arrest+Year+time.tag,data=train)  
crime_ctree4<-ctree(Domestic~Year+time.tag,data=train)
```

00-06 -> 1

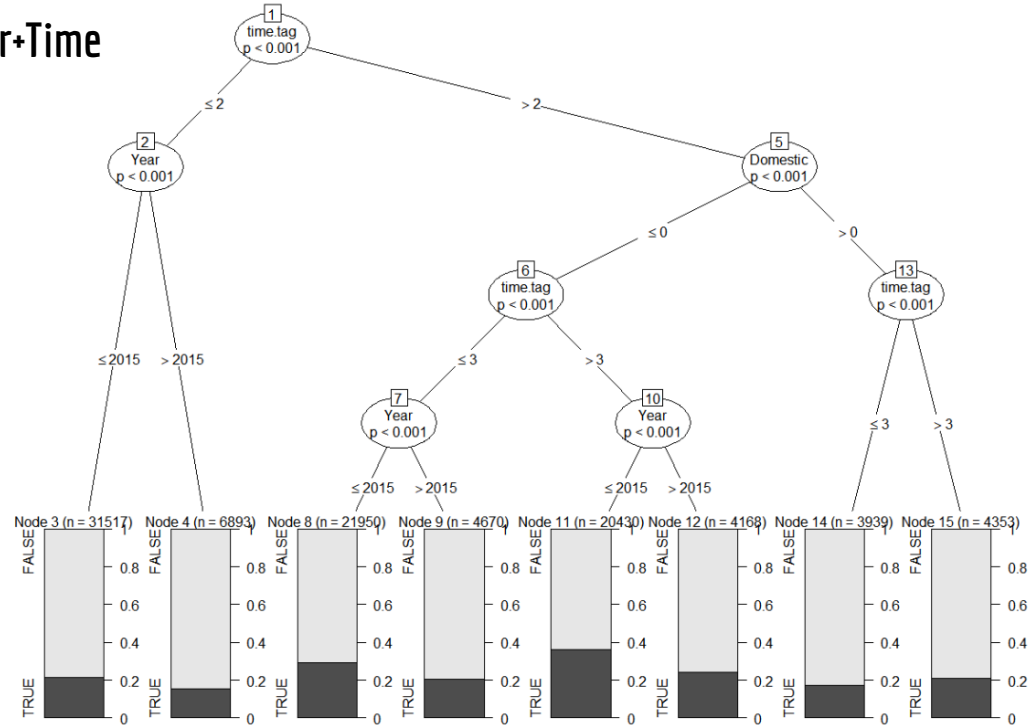
06-12 -> 2

12-18 -> 3

18-24 -> 4

# Modeling - Decision Tree

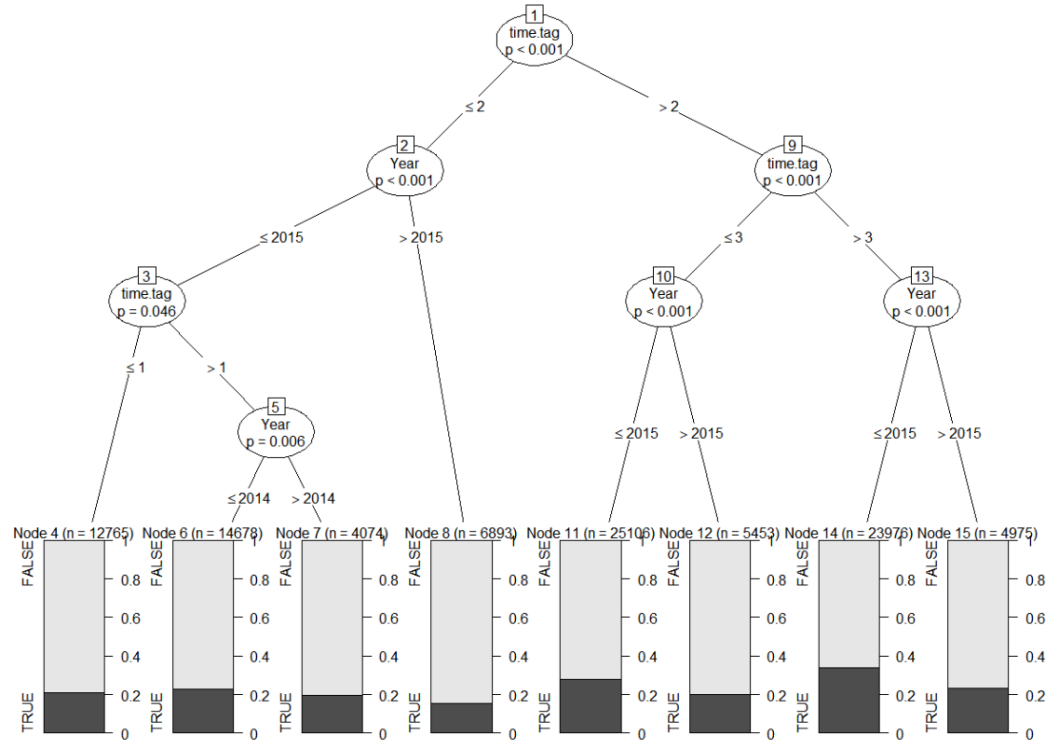
## 1. Arrest - Domestic+Year+Time





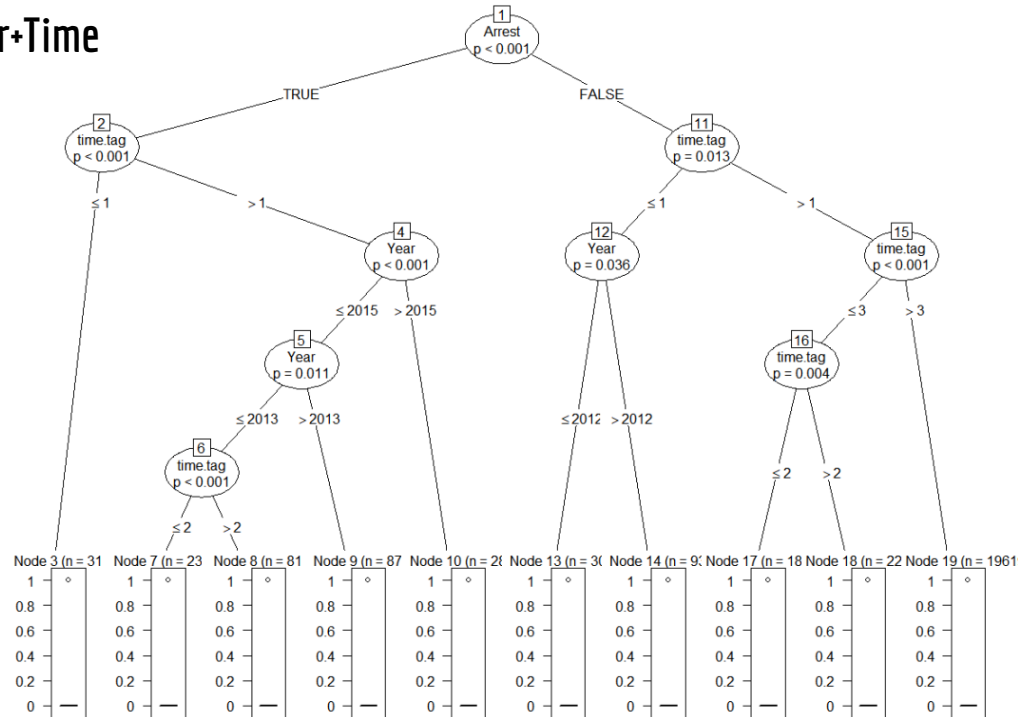
# Modeling - Decision Tree

## 2. Arrest - Year+Time



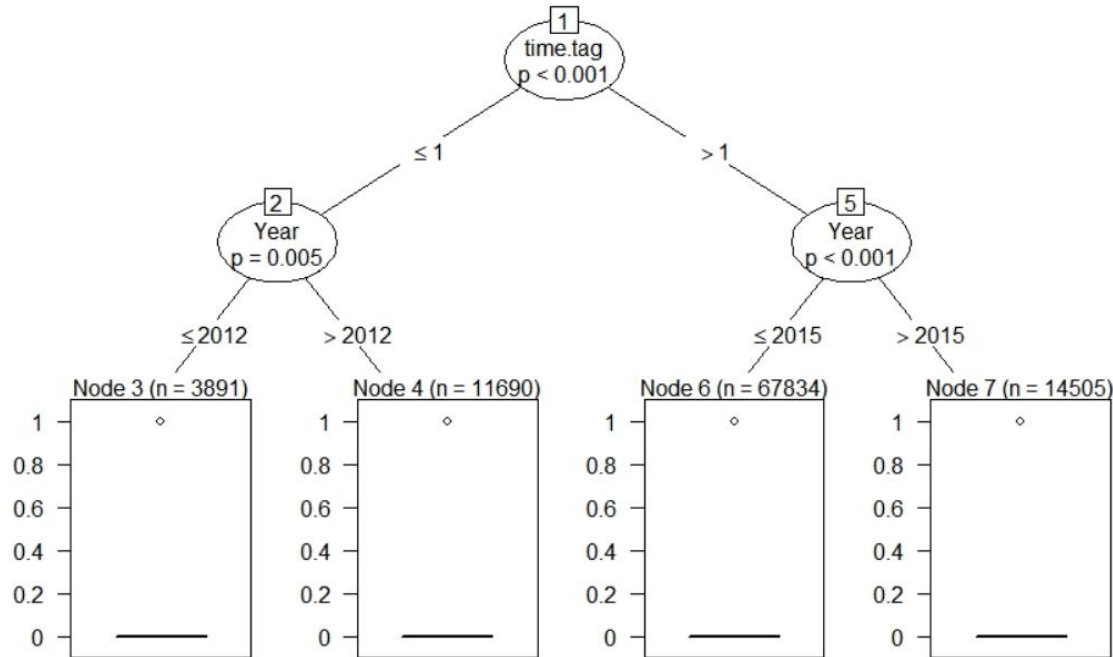
# Modeling - Decision Tree

## 3. Domestic - Arrest+Year+Time



# Modeling - Decision Tree

## 4. Domestic - Year+Time



# Modeling - 범죄발생가능성 지수

## 1. 범죄유형 - 재범률에 따른 점수 부여

Drug, Property <- 5

Others, Public, White, Violent <- 4

Sex <- 2

```
chicago$type_score<-0+
↓
for(i in 1:nrow(chicago)){
  if(chicago$Primary.Type[i]=="drug"|chicago$Primary.Type[i]=="property"){+
    chicago$type_score[i]<-5+
  }+
  if(chicago$Primary.Type[i]=="others"|chicago$Primary.Type[i]=="public"|chicago$Primary.Type[i]=="white"|chicago$Primary.Type[i]=="violent"){+
    chicago$type_score[i]<-4+
  }+
  if(chicago$Primary.Type[i]=="sex"){+
    chicago$type_score[i]<-2+
  }+
}
```

수행자 특징	출소자의 비율	출소 후 3년 내 재범률			
		재범률	재유죄판결	새로운 징역형으로 인한 재범률	전체 재범률
총 출소자	100	67.5	46.9	25.4	51.8
폭력 범죄	22.5	61.7	39.9	20.4	48.8
살인	1.7	40.7	20.5	10.8	31.4
납치	0.4	59.4	37.8	25.1	29.5
강간	1.2	46.0	27.4	12.6	43.5
다른 성범죄	2.4	41.4	22.3	10.5	36.0
강도	9.9	70.2	43.5	25.0	54.7
폭행	6.5	66.1	44.2	21.0	51.2
가난 폭력범죄	0.4	51.7	29.8	12.7	40.9
재산범죄	33.5	73.8	53.4	30.5	56.4
주거침입 절도	15.2	74.0	54.2	30.8	56.1
단순 절도	9.7	74.6	55.7	32.6	60.1
차량 절도	3.5	78.8	54.3	31.3	58.1
방화	0.5	57.7	41.0	20.1	38.7
사기	2.9	66.3	42.1	22.8	45.4
장물	1.4	77.4	57.2	31.8	62.1
가난 재산범죄	0.3	71.1	47.6	28.5	40.0
약물 범죄	32.6	66.7	47.0	25.2	49.2
약물 소지	7.5	67.5	46.6	23.9	43.6
마약 밀매	20.2	64.2	44.0	24.8	46.1
가난/불특정	4.9	75.5	60.5	28.8	71.8
공공질서침해	9.7	62.2	42.0	21.6	48.0
무기	3.1	70.2	46.6	24.3	55.5
음주 운전	3.3	51.5	51.7	16.6	43.7
가난 공공질서침해	3.3	65.1	48.0	24.4	43.6
가난 범죄	1.7	64.7	42.1	20.7	66.9

# Modeling - 범죄발생가능성 지수

## 2. 범죄발생장소 - 빈도에 따른 점수 부여

0.2 이상 <- 5                      0.02 이상 0.1 미만 <- 3                      0.01 미만 <- 1  
0.1 이상 0.2 미만 <- 4            0.01 이상 0.02 미만 <- 2

```
chicago$location_score<-0↓  
y<-summary(chicago$Location.Description)/nrow(chicago)↓  
↓  
for(i in 1:nrow(chicago)){↓  
  x<-chicago$Location.Description[i]↓  
  prob<-unname(y[which(x==names(y))])↓  
  if(length(prob)>=1){↓  
    chicago$location_score[i]<-ifelse(prob>=0.2,5,ifelse(prob>=0.1,4,ifelse(p  
rob>=0.02,3,ifelse(prob>=0.01,2,1))))↓  
  } else {↓  
    chicago$location_score[i]<-1↓  
  }↓  
}↓
```

# Modeling - 범죄발생가능성 지수

---

## 3. 가정범죄 여부에 따른 점수 부여

가정범죄 <- 5

가정범죄 이외의 범죄 유형 <- 0

```
chicago$domestic_score<-0  
↓  
for(i in 1:nrow(chicago)){  
  chicago$domestic_score[i]<-ifelse(chicago$Domestic[i]==TRUE,5,0)  
}
```

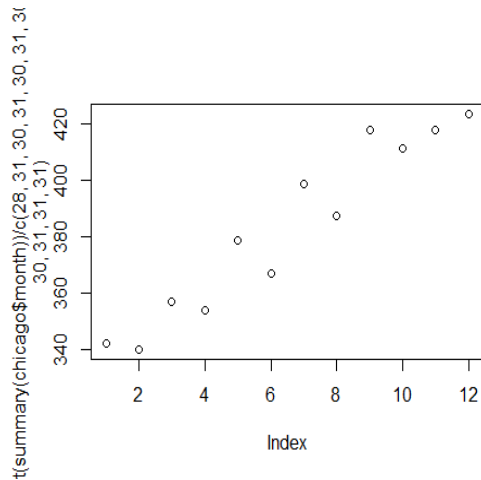
# Modeling - 범죄발생가능성 지수

## 4. 범죄 발생 달 - 빈도에 따른 점수 부여

```
> sort(table(crime$month))
```

Feb	Dec	Nov	Jan	Apr	Mar	Sep	Oct	Jun	May	Aug	Jul
9587	10546	10712	10980	11361	11377	11958	12015	12535	12747	12943	13125

```
chicago$month_score<-0+
+
for(i in 1:nrow(chicago)){+
  if(chicago$month[i]=="Jul" | chicago$month[i]=="Jun" | chicago$month[i]=="Aug" |
chicago$month[i]=="May"){+
    chicago$month_score[i]<-5+
  }+
  if(chicago$month[i]=="Oct" | chicago$month[i]=="Sep"){+
    chicago$month_score[i]<-4+
  }+
  if(chicago$month[i]=="Mar" | chicago$month[i]=="Apr"){+
    chicago$month_score[i]<-3+
  }+
  if(chicago$month[i]=="Jan" | chicago$month[i]=="Nov"){+
    chicago$month_score[i]<-2+
  }+
  if(chicago$month[i]=="Feb" | chicago$month[i]=="Dec"){+
    chicago$month_score[i]<-1+
  }+
}
```



5월, 6월, 7월, 8월 <- 5

9월, 10월 <- 4

4월, 5월 <- 3

1월, 11월 <- 2

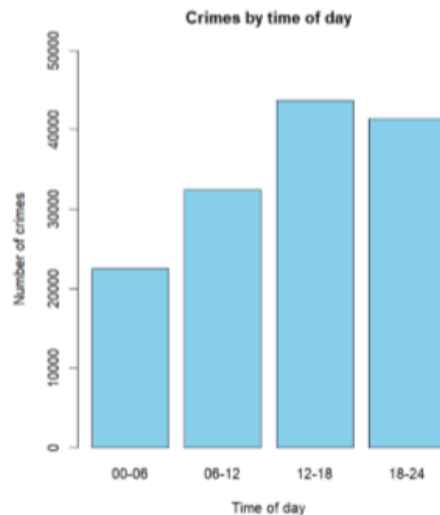
2월, 12월 <- 1

# Modeling - 범죄발생가능성 지수

## 5. 시간대 - 빈도에 따른 점수 부여

```
12-18 <- 5  
18-24 <- 4  
06-12 <- 3  
00-06 <- 2
```

```
chicago$time_score<-0  
↓  
for(i in 1:nrow(chicago)){  
  x<-chicago$time.tag[i]  
  chicago$time_score[i]<-ifelse(x=="12-18",4.5,ifelse(x=="18-24",4,ifelse(x=="06-12",3,ifelse(x=="00-06",2,0))))  
}↓
```





# Modeling - 범죄발생가능성 지수

---

## 최종 범죄 발생 가능성 지수

범죄유형: 0.4

범죄 발생 장소: 0.25

가정 범죄 여부: 0.1

범죄 발생 달: 0.1

시간대: 0.15

```
chicago$crime_score<-0.4*chicago$type_score+0.25*chicago$location_score+0.1*c  
hicago$domestic_score+0.1*chicago$month_score+0.15*chicago$time_score↓
```

# Modeling - 범죄발생가능성 지수

## 최종 범죄 발생 가능성 지수

- 최종 범죄 발생 가능성 지수가 가장 높은 사례 추출

```
chicago[which.max(chicago$crime_score),]
```

```
##      X.1  X1  X1_1      X      ID Case.Number      Date↓
## 1454 1454 1454 200719 2534413 8626214      HV299466 05/23/2012 01:00:00 PM↓
##                               Block IUCR Primary.Type Description↓
## 1454 116XX S VINCENNES AVE 1320      property  TO VEHICLE↓
##      Location.Description Arrest Domestic Beat District Ward↓
## 1454      STREET FALSE      TRUE 2212      22  34↓
##      Community.Area FBI.Code X.Coordinate Y.Coordinate Year↓
## 1454      75      14      1165794      1827755 2012↓
##                               Updated.On Latitude Longitude↓
## 1454 02/04/2016 06:33:39 AM 41.68294 -87.66872↓
##                               Location      newdate      time time.tag↓
## 1454 (41.682938571, -87.668723088) 2012-05-23 13:00:00 12-18↓
##                               newdate1 month day type_score location_score domestic_score↓
## 1454 2012-05-23 May Wed      5      5      5↓
##                               month_score time_score crime_score↓
## 1454      5      4.5      4.925↓
```

- 빈도수를 주로 활용하였기 때문에 빈도가 가장 큰 경우(ex. 재산범죄, 거리에서 범죄 발생)를 모두 포함한 범죄사건이 가장 높은 범죄발생 가능성 지수를 받음.

- 범죄 유형을 제외하고는 재발가능성을 고려하지 않았기 때문에 이러한 결과가 나온 것으로 보임

 범죄 관련 논문을 살펴볼 필요

# Modeling - 범죄발생가능성 지수

## 최종 범죄 발생 가능성 지수

- 달을 11월로 고정

```
##          X.1    X1    X1_1      X      ID Case.Number↓
## 15573 15573 15573 1164988 3501516 10314439  HY503595↓
##                               Date      Block IUCR Primary.Type↓
## 15573 11/16/2015 01:30:00 PM 100XX S MICHIGAN AVE 1320  property↓
##      Description Location.Description Arrest Domestic Beat District Ward↓
## 15573 TO VEHICLE                STREET FALSE TRUE 511      5      9↓
##      Community.Area FBI.Code X.Coordinate Y.Coordinate Year↓
## 15573              49      14      1179018      1838286 2015↓
##                               Updated.On Latitude Longitude↓
## 15573 11/23/2015 03:53:26 PM 41.71155      -87.62↓
##                               Location      newdate      time time.tag↓
## 15573 (41.711546981, -87.619995865) 2015-11-16 13:30 13:30:00 12-18↓
##      newdate1 month day type_score location_score domestic_score↓
## 15573 2015-11-16 Nov Mon      5      5      5↓
##      month_score time_score crime_score↓
## 15573      2      4.5      4.625↓
```

- 여전히 빈도가 높은 경우들을 모두 포함하는 사례가 추출됨.



**THANK YOU**